ISSN 2085-4579

# Ultima InfoSys

Jurnal Sistem Informasi

SI Sistem Informasi
Universitas Multimedia Nusantara



VOL 17 No.1

Engineering | Vol.17 | No.1 | Page 01-86 | June 2026 | ISSN 2085-4579

ISSN 2085-4579

# EDITORIAL BOARD



# EDITORIAL ADDRESS

Universitas Multimedia Nusantara (UMN)
Jl. Scientia Boulevard
Gading Serpong
Tangerang, Banten - 15811
Indonesia
Phone. (021) 5422 0808
Fax. (021) 5422 0800
Email : ultimainfosys@umn.ac.id


**Ultima InfoSys : Jurnal Ilmu Sistem Informasi** is a Journal of Information Systems which presents scientific research articles in the field of Information Systems, as well as the latest theoretical and practical issues, including database systems, management information systems, system analysis and development, system project management information, programming, mobile information system, and other topics related to Information Systems. ULTIMA InfoSys Journal is published regularly twice a year (June and December) by Faculty of Engineering and Informatics in cooperation with UMN Press.

# Call for Papers


**International Journal of New Media Technology (IJNMT)** is a scholarly open access, peer-reviewed, and interdisciplinary journal focusing on theories, methods and implementations of new media technology. Topics include, but not limited to digital technology for creative industry, infrastructure technology, computing communication and networking, signal and image processing, intelligent system, control and embedded system, mobile and web based system, and robotics. IJNMT is published twice a year by Faculty of Engineering and Informatics of Universitas Multimedia Nusantara in cooperation with UMN Press.


**Ultimatics : Jurnal Teknik Informatika** is the Journal of the Informatics Study Program at Universitas Multimedia Nusantara which presents scientific research articles in the fields of Analysis and Design of Algorithm, Software Engineering, System and Network security, as well as the latest theoretical and practical issues, including Ubiquitous and Mobile Computing, Artificial Intelligence and Machine Learning, Algorithm Theory, World Wide Web, Cryptography, as well as other topics in the field of Informatics.


**Ultima Computing : Jurnal Sistem Komputer** is a Journal of Computer Engineering Study Program, Universitas Multimedia Nusantara which presents scientific research articles in the field of Computer Engineering and Electrical Engineering as well as current theoretical and practical issues, including Edge Computing, Internet-of-Things, Embedded Systems, Robotics, Control System, Network and Communication, System Integration, as well as other topics in the field of Computer Engineering and Electrical Engineering.


**Ultima InfoSys : Jurnal Ilmu Sistem Informasi** is a Journal of Information Systems Study Program at Universitas Multimedia Nusantara which presents scientific research articles in the field of Information Systems, as well as the latest theoretical and practical issues, including database systems, management information systems, system analysis and development, system project management information, programming, mobile information system, and other topics related to Information Systems.

ISSN 2085-4579

# FOREWORD

Greetings!

Ultima InfoSys : Jurnal Ilmu Sistem Informasi is a Journal of Information Systems which presents scientific research articles in the field of Information Systems, as well as the latest theoretical and practical issues, including database systems, management information systems, system analysis and development, system project management information, programming, mobile information system, and other topics related to Information Systems. ULTIMA InfoSys Journal is published regularly twice a year (June and December) by Faculty of Engineering and Informatics in cooperation with UMN Press.

In this June 2026 edition, ULTIMA InfoSys enters the 1st Edition of Volume 17. In this edition there are nine scientific papers from researchers, academics and practitioners in the fields covered by Ultima Infosys. Some of the topics raised in this journal are: Analyzing Sentiments in Social Media Posts About Indonesian Tourism with Naïve Bayes; Applied Machine learning for Pediatric Nutrition A K-Means Clustering Application Based on WHO Z-Scores; E-Service Quality Analysis of the MyTelkomsel Application Using CSI, IPA, and PGCV; An Integrated Web-Based Waste Management Platform for Urban Collection Schedulling and Fee Payment Services Pekanbaru City; A Cloud-Integrated Business Process Redesign Framework for Digital Transformation in Traditional SMEs; Redundancy-Aware Feature Selection using mRMR and F-Test for EEG Emotion Classification; Enterprise Architecture Design Using TOGAF ADM for Digital Transformation at PT Profito Inovasi Kreatif; Depression Risk Classification Based on Self-Reported Psychosocial and Behavioral Survey Data Using Machine Learning; ERP Human Resource Management Readiness Framework Integrating Prioritization, Roadmaps, Risk Mitigation, McKinsey 7S.

On this occasion we would also like to invite the participation of our dear readers, researchers, academics, and practitioners, in the field of Engineering and Informatics, to submit quality scientific papers to: International Journal of New Media Technology (IJNMT), Ultimatics : Jurnal Teknik Informatics, Ultima Infosys: Journal of Information Systems and Ultima Computing: Journal of Computer Systems. Information regarding writing guidelines and templates, as well as other related information can be obtained through the email address ultimainfosys@umn.ac.id and the web page of our Journal here.

Finally, we would like to thank all contributors to this December 2024 Edition of Ultima Infosys. We hope that scientific articles from research in this journal can be useful and contribute to the development of research and science in Indonesia.

June 2026,

**Monica Pratiwi, S.ST, MT.**
Editor-in-Chief

ISSN 2085-4579

# TABLE OF CONTENT

ISSN 2085-4579

# Sentiment Analysis of Indonesian Tourism Social Media Using Naïve Bayes for Decision Support

Nathaniel Felix Fraderic[1], Aditiya Hermawan[2], Junaedi[3]
[1,2] Departement of Informatics Engineering, Universitas Buddhi Dharma, Tangerang, Indonesia
[1,2] Departement of Information System, Universitas Buddhi Dharma, Tangerang, Indonesia
[1]nathaniel.felix@ubd.ac.id
[2]aditiya.hermawan@ubd.ac.id
[3]junaedi@ubd.ac.id

Accepted on April 07th, 2025
Approved on January 07th, 2026

***Abstract*— The tourism sector is still the leading alternative sector to boost the Indonesian economy. However, the development of Indonesia's tourism sector still faces challenges, particularly in communication and publication, which remain inadequate. The purpose of this research is to improve the quality of tourist attractions by analyzing responses from visitors to assess the community's feedback. This research aims to employ a data analysis technique that automatically processes these responses, enabling effective sentiment classification and providing actionable insights for enhancing tourism experiences. The method of this study is sentiment analysis using text mining, specifically utilizing the Naïve Bayes algorithm to classify sentiment efficiently. The data for this process is collected using the API provided by the X platform. X was selected as the research object because it facilitates rapid data retrieval through keywords or hashtags, offering a rich source of public opinion specifically on Indonesian tourist destinations. Clearly, the results of this study demonstrate the successful implementation of a website created using the Python programming language. This website visualizes the analysis of people's responses to tourist attractions in graphical form, presenting the percentage of each sentiment class. The text mining process achieved an accuracy of 87%. The conclusion of this study highlights its novelty in applying the Naïve Bayes algorithm for sentiment analysis of Indonesian-language tweets related to tourism, a context that has not been widely explored before. Future research can enhance the sentiment analysis process by incorporating a comprehensive list of positive and negative words in Indonesian, further improving classification precision.**

***Index Terms*— Analyzing sentiment; Naïve Bayes; Text mining; Tourism.**

## I. INTRODUCTION

The tourism sector serves as a crucial alternative driver for Indonesia's economic growth. In 2022, Indonesia's tourism sector contributed 3.6% to the national GDP, reflecting a decline from 4.2% in the previous year [1]. This upward trend demonstrates the sector's gradual recovery and growing significance to the national economy after a period of challenges. However, the development of Indonesia's tourism sector still faces various challenges that must be addressed to support national economic growth [2]. Key among these challenges are deficiencies in communication and publication, which are critical for promoting tourist destinations and attracting visitors. Effective communication and publication are essential to ensure positive feedback and sentiments from visiting tourists, which play a crucial role in shaping perceptions of tourist destinations.

Understanding consumer perceptions is particularly vital in the context of tourism destination marketing, as it influences decisions to visit these destinations [3]. The growing presence of digital tourists highlights the importance of media in shaping interests and preferences, contributing to the popularity of certain themes and destinations [4]. This has led to the rise of social media as a powerful platform for gauging public opinion. Analyzing traveler sentiments, therefore, is essential for enhancing the visibility and appeal of tourist destinations. Leveraging the rapid development of information technology, social media platforms like X have emerged as valuable data sources. With Indonesia ranking fifth globally in terms of X users (18.45 million users as of 2022) [5], X offers rich data for sentiment analysis, facilitated by its public API, which enables efficient text data retrieval. This makes X an ideal medium to explore real-time public opinions on tourist attractions, contributing to a dynamic approach to tourism promotion.

Sentiment analysis, or opinion mining, involves analyzing opinions, evaluations, attitudes, judgments, and feelings about products, services, organizations, events, or topics. Although the field is well-established, its application in tourism, especially in Indonesia, has been insufficiently explored. In recent years, there has been an increase in sentiment analysis studies applied to tourism; however, these have primarily focused on broader or international contexts without addressing the unique linguistic and cultural

nuances of Indonesian tourism discourse [6]. For example, Mehraliyev et al, reviewed sentiment analysis methodologies in tourism, emphasizing the need for localized studies in emerging markets [7]. Similarly, Puh and Bagić Babac (2023) employed machine learning for sentiment analysis in tourism but focused on European destinations, leaving a gap in Indonesian-specific research [8].

This study aims to address these gaps by applying the Naïve Bayes algorithm to analyze sentiments in Indonesian-language tweets about tourist destinations. Previous research using the Naïve Bayes method has demonstrated promising results, such as an accuracy of 81.77% in sentiment analysis [9]. However, despite these advancements, there is still limited research specifically targeting Indonesian-language tweets in the tourism sector. Al-Bakri et al. (2022) compared Naïve Bayes with other algorithms, such as KNN, highlighting the superior performance of Naïve Bayes in specific contexts, though not in Indonesian tourism [10]. Similarly, Singgalen (2024) demonstrated the effectiveness of the Naïve Bayes Classifier in classifying sentiments related to over-tourism issues, highlighting its ability to handle nuanced public perceptions and facilitate sustainable tourism practices by addressing imbalances in data distribution through techniques like SMOTE [11]. Kalaivani et al. (2023) further highlighted the reliability of machine learning techniques like Naïve Bayes in hotel review sentiment analysis [12]. Additionally, Bi et al. (2024) provided an extensive review of text analysis methods in tourism, emphasizing the use of text mining techniques like sentiment analysis and topic modeling. Their study highlighted the utility of these approaches for understanding complex tourist behaviors, predicting industry trends, and managing dynamic challenges in tourism and hospitality [13].

The study differs from prior works in several key aspects. First, it focuses exclusively on Indonesian-language tweets, addressing the linguistic and cultural context that has been largely overlooked in previous studies. Second, it integrates recent advancements in machine learning, specifically in the area of data preprocessing. Techniques like case folding, tokenization, stopword removal, stemming, and TF-IDF weighting were employed to improve data quality and feature extraction. These preprocessing steps ensure that the text data is cleaned and structured effectively, enhancing the quality of the sentiment classification. Third, this study emphasizes practical applications, such as visualizing sentiment analysis results to aid decision-making for potential tourists and industry stakeholders. By leveraging social media data from X, this research not only contributes to the growing body of sentiment analysis literature but also offers practical tools for enhancing tourism marketing strategies and improving tourism experiences in Indonesia.

## II. METHODOLOGIES

The research object used in this study is the response from the public obtained from X. This method follows a similar approach described by Hasanati [14], where sentiment analysis was performed using data from X to assess public opinions, and a series of preprocessing steps, such as text normalization and classification, were employed to structure the data for analysis. The research method is depicted in the following diagram in Fig. 1 below.

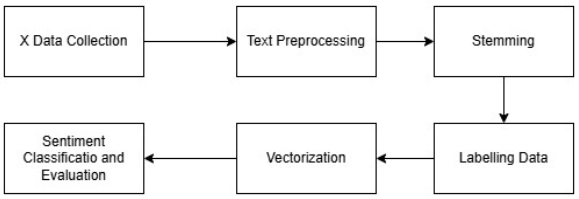


Fig. 1. Research methods diagram

### A. X Data Collection

The data for this study was retrieved using the X platform's public API, requiring an API Key and an Access Token to access the data. Data collection was conducted using a predefined set of keywords related to popular Indonesian tourist destinations, such as "Candi Prambanan," "Borobudur," and "Malioboro," which were selected based on their relevance to tourism discussions on X. The data was gathered over a period from March 2023 to June 20?3, during which a total of 2,717 tweets were collected  The following is the raw data crawling results shown in Table I.

TABLE I. RAW DATA

| Timestamp | Raw_Text | Label | Hastag |
|---|---|---|---|
| 2023-03-17 19:33:05 | berapa harga tiket candi prambanan | NaN | Candi Prambanan |
| 2023-03-17 19:33:05 | RT @media_twc: PT TWC melakukan penutupan dest… | NaN | Candi Prambanan |
| 2023-03-17 19:33:05 | RT @media_twc: PT TWC melakukan penutupan dest… | NaN | Candi Prambanan |
| 2023-03-17 19:33:05 | RT @cineyar: Terima kasih telah meliburkan are… | NaN | Candi Prambanan |
| 2023-03-17 19:33:05 | RT @Fredy_siTom: @VeelaaRhie Candi Borobudur d… | NaN | Candi Prambanan |

Table I presents the initial results of the data collection process, consisting of raw textual data retrieved from the X social media platform through its public API using the keyword “Candi Prambanan.” The raw text attribute contains the original, unprocessed content of each social media post as published by users, including informal language, abbreviations, and contextual expressions that reflect authentic public discourse. As such, this attribute may include noise such as reposts, user mentions, hashtags, URLs, and other non-informative elements, and therefore has not yet undergone text cleaning or sentiment classification.

The hashtag attribute indicates the tourism-related keyword or destination associated with each post and is used to link individual texts to specific tourist attractions during the analysis. The Label column is intentionally left empty at this stage, as sentiment categories (positive, neutral, or negative) are assigned only after the preprocessing and lexicon-based labeling procedures are completed. Additionally, the timestamp attribute records when each post was collected, enabling temporal traceability of public responses. Overall, this table illustrates the role of raw social media data as the foundational input for subsequent sentiment analysis using the Naïve Bayes algorithm to computationally assess public perceptions of Indonesian tourist destinations.

### *B. Text Preprocessing Process*

After the data crawling process from X, text processing is conducted as a crucial step in sentiment analysis. This stage aims to clean and transform unstructured data into a structured format that can be efficiently processed by the system, enabling algorithms to calculate and classify sentiments with greater accuracy [15]. The preprocessing process involves several stages to prepare the text for analysis. Firstly, text cleaning is performed to eliminate non-textual characters like numbers, punctuation, and unnecessary spaces. Subsequently, case folding is applied to ensure uniformity by converting all text to lowercase, preventing distinctions based on capitalization. Tokenization follows, breaking down sentences into individual words or tokens, facilitating further processing. This step enables thorough cleaning of each component within the text, such as tweets. Finally, stopword removal occurs, where common connecting words like "in," "to," "which," and "and" are filtered out using a predefined stoplist, refining the text for subsequent analysis. Each stage contributes to enhancing the quality and relevance of the text data for subsequent tasks.
The results of this process will be displayed in Table II below:

TABLE II. PREPROCESSING RESULTS

| Timestamp | Text_clean |
|---|---|
| 2023-03-17 19:33:05 | berapa harga tiket candi prambanan |
| 2023-03-17 19:33:05 | PT TWC melakukan penutupan destinasi Taman Wi… |
| 2023-03-17 19:33:05 | PT TWC melakukan penutupan destinasi Taman Wi… |
| 2023-03-17 19:33:05 | Terima kasih telah meliburkan area Candi Pram… |
| 2023-03-17 19:33:05 | Candi Borobudur dibangun oleh Raja dari Dinas… |

### *C. Stemming*

After the text preprocessing in Table II is performed, the stemming process continues, converting the words into their root words. The stemming results are displayed in Table III below:

TABLE III. STEMMING RESULTS

| Text_clean | Text_stemming |
|---|---|
| berapa harga tiket candi prambanan | berapa harga tiket candi prambanan |
| PT TWC melakukan penutupan destinasi Taman Wi… | pt twc laku tutup destinasi taman wisata candi… |
| PT TWC melakukan penutupan destinasi Taman Wi… | pt twc laku tutup destinasi taman wisata candi… |
| Terima kasih telah meliburkan area Candi Pram… | terima kasih telah libur area candi prambanan… |
| Candi Borobudur dibangun oleh Raja dari Dinas… | candi borobudur bangun oleh raja dari dinasti… |

### *D. Labelling Data*

During the data labeling process, words are assigned scores based on a predefined list of positive and negative words. These scores determine the polarity of the text, thereby classifying each sentence accordingly: sentences with polarity scores > 0 are classified as positive, those with scores = 0 are considered neutral, and those with scores < 0 are labeled as negative [16].

### *E. Vectorization*

After the words resulting from Preprocessing and Stemming are collected, then the words will be weighted using the TF-IDF (Term Frequency-Inverse Document Frequency ) algorithm. The purpose of this process is to assign a weight value to the words in the document so that the classification process can be carried out. The following is an example of the top value of the TF-IDF weighting after weighting as shown in Table IV below:

TABLE IV. TOP TERM TF-IDF

| Term | TF-IDF |
|---|---|
| Borobudur | 1.000000 |
| Malioboro | 1.000000 |
| Maksutmu | 0.968713 |
| Enter | 0.968713 |
| Puji | 0.963584 |

### *F. Sentiment Classification and Evaluation*

The data collection phase involved utilizing the X API to gather information, focusing on five specific keywords: Malioboro Street, Kraton Yogyakarta, Borobudur Temple, Prambanan Temple, and Tugu Yogyakarta. This endeavor yielded a dataset comprising 2717 entries. Following data collection, the text preprocessing stage ensued. Here, collected text underwent several procedures including text cleaning,

case folding, noise removal, tokenization, filtering (stopword removal), and stemming. Next, in the data labeling step, sentiment analysis was conducted based on lexicon-based methodologies. This involved associating sentiments with their semantic values and assessing the intensity of words within sentences. Utilizing a predetermined lexicon consisting of positive and negative words, the data labeling process was automated. Subsequently, each sentence was assigned a score based on the presence of positive and negative words, ultimately categorizing them into positive, negative, or neutral classes.

Moving forward, the model classification phase employed the Multinomial Naïve Bayes algorithm for sentiment analysis [17]. The general equation 1 for Naïve Bayes classification is given by:

$$P(c|d) = P(c) * \prod_{(w \in d)} P(w|c) \quad (1)$$

Here, P(c|d) represents the posterior probability of class c given document d. To handle the issue of zero probabilities, Laplace smoothing is applied, which adds a +1 to the numerator and |V| (the size of the vocabulary) to the denominator:

$$P(w_i|c) = (count(w_i, c) + 1) / (count(c) + |V|) \quad (2)$$

In this formula, P(w_i|c) is the probability of the word w_i occurring in class c. count(w_i, c) is the number of times the word w_i occurs in class c, while count(c) is the total number of occurrences of all words in class c, and |V| is the total number of unique terms (or features) in the vocabulary.

The Naïve Bayes algorithm uses these formulas to classify sentiments in tourism-related content by constructing a predictive model, assigning probability scores to individual words, and categorizing them into distinct sentiment classes: positive, negative, or neutral. In this study, the Multinomial Naïve Bayes variant was employed with Laplace smoothing to address zero-probability issues and enhance model robustness when handling sparse textual features. The classification process was preceded by a parameterized feature extraction stage using TF–IDF weighting with unigram representations, allowing the model to emphasize discriminative terms relevant to tourism discourse.

Utilizing a supervised machine learning approach, this study categorized sentiments within tourism-related content by constructing a predictive model. This model assigned probability scores to individual words, enabling the accurate classification of text into distinct sentiment categories, thereby offering insights into feedback and perceptions [18]. The training labels were generated through an Indonesian sentiment lexicon–based labeling strategy, which serves as an improvement over manual labeling by ensuring consistency and scalability across large datasets. The dataset was divided into training (75%) and testing (25%) data subsets for model training and evaluation.

Finally, evaluation of the sentiment analysis model was conducted. This involved assessing the performance metrics including accuracy, precision, recall, and F1-Score through the computation of the confusion matrix. Additionally, the evaluation results underwent validation using the 10 K-Fold Cross Validation technique. This validation scheme was selected to ensure performance stability across different data partitions and to demonstrate the generalizability of the proposed parameterized sentiment analysis framework. The reason for using 10-fold cross-validation is to provide a reliable estimate of model performance by splitting the data into 10 subsets, ensuring that each subset is used for both training and testing. This approach helps reduce the risk of overfitting and ensures the model's generalizability. The choice of 10 folds is based on standard practice in machine learning, where it offers a good balance between computation time and performance reliability [19].

## III. Result and Discussion

System implementation and analysis of the results is carried out using the Python programming language. The implementation results in each stage are as follows:

### *A. Classification Process*

The Multinomial Naïve Bayes algorithm was utilized to classify sentiments in tourism-related content. By leveraging Scikit-learn, an open-source library, the model was trained to calculate word probability scores, enabling efficient categorization of documents into their respective sentiment classes [20]. Scikit-learn offers a wide range of algorithms including KNN, SVM, Random Forest, and Naïve Bayes. Beyond classification, it supports preprocessing, regression, clustering, and model selection. The training data comprises 75%, while 25% is reserved for testing purposes. Multinomial naive Bayes is a method that works by calculating the frequency of each term in the document. In Multinomial Naive Bayes, the order in which words appear in the document is ignored so that the document is treated as a "bag of words", so that each word is processed using a multinomial distribution.

Classification results using the Multinomial Naïve Bayes algorithm are shown by the following Table V confusion matrix:

TABLE V. Naïve Bayes Confusion Matrix Results

| | Precision | Recall | F1-score | Support |
|---|---|---|---|---|
| Negative | 0.94 | 0.58 | 0.71 | 417 |
| Netral | 0.85 | 0.97 | 0.91 | 1526 |
| Positive | 0.90 | 0.84 | 0.87 | 774 |
| Accuracy | - | - | 0.87 | 2717 |
| Macro average | 0.90 | 0.79 | 0.83 | 2717 |

| | | | | |
|---|---|---|---|---|
| Weighted average | 0.88 | 0.87 | 0.87 | 2717 |

*B. Evaluation*

The first evaluation, performed with matrix confusion to measure the results of the experiment. To find true positive (TP), true negative (TN), false positive (FP), and false negative values (FN). Below is a matrix confusion by comparing columns of labelling data with lexicon-based and classification results with Naïve bayes algorithms.

From the results of Table VI above there are three classes, namely, Negative, Neutral, and Positive. In this case, True positive = 647, True negative = 240, False positive = 127, False negative = 177, True neutral = 1481 False neutral = 45 and Total = 2717. To find an accuracy value, you can use the following Formula (3), Formula (4) to find precision, Formula (5) to find recall and last formula (6) to get f1 score :

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} * 100\% \quad (3)$$

$$Accuracy = \frac{647+240+1481}{647+240+1481+127+177+45} * 100\%$$

$$= 87{,}15\%$$

$$Precision = \frac{TP}{TP+FP} * 100\% \quad (4)$$

$$Precision = \frac{647}{647+127} * 100\%$$

$$= 83{,}59\%$$

$$Recall = \frac{TP}{TP+FN} * 100\% \quad (5)$$

$$Recall = \frac{647}{647+177} * 100\%$$

$$= 78{,}51\%$$

$$F1 - Score = \frac{2*precision*recall}{precision+recall} \quad (6)$$

$$F1 - Score = \frac{2*0{,}8359*0{,}7851}{0{,}8359+0{,}7851}$$

$$= 80{,}97\%$$

TABLE VI. CONFUSION MATRIX NAÏVE BAYES EVALUATION RESULTS

| | | Prediction | | | |
|---|---|---|---|---|---|
| | | **Negative** | **Netral** | **Positive** | **Total** |
| **Actual** | **Negative** | 240 | 146 | 31 | 417 |
| | **Netral** | 4 | 1481 | 41 | 1526 |
| | **Positive** | 12 | 115 | 647 | 774 |
| | **Total** | 269 | 1742 | 719 | 2717 |

This evaluation was carried out using the 10 K-fold cross validation method to obtain average values of accuracy, recision, recall and F1-Score. With the 10 K-fold cross validation method, the data is divided into 10 parts. With the 10-K-fold Cross Validation method, it is calculated by dividing the data into K subsets, In this case K = 10, the dataset will be divided equally into random positions, where each subset is alternately used as test data while the other subset is used as training data. This process is carried out repeatedly K times or 10 times [21], so that each subset is used as test data once.


Fig. 2. 10-k-fold cross validation process [21]

So by using the 10 K-fold cross validation evaluation method like in **Error! Reference source not found.Error! Reference source not found.**, it can detect the occurrence of overfitting or underfitting. Overfitting occurs when a model can have excellent performance on a specific subset in several iterations, instead Underfitting happens if a model has poor performance in a particular subset.

Then a random test is performed on testing data and training data that yields the following Table VII:

TABLE VII. EVALUATION 10 K-FOLD VALIDATION RESULTS

| Fold | Accuracy | F1-Score | Recall | Precision |
|---|---|---|---|---|
| 1 | 0.875000 | 0.835944 | 0.80248 | 0.890807 |
| 2 | 0.845588 | 0.812332 | 0.782112 | 0.875484 |
| 3 | 0.900735 | 0.87692 | 0.85291 | 0.919553 |
| 4 | 0.841912 | 0.787672 | 0.75646 | 0.873181 |
| 5 | 0.889706 | 0.85527 | 0.824056 | 0.911976 |
| 6 | 0.830882 | 0.769654 | 0.730701 | 0.86676 |
| 7 | 0.900735 | 0.853745 | 0.816694 | 0.922976 |
| 8 | 0.878229 | 0.839525 | 0.809758 | 0.902984 |
| 9 | 0.878229 | 0.825891 | 0.793823 | 0.896928 |

| | | | | |
|---|---|---|---|---|
| 10 | 0.874539 | 0.833785 | 0.79608 | 0.89552 |
| Average | 0.871556 | 0.829074 | 0.796508 | 0.895617 |

After testing 10 times, the accuracy, F1-score, precision and recall values will be calculated as the average of each iteration carried out using 10 K-Fold Cross Validation. The following are the results of K-fold Cross Validation, as shown in Table VIII below:

TABLE VIII. K-FOLD CROSS VALIDATION RESULTS

| Evaluation Method | Accuracy | F1-Score | Recall | Precision |
|---|---|---|---|---|
| Manual | 0.871549 | 0,809703 | 0,785194 | 0,835917 |
| 10-fold | 0.871556 | 0.829074 | 0.796508 | 0.895617 |

From Table VIII, it is obtained and then evaluated using K-Fold Cross Validation, with a total of 10 folds. The validation results show an average accuracy of 0.871556, which is considered to be a reliable and good performance based on K-Fold Cross Validation techniques, as reported by several studies. For instance, Majbur et al. [22], found that Naïve Bayes achieved an average accuracy of 82.33% using similar methods. Based on these findings, several suggestions are proposed to improve the performance of the Naive Bayes classification model. First, to improve the accuracy, F1-score, precision, and recall produced by the model, it is recommended to enhance the application of this method by involving the addition of a list of positive and negative words. This would enrich the labeling process of the training data used, allowing the model to become more sensitive in recognizing positive and negative nuances within the text. Second, regarding the size of the training data, it is advisable to ensure a balance between the amount of data with positive and negative labels. By utilizing a large amount of training data that maintains a balanced label distribution, the model would become more effective in detecting sentences that reflect either positive or negative sentiment. Thus, this approach would enable an improvement in accuracy when classifying these texts more effectively.

To better understand the implications of the model's performance and the proposed improvements, Figure 3 presents a visualization of sentiment classification results for the hashtag Candi Prambanan. This visualization offers a clear representation of the sentiment distribution and complements the previous evaluation results.


Fig. 3 Visualization of sentiment classification results

This study presents significant contributions to the field of sentiment analysis in tourism, particularly through its contextual and methodological advancements. The primary novelty lies in the development of an Indonesian-language corpus derived from the X platform, collected systematically using targeted tourism-related keywords such as Borobudur, Malioboro, and Candi Prambanan. This corpus captures authentic public discourse about Indonesian tourist destinations, thereby addressing a notable gap identified in previous international studies [7], [8] that predominantly focused on Western tourism contexts and English-language data. By localizing sentiment analysis to the Indonesian linguistic and cultural framework, this research enhances the applicability of machine learning models in understanding domestic tourism sentiment and provides a foundation for regional digital analytics. Furthermore, the study operationalizes a complete text-mining pipeline spanning data cleaning, lexicon-based sentiment labeling, TF-IDF vectorization, and classification using the Multinomial Naïve Bayes algorithm to achieve high computational efficiency and reproducibility. Rather than solely emphasizing model performance, this study positions the sentiment classification results as a means to interpret public perceptions and communication effectiveness within Indonesian tourism. The model's performance, validated through 10-fold cross-validation with an average accuracy of 87.15%, surpasses comparable studies utilizing SVM [14] or Doc2Vec approaches [16], thus reinforcing the robustness of Naïve Bayes for sentiment classification in morphologically rich languages like Indonesian.

In terms of substantive findings, the sentiment distribution extracted from 2,717 social media posts reveals that public perception toward Indonesian tourist destinations is predominantly positive and neutral, with approximately 84% of discussions reflecting favorable impressions or neutral informational exchanges, and only about 15% expressing dissatisfaction or critical opinions. The positive sentiments commonly highlight cultural heritage, hospitality, and natural beauty, indicating that Indonesia's core tourism assets are generally well perceived by the public. Neutral sentiments are largely informational, reflecting inquiries about ticket prices, access, and operational details, which underscores the role of social media as an important communication channel for tourism-related information. Meanwhile, negative sentiments are mostly associated with infrastructure limitations, crowding, and inconsistent pricing, directly mirroring the communication and service management challenges identified in the

introduction. These findings explicitly answer the research problem stated in the title and introduction by demonstrating how public sentiment reflects both the strengths and persistent challenges of Indonesian tourism. Accordingly, sentiment analysis is shown to function not merely as a classification task, but as an empirical diagnostic tool for identifying priority areas in tourism communication, service improvement, and destination management. Moreover, the integration of a Python-based web visualization system transforms these findings into actionable intelligence, enabling tourism authorities and stakeholders to monitor public sentiment trends and adapt promotional and managerial strategies accordingly. Thus, beyond methodological innovation, this research provides substantive, data-driven evidence of how sentiment analysis can contribute to enhancing Indonesia's tourism competitiveness through a deeper understanding of visitor perceptions and service gaps.

## IV. CONCLUSION

This research successfully developed a system for sentiment analysis of tweets, achieving an accuracy rate of 87.15% with the Naïve Bayes algorithm. The use of TF-IDF for text vectorization enhanced the model's ability to identify sentiment as positive, negative, or neutral, offering valuable insights for decision-making in tourism. The system demonstrated reliability through rigorous 10 K-Fold cross-validation, yielding an F1-score of 82%, recall of 79%, and precision of 89%. These results highlight the system's robustness in handling varied tweet content and ensuring consistent sentiment classification.

The system's ability to process diverse social media data efficiently makes it a promising tool for sentiment-based decision support, particularly in the tourism industry. While the integration of Naïve Bayes is effective, future research should explore other machine learning techniques such as Random Forest or Deep Learning for comparison. Additionally, expanding to multilingual sentiment analysis and incorporating advanced features like emotion detection could further enhance the system's performance and applicability. This work lays a solid foundation for developing more comprehensive sentiment analysis systems in various domains.

## REFERENCES


[1] I. M. Hasibuan, S. Mutthaqin, R. Erianto, and I. Harahap, "Kontribusi Sektor Pariwisata Terhadap Perekonomian Nasional," *urnal Masharif al-Syariah J. Ekon. dan Perbank. Syariah*, vol. 8, no. 2, pp. 1200–1217, 2023.

[2] R. Kurniawan, "Strategi Pemasaran Pariwisata Untuk Meningkatkan Pariwisata Lokal," vol. 7, pp. 9867–9877, 2024, doi: https://doi.org/10.31004/jrpp.v7i3.31431.

[3] D. Setiadi and Y. I. Mukti, "Electronic Tourism Using Decision Support Systems to Optimize the Trips," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 23, no. 1, pp. 183–200, 2023, doi: 10.30812/matrik.v23i1.3331.

[4] Y. A. Singgalen, "Analisis Sentimen dan Pemodelan Topik dalam Optimalisasi Pemasaran Destinasi Pariwisata Prioritas di Indonesia," *J. Inf. Syst. Informatics*, vol. 3, no. 3, pp. 459–470, 2021, doi: 10.51519/journalisi.v3i3.171.

[5] R. T. B. Ardianto and Y. Nataliani, "Analisis Jaringan Sosial Pengguna Twitter Dalam Skema Penyediaan Laptop Bagi Toko Komputer," vol. 02, no. 3, pp. 208–218, 2023, doi: https://doi.org/10.24246/itexplore.v2i3.2023.pp208-218.

[6] Y. Mao, Q. Liu, and Y. Zhang, "Sentiment analysis methods, applications, and challenges: A systematic literature review," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 36, no. 4, p. 102048, 2024, doi: https://doi.org/10.1016/j.jksuci.2024.102048.

[7] F. Mehraliyev, I. C. C. Chan, and A. Kirilenko, "Sentiment analysis in hospitality and tourism: a thematic and methodological review," *Int. J. Contemp. Hosp. Manag.*, vol. ahead-of-p, 2021, doi: 10.1108/IJCHM-02-2021-0132.

[8] K. Puh and M. Bagić Babac, "Predicting sentiment and rating of tourist reviews using machine learning," *J. Hosp. Tour. Insights*, vol. 6, no. 3, pp. 1188–1204, 2023, doi: 10.1108/JHTI-02-2022-0078.

[9] C. Villavicencio, J. J. Macrohon, X. A. Inbaraj, J. H. Jeng, and J. G. Hsieh, "Twitter sentiment analysis towards covid-19 vaccines in the Philippines using naïve bayes," *Inf.*, vol. 12, no. 5, 2021, doi: 10.3390/info12050204.

[10] N. F. Al-Bakri, J. F. Yonan, A. T. Sadiq, and A. S. Abid, "Tourism companies assessment via social media using sentiment analysis," *Baghdad Sci. J.*, vol. 19, no. 2, pp. 422–429, 2022, doi: 10.21123/BSJ.2022.19.2.0422.

[11] Y. Afrianto Singgalen, "Sentiment Classification of Over-Tourism Issues in Responsible Tourism Content using Naïve Bayes Classifier," *J. Comput. Syst. Informatics*, vol. 5, no. 2, pp. 275–285, 2024, doi: 10.47065/josyc.v5i2.4904.

[12] M. V. Kalaivani, Shamini B, S. R, and T. S. D, "Hotel Reviews Sentiment Analysis Using Machine Learning," *Int. Res. J. Mod. Eng. Technol. Sci.*, no. 03, pp. 3047–3056, 2023, doi: 10.56726/irjmets34902.

[13] J. W. Bi, X. E. Zhu, and T. Y. Han, "Text Analysis in Tourism and Hospitality: A Comprehensive Review," *J. Travel Res.*, no. May, 2024, doi: 10.1177/00472875241247318.

[14] N. Hasanati, Q. Aini, and A. Nuri, "Implementation of Support Vector Machine with Lexicon Based for Sentiment Analysis on Twitter," in *2022 10th International Conference on Cyber and IT Service Management (CITSM)*, 2022, pp. 1–4. doi: 10.1109/CITSM56380.2022.9935887.

[15] A. M. Pravina, "Sentiment Analysis of Delivery Service Opinions on Twitter Documents using K-Nearest Neighbor," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 2, pp. 996–1012, 2022, doi: 10.35957/jatisi.v9i2.1899.

[16] T. H. Jaya Hidayat, Y. Ruldeviyani, A. R. Aditama, G. R. Madya, A. W. Nugraha, and M. W. Adisaputra, "Sentiment analysis of twitter data related to Rinca Island development using Doc2Vec and SVM and logistic regression as classifier," *Procedia Comput. Sci.*, vol. 197, pp. 660–667, 2022, doi: https://doi.org/10.1016/j.procs.2021.12.187.

[17] A. Sabrani, I. G. W. Wedashwara W., and F. Bimantoro, "Multinomial Naïve Bayes untuk Klasifikasi Artikel Online tentang Gempa di Indonesia," *J. Teknol. Informasi, Komputer, dan Apl. (JTIKA )*, vol. 2, no. 1, pp. 89–100, 2020, doi: 10.29303/jtika.v2i1.87.

[18] A. R. Alaei, S. Becken, and B. Stantic, "Sentiment Analysis in Tourism: Capitalizing on Big Data," *J. Travel Res.*, vol. 58, no. 2, pp. 175–191, 2021, doi: 10.1177/0047287517747753.

[19] A. Rifa'i, H. Sujaini, and D. Prawira, "Sentiment Analysis Objek Wisata Kalimantan Barat Pada Google Maps Menggunakan Metode Naive Bayes," *J. Edukasi dan Penelit. Inform.*, vol. 7, no. 3, p. 400, 2021, doi: 10.26418/jp.v7i3.48132.

[20] C. Muhamad, S. Ramdani, A. N. Rachman, and R. Setiawan, "Comparison of the Multinomial Naive Bayes Algorithm and Decision Tree with the Application of AdaBoost in Sentiment Analysis Reviews PeduliLindungi Application," *Int. J. Inf. Syst. Technol. Akreditasi*, vol. 6, no. 158, pp. 419–430, 2022, doi: https://doi.org/10.30645/ijistech.v6i4.257.

[21] H. Belyadi and A. Haghighat, "Machine Learning Guide for Oil and Gas Using Python: A Step-by-Step Breakdown with Data, Algorithms, Codes, and Applications.," H. Belyadi and A. B. T.-M. L. G. for O. and G. U. P. Haghighat, Eds., Gulf Professional Publishing, 2021, p. iii. doi: https://doi.org/10.1016/B978-0-12-821929-4.01001-5.

[22] R. F. Majbur, F. Yanuar, D. Devianto, D. Matematika, and U. Andalas, "Penerapan Metode Naïve Bayes Classifier Dalam Menganalisis Sentimen Pada Media Sosial X Terhadap Pilpres 2024 Di Indonesia," *J. Ilm. Pendidik. Mat. Mat. dan Stat.*, vol. 5, no. 3, pp. 2012–2025, 2024, doi: 10.46306/lb.v5i3.

ISSN 2085-4579

# Applied Machine learning for Pediatric Nutrition A K-Means Clustering Application Based on WHO Z-Scores

Julius Sepadan Maruli Tua Manurung[1], Sri Agustina Rumapea[1], Edward Rajagukguk[1], Fernando Rumapea[1]
[1] Information System Study Program, Universitas Methodist Indonesia, Medan, Indonesia
srirumape78@gmail.com

Accepted on November 21st, 2025
Approved on March 27th, 2026

***Abstract*—Nutritional status in early childhood is a key indicator of public health and serves as a basis for targeted nutritional interventions. Children's nutrional status is significantly influenced by family food adequacy and parental nutritional literacy. Nutritional problems in children can come from primary and secondary factors. Primary factors include the non-fulfillment of nutrient intake, both in quantity and quality. Meanwhile, secondary factors occur due to disturbances in the child's body in carrying out its natural functions. This study applies the K-Means Clustering algorithm to anthropometric data—including weight, height, and head circumference—to classify children's nutritional status into five categories: severely undernourished (−3 SD), moderately undernourished (−2 SD), normal, mildly overnourished (+1 SD), and moderately overnourished (+2 SD). The clustering process is based on Z-scores derived from WHO standards, namely Weight-for-Age (WAZ), Height-for-Age (HAZ), and Head Circumference-for-Age (HCAZ), which serve as input features for the clustering algorithm. The number of clusters (k = 5) is aligned with national nutritional classification guidelines. The clustering results are visualized to illustrate the data distribution across the three indicators, and evaluated using Euclidean distance to assess the proximity of each data point to its assigned cluster centroid. The results demonstrate that K-Means Clustering can effectively classify children's nutritional status in a manner consistent with manual classification based on WHO thresholds. This study highlights the potential of data mining approaches in supporting health information systems for the early and automated detection of child malnutrition.**

***Index Terms*—Nutritional Status; Anthropometry; Z-Score; K-Means Clustering**

## I. INTRODUCTION

Nutritional status during early childhood forms the foundation for a child's physical growth and cognitive development. During this critical growth window (ages 0–5), any deviation—whether undernutrition or overnutrition—can have long-term consequences on a child's overall quality of life. To monitor child development objectively and in a standardized manner, the World Health Organization (WHO) recommends the use of three Z-score indicators: Weight-for-Age (WAZ), Height-for-Age (HAZ), and Head Circumference-for-Age (HCAZ) [1]. These indicators assess a child's weight, height, and head circumference relative to a healthy reference population and have become a global standard for nutritional surveillance [2].

In practice, however, nutritional classification in many healthcare facilities—especially in remote or resource-limited areas—is still performed manually. This approach is time-consuming, prone to human error, and difficult to scale efficiently within a data-driven health system. As a result, nutritional interventions are often delayed or misaligned due to the lack of rapid and reliable analysis tools [3].

With the rise of digital health technologies, data-driven methods such as machine learning have increasingly been applied in public health. One promising approach is K-Means Clustering, an unsupervised learning algorithm that can group data based on patterns or similarities in specific features [4]. Several studies have applied this method to classify children's nutritional status, used K-Means to cluster children based on weight and age, but without incorporating WHO Z-score standards [5]. Clustering to raw anthropometric values but failed to align the results with clinical thresholds [6]. In previous studies, Z-score indicator has been used, although only one or two variables were included [7].

Clustering in the context of stunting but excluded HCAZ, which is essential for assessing neurological development. Furthermore, most existing studies lack alignment with WHO nutritional cutoffs, limiting their applicability in clinical settings [8].

This highlights a clear research gap: no prior study has integrated all three Z-score indicators—WAZ, HAZ, and HCAZ—into a unified clustering model to classify children's nutritional status. The present study seeks to fill this gap by applying the K-Means Clustering algorithm to all three indicators simultaneously. Children are grouped into five categories: severely undernourished (−3 SD), moderately undernourished (−2 SD), normal, mildly overnourished (+1 SD), and moderately overnourished (+2 SD), consistent with WHO standards.

The objectives of this study are to (1) develop a data-driven nutritional classification model using K-Means Clustering with WHO-based Z-scores, and (2) evaluate its performance in comparison to conventional classification approaches. The outcomes are expected to assist healthcare providers in the early detection of growth deviations based on anthropometric data and support timely, targeted interventions.

## II. LITERATUR REVIEW

Nutritional status during early childhood forms the foundation for a child's physical growth, cognitive development, and long-term quality of life. In this critical growth window (ages 0–5), both undernutrition and overnutrition can lead to persistent deficits in health, learning capacity, and productivity later in life. To monitor child development objectively and in a standardized manner, the World Health Organization (WHO) recommends the use of three Z-score indicators—Weight-for-Age (WAZ), Height-for-Age (HAZ), and Head Circumference-for-Age (HCAZ)—which compare a child's weight, height, and head circumference against a healthy reference population [9]. These indicators have become the global standard for nutritional surveillance and early detection of growth deviations.

In many healthcare facilities, particularly in remote or resource-limited settings, nutritional classification is still performed manually. Health workers often rely on manual plotting or basic spreadsheets to interpret WHO growth charts, which is time-consuming, prone to human error, and difficult to scale in a data-driven health system [10]. As the volume of routine anthropometric data increases, this manual process can delay the identification of children at risk and contribute to interventions that are either late or poorly targeted [11].

With the rise of digital health technologies, data-driven methods such as machine learning have increasingly been applied in public health to address these challenges. One promising approach is K-Means clustering, an unsupervised learning algorithm that groups data based on similarities in selected features [12]. Several studies have used K-Means to classify children's nutritional status or to explore patterns of malnutrition, generally using weight, height, or age as input variables [13]. Some studies have clustered children based on raw anthropometric values or on one or two Z-score indicators, yet their cluster definitions are not always aligned with WHO cut-off points, which limits their direct applicability in clinical decision-making[14]. Other works have demonstrated that simple anthropometric features can also be used in supervised models, such as K-Nearest Neighbours, to support early detection of stunting and abnormal nutritional status [15] .

Despite these advances, important gaps remain. Most existing clustering studies do not integrate all three WHO Z-score indicators—WAZ, HAZ, and HCAZ—into a single model, even though head circumference is closely related to brain growth and is a sensitive marker of early neurodevelopmental risk [16]. In many cases, HCAZ is either ignored or analysed separately, and the resulting clusters do not map directly to the five WHO nutritional categories. This misalignment makes the outputs harder to interpret in routine practice and reduces their usefulness for front-line health workers [17].

These limitations highlight a clear research gap: there is a need for an integrated clustering model that simultaneously uses WAZ, HAZ, and HCAZ and produces clusters that are consistent with WHO nutritional cutoffs. To address this gap, the present study applies the K-Means clustering algorithm to all three Z-score indicators at once. Children are grouped into five categories—severely undernourished (≤ −3 SD), moderately undernourished (≤ −2 SD), normal, mildly overnourished (≥ +1 SD), and moderately overnourished (≥ +2 SD)—in accordance with WHO standards. The objectives of this study are to (1) develop a data-driven nutritional classification model using K-Means clustering with WHO-based Z-scores and (2) evaluate its performance in comparison with conventional classification approaches. The expected outcome is a more systematic and scalable way to detect growth deviations based on anthropometric data, thereby supporting healthcare providers in delivering timely and targeted interventions.

Assessing the nutritional status of children through anthropometric data is a foundational practice in global child health monitoring. The World Health Organization (WHO) recommends the use of standardized growth indicators such as Weight-for-Age Z-score (WAZ), Height-for-Age Z-score (HAZ), and Head Circumference-for-Age Z-score (HCAZ). These indicators quantify the deviation of a child's anthropometric measurements from the median values of a healthy reference population, adjusted for age and sex, and provide a common framework for comparing nutritional status across populations and over time.

Each indicator captures a specific dimension of growth. WAZ reflects a child's general weight status and is used to detect underweight conditions [18]. HAZ

identifies stunting, which indicates chronic malnutrition or long-term growth restriction and has been associated with adverse health and educational outcomes [19]. HCAZ reflects head growth and serves as an early indicator of neurological development and brain growth, especially in the first two years of life, when rapid brain development occurs [20]. Evidence shows that low head circumference-for-age is associated with increased risk of suboptimal neurodevelopment, making HCAZ an important complementary indicator alongside WAZ and HAZ. In this study, these Z-scores are categorized into five clinical groups: severely undernourished ($Z < -3$ SD), moderately undernourished ($-3$ SD $\leq Z < -2$ SD), normal ($-2$ SD $\leq Z \leq +1$ SD), mildly overnourished ($+1$ SD $< Z \leq +2$ SD), and moderately overnourished ($Z >$ $+2$ SD), consistent with WHO recommendations[21].

In recent years, machine learning—particularly clustering techniques—has been adopted to enhance the efficiency and objectivity of nutritional classification. Clustering is an unsupervised learning approach that groups data based on similarity across multiple features without requiring labeled training data. Among various clustering methods, K-Means clustering is one of the most widely used due to its conceptual simplicity and computational efficiency. It partitions data into k clusters by iteratively assigning observations to the nearest centroid and updating centroids to minimize within-cluster variance. This makes K-Means suitable for exploring patterns in large anthropometric datasets and for segmenting children into risk-based groups.

Several studies have demonstrated the applicability of K-Means clustering in child nutrition classification. Improved upon these early efforts by using Z-score indicators in their models; however, they typically restricted the analysis to one or two dimensions (for example, only WAZ or HAZ), which reduces the ability of the model to capture the multidimensional nature of child growth [22].

More recent work has expanded the use of K-Means to different datasets and levels of aggregation. Several studies have used K-Means to cluster toddlers based on combinations of anthropometric indicators and related variables for the purpose of identifying malnutrition or stunting risk at the individual level. Other studies have applied K-Means to classify districts or cities based on stunting prevalence or malnutrition indicators, providing a basis for spatial targeting and program prioritization [23]. In parallel, supervised learning methods such as K-Nearest Neighbours (KNN) have been developed to classify toddlers' nutritional status and support early detection of stunting using simple anthropometric inputs. Together, these studies show that routine growth data can be transformed into decision-support tools that assist health workers in identifying children and areas at higher risk.

However, several important limitations are evident across the existing literature. First, many clustering studies operate on raw anthropometric values or on a limited subset of indicators (e.g., only WAZ or HAZ) and do not integrate all three WHO growth indicators—WAZ, HAZ, and HCAZ—within a single model. Second, head circumference is frequently omitted, even though HCAZ is a sensitive marker of early brain development and can provide additional insight into neurodevelopmental risk. Third, not all studies explicitly align the resulting clusters with WHO nutritional cut-offs, which weakens their clinical relevance and makes it difficult for practitioners to directly use the outputs in routine child health services. These limitations reduce the extent to which existing models can be integrated into digital health applications or standard growth monitoring workflows.

To date, and to the best of our knowledge, there is no documented study that simultaneously incorporates WAZ, HAZ, and HCAZ into a unified K-Means clustering model for individual-level nutritional classification. This gap presents an opportunity to develop a more holistic, data-driven framework that captures both physical growth and early neurodevelopmental indicators while remaining aligned with WHO clinical categories.

The present study seeks to address this gap by applying the K-Means clustering algorithm to all three Z-scores—WAZ, HAZ, and HCAZ—simultaneously. The aim is to classify children's nutritional status into clinically relevant groups that are consistent with WHO cut-offs and to enable early detection of deviations from normal growth. This approach is expected to provide a more comprehensive and interpretable classification than models based on a single indicator, while also offering a scalable analytic component for digital health applications in early childhood nutrition monitoring.

## III. METHODOLOGY

This study implements the K-Means Clustering algorithm to classify children's nutritional status using three anthropometric indicators standardized by the World Health Organization (WHO): Weight-for-Age Z-score (WAZ), Height-for-Age Z-score (HAZ), and Head Circumference-for-Age Z-score (HCAZ). The methodological process consists of four main stages: (1) data preparation, (2) Z-score calculation, (3) clustering using K-Means, and (4) system implementation and visualization.

### *A. Data Preparition*

The dataset consists of 77 records of children aged 0 to 60 months, collected from a local pediatric clinic. Each record includes the child's sex, age (in months), weight (kg), height (cm), and head circumference (cm). The sex and age of each child are used to select the appropriate WHO reference data during Z-score computation.

### B. Z-Score Calculation

Three nutritional indices are calculated based on WHO child growth standards:

- **WAZ (Weight-for-Age Z-score)**: Indicates overall underweight status.
- **HAZ (Height-for-Age Z-score)**: Identifies stunting or linear growth delay.
- **HCAZ (Head Circumference-for-Age Z-score)**: Serves as a proxy for brain development.

The Z-score for each indicator is calculated using the following general formula:

(1)

Where:

- XXX is the child's measured value (e.g., weight, height, head circumference),
- μ\muμ is the median value for the same age and sex from WHO growth standards,
- σ\sigmaσ is the standard deviation for that measurement.
- Separate computations are performed for male and female children based on WHO growth charts.

### C. K-Means Clustering

The calculated WAZ, HAZ, and HCAZ scores are used as input features for the K-Means Clustering algorithm. The algorithm performs the following steps:

1. **Initialize centroids**: Five initial centroids are chosen, representing the five nutritional categories:
   - Cluster 1: Severely undernourished (Z < −3 SD)
   - Cluster 2: Moderately undernourished (−3 SD ≤ Z < −2 SD)
   - Cluster 3: Normal (−2 SD ≤ Z ≤ +1 SD)
   - Cluster 4: Mildly overnourished (+1 SD < Z ≤ +2 SD)
   - Cluster 5: Moderately overnourished (Z > +2 SD)
2. **Distance calculation**: Each data point is assigned to the nearest centroid using the Euclidean distance:

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \quad (2)$$

Where xxx is the Z-score vector of a child and yyy is the centroid vector.

3. **Centroid update**: New centroids are calculated by averaging the members of each cluster:

$$\mu j = \frac{1}{n_j} \sum_{i=1}^{n_j} xi \quad (3)$$

Where njn_jnj is the number of data points in cluster jjj.

4. **Iteration**: Steps 2 and 3 are repeated until convergence is reached, i.e., no significant changes in centroid positions.

### D. System Implementation and Visualization

To support practical deployment, a prototype system with a simple graphical user interface (GUI) was developed. The system provides the following functionalities:

- Input form for anthropometric data
- Automatic computation of WAZ, HAZ, and HCAZ using WHO reference values
- Clustering of new data points based on trained K-Means model
- Display of results including assigned cluster and nutritional category
- Visualization using scatter plots for WAZ, HAZ, and HCAZ distributions across clusters
- Summary statistics for each cluster (count, percentage)

This system can assist healthcare workers in performing rapid assessments and identifying children at risk for growth deviations or abnormal development.

## IV. Results and Discussion

This section presents the results of the Z-score computations, K-Means clustering process, system implementation, and visualization of classification outcomes. The discussion highlights the effectiveness of the proposed approach in identifying growth deviations and its alignment with WHO-based nutritional categories.

### A. Z-Score Computation Results

The first step involved calculating WAZ, HAZ, and HCAZ for each of the 77 children in the dataset using WHO reference data. Separate computations were performed for male and female children to account for sex-specific growth standards.

Table I shows a sample of the computed Z-scores for five children:

TABLE I Sample Z-Score Calculation (WAZ, HAZ, HCAZ)

| ID | Gender | Age | WAZ | HAZ | HCAZ |
|---|---|---|---|---|---|
| M1 | Female | 44 | -1,21 | -1,05 | 0,29 |
| M2 | Female | 24 | -0,38 | 0,09 | -0,14 |
| M3 | Male | 38 | 0,26 | -0,37 | -0,79 |
| M4 | Female | 60 | -1,58 | -2 | -0,64 |

| M5 | Male | 40 | -1,53 | -2,72 | -0,5 |
|---|---|---|---|---|---|
| M6 | Male | 39 | -0,41 | -0,79 | 0,5 |
| **ID** | **Gender** | **Age** | **WAZ** | **HAZ** | **HCAZ** |
| M7 | Female | 31 | -0,6 | -1,26 | -0,71 |
| M8 | Male | 53 | -1,57 | -3,2 | -1,43 |
| M9 | Male | 24 | -0,93 | 0,13 | -0,36 |
| M10 | Male | 59 | -1,04 | -1,17 | 0,33 |
| … | … | … | … | … | … |
| M77 | Female | 27 | 0,19 | 0,5 | -0,43 |

The complete Z-score data served as the input for clustering.

### B. K-Means Clustering Results

The K-Means algorithm was applied with k = 5 to reflect the five clinical nutrition categories defined by WHO. Initial centroids were selected based on representative values from the Z-score ranges for each group.

The algorithm converged after three iterations, producing five clusters that correspond to the intended nutritional classifications. Table II summarizes the final centroid values for each cluster:

TABLE II FINAL CLUSTER CENTROIDS (WAZ, HAZ, HCAZ)

| Kode | Category | WAZ | HAZ | HCAZ |
|---|---|---|---|---|
| C1 | Saverely Undernourished | -4,3 | -7,105 | -5 |
| C2 | Moderately Undernourished | -1,47189 | -2,08242 | -0,87014 |
| C3 | Normal | -0,29039 | -0,39824 | -0,01824 |
| C4 | Mildly Overnourished | 2,59 | 0,98 | -0,43 |
| C5 | Moderately Overnourished | 2,605 | 3,07 | 1,605 |

The cluster distribution is summarized in Table III.

TABLE III CLUSTER DISTRIBUTION AND PERCENTAGE

| Cluster | Category | Children | Percentage |
|---|---|---|---|
| C1 | Saverely Undernourished | 2 | 2,60% |
| C2 | Moderately Undernourished | 22 | 28,57% |
| C3 | Normal | 49 | 63,64% |
| C4 | Mildly Overnourished | 2 | 2,60% |
| C5 | Moderately Overnourished | 2 | 2,60% |

The results show that the majority of children (63.64%) fall within the normal category, while a significant portion (28.57%) are moderately undernourished, highlighting the importance of early identification and intervention.

### C. Visualization of Cluster Results

To enhance interpretability, a 3D scatter plot was generated using WAZ, HAZ, and HCAZ as axes. Each cluster is represented with a distinct color (e.g., red for severely undernourished, blue for normal).

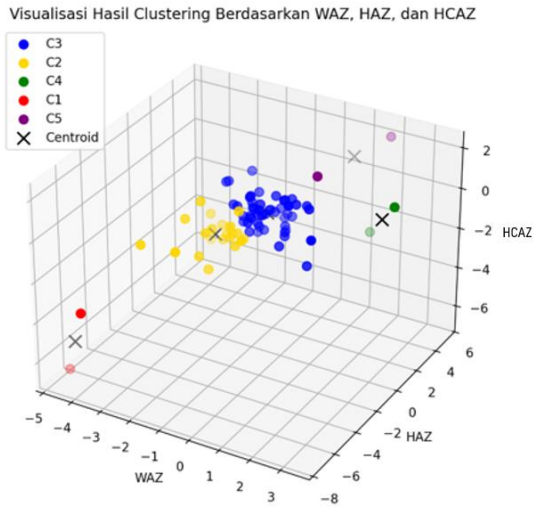


Fig. 1. Scatter Plot of Cluster Based on Z-Scores

The plot demonstrates clear group separation, with distinct boundaries between malnourished and normal children. Outliers were successfully captured in extreme clusters (C1 and C5), validating the effectiveness of the distance-based grouping.

### D. System Implementation and Use Case

To bridge the gap between algorithmic output and real-world application, a simple yet functional interface was developed as part of this study. Health workers can enter basic anthropometric information—such as age, sex, weight, height, and head circumference—into the system. The system then automatically calculates the Z-scores (WAZ, HAZ, and HCAZ), performs clustering, and returns the child's nutritional classification based on the nearest cluster.

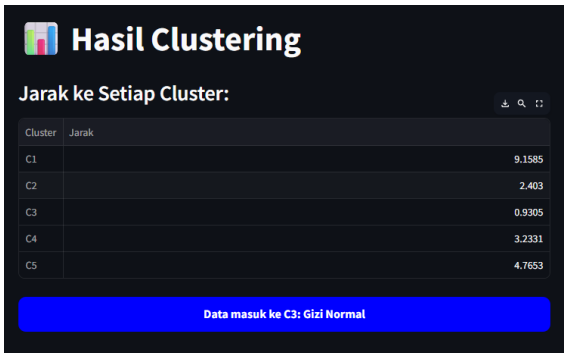


Fig. 2 GUI Interface and Nutritional Classification Results

The system was designed with usability in mind, especially for field settings or primary healthcare clinics where time and resources are limited. By providing quick, standardized feedback based on WHO guidelines, it helps frontline health staff make faster, data-informed decisions—without the need for advanced technical knowledge.

### *E. Discussion*

The results of this study reinforce the potential of machine learning—particularly K-Means Clustering—as a practical tool for early childhood growth monitoring. By using three key anthropometric indicators (WAZ, HAZ, and HCAZ) in combination, this model offers a more holistic view of a child's health status compared to single-indicator methods.

1. Comparison with Previous Studies

Earlier work by Kumar et al. [4] focused on clustering children based on weight and age but did not use Z-scores or reference growth standards, which limited the clinical interpretability of their results. In contrast, our model is rooted in WHO-standardized indicators, making the clusters directly relevant to real-world classifications of undernutrition and overnutrition.

Huang et al. [5] and Ranjawali et al. [5] also explored clustering based on raw anthropometric data, but again, without mapping their output to WHO-defined thresholds. Dewi et al. f and Muhammad et al. improved on this by introducing WAZ and HAZ scores—but stopped short of incorporating HCAZ, an important signal for early brain development. By including all three indicators, our approach captures a more complete picture of a child's nutritional and developmental status.

Another key differentiator of this study is its emphasis on practical usability. While most prior studies ended at the algorithmic or analytical stage, this work goes a step further by developing a working system that can be used by practitioners. The interface is intuitive and requires minimal training, making it accessible even in lower-resource environments.

2. Key Contributions

In summary, this study contributes to the field in three important ways:

- It integrates all three WHO-recommended Z-score indicators into the clustering process, offering a richer and more reliable classification.
- It aligns cluster groupings with clinically accepted nutritional categories, making the results immediately useful in healthcare settings.

It provides a ready-to-use digital tool for real-time classification, empowering health workers with faster, data-based insights.

3. Limitations and Future Directions

Despite these strengths, the study does have limitations. The dataset is relatively small (77 records), which may affect the generalizability of the model. In future work, the model should be trained and tested on larger, more diverse datasets across multiple regions to increase robustness.

Other clustering methods—such as DBSCAN or Gaussian Mixture Models—could also be explored for comparison, especially in handling outliers or non-spherical data. Additionally, integrating this model into national electronic health record systems could extend its impact by enabling large-scale, automated nutritional screening.

## V. CONCLUSION

This study examined the application of the K-Means Clustering algorithm to classify children's nutritional status using three WHO-standard anthropometric indicators: Weight-for-Age Z-score (WAZ), Height-for-Age Z-score (HAZ), and Head Circumference-for-Age Z-score (HCAZ). By integrating all three indicators into one model, the study offers a more comprehensive and clinically meaningful approach to identifying nutritional deviations during early childhood. The clustering process successfully grouped children into five categories consistent with WHO growth standards. Most children were classified as having normal nutritional status, while a significant proportion were found to be moderately undernourished. This confirms that using multiple Z-score indicators provides a clearer and more accurate picture of a child's growth pattern.

A key contribution of this work lies in the development of a functional system that enables health workers to input basic anthropometric data and instantly receive a nutritional classification. This tool supports early detection and timely intervention, especially in low-resource settings where manual classification may be impractical or inconsistent. Compared to previous studies that relied on limited indicators or remained at the modeling stage, this study not only improves classification accuracy but also bridges the gap between data analysis and practical application. The inclusion of HCAZ as a developmental metric adds further value, especially for assessing brain growth in younger children.

An important strength of the clustering approach used in this research is its ability to quantify how far each group deviates from the standard growth curve. The model revealed that 36.36% of children fell outside the WHO-defined normal nutritional range—either as undernourished or overnourished—while 63.64% were

classified within normal limits. Among those outside the normal range, the majority belonged to the moderately undernourished group, indicating early signs of growth faltering. This method not only categorizes but also provides measurable insight into the severity and distribution of growth deviations. It enables a more targeted and data-informed nutritional intervention strategy, especially in contexts where time, accuracy, and simplicity are critical.

Although the dataset used in this study was relatively small, the findings are promising. Future work should involve larger and more diverse datasets, as well as comparisons with other clustering algorithms to evaluate robustness and scalability. Integrating this model into national health systems or mobile health platforms could further amplify its practical impact, making machine learning a valuable tool in improving child health outcomes.

## REFERENCES


[1] I. S. Tinendung and I. Zufria, "Pengelompokan Status Stunting Pada Anak Menggunakan Metode K-Means Clustering," J. Media Inform. Budidarma, vol. 7, no. 4, p. 2014, 2023, doi: 10.30865/mib.v7i4.6908.

[2] R. Husna, V. Riyanto, F. Teknik, S. Informasi, U. Bina, and S. Informatika, "Klasterisasi Status Gizi Balita Menggunakan Algoritma K-Means Melalui Pendekatan Soft System Methodology," vol. 13, no. September, pp. 1–9, 2024, doi.org/10.70309/ticom.v13i1.120.

[3] R. Ranjawali, A. C. Talakua, and R. T. Abineno, "Clustering Stunting Pada Balita Dengan Metode K- Means Di Puskesmas Kanatang," SATI Sustain. Agric. Technol. Innov., pp. 80–92, 2023, [Online]. Available: https://ojs.unkriswina.ac.id/index.php/semnas-FST/article/view/587/324

[4] M. N. Zhafar et al., "Penerapan Metode Clustering Dengan Algoritma K-Means Untuk Analisa Persebaran Varian Covid-19 ( Studi Kasus Kelurahan Antapani Kidul )," e-Proceeding Eng., vol. 2, no. 1, p. 1042, 2023, [Online]. Available: http://jurnal.mdp.ac.id

[5] E. N. Tenggara, D. Kusumo, and K. V. Id, "Gut microbiota differences in stunted and normal-lenght children aged 36 – 45 months in," pp. 1–20, 2024, doi: 10.1371/journal.pone.0299349.

[6] H. Li et al., "Prevalence and associated factors for stunting , underweight and wasting among children under 6 years of age in rural Hunan Province , China : a community-based cross-sectional study," BMC Public Health, pp. 1–12, 2022, doi: 10.1186/s12889-022-12875-w.

[7] L. Riaz, A. M. Siddiqui, M. N. Rashid, R. Ahmed, N. Huda, and N. Shahab, "A Study of Estimation of Stature by Anthropometric Measurements of the Head ABSTRACT," vol. 36, no. 4, pp. 40–42, 2025, doi: 10.60110/medforum.360409.INTRODUCTION.

[8] R. M. Sari, A. Rizka, N. A. Putri, and A. Efriana, "Penerapan Data Mining Untuk Analisis Stunting Pada Balita," vol. 13, no. November, pp. 1717–1728, 2024, doi.org/10.33395/jmp.v13i2.14218.

[9] H. Munawaroh et al., "Peranan Orang Tua Dalam Pemenuhan Gizi Seimbang Sebagai Upaya Pencegahan Stunting Pada Anak Usia 4-5 Tahun," Sentra Cendekia, vol. 3, no. 2, p. 47, 2022, doi: 10.31331/sencenivet.v3i2.2149.

[10] B. S. Kaka et al., "Penerapan Metode K-Means Clustering Bagi Balita Penderita," vol. 10, no. 4, pp. 256–269, 2023, doi.org/10.35957/jatisi.v10i4.6085.

[11] D. K. Geyer, "Penerapan Algoritma K-Means Dalam Proses Klasterisasi Balita Stunting," vol. 31, no. 2, pp. 280–289, 2025, doi: 10.36309/goi.v31i2.420.

[12] M. Miranda, N. Rahaningsih, and ..., "Analisis Clustering Data Anak Balita di Posyandu Kampung Sukarame Menggunakan Algoritma K-Means," J. Inform. …, vol. 6, no. 1, pp. 136–141, 2024, [Online]. Available: https://publikasiilmiah.unwahas.ac.id/JINRPL/article/view/10256

[13] U. Bina, D. Jl, and J. A. Yani, "Analisis Clustering Data Gizi Pada Puskesmas 1 Ulu Menggunakan Metode Algoritma K-Means," vol. 5, no. 2, pp. 209–217, 2025, doi: 10.35957/algoritme.v5i2.10707.

[14] Endang Sawitri, Setianingsih, and Indri Septiana, "Gambaran Status Gizi Pada Anak Usia Pra Sekolah Di Tk Pertiwi Tangkil," TRIAGE J. Ilmu Keperawatan, vol. 10, no. 1, pp. 30–36, 2023, doi: 10.61902/triage.v10i1.652.

[15] A. S. Ritonga and I. Muhandhis, "Aplikasi Berbasis Website Untuk Mendeteksi Status Gizi Balita Menggunakan Metode K-Nearest Neighbors ( KNN )," vol. 5, no. 1, pp. 44–55, 2024, doi.org/10.61628/jsce.v5i1.1081.

[16] I. Iskandar and D. Jollyta, "Penerapan Algoritma K-Means Untuk Clusterisasi Kasus Stunting Di Provinsi Riau," vol. 6, no. 1, pp. 18–23, 2024, doi.org/10.35145/jmapteksi.v6i1.4130.

[17] A. Stunting, "PENERAPAN METODE K-MEANS UNTUK ANALISIS STUNTING GIZI PADA BALITA :," vol. 2, no. 1, 2022, doi: 10.20885/snati.v2i1.18.

[18] N. Made and D. Kansa, "Implementasi K-Means Untuk Pengelompokan Status Gizi Balita ( Studi Kasus Banjar Titih ) Implementation of K-Means for Clustering the Nutritional Status of Toddlers ( Banjar Titih Case Study )," vol. 1, no. 2, pp. 92–101, 2021, doi: 10.25008/janitra.

[19] A. R. Romadhoni, A. Romadhoni, D. Taqiyyudin, A. Abid, and U. P. Bangsa, "PENERAPAN ALGORITMA K-MEANS DALAM BAYI DAN BALITA DI KOTA SEMARANG Wawasan dan Rencana Pemecahan Masalah," vol. 3, no. 1, pp. 32–41, 2024, https://jurnal.universitasputrabangsa.ac.id/index.php/ijasta/article/view/834/377.

[20] A. Ali, "Clustering Data Antropometri Balita Untuk Menentukan Status Gizi Balita Di Kelurahan Jumput Rejo Sukodono Sidoarjo," JATISI (Jurnal Tek. Inform. dan Sist. Informasi), vol. 7, no. 3, pp. 395–407, 2020, doi: 10.35957/jatisi.v7i3.530.

[21] H. Sulastri, H. Mubarok, J. Informatika, F. Teknik, and U. Siliwangi, "Implementasi Algoritma Machine Learning Untuk Penentuan Cluster Status Gizi Balita," vol. 5, no. 2, pp. 184–191, 2021, doi.org/10.30872/jurti.v5i2.6779.

[22] S. Dewi, Y. Syahra, and F. Taufik, "Pengelompokan Status Gizi Pada Anak Usia 3–5 Tahun Menggunakan Metode K-Means Clustering," J. Sist. Inf. Triguna Dharma (JURSI TGD), vol. 3, no. 2, pp. 201–212, 2024, [Online]. Available: http://ojs.trigunadharma.ac.id/index.php/jsi/article/view/5939

[23] A. K-means, A. Fadilah, M. N. P, S. Lumbanbatu, and S. Defiyanti, "Pengelompokan kabupaten/kota di indonesia berdasarkan faktor penyebab stunting pada balita menggunakan algoritma k-means," vol. 6, no. 2, pp. 223–230, 2022, doi.org/10.26798/jiko.v6i2.581.

ISSN 2085-4579

# E-Service Quality Analysis of the MyTelkomsel Application Using CSI, IPA, and PGCV

Ananda Khoirunnisa[1], Dwi Rosa Indah[2*], Ardina Ariani[3] , Ari Wedhasmara[4], Naretha Kawadha Pasemah Gumay[5], M. Rudi Sanjaya[6], Mukhlis Febriady[7]
[1-7] Department of Information System, Universitas Sriwijaya, Palembang, Indonesia
indah812@unsri.ac.id

Accepted on November 22nd, 2025
Approved on May 5th, 2026

***Abstract*— This study evaluates the e-service quality of the MyTelkomsel application in response to ongoing user Ccerns, particularly slow service response and the limited effectiveness of the automated support system. A structured questionnaire based on the five SERVQUAL dimensions was used and demonstrated strong validity and reliability. Data were collected from one hundred users in Java and Sumatera. The analysis combined three methods: the Customer Satisfaction Index, Importance Performance Analysis, and Potential Gain in Customer Value. The findings show that the application achieved a Customer Satisfaction Index score of 67.75 percent, indicating that users are generally satisfied but expect further improvement. The Importance Performance Analysis recorded a suitability level of 74.54 percent, with several attributes placed in the high-importance low-performance quadrant, including login security, chatbot speed, complaint handling, and personal data protection. The Potential Gain in Customer Value results indicate that chatbot-related attributes and transaction reliability provide the highest potential for increasing customer value. Overall, the study highlights specific service attributes that require priority enhancement to strengthen user satisfaction and service quality. Furthermore, the identified service gaps suggest that chatbot-related issues are not solely operational but are also associated with inherent limitations of conventional rule-based or intent-based chatbot systems, which often lack contextual understanding, adaptability, and effective problem-solving capabilities. This indicates the need for more advanced, human-centered AI approaches to improve service performance and user experience.**

***Index Terms*— e-service quality; customer satisfaction; importance performance analysis; potential gain in customer value; servqual**

## I. INTRODUCTION

The current landscape, characterized by rapid digital transformation, underscores the critical role of the telecommunications sector in driving national digitalization initiatives, such as the Indonesia Digital Network program [1]. This technological advancement, coupled with infrastructural progress like the initiation of 5G network implementation [2], has generated novel means of communication and daily operation [3]. As the dominant mobile telecommunications provider in Indonesia, serving an extensive customer base exceeding 170 million [4], PT Telekomunikasi Selular (Telkomsel) introduced and subsequently developed the MyTelkomsel application. This application was originally launched to streamline customer-company interactions and has since evolved into an integrated "superapp" consolidating diverse service streams [5].

Despite its strategic evolution, the MyTelkomsel application is consistently subjected to significant customer discontent pertaining to its e-service quality. The fundamental challenge revolves around the operational efficiency of customer support and issue resolution. Frequent complaints detail protracted response times and a documented inadequacy of the automated chatbot, Veronika, whose standardized replies often fail to provide substantive solutions. This frequently compels customers to seek assistance via less convenient physical channels [6]. While chatbots are widely implemented as cost-efficient customer service solutions, traditional rule-based or intent-based chatbots often lack contextual understanding, reasoning capability, and adaptive interaction. As a result, user dissatisfaction may originate not merely from service execution, but from structural technological constraints embedded in conventional chatbot architectures. Research on e-service quality has widely emphasized its importance in shaping user satisfaction and loyalty, especially within digital platforms and mobile applications. Demonstrated this relationship in the context of internet banking, showing that higher e-service quality significantly increases customer satisfaction, which subsequently strengthens loyalty. Their findings highlight the critical role of service dimensions such as responsiveness, reliability, and system functionality in determining user

perceptions within digital environments [7]. Similar conclusions are found in studies evaluating various digital applications, reinforcing the relevance of assessing e-service quality for platforms like MyTelkomsel.

In the area of customer satisfaction measurement, several studies have employed the Customer Satisfaction Index (CSI) to quantify overall user satisfaction levels. Previous research used CSI to evaluate logistics service quality and found that customers were generally satisfied, although several key service attributes still required improvement [8]. Another study applied CSI to assess service strategies on digital platforms and reported positive satisfaction levels, yet noted that certain underperforming components continued to influence the overall user experience. Collectively, these findings highlight the value of CSI as a standardized and reliable quantitative measure for assessing customer satisfaction [9].

Beyond CSI, many researchers have applied Importance–Performance Analysis (IPA) to identify service attributes that require managerial attention. Studying food delivery applications, found that several features considered highly important by users were performing below expectations, placing them in Quadrant I as main priorities for improvement [10]. Similarly, Previous research used IPA to analyze the quality of OVO electronic payment services in Tokopedia and identified specific attributes that needed immediate enhancement. These results show that IPA is effective for mapping performance gaps in digital services [11].

To complement IPA and quantify the potential value of service improvements, prior studies have also utilized the Potential Gain in Customer Value (PGCV) method. Previous research applied PGCV alongside other satisfaction measurement tools, identifying key attributes with the highest potential value gain in higher education services [12]. Another study, demonstrated that the integration of IPA and PGCV could effectively determine customer value priorities in public service institutions[13]. Similarly, previous research confirmed the effectiveness of the PGCV method by identifying underperforming attributes in OVO's service that offered the greatest potential for improvement when optimized [11].

Overall, previous research consistently shows that e-service quality strongly influences user satisfaction and that the combined use of CSI, IPA, and PGCV provides a powerful analytical framework for evaluating performance gaps and prioritizing strategic improvements. However, only a limited number of studies have applied this full combination of methods within the telecommunications sector, especially for evaluating a large-scale mobile application such as MyTelkomsel. Furthermore, prior studies have predominantly focused on methodological approaches, with limited attention to the technological limitations of underlying systems, particularly conventional chatbot architectures. There is still a lack of research that examines e-service quality from an AI-oriented and human-centered perspective, especially in understanding how the constraints of rule-based or intent-based chatbots affect user satisfaction. Therefore, this gap provides a strong rationale for the present study.

The distinct contribution of this research lies in the integrated use of the CSI, IPA, and PGCV models, which together provide a more comprehensive and value-oriented diagnostic assessment compared to studies that rely on a single analytical method. This combined approach produces more actionable insights for improving service performance. The sampling frame is intentionally focused on users in Java and Sumatera, which together account for more than 70% of Indonesia's population [14] and record some of the country's highest levels of digital adoption, with internet penetration rates of 83.64% in Java and 77.34% in Sumatera [15]. This regional focus enhances the representativeness and relevance of the findings and enables Telkomsel to formulate more targeted and effective strategies for improving the e-service quality of the MyTelkomsel application.

## II. METHODOLOGY


Fig. 1. Methodological Flowchart

Figure 1. illustrates the methodological flowchart employed in this study. Figure 1. presents the research workflow applied to assess the service quality of the MyTelkomsel application using the Customer Satisfaction Index (CSI), Importance Performance Analysis (IPA), and Potential Gain in Customer Value (PGCV) methods. The following subsections explain each stage in detail.

*A. Planning*

1. Literature Review

The study begins with a review of key theories and prior research on service quality, customer satisfaction, and analytical methods such as CSI, IPA, and PGCV, forming the conceptual and theoretical foundation for the analysis.

2. Problem Identification

After the literature review, the researcher identifies key issues related to user perceptions and satisfaction with MyTelkomsel's service quality, which then form the basis of the research problem.

3. Determining Research Objectives and Benefits

The research objectives are formulated to evaluate the service quality of the MyTelkomsel application based on user assessments, while both practical and academic contributions are outlined to emphasize the study's significance.

4. Determining Population and Sample Size

The research population consists of all users of the MyTelkomsel application. A population is a generalization area containing subjects or objects with specific characteristics determined by the researcher, while a sample represents a portion of that population with similar characteristics, symbolized by *n*. Several sampling techniques may be applied to determine an appropriate sample size [16]. In this study, the Slovin formula was used because it is commonly applied in quantitative research and does not require complex statistical tables. Based on data from the Central Bureau of Statistics , the population of Sumatera is 62,259,500 and Java is 158,079,100, resulting in a combined population of 220,368,600 [14].

Using the Slovin formula with a 10 percent error tolerance, the calculation is as follows. The use of a 10% error margin is considered acceptable given the exploratory and diagnostic nature of this study, which aims to identify general patterns and service gaps rather than achieve high-precision estimation.

$$n = \frac{N}{1 + Ne^2} \quad (1)$$

Description:

*n* : sample size
*N* : population size
*e* : error tolerance limit (10% or 0,1)

$$n = \frac{220.368.600}{1 + 220.368.600\ (0{,}1)^2}$$

$$= \frac{220.368.600}{2.203.687} = 99{,}98$$

Thus, the minimum required sample is 100 respondents. This number is deemed sufficient to represent the population while anticipating incomplete or invalid data. Respondents were selected using purposive sampling with the inclusion criterion of being active MyTelkomsel users for at least one month.

Sampling refers to the process of selecting a subset of elements from a population [17]. This study applies Proportionate Stratified Random Sampling, which divides the population into proportional strata and selects samples randomly within each stratum. The unit of analysis consists of active MyTelkomsel users who meet the established inclusion criteria for the CSI, IPA, and PGCV assessments.

The formula for determining the number of samples in each stratum is as follows [18]:

$$ni = \frac{Ni}{N} \times n \quad (2)$$

*ni* : sample in each stratum
*n* : total sample (100)
*Ni* : population of each stratum
*N* : total population (220,368,600)

The proportional distribution of samples across Java and Sumatera is shown below:

- Java Island

1. DKI Jakarta:

$$ni = \frac{10.678.000}{220.368.600} \times 100$$
$$ni = 4{,}84 \approx 5$$

2. West Java:

$$ni = \frac{50.759.000}{220.368.600} \times 100$$
$$ni = 23{,}03 \approx 23$$

3. Central Java:

$$ni = \frac{38.233.900}{220.368.600} \times 100$$
$$ni = 17{,}34 \approx 17$$

4. DI Yogyakarta:

$$ni = \frac{3.781.500}{220.368.600} \times 100$$
$$ni = 1{,}71 \approx 2$$

5. East Java:

$$ni = \frac{42.089.300}{220.368.600} \times 100$$
$$ni = 19{,}09 \approx 19$$

6. Banten:

$$ni = \frac{12.537.400}{220.368.600} \times 100$$
$$ni = 5{,}68 \approx 6$$

- Sumatera Island

1. Aceh:

$$ni = \frac{5.626.000}{220.368.600} \times 100$$
$$ni = 2{,}55 \approx 2$$

2. North Sumatera:

$$ni = \frac{15.785.800}{220.368.600} \times 100$$
$$ni = 7{,}16 \approx 7$$

3. West Sumatera:

$$ni = \frac{5.914.300}{220.368.600} \times 100$$
$$ni = 2{,}68 \approx 3$$

4. Riau:

$$ni = \frac{6.811.200}{220.368.600} \times 100$$
$$ni = 3{,}09 \approx 3$$

5. Jambi:

$$ni = \frac{3.768.500}{220.368.600} \times 100$$
$$ni = 1{,}71 \approx 2$$

6. South Sumatera:

$$ni = \frac{8.928.500}{220.368.600} \times 100$$
$$ni = 4{,}05 \approx 4$$

7. Bengkulu:

$$ni = \frac{2.138.000}{220.368.600} \times 100$$
$$ni = 0{,}97 \approx 1$$

8. Lampung:

$$ni = \frac{9.522.900}{220.368.600} \times 100$$
$$ni = 4{,}32 \approx 4$$

9. Bangka Belitung Islands:

$$ni = \frac{1.550.800}{220.368.600} \times 100$$
$$ni = 0{,}70 \approx 1$$

10. Riau Islands:

$$ni = \frac{12.537.400}{220.368.600} \times 100$$
$$ni = 1{,}00 \approx 1$$

This proportional allocation ensures that the sample reflects the distribution of the population across both islands.

## B. *Design*

### 1. Questionnaire Development Based on SERVQUAL Dimensions

According to Lovelock and Wirtz, service quality is a long-term cognitive evaluation by consumers regarding the performance of the services provided by a company [19]. E-service quality refers to the quality of services delivered through internet-based platforms, enabling interactions between sellers and buyers so that transactions and shopping activities can be carried out more effectively and efficiently [20]. The questionnaire is constructed based on the five SERVQUAL dimensions Tangibles, Reliability, Responsiveness, Assurance, and Empathy and measures two aspects: importance and performance [21].

TABLE I. RESEARCH QUESTIONNAIRE INDICATORS

| Dimension | Code | Variable |
|---|---|---|
| *Tangibles* | A1 | The application's user interface (UI) is visually appealing and easy to understand. |
| | A2 | Navigation within the application is easy to use and not confusing. |
| | A3 | Service information and features in the application are displayed clearly and informatively. |
| | A4 | The application rarely experiences bugs or errors during use. |
| | A5 | The visual quality of the application meets current digital application standards. |
| *Reliability* | B1 | The application can be accessed at any time without significant disruptions. |
| | B2 | The features in the application function as promised. |
| | B3 | Billing and payment history is displayed accurately and completely. |
| | B4 | Notifications from the application are delivered in a timely and relevant manner. |
| | B5 | The login system and application security operate properly without errors. |
| *Responsiveness* | C1 | The application provides a responsive and easily accessible issue-reporting feature. |
| | C2 | Responses to complaints submitted through the application are quick and solution-oriented. |
| | C3 | The application is updated regularly to improve functionality. |
| | C4 | The application displays real-time status updates on complaint resolution. |
| | C5 | Detailed instructions for using the application's features are provided. |
| *Assurance* | D1 | Users' personal data is securely stored within the application. |
| | D2 | Transactions within the application (such as payments) are safe and trustworthy. |
| | D3 | Privacy policies and terms of use are easy to access. |
| | D4 | Users feel comfortable using the application without concerns about data leaks. |
| | D5 | The application provides transparent information regarding user rights and obligations. |
| *Empathy* | E1 | The application offers active and responsive customer support or assistance features. |
| | E2 | The application provides offers or promotions relevant to users' needs. |
| | E3 | The application adjusts language and content based on users' location or profile. |
| | E4 | User feedback is followed up by the service provider. |

| | | |
|---|---|---|
| | E5 | User experience is prioritized in the development of new features. |

2. Content Validity Assessment

Content validity refers to the extent to which each element of an assessment instrument is relevant and adequately represents the intended construct [22]. A highly valid instrument is essential to ensure scientifically reliable research results [23]. According to Suryadi, several methods are commonly used to evaluate content validity, including I-CVI, S-CVI/Ave, S-CVI/UA, CVR, and CVI [24]. Before distribution, the questionnaire undergoes a content validity test (expert judgment) to ensure clarity, relevance, and representativeness of each item. Validation is carried out by the academic supervisor and subject-matter experts in digital service quality and e-commerce.

1) I-CVI (*Item-Content Validity Index*)

$$I-CVI = \frac{Number\ of\ Approved\ Items}{Number\ of\ Reviewers} \quad (3)$$

2) S-CVI/Ave (*Scale-Content Validity Index/Average*)

$$S-\frac{CVI}{Ave} = \frac{\sum I-CVI}{Number\ of\ Questions} \quad (4)$$

3) S-CVI/UA (*Scale-Content Validity Index/Universal Agreement*)

$$S-CVI/UA = \frac{\sum Skor\ UA}{Jumlah\ Item\ Pertanyaan} \quad (5)$$

4) CVR (*Content Validity Ratio*)

$$CVR = \frac{ne-\frac{n}{2}}{\frac{N}{2}} \quad (6)$$

5) CVI (*Content Validity Index*)

$$CVI = \frac{\sum CVR}{Number\ of\ Questions} \quad (7)$$

3. Data Collection

Validated questionnaires are distributed to qualified respondents to gather data on user perceptions and experiences when using the MyTelkomsel application.

*C. Analysis*

1. Validity and Reliability Testing

All collected data undergo validity and reliability tests to ensure the measurement instrument is accurate and consistent. Items that do not meet the criteria are revised accordingly.

2. Data Processing Based on Respondent Demographics

Valid and reliable responses are processed by classifying demographic characteristics such as province and duration of MyTelkomsel usage. This supports a more contextual interpretation of the findings.

3. Analytical Procedures

Three analytical methods are used:

- Customer Satisfaction Index (CSI)
  Customer Satisfaction Index (CSI) is a qualitative measure expressed as the percentage of customers who report satisfaction in a customer satisfaction survey [25]. CSI is used to evaluate satisfaction by considering both the importance and performance of service attributes. According to Aritonang, the steps for calculating CSI include [26]:

- Mean Importance Score (MIS)
  Average importance rating given by respondents:

$$MIS = \frac{\sum_{i=1}^{n} Y_i}{n} \quad (8)$$

- Weight Factor (WF)
  The proportion of each MIS relative to the total MIS:

$$WF = \frac{MIS_i}{\sum_{i=1}^{n} MIS_i} \quad (9)$$

- Weight Score (WS)
  Calculated by multiplying WF by the Mean Satisfaction Score (MSS):

$$WS = WF_i \times MSS \quad (10)$$

- Customer Satisfaction Index (CSI)
  Overall satisfaction level based on weighted scores:

$$CSI = \frac{\sum_{k=1}^{p} WSk}{HS} \times 100\% \quad (11)$$

TABLE II. CSI INDEX TABLE

| CSI Value | Category |
|---|---|
| 81% - 100% | Very Satisfied |
| 66% - 80,99% | Satisfied |
| 51% - 65,99% | Fairly Satisfied |
| 35% - 50,99% | Less Satisfied |
| 0% - 34,99% | Not Satisfied |

*Source*: Aritonang (2005)

- Importance Performance Analysis (IPA)
  The Importance Performance Analysis (IPA) method is used to determine the level of alignment between user expectations and perceived performance, which is then mapped onto a Cartesian diagram [27]. IPA combines the measurement of importance and performance into a two-dimensional analysis, allowing easier interpretation and practical recommendations [28]. In this study, variable X represents the perceived performance of MyTelkomsel's service quality, while variable Y reflects the level of importance or user expectations.
  According to Tjiptono, the IPA framework consists of four strategic quadrants [29]:
  1) Quadrant I (Concentrate Here): High importance, low performance; attributes requiring immediate managerial attention.

2) Quadrant II (Keep Up the Good Work): High importance, high performance; attributes that should be maintained.
3) Quadrant III (Low Priority): Low importance, low performance; attributes with minimal impact on users.
4) Quadrant IV (Possible Overkill): Low importance, high performance; attributes performed well but considered less essential by users.

- Potential Gain in Customer Value (PGCV)
  The analysis of importance and performance helps identify priority areas for improvement [30]. This approach allows Importance and Performance scores to be compared quantitatively in a more detailed manner. The steps of the PGCV index are as follows:
  - Achieved Customer Value (ACV)
    ACV reflects the value perceived by customers based on their assessments of performance and importance.
    $ACV = \bar{X} \times \bar{Y}$ (12)
    $\bar{X}$ : average performance score
    $\bar{Y}$ : average importance score
  - Ultimately Desired Customer Value (UDCV)
    UDCV represents the value ideally expected by customers. It is obtained by multiplying the average importance score by the maximum Likert scale value.
    $UDCV = \bar{Y} \times P\,Max$ (13)
    $\bar{Y}$ : average importance score
    $Pmax$ : highest Likert scale value
  - PGCV Index
    The PGCV value indicates the gap between desired and achieved value. Higher PGCV values reflect attributes that require higher priority improvements.
    $PGCV\ Index = UDCV - ACV$ (14)
    This index helps identify which service attributes or features should be improved first to enhance overall customer satisfaction.

### D. Evaluation

This stage presents the outcomes of the analytical processes conducted in the study. The results include the calculated Customer Satisfaction Index, the distribution of service attributes across the four Importance Performance Analysis quadrants, and the priority values generated through the PGCV index. These findings reflect users' perceived performance, expectations, and the attributes requiring improvement within the MyTelkomsel application.

## III. Result and Discussion

This section presents the empirical findings of the study in accordance with the research stages described previously, namely Planning, Design, Data Collection, Analysis, and Evaluation. The results cover the content validity process, respondent characteristics, validity and reliability testing, and the outcomes of the CSI, IPA, and PGCV analyses for the MyTelkomsel application.

### A. Planning

The planning stage produced a final sampling design and a clear profile of the respondents involved in this study. Questionnaires were distributed to 100 users of the MyTelkomsel application residing in Java and Sumatera.

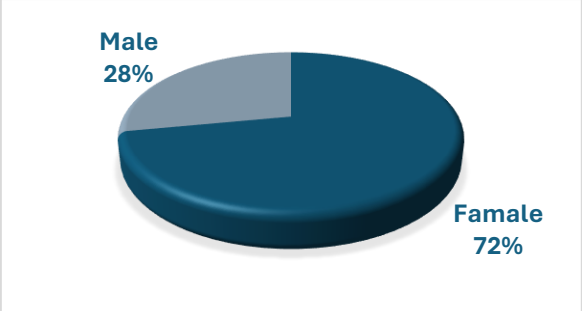


Fig. 2. Gender of Customer Respondents

Based on Figure 2., 28 respondents (28%) were male and 72 respondents (72%) were female.

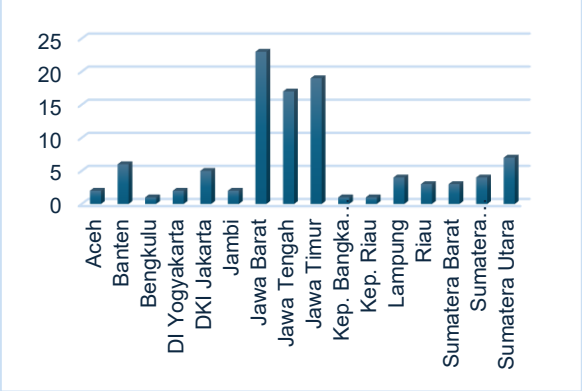


Fig. 3. Respondent Domicile Distribution

Figure 3. shows that all respondents live in provinces located in Java and Sumatera, with the highest proportion coming from West Java Province, while the remaining respondents are fairly distributed across other provinces. This pattern is consistent with the use of Proportionate Stratified Random Sampling, indicating that each province in both islands is proportionally represented.

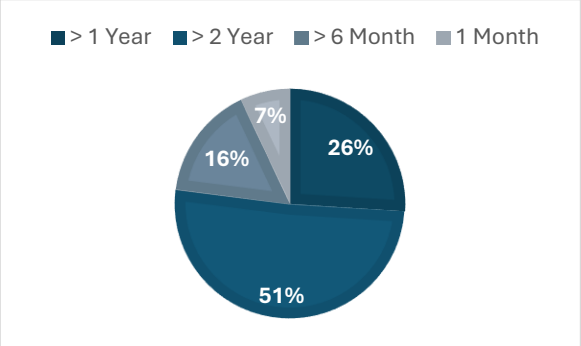


Fig. 4. Distribution of Long-Term Application Usage

Furthermore, Figure 4. describes the length of app usage. The majority of respondents (51%) have used the MyTelkomsel application for more than two years, 26% for more than one year, 16% for more than six months, and 7% for approximately one month. These findings confirm that most respondents are experienced users, so their assessments are highly relevant for measuring perceived service quality and customer satisfaction.

*B. Design*

The design stage focused on developing and refining the measurement instrument through content validity and psychometric testing.

1. Content Validity
   The content validity stage was conducted to ensure that each item in the research instrument adequately represented the intended construct. In this study, content validity was assessed by two experts, Mr. Ari Wedhasmara, S.Kom., M.T.I., and Ms. Naretha Kawadha Pasemah Gumay, S.Si., M.Kom. Both experts evaluated the relevance and clarity of each indicator using a four-point rating scale.

TABLE III. CONTENT VALIDITY CALCULATION RESULTS AFTER REVISION

| Code | Expert 1 | Expert 2 | I-CVI | Agreement |
|---|---|---|---|---|
| A1 | 4 | 4 | 1 | Accepted |
| A2 | 4 | 4 | 1 | Accepted |
| A3 | 4 | 4 | 1 | Accepted |
| A4 | 4 | 4 | 1 | Accepted |
| A5 | 4 | 4 | 1 | Accepted |
| B1 | 4 | 4 | 1 | Accepted |
| B2 | 4 | 4 | 1 | Accepted |
| B3 | 4 | 4 | 1 | Accepted |
| B4 | 4 | 4 | 1 | Accepted |
| B5 | 4 | 4 | 1 | Accepted |
| C1 | 4 | 4 | 1 | Accepted |
| C2 | 4 | 4 | 1 | Accepted |
| C3 | 4 | 4 | 1 | Accepted |
| C4 | 4 | 4 | 1 | Accepted |
| C5 | 4 | 4 | 1 | Accepted |
| D1 | 4 | 4 | 1 | Accepted |
| D2 | 4 | 4 | 1 | Accepted |
| D3 | 4 | 4 | 1 | Accepted |
| D4 | 4 | 4 | 1 | Accepted |
| D5 | 4 | 4 | 1 | Accepted |
| E1 | 4 | 4 | 1 | Accepted |
| E2 | 1 | 4 | 1 | Accepted |
| E3 | 4 | 4 | 1 | Accepted |
| E4 | 4 | 4 | 1 | Accepted |
| S-CVI | | | 1 | Accepted |

After incorporating expert feedback and completing two rounds of revisions, the instrument was re-evaluated. The results in Table 3. show that all indicators achieved I-CVI = 1.00 and S-CVI = 1.00, indicating perfect agreement between experts that the items are clear, relevant, and accurately represent their respective constructs. Therefore, the final questionnaire fully meets the criteria for strong content validity and is appropriate for subsequent testing.

2. Instrument Structure
   The finalized questionnaire consists of 24 indicators distributed across the five SERVQUAL dimensions (Tangibles, Reliability, Responsiveness, Assurance, and Empathy), each measured with importance and performance/satisfaction scales using a five-point Likert format. This structure directly supports the application of the CSI, IPA, and PGCV methods in subsequent stages.

3. Data Collection

   Data collection produced a complete distribution of responses for each indicator on both importance and satisfaction dimensions.

TABLE IV. DISTRIBUTION OF RESPONDENTS' ANSWERS LEVEL OF IMPORTANCE

| Code | TP 1 | KP 2 | CP 3 | P 4 | SP 5 | *Scale* |
|---|---|---|---|---|---|---|
| A1 | 1 | 2 | 7 | 33 | 56 | *Tangibles* |
| A2 | 0 | 0 | 7 | 34 | 59 | *Tangibles* |
| A3 | 0 | 1 | 7 | 28 | 64 | *Tangibles* |
| A4 | 0 | 1 | 6 | 26 | 67 | *Tangibles* |
| A5 | 0 | 0 | 4 | 37 | 59 | *Tangibles* |
| B1 | 0 | 1 | 13 | 18 | 68 | *Reliability* |
| B2 | 0 | 1 | 6 | 25 | 68 | *Reliability* |
| B3 | 1 | 6 | 10 | 23 | 60 | *Reliability* |
| B4 | 0 | 2 | 9 | 29 | 60 | *Reliability* |
| B5 | 0 | 0 | 14 | 21 | 65 | *Reliability* |
| C1 | 0 | 0 | 10 | 32 | 58 | *Responsiveness* |
| C2 | 0 | 0 | 10 | 29 | 61 | *Responsiveness* |
| C3 | 0 | 2 | 9 | 25 | 64 | *Responsiveness* |
| C4 | 0 | 0 | 12 | 27 | 61 | *Responsiveness* |
| C5 | 0 | 0 | 10 | 27 | 63 | *Responsiveness* |
| D1 | 0 | 2 | 5 | 21 | 72 | *Assurance* |
| D2 | 0 | 0 | 6 | 23 | 71 | *Assurance* |
| D3 | 0 | 0 | 7 | 24 | 69 | *Assurance* |
| D4 | 0 | 0 | 8 | 28 | 64 | *Assurance* |
| D5 | 0 | 0 | 8 | 28 | 64 | *Assurance* |
| E1 | 0 | 0 | 7 | 26 | 67 | *Empathy* |
| E2 | 0 | 0 | 6 | 29 | 65 | *Empathy* |
| E3 | 0 | 0 | 8 | 30 | 62 | *Empathy* |
| E4 | 0 | 0 | 9 | 23 | 68 | *Empathy* |

Table 4. presents the frequency distribution of importance scores. Most responses are concentrated in the "important" (4) and "very important" (5) categories across all attributes, confirming that the selected indicators are indeed perceived as relevant by users.

TABLE V. DISTRIBUTION OF RESPONDENTS' ANSWERS LEVEL OF SATISFACTION

| Code | TP | KP | CP | P | SP | *Scale* |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | |
| A1 | 8 | 10 | 13 | 48 | 21 | *Tangibles* |
| A2 | 6 | 13 | 10 | 52 | 19 | *Tangibles* |
| A3 | 7 | 7 | 12 | 48 | 26 | *Tangibles* |
| A4 | 5 | 12 | 9 | 40 | 34 | *Tangibles* |
| A5 | 7 | 7 | 11 | 50 | 25 | *Tangibles* |
| B1 | 9 | 12 | 20 | 39 | 20 | *Reliability* |
| B2 | 8 | 13 | 8 | 49 | 22 | *Reliability* |
| B3 | 15 | 22 | 16 | 31 | 16 | *Reliability* |
| B4 | 9 | 10 | 16 | 36 | 29 | *Reliability* |
| B5 | 12 | 16 | 18 | 33 | 21 | *Reliability* |
| C1 | 28 | 12 | 14 | 31 | 15 | *Responsiveness* |
| C2 | 29 | 10 | 18 | 27 | 16 | *Responsiveness* |
| C3 | 30 | 11 | 13 | 29 | 17 | *Responsiveness* |
| C4 | 30 | 11 | 17 | 26 | 16 | *Responsiveness* |
| C5 | 29 | 12 | 12 | 30 | 17 | *Responsiveness* |
| D1 | 9 | 28 | 12 | 28 | 23 | *Assurance* |
| D2 | 9 | 10 | 12 | 37 | 32 | *Assurance* |
| D3 | 10 | 10 | 15 | 37 | 28 | *Assurance* |
| D4 | 13 | 25 | 13 | 29 | 20 | *Assurance* |
| D5 | 8 | 15 | 12 | 41 | 24 | *Assurance* |
| E1 | 29 | 9 | 12 | 29 | 21 | *Empathy* |
| E2 | 12 | 10 | 8 | 31 | 39 | *Empathy* |
| E3 | 13 | 25 | 25 | 24 | 23 | *Empathy* |
| E4 | 9 | 26 | 12 | 28 | 25 | *Empathy* |

Table 5. shows the distribution of satisfaction scores. Compared to the importance dimension, satisfaction responses are more dispersed, with noticeable portions in the "neutral" and "less satisfied" categories for several attributes, especially within the Responsiveness dimension. This difference between importance and satisfaction distributions suggests potential gaps between user expectations and perceived performance, which are explored further using CSI, IPA, and PGCV. In this context, chatbot-related attributes reflect the effectiveness of automated service intelligence, particularly in delivering responsiveness and perceived empathy within digital interactions.

## C. *Analysis*

1. Content Validity

The validity test was conducted to examine whether each questionnaire item accurately measures the intended construct. Item validity was assessed by comparing the correlation coefficient (r-count) with the critical value (r-table).

TABLE VI. VALIDITY TEST RESULTS (30 RESPONDENTS)

| Code | I | S | R Tabel (5%) | Validity |
|---|---|---|---|---|
| A1 | 0.711 | 0.534 | 0,361 | Valid |
| A2 | 0.849 | 0.759 | 0,361 | Valid |
| A3 | 0.786 | 0.611 | 0,361 | Valid |
| A4 | 0.746 | 0.592 | 0,361 | Valid |
| A5 | 0.810 | 0.727 | 0,361 | Valid |
| B1 | 0.841 | 0.634 | 0,361 | Valid |
| B2 | 0.907 | 0.766 | 0,361 | Valid |
| B3 | 0.728 | 0.913 | 0,361 | Valid |
| B4 | 0.789 | 0.822 | 0,361 | Valid |
| B5 | 0.747 | 0.920 | 0,361 | Valid |
| C1 | 0.836 | 0.847 | 0,361 | Valid |
| C2 | 0.865 | 0.853 | 0,361 | Valid |
| C3 | 0.886 | 0.905 | 0,361 | Valid |
| C4 | 0.854 | 0.899 | 0,361 | Valid |
| C5 | 0.805 | 0.730 | 0,361 | Valid |
| D1 | 0.825 | 0.517 | 0,361 | Valid |
| D2 | 0.880 | 0.781 | 0,361 | Valid |
| D3 | 0.925 | 0.865 | 0,361 | Valid |
| D4 | 0.869 | 0.440 | 0,361 | Valid |
| D5 | 0.884 | 0.812 | 0,361 | Valid |
| E1 | 0.839 | 0.810 | 0,361 | Valid |
| E2 | 0.816 | 0.697 | 0,361 | Valid |
| E3 | 0.871 | 0.759 | 0,361 | Valid |
| E4 | 0.892 | 0.608 | 0,361 | Valid |

TABLE VII. VALIDITY TEST RESULTS (100 RESPONDENTS)

| Code | I | S | R Tabel (5%) | Validity |
|---|---|---|---|---|
| A1 | 0.506 | 0.744 | 0,196 | Valid |
| A2 | 0.743 | 0.838 | 0,196 | Valid |
| A3 | 0.660 | 0.818 | 0,196 | Valid |
| A4 | 0.657 | 0.757 | 0,196 | Valid |
| A5 | 0.720 | 0.799 | 0,196 | Valid |
| B1 | 0.689 | 0.753 | 0,196 | Valid |
| B2 | 0.731 | 0.864 | 0,196 | Valid |
| B3 | 0.746 | 0.726 | 0,196 | Valid |
| B4 | 0.802 | 0.735 | 0,196 | Valid |
| B5 | 0.850 | 0.747 | 0,196 | Valid |
| C1 | 0.841 | 0.788 | 0,196 | Valid |
| C2 | 0.864 | 0.837 | 0,196 | Valid |
| C3 | 0.863 | 0.830 | 0,196 | Valid |
| C4 | 0.874 | 0.839 | 0,196 | Valid |
| C5 | 0.829 | 0.827 | 0,196 | Valid |
| D1 | 0.675 | 0.852 | 0,196 | Valid |
| D2 | 0.796 | 0.843 | 0,196 | Valid |
| D3 | 0.829 | 0.826 | 0,196 | Valid |
| D4 | 0.700 | 0.848 | 0,196 | Valid |
| D5 | 0.779 | 0.856 | 0,196 | Valid |
| E1 | 0.768 | 0.817 | 0,196 | Valid |
| E2 | 0.745 | 0.764 | 0,196 | Valid |
| E3 | 0.819 | 0.889 | 0,196 | Valid |
| E4 | 0.726 | 0.835 | 0,196 | Valid |

A preliminary test with 30 respondents showed that all 24 items had r-count > 0.361, indicating that every item was valid. The main validity test with 100 respondents also confirmed that all indicators had r-count > 0.196 with $p < 0.05$ for both importance and satisfaction dimensions. These results demonstrate that all questionnaire items are

statistically valid and suitable for use in the CSI, IPA, and PGCV analyses.

2. Reliability Test

Reliability was assessed using Cronbach's Alpha to evaluate internal consistency. An instrument is considered reliable when $\alpha > 0.60$.

TABLE VIII. RELIABILITY TEST RESULTS (30 RESPONDENTS)

| Variable | Cronbach Alpha | N of Items | Desc |
|---|---|---|---|
| **Importance** | 0,963 | 24 | Reliable |
| **Satisfaction** | 0,980 | 24 | Reliable |

TABLE IX. RELIABILITY TEST RESULTS (100 RESPONDENTS)

| Variable | Cronbach Alpha | N of Items | Desc |
|---|---|---|---|
| **Importance** | 0,967 | 24 | Reliable |
| **Satisfaction** | 0,977 | 24 | Reliable |

In the preliminary test (30 respondents), Cronbach's Alpha values were 0.963 (importance) and 0.980 (satisfaction). In the main test (100 respondents), reliability remained high with $\alpha = 0.967$ (importance) and 0.977 (satisfaction). All coefficients far exceed the minimum threshold, indicating that the instrument is highly reliable and consistent in me asuring both user importance and satisfaction toward MyTelkomsel's service quality.

3. Customer Satisfaction Index (CSI)

CSI was computed using the Mean Importance Score (MIS), Weight Factor (WF), Mean Satisfaction Score (MSS), and Weighted Score (WS) for each attribute. The total WS is 338.77. With a maximum Likert scale of 5, the CSI is:

$$CSI = \frac{338,77}{5} \times 100\% = 67,75 \times 100\% = 67,75$$

According to Aritonang's CSI classification, a score of 67.75% falls into the "Satisfied" category (66%–80.99%). This indicates that, overall, users in Java and Sumatera are satisfied with the e-service quality of the MyTelkomsel application. Nevertheless, the score has not yet reached the "very satisfied" category, implying that there is still room for improvement, particularly on attributes with high importance but relatively lower performance.

4. Importance Performance Analysis (IPA)

a) Attribute Suitability Level Analysis

Based on the calculation of the level of conformity between the level of importance and the level of performance, the average level of conformity was 74.54%. which falls into the "appropriate" category (70%–79%).

- Attributes with the highest conformity include A4 (84.10%) and A3/A5 (83.30%), showing that ease of use and clarity of information are well aligned with user expectations.
- The lowest conformity is found in C4 (64.06%), followed by C2 (64.38%) and C3 (64.60%), all related to the Responsiveness of the Chatbot Veronica and help features.

b) Cartesian Diagram Analysis

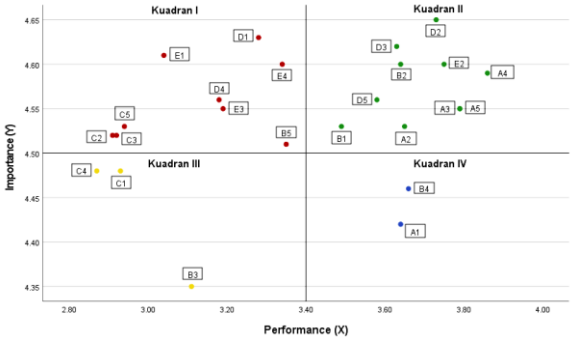


Fig. 5. Cartesian Diagram

The Cartesian diagram in Figure 4.4 then classifies attributes into four quadrants:

1. Quadrant I – Main Priorities: B5, C2, C3, C5, D1, D4, E1, E3, and E4 (high importance, low performance). These attributes involve login and security reliability, chatbot speed and accuracy, completeness of guidance, data protection, and follow-up on feedback. They require urgent improvement.
2. Quadrant II – Maintain Performance: A2, A3, A4, A5, B1, B2, D2, D3, D5, E2, and E5 (high importance, high performance), which should be maintained through continuous monitoring and system maintenance.
3. Quadrant III – Low Priority: B3, C1, and C4 (low importance, low performance). Improvements here can be carried out gradually.
4. Quadrant IV – Possible Overinvestment: A1 and B4 (low importance, high performance), indicating that current performance is already very good relative to their perceived importance.

c) Analysis of Chatbot Performance Limitations

The consistent underperformance of chatbot-related attributes in Quadrant I

indicates that the issue is not solely operational but also structural in nature. Conventional chatbot systems, which are typically rule-based or intent-based, have limited capability in understanding user context, handling complex queries, and providing adaptive responses. As a result, users often receive generic, repetitive, or irrelevant responses, particularly when their inquiries fall outside predefined scenarios. This limitation directly affects perceived responsiveness, accuracy, and empathy key dimensions in e-service quality. Therefore, the improvement of chatbot performance requires not only operational enhancements but also the adoption of more advanced conversational AI technologies that enable context-aware, flexible, and human-centered interactions.

5. Potential Gain in Customer Value (PGCV)

TABLE X. POTENTIAL CUSTOMER VALUE INCREASE (PGCV) RESULTS

| Code | Avg (Y) | Avg (X) | ACV (X*Y) | UDCV (Y*5) | UDCV - ACV |
|---|---|---|---|---|---|
| B5 | 4,51 | 3,35 | 15,10 | 22,55 | 7,4415 |
| C2 | 4,52 | 2,91 | 13,15 | 22,6 | 9,4468 |
| C3 | 4,52 | 2,92 | 13,19 | 22,6 | 9,4016 |
| C5 | 4,53 | 2,94 | 13,31 | 22,65 | 9,3318 |
| D1 | 4,63 | 3,28 | 15,18 | 23,15 | 7,9636 |
| D4 | 4,56 | 3,18 | 14,50 | 22,8 | 8,2992 |
| E1 | 4,61 | 3,04 | 14,01 | 23,05 | 9,0356 |
| E3 | 4,55 | 3,19 | 14,51 | 22,75 | 8,2355 |
| E4 | 4,60 | 3,34 | 15,36 | 23 | 7,636 |

PGCV analysis quantifies the gap between Actual Customer Value (ACV) and Ultimate Desired Customer Value (UDCV) for key attributes (Table 4.10). The larger the difference (UDCV – ACV), the greater the potential gain in customer value if performance is improved.

The highest PGCV values are:

- C2 – Chatbot responsiveness (≤ 1 minute): 9.4468
- C3 – Accuracy of chatbot responses: 9.4016
- C5 – Chatbot's ability to provide satisfactory solutions: 9.3318
- E1 – Smoothness of transactions via MyTelkomsel: 9.0356

These four attributes represent the primary leverage points for increasing customer value. Other attributes with high PGCV values include D4 (feeling safe from data leakage), E3 (up-to-date and relevant information), D1 (security of personal data), E4 (ease of finding features), and B5 (login and security system reliability). Improvements in these areas will significantly enhance perceived value, trust, and satisfaction. The dominance of chatbot-related attributes (C2, C3, and C5) among the highest PGCV values indicates that automated service quality plays a critical role in shaping customer value. This suggests that users place high expectations not only on the speed of chatbot responses but also on their accuracy and problem-solving capabilities. From a technological perspective, these findings highlight the limitations of conventional rule-based or intent-based chatbot systems, which often fail to meet user expectations due to their lack of contextual understanding and adaptability. In contrast, the adoption of more advanced conversational AI can significantly enhance customer value by enabling more intelligent, responsive, and personalized interactions.

### *D. Evaluation*

The results of this study summarize the key findings from the CSI, IPA, and PGCV analytical methods. The CSI score of 67.75% places MyTelkomsel in the "Satisfied" category, indicating that service performance meets user expectations but has not yet reached the "Very Satisfied" level. The IPA results show an average conformity of 74.54%, with the highest alignment found in UI-related attributes and the lowest in attributes related to chatbot responsiveness. The Cartesian diagram identifies several attributes particularly B5, C2, C3, C5, D1, D4, E1, E3, and E4 as main priorities for improvement. PGCV analysis reinforces these findings, with the largest potential gains observed in C2, C3, C5, and E1, indicating that chatbot performance, transaction smoothness, and data security are the attributes most in need of enhancement to increase customer value and satisfaction.

## IV. CONCLUSION

This study delivers an all-encompassing assessment of MyTelkomsel's e-service quality by combining CSI, IPA, and PGCV methods. The 67.75% CSI score shows users are reasonably satisfied, while the 74.54% conformity figure indicates performance generally matches what users expect, though significant shortfalls exist, especially regarding Responsiveness and Security. IPA results pinpoint chatbot performance, data protection, and how complaints are handled as the main improvement priorities. PGCV analysis underscores that enhancing chatbot responsiveness, accuracy, solution quality, and transaction ease would generate the most customer value. The findings suggest that the primary service quality gaps are structurally linked to the limitations of conventional chatbot-based systems. Therefore, future development should not only focus on improving existing chatbot performance but also consider

adopting human-centered conversational AI capable of contextual understanding, adaptive interaction, and effective problem resolution. Addressing these critical areas while maintaining currently strong features will elevate overall satisfaction and bolster MyTelkomsel's reliability. Future efforts should concentrate on regular system upgrades, improved responsiveness, and stronger user-oriented improvements to deliver a more seamless and protected digital service. However, this study has several limitations. First, the data were collected from a limited number of respondents and were geographically restricted to users in Java and Sumatra, thus not fully representing all MyTelkomsel users in Indonesia. Second, the study focuses only on specific service quality dimensions and may not capture other factors influencing user satisfaction, such as pricing or network performance. Therefore, future research is recommended to involve a larger and more diverse sample, expand the study area to include users across Indonesia, incorporate additional variables, and explore comparative studies with other digital service applications to provide more comprehensive insights into e-service quality improvement.

## REFERENCES


[1] F. A. Al Putra, S. G. Prakoso, and N. D. Ardita, "A Literature Study on Network Society in Indonesia 1977-2022," *Komuniti: Jurnal Komunikasi dan Teknologi Informasi*, vol. 16, no. 1, pp. 25–40, Mar. 2024, doi: doi.org/10.23917/komuniti.v16i1.3097.

[2] S. H. Komariah, R. Rd. Saedudin, R. Y. Arumsari, and U. Y. Ksp, "Implementation of 5G Telecommunication Network Services in Indonesia based on Techno-economic Analysis," *International Journal on Informatics Visualization*, vol. 7, no. 4, pp. 2569–2577, Dec. 2023, doi: dx.doi.org/10.62527/joiv.7.4.2447.

[3] Z. Setiawan, N. Jauhar, D. A. Putera, A. D. Santosa, and K. Fenanlampir, *KEWIRAUSAHAAN DIGITAL*. Padang: PT GLOBAL EKSEKUTIF TEKNOLOGI, 2023. [Online]. Available: https://www.researchgate.net/publication/371724102

[4] PT Telkomsel, "Transforming Telecommunications in Indonesia," telkomsel.com.

[5] F. Nugraha, A. R. Dewi, I. Bastian, and P. Korespondensi, "Analisis Tingkat Kepuasan Pengguna Aplikasi MyTelkomsel Menggunakan Evaluasi Heuristik dan Metode PIECES (Studi Kasus : Mahasiswa Kampus Karawaci Universitas Gunadarma)," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, vol. 9, no. 3, pp. 463–468, Jun. 2022, doi: 10.25126/jtiik.202294403.

[6] A. Ibrahim, F. S. Elisa, J. Fernando, L. Salsabila, N. Anggraini, and S. N. Arafah, "Pengaruh E-Service Quality Terhadap Loyalitas Pengguna Aplikasi MyTelkomsel," *Building of Informatics, Technology and Science (BITS)*, vol. 3, no. 3, pp. 302–311, Dec. 2021, doi: 10.47065/bits.v3i3.1076.

[7] I. Sasono *et al.*, "The Impact of E-Service Quality and Satisfaction on Customer Loyalty: Empirical Evidence from Internet Banking Users in Indonesia," *Journal of Asian Finance, Economics and Business*, vol. 8, no. 4, pp. 465–473, 2021, doi: 10.13106/jafeb.2021.vol8.no4.0465.

[8] B. E. Ogunnowo and S. S. Sule, "Measurement of Customer Perceptions of Logistics Service Quality," *Jurnal Teknik Industri*, vol. 22, no. 1, pp. 43–56, Feb. 2021, doi: 10.22219/jtiumm.vol22.no1.43-56.

[9] H. M. Hana, "Customer Satisfaction Analysis through Service Quality for Service Strategy Improvement Mutiareads," *Management Analysis Journal*, vol. 11, no. 1, pp. 39–45, Mar. 2022, [Online]. Available: http://maj.unnes.ac.id

[10] M. Wahyudin, C. C. Chen, H. Yuliando, N. Mujahidah, and K. M. Tsai, "Importance–performance and potential gain of food delivery apps: in view of the restaurant partner perspective," *British Food Journal*, vol. 126, no. 5, pp. 1981–2003, Apr. 2024, doi: 10.1108/BFJ-11-2022-1003.

[11] E. Septiani, E. Pujiyanto, and M. Hisjam, "A Design to Improve the Quality of OVO Electronic Money Payment Services in Tokopedia using IPA and PGCV," *Jurnal Teknik Industri*, vol. 21, no. 2, pp. 153–162, Aug. 2020, doi: 10.22219/jtiumm.vol21.no2.153-162.

[12] T. H. Raharjo, K. B. Mulyono, I. Ismiyati, and A. Jaenudin, "HEISQUAL – ACSI – IPA – PGCV: Synthesis of higher education service satisfaction measurements," *Asian Management and Business Review*, no. 2, pp. 121–137, Aug. 2023, doi: 10.20885/ambr.vol3.iss2.art2.

[13] R. Alfatiyah and S. Bastuti, "Improving the Quality of Service Using the IPA and PGCV Methods at BPJS Kesehatan, South Tangerang," *Jurnal Ilmiah Teknik Industri*, vol. 22, no. 1, pp. 25–32, Jun. 2023, doi: 10.23917/jiti.v22i1.21685.

[14] BPS-Statistics Indonesia, "Statistik Indonesia 2025," 2025.

[15] N. Huda, Dyah, A., Rani, S. E. Bhima, and Y. Adhinegara, "Digital Economy Outlook 2025," Jakarta, Indonesia, 2024. [Online]. Available: www.celios.co.id

[16] Sugiyono, *Metode Penelitian Kuantitatif Kualitatid dam R&D*, 19th ed. Bandung: Alfabeta Bandung, 2013.

[17] U. Sekaran, *Research Methods for Business: A Skill Building Approach*. Hoboken, NJ: John Wiley & Sons, 2006.

[18] S. Natsir, "Ringkasan Disertasi : Pengaruh Gaya Kepemimpinan Terhadap Kinerja Karyawan di Sulawesi Tengah," Universitas Airlangga, Surabaya, 2004.

[19] J. Wirtz and C. Lovelock, *Services Marketing*, 8th ed. USA: World Scientific Publishing Co. Inc., 2016. doi: 10.1142/y0001.

[20] A. Parasuraman, V. A. Zeithaml, and L. L. Berry, "SERVQUAL A Multiple-item Scale for Measuring Consumer Perceptions of Service Quality," *Journal of Retailing*, vol. 64, no. 1, Jan. 1988, [Online]. Available: https://www.researchgate.net/publication/200827786

[21] D. A. Cook and T. J. Beckman, "Current concepts in validity and reliability for psychometric instruments: Theory and application," *Am J Med*, vol. 119, no. 2, pp. 166.e7-166.e16, Mar. 2006, doi: 10.1016/j.amjmed.2005.10.036.

[22] M. S. B. Yusoff, "ABC of Content Validation and Content Validity Index Calculation," *Education in Medicine Journal*, vol. 11, no. 2, pp. 49–54, Jun. 2019, doi: 10.21315/eimj2019.11.2.6.

[23] T. Suryadi, F. Alfiya, M. Yusuf, R. Indah, T. Hidayat, and K. Kulsum, "CONTENT VALIDITY FOR THE RESEARCH INSTRUMENT REGARDING TEACHING METHODS OF THE BASIC PRINCIPLES OF BIOETHICS," *Jurnal Pendidikan Kedokteran Indonesia: The Indonesian Journal of Medical Education*, vol. 12, no. 2, p. 186, Jul. 2023, doi: 10.22146/jpki.77062.

[24] I. gede kt. T. P. Budhi and N. K. Sumiari, "Pengukuran Customer Satisfaction Index Terhadap Pelayanan di Century Gym," *Jurnal Ilmiah SISFOTENIKAJ*, vol. 7, no. 1, pp. 25–37, Jan. 2017.

[25] L. R. Aritonang and R. Lerbin, *Kepuasan Pelanggan*. Gramedia. Jakarta, 2005.

[26] N. Wisudawati, M. G. Irfani, M. Hastarina, and B. Santoso, "Penggunaan Metode Importance Performance Analysis (IPA) Untuk Menganalisis Kepuasan Masyarakat Terhadap Pelayanan Administrasi Kependudukan Kecamatan Lengkiti," *Integrasi Jurnal Ilmiah Teknik Industri*, vol. 8, no. 1, pp. 32–39, 2023, [Online]. Available: http://jurnal.um-palembang.ac.id/index.php/integrasi

[27] T. Moladia and T. Kristiana, "Penerapan Metode Importance Performance Analysis (IPA) untuk

Menganalisis Kualitas Aplikasi Tokopedia Berdasarkan Kepuasan Pelanggan," *KOMPUTA : Jurnal Ilmiah Komputer dan Informatika*, vol. 12, no. 1, pp. 9–18, Apr. 2023.

[28] Tjiptono and Aditya, *Pengaruh Kualitas Pelayanan dan Kualitas Produk terhadap Keputusan Pembelian*. Jakarta: Andi Offset, 2012.

[29] S. F. Siregar, "Analisis Tingkat Kualitas Pelayanan dengan Metode Index Potential Gain Customer Value (PGCV) di PT Bank Muamalat Indonesia Cabang Medan," *Jurnal Sistem Teknik Industri*, vol. 7, no. 4, pp. 40–47, 2006.

ISSN 2085-4579

# An Integrated Web-Based Waste Management Platform for Urban Collection Schedulling and Fee Payment Services Pekanbaru City

Ibnu Surya[1], Juni Nurma Sari[2], Satria Perdana Arifin[3] , Thesa Nabila Balqis Pardede[1]
[1]Program Study of Computer Engineering Technology, Department of Information Technology, Politeknik Caltex Riau, Pekanbaru, Indonesia
ibnu@pcr.ac.id, thesa@pcr.ac.id
[2] Program Study of Master of Applied Computer Engineering, Department of Information Technology, Politeknik Caltex Riau, Pekanbaru, Indonesia
juni@pcr.ac.id
[3]Program Study of Information System, Department of Information Technology, Politeknik Caltex Riau, Pekanbaru, Indonesia
satria@pcr.ac.id

Accepted on December 15th, 2025
Approved on June 9th, 2026

***Abstract*—Waste management has become a global issue, encompassing various challenges such as the separation of organic and inorganic waste, waste collection scheduling, fee collection, and the accumulation of waste at final disposal sites (TPA). In Pekanbaru City, the waste management mechanism involves empowering Waste Collection Agencies (LPS) across 83 sub-districts. LPS is responsible for collecting household waste and transporting it to the TPA, as well as managing the waste collection fees paid by residents. Since the fee management process is still carried out manually, an information system is needed to manage resident data and monitor fee payments effectively. This study developed the SilepasPKU Waste Management Information System using the prototype method, featuring waste collection management and fee management modules. The system was tested through three iterations with LPS UmbanSari users: the first iteration focused on design testing, while the second tested system functionality. Several improvements were made based on user feedback. In the third iteration, all system functionalities operated successfully. The testing method is Blackbox Testing, User Acceptance Test (UAT) and Usability Testing. The UAT result is user accept all the functional system. The usability test is held to take test for three role which is waste fee collector, waste collector and LPS administrator, and the processing results of the questionnaire data, it shows that the system has a good level of usability and is acceptable to users. Currently, SilepasPKU is being used by LPS UmbanSari in their waste management operations.**

***Index Terms*—, Household waste; Waste collection agency; Waste collection fee; and Web-based management system.**

## I. INTRODUCTION

Waste is a global issue that affects many parts of the world. In developing countries, waste management systems remain poorly organized due to limited public knowledge on proper waste handling, inadequate household-level management practices, and the lack of adequate infrastructure for final disposal sites (TPA). Many TPAs still rely on open dumping, while controlled landfill and sanitary landfill methods are rarely implemented. Other challenges include insufficient recycling facilities, excessive use of single-use plastics, and weak coordination between existing institutions and local governments, resulting in ineffective waste management systems [1].

According to the 2023 National Waste Management Information System (SIPSN) report [2], the largest portion of waste is disposed of in TPAs (60%–70%), followed by burning (15%–20%), recycling (10%–15%), composting (5%–10%), and uncontrolled dumping (10%). Waste burning is commonly practiced due to the absence of waste transportation facilities, inherited traditions, and limited awareness of the environmental risks associated with open burning.

Since July 2025, waste management in Pekanbaru has been carried out through Waste Collection Agencies (LPS), which operate at the sub-district level to collect household waste. Based on interviews with LPS officers in Umbansari, waste collection was previously managed by private companies. However, frequent delays in collection often resulted in waste accumulation in residential areas. Consequently, the establishment of LPS was initiated by the local sub-district leader and operated by community members. This system aims to regulate household waste collection and ensure its transport to the TPA, contributing to a cleaner and healthier environment. LPS has become one of the city's flagship programs. In addition to collecting waste, LPS is responsible for managing waste collection fees. To facilitate these responsibilities, the SilepasPKU Waste Management Information System was developed using the prototype

method, featuring modules for waste collection management and fee management.

Several previous studies have explored waste management information systems. Study [3], [4], [5], [6] developed a mobile-based system designed to simplify household waste collection, where waste had already been sorted into organic and inorganic categories, and pickup was requested directly by residents or coordinated and assembled in one place. Another study [7], [8] developed a web-based waste management system.

Studies [9], [10], [11], [12] developed web-based waste bank applications using the 3R (Reduce, Recycle, Safe Disposal) approach, aiming to educate communities on reducing household waste generation. These systems include roles for administrators, officers, and residents, enabling convenient waste exchange transactions.

Another study [13] introduced an application featuring route search for disposal sites, real-time updates on waste collector arrival times, and reporting capabilities for local planning agencies (Bappeda) to monitor collected waste volumes. The application, named E-waste, serves as a solution to optimize waste management in Gorontalo City.

From the analysis of previous research, it can be concluded that most of the systems that have been developed focus on aspects of waste collection, waste banks, disposal routes, reporting waste quantities, or community education. Although these features are important in supporting waste management, there are still limitations in the integration between waste collection schedule management and resident waste fee management, especially in waste services at the sub-district level.

The research being conducted currently is the development of waste management system that integrates waste collection scheduling and waste fee management for residents in UmbanSari sub-district. The waste collection module manages scheduling and daily collection recording, while the fee management module records and organizes resident fee payments. The primary contribution of this study is the integration of waste collection operational management and waste fee administration functionalities within a unified web-based platform.

The waste management application that developed in this research integrates specialized functionalities for waste collection operations and household waste fee administration in UmbanSari Village. The system was commissioned by the management of LPS UmbanSari to streamline and enhance local waste management activities. The System, named SilepasPKU, was developed using the Prototype methodology, which emphasizes continuous user involvement throughout the development process to ensure conformity with organizational business processes and stakeholder requirements. The Prototype methodology, consisting of the following phases: identifying existing business processes and roles, proposing improved workflows, designing the system and user interfaces, conducting user evaluations, implementing the system using the Laravel framework with MySQL database, and performing black box testing to ensure all functionalities work as expected. Testing was conducted with LPS administrators. Results indicate that SilepasPKU effectively supports waste collection and fee management activities, helping LPS monitor waste collection, disposal, and resident fee contributions more efficiently.

## II. Research Methodology

The development of SilepasPKU employed the Prototype software development method, which consists of the following phases: identification, design, prototyping, evaluation, implementation, and testing. The methodology used in this study is described below.

### *A. Identification*

The identification phase focused on gathering user requirements for SilepasPKU. The users in question were the waste management administrators (LPS) of Umbansari Sub-district. Observations were carried out through interviews and evaluations of the existing LPS operational processes. The observation process was conducted through interviews with LPS managers using a predefined set of questions to gather user requirements. During these interviews, an evaluation of the current operational processes of the LPS was also carried out. A gap analysis of the current system provided the foundation for developing the new information system. A comparison between the current (As-Is) business processes and the proposed (To-Be) processes is presented in Table 1.

TABLE I Comparison Of The Existing And Proposed Business Processes

| No | As-Is | To-Be |
|---|---|---|
| 1. | Viewing contribution data and waste collection progress is done manually. | Displays statistical data on contributions and progress of automatic waste collection in the system. |
| 2. | Management of sub-district, village, and family card data is still done manually. | Management of sub-district, village and family card data is carried out directly in the system, such as adding, changing and deleting. |
| 3. | Manual recording and monitoring of the status of annual citizen contribution payments. | Recording and monitoring of payment status is done on the system. |
| 4. | Manual recording of daily waste collection status per house. | Recording daily waste collection status is neater and easier in the system. |

During this stage, the business value of the proposed waste management information system was identified, namely improving efficiency in data

management and waste fee recording within LPS. The resulting business requirements are:

a. The system must be able to record and store resident waste fee data.
b. The system must allow monitoring and recording of daily waste collection progress.
c. The system must be able to record resident data registered under LPS.
d. The system must support the management of sub-district and village data used in SilepasPKU

Based on the identification process, the users of SilepasPKU consist of the super admin, LPS admin, waste collection officers, and fee collection officers, each with the following access rights:

1. The LPS super admin is a user with full access rights, responsible for managing master data (sub-districts and villages) and user accounts.
2. The LPS admin is an officer or staff member at the sub-district level who manages family card data, waste fee data, and waste collection activities within their designated area.
3. The waste collection officer is a user assigned to the Waste Collection feature, responsible for marking the waste collection status in the designated zones within their assigned village.
4. The fee collection officer uses the Waste Fee Payment feature to record payments received from residents in the village where they are assigned.

Based on these roles, the proposed business processes for the SilepasPKU system include monitoring waste management and fee collection (Figure 1), waste collection activities (Figure 2), and waste fee collection processes (Figure 3).

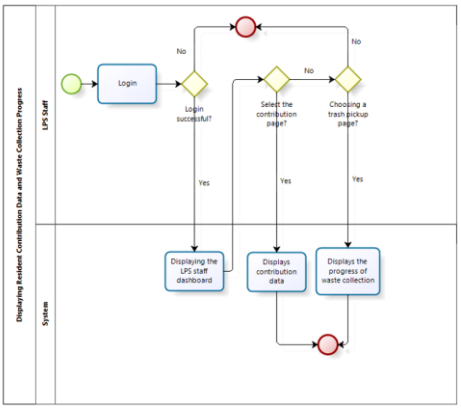


Fig. 1 Business Process for Monitoring Waste Management and Waste Fee Collection

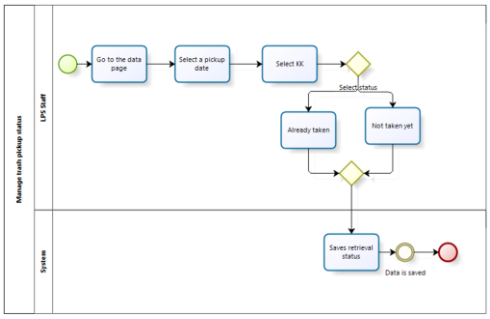


Fig. 2 Business Process for Waste Collection

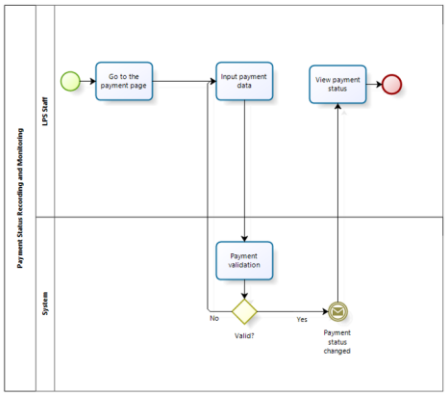


Fig. 3 Business Process for Waste Fee Collection

## B. *Design*

In this stage, the SilepasPKU system was designed according to the proposed business processes. The design includes system architecture, use case modeling, database design, and user interface design:

### 1. *System Architecture,*

The SilepasPKU system architecture adopts a Client–Server model in which users interact with the system through their respective devices. The system architecture is presented in Figure 4.

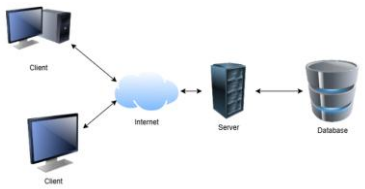


Fig. 4 System Architecture

### 2. *Use case Diagram*

The use case diagram in this study illustrates that the Super Admin is able to manage master data, including sub-districts and villages, as well as AdminLPS user accounts. Meanwhile, the AdminLPS user can manage zone data, resident data, waste fee data, payment validation, and the waste collection schedule within their designated area. The use case diagram is presented in Figure 5.

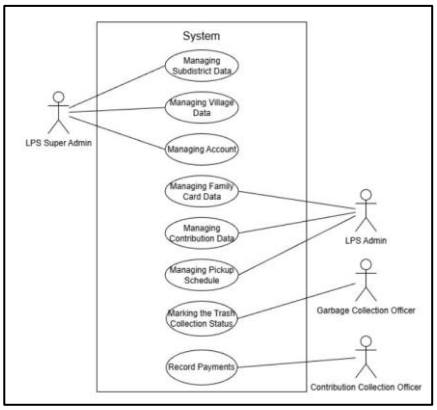


Fig. 5 Use Case Diagram of the SilepasPKU Application

*3. Database Design*

The SilepasPKU system is designed to manage resident data, waste collection activities, and waste fee records; therefore, several database tables were created, including Sub-district, Village, Family Card, Schedule (Waste Collection Schedule), Collection (Waste Collection), Fee, Payment (Waste Fee Payment), and User for account management. The database design is shown in Figure 6.

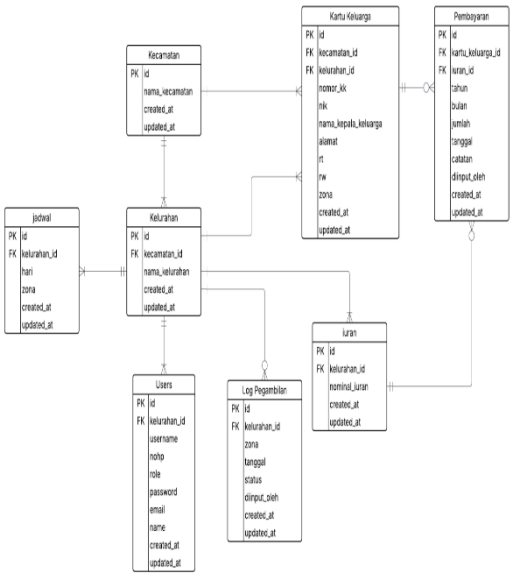


Fig. 6 Database Design

*4. User Interface Design*

The user interface design of SilepasPKU was developed based on the use case diagram. Several interface prototypes of the SILEPASPKU system are shown in the figures below. Figure 7 presents the dashboard prototype, Figure 8 illustrates the waste collection recording interface, and Figure 9 displays the waste fee data interface.


Fig. 7 Dashboard


Fig. 8 Waste Collection

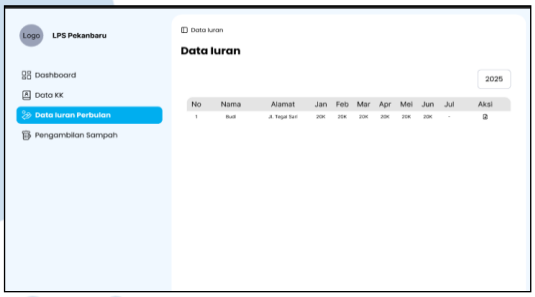


Fig. 9 Waste Fee Data

*C. Prototype Evaluation by Users*

The user interface design presented in subsection II.B.4 was evaluated by the LPS Super Admin. The evaluation results indicated that the Super Admin approved the interface design, allowing the system development process to proceed to the implementation stage.

*D. Implementation*

The completed system design was implemented using the PHP programming language with the Laravel framework, while MySQL was used as the database management system.

*E. Testing*

The SilepasPKU web application was tested by the LPS Super Admin using the black box testing method.

## III. RESULT AND DISCUSSION

After the user interface design was approved by the LPS Super Admin, the development of the SilepasPKU web-based system continued using the PHP

programming language with the Laravel framework and MySQL as the database management system.

### *A. Implementation of the SilepasPKU Web Application*

Based on the proposed business processes, the SilepasPKU web application was developed with several key features, including a dashboard for displaying waste fee statistics and waste collection progress, data management modules, and interfaces for recording and monitoring both waste fee payments and waste collection activities.

The SilepasPKU system supports multiple user accounts according to the defined business roles. Users are required to log in through the login page, as shown in Figure 10, before accessing the system.

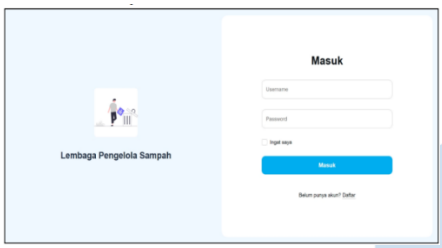


Fig. 10 Login Page

The SilepasPKU application is managed by the Super Admin account, which has full access rights to oversee all system master data, including sub-districts, villages, and AdminLPS user accounts. In addition, the Super Admin can monitor waste management and waste fee collection activities through the dashboard, as shown in Figure 11. The dashboard presents a summary of operational data in Pekanbaru, consisting of the following components:

1. Statistical Cards, displaying:
   a. Annual Fee Potential: the total projected revenue for one year assuming full payment compliance by all registered households.
   b. Fee Revenue (This Year): the actual accumulated revenue collected up to the current period of the year.
   c. Total Family Card: the total number of households registered across the city.
2. Annual Fee Revenue Chart: which visualizes monthly waste fee revenue trends from January to December, helping to assess financial performance over time.
3. Weekly Waste Collection Progress Chart: which illustrates the number of households whose waste was successfully collected each day during the past week, providing an overview of operational performance.

The sub-district data management interface is shown in Figure 12, while Figure 13 presents an example of the account management interface.

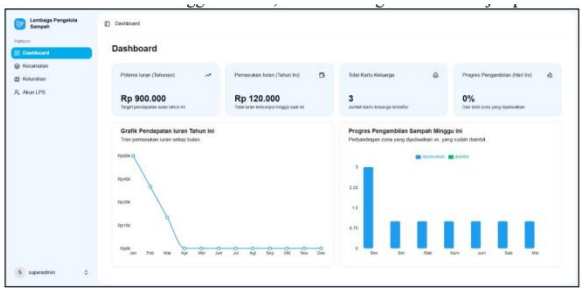


Fig. 11 Dashboard Page

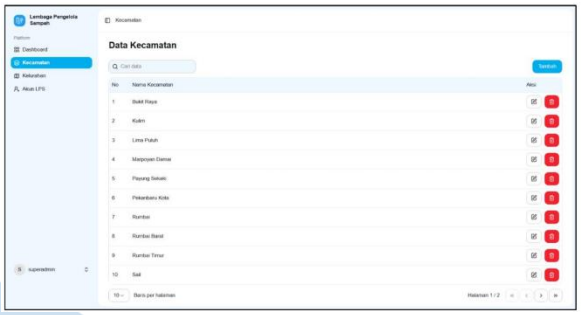


Fig. 12 Sub-District Data Management

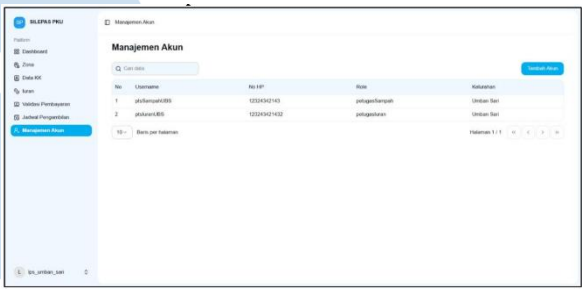


Fig. 13 Account Management

The next user role in the SilepasPKU system is the LPS Admin, who can view and manage data exclusively within the sub-district to which they are assigned. The LPS Admin can access the dashboard for their respective sub-district, perform create, read, update, and delete (CRUD) operations on Family Card data, manage the waste collection schedule (Figure 14), validate daily waste fee deposits (Figure 15), define collection zones for each area (Figure 16), and set monthly waste fee amounts in accordance with applicable regulations (Figure 17). In addition, the LPS Admin is responsible for managing the accounts of waste fee collectors and waste collection officers.

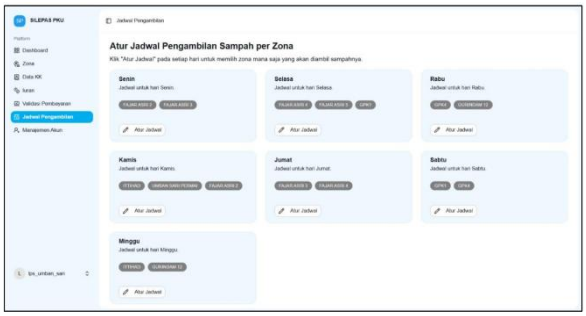


Fig. 14 Waste Collection Recording Page

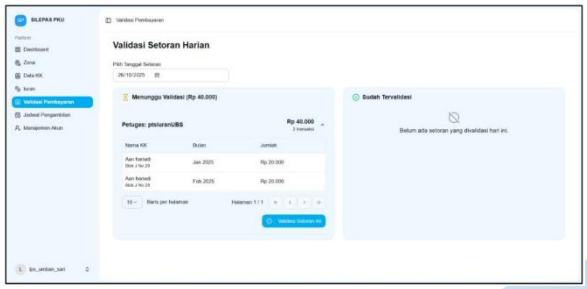


Fig. 15 Daily Deposit Validation


Fig. 16 Zone Assignment Page

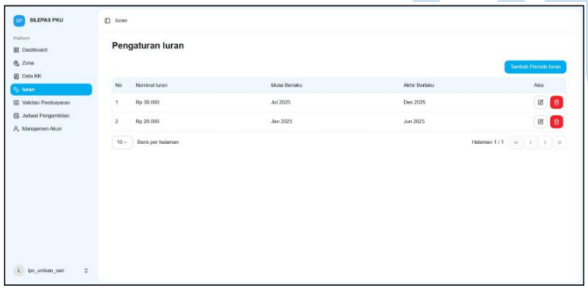


Fig. 17 Waste Fee Determination Page

The next user role in the SilepasPKU system is the fee collection officer, who is responsible for recording resident waste fee payments into the system, as illustrated in Figure 18.

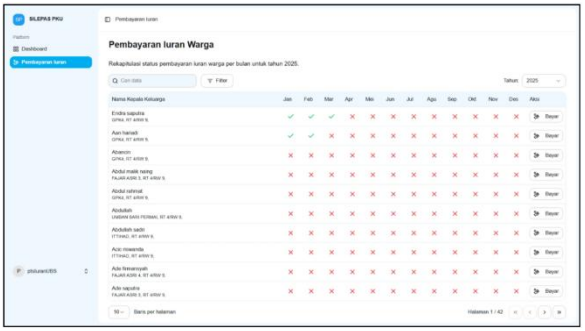


Fig. 18 Waste Fee Payment Recording Page

The final user role in the SilepasPKU system is the waste collection officer, who is responsible for recording household waste collection activities into the system, as shown in Figure 19. This menu functions as a digital checklist to support daily waste collection operations.


Fig. 19 Waste Collection Checklist Page

Each user account performs its respective role according to the access rights assigned. After completing their tasks, users can log out of the system, as shown in Figure 20.


Fig. 20 Logout Menu

## B. *System Testing*

The final phase of the Prototype methodology consists of system functionality testing and user interface (UI) evaluation using the System Usability Scale (SUS) to assess the system's usability

1. User Acceptance Test

Comprehensive system testing was conducted to evaluate the functionality of the system from the users' perspective, specifically the LPS management acting in the role of Super Administrator. This testing phase, known as the **User Acceptance Test (UAT)**, involved the LPS Super Administrator, who has access to all functionalities available within the SilepasPKU system

The results of the UAT are presented in Table II.

TABLE II UAT TESTING

| Function | Scenario | Expected Result | Test Result | Conclusion |
|---|---|---|---|---|
| Displaying automatic waste fee statistics and waste collection progress in the system | The Super Admin selects the Dashboard menu and views the waste fee statistics and waste collection progress | The waste fee statistics and waste collection progress are displayed as expected | Successful | Valid |
| Managing sub-district, village, and family card data directly in the system, including adding, modifying, and deleting | The Super Admin selects the management menu for sub-districts, villages, and family cards, and performs add, edit, and delete operations | Sub-district, village, and family card data can be properly managed | Successful | Valid |
| Recording and monitoring payment status in the system | The Super Admin selects the payment recording and payment monitoring menus | Both the recording and monitoring menus function correctly | Successful | Valid |
| Recording daily waste collection status in a clearer and more organized manner | The Super Admin selects the daily waste collection status recording menu | The daily waste collection status menu functions correctly | Successful | Valid |

During the UAT phase, the Super Admin user performed several test scenarios to evaluate the system's functionalities, including monitoring the progress of waste collection and waste fee collection, recording waste collection activities, recording waste fee payments, and managing various data within the SilepasPKU application. The results showed that all functions operated correctly and were validated successfully. These findings indicate that the SilepasPKU application meets the defined business requirements, namely: the ability to record and store resident waste fee data, monitor and document daily waste collection progress, record resident data registered under LPS, and manage sub-district, village, and user account data within the system. Therefore, SilepasPKU provides increased efficiency in data management and waste fee recording for LPS operations.

2. Ussability Testing

Usability testing was performed for all user roles within the SilepasPKU system, including the LPS Administrator, Waste Collector, and Fee Collector. The assessment employed the System Usability Scale (SUS) method, using the questionnaire items listed in Table III

*Table III Ussability Testing*

| No | Question | Score | | | | |
|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 |
| 1 | I think that I would like to use this system frequently | | | | | |
| 2 | I found the system unnecessarily complex | | | | | |
| 3 | I thought the system was easy to use. | | | | | |
| 4 | I think that I would need the support of a technical person to be able to use this system. | | | | | |
| 5 | I found the various functions in this system were well integrated baik | | | | | |
| 6 | I thought there was too much inconsistency in this system. | | | | | |
| 7 | Saya merasa sebagian besar orang akan cepat mempelajari cara I would imagine that most people would learn to use this system very quickly | | | | | |
| 8 | I found the system very cumbersome to use | | | | | |
| 9 | I felt very confident using the system | | | | | |
| 10 | I needed to learn a lot of things before I could get going with this system | | | | | |
| **Questionnaire for the LPS Administrator Role** | | | | | | |
| 1 | I was able to log in without difficulty | | | | | |
| 2 | The resident data menu is easy to find | | | | | |
| 3 | The fee recording process is easy to perform. | | | | | |
| 4 | The deposit validation process was easy to understand | | | | | |
| 5 | Proses validasi setoran mudah dipahami | | | | | |
| 6 | The deposit validation process was easy to understand | | | | | |
| **Questions for the Fee Collection Officer** | | | | | | |
| 1 | I was able to log in without difficulty | | | | | |
| 2 | The resident list was easy to find | | | | | |
| 3 | Recording payments was easy to perform | | | | | |
| 4 | The deposit reporting process was easy to understand. | | | | | |
| **Waste Collection Officer** | | | | | | |
| 1 | I was able to log in without difficulty. | | | | | |
| 2 | The collection schedule was easy to find | | | | | |
| 3 | The collection checklist was easy to complete | | | | | |
| 4 | The collection checklist was easy to complete | | | | | |

The SilepasPKU system was tested for ease of use and usefulness using the System Usability Scale

(SUS) method. The testing involved all user roles present in the system, namely LPS Admin, Contribution Collection Officer, and Waste Collection Officer. Each role in the system filled out the SUS questionnaire with questions as shown in Table III using a 1–5 Likert scale, where :

- 1 indicates Strongly Disagree
- 2 indicates Disagree
- 3 indicates Neutral
- 4 indicates Agree
- 5 indicates Strongly Agree

Based on the processing results of the questionnaire data, it shows that the system has a good level of usability and is acceptable to users. This result indicates that users find the system easy to learn, easy to use, and able to support the implementation of operational tasks in waste management and contribution administration. Additionally, users also assessed that the available features are well integrated and align with the operational needs of LPS Umbansari.

Therefore, it can be concluded that the SilepasPKU system has a good level of usability and is suitable for use as a supporting tool.

## IV. CONCLUSION

Waste is an everyday issue, particularly household waste, which includes food ingredients, leftovers, plants, and other organic materials. Proper waste management should be coordinated by local authorities to ensure that it does not negatively impact the environment. Several initiatives have been implemented by local governments, such as using applications for waste collection, organizing recyclable waste management (waste banks), and conducting regular household waste pickups. These efforts motivate the development of innovations that can simplify waste collection coordination and facilitate the collection of waste fees used to support waste management operations.

In this study, the innovation that build is the SilepasPKU web based application, which serves as a system for monitoring and recording waste collection activities and waste fee payments in UmbanSari Village at Pekanbaru Indonesia. SilepasPKU, a web-based waste management information The main contribution of this research is the integration of waste collection management and waste fee administration into a single platform that can be accessed by multiple user roles, including LPS Super Admins, LPS Admins, waste collection officers, and fee collection officers. The system was developed using the Prototype methodology, which enabled continuous user involvement throughout the development process to ensure that the implemented features aligned with the operational requirements of LPS Umbansari. The evaluation results indicate that the system functions properly based on Black Box Testing, is accepted by users through User Acceptance Testing (UAT), and demonstrates a satisfactory level of usability based on System Usability Scale (SUS) assessment. These results suggest that SilepasPKU can improve the efficiency of waste data management, waste collection monitoring, and waste fee recording processes.

Despite these positive results, the current system has several limitations. The application is only available as a web-based platform and does not yet support mobile devices. In addition, it does not provide automatic notification features for fee payments and the system hasn't integrate GPS-based location tracking to monitor waste collection activities in real time.

Future research may focus on developing a mobile application version, implementing notification services, integrating GPS and mapping technologies, and evaluating the system in a wider operational environment involving multiple villages or waste management organizations. These enhancements could further improve the effectiveness, scalability, and usability of the system.

## ACKNOWLEDGMENT

The authors would like to express their gratitude to Politeknik Caltex Riau for providing funding for this research through the 2025 Internal Research Grant.

## REFERENCES

[1] Y. Hendra, "Comparison of Waste Management Systems in Indonesia and South Korea: A Study of Five Waste Management Aspects," *Aspirasi*, vol. 7, no. 1, Jun. 2016, doi: 10.46807/aspirasi.

[2] N. Larasati and L. Fitria, "Analysis of the Organic Waste Management System at the University of Indonesia (A Case Study on the Effectiveness of the Waste Processing Unit at UI Depok)," *Jurnal Nasional Kesehatan Lingkungan Global*, vol. 1, no. 2, Jun. 2020, doi: 10.7454/jnklg.v1i2.1032.

[3] A. N. Kholili and D. Redaksi, "Mobile-Based Household Waste Management Information System," *Jurnal Informatika dan Teknologi (INTECH)*, vol. 4, no. 1, pp. 28–34, May 2023, doi: https://doi.org/10.54895/intech.v4i1.1982.

[4] Raghavendra D and M. Srinivasulu, "Mobile Application Model for Solid Waste Collection System," *International Journal for Research Trends and Innovation (www.ijrti.org)*, vol. 7, no. 9, p. 237, 2022, doi: 10.6084/m9.doione.IJRTI2209031.

[5] C. C. Cruz, H. C. R. Villafuerte, R. B. Avila, and K. L. S. Derla, "Mobile Application for Monitoring Trash Collection Schedules and Promote Efficient Solid Waste Management," *Cognizance Journal of Multidisciplinary Studies*, vol. 4, no. 3, pp. 22–31, Mar. 2024, doi: 10.47760/cognizance.2024.v04i03.003.

[6] A. Paudel, A. Pant, A. Manandhar, and B. Gautam, "Towards Effective Solid Waste Management: A Mobile Application for Coordinated Waste Collection and User-official Interaction," *International Journal of Information Technology and Computer Science*, vol. 16, no. 1, pp. 13–25, Feb. 2024, doi: 10.5815/ijitcs.2024.01.02.

[7] P. D. S. S. Reddy, B. Hussain, N. Ravi, V. Praveen, and A. M. Khan, "Web Based Waste Management System," *INTERANTIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT*, vol. 09, no. 01, pp. 1–9, Jan. 2025, doi: 10.55041/IJSREM40815.

[8] A. Srivastav, G. Gupta, I. Tyagi, K. Soni, and B. Rani, "Web Technology based Waste Management System: A Real Time on Demand Collection of Waste," *Int J Res Appl Sci Eng Technol*, vol. 9, no. VII, pp. 1289–1294, Jul. 2021, doi: 10.22214/ijraset.2021.36581.

[9] D. Wulandari, S. Hadi Utomo, and B. S. Narmaditya, "Waste Bank Waste Management Model in Improving Local Economy," *International Journal of Energy Economics and Policy* |, vol. 7, no. 3, pp. 36–41, 2017, Accessed: Dec. 15, 2025. [Online]. Available: https://dergipark.org.tr/tr/download/article-file/361762

[10] D. Choirrunnisa and N. Ngatindriatun, "The Impact of Waste Bank on Waste Processing Behavior and Income," *Efficient: Indonesian Journal of Development Economics*, vol. 4, no. 2, pp. 1201–1216, Jun. 2021, doi: 10.15294/efficient.v4i2.46061.

[11] S. Widaningsih and A. Suheri, "Web-Based Waste Bank Data Management Information System in Cianjur Regency," *IJCIT (Indonesian Journal on Computer and Information Technology)*, vol. 4, no. 2, Oct. 2019, doi: 10.31294.

[12] Y. Yunita, M. Adrianshyah, and H. Amalia, "Waste Bank Information System Using a Prototype Model," *INTI Nusa Mandiri*, vol. 16, no. 1, pp. 15–24, Aug. 2021, doi: 10.33480/inti.v16i1.2269.

[13] R. Taufik RL Bau, H. A, B. Ahaiki, and A. Engelen, "Digital Inovation in Waste Management:Designing an Waste Prototype as Solustion to Optimize Waste Management in Gorontalo City," *International Journal of Computer and Information System (IJCIS)*, vol. 4, Dec. 2023, Accessed: Dec. 09, 2025. [Online]. Available: https://www.ijcis.net/index.php/ijcis/article/view/148/124

ISSN 2085-4579

# A Cloud-Integrated Business Process Redesign Framework for Digital Transformation in Traditional SMEs

Yulyanty Chandra1, Lorio Purnomo2
1 Information Systems Management Department
BINUS Undergraduate Program – School of Information System, Bina Nusantara University Jakarta, Indonesia
2 Binus Entrepreneurship Centre, Management, Binus Business School, Bina Nusantara University Jakarta, Indonesia
1yuliyantychandra@binus.edu
2lorio.purnomo@binus.ac.id

Accepted on January 23rd, 2026
Approved on June 18th, 2026

***Abstract*—Fast moving digital technologies have forced small and medium sized companies (SMEs) to restructure their business processes to increase sustainability and competitiveness. Yet many traditional SMEs - particularly in the food and beverage industry - still rely on manual and fragmented processes that impede efficiency, scalable and data integration. This research describes a cloud-integrated methodology to reform business process in a conventional Bakery using Suisse bakery Jakarta as qualitative case study. Data were collected through direct observation and semi structured interviews with operational stakeholders to identify current company operations, operational challenges and digitization needs. These results were mapped into a restructured business process model based on value chain analysis, use case modeling and cloud-based system architecture design. The suggested redesign combines online ordering, digital product catalog management, social media marketing and delivery services with cloud-based inventory, finance, warehousing and reporting systems. The resulting architecture combines corporate data and provides real-time information access to business processes. The research concludes that process visibility / data interoperability / operational flexibility / organizational scalability through cloud-based business process redesign improves process visibility / data interoperability and reduces dependency on physical IT infrastructure. This research mainly contributes a systematic digital transformation framework that makes business process redesign a precondition for cloud adoption and not just an outcome of technology installation. This methodology helps conventional SMEs to adopt cloud-enabled operations and contributes to the literature on digital transformation by providing a systematic approach to align business processes with cloud computing technology.**

***Index Terms*— Business Process Redesign; Cloud Computing; Bakery Management System; SME Digital Transformation; Qualitative Study.**

## I. INTRODUCTION

We all know that the food and beverage business world, especially the bakery homemade business is growing bigger from time to time. Most of SMEs have developed and consolidation the business according to the new era of technology[1], but some do not integrate with technology, because of the lack of knowledge of the system. The food and beverage industry in particular the bakery sector has expanded due to increasing consumer demands for service efficiency, accessibility and digital engagement. In response to these changes many small and medium sized firms (SMEs) started adopting digital technology for their operational activities, customer engagement and managerial decision making. Despite this trend, many conventional SMEs still use manual business processes because they lack technology expertise, budget constraints or fear of organizational change.

To enlarge the business, they must figure out and prepare of the several ways to support the supporting business process [2]from the marketing strategy, store location, human resources, variative menu of the food and beverage and the most important thing is the infrastructure of the system they build to enhance the daily operational business process activity. In many SMEs digitalization initiatives focus on system implementation rather than business process change resulting in slow technology adoption and low performance improvements.[3] The infrastructure itself must be in support not only for the business proces, but also for the activities and design of manufacture [4]. With a good infrastructure system, it will help to integrate the whole system for selling and buying at the bakery, to provide a better service for customers and to win the competition among other bakeries[5].

SMEs looking for economical, scalable and adaptable information solutions have turned to cloud computing. Unlike on-premises architecture, cloud-

based systems let enterprises access data and applications remotely, reduce hardware maintenance costs and dynamically alter computing resources as needed.[6] Such attributes make cloud computing particularly useful for SME operating in dynamic sectors like the food and beverage industry. Successful cloud deployment demands an absolute alignment between business processes and technical architecture - an area often not adequately investigated during traditional baking operations.[7]

For this research, the business process design is designed to Suisse Bakery, a very well-known old-fashioned bakery located in Hayam Wuruk, Jakarta since 1978. Although it is a well-known old-fashioned bakery, it is also important that they need to develop a good management system to survive and to maintain the customer from other bakery competition in Jakarta. Suisse bakery has not been integrated with technology since 1978, all the business process they are running is still manually. No website to promote the bakery, no online ordering system (OOS) and no system management to manage the customer's order and payment. The payment is only in cash, so the customer sometimes finds it difficult to make the payment other than cash.

This study aims to produce a design of business process in terms of technological infrastructure embedded with Cloud computing in the form of diagrams, catalogues, and matrices that can help Suisse Bakery in enlarge their business and to help to maintain the business management system also can help bakery staff manage data, including information on products, transactions, raw materials, and reports. Reports are available for owners to view without visiting the business due to the application's internet connection[8]. With the boom of electronic technology, medium-sized and small firms (SMEs) are becoming very competitive in the food & drink segment as well. Digital transformation facilitates operational efficiency, customer engagement and also managerial decision making for the firms by using information technologies. Many traditional SMEs still use manual, disconnected and non-integrated business processes which cause operational inefficiencies, restricted scalability, data inconsistencies and lower response to market changes. Cloud computing is an appealing technology solution because of its scalability, flexibility, accessibility and cost effectiveness to SMEs. In previous research, benefits of cloud adoption for SMEs were debated and identified as better data management, lower infrastructure costs and greater organisational agility. But prior research has focused mainly on factors affecting the implementation and adoption of technology and less on the redesign of underpinning business processes - a key indicator of whether cloud technologies can deliver organizational value.

This reveals a research gap. Business process redesign and cloud computing are acknowledged accelerators of digital transformation but little is known about their integration in traditional SMEs with manual workflows. Current research often considers cloud adoption merely a technological upgrade, and not a transformation effort that requires a synchronization of organizational processes with digital infrastructure. Thus the ignorance remains as to how traditional SMEs should restructure their operation before adopting cloud-based solutions. Herein, a cloud-integrated business process redesign methodology with Suisse Bakery Jakarta as case study fills this gap. Research highlights the strategic role of business process redesign for cloud adoption combining value chain analysis, use case modeling and cloud-based architectural design. This work contributes to the digital transformation literature through providing a structured framework combining operational processes with cloud technology and offering practical assistance to traditional SMEs to achieve sustainable digital transformation.

## II. LITERATURE REVIEW

A good business in this era of development should follow the world's technology that is always changing so fast. It is not only about how to make a good selling point, but also important to maintain the loyalty of customer satisfaction.

### *A. Business Process*

According to According to [9], in his book "Business Process Change", the definition of business processes is a series of activities carried out by a business which includes the initiation of inputs, transformation of information, and producing outputs. The output can be valuable to the business or market customer, it can also be valuable to other processes (within the organization). A business process can be broken down into several sub-processes that each have their own attributes that contribute to achieving the objectives of the parent process. Sub-processes can be broken down into activities, which are the smallest sub-processes that can consist of one or more steps that must be included in the business process. With the business process maturity model, it can help the daily operational process, management process, supporting process more effective and efficient [10].

BPR is a fundamental technique for enhancing organizational performance through fundamental rethinking / rebuilding of operational processes leading to dramatic improvements in key performance measures such as efficiency / quality of service / responsiveness / cost savings. This Business Process Redesign is relevant in the food and beverage industry because the operational operations involve several interrelated processes such as procurement / inventory management / production / sales / delivery and customer service / etc. Poor coordination between these tasks often results in delays, inaccurate inventory records, higher operating expenses and lower customer satisfaction.

Previous studies have demonstrated that BPR can dramatically improve operational performance of food-related organisations. BPR techniques together with digital process integration enabled food SMEs to improve business processes and cut processing time by 48% by using automated mechanisms for sales and inventory management.[11] Similarly, research in food and beverage industry showed that process reform, innovation thinking and alignment of organisational resources improve operational performance and competitive advantage. Results indicate that constant redesigning corporate activities is necessary for efficiency and competitiveness in dynamic market contexts. New literature on digital transformation further demonstrates that technology alone cannot create organizational value unless accompanied by structured business processes. Business process Management (BPM) and digital transformation are therefore complementary projects where Process reform is the basis for a successful adoption of digital technology.[12] This is especially relevant for traditional SMEs in the food and beverage industry where many operations are still manual and fragmented. Therefore, business process redesign is required before introducing cloud-based solutions to enable process standardisation, data interoperability and organisational readiness for digital transformation.[13].

*B. Cloud Computing*

Cloud computing is always giving the new concept of integrated information system by providing scalable and virtual resources for the company. Good techniques and resources, live migration, will vastly improved data center based cloud with little of performance overhead. [14]. Next In figure 1, shows there are three main components inside of cloud computing, like infrastructure, platform, and also application. And it is all connected to Laptops, Servers, desktop, tablets, phones [15]

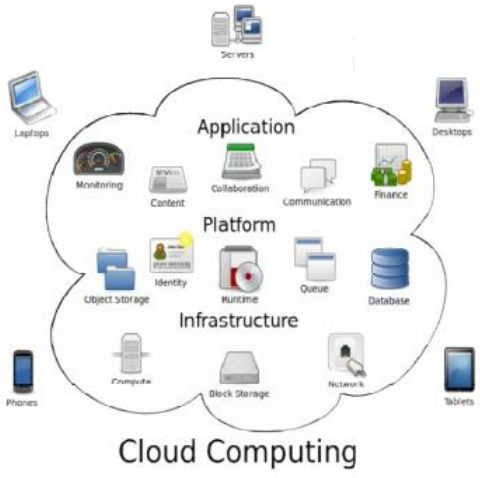


Fig 1. Cloud Computing

An information system called the Computing-Based Management Information System (CMIS) may perform a range of management-related duties, provide accurate information to all levels of management within an organization, and evaluate different types of data. Cloud computing concepts state that it is primarily built on five distinct qualities, including pay-as-you-go and self-provisioning of resources, along with shared resources, extreme scalability, and flexibility. The concept of cloud computing is distinct from previous computing paradigms that presumed restricted resources. Rather, the foundation of cloud computing is a business model that shares resources at the network, host, and application levels. Since cloud solutions enable organizations to lower the cost of computer resources, there is an increasing demand for cloud computing services.

The Information System becomes more superior and cost-effective with cloud computing, increasing its overall usefulness to a company. The following are the primary ways that cloud technology has affected information systems:

- Reasonably priced infrastructure: Companies may save a significant amount of money by avoiding the need to buy and maintain hardware. An enterprise can choose any of the three alternatives based on their needs and make the appropriate payment if they require public or hybrid cloud.

- Quick access to information: In this sense, cloud computing has completely transformed everything by enabling on-demand access to all company data from any location in the world without the need for physical presence. Whether you are at home, on vacation, or in route to or from the office, all of your information is accessible via an internet-connected device.

- The ability to scale: A cloud-based information system only requires a request to the cloud service provider; everything else is taken care of for you. In contrast, a traditional management information system would require the purchase and installation of hardware to fulfill the update. The service provider promptly updates cloud storage systems and apps to guarantee scalability and save a significant amount of time and work.

- Enhanced collaboration and integration: It might also help the business give its customers better service. Without the cloud, this kind of collaboration or integration would not be feasible since it would be necessary to install third-party apps on the data-containing systems and grant physical access to the systems to their IT staff to make the necessary deductions.

One of the main worries that cloud users have since the cloud's introduction is security. Numerous dangers could jeopardize the security and safety of user data. Recent years have seen a rise in cyberattacks and data breaches, putting cloud service providers at danger of losing the private data of their customers. The prolem can be solved by always maintaining the rapid

development of cloud computing, to protect the most crucial things like data privacy and service availability [16]. Among many other factors, weak or insecure passwords, inadequate firewalls, and unprotected APIs are the most common causes of cyberattacks and account theft. [17]

As a critical enabler of digital transformation, cloud computing provides SMEs with on-demand, scalable, flexible and cost-effective access to computer resources, apps and data storage through internet-delivered services. In contrast to conventional on-premises infrastructure, cloud-based systems require no major hardware or software investments and thus enable SMEs to adopt new digital technologies despite financial and technological limitations. Operations - cloud computing centralizes data administration, allows real time information exchange and facilitates communication across corporate divisions. Integrated cloud systems manage company functions such as inventory management, sales transactions, customer ordering, financial reporting, production planning, delivery management & supply chain coordination in a digital ecosystem. Dieser integration reduces data redundancy, manual processing errors, process visibility and prompt and informed managerial decision-making. Also, through cloud computing the organizational agility, scalability and operational reliability are also increased as the cloud organization can alter the resources according to the variation in the market demands. Cloud adoption improves SME performance through optimizing operations, lowering infrastructure costs, improving scalability and making data more easily accessible across organizational units. [18]Research indicates that cloud computing benefits are maximized when technology implementation fits organizational processes and operational requirements. Consequently cloud computing needs to be considered more than an information technology infrastructure, but an enabler of business process integration and a platform for long term digital transformation, particularly for traditional food & beverage SMEs looking to improve operational performance and service quality.

### *C. Bakery Management Information System (BMIS)*

BMIS is a computerized Bakery Management Information System. The system's customer database handles the issues of registering new agents on a daily basis, managing stock levels, and handling refunds, deliveries, product sales, and rates. Following registration, a consumer receives a special code that must be entered each time a purchase is made. By analyzing report trends, including stock taking, items sold, costs, and profits, the bakery can forecast supply and demand and improve its production and sales departments, among other areas, with the use of BMIS. The inputs and outputs are processed by the Bakery Management Information System (BMIS). It will help the admin to make the job more easier such as managing the production data, orders and payment [16], [19]. Details about data inputs include product sales through purchases, which will require filling out forms for sales, products, deliveries, returns, and agents. Forms such as completed product forms and stocking details are filled out to handle manufacturing specifics. Sales reports, stocktaking reports, completed product reports, delivery reports, return reports, and all searches, such as returns searches, are among the outputs that will be produced [20].

One of example from BMIS is Bakeroo as seen in figure 2 and Bakeroo managing plan in figure 3. Using a single platform, manage the workforce management, production management, inventory, workplace health and safety, and food safety for the bakery.

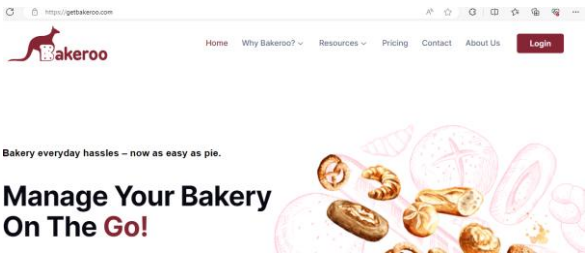


Fig 2. Bakeroo Website

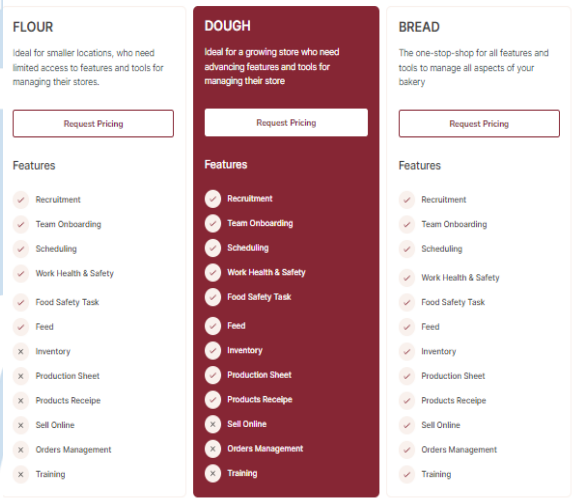


Fig 3. Bakeroo Managing Plan

Although the Bakeroo system has its ability to integrated all the system in a bakery, the system also needs few attention in the system limitation such as the ability of the users or bakery admin to have a knowledge or understanding the process of using the bakeroo system, and mastering the solutions for the system's problem. Also from the Vendor's side (Bakeroo system), the storage of customer's data such as daily transaction, customer's personal data, only stored inside of the Bakeroo's platform and give limitation to the admin or users when the data wished to be store outside of the platform. Using Vanilla implementaion for the Bakeroo, none of modification of the system can be made, forcing the bakery to use the default system. This process can trigger mismatch between the organization's business processes and standard processes applied in vanilla implementation.

## III. METHODOLOGY

The methods apply for this research are qualitative methods which are observation by visiting the bakery store as a customer and take a journey to have a tour at the bakery environment, see on how the business process is running with observation and interviewing few of the employees such as cashier and waiters [21].

1. Research Information

The research involving three key informants who were directly in touch inside of the Suisse Bakery's operational activity, they are consist of one store supervisor, one of the cashier and one of the kitchen staf.

2. Interview technics

The semi – structured interview were conducted to explore existing business processes, operational flow and information flow. The interview is about 15- 20 minutes at the bakery.

3. Observation

The research were observated on the daily operational activities such as customer ordering, cashier transaction, production activities and reporting processes. Field notes were recorded to document workflow sequences, communication patterns and operational bottlenecks.

4. Data Validation

To improve data credibility, source triangulation has been applied, by comparing information obtained from interviews, direct observations, and organizational documents.

5. Data analysis procedure

The analysis was held in four stages:

a. Identification the existing business processes through observations and interviews
b. Mapping the operational activities using value chain analysis
c. System development for the system requirement using the use case modelling
d. Design of cloud-integrated business process architecture based on identified operatonal requirements.

The Research design stages are shown in Figure 4, it starts with an explanation about the main activities and supporting process that include in the management of Suisse Bakery. Next, using a use case diagram to clearly identify the system's relevant part, define the various components of the business process and ascertain how information systems link to the business processes of the management of the Suisse Bakery.

The final design is to create an architecture-based cloud computing infrastructure for the Bakery's internal stages. Bakery environment is very important to determine the integration for whole main role for the business process like customer and the bakery staff or worker [22].

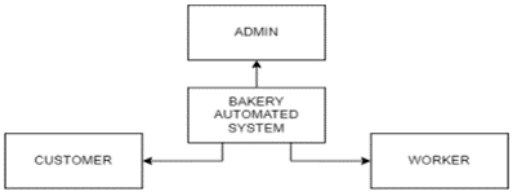


Fig 4. Bakery Environment

The research is using qualitative methods with interview and observations. For the figure 5 below is showing the research design stages.

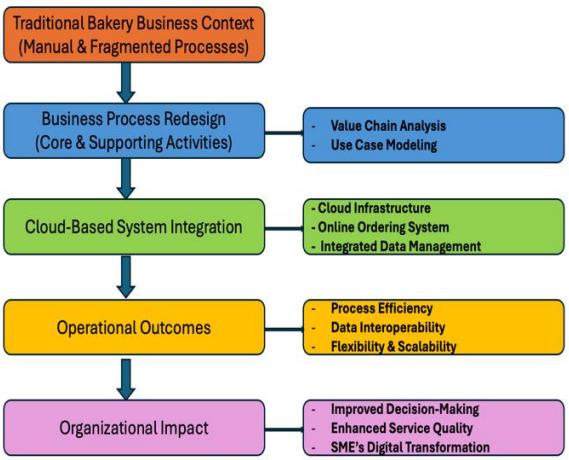


Fig 5. Framework Study

This work's conceptual framework defines the relationship between business process redesign and cloud-based system integration for digital transformation in conventional small and medium-sized enterprises (SMEs). The paradigm operates against the background of a conventional baking firm with manual, disjointed and non-integrated operational processes.[23]

- Initially the framework considers the conventional bakery business environment where all operational tasks like ordering / payment / inventory / reporting are performed manually. This state often causes data inconsistencies, restricted process visibility & operational inefficiencies which limit organizational scalability & market - readiness.[24]
- The second stage considers business process reform an essential tool for digital transformation. It considers business process redesign as an intermediary step rather than directly integrating technology. Value chain analysis and use case modelling identify and restructure key activities like online ordering, delivery services and customer interaction as well as supporting activities like kitchen operations, finance, marketing and warehouse management. Dieser process-oriented methodology guarantees that technical solutions correspond to real operational needs.[25]
- As the facilitating technology layer, the framework adds cloud-based systems after process redesign. Cloud computing offers scalable infrastructure, centralized data storage and real time access enabling optimized business processes to work

across roles. Cloud infrastructure provides system functions independent of physical IT assets for online ordering systems, inventory, transaction processing and management reporting.[26]

- Playing with restructured business processes in combination with cloud-based systems yields better operation outcomes like process/data efficiency, data interoperability, flexibility and scalability. These results allow SMEs to adapt operational capacity to changing demand conditions, relevant in the food and beverage industry.[26]
- It demonstrates how such operational enhancements induce large organizational benefits such as improved managerial decision making, improved service quality and sustained digital transformation. Such a paradigm places business process redesign as a precondition for cloud adoption, building upon the existing literature on SME digitization and providing a systematic path for traditional firms moving to cloud-enabled operations.[23]

## IV. Result and Discussion

The result is in this section for designing the business process for Suisse Bakery Management system, through the Value chain to explain about the whole process in main activities and supporting activities (seen on Figure 6). For main activities there are Catalogue Menu, Social Media Campaign, Delivery courier, OOS (Online Ordering System). For the Supporting activities there are Kitchen, Marketing/Sales, Finance/Accounting, also Warehouse.

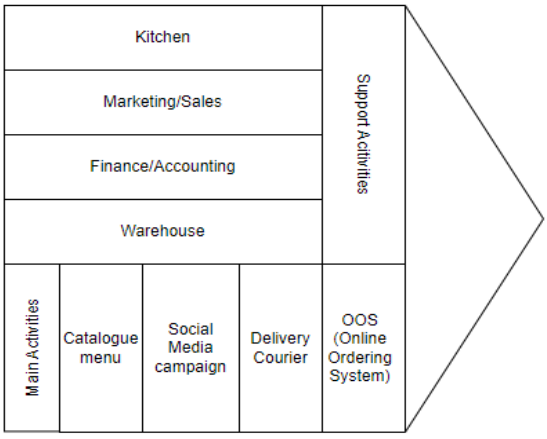


Fig 6. Value chain of main activities and support activities

The catalogue menu for bakery is a product catalogue that is filled with pictures of the bakery menu, detailed with the description of each menu and complete with the price to make customer easier to choose what kind of menu they want to order. The catalogue menu is presented in full color, divided based on the characteristics of the menu, for example the food and beverage section is different to make the customer easy to find [27].

Social media campaign is social media promotion [28]. Every social media networking site caters to a certain demographic and employs varying methods to enable users to exchange ideas, videos, images, and links, forming a network of individuals linked by shared interests. The growing quantity of people on these platforms has drawn businesses to incorporate online advertising into their marketing strategies because of social media's ability to target consumers and the fact that it is frequently less expensive than TV or print advertising. Social media networking services like Facebook, Twitter, and YouTube allow millions of people to share information, disseminating news from across the globe as well as news about their own lives [29].

Next is Delivery Courier, we call it to provide a good service to the customer. For running the business nowadays, the customer is not only on a walk-in to the store, but after they ordered by phone or website, the store will prepare the products and ready to deliver to the customer. Will be an advantage for customers not to spend so much time on their way to the bakery. The intention is to give clients the pleasure of placing their favorite meal orders in the comfort of their own homes[30]

And, Online Ordering System (OOS), The Internet has altered not only the social interaction pattern but also the corporate and economic interaction patterns in this rapidly evolving world. Online ordering systems, or OOS for short, are sophisticated organizational-technical systems designed to enable the processes involved in creating, altering, sending, controlling access, and delivering orders. It offers online services that let users browse for products, make orders, and purchase goods and services via the Internet. Therefore, to sell their items, the merchants would not need to own or rent a physical site. Customers may order cakes anytime, anyplace, and without having to waste time traveling to the bakery and waiting in line to place their orders, demonstrating the flexibility of this system [30].

While for the supporting activities are Kitchen, the kitchen is the busy place at the bakery. The bakers need to prepare the ingredients, bake, knead the dough, also make sure the bread display is always full so the customer would not wait too long to have the bread. The store supervisor needs to maintain the food stock and communicate with the baker about the food availability.

Next one is Marketing/Sales, to promote the bakery, marketing and sales hold an important assignment to make the marketing program every year, define the target for customer, designing an eye-catching advertisement can help to reach out the customer and interest the customer to come to the bakery or delivery online. Through digital marketing [31], Businesses can give better client satisfaction and have access to more efficient CRM technologies. According to [32], there are five determinants of loyalty identified by Reichheld and Schefter (2000):

Quality customer support, On-time delivery, Compelling product presentations, Convenient and Reasonable price shipping and handling, also Clear trustworthy privacy policies.

Moving forward to Finance/Accounting, work to record the financial statements of customers, especially for reports income from bakery every month, recap the expense for buying the raw material, operational expenses, etc. Preparing the annual bakery budget data collection for various needs.

The last one is Warehouse, Suisse Bakery continues to manage its business processes manually, utilizing several traditional procedures such purchasing raw materials based only on customer orders, which frequently results in mismatches. Between the required raw resources and the output at the point of production. Subsequently, there is still no unique recording made during the production phase, which encompasses the utilization of raw materials and the outcomes of finished items. As a result, after production is completed, the raw material calculations frequently contain inaccuracies. The procedure for reporting the distribution of manufactured goods presents another issue [33].

Next step is defining the business process parts that are including the details of each business process that can be seen in the use case diagram below (see Figure 7).

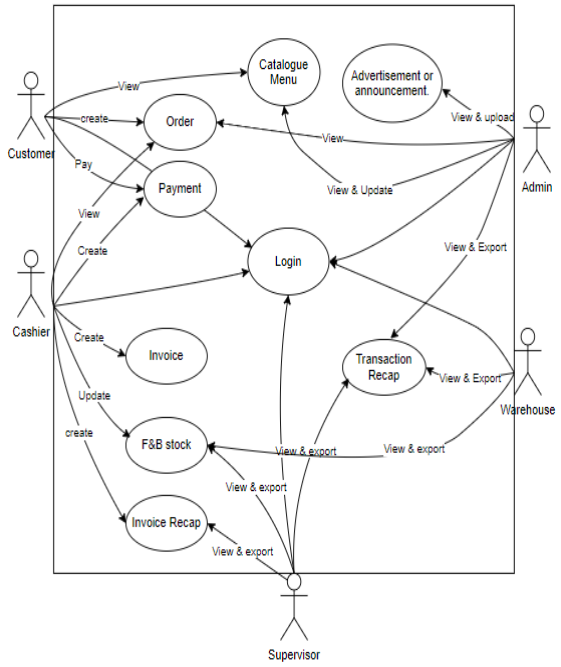


Fig 7. Use Case Diagram Business Process

The use case diagram above consists of five actors that are involved in the business process[34], they are Supervisor, cashier, customer, admin, warehouse. They all can login into the Bakery management system. The use case diagram covers all the main activities and supporting activities as seen in figure 6. The supervisor has access to manage the whole selling and buying activities, which is to view and export the data from transactions recap, invoice recap, food, and beverage stock. The data which has been exported can be used to make a report to the manager or to analyze and make forecasting for future transactions. Cashier has access to view the order from customer, and then create the bills/payment for the customer to pay, create invoice, updating the stock, and recap the invoice routinely. The customer also can view the catalogue menu to choose the food and beverage they want to order, after that they make an order through the system, after bills are released from cashier, they can do the payment. The admin has access to routinely updating the catalogue menu on the system, also upload the announcement and advertisement with the new one, so customer will always get up to date for the information. Admin also can view and export the transaction recap to report to the supervisor. The last is warehouse, they can view and export the food and beverage stock on the system so they can prepare to order the raw materials to the supplier. They also can view and export the transaction recap to compare and adjust the stock and selling data [35] Because bakeries are in areas that are associated with the food and beverage business, and because there is a possibility of physical damage or loss, cloud architecture was selected over on-premises design. Their computer solutions may be scaled to meet their needs thanks to cloud-based design. The bakery can reduce its computer resources to increase efficiency during the weekdays or at night as fewer people use the system to place orders than they do on the weekends and in the morning. Additionally, this method makes the workplace flexible. Due to the cloud's ability to keep all data, both the bakery and its customers can conduct business from any location, particularly when they are unable to visit the bakery. With all these benefits, cloud-based architecture has the potential to boost output.

Figure 8 shows the cloud-based architecture. Cloud architecture is preferred over on-premises architecture because the bakery is a public place and there is concern that there may be damage or loss of property.

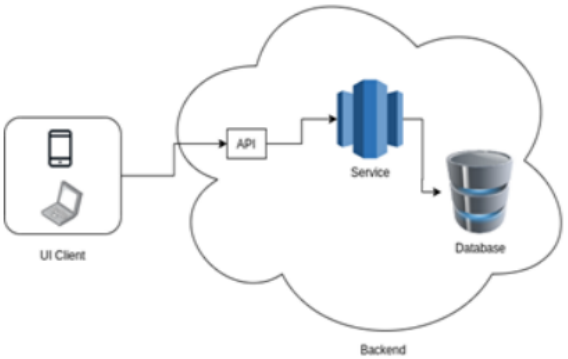


Fig 8. Infrastructure Architecture-Based Cloud Computing

Thanks to the cloud-based architecture, it is possible to develop IT solutions according to its needs. When the bakery is closed or out of operational working hours, customers cannot order through the bakery management system, so the bakery can reduce IT resources to operate more efficiently. This solution also allows for the creation of a flexible workplace. Because all information is stored in the cloud, admin and customers can continue to operate from anywhere, especially when unable to come to the store. With all these benefits, cloud-based architecture can increase productivity.

| Business Function | Current business process for Suisse Bakery | Bakeroo system | Proposed framework |
|---|---|---|---|
| Product ordering | Walk-in and telephone orders | Digital order management | OOS |
| product catalog | Manual printed information | Digital catalog | Digital catalog integrated with bakery's social media |
| Inventory control | Manual stock record | Inventory management module | Cloud-based inventory integrated with sales and warehouse |
| Financial Reporting | Manual recap and report | Standard operational reports | Real-time cloud reporting |
| Delivery management | Manual coordination | production-oriented workflow | Delivery integrated with ordering process |
| Data Access | Location - dependent | Cloud access available | Fully cloud-integrated architecture |
| Business process alignment | Existing manual workflow | Generic bakery workflow | Customized to Suisse Bakery operations |

Fig 9 Comparative analysis of business process capabilities

On Figure 9, compares the existing operational environment of Suisse Bakery, the Bakeroo bakery management platform, and the proposed cloud-integrated framework. The comparison shows that Bakeroo provides various digital management capabilities including inventory management, workforce management, production monitoring, and food safety management. However, Bakeroo functions primarily as a generic bakery management application. In contrast, the proposed framework was developed based on field observations and interviews conducted at Suisse Bakery and therefore reflects the specific operational requirements of the organization. The framework not only introduces digital tools but also redesigns business processes through the integration of online ordering, social media marketing, inventory control, delivery services, financial reporting, and cloud computing infrastructure. Consequently, the contribution of this research lies in providing a structured business process redesign framework that aligns cloud technology adoption with organizational needs.

| Area | Existing Condition | Proposed Condition | Expected Impact |
|---|---|---|---|
| Ordering Process | Manual order recording | Automated online ordering | Faster order processing |
| Inventory Monitoring | Manual checking | Real-time inventory visibility | Reduced stock discrepancies |
| Reporting | Manual compilation | Automated reporting | Improved managerial decision-making |
| Customer Reach | Physical store only | Social media and online ordering | Increased market reach |
| Data Management | Separate records | Centralized cloud database | Better data consistency |

Fig 10. Expected improvements

Figure 10 presents the expected operational improvements resulting from the proposed cloud-integrated business process redesign framework. The comparison indicates that several operational activities currently performed manually at Suisse Bakery can be enhanced through process automation and cloud-based integration. In the ordering process, manual order recording can be replaced with an automated online ordering system, which is expected to accelerate order processing and improve customer convenience.

Furthermore, inventory monitoring can be improved through real-time inventory visibility, reducing stock discrepancies and enabling better control of raw material availability. The proposed automated reporting mechanism is also expected to reduce the time required for report preparation while supporting faster and more informed managerial decision-making. In terms of customer engagement, the integration of social media and online ordering channels may expand market reach beyond physical store visitors. Finally, the use of a centralized cloud database is expected to improve data consistency by eliminating fragmented records and providing a single source of information across business functions. Overall, the proposed framework demonstrates the potential to enhance operational efficiency, information accessibility, and organizational scalability in support of digital transformation initiatives for traditional SMEs

## V. CONCLUSION

It shows that business process redesign is necessary for cloud adoption and digital transformation in traditional small and medium-sized organizations (SMEs). Analyzing a manual, fragmented and non-integrated bakery setting demonstrates that digital technology alone is inadequate for transformation. Essential and supporting business processes must be redesigned to ensure that cloud-based technologies meet actual operational requirements and organisational competences.

The proposed conceptual framework contributes to the literature on digital transformation/business process management by making business process redesign a precondition for cloud implementation and not a result. The methodology demonstrates how value chain analysis & use case modeling of cloud-based systems can enhance process efficiency, data interoperability, flexibility and scalability with respect to restructured processes. All these operational enhancements bring great organizational benefits in terms of refined managerial decision-making improved service quality and organisational readiness for adjusting to changing demand conditions - especially in the food and beverage sector.

This study offers a methodical approach for traditional SMEs moving towards cloud-enabled activities while preserving business continuity. Managers and practitioners are advised to prioritise process analysis and redesign before investing in

digital infrastructure to gain maximum benefits of cloud technologies. Though confined to one qualitative example, this study offers a general transferable framework applicable to many SME industries. Later research might extend this study with quantitative or mixed methods to empirically validate the framework across sectors and organisational settings.

Its qualitative single-case design constrains generalizability of findings to other SME sectors. In addition, operational enhancements are derived from process analysis instead of quantitatively measured. Later study may empirically support the proposed paradigm through quantitative or mixed method approaches across SMEs. Further longitudinal studies could explore whether business process reform enduringly impacts cloud adoption and organizational performance. Opening this framework to other industries/adding new technologies such as analytics or artificial intelligence is one possibility for future research.

In this study we see how business process reform is needed to make cloud adoption/digital transformation work in traditional SMEs. The research used Suisse Bakery as case study to develop cloud-integrated business process redesign methodology to connect operational operations to digital infrastructure. Proposed framework links customer ordering / inventory management financial reporting / marketing and delivery services in a centralized digital environment via value chain analysis / use case modeling and cloud architecture design. They enrich the literature on digital transformation and business process management by addressing the problem of business process redesign/cloud computing not being sufficiently incorporated in conventional SMEs. The report says that cloud adoption adds organizational value when accompanied by restructured business processes for operational / data interoperability / process visibility / scalability / managerial decision making.

But despite all these developments this research is constrained by many things. It is a qualitative single-case study, which limits generalizability of results to other SME industries/organisations. Its suggested framework was designed at design level and has not been empirically evaluated in full system implementation/performance assessment. Further study with a quantitative or mixed method approach in other SMEs and industry sectors should confirm the framework. Future system development may include business analytics / artificial intelligence / demand forecasting / customer relationship management / real time supply chain monitoring / data driven decision making. Longitudinal studies are recommended to assess if cloud-integrated business process redesign has lasted effects on organizational performance / innovation capacity / digital transformation maturity.

## AUTHOR CONTRIBUTION

Yulyanty Chandra: Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Resources, Supervision, Validation, Visualization, Writing–Original Draft Preparation, Writing–Review and Editing; Lorio Purnomo: Formal Analysis, Methodology, Writing–Original Draft Preparation, Writing–Review and Editing.

## DATA AVAILABILITY

https://doi.org/10.5281/zenodo.18347335

## REFERENCES


[1] S. Chong, "Business process management for SMEs: an exploratory study of implementation factors for the Australian wine industry," *Journal of Information Systems and Small Business*, vol. 1, no. 1–2, pp. 41–58, 2007.

[2] M. Kajba, B. Jereb, and R. Gumzej, "Business process reengineering–process optimization of boutique production SME," *Montenegrin Journal of Economics*, vol. 18, no. 4, pp. 117–140, 2022.

[3] I. Zaballa, M. Merino, and J. D. Villarroel, "Children's pictorial expression of plant life and its connection with school-based greenness," *Sustainability (Switzerland)*, vol. 13, no. 9, May 2021, doi: 10.3390/su13094999.

[4] B. J. Hicks, S. J. Culley, C. A. McMahon, and P. Powell, "Understanding information systems infrastructure in engineering SMEs: A case study," *Journal of Engineering and Technology Management*, vol. 27, no. 1–2, pp. 52–73, 2010.

[5] S. Saidi, L. Kattan, P. Jayasinghe, P. Hettiaratchi, and J. Taron, "Integrated infrastructure systems—A review," *Sustain. Cities Soc.*, vol. 36, pp. 1–11, 2018.

[6] M. Á. Gómez, R. Medina, A. S. Leicht, S. Zhang, and A. Vaquera, "The performance evolution of match play styles in the Spanish professional basketball league," *Applied Sciences (Switzerland)*, vol. 10, no. 20, pp. 1–9, Oct. 2020, doi: 10.3390/app10207056.

[7] J. González-Ramírez, H. Cheng, and S. Arral, "Funding campus sustainability through a green fee—estimating students' willingness to pay," *Sustainability (Switzerland)*, vol. 13, no. 5, pp. 1–15, Mar. 2021, doi: 10.3390/su13052528.

[8] I. Odun-Ayo, M. Ananya, F. Agono, and R. Goddy-Worlu, "Cloud computing architecture: A critical analysis," in *2018 18th international conference on computational science and applications (ICCSA)*, IEEE, 2018, pp. 1–7.

[9] P. Harmon, *Business process change: a manager's guide to improving, redesigning, and automating processes*. Morgan Kaufmann, 2003.

[10] M. Andriani, T. M. A. A. Samadhi, J. Siswanto, and K. Suryadi, "Aligning business process maturity level with SMEs growth in Indonesian fashion industry," *International Journal of Organizational Analysis*, vol. 26, no. 4, pp. 709–727, 2018.

[11] F. N. Pratiwi and M. Dachyar, "SME's Business Process Improvement in Food Industry Using Business Process Re-Engineering Approach," *International Journal of Advanced Science and Technology*, vol. 29, no. 7s, pp. 3665–3674, 2020.

[12] P. Q. Huy and V. K. Phuc, "Unveiling how business process management capabilities foster dynamic decision-making for effectiveness of sustainable digital transformation," *Business Process Management Journal*, vol. 31, no. 8, pp. 67–103, 2025, doi: 10.1108/BPMJ-06-2024-0467.

[13] O. T. Olajide and O. I. Okunbanjo, "Effects of Business Process Reengineering on Organisational Performance in the Food and Beverage Industry in Nigeria," *Journal of*

*Business and Management Research*, vol. 3, no. 1–2, pp. 57–74, Oct. 2020, doi: 10.3126/jbmr.v3i1.32029.

[14] A. N. Toosi, R. N. Calheiros, and R. Buyya, "Interconnected cloud computing environments: Challenges, taxonomy, and survey," *ACM Computing Surveys (CSUR)*, vol. 47, no. 1, pp. 1–47, 2014.

[15] A. P. Rajan, "Evolution of cloud storage as cloud computing infrastructure service," *arXiv preprint arXiv:1308.1303*, 2013.

[16] W. Liu, "Research on cloud computing security problem and strategy," in *2012 2nd international conference on consumer electronics, communications and networks (CECNet)*, IEEE, 2012, pp. 1216–1219.

[17] Hamed Taherdoost, "An Overview of Trends in Information Systems: Emerging Technologies that Transform the Information Technology Industry," *Cloud Computing and Data Science*, pp. 1–16, Aug. 2022, doi: 10.37256/ccds.4120231653.

[18] A. Mkhize, K. D. Mokhothu, M. Tshikhotho, and B. A. Thango, "Evaluating the Impact of Cloud Computing on SME Performance: A Systematic Review," *Businesses*, vol. 5, no. 2, p. 23, May 2025, doi: 10.3390/businesses5020023.

[19] R. Saputra, "Goods Invontory Management System at Nusantara Bakery Company Based on Android," *Journal of Scientech Research and Development*, vol. 6, no. 1, pp. 384–393, 2024.

[20] K. Nickson, "A BAKERY MANAGEMENT INFORMATION SYSTEM (BMIS) CASE STUDY :( HOT LOAF BAKERY MBARARA BRANCH) A PROJECT REPORT SUBMITTED TO THE SCHOOL OF COMPUTER STUDIES FOR THE STUDY LEADING TO A PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF A DEGREE OF BACHELORS OF INFORMATION TECHNOLOGY AT KAMPALA INTERNATIONAL UNIVERSITY," 2009.

[21] R. K. Pundir and P. Jain, "Qualitative and quantitative analysis of microflora of Indian bakery products," *Journal of Agricultural Technology*, vol. 7, no. 3, pp. 751–762, 2011.

[22] S. A. A. Ravishan, E. M. A. I. B. Ekanayaka, B. M. Edirisinghe, N. H. Manimendra, D. I. De Silva, and S. M. D. T. H. Dias, "Automated Bakery Management System," *International Journal of Engineering and Management Research*, vol. 12, no. 5, 2022, doi: 10.31033/ijemr.12.5.67.

[23] T. R. Merlo, F. Fard, and S. Hawamdeh, "Cloud Computing's Impact on the Digital Transformation of the Enterprise: A Mixed-Methods Approach," *Sustainability (Switzerland)*, vol. 17, no. 13, Jul. 2025, doi: 10.3390/su17135755.

[24] L. Farida Panduwinata, W. T. Subroto, and N. C. Sakti, "Digitalization on Micro, Small, and Medium Enterprises (MSMEs) : A Systematic Literature Review," *International Journal of Economics*, pp. 397–409, 2025, doi: 10.62951/ijecm.v1i4.435.

[25] A. Kargas *et al.*, "A Systematic Literature Review on SMEs Digital Transformation," Oct. 27, 2025. doi: 10.20944/preprints202510.2077.v1.

[26] A. Mkhize, K. D. Mokhothu, M. Tshikhotho, and B. A. Thango, "Evaluating the Impact of Cloud Computing on SME Performance: A Systematic Review," *Businesses*, vol. 5, no. 2, p. 23, May 2025, doi: 10.3390/businesses5020023.

[27] A. N. Raside and A. A. Ramli, "Development of Online Bakery Shop Web Application (e-Bakery)," *Applied Information Technology And Computer Science*, vol. 2, no. 2, pp. 1846–1859, 2021.

[28] E. S. Soegoto, F. A. Purnama, and A. Hidayat, "Role of internet and social media for promotion tools," in *IOP Conference Series: Materials Science and Engineering*, IOP Publishing, 2018, p. 012040.

[29] K. Vonderschmitt, "The growing use of social media in political campaigns: How to use Facebook, Twitter and YouTube to create an effective social media campaign," 2012.

[30] L. Yan Yun, M. Hamdi Irwan Hamzah, and F. Sains Komputer dan Teknologi Maklumat, "The Development of Caketisserie Bakery & Café Online Cake Ordering System," *Applied Information Technology And Computer Science*, vol. 4, no. 2, pp. 1412–1430, 2023, doi: 10.30880/aitcs.2023.04.02.080.

[31] S. S. Veleva and A. I. Tsvetanova, "Characteristics of the digital marketing advantages and disadvantages," in *IOP Conference Series: Materials Science and Engineering*, IOP Publishing, 2020, p. 012065.

[32] D. Chaffey and P. R. Smith, *Digital marketing excellence: planning, optimizing and integrating online marketing*. Taylor & Francis, 2022.

[33] R. Fauzan, M. F. Shiddiq, and N. R. Raddlya, "The designing of warehouse management information system," in *IOP Conference Series: Materials Science and Engineering*, IOP Publishing, 2020, p. 012054.

[34] R. Fauzan, D. Siahaan, S. Rochimah, and E. Triandini, "Use case diagram similarity measurement: A new approach," in *2019 12th International Conference on Information & Communication Technology and System (ICTS)*, IEEE, 2019, pp. 3–7.

[35] A. D. Pop, G. Rus, and R. F. Drența, "Modeling and simulation of technological factors in bakery industry," in *Advances in Manufacturing Engineering and Materials: Proceedings of the International Conference on Manufacturing Engineering and Materials (ICMEM 2018), 18–22 June, 2018, Nový Smokovec, Slovakia*, Springer, 2019, pp. 531–538.

ISSN 2085-4579

# Redundancy-Aware Feature Selection using mRMR and F-Test for EEG Emotion Classification

Ira Febrianti[1], Carens Chanda Claudhyta Hasan[2], Nadzifatu Chomtsa[3], Hanifa Khairunisa[4], Deandra Faysa Mardatila[5], Yuri Pamungkas[6*]
[1-6] Department of Medical Technology, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
yuri@its.ac.id

Accepted on April 04th, 2026
Approved on June 21st, 2026

***Abstract*—Emotions play an essential role in human interaction, driving the development of reliable automatic emotion recognition systems. Electroencephalography (EEG) offers a noninvasive method to record neural activity related to emotional states; however, many existing studies focus on limited feature configurations or binary classification problems. This research examines the influence of feature dimensionality and classifier selection on three-class EEG-based emotion recognition involving positive, neutral, and negative categories. The primary contribution of this study is a systematic assessment of feature and classifier compatibility across 28 experimental scenarios within a unified evaluation framework. Using a publicly available EEG dataset containing statistical and spectral features, selection was conducted using F-test and Minimum Redundancy Maximum Relevance (mRMR) methods, isolating the top 5, 10, and 15 features alongside the complete set. Four classifiers (Random Forest, Support Vector Machine, K-Nearest Neighbors, and Neural Networks) were evaluated via a 70/30 hold-out validation scheme using accuracy, F1-score, and Area Under the Curve (AUC). Results indicate that Random Forest trained with the full feature set achieved the highest performance, reaching 99.53% accuracy and 0.9994 AUC. These findings suggest that ensemble-based models demonstrate greater robustness when handling high-dimensional EEG features in multi-class emotion recognition.**

***Index Terms*—EEG; Emotion Recognition; Machine Learning; Feature Selection**

## I. INTRODUCTION

Emotions are fundamental to human social interaction and societal functioning because they support coordination, communication, and group cohesion that developed alongside the growth of human communities over evolutionary timescales [1]. Beyond internal experiences, emotions also operate interpersonally and are shaped by social context, while simultaneously shaping others' responses and decisions [2]. The emotions as social information model further explains that emotional expressions convey cues about the expresser's feelings, goals, and intentions, which then influence observers through affective responses and cognitive inferences [3]. These perspectives motivate the need for reliable emotion recognition methods, especially those that can capture emotional states objectively rather than relying solely on self-report or behavioral observation.

Electroencephalography has become a prominent approach for emotion recognition because it is non-invasive and provides real-time measurements of brain electrical activity related to central nervous system dynamics [4]. EEG-based emotion recognition is commonly formulated as a signal-processing and pattern-recognition pipeline that includes acquisition, preprocessing, feature extraction, feature selection, and classification [5]. However, a persistent research problem remains that many EEG-derived features are high-dimensional and redundant, while EEG signals themselves are noisy and non-stationary. This combination can degrade generalization performance, inflate computational cost, and complicate interpretability, particularly when models are trained on limited samples relative to the number of extracted features [6].

State-of-the-art studies have explored diverse feature representations and classifiers to improve performance. Multi-feature fusion in the time domain and composite representations based on wavelet analysis and entropy have reported accuracies above 70% in common emotion dimensions such as valence and arousal [7]. Other work has investigated spectral and time-frequency strategies including Gabor-based spectral descriptors, wavelet transforms, and general time-frequency analysis, as well as entropy-driven feature families [8], [9], [10]. In parallel, machine learning classifiers such as neural networks, support vector machines, and random forests have been widely adopted for EEG emotion classification and have achieved competitive results across datasets [11], [12], [13]. These efforts collectively show that both feature engineering and model choice strongly affect recognition accuracy, yet they also suggest that improvements are often constrained by the quality and compactness of the selected feature set.

Prior studies highlight the importance of feature selection and discriminative representations, especially in relation to frequency bands. A mutual information-based feature selection combined with kernel classifiers achieved around 73% accuracy for both arousal and valence, indicating that reducing irrelevant information can improve prediction [14]. Differential entropy features have been shown to outperform conventional energy spectrum descriptors, with reported accuracy above 84%, and gamma-band activity was found to be particularly associated with emotional variation [15]. Comparisons across multiple feature extraction families indicate that time-domain statistical characteristics can provide strong results, with accuracies around the high 70% range for valence, arousal, and dominance [16]. More recent evaluations comparing extraction, reduction, and classification approaches reported strong binary performance for kernel spectral regression with random forest, while again reaffirming gamma rhythm as informative for emotional changes [17]. Additional hybrid reduction schemes such as combining PCA with statistical screening have also reported mid-80% performance with SVM, and Hjorth-parameter-based pipelines with ANOVA-based selection have achieved accuracy in the mid-70% range using ensemble classifiers [18], [19]. Cross-dataset analyses further suggest that the best-performing features and classifiers can vary by dataset and protocol, although certain families such as fractal-dimension features can be consistently competitive when paired with appropriate classifiers [20].

Despite these advances, the research gap is that many pipelines either rely on large feature sets that are prone to redundancy or apply a single selection criterion that may not simultaneously maximize relevance to the target labels and minimize overlap among chosen features. This study addresses that gap by proposing a redundancy-aware feature selection strategy that integrates minimum redundancy maximum relevance with F-test screening for EEG emotion classification. The solution is designed to prioritize features that are both discriminative and non-redundant, thereby reducing dimensionality while preserving informative variability related to emotional states. We extract a broad set of statistical and mathematical descriptors including mean, minimum, maximum, covariance matrix representations, matrix logarithm features, correlation measures, and frequency-domain components via Fast Fourier Transform, and then apply the proposed selection strategy prior to classification.

The novelty of this work lies in combining redundancy-aware selection with statistical significance screening on a heterogeneous feature pool spanning time-domain statistics, matrix-based descriptors, and frequency components, aiming to yield a compact and informative subset for emotion prediction. The contribution of the research is to introduce and evaluate an integrated feature selection pipeline using minimum redundancy maximum relevance and F-test screening to improve EEG emotion classification performance while reducing feature redundancy and computational burden. The contribution of the research is also to provide an empirical analysis of how the selected non-redundant features impact classifier performance for emotion dimensions such as valence and arousal using common machine learning models, thereby supporting more robust and interpretable EEG-based affect recognition.

## II. Prior Works

Table 1 provides a consolidated summary of representative studies published between 2018 and 2025 on EEG-based emotion recognition. It brings together recent contributions that reflect how the field has progressed through different combinations of feature engineering, feature selection, and machine learning classification, while reporting performance using commonly adopted metrics. This synthesis is intended to offer a clear snapshot of current research directions and typical performance ranges reported in recent literature.

TABLE I. Summary of Previous Research

| Authors | Year | Method | Evaluation | Key finding |
|---|---|---|---|---|
| Vempati, et al. [21] | 2023 | mRMR feature selection plus ensemble ML (subspace KNN, SVM, ANN) on multichannel EEG rhythms | Accuracy, LOSOCV | mRMR-selected gamma rhythm features with ensemble KNN reached 93.5–99.8%, confirming mRMR effectiveness for EEG emotion classification. |
| Zhang, et al. [22] | 2023 | ML (RF, SVM, KNN, XGBoost) with DWT and PCA | Accuracy, Recall, Precision, F1, ROC | Higher-frequency bands, especially gamma, were most informative. Beta and gamma correlated with emotional dynamics, with frontal sites reflecting agitation. |
| Abdumalikov, et al. [23] | 2024 | ML (RF, LR, XGBoost, SVM) plus feature selection (ANOVA F-test/LASSO/GA) with Bayesian optimization | Accuracy, hold-out validation | RF with LASSO achieved 99.39% on EEG Emotion dataset, consistently outperforming baseline across three feature selection methods. |
| Hamzah, et al. [24] | 2024 | Deep learning with time, frequency, and time-frequency features | Not stated explicitly | Review indicates increasing use of deep learning for complex EEG patterns, with feature choice remaining critical. |
| Rios, et al. [25] | 2024 | ML with PCA and PSD, RF and Extra-Trees | Accuracy, F1-score | Real-time VAD estimation system. Extra-Trees reached ~94% overall accuracy, with optimized channels including frontal, temporal, parietal, and occipital sites. |

| Authors | Year | Method | Evaluation | Key finding |
|---|---|---|---|---|
| Wang, et al. [26] | 2024 | GAN-based deep model with self-attention and residual design | Emotion detection accuracy | GAN framework improved emotion detection accuracy and addressed training instability such as vanishing gradients. |
| Mouazen, et al. [27] | 2025 | Hybrid ML and DL models, CNNs, RNNs, transformers, autoencoders, SHAP | Peak accuracy range | Hybrid approaches reported peak accuracy around 85–95% and emphasized interpretability and real-time feasibility. |

## III. METHODOLOGY

The workflow of this study is organized as a sequential pipeline that transforms EEG recordings into final emotion classification outcomes. It begins with the available EEG data representing three emotional states, namely positive, neutral, and negative, and proceeds through standardized processing stages that ensure the data are suitable for machine learning analysis and performance evaluation. This structure follows common EEG-based emotion recognition practice, where each stage is designed to reduce noise and complexity while preserving emotion-related information needed for accurate prediction.

The preprocessing stage converts continuous EEG recordings into consistent and analyzable segments. In the original data collection protocol, the raw signals were resampled and downsampled to a uniform sampling rate, then segmented using a sliding-window approach to handle the non-stationary nature of EEG. Feature extraction was subsequently performed to generate a high-dimensional feature representation, yielding 2,549 attributes that summarize relevant temporal and spectral characteristics of brain activity across standard EEG frequency bands. This feature-based representation supports efficient learning while maintaining a broad coverage of signal properties relevant to emotional variation.

Dimensionality reduction in this study is achieved through feature selection rather than transformation-based methods. Two statistical strategies are applied to identify informative features and reduce redundancy. The F-test ranks features by how strongly they differ across emotion classes, while minimum redundancy maximum relevance selects features that are highly associated with the target labels and minimally overlapping in information content. This combination is intended to produce compact feature subsets that remain discriminative, improve generalization, and reduce computational burden during model training and testing.

The selected features are then used as inputs to multiple classifiers to evaluate how feature selection influences emotion recognition performance. The study employs support vector machines, k-nearest neighbors, random forests, and neural networks, providing a comparative perspective across commonly used machine learning paradigms for EEG classification tasks. Model performance is assessed using complementary metrics that include precision, recall, F1-score, area under the curve, true negative rate, and accuracy, enabling both class-wise discrimination assessment and overall correctness evaluation. This evaluation strategy supports a systematic comparison of classifier behavior under different feature selection settings and feature subset sizes.

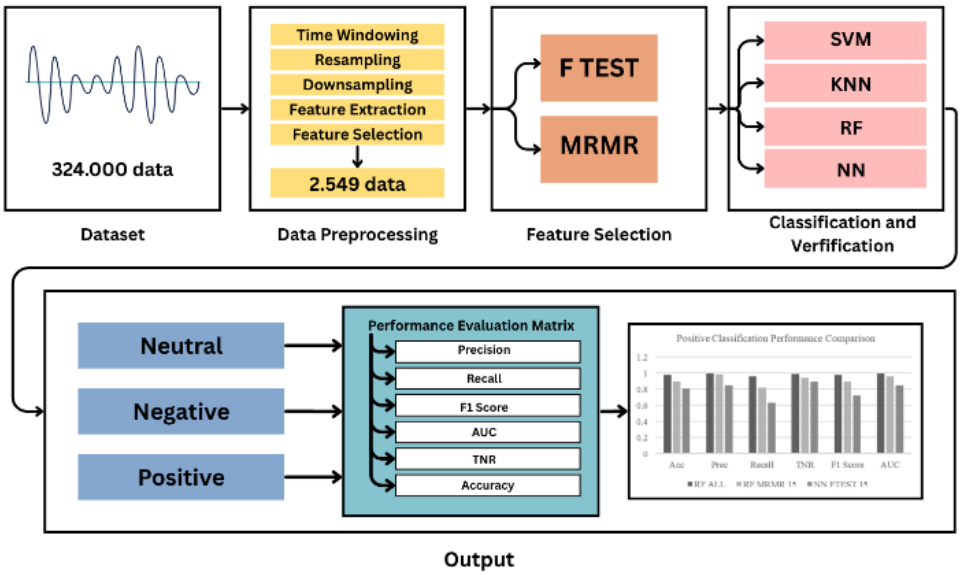


Fig. 1. Methodology of the research

### A. Dataset

This study used a publicly available dataset entitled "Detecting Emotions Using EEG Waves", which was developed by Karn Suksomkit and released on Kaggle [28]. The dataset contains electroencephalogram recordings that are labeled into three emotion categories, namely positive, neutral, and negative. Each sample is represented by a set of features that were extracted from the underlying raw EEG signals, enabling supervised learning for EEG-based emotion classification. The feature set includes common time-domain descriptive statistics that summarize the distribution of EEG amplitudes. These features include the mean as an estimate of average signal level, as well as minimum and maximum values that capture extreme amplitudes within the signal segment. Such statistical descriptors provide compact summaries of overall signal behavior and are frequently used as baseline representations in emotion-related EEG modeling.

In addition to descriptive statistics, the dataset provides matrix-based and inter-channel relationship features. Covariance matrices are included to represent signal variance and cross-channel dependencies, reflecting how EEG channels co-vary during emotional processing [29]. The matrix logarithm is also provided as a transformation applied to covariance matrices to map them into a space that is often more suitable for learning algorithms [30]. Correlation coefficients further quantify linear relationships among channels, offering additional information about functional coupling patterns that may vary with emotional states [31]. The dataset also incorporates frequency-domain information derived using the Fast Fourier Transform, which converts EEG signals from the time domain into spectral components. These features aim to capture brainwave-related patterns in the frequency spectrum that are relevant to emotional processing [32]. By combining time-domain statistical descriptors, inter-channel measures, and spectral representations, the dataset supports comprehensive modeling of EEG dynamics for emotion recognition tasks.

### B. Preprocessing

The EEG dataset used in this study was originally reported in earlier work that investigated emotion classification using a Muse headband device. Data were collected from two participants, one male and one female, aged 20 to 22 years. For each emotional condition, EEG activity was recorded for three minutes. In addition, six minutes of neutral resting-state data were obtained, resulting in recordings representing positive, negative, and neutral emotional states. EEG acquisition was conducted using a commercial Muse headband equipped with four dry electrodes placed at TP9, AF7, AF8, and TP10 in accordance with the international EEG placement system. Emotional states were elicited through audiovisual stimuli [33]. Negative emotions were induced using selected scenes from Marley and Me, Up, and My Girl, while positive emotions were elicited using La La Land, Slow Life, and Funny Dogs. The neutral condition was recorded without external stimuli and was performed before the emotional sessions to reduce potential carryover effects that could influence subsequent recordings [34].

The raw EEG signals were resampled and down-sampled to a uniform sampling rate of 150 Hz, producing a total of 324,000 data points [35]. To handle the non-stationary nature of continuous EEG streams, the recordings were segmented using a one-second sliding window with a 0.5-second overlap, generating shorter epochs suitable for analysis. From these segments, a total of 2,549 attributes were extracted, including statistical descriptors computed across multiple canonical frequency bands comprising delta, theta, alpha, beta, and gamma. Feature selection was subsequently applied using evaluators such as Information Gain and Symmetrical Uncertainty to identify the most informative predictors and improve computational efficiency [36]. To reduce electromyographic contamination, participants were instructed to avoid intentional movements during the recording sessions [37]. At the same time, natural behaviors such as blinking were neither explicitly encouraged nor suppressed in order to preserve ecological validity. Prior evidence indicates that blink dynamics are associated with cognitive demand and attentional processes, and thus may provide informative signals that contribute to emotion-related classification. Participants were also instructed to keep their eyes open during the tasks to maintain consistent recording conditions across sessions [38].

### C. Feature Selection

Unlike studies that start from raw EEG signal processing, this research used a structured dataset obtained from Kaggle that already provides pre-extracted features derived from EEG recordings. The available attributes include both statistical descriptors and spectral characteristics, allowing the analysis to proceed directly at the feature level rather than at the continuous-signal level [39]. Because the dataset is feature-based, preprocessing in this study emphasized selecting informative variables and building classification models, rather than applying signal filtering, artifact correction, or other waveform-level operations [40]. All preprocessing and analysis were implemented in MATLAB 2024. Given the high dimensionality of the feature space, the study applied feature selection to reduce complexity and improve learning stability. Two statistical selection approaches were used, namely the F-test and minimum redundancy maximum relevance. The F-test was used to assess the discriminative association between individual features and the emotion labels, supporting the identification of features that vary significantly across classes [41].

Minimum redundancy maximum relevance was then employed to prioritize features that are strongly related to the emotion class labels while simultaneously

minimizing redundancy among the selected features. This strategy aims to retain complementary information by reducing overlap between features that carry similar patterns, thereby improving the compactness and interpretability of the final feature subset used for classification [42]. The ranked feature list produced by the minimum redundancy maximum relevance procedure is shown and summarized in Figure 2 and Table 2.

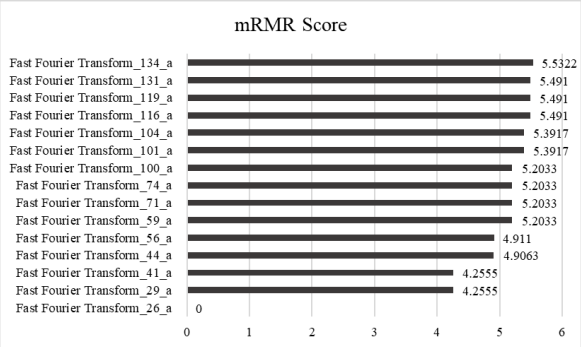


Fig. 2. Top 15 features selected using mRMR Method

TABLE II. TOP 15 FEATURES SELECTED USING MRMR METHOD

| Rank | Features | mRMR Score |
|---|---|---|
| 1 | Fast Fourier Transform_746_b | 5.5322 |
| 2 | Mean_0_a | 5.4910 |
| 3 | Fast Fourier Transform_748_b | 5.4910 |
| 4 | Log Magnitude_72_a | 5.4910 |
| 5 | Covariance Matrix_50_a | 5.3917 |
| 6 | Log Magnitude_60_a | 5.3917 |
| 7 | Covariance Matrix_30_a | 5.2033 |
| 8 | Covariance Matrix_99_b | 5.2033 |
| 9 | Fast Fourier Transform_233_a | 5.2033 |
| 10 | Log Magnitude_4_b | 5.2033 |
| 11 | Correlate_59_b | 4.9110 |
| 12 | Moments_2_a | 4.9063 |
| 13 | Fast Fourier Transform_491_a | 4.2555 |
| 14 | Minimum Quantile_19_a | 4.2555 |
| 15 | Maximum Quantile_17_b | 3.5684 |

After applying minimum redundancy maximum relevance, the F-test was used to further evaluate the statistical relevance of individual features across the emotion classes [43]. This step aimed to quantify how strongly each feature differs among class groups, providing an additional perspective for prioritizing discriminative attributes before classification. The F-test assesses the ratio between between-class variance and within-class variance for a given feature. Features that produce higher F-scores indicate larger differences among class means relative to the variability within each class, and are therefore considered to have stronger discriminative capability for emotion classification [44]. The resulting feature rankings obtained from the F-test are summarized in Table 3.

TABLE III. FEATURE RANKING RESULTS USING F-TEST METHOD

| Rank | Features | F-test Score |
|---|---|---|
| 1 | Fast Fourier Transform_26_a | Inf |
| 2 | Fast Fourier Transform_29_a | Inf |
| 3 | Fast Fourier Transform_41_a | Inf |
| 4 | Fast Fourier Transform_44_a | Inf |
| 5 | Fast Fourier Transform_56_a | Inf |
| 6 | Fast Fourier Transform_59_a | Inf |
| 7 | Fast Fourier Transform_71_a | Inf |
| 8 | Fast Fourier Transform_74_a | Inf |
| 9 | Fast Fourier Transform_100_a | Inf |
| 10 | Fast Fourier Transform_101_a | Inf |
| 11 | Fast Fourier Transform_104_a | Inf |
| 12 | Fast Fourier Transform_116_a | Inf |
| 13 | Fast Fourier Transform_119_a | Inf |
| 14 | Fast Fourier Transform_131_a | Inf |
| 15 | Fast Fourier Transform_134_a | Inf |

The results indicate that all selected features produced infinite F-scores, which suggests extremely strong separation among the emotion classes under the F-test criterion. In practical terms, an infinite F-score typically arises when the within-class variance for a feature approaches zero while between-class differences remain non-zero, leading to a ratio that diverges. This outcome implies that, for the evaluated samples, the corresponding features exhibit highly consistent values within each class while differing markedly across classes, and therefore appear to be highly discriminative according to the statistical test [45]. To examine how classification performance changes as the feature space becomes more compact or more expressive, subsets of the highest-ranked features were formed based on the F-test ordering. Specifically, three feature dimensions were evaluated by selecting the top 5, top 10, and top 15 ranked features for subsequent classification experiments. This staged selection supports a systematic comparison of model performance under different levels of dimensionality and helps assess whether a smaller subset can achieve comparable accuracy to larger feature sets, while potentially improving efficiency and reducing redundancy effects [46].

### *D. Classification Models*

In this study, emotion classification was formulated as a single-label multi-class task in which each sample was assigned to exactly one category among positive,

neutral, and negative emotional states. All classification procedures were implemented in MATLAB 2024 using the Classification Learner App, ensuring a consistent workflow for training, validation, and performance reporting across models. To examine how feature selection influences predictive performance, four widely used classifiers were evaluated, namely K-Nearest Neighbors, Support Vector Machine, Neural Network, and Random Forest. Each classifier was first trained using the complete feature set to establish a baseline. The same models were then re-trained and re-tested using reduced feature subsets produced by the F-test and minimum redundancy maximum relevance methods. For both selection strategies, three subset sizes were considered, consisting of 5, 10, and 15 features, to assess performance under different dimensionalities.

Across the full-feature baseline and the reduced-feature settings, a total of 28 experimental scenarios were conducted to enable a structured comparison of classifier behavior across alternative feature configurations. Model evaluation followed a hold-out validation strategy in which 70% of the samples were used for training and the remaining 30% were reserved for testing. All analyses were performed within the MATLAB environment, including confusion matrix generation and metric computation to support consistent and reproducible reporting. The primary performance measure was the area under the curve, computed on a per-class basis to reflect discrimination capability for each emotional category. Accuracy and F1-score were also reported to provide a complementary view of overall correctness and class-sensitive predictive balance, thereby enabling a more comprehensive interpretation of classification performance across models and feature selection conditions.

The following hyperparameters were applied to each classifier using the default presets in MATLAB 2024 Classification Learner. Random Forest (Bagged Trees): learner type = Decision Tree, number of learners = 30, maximum number of splits = 1492. Support Vector Machine (SVM): kernel = Linear, box constraint C = 1, kernel scale = automatic, multiclass coding = One-vs-One, with data standardization enabled. K-Nearest Neighbors (KNN): preset = Fine KNN, number of neighbors K = 1, distance metric = Euclidean, distance weight = equal, with data standardization enabled. Neural Network: preset = Wide Neural Network, hidden layer sizes = 100 (1 layer), activation = ReLU, iteration limit = 1000. These settings were held constant across all 28 experimental scenarios to ensure fair comparison.

### *E. Feature Extraction*

In this study, feature extraction was not performed directly from raw EEG signals because the analysis relied on a publicly available Kaggle EEG dataset that already provides a rich set of pre-extracted features. The feature set is designed to represent multiple aspects of EEG activity under different emotional conditions, including statistical descriptors, inter-channel relationship measures, and frequency-domain characteristics. This approach allows the workflow to focus on feature selection and classification while still capturing essential temporal and spectral information relevant to emotion recognition [47].

Basic statistical descriptors such as mean, minimum, and maximum were included to summarize the amplitude characteristics of EEG segments. The mean reflects the average signal amplitude within a segment, while the minimum and maximum values represent the lowest and highest amplitudes, capturing the intensity and variability of neural activity. These features provide compact temporal summaries and are commonly used in EEG-based emotion recognition because they can reflect changes in baseline activity and amplitude fluctuations that occur across emotional states [48]. The mean of a signal segment can be expressed as

$$Mean = \frac{1}{N}\sum_{i=1}^{N} x_i \quad (1)$$

where $x_i$ denotes the signal samples and $N$ is the number of samples in the segment.

To represent coordinated activity across EEG channels, the dataset also includes covariance matrices. Covariance captures linear co-variation between channels and can reflect shared neural dynamics that may change under emotional stimulation. Because covariance matrices are positive semi-definite and are often treated as lying on a Riemannian manifold, applying the matrix logarithm maps them into a Euclidean space, which facilitates their use in conventional machine learning pipelines. The covariance between two channel signals $x$ and $y$ can be written as

$$\mathrm{Cov}(x,y) = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y}) \quad (2)$$

where $x_i$ and $y_i$ are samples from two channels, $N$ is the number of samples, and $\bar{x}$ and $\bar{y}$ are the corresponding channel means.

Correlation features were further used to quantify the similarity of signal patterns across channels in a normalized manner. Correlation can help identify synchronization or functional coupling, which may vary depending on emotional processing and stimulus engagement. The correlation coefficient is defined as

$$\mathrm{Corr}(x,y) = \frac{\mathrm{Cov}(x,y)}{\sigma_x \sigma_y} \quad (3)$$

where $\sigma_x$ and $\sigma_y$ are the standard deviations of $x$ and $y$, respectively. The coefficient ranges from $-1$ to 1, with values closer to $\pm 1$ indicating stronger linear association.

In addition to time-domain and relational descriptors, frequency-domain features were derived using the Fast Fourier Transform to compute spectral

representations efficiently [49]. FFT enables transformation of EEG signals from the time domain into the frequency domain so that energy or power distribution across frequency bins can be analyzed. Such spectral information is highly relevant to emotion recognition because canonical EEG bands including theta, alpha, beta, and gamma are linked to cognitive and affective processes such as relaxation, alertness, and emotional arousal [50]. The discrete Fourier transform of a length-$N$ discrete signal $x(n)$ is given by

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-j\frac{2\pi kn}{N}}, \quad k = 0,1,\dots,N-1 \quad (4)$$

where $x(n)$ is the time-domain EEG signal, $k$ is the frequency index, $N$ is the number of samples, and $j$ denotes the imaginary unit.

### F. Dimensionality Reduction

Rather than relying on transformation-based dimensionality reduction techniques such as Principal Component Analysis, this study adopted statistical feature selection to reduce dimensionality at the feature level. The main objective was to identify a compact subset of informative features from the high-dimensional EEG feature space while preserving interpretability, improving classification effectiveness, and reducing computational cost during model training and testing. Two complementary selection strategies were applied, namely the F-test and minimum redundancy maximum relevance. The F-test was used to quantify how strongly each feature differentiates the target emotion classes by evaluating the ratio between between-class variation and within-class variation. Features with larger F-scores were considered more discriminative and were therefore prioritized for downstream classification.

Minimum redundancy maximum relevance was used to address a common limitation of purely univariate ranking methods, which may select features that are individually strong yet highly overlapping in information content. This method selects features that are highly associated with the class labels while simultaneously minimizing redundancy among the chosen features, encouraging a feature subset that is both relevant and diverse in the information it carries. Through these selection procedures, the original feature space containing more than 2,500 features was reduced into several compact subsets. The top-ranked 5, 10, and 15 features were retained as inputs for classification experiments, enabling a systematic evaluation of model performance under different feature dimensionalities and supporting a clear assessment of the trade-off between compactness and predictive accuracy.

## IV. RESULTS

In the initial phase of this study, four classifiers were evaluated, comprising Random Forest, Neural Network, K-Nearest Neighbors, and Support Vector Machine, using feature subsets generated by the F-test and minimum redundancy maximum relevance methods. To assess the influence of dimensionality and to understand each classifier's sensitivity to feature reduction, experiments were conducted using the full feature set and reduced subsets consisting of the top 5, top 10, and top 15 ranked features. In total, 28 experimental scenarios were executed: four configurations used the full feature set (one per classifier, without feature selection), while the remaining 24 reflected combinations of two feature selection methods (F-test and mRMR), three subset sizes (top 5, 10, and 15 features), and four classifier types.

The results showed notable variability across configurations, indicating that classification performance depends strongly on the interaction between the learning algorithm and the selected feature representation. Among all evaluated settings, three configurations consistently produced high performance across the positive, neutral, and negative emotion categories. These top-performing settings were Random Forest trained on the full feature set, Random Forest trained using 15 features selected by minimum redundancy maximum relevance, and Neural Network trained using 15 features selected by the F-test. The corresponding performance summaries are presented in Table 4 and Figure 3 for the positive class, Table 5 and Figure 4 for the neutral class, and Table 6 and Figure 5 for the negative class.

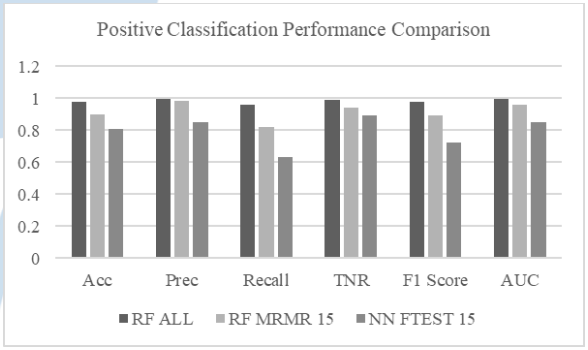


Fig. 3. Performance on Positive Emotion Classification

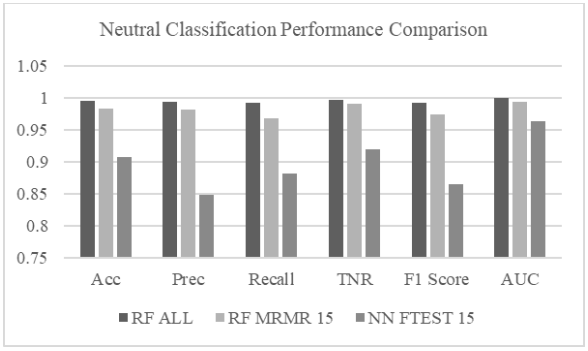


Fig. 4. Performance on Neutral Emotion Classification

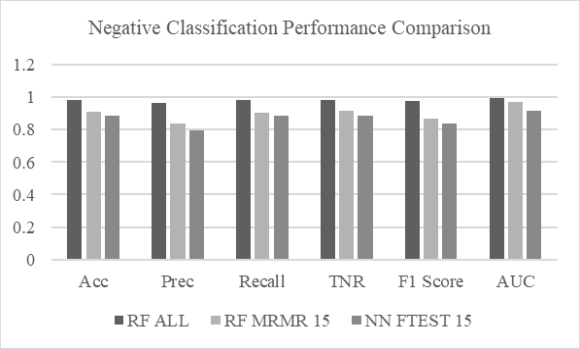


Fig. 5. Performance on Negative Emotion Classification

TABLE IV. PERFORMANCE COMPARISON OF CLASSIFIERS FOR POSITIVE EMOTION CLASSIFICATION

| Metrics Evaluation | RF ALL | RF mRMR 15 | NN FTEST 15 |
|---|---|---|---|
| Accuracy | 0.97790 | 0.90221 | 0.80643 |
| Precision | 0.99400 | 0.98178 | 0.84837 |
| Recall | 0.95766 | 0.82056 | 0.63105 |
| TNR | 0.98796 | 0.94283 | 0.89368 |
| F1 Score | 0.97549 | 0.89396 | 0.72375 |
| AUC | 0.99796 | 0.96020 | 0.85180 |

TABLE V. PERFORMANCE COMPARISON OF CLASSIFIERS FOR NEUTRAL EMOTION CLASSIFICATION

| Metrics Evaluation | RF ALL | RF mRMR 15 | NN FTEST 15 |
|---|---|---|---|
| Accuracy | 0.99531 | 0.98326 | 0.90757 |
| Precision | 0.99400 | 0.98178 | 0.84837 |
| Recall | 0.99202 | 0.96806 | 0.88224 |
| TNR | 0.99698 | 0.99093 | 0.92036 |
| F1 Score | 0.99301 | 0.97487 | 0.86497 |
| AUC | 0.99943 | 0.99460 | 0.96420 |

TABLE VI. PERFORMANCE COMPARISON OF CLASSIFIERS FOR NEGATIVE EMOTION CLASSIFICATION

| Metrics Evaluation | RF ALL | RF mRMR 15 | NN FTEST 15 |
|---|---|---|---|
| Accuracy | 0.98259 | 0.91092 | 0.88547 |
| Precision | 0.96443 | 0.83925 | 0.79385 |
| Recall | 0.98387 | 0.90524 | 0.88508 |
| TNR | 0.98195 | 0.91374 | 0.88566 |
| F1 Score | 0.97405 | 0.87100 | 0.83699 |
| AUC | 0.99643 | 0.97250 | 0.91930 |

For positive emotion classification (Table 4, Figure 3), RF ALL achieved the highest performance (Recall: 0.95766, F1-score: 0.97549, AUC: 0.99796). RF mRMR 15 yielded a Recall of 0.82056, F1-score of 0.89396, and AUC of 0.96020. NN FTEST 15 recorded the lowest performance among the three (Recall: 0.63105, F1-score: 0.72375, AUC: 0.85180).

For neutral emotion classification (Table 5, Figure 4), RF ALL achieved a Recall of 0.99202, F1-score of 0.99301, and AUC of 0.99943. RF mRMR 15 yielded a Recall of 0.96806, F1-score of 0.97487, and AUC of 0.99460. NN FTEST 15 recorded a Recall of 0.88224, F1-score of 0.86497, and AUC of 0.96420.

For negative emotion classification (Table 6, Figure 5), RF ALL achieved a Recall of 0.98387, F1-score of 0.97405, and AUC of 0.99643. RF mRMR 15 yielded a Recall of 0.90524, F1-score of 0.87100, and AUC of 0.97250. NN FTEST 15 recorded a Recall of 0.88508, F1-score of 0.83699, and AUC of 0.91930.

These findings demonstrate that classification performance varies notably across configurations, with Random Forest using the full feature set consistently achieving the highest results across all three emotion classes, while reduced-feature subsets and alternative classifiers showed varying degrees of performance decline depending on the selection method and subset size.

## V. DISCUSSIONS

The experimental results demonstrate that EEG-based emotion recognition is strongly influenced by both the selected feature subset and the classifier used for learning. In this study, Random Forest, Neural Network, K-Nearest Neighbors, and Support Vector Machine were evaluated using two commonly applied feature selection strategies, namely the F-test and minimum redundancy maximum relevance. Each classifier was assessed using multiple feature dimensionalities, including the top 5, top 10, top 15, and the full feature set, to examine the trade-off between compact representations and predictive effectiveness for positive, neutral, and negative emotional states.

Across the 28 experimental scenarios, Random Forest trained using the full feature set, denoted as RF ALL, consistently achieved the best overall performance for all emotion classes. Its strongest outcomes were observed for the neutral class, where it reached an accuracy of 99.53 percent, an F1 score of 0.9930, and an AUC of 0.9994. Comparable results for the positive and negative classes further indicate that Random Forest remains robust in high-dimensional EEG feature spaces, likely because its ensemble structure can accommodate feature redundancy and implicitly exploit feature importance during split optimization.

When stronger dimensionality reduction was applied, performance differences became more apparent across classifiers. Neural Network trained with the top 15 features selected by the F-test, denoted as NN FTEST 15, showed noticeably lower performance, particularly for the positive class, where the F1 score and AUC declined to 0.72375 and 0.8518, respectively. This pattern suggests that some classifiers are more sensitive to reduced feature representations, especially when informative interactions among features are removed or when the remaining subset is insufficient to support stable decision boundaries. In

such cases, Neural Network models may require architectural adjustment or additional regularization to better exploit limited input representations, particularly when the features are not learned end to end from the raw signal but provided as engineered descriptors.

At the same time, Random Forest trained using the top 15 features selected by minimum redundancy maximum relevance, denoted as RF mRMR 15, remained competitively strong. This indicates that redundancy-aware feature selection can preserve a large proportion of discriminative information when paired with an appropriate learner, while also reducing dimensionality and computational cost. The stable performance of Random Forest under both the full and reduced feature sets highlights its versatility and supports its suitability for settings where either maximum accuracy or a compact feature set is required.

The Support Vector Machine results showed notable sensitivity to feature dimensionality. Under the full feature set, SVM produced lower performance compared to configurations using reduced subsets selected by mRMR or F-test, which is consistent with the well-documented challenge of margin estimation in high-dimensional spaces where training samples are limited relative to feature count. The choice of kernel function and regularization parameter C are particularly important for SVM performance and should be carefully tuned for each feature configuration. K-Nearest Neighbors exhibited comparable sensitivity, performing more competitively under mRMR-reduced feature subsets where the non-redundant feature space supports more meaningful distance-based similarity computations. However, KNN was consistently outperformed by Random Forest across all configurations, likely due to KNN's susceptibility to irrelevant or overlapping features and its reliance on local neighborhood structures that may not generalize well in multi-class EEG emotion classification. These findings underline the importance of aligning feature selection strategy with the strengths and limitations of each classifier.

A broader comparison with prior work is summarized in Table 1, which reports recent studies on EEG-based emotion classification. Many prior approaches report moderate performance ranges and often focus on binary labeling or limited evaluation metrics. In contrast, the approach presented in this study achieves high three-class performance with strong discrimination measured by AUC and complementary reporting using precision, recall, and F1-score, providing a more complete assessment of classification behavior across emotion categories. This comparison supports the view that robust ensemble-based learning combined with well-chosen features can yield strong results even without end-to-end deep architectures.

Compared with previous studies summarized in Table 1, the proposed approach demonstrates several important advantages. First, this study evaluates a comprehensive combination of classifiers and feature selection strategies within a unified experimental framework consisting of 28 different scenarios, enabling a more systematic analysis of classifier–feature compatibility than most prior works, which commonly focus on a single model or limited feature configuration. Second, unlike studies that primarily emphasize binary emotion recognition, this work addresses a more challenging three-class classification problem involving positive, neutral, and negative emotional states while still achieving very high-performance metrics.

Another important advantage lies in the integration of redundancy-aware feature selection using mRMR combined with F-test evaluation. This approach not only reduces dimensionality but also preserves discriminative information effectively, as demonstrated by the strong performance of RF mRMR 15 despite using a substantially smaller feature subset. In comparison, several previous studies relied on larger feature spaces, deep architectures with higher computational requirements, or transformation-based reduction techniques that may reduce interpretability.

Furthermore, the proposed framework provides strong classification performance without requiring computationally expensive end-to-end deep learning architectures. The Random Forest model using the complete feature set achieved 99.53% accuracy and 0.9994 AUC for neutral emotion classification, outperforming or remaining competitive with many recent studies summarized in Table 1. These findings suggest that carefully optimized machine learning pipelines combined with effective feature selection strategies can achieve highly robust EEG emotion recognition while maintaining lower computational complexity and better interpretability.

The findings emphasize the importance of compatibility between the classifier and the selected feature representation. Feature selection can improve interpretability and efficiency, but its benefits depend on whether the classifier can effectively exploit the remaining information after dimensionality reduction. For high-resource scenarios, RF ALL provides the strongest option, whereas reduced-feature configurations such as RF mRMR 15 offer practical alternatives for resource-constrained deployment. Future work could investigate hybrid selection strategies, richer temporal and spectral representations, and models tailored for low-dimensional inputs, as well as validation on larger and more diverse datasets to improve generalizability and robustness.

Several limitations of this study merit acknowledgment. First, the EEG dataset was collected from only two participants using a consumer-grade Muse headband with four electrodes, which limits generalizability to broader populations and diverse acquisition settings. The small participant count and limited electrode coverage may not capture the full spatial distribution of emotion-related neural dynamics accessible through clinical-grade high-density EEG systems. Second, emotion elicitation relied on audiovisual stimuli with individually variable emotional responses, introducing subjectivity in label quality. Third, all classifiers were trained and

evaluated within a single-dataset framework without cross-dataset validation, meaning transferability of findings to different EEG protocols or emotional paradigms remains untested. Fourth, the use of pre-extracted engineered features rather than end-to-end deep learning pipelines may constrain the ceiling of achievable performance. Future studies should address these limitations through larger and more diverse participant cohorts, higher-density EEG systems, cross-dataset generalization experiments, and deep learning architectures that can learn directly from raw EEG signals.

## VI. CONCLUSION

This study examined how classifier choice and feature dimensionality influence EEG-based emotion recognition in a multi-class setting. Across the tested configurations, Random Forest trained using the full feature set consistently delivered the strongest performance for all emotion categories, achieving high accuracy and AUC values that exceeded those reported in several earlier studies. These findings indicate that ensemble learning methods can effectively exploit rich EEG feature representations and remain robust in high-dimensional feature spaces. Feature selection using the F-test and minimum redundancy maximum relevance substantially reduced the input dimensionality; however, the resulting impact on performance varied depending on the classifier. In particular, the Random Forest model using 15 mRMR-selected features maintained competitive accuracy and discrimination performance while using a significantly smaller feature subset, demonstrating that redundancy-aware feature selection can preserve important class-related information while improving computational efficiency. Nevertheless, the results also indicate that dimensionality reduction does not always guarantee superior classification performance, since the highest overall performance in this study was still achieved using the complete feature set.

In contrast, Neural Network models were more sensitive to feature reduction, particularly for the positive emotion class, suggesting that aggressive dimensionality reduction may remove informative feature interactions required to construct stable decision boundaries. Overall, these findings emphasize that EEG emotion recognition performance is influenced not only by the classifier or feature selection method individually, but also by the compatibility between the learning model and the selected feature representation. Random Forest appears particularly suitable for scenarios where classification accuracy is prioritized and computational resources allow the use of high-dimensional features, whereas reduced-feature configurations may provide practical advantages for real-time or resource-constrained implementations. This study also has several limitations, including the limited number of participants and the reliance on pre-extracted features from a public dataset, which may affect generalizability and restrict direct control over EEG preprocessing procedures. Future research should therefore investigate adaptive or hybrid feature selection approaches, more advanced deep learning architectures tailored to compact feature representations, and broader validation using larger and more diverse EEG datasets.

## ACKNOWLEDGMENT

The authors gratefully acknowledge the Department of Medical Technology, Institut Teknologi Sepuluh Nopember, for providing the facilities and institutional support that contributed to the completion of this research.

## REFERENCES

[1] M. M. N. Bieńkiewicz *et al.*, "Bridging the gap between emotion and joint action," *Neurosci. Biobehav. Rev.*, vol. 131, pp. 806–833, Dec. 2021, doi: 10.1016/J.NEUBIOREV.2021.08.014.

[2] M. Tekke, "Beyond emotion: social awareness as a key to wellbeing and reduced violence tendency in university students," *Humanit. Soc. Sci. Commun.*, vol. 13, no. 1, 2026, doi: 10.1057/s41599-025-06475-3.

[3] L. O. H. Kroczek, A. Lingnau, V. Schwind, C. Wolff, and A. Mühlberger, "Observers predict actions from facial emotional expressions during real-time social interactions," *Behav. Brain Res.*, vol. 471, p. 115126, 2024, doi: 10.1016/j.bbr.2024.115126.

[4] A. Chaddad, Y. Wu, R. Kateb, and A. Bouridane, "Electroencephalography Signal Processing: A Comprehensive Review and Analysis of Methods and Techniques," *Sensors*, vol. 23, no. 14, p. 6434, 2023, doi: 10.3390/s23146434.

[5] E. Gkintoni, A. Aroutzidis, H. Antonopoulou, and C. Halkiopoulos, "From Neural Networks to Emotional Networks: A Systematic Review of EEG-Based Emotion Recognition in Cognitive Neuroscience and Real-World Applications," *Brain Sci.*, vol. 15, no. 3, p. 220, 2025, doi: 10.3390/brainsci15030220.

[6] A. Mehrabi, N. Sreenivasan, U. Gunawardana, and G. Gargiulo, "Hybrid Spike-Encoded Spiking Neural Networks for Real-Time EEG Seizure Detection: A Comparative Benchmark," *Biomimetics*, vol. 11, no. 1, p. 75, 2026, doi: 10.3390/biomimetics11010075.

[7] F. M. Heir, "A hybrid CNN and reinforcement learning framework for speaker identification using Mel-Spectrogram and continuous wavelet transform features," *Sci. Rep.*, 2026, doi: 10.1038/s41598-026-35858-y.

[8] A. Orhan, "A Comparative Study of Time–Frequency Representations for Bearing and Rotating Fault Diagnosis Using Vision Transformer," *Machines*, 2025, doi: 10.3390/machines13080737.

[9] N. Filimonova, "Determination of the Time-frequency Features for Impulse Components in EEG Signals," *Neuroinformatics*, 2025, doi: 10.1007/s12021-024-09698-y.

[10] R. Ahmed, "A Lightweight Hybrid Gabor Deep Learning Approach and its Application to Medical Image Classification," *Int. J. Comput. Vis.*, 2026, doi: 10.1007/s11263-025-02658-2.

[11] S. S. Al-Hadithy, "Emotion recognition in EEG Signals: Deep and machine learning approaches, challenges, and future directions," *Comput. Biol. Med.*, 2025, doi: 10.1016/j.compbiomed.2025.110713.

[12] M. Saeidi, "Neural Decoding of EEG Signals with Machine Learning: A Systematic Review," *Brain Sci.*, 2021, doi: 10.3390/brainsci11111525.

[13] A. Prakash, "Electroencephalogram-based emotion recognition: a comparative analysis of supervised machine learning algorithms," *Data Sci. Manag.*, 2025, doi:

10.1016/j.dsm.2024.12.004.

[14] H. Roubhi, "Mutual Information-based Feature Selection Strategy for Speech Emotion Recognition using Machine Learning Algorithms Combined with the Voting Rules Method," *ETASR*, 2025, doi: 10.48084/etasr.9066.

[15] Y. Xu, "Emotional recognition of EEG signals utilizing residual structure fusion in bi-directional LSTM," *Complex Intell. Syst.*, 2024, doi: 10.1007/s40747-024-01682-y.

[16] G. S. Kumar, "Classification of Human Emotional States Based on Valence-Arousal Scale using Electroencephalogram," *J. Med. Signals Sensors*, 2023, doi: 10.4103/jmss.jmss_169_21.

[17] J. Barrowclough, "Personalised Affective Classification Through Enhanced EEG Signal Analysis," *Appl. Artif. Intell.*, 2025, doi: 10.1080/08839514.2025.2450568.

[18] K. Slama, "Comprehensive review of machine learning and deep learning techniques for epileptic seizure detection and prediction based on neuroimaging modalities," *VCIBA*, 2025, doi: 10.1186/s42492-025-00208-8.

[19] N. Bhadra, "Multiclass classification of environmental chemical stimuli from unbalanced plant electrophysiological data," *PLoS One*, 2023, doi: 10.1371/journal.pone.0285321.

[20] M. H. Kabir, "Exploring Feature Selection and Classification Techniques to Improve the Performance of an Electroencephalography-Based Motor Imagery Brain–Computer Interface System," *Sensors*, 2024, doi: 10.3390/s24154989.

[21] R. Vempati and L. D. Sharma, "EEG rhythm based emotion recognition using multivariate decomposition and ensemble machine learning classifier," *J. Neurosci. Methods*, vol. 393, p. 109879, 2023, doi: https://doi.org/10.1016/j.jneumeth.2023.109879.

[22] J. Zhang, Z. Yin, and P. Chen, "EEG-based Affect Classification with Machine Learning Algorithms," *IFAC-PapersOnLine*, vol. 56, no. 2, pp. 11627–11632, 2023, doi: 10.1016/j.ifacol.2023.10.486.

[23] S. Abdumalikov, J. Kim, and Y. Yoon, "Performance Analysis and Improvement of Machine Learning with Various Feature Selection Methods for EEG-Based Emotion Classification," *Appl. Sci.*, vol. 14, no. 22, Nov. 2024, doi: 10.3390/app142210511.

[24] H. A. Hamzah and K. K. Abdalla, "EEG-based emotion recognition systems; comprehensive study," *Heliyon*, vol. 10, no. 10, p. e31485, 2024, doi: 10.1016/j.heliyon.2024.e31485.

[25] M. A. Blanco-Ríos *et al.*, "Real-time EEG-based emotion recognition for neurohumanities: perspectives from principal component analysis and tree-based algorithms," *Front. Hum. Neurosci.*, vol. 18, 2024, doi: 10.3389/fnhum.2024.1319574.

[26] L. I. Wang, J. Go, and X. Chen, "AN EEG-BASED EMOTION RECOGNITION MODEL USING AN INTERACTION DESIGN FRAMEWORK and DEEP LEARNING," *J. Mech. Med. Biol.*, vol. 24, no. 2, pp. 1–16, 2024, doi: 10.1142/S0219519424400220.

[27] B. Mouazen, A. Benali, N. T. Chebchoub, E. H. Abdelwahed, and G. De Marco, "Enhancing EEG-Based Emotion Detection with Hybrid Models: Insights from DEAP Dataset Applications," *Sensors*, vol. 25, no. 6, pp. 1–22, 2025, doi: 10.3390/s25061827.

[28] D. Gao, "A multimodal functional structure-based graph neural network for fatigue detection," *Brain Res. Bull.*, 2025, doi: 10.1016/j.brainresbull.2025.111420.

[29] A. Mellot, "Harmonizing and aligning M/EEG datasets with covariance-based techniques to enhance predictive regression modeling," *Imaging Neurosci.*, 2023, doi: 10.1162/imag_a_00040.

[30] F. Yu, "Multidimensional structural-functional coupling uncovers network dysregulation and predicts binge-eating severity in bulimia nervosa," *BMC Med.*, 2025, doi: 10.1186/s12916-025-04556-3.

[31] I. Stancin, "A Review of EEG Signal Features and Their Application in Driver Drowsiness Detection Systems," *Sensors*, 2021, doi: 10.3390/s21113786.

[32] M. Blanco-Ruiz, "Emotion Elicitation Under Audiovisual Stimuli Reception: Should Artificial Intelligence Consider the Gender Perspective?," *IJERPH*, 2020, doi: 10.3390/ijerph17228534.

[33] Y. Mashiki, "Affective dimensions of emotional memory shape corticospinal excitability in the upper and lower limbs in young males," *Neuroscience*, 2025, doi: 10.1016/j.neuroscience.2025.10.034.

[34] D. Mikhaylov, "Comparison of EEG Signal Spectral Characteristics Obtained with Consumer- and Research-Grade Devices," *Sensors*, 2024, doi: 10.3390/s24248108.

[35] A. Tukhtaev, "Feature Selection and Machine Learning Approaches for Detecting Sarcopenia Through Predictive Modeling," *Mathematics*, 2024, doi: 10.3390/math13010098.

[36] K. Bhadra, "Learning to operate an imagined speech Brain-Computer Interface involves the spatial and frequency tuning of neural activity," *Commun. Biol.*, 2025, doi: 10.1038/s42003-025-07464-7.

[37] S. M. Ceh, "Neurophysiological indicators of internal attention: An electroencephalography–eye-tracking coregistration study," *Brain Behav.*, 2020, doi: 10.1002/brb3.1790.

[38] A. Jabbar, "Spectral feature modeling with graph signal processing for brain connectivity in autism spectrum disorder," *Sci. Rep.*, 2025, doi: 10.1038/s41598-025-06489-6.

[39] B. Mdluli, "Signal Preprocessing, Decomposition and Feature Extraction Methods in EEG-Based BCIs," *Appl. Sci.*, 2025, doi: 10.3390/app152212075.

[40] M. H. T. Najaran, "A deep learning feature mapping algorithm for emotion detection via facial and audio signals," *Appl. Soft Comput.*, 2026, doi: 10.1016/j.asoc.2026.114998.

[41] M. Habibollahi, "How PCA helps multi-criteria decision making for feature selection: A feature fusion approach in bioinformatics and gene expression data," *Alexandria Eng. J.*, 2025, doi: 10.1016/j.aej.2025.09.028.

[42] I. K. Ihianle, "Minimising redundancy, maximising relevance: HRV feature selection for stress classification," *Expert Syst. Appl.*, 2024, doi: 10.1016/j.eswa.2023.122490.

[43] M. Bilucaglia, "Applying machine learning EEG signal classification to emotion-related brain anticipatory activity," *F1000Research*, 2021, doi: 10.12688/f1000research.22202.3.

[44] J. B. Nasejje, "Statistical approaches to identifying significant differences in predictive performance between machine learning and classical statistical models for survival data," *PLoS One*, 2022, doi: 10.1371/journal.pone.0279435.

[45] M. Malekipirbazari, "Performance comparison of feature selection and extraction methods with random instance selection," *Expert Syst. Appl.*, 2021, doi: 10.1016/j.eswa.2021.115072.

[46] Z. Wang, "Emotion recognition based on multimodal physiological electrical signals," *Front. Neurosci.*, 2025, doi: 10.3389/fnins.2025.1512799.

[47] F. Albasu, "Electroretinogram Analysis Using a Short-Time Fourier Transform and Machine Learning Techniques," *Bioengineering*, 2024, doi: 10.3390/bioengineering11090866.

[48] N. S. Suhaimi, "EEG-Based Emotion Recognition: A State-of-the-Art Review of Current Trends and Opportunities," *Comput. Intell. Neurosci.*, 2020, doi:

10.1155/2020/8875426.

[49] M. Henry, "An ultra-precise Fast Fourier Transform," *Measurement*, 2023, doi: 10.1016/j.measurement.2023.113372.

[50] L. Gong, M. Li, T. Zhang, and W. Chen, "EEG emotion recognition using attention-based convolutional transformer neural network," *Biomed. Signal Process. Control.*, vol. 84, p. 104835, 2023, doi: 10.1016/j.bspc.2023.104835.

ISSN 2085-4579

# Enterprise Architecture Design Using TOGAF ADM for Digital Transformation at PT Profito Inovasi Kreatif

Allegra Aretha Putri[1], Nadine Aurelia[1], Vera Veronika[1*], Fransiska Eka Putri Wiriady[1], Afifah Trista Ayunda[1]

[1]Business Information System Program, Faculty of Science and Technology, Pradita University, Indonesia
allegra.aretha@student.pradita.ac.id, nadine.aurelia@student.pradita.ac.id, vera.veronika@student.pradita.ac.id, fransiska.eka@student.pradita.ac.id, afifah.trista@pradita.ac.id

Accepted on April 30th, 2026
Approved on June 1st, 2026

**Abstract— Rapid advancements in information technology are driving organizations to improve operational integration and business performance through enterprise systems and process automation. PT Profito Inovasi Kreatif is a B2B distribution company that utilizes ERP, CRM, and cloud-based systems to support operational activities and distribution management. However, the company faces several operational challenges, including declining ERP performance due to increasing transaction volumes, the complexity of managing more than 6,000 product SKUs, manual distribution planning and delivery coordination, and limited integration among operational systems. These challenges affect operational scalability, inventory coordination, and real-time decision-making. This study aims to design an enterprise architecture using the TOGAF ADM framework to support digital transformation and operational optimization. A qualitative case study approach was employed through semi-structured interviews, operational data analysis, and system documentation reviews. The analysis covered business, data, application, and technology architectures using as-is and to-be models to identify operational gaps and integration requirements. The proposed enterprise architecture includes ERP Develop Internal, Transportation Management System (TMS) integration for automated routing, API-based system integration, and centralized dashboard monitoring to improve operational coordination. The resulting architecture provides a structured blueprint for enhancing system integration, operational scalability, and data-driven decision-making, thereby supporting the company's digital transformation and long-term business growth.**

***Index Terms*— Enterprise Architecture; TOGAF ADM; Digital Transformation; ERP; System Integration; Transportation Management System (TMS)**

## I. INTRODUCTION

Digital transformation is an essential strategic option for organizations that are looking to improve operational efficiency, business agility, and the competitive advantage of their firms in dynamic business environments [1]. Enterprise architecture (EA) facilitates alignment between business processes and information technology by integrating the management of systems, data, applications, and technology infrastructure [2]. With proper Enterprise Architecture planning, organizations can integrate systems, streamline operations, and support sustainable digital transformation initiatives [3].

PT Profito Inovasi Kreatif is a growing B2B distribution company specializing in industrial materials, interior products, and furniture. The company has implemented several technologies, including Accurate ERP, CRM, AWS cloud services, dashboard technology and GPS device-based delivery tracking. Nonetheless, operational issues remain in manual distribution planning, performance degradation of ERP with increasing transaction volumes, and managing more than 6,000 product types (SKUs). Such scenarios reduce productivity and create challenges in managing scalable, cohesive business processes.

In this study digital transformation refers to the digitalization of distribution business processes through ERP integration, SKU data management, and TMS-based routing automation to enhance operational efficiency. TOGAF ADM was chosen because it offers a systematic and adaptable method for organizing

business processes, information systems, data, and technology infrastructure within complex organizational settings [4]. By using TOGAF ADM, organizations can create an enterprise architecture framework that facilitates business process integration and digital transformation efforts [5].

This research aims to design an Enterprise Architecture using the TOGAF ADM Framework for PT Profito Inovasi Kreatif to support digital transformation and improve operational efficiency in distribution operations.

## II. LITERATURE STUDY

### A. *THEORETICAL FRAMEWORK*

Enterprise Architecture (EA) is a structured approach to linking business processes and information systems to an organization's overall objectives. EA helps improve organizational efficiency and effectiveness by providing a high-level overview of the relationships among business strategy, enterprise-level information systems, and IT infrastructure [6]. In addition, EA eliminates redundancy, integrates information systems, and assists strategic decision-making [7]. EA also enables organizations to conduct a gap analysis from the "as-is" to the "to-be" and to design better systems [8]. TOGAF (The Open Group Architecture Framework) is one of the most popular frameworks for Enterprise Architecture design. The Architecture Development Method (ADM) within TOGAF is designed to help businesses integrate business demands and information technology across the organization. The method consists of several interconnected phases that build on each other to ensure the architecture meets the organization's needs and supports business processes. TOGAF ADM also supports sustainable enterprise architecture development and improves efficiency in system development and change management [6].

### B. *Previous Research*

TOGAF ADM has been widely applied in government services [9], healthcare [10], and retail and industrial sectors [11], [12] to support system integration and business process optimization. In the context of digital transformation, Enterprise Architecture (EA) plays a crucial role in aligning business processes, information systems, and cross-functional data integration through ERP modernization and integrated system design [13], [14], [15]. In addition, EA supports operational scalability and interoperability of information systems to improve efficiency and data-driven decision making [16], [17]. However, there is still a research gap in Enterprise Architecture design for B2B distribution companies with high transaction volumes and SKU complexity. Limited studies address specific issues such as ERP performance degradation caused by large transactions and high data volumes, managing more than 6,000 SKUs, and integration between ERP and Transportation Management Systems (TMS) to support automated distribution routing. Therefore, this study aims to design an enterprise architecture using TOGAF ADM at PT Profito Inovasi Kreatif to support digital transformation in distribution operations through ERP integration, operational scalability, and process automation.

## III. METHODOLOGIES

### A. *Research Type and Approach*

The purpose of this research is to conduct a descriptive case study with a qualitative approach to design and analyze information systems and business processes at PT Profito Inovasi Kreatif, a B2B distribution company [18]. Interviews, observations, and analyses of operational documents were conducted to describe the current state of the company's systems and business processes. This study uses the TOGAF ADM framework to model an integrated enterprise architecture that enables digital transformation through system integration and operational scalability [8].

### B. *Data Collection Sources and Techniques*

To determine the business process flows, operational challenges, and information systems used in the company, data were collected through semi-structured interviews with the Director of PT Profito Inovasi Kreatif [19]. The interview lasted approximately 60 minutes and focused on ERP performance, inventory management, distribution processes, system integration challenges, and digital transformation initiatives. The Director was selected as the key informant due to direct involvement in strategic and operational decision-making across business functions. Semi-structured interviews were adopted because they allow researchers to explore participants' perspectives while maintaining alignment with the research objectives [19].

In addition to interviews, secondary operational data were collected to support and validate the findings. Transaction records were reviewed to verify the high

volume of business transactions and the ERP performance challenges reported during the interview. Product and SKU data were analyzed to validate the complexity of inventory management, particularly the management of more than 6,000 product SKUs. System documentation was examined to confirm the existing information systems, application landscape, and operational workflows, including the use of ERP, CRM, dashboard systems, and distribution monitoring systems. Data triangulation was conducted by comparing interview findings with these operational records and documents to ensure the consistency, credibility, and accuracy of the results used as the basis for the enterprise architecture design.

### *C. System Analysis Method*

System analysis in this research was conducted to evaluate the condition of information systems and business processes at PT Profito Inovasi Kreatif, including the management of more than 6,000 product SKUs with high operational complexity, as well as the use of the Accurate ERP system, CRM, and distribution systems. The analysis focused on operational challenges, ERP performance, process efficiency, and system integration. The results of the analysis were used to identify system requirements as the basis for designing enterprise architecture using the TOGAF ADM framework [8].

### *D. System Design Method*

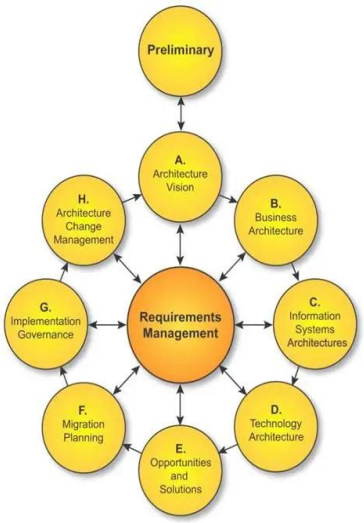


Fig. 1. TOGAF ADM Phases

Figure 1 shows the TOGAF ADM phases. This research implements an Enterprise Architecture (EA) approach using the TOGAF ADM framework to align business strategy and technology and produce an integrated architecture that supports PT Profito Inovasi Kreatif in its digital transformation [20]. The design process follows the TOGAF ADM cycle as outlined below:

1. Preliminary Phase and Architecture Vision. The purpose is to determine the boundaries and aims of architecture (ERP, CRM, website, distribution systems).

2. Business Architecture. Analyzing ordering, inventory, and distribution processes to identify existing problems.

3. Information Systems Architecture. Designing data and application management, including product data (>6,000 SKUs), customer data, and transaction data, as well as the integration of ERP, CRM, and distribution systems to support real-time data synchronization.

4. Technology Architecture. Determining the technology infrastructure, such as AWS and GPS-based monitoring systems.

5. Opportunities & Solutions. Developing technology solutions, such as ERP-integrated TMS, that assist in the automation of routing and delivery coordination.

6. Migration Planning. Planning the system implementation phases, priorities, and resource allocation.

7. Implementation Governance. Implementation of monitoring systems in accordance with plans, standards and digital transformation objectives.

8. Architecture Change Management. Handling changes to keep the system flexible in line with business needs.

Moreover, Requirements Management is performed to manage system requirements throughout the design process, enabling the production of an integrated architecture as a basis for system development. The requirements listed are aligned with the potential solutions via mapping to the operational problems and system requirements identified during the analysis step, as shown in Table 1.

TABLE 1. REQUIREMENTS TRACEABILITY

| Problem Identified | Requirement | Proposed Solution |
|---|---|---|
| ERP performance decreases due to large SKU data | Scalable and integrated system | ERP optimization and integration |
| Manual distribution process | Automated route management | Transportation Management System (TMS) |
| Data is not fully integrated | Real-time system integration | API-based integration |
| Difficulty in monitoring operational data | Centralized monitoring system | Dashboard and reporting system |

## IV. RESULTS AND DISCUSSION

### A. *Preliminary Phase*

The analysis revealed that PT Profito Inovasi Kreatif has a relatively integrated system. However, several issues remain, particularly in ERP performance, SKU data management, distribution planning, inventory management, and ERP-distribution system integration. Therefore, this study focuses not only on system integration but also on improving operational efficiency, enabling automation, and optimizing distribution processes. These requirements are translated into the Architecture Principles presented in Table 2, which serve as the foundation for the proposed solutions in the Opportunities and Solutions section.

TABLE 2. ARCHITECTURE PRINCIPLES

| No | Architecture | Principle | Description |
|---|---|---|---|
| 1 | Business | Business Process Optimization | Business processes must be optimized to improve efficiency in distribution, inventory, and customer service. |
| | | Data-Driven Decision Making | Decisions must be based on accurate and real-time data. |
| 2 | Data | Data Quality | Data must be accurate, consistent, and relevant. |
| | | Data Integration | Data must be integrated across systems (e.g., ERP, CRM) to avoid redundancy. |
| | | Data Availability | Data must be available in real-time. |
| 3 | Application | Business Process Automation | Systems must automate key business processes. |
| | | System Scalability | Applications must support increased data and transaction volumes. |
| 4 | Technology | System Performance | Systems must be fast and stable. |
| | | Intelligent System | Technology must support analytics and optimization. |
| | | Infrastructure Security | Systems must be secure from threats. |

### B. *Architecture Vision*

The enterprise architecture design for PT Profito Inovasi Kreatif was developed in this phase to support the company in its digital transformation, enhancing operational efficiency and business competitiveness. Based on interviews with company representatives, the company's goal is to be the leading ecosystem solution provider through technology-driven innovation in the interior and furniture industry, with a targeted annual growth of 25%. Currently, the company uses several information systems, including Accurate ERP and an internal CRM solution, as well as various dashboards and AWS infrastructure. Nonetheless, a few concerns were raised such as declining ERP performance caused by increasing data volume, manual data analysis processes, and inefficient distribution processes. Hence, various strategic initiatives were suggested, including internal ERP development to improve scalability and performance, a Transportation Management System (TMS) for automated route optimization, enhancements to company websites to support online transactions, and integration of data across systems to enable real-time business analytics.

### C. *Business Architecture*

The Business Architecture phase analyzes the core business processes of PT Profito Inovasi Kreatif by identifying the current (as-is) processes and proposing improvements for the future (to-be) processes, as shown in Fig. 2.

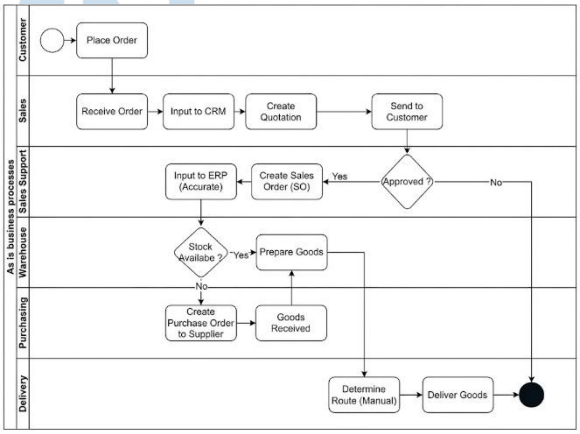


Fig. 2. Current (as-is) business processes.

The proposed business process (to-be) is presented, as shown in Fig. 3.

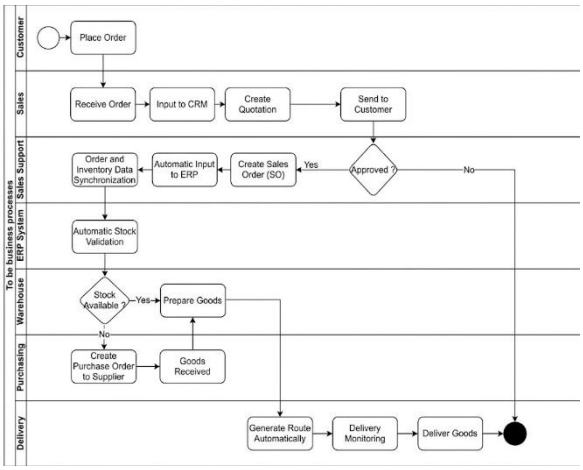


Fig. 3. Proposed (to-be) business processes

The proposed business process was designed based on operational requirements related to ERP performance, manual distribution processes, and the complexity of SKU management. Currently, stock validation and delivery route planning are performed manually, leading to delays and potential human errors. The future process integrates ERP and TMS to support automated stock validation, route generation, and real-time distribution coordination. In addition, data synchronization enables real-time processing and improves operational coordination. These improvements are expected to reduce errors, improve operational processes, and support data-driven decision-making.

## *D. Information Systems Architecture*

### *1. Data Architecture*

The Data Architecture phase aims to define the key data entities required to support the business processes and proposed system, as shown in Fig. 4.

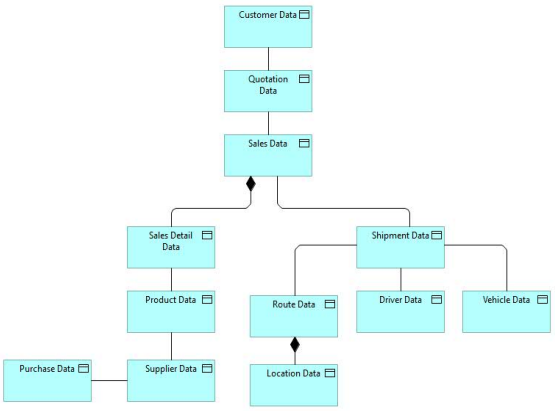


Fig. 4. Data Architecture using ArchiMate

The Data Architecture phase identifies key data objects designed to support the proposed enterprise architecture, as shown in Fig. 4. These data objects are derived from the company's sales, inventory, purchasing, and distribution processes. Product and sales data support SKU and transaction management, while shipment, route, driver, vehicle, and location data support distribution planning and automated routing. The proposed architecture supports ERP, CRM, and TMS integration to reduce manual processes and fragmented data processing. ArchiMate relationships are used to represent data dependencies within the distribution process.

### *2. Application Architecture*

The Application Architecture integrates various applications that support the business processes of PT Profito Inovasi Kreatif, including ERP, CRM, dashboard systems, websites, and the Transportation Management System (TMS). Application architecture supports information system integration within the organization [8]. In addition, TOGAF ADM supports the alignment between business needs, application integration, and digital transformation initiatives [21]. In this study, digital transformation focuses on improving distribution operations through ERP integration, TMS-based routing automation, and real-time operational monitoring. The analysis compares the current condition (as-is) with the proposed condition (to-be), as presented in Table 3.

TABLE 3. COMPARISON OF APPLICATION ARCHITECTURE (AS-IS AND TO-BE)

| Aspect | As-Is | To-Be |
|---|---|---|
| ERP | Using Accurate with declining performance due to high data volume | Development of an internal ERP to improve performance and scalability |
| CRM | Internal CRM for customer and quotation management | Further integration with ERP to improve efficiency |
| Website | Online ordering with real-time stock updates | Development of a new website to improve service quality |
| Dashboard | Used for operational data access, analysis is still manual | System enhancement for more integrated data analysis |
| Distribution | Manual route planning | TMS implementation for automated route |

| | | |
|---|---|---|
| | | planning and analysis |

As presented in Table 3, the proposed Application Architecture is developed based on the company's distribution business processes, focusing on improving system integration, operational efficiency, real-time data synchronization, and automated distribution coordination. This includes the integration of ERP, CRM, dashboard, and TMS to support the digital transformation of distribution operations.

### *E. Technology Architecture*

The Technology Architecture aims to examine the infrastructure that supports the operation of information systems at PT Profito Inovasi Kreatif. Table 4 shows the current condition (as-is) and a proposed condition (to-be), on which this analysis is based.

TABLE 4. COMPARISON OF TECHNOLOGY ARCHITECTURE (AS-IS AND TO-BE)

| Technology | As-Is | To-Be |
|---|---|---|
| Cloud Computing (AWS) | Server and database infrastructure | Scalable cloud infrastructure |
| Database System | Operational data storage | Improved data management and analysis |
| API Integration | ERP and website integration | Real-time system integration |
| GPS Tracking System | Vehicle tracking | TMS integration |
| Internet and WiFi Network | Network connectivity | Improved network stability |
| Google Workspace | Cloud-based collaboration platform | Integrated collaboration system |

As presented in Table 4, the existing technology infrastructure has supported operational activities. However, challenges remain in system performance, data processing, and real-time system integration. Therefore, the proposed Technology Architecture focuses on optimizing cloud infrastructure, improving database management, and implementing API-based integration to support real-time integration between the internal ERP system and the Transportation Management System (TMS).

### *F. Opportunities and Solutions*

During the Opportunities and Solutions phase, several initiatives are proposed, including the implementation of a Transportation Management System (TMS) for automated routing, the development of an internal Enterprise Resource Planning (ERP) system to improve scalability and system performance, and API-based integration to support cross-functional data exchange. These initiatives aim to enhance operational efficiency, reduce manual data processing, and support the digital transformation of distribution operations. The development of an internal ERP was selected because the company requires greater flexibility to manage more than 6,000 product SKUs, increasing transaction volumes and future integration requirements. Although upgrading the existing ERP could improve performance, it would provide limited customization and scalability for long-term business needs, particularly in supporting the integration of ERP, CRM, TMS, and other operational systems. Therefore, an in-house ERP solution was considered more suitable to accommodate the company's specific operational requirements and long-term digital transformation strategy.

Moreover, the proposed architecture was reviewed by a lecturer with expertise in Enterprise Architecture and TOGAF ADM. The validation process focused on assessing the alignment between business requirements and the proposed architecture, the consistency of architectural components across TOGAF ADM phases, the feasibility of implementation within the company's operational environment, and the ability of the architecture to support future scalability and system integration. Feedback and recommendations obtained during the review process were incorporated into the final architecture design. Based on the evaluation, the proposed architecture was considered appropriate for supporting the company's operational requirements and digital transformation objectives.

### *G. Migration Planning*

During the system implementation planning phase, as presented in Table 5, each step is planned

based on priority, aligned with the company's business and operational needs.

TABLE 5. MIGRATION PLANNING ROADMAP

| Phase | Implementation Focus | Objective | Priority |
|---|---|---|---|
| Phase 1 | Internal ERP Development | Improve ERP system scalability and performance | High |
| Phase 2 | TMS Implementation | Support automated route planning and distribution coordination | High |
| Phase 3 | System Data Integration | Support real-time data synchronization across systems | Medium |
| Phase 4 | Website Enhancement | Improve customer service and support online transactions | Medium |
| Phase 5 | Dashboard Optimization | Support more integrated and real-time business analytics | Low |

Table 5 shows that priorities are established based on the urgency of the items, the impact on business processes, and dependencies between systems. The creation of an in-house ERP and a Transportation Management System (TMS) are considered high priorities because they are directly related to ERP performance issues and manual distribution processes. Furthermore, system data integration is carried out to support real-time data synchronization, while website enhancement and dashboard optimization are focused on improving service quality and data analysis.

### *H. Implementation Governance*

In this phase, architecture implementation is carried out in a controlled manner to ensure alignment between the proposed architecture and the company's business objectives. Using the TOGAF ADM approach, the implementation process is managed in a structured manner to support business needs and information technology development. The implementation includes the development of an internal ERP system, Transportation Management System (TMS), website enhancement, and system data integration. The governance process involves management evaluation, system testing, quality assurance, and periodic reviews to ensure that the implemented architecture remains aligned with operational needs and business objectives.

## V. CONCLUSION

This research generates an enterprise architecture design based on TOGAF ADM framework to accommodate the digital transformation needs of PT Profito Inovasi Kreatif through the alignment between business processes and information technology. Through interviews and analysis of the current system, several key problems were identified, including declining ERP performance, manual data processing, and inefficient distribution processes.

The architecture is developed and includes the development of an internal ERP system to support scalability and performance, the implementation of a Transportation Management System (TMS) for automated route optimization, and improved system integration to support real-time data processing. The implementation of this architecture is expected to improve operational efficiency, reduce manual processing errors, and support data-driven decision-making.

Overall, the proposed enterprise architecture provides a structured and integrated blueprint as a reference for system development and digital transformation initiatives at PT Profito Inovasi Kreatif.

## REFERENCES

[1] M. Yusuf, B. Surya, F. Menne, M. Ruslan, S. Suriani, and I. Iskandar, "Business Agility and Competitive Advantage of SMEs in Makassar City, Indonesia," Sustainability, vol. 15, no. 1, p. 627, 2023, doi: 10.3390/su15010627.

[2] G. Fuentes-Quijada, F. Ruiz-González, and A. Caro, "Enterprise Architecture and IT Governance to Support the BizDevOps Approach: a Systematic Mapping Study," *Information Systems Frontiers*, vol. 27, pp. 865–888, 2025, doi: 10.1007/s10796-024-10473-2.

[3] E. Niemi and S. Pekkola, "The Benefits of Enterprise Architecture in Organizational Transformation," Business & Information Systems Engineering, vol. 62, no. 6, pp. 585–597, 2020, doi: 10.1007/s12599-019-00605-3.

[4] D. C. Ruiz, G. V. Araujo, and J. D. Alarcon, "Enterprise Architecture to Optimize the Sales Process Using the TOGAF ADM Cycle in Companies in the Retail Sector," in Proceedings of the 20th International Conference on Web Information Systems and Technologies (WEBIST 2024), 2025, pp. 218–225, doi: 10.5220/0012928500003825.

[5] R. S. B. Cokro and G. Wang, "Improving Architecture of Legal System by using TOGAF-ADM," International Journal of Science and Applied Information Technology, vol. 12, no. 5, pp. 58–62, 2023, doi: 10.30534/ijsait/2023/041252023.

[6] D. N. A. Sista, I. M. Candiasa, and I. G. A. Gunadi, "Enterprise Architecture Design of Information Systems Using TOGAF ADM at SMA Negeri 1 Singaraja," *JST (Jurnal Sains dan Teknologi)*, vol. 10, no. 2, pp. 316–328, 2021. doi: 10.23887/jstundiksha.v10i2.37137.

[7] T. Rohman, Hermanto, S. Assani, and A. Hendi, "Enterprise Architecture Design Using TOGAF ADM at Universitas Qomaruddin," *Jurnal Tekno Kompak*, vol. 19, no. 2, pp. 91–103, 2025. doi: 10.33365/jtk.v19i2.51.

[8] N. I. Mohd Rahimi, S. Mat Yatya, and N. A. Abu Bakar, "Enterprise Architecture: Enabling Digital Transformation for Healthcare Organization," Open International Journal of Informatics, vol. 11, no. 1, pp. 67–73, 2023, doi: 10.11113/oiji2023.11n1.246.

[9] M. I. Alhari and A. A. N. Fajrilah, "Enterprise Architecture: A Strategy to Achieve e-Government Dimension of Smart Village Using TOGAF ADM 9.2," International Journal on Informatics Visualization, vol. 6, no. 2-2, pp. 540–545, 2022, doi: 10.30630/joiv.6.2-2.1147.

[10] H. Al Omari, A. Alkhateeb, and B. Hammo, "Applying TOGAF-Based Enterprise Architecture in the Healthcare Sector: A Case Study of the National Center for Diabetes in Jordan," Jordanian Journal of Computers and Information Technology (JJCIT), vol. 10, no. 2, pp. 182–197, 2024, doi: 10.5455/jjcit.71-1705704023.

[11] M. I. Fianty, "Designing an Enterprise Architecture Using TOGAF ADM Framework (Case Study: PT Sumber Alfaria Trijaya)," G-Tech Journal, vol. 7, no. 2, pp. 601–610, 2023, doi: 10.33379/gtech.v7i2.2409.

[12] D. Y. Prawira et al., "Enterprise Architecture Design Using TOGAF ADM: The Case of KotaKita," Jurnal Teknologi Sistem Informasi, vol. 6, no. 2, pp. 214–228, 2023, doi: 10.32493/jtsi.v6i2.29416.

[13] M. van de Wetering, "The Role of Enterprise Architecture for Digital Transformations," Sustainability, vol. 13, no. 4, p. 2237, 2021, doi: 10.3390/su13042237.

[14] M. Jusuf and S. Kurniawan, "Enterprise Architecture Model for Integrated Data Management in Industrial Companies," International Journal of Information Systems and Informatics, vol. 5, no. 1, pp. 44–56, 2023, doi: 10.51519/ijisinfo.v5i1.412.

[15] S. Kumar and R. Singh, "Modernizing Legacy ERP Systems Using TOGAF ADM: A Roadmap for Digital Excellence," Journal of Enterprise Information Management, early access, 2024, doi: 10.1108/JEIM-01-2024-0031.

[16] A. R. S. Sembiring et al., "Design of Enterprise Architecture for ERP System Implementation Using TOGAF ADM Framework," IOP Conference Series: Materials Science and Engineering, vol. 1003, no. 1, 2021, doi: 10.1088/1757-899X/1003/1/012104.

[17] J. Zhou, "API-Driven Enterprise Architecture for Digital Transformation: Ensuring System Interoperability," International Journal of Information Management, vol. 65, p. 102497, 2023, doi: 10.1016/j.ijinfomgt.2022.102497.

[18] S. Susilawati, I. Maita, F. N. Salisah, and M. Fronita, "Perancangan Enterprise Architecture Menggunakan Framework TOGAF ADM pada Dinas Koperasi dan UKM Kota Pekanbaru," Jurnal Sistem Informasi dan Informatika (SIMIKA), vol. 8, no. 1, pp. 111–126, 2025, doi: 10.47080/simika.v8i1.3731.

[19] R. Ruslin, S. Mashuri, M. S. A. Rasak, F. Alhabsyi, and H. Syam, "Semi-structured Interview: A Methodological Reflection on the Development of a Qualitative Research Instrument in Educational Studies," IOSR Journal of Research & Method in Education (IOSR-JRME), vol. 12, no. 1, pp. 22–29, 2022, doi: 10.9790/7388-1201052229.

[20] R. Riztriano, "Enterprise Architecture Design Using the TOGAF Architecture Development Method (TOGAF-ADM) Framework at The Library of SMAN 22 Jakarta," Jurnal Komputer, Informasi dan Teknologi, vol. 5, no. 2, pp. 1–26, 2025, doi: 10.53697/jkomitek.v5i2.3327.

[21] L. A. D. Sari, R. Mulyana, and I. Y. Mukti, "A TOGAF 10-Based Enterprise Architecture Framework for Digital Transformation in SME Banks," Jurnal Teknik Informatika (JUTIF), vol. 6, no. 2, pp. 673–690, 2025, doi: 10.52436/1.jutif.2025.6.2.4329.

ISSN 2085-4579

# Depression Risk Classification Based on Self-Reported Psychosocial and Behavioral Survey Data Using Machine Learning

Marcelinus Jonathan Salim[1], Tegar Anugrah Firdaus[2], Carens Chanda Claudhyta Hasan[3], Yuri Pamungkas[4*]

[1-4]Department of Medical Technology, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
yuri@its.ac.id

Accepted on March 27th, 2026
Approved on June 29th, 2026

***Abstract*—Mental health disorders, particularly depression, remain a major public health concern, while early risk identification outside clinical settings continues to pose significant challenges. The increasing availability of survey-based mental health data provides opportunities for data-driven risk screening, yet effective modeling strategies and evaluation practices remain underexplored. This study presents a comparative evaluation of multiple machine learning algorithms for depression risk classification using a publicly available mental health survey dataset. Rather than predicting clinical depression, the target variable is formulated as a risk proxy derived from social weakness indicators to support screening-oriented analysis. A quantitative experimental framework is employed to compare Logistic Regression, Random Forest, Support Vector Machine, and Extreme Gradient Boosting under consistent preprocessing and data partitioning conditions. Model performance is evaluated using complementary metrics, including accuracy, recall for High-risk cases, and the area under the receiver operating characteristic curve (ROC-AUC). Threshold optimization based on ROC analysis is applied to align model outputs with screening objectives that prioritize sensitivity. The results demonstrate that Logistic Regression and Support Vector Machine consistently achieve superior or comparable performance across all evaluation dimensions, including high overall accuracy, near-perfect sensitivity for High-risk detection, and strong discriminative capability. In contrast, more complex ensemble and distance-based models show mixed outcomes, indicating diminishing performance gains from increased algorithmic complexity. These findings highlight that simple and interpretable models can effectively support depression risk screening using survey-based data, offering a practical balance between predictive performance, transparency, and computational efficiency.**

***Index Terms*—Depression Risk Classification; Machine Learning; Mental Health Screening; Risk Proxy; Survey-based Data**

## I. INTRODUCTION

Mental health disorders, particularly depression and closely related conditions, remain a significant global health challenge due to their persistent effects on emotional well-being, social functioning, and daily productivity [1],[2],[3]. Depression is associated not only with clinical impairment but also with broader social and economic consequences that often extend beyond the healthcare system [4],[5],[6]. Despite increased awareness, a substantial proportion of individuals experiencing depressive symptoms remain undetected, especially outside formal clinical environments [7],[8]. In recent years, large-scale survey-based mental health datasets have become increasingly available, enabling population-level analysis of psychological well-being and related risk factors [9],[10]. Such datasets provide an opportunity to explore early indicators of depression risk using self-reported behavioral and psychosocial attributes, particularly in contexts where clinical data are limited or unavailable [11],[12]. However, survey-based data are inherently heterogeneous and subjective, requiring analytical approaches capable of handling noisy and high-dimensional feature spaces [13].

Machine learning techniques have been widely adopted to address these challenges due to their ability to model complex, non-linear relationships among multiple variables [14],[15]. Compared to conventional statistical approaches, machine learning models can accommodate categorical attributes, capture interaction effects, and scale efficiently to larger datasets [16],[17]. Nevertheless, these models are not intended to replace clinical diagnosis, but rather to support exploratory analysis and early-stage risk screening within data-driven mental health research [18],[19]. Although numerous studies have applied machine learning to depression-related datasets, many focus on a single algorithm or rely primarily on accuracy as the sole evaluation metric. In imbalanced or perception-based datasets, such practices may obscure important performance trade-offs, as misclassification of high-risk individuals carries greater implications than overall correctness [20],[21]. Therefore, a more comprehensive evaluation framework that incorporates sensitivity-oriented and discrimination-based metrics is required to

properly assess model behavior in mental health applications.

Different machine learning algorithms exhibit distinct learning mechanisms and inductive biases, which may lead to varying predictive performance when applied to the same dataset [9],[10]. Linear models such as Logistic Regression offer interpretability and stability, while kernel-based methods like Support Vector Machines are effective in high-dimensional feature spaces [22],[23],[24]. Ensemble approaches, including Random Forest and Extreme Gradient Boosting (XGBoost), aim to improve robustness and generalization by aggregating multiple weak learners [22],[23]. In this study, a classification framework is developed using a secondary mental health survey dataset obtained via Mendeley Data [25]. Rather than predicting clinical depression, the target variable is formulated as a risk proxy derived from responses related to social weakness, which has been associated with psychosocial vulnerability and elevated depressive tendencies in population-based studies [17]. This proxy-based formulation is adopted to support methodological comparison and early risk screening, rather than clinical or diagnostic decision-making [18].

Model performance is evaluated using multiple complementary metrics to capture different aspects of predictive behavior [19]. Overall accuracy is reported to reflect general classification correctness, while recall is emphasized to assess sensitivity in identifying potential risk cases, particularly those belonging to the High-risk group. In addition, the area under the receiver operating characteristic curve (ROC-AUC) is employed to evaluate discriminative capability across varying decision thresholds [20]. To further align model outputs with screening-oriented objectives, threshold optimization based on ROC analysis is applied, prioritizing the reduction of missed High-risk cases [21]. The primary objective of this study is to conduct a systematic performance comparison of Logistic Regression, Random Forest, Support Vector Machine, and XGBoost under a consistent experimental framework [9]. By providing a transparent and reproducible comparative analysis, this study aims to highlight performance trade-offs among widely used machine learning algorithms and to contribute methodological insights for future research in mental health risk modeling [10].

## II. Material & Methods

This methods section describes the study workflow for depression risk classification using machine learning, focusing on how the dataset was defined and analyzed within the chosen study design, how data were cleaned and transformed into model-ready representations, and how the train–test split strategy was applied to ensure fair and reproducible comparisons. It then details the machine learning models evaluated in this performance study, including their training procedures and key configurations. Finally, the evaluation metrics are specified to quantify and compare predictive performance in a consistent manner across all models.

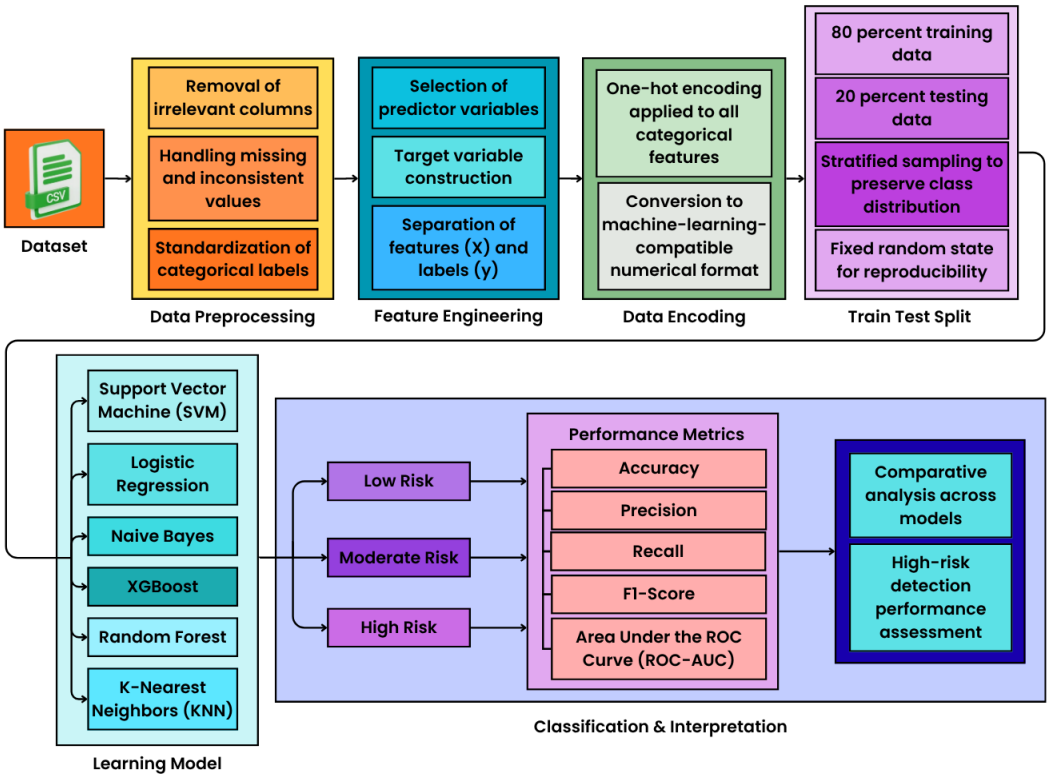


Fig. 1. Methodology of the research.

### A. Dataset & Study Design

This study adopts a quantitative experimental design using a publicly available secondary dataset, originally titled "The RHMCD-20 Datasets for Depression and Mental Health Data Analysis with Machine Learning," published by Salehin and Amin via Mendeley Data [25]. The dataset consists of 824 structured survey responses, comprising 12 categorical predictor variables and one target variable (Social_Weakness). The predictor variables represent demographic, psychosocial, behavioral, and lifestyle-related factors associated with mental health conditions, including Age, Gender, Occupation, Days Indoors, Growing Stress, Quarantine Frustrations, Changes in Habits, Mental Health History, Weight Change, Mood Swings, Coping Struggles, and Work Interest. The target variable (Social_Weakness) consists of three response categories (Yes, Maybe, and No), with a distribution of 259 (31.4%), 287 (34.8%), and 278 (33.7%) samples, respectively. Since all predictor variables are categorical, one-hot encoding was applied during preprocessing to convert categorical responses into machine-readable binary representations for model development. Rather than directly predicting clinical depression, this study formulates a three-class depression risk proxy for classification based on the target variable, as described in the subsequent preprocessing stage.

### B. Data Preprocessing and Representation

Prior to model training, the dataset undergoes a preprocessing stage to ensure compatibility with machine learning algorithms and consistency across experimental procedures. Because this study utilizes survey-based data and aims to support early risk screening rather than formal clinical diagnosis, the target variable is defined as a depression risk proxy rather than a definitive medical condition. This proxy was constructed directly from the respondents' answers to the Social Weakness attribute, which serves as the target variable in this study. The Social Weakness attribute was selected because it reflects respondent's perceived psychosocial vulnerability and provides an interpretable surrogate outcome for comparative machine learning analysis when clinical diagnostic labels are unavailable. The original response categories were mapped into three risk levels: Low-risk (No), Moderate-risk (Maybe), and High-risk (Yes). This proxy-based categorization enables comparative evaluation of machine learning models for early depression risk screening while avoiding interpretation as a clinical diagnosis.

Regarding the predictor variables, given the survey-based nature of the data, they are treated as categorical attributes reflecting respondents' discrete choices and self-reported conditions [26],[27]. These categorical features are transformed using one-hot encoding, a common and effective approach for representing nominal variables in machine learning models [28],[29],[30]. This technique converts each category into a binary indicator, allowing models to process categorical information without imposing artificial ordinal relationships that may bias learning outcomes [31],[32],[33]. Such a representation is particularly appropriate for survey-based datasets, where categorical responses do not imply inherent ordering. For instance, critical survey attributes such as Growing_Stress consist of responses ('Yes', 'No', 'Maybe'), Coping_Struggles consist of ('Yes', 'No'), and Work_Interest consist of ('Yes', 'No', 'Maybe'). After transforming all the categorical variables using one-hot encoding, the feature space expanded to a total of 39 independent features. Finally, all preprocessing steps are implemented within a unified processing pipeline to ensure identical transformations are applied to both training and testing data, thereby reducing the risk of data leakage and supporting reproducibility. No feature scaling or dimensionality reduction is applied, as the resulting binary-valued features are suitable for direct use in the selected classification algorithms.

### C. Train Test Split Strategy

The analytical workflow began with data ingestion and risk-proxy target formulation, followed by one-hot encoding of all categorical features. Regarding the target variable, the risk proxy formulation yielded a relatively balanced class distribution: Moderate-risk (287 samples), Low-risk (278 samples), and High-risk (259 samples), resulting in a calculated class imbalance ratio of 1.11. To evaluate model performance under consistent and reproducible conditions, the dataset is partitioned into training and testing subsets using an 80:20 split ratio. This proportion is commonly adopted in supervised learning studies to provide sufficient data for model training while retaining an independent subset for unbiased performance evaluation [34],[35],[36]. Given the presence of class imbalance in the target variable, stratified sampling is applied during the splitting process to preserve the original class distribution across both subsets. This strategy helps prevent biased evaluation results that may arise when minority classes are underrepresented in either the training or testing data [37],[38],[39]. A fixed random seed is used to ensure reproducibility, and the same data partition is maintained across all evaluated models. By applying an identical train–test split for each algorithm, the comparative analysis reflects differences in model behavior rather than variations in data allocation [39].

### D. Machine Learning Models

This study evaluates six supervised machine learning algorithms representing complementary classification paradigms. All models were trained and evaluated under identical preprocessing and data partitioning conditions to ensure a fair and consistent comparison. Logistic Regression (LR), Support Vector Machine (SVM), Random Forest (RF), Extreme Gradient Boosting (XGBoost), Naïve Bayes (NB), and K-Nearest Neighbors (KNN) were selected to represent linear, margin-based, ensemble, probabilistic, and distance-based learning approaches that are widely applied in structured tabular data analysis. This

selection encompasses linear and margin-based models (LR, SVM) known for stability in high dimensions, tree-based and boosting ensembles (Random Forest, XGBoost) recognized for handling complex interactions, as well as probabilistic and distance-based baselines (Naïve Bayes, KNN).

To ensure reproducibility, all experiments were implemented using Python 3.10 and the Scikit-learn library (version 1.2.2). A fixed random seed (random_state = 42) was applied consistently across all data partitioning and model initialization steps. Furthermore, to prevent data leakage and overfitting during hyperparameter tuning, a Stratified 5-Fold Cross-Validation approach was utilized on the training set prior to the final evaluation on the test set. The final hyperparameter settings for each machine learning algorithm were determined through the tuning process and subsequently used for the comparative evaluation. A summary of the selected hyperparameters is presented in Table I.

TABLE I. HYPERPARAMETER CONFIGURATION OF THE EVALUATED MACHINE LEARNING MODELS

| Algoritm | Hyperparameter Configuration |
|---|---|
| Logistic Regression | solver = liblinear, C = 1.0, class_weight = balanced |
| Support Vector Machine | kernel = RBF, C = 1.0, probability = True, class_weight = balanced |
| Random Forest | n_estimators = 100, max_depth = None, class_weight = balanced |
| XGBoost | learning_rate = 0.1, max_depth = 6, objective = multi-softmax |
| K-Nearest Neighbors | n_neighbors = 5, weights = uniform, metric = minkowski |
| Naïve Bayes | alpha = 1.0, binarize = 0.0, fit_prior = True |

### 1. *Logistic Regression (LR)*

Logistic Regression is employed as a linear baseline classifier due to its interpretability and suitability for binary risk modeling. The model estimates the probability of the positive class using a logistic function applied to a linear combination of input features [40],[41]:

$$P(y = 1 \mid x) = \frac{1}{1 + \exp\left(-(w^{\top}x + b)\right)} \quad (1)$$

where $x$ denotes the feature vector, $w$ is the weight vector, and $b$ is the bias term. To mitigate class imbalance, balanced class weights are applied during training.

### 2. *Support Vector Machine (SVM)*

Support Vector Machine is a margin-based classifier that seeks an optimal hyperplane maximizing the separation between classes [23],[42]. The optimization objective can be expressed as:

$$\min_{w,b} \frac{1}{2}\|w\|^2 \text{ subject to } y_i(w^{\top}x_i + b) \geq 1 \quad (2)$$

In this study, SVM is configured to produce probabilistic outputs to support ROC-based evaluation, and class imbalance is handled using class weighting.

### 3. *Random Forest*

Random Forest is an ensemble learning method that constructs multiple decision trees using bootstrap sampling and aggregates their predictions to improve robustness and generalization [43],[44]. The final prediction is obtained through majority voting:

$$\hat{y} = mode\{h_1(x), h_2(x), \ldots, h_T(x)\} \quad (3)$$

where $h_T(x)$ denotes the prediction of the ttt-th decision tree and TTT represents the total number of trees. Balanced class weighting is employed to address class imbalance.

### 4. *Extreme Gradient Boosting*

XGBoost is a gradient boosting ensemble method that sequentially builds weak learners to minimize a regularized objective function, making it effective for structured tabular data [43],[45]. The objective function is defined as:

$$\mathcal{L} = \sum_{i=1}^{n} \ell(y_i, \hat{y}_i) + \sum_{k}^{\Omega}(f_k) \quad (4)$$

with the regularization term:

$$\Omega(f) = \gamma^T + \frac{1}{2}\lambda||w||^2 \quad (5)$$

where $\ell(\cdot)$ denotes the loss function and $\Omega(\cdot)$ controls model complexity. Class imbalance is handled using the *scale_pos_weight* parameter derived from class proportions.

### 5. *Additional Baseline Models*

To provide a broader comparative context, K-Nearest Neighbors (KNN) and Bernoulli Naïve Bayes classifiers are included as supplementary baselines. These models are evaluated using the same preprocessing pipeline and train-test split to maintain comparability. Specifically, the K-Nearest Neighbors (KNN) algorithm was configured with k=5 neighbors (n_neighbors=5) using the standard Minkowski distance metric and uniform neighbor weighting. Meanwhile, the Bernoulli Naïve Bayes model was implemented with additive Laplace smoothing (α = 1.0) and binarization thresholds matching the one-hot encoded input feature space. The detailed parameter specifications for all evaluated models are summarized in Table I.

The overall experimental workflow consisted of four main stages: dataset preparation and risk proxy construction, categorical feature encoding through one-hot encoding, stratified train-test splitting, and comparative evaluation of six supervised machine learning algorithms. All models were trained under

identical preprocessing and data partitioning conditions and evaluated using accuracy, macro recall, macro F1-score, High-risk recall, and ROC-AUC to ensure a fair and consistent comparison.

## III. Results and Discussions

This chapter reports the classification results obtained from the evaluated machine learning models and discusses their performance using accuracy, recall, and ROC-AUC metrics. The analysis focuses on comparative performance, sensitivity to high-risk cases, and the relationship between model complexity and predictive effectiveness.

### *A. Overall Classification Performance*

The overall classification results indicate that the evaluated models exhibit distinct performance profiles when applied to survey-based depression risk data. Figure 2 provides a visual summary of overall classification performance across the selected models using accuracy, macro-averaged recall, and macro F1-score. Logistic Regression and Support Vector Machine demonstrate the most consistent performance across all evaluation metrics, achieving the highest accuracy values (0.95) alongside strong macro-averaged recall and F1-scores. Despite their relatively simple decision boundaries, both models effectively capture the dominant patterns embedded in the one-hot encoded categorical features. This suggests that the underlying structure of the dataset allows linear or near-linear classifiers to perform competitively.

Random Forest achieves slightly lower overall performance, with an accuracy of 0.91 and reduced macro-averaged recall compared to the linear models. Although ensemble learning improves robustness by aggregating multiple decision trees, the observed performance gain remains limited in this context. The radar chart in Figure 1 highlights this trade-off, where Random Forest exhibits a smaller performance area despite its higher model complexity. Overall, the results show that increasing algorithmic complexity does not necessarily translate into superior classification performance. Instead, simpler and more interpretable models, such as Logistic Regression and Support Vector Machine, provide a favorable balance between predictive accuracy and model stability for depression risk classification based on survey data.

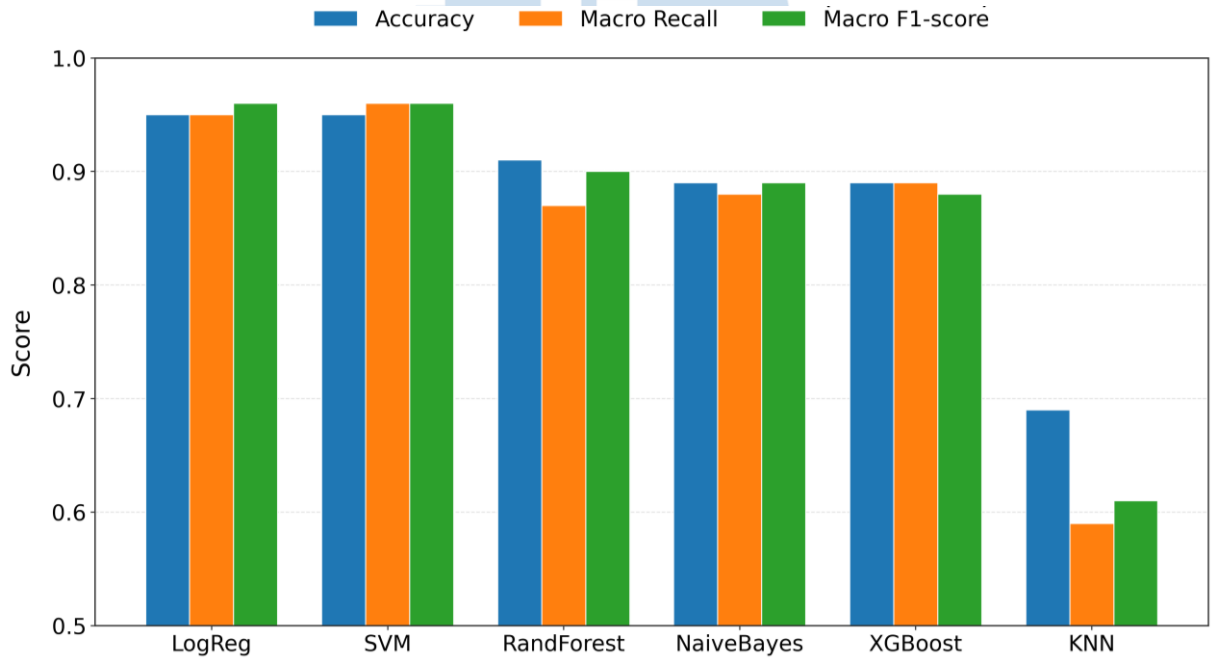


Fig. 2. Overall classification performance including accuracy, macro recall, and macro F1-score for all evaluated models.

### *B. Sensitivity to High Risk Cases (Recall)*

Sensitivity to high-risk cases is a critical evaluation aspect in depression risk classification, as false negatives may delay timely intervention. Therefore, this subsection focuses on recall for the High-risk class to assess each model's ability to correctly identify individuals with elevated risk. The results provided in Figure 3 indicate that Support Vector Machine achieves perfect recall (1.00) for the High-risk class, followed closely by Logistic Regression (0.96). These findings suggest that both margin-based and linear classifiers are highly effective in capturing discriminative patterns associated with high-risk responses in the survey data. Their strong performance aligns with the objective of risk screening, where prioritizing sensitivity over specificity is often preferred.

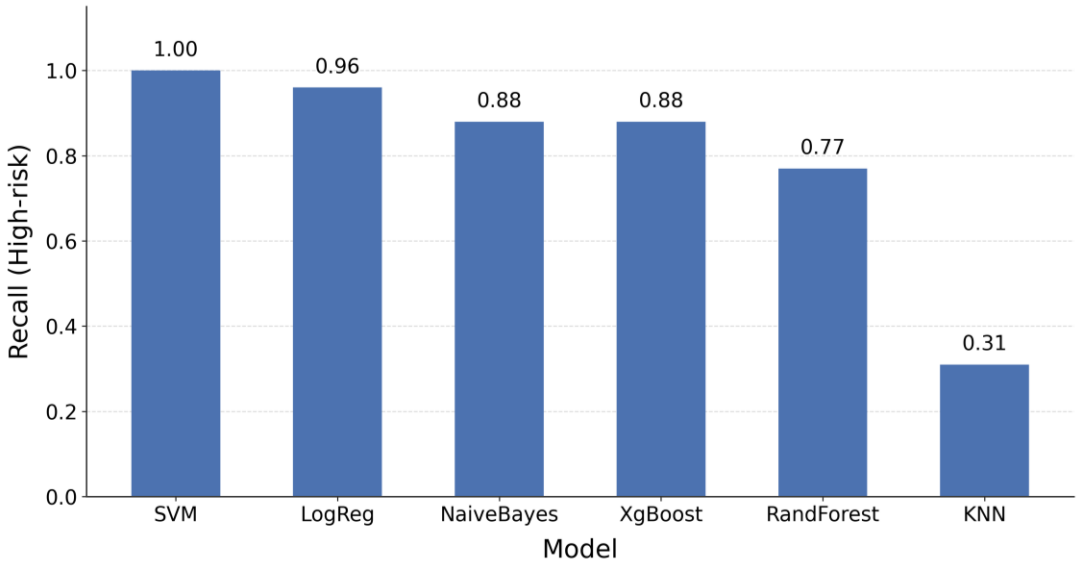


Fig. 3. Sensitivity to high-risk cases for all evaluated models.

Naïve Bayes and Decision Tree demonstrate moderate sensitivity, each achieving a recall of 0.88. While these models correctly identify a substantial proportion of high-risk cases, their performance remains slightly inferior to that of Logistic Regression and SVM. Random Forest shows a lower recall value (0.77), indicating that a portion of high-risk instances is misclassified into adjacent categories, most notably the Moderate-risk class. In contrast, K-Nearest Neighbors exhibits the lowest sensitivity (0.31) for high-risk detection. This substantial drop highlights the limitations of distance-based classifiers in high-dimensional, sparse feature spaces generated by one-hot encoding. Overall, the results emphasize that model selection plays a crucial role in high-risk screening tasks, with simpler linear and margin-based approaches offering superior sensitivity compared to more complex or distance-dependent methods.

### *C. Discriminative Ability (ROC-AUC)*

Discriminative ability was further examined using receiver operating characteristic (ROC) analysis for the High-risk class under a one-vs-rest setting. Unlike accuracy or recall, ROC-AUC provides a threshold-independent view of how well a model separates high-risk instances from the remaining categories across a continuum of decision thresholds. Overall, the ROC profiles indicate that Logistic Regression, Support Vector Machine, and Random Forest achieve near-perfect discrimination, with AUC values approaching unity. This suggests that these models consistently assign higher risk scores to High-risk cases than to Low or Moderate cases, even when the classification threshold is varied. The strong separation shown by these curves supports their suitability for screening-oriented applications where robust ranking performance is desirable.

In comparison, Naïve Bayes and Decision Tree show high but slightly reduced discriminative performance, indicating minor overlap in score distributions between High-risk and non–High-risk observations. The most noticeable performance gap is observed for K-Nearest Neighbors, which yields a substantially lower AUC and a ROC curve closer to the diagonal reference line. This pattern reflects weaker ranking capability in high-dimensional sparse feature spaces generated by one-hot encoding, consistent with its lower sensitivity reported in the previous subsection. Taken together, the ROC analysis reinforces the earlier findings on sensitivity: models with linear or margin-based decision structures provide strong and stable discrimination for High-risk detection, whereas distance-based classification is less reliable in this setting.

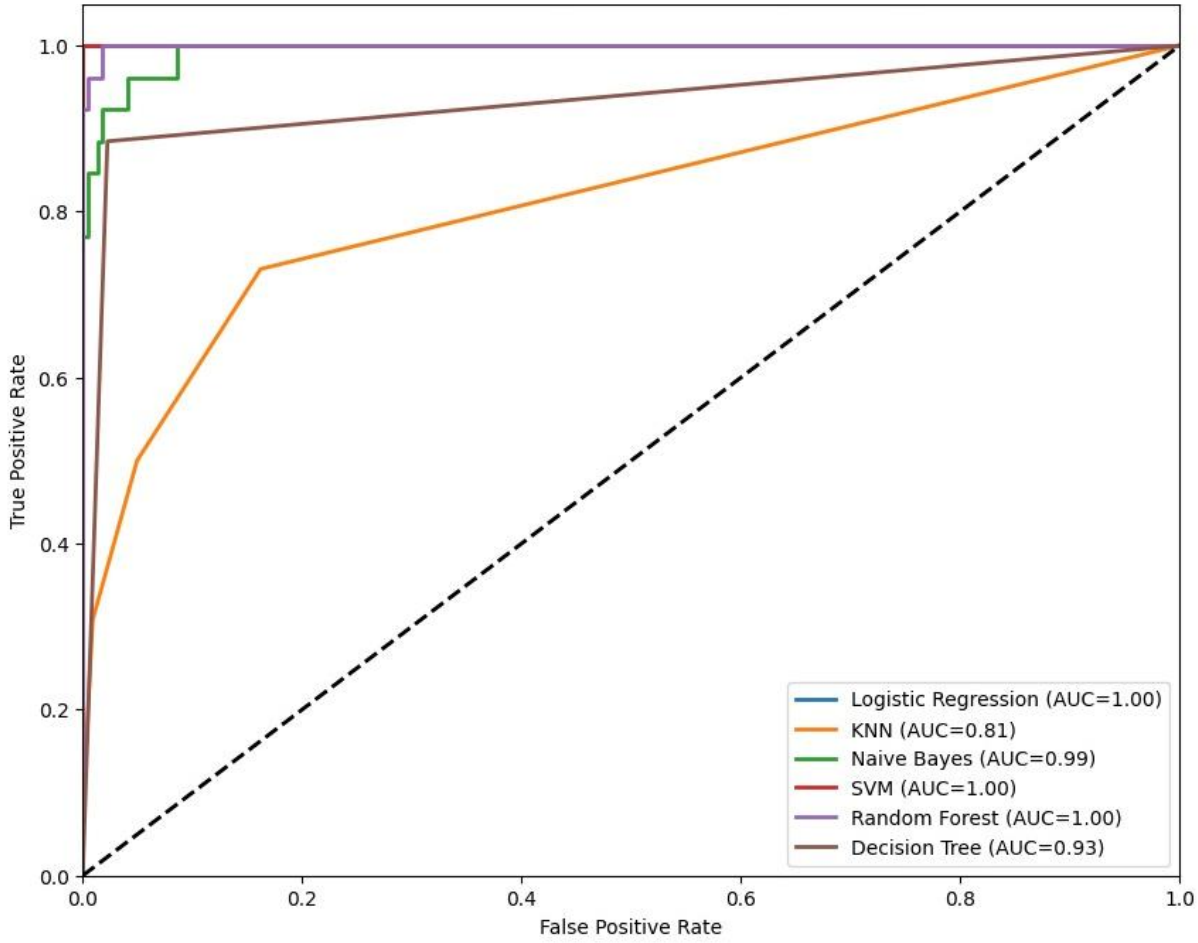


Fig. 4. Combined ROC curves for all evaluated models in detecting the high-risk class.

*D. Model Complexity vs Performance*

The relationship between model complexity and classification performance is a key consideration in practical depression risk assessment. While more complex algorithms are often expected to yield superior results, the findings of this study reveal that increased model complexity does not necessarily lead to improved performance when applied to survey-based mental health data. Across all evaluation dimensions overall performance, sensitivity to High-risk cases, and discriminative ability Logistic Regression and Support Vector Machine consistently match or outperform more complex models, including Random Forest and XGBoost. In contrast, ensemble-based and distance-based approaches exhibit mixed outcomes, indicating diminishing returns from increased structural complexity.

The superior performance of these simpler, linear, or margin-based models can be directly attributed to the characteristics of the preprocessed dataset. Survey data, which heavily relies on categorical responses, was transformed using one-hot encoding, inherently generating a high-dimensional and highly sparse feature space. In such sparse environments, linear models and SVMs excel because they can efficiently identify a global separating hyperplane without being heavily compromised by the sparsity; their strong performance suggests that the underlying relationships within the dataset are sufficiently linear or margin-separable. Conversely, tree-based ensemble models rely on greedy, recursive data splitting. When features are sparse and binary, individual splits yield very low information gain, making it difficult for tree architectures to construct robust decision paths, which often leads to suboptimal generalization. Similarly, distance-based algorithms like K-Nearest Neighbors show notable performance degradation and suffer significantly from the 'curse of dimensionality' in this one-hot encoded space, as the Euclidean distance between sparse binary vectors becomes uniform and loses its discriminative power. Thus, the inherent linearity of the transformed categorical data makes simpler models structurally better suited for this specific dataset.

From a practical perspective, the results emphasize the advantages of simpler and more interpretable models for depression risk screening. Models such as Logistic Regression and Support Vector Machine not only achieve high predictive accuracy but also offer greater transparency, ease of deployment, and lower computational overhead. These characteristics are

especially important in real-world mental health applications, where interpretability and reliability are often as critical as predictive performance.

## IV. CONCLUSION

This study presents a comparative evaluation of multiple machine learning algorithms for depression risk classification using survey-based mental health data. Model performance was assessed in terms of overall accuracy, sensitivity to High-risk cases, discriminative ability, and the relationship between model complexity and predictive performance. The results show that Logistic Regression and Support Vector Machine consistently outperform or match more complex models across all evaluation dimensions. Both approaches achieve high accuracy, strong sensitivity for High-risk detection, and excellent discriminative capability, indicating that linear and margin-based classifiers are sufficient to capture dominant patterns in categorical mental health survey data. In contrast, ensemble and distance-based methods demonstrate mixed performance, with no clear advantage from increased algorithmic complexity. From a practical perspective, these findings suggest that simple and interpretable models provide an effective balance between performance and usability for depression risk screening. Although this study is limited by its reliance on a single self-reported dataset, future work may explore external validation and alternative feature representations to further enhance model generalizability.

## ACKNOWLEDGMENT

The authors would like to acknowledge the Department of Medical Technology, Institut Teknologi Sepuluh Nopember, for the facilities and support in this research.

## REFERENCES

[1] C. Stecher, S. Cloonan, and M. E. Domino, "The Economics of Treatment for Depression," Annu. Rev. Public Health, vol. 45, no. 1, pp. 527–551, May 2024, doi: 10.1146/annurev-publhealth-061022-040533.

[2] D. Proudman, P. Greenberg, and D. Nellesen, "The Growing Burden of Major Depressive Disorders (MDD): Implications for Researchers and Policy Makers," Pharmacoeconomics, vol. 39, no. 6, pp. 619–625, Jun. 2021, doi: 10.1007/s40273-021-01040-7.

[3] K. S. Chaudhari, M. P. Dhapkas, A. Kumar, and R. G. Ingle, "Mental Disorders – A Serious Global Concern that Needs to Address," International Journal of Pharmaceutical Quality Assurance, vol. 15, no. 02, pp. 973–978, Jun. 2024, doi: 10.25258/ijpqa.15.2.66.

[4] A. Kupferberg and G. Hasler, "The social cost of depression: Investigating the impact of impaired social emotion regulation, social cognition, and interpersonal behavior on social functioning," J. Affect. Disord. Rep., vol. 14, p. 100631, Dec. 2023, doi: 10.1016/j.jadr.2023.100631.

[5] P. Greenberg and others, "The Economic Burden of Adults with Major Depressive Disorder in the United States (2019)," Adv. Ther., vol. 40, no. 10, pp. 4460–4479, Oct. 2023, doi: 10.1007/s12325-023-02622-x.

[6] P. Cuijpers and C. F. Reynolds, "Increasing the Impact of Prevention of Depression—New Opportunities," JAMA Psychiatry, vol. 79, no. 1, p. 11, Jan. 2022, doi: 10.1001/jamapsychiatry.2021.3153.

[7] G. Purebl, K. Schnitzspahn, and É. Zsák, "Overcoming treatment gaps in the management of depression with non-pharmacological adjunctive strategies," Front. Psychiatry, vol. 14, Nov. 2023, doi: 10.3389/fpsyt.2023.1268194.

[8] A. Faisal-Cury, C. Ziebold, D. M. de O. Rodrigues, and A. Matijasevich, "Depression underdiagnosis: Prevalence and associated factors. A population-based study," J. Psychiatr. Res., vol. 151, pp. 157–165, Jul. 2022, doi: 10.1016/j.jpsychires.2022.04.025.

[9] K. L. Mansfield, S. Puntis, E. Soneson, A. Cipriani, G. Geulayov, and M. Fazel, "Study protocol: the OxWell school survey investigating social, emotional and behavioural factors associated with mental health and well-being," BMJ Open, vol. 11, no. 12, p. e052717, Nov. 2021, doi: 10.1136/bmjopen-2021-052717.

[10] J. J. Newson, J. Bala, J. N. Giedd, B. Maxwell, and T. C. Thiagarajan, "Leveraging big data for causal understanding in mental health: a research framework," Front. Psychiatry, vol. 15, Feb. 2024, doi: 10.3389/fpsyt.2024.1337740.

[11] M. Z. Uddin, K. K. Dysthe, A. Følstad, and P. B. Brandtzaeg, "Deep learning for prediction of depressive symptoms in a large textual dataset," Neural Comput. Appl., vol. 34, no. 1, pp. 721–744, Jan. 2022, doi: 10.1007/s00521-021-06426-4.

[12] M. Garg, "Mental Health Analysis in Social Media Posts: A Survey," Archives of Computational Methods in Engineering, vol. 30, no. 3, pp. 1819–1842, Apr. 2023, doi: 10.1007/s11831-022-09863-z.

[13] M. Salama, H. A. Kader, and A. Abdelwahab, "An analytic framework for enhancing the performance of big heterogeneous data analysis," International Journal of Engineering Business Management, vol. 13, Jan. 2021, doi: 10.1177/1847979021990523.

[14] D. Feldner-Busztin and others, "Dealing with dimensionality: the application of machine learning to multi-omics data," Bioinformatics, vol. 39, no. 2, Feb. 2023, doi: 10.1093/bioinformatics/btad021.

[15] L. K. Andersen and B. J. Reading, "A supervised machine learning workflow for the reduction of highly dimensional biological data," Artificial Intelligence in the Life Sciences, vol. 5, p. 100090, Jun. 2024, doi: 10.1016/j.ailsci.2023.100090.

[16] W. Li and K. M. Kockelman, "How does machine learning compare to conventional econometrics for transport data sets? A test of ML versus MLE," Growth Change, vol. 53, no. 1, pp. 342–376, Mar. 2022, doi: 10.1111/grow.12587.

[17] S. K. Dhillon, M. D. Ganggayah, S. Sinnadurai, P. Lio, and N. A. Taib, "Theory and Practice of Integrating Machine Learning and Conventional Statistics in Medical Data Analysis," Diagnostics, vol. 12, no. 10, p. 2526, Oct. 2022, doi: 10.3390/diagnostics12102526.

[18] Md. M. Islam, S. Hassan, S. Akter, F. A. Jibon, and Md. Sahidullah, "A comprehensive review of predictive analytics models for mental illness using machine learning algorithms," Healthcare Analytics, vol. 6, p. 100350, Dec. 2024, doi: 10.1016/j.health.2024.100350.

[19] D. Ostojic, P. A. Lalousis, G. Donohoe, and D. W. Morris, "The challenges of using machine learning models in psychiatric research and clinical practice," European Neuropsychopharmacology, vol. 88, pp. 53–65, Nov. 2024, doi: 10.1016/j.euroneuro.2024.08.005.

[20] J. T. Hancock, T. M. Khoshgoftaar, and J. M. Johnson, "Evaluating classifier performance with highly imbalanced Big Data," J. Big Data, vol. 10, no. 1, p. 42, Apr. 2023, doi: 10.1186/s40537-023-00724-5.

[21] C. Mosquera, L. Ferrer, D. H. Milone, D. Luna, and E. Ferrante, "Class imbalance on medical image classification: towards better evaluation practices for discrimination and calibration performance," Eur. Radiol., vol. 34, no. 12, pp. 7895–7903, Jun. 2024, doi: 10.1007/s00330-024-10834-0.

[22] L. Wu, R. Huang, I. V Tetko, Z. Xia, J. Xu, and W. Tong, "Trade-off Predictivity and Explainability for Machine-

Learning Powered Predictive Toxicology: An in-Depth Investigation with Tox21 Data Sets," Chem. Res. Toxicol., vol. 34, no. 2, pp. 541–549, Feb. 2021, doi: 10.1021/acs.chemrestox.0c00373.

[23] K.-L. Du, B. Jiang, J. Lu, J. Hua, and M. N. S. Swamy, "Exploring Kernel Machines and Support Vector Machines: Principles, Techniques, and Future Directions," Mathematics, vol. 12, no. 24, p. 3935, Dec. 2024, doi: 10.3390/math12243935.

[24] M. Gimeno, K. S. del Real, and A. Rubio, "Precision oncology: a review to assess interpretability in several explainable methods," Brief. Bioinform., vol. 24, no. 4, Jul. 2023, doi: 10.1093/bib/bbad200.

[25] I. Salehin and N. Amin, "The RHMCD-20 Datasets for Depression and Mental Health Data Analysis with Machine Learning," 2024. doi: 10.17632/pxjmjyfdh2.

[26] W. L. Ku and H. Min, "Evaluating Machine Learning Stability in Predicting Depression and Anxiety Amidst Subjective Response Errors," Healthcare, vol. 12, no. 6, p. 625, Mar. 2024, doi: 10.3390/healthcare12060625.

[27] A. Dixit, A. K. Gupta, N. Chaplot, and V. Bharti, "Emoji Support Predictive Mental Health Assessment Using Machine Learning: Unveiling Personalized Intervention Avenues," International Journal of Experimental Research and Review, vol. 42, pp. 228–240, Aug. 2024, doi: 10.52756/ijerr.2024.v42.020.

[28] M. R. Rezvan, A. G. Sorkhi, J. Pirgazi, and M. M. P. Kallehbasti, "AdvanceSplice: Integrating N-gram one-hot encoding and ensemble modeling for enhanced accuracy," Biomed. Signal Process. Control, vol. 92, p. 106017, Jun. 2024, doi: 10.1016/j.bspc.2024.106017.

[29] S. Bagui, D. Nandi, S. Bagui, and R. J. White, "Machine Learning and Deep Learning for Phishing Email Classification using One-Hot Encoding," Journal of Computer Science, vol. 17, no. 7, pp. 610–623, Jul. 2021, doi: 10.3844/jcssp.2021.610.623.

[30] F. Pargent, F. Pfisterer, J. Thomas, and B. Bischl, "Regularized target encoding outperforms traditional methods in supervised machine learning with high cardinality features," Comput. Stat., vol. 37, no. 5, pp. 2671–2692, Nov. 2022, doi: 10.1007/s00180-022-01207-6.

[31] W. Yustanti, N. Iriawan, and I. Irhamah, "Categorical encoder based performance comparison in pre-processing imbalanced multiclass classification," Indonesian Journal of Electrical Engineering and Computer Science, vol. 31, no. 3, p. 1705, Sep. 2023, doi: 10.11591/ijeecs.v31.i3.pp1705-1715.

[32] Y. Sun and others, "Modifying the one-hot encoding technique can enhance the adversarial robustness of the visual model for symbol recognition," Expert Syst. Appl., vol. 250, p. 123751, Sep. 2024, doi: 10.1016/j.eswa.2024.123751.

[33] M. Klimo, P. Lukáč, and P. Tarábek, "Deep Neural Networks Classification via Binary Error-Detecting Output Codes," Applied Sciences, vol. 11, no. 8, p. 3563, Apr. 2021, doi: 10.3390/app11083563.

[34] M. Fallahpoor, S. Chakraborty, M. T. Heshejin, H. Chegeni, M. J. Horry, and B. Pradhan, "Generalizability assessment of COVID-19 3D CT data for deep learning-based disease detection," Comput. Biol. Med., vol. 145, p. 105464, Jun. 2022, doi: 10.1016/j.compbiomed.2022.105464.

[35] M. Sivakumar, S. Parthasarathy, and T. Padmapriya, "Trade-off between training and testing ratio in machine learning for medical image processing," PeerJ Comput. Sci., vol. 10, p. e2245, Sep. 2024, doi: 10.7717/peerj-cs.2245.

[36] S. N. Selamat, N. A. Majid, and A. M. Taib, "A Comparative Assessment of Sampling Ratios Using Artificial Neural Network (ANN) for Landslide Predictive Model in Langat River Basin, Selangor, Malaysia," Sustainability, vol. 15, no. 1, p. 861, Jan. 2023, doi: 10.3390/su15010861.

[37] M. Kim and K.-B. Hwang, "An empirical evaluation of sampling methods for the classification of imbalanced data," PLoS One, vol. 17, no. 7, p. e0271260, Jul. 2022, doi: 10.1371/journal.pone.0271260.

[38] J. Sadaiyandi, P. Arumugam, A. K. Sangaiah, and C. Zhang, "Stratified Sampling-Based Deep Learning Approach to Increase Prediction Accuracy of Unbalanced Dataset," Electronics (Basel)., vol. 12, no. 21, p. 4423, Oct. 2023, doi: 10.3390/electronics12214423.

[39] A. Jude and J. Uddin, "Explainable Software Defects Classification Using SMOTE and Machine Learning," Annals of Emerging Technologies in Computing, vol. 8, no. 1, pp. 36–49, Jan. 2024, doi: 10.33166/AETiC.2024.01.004.

[40] S. Kumar and V. Gota, "Logistic regression in cancer research: A narrative review of the concept, analysis, and interpretation," Cancer Research, Statistics, and Treatment, vol. 6, no. 4, pp. 573–578, Oct. 2023, doi: 10.4103/crst.crst_293_23.

[41] A. Zaidi and A. S. M. Al Luhayb, "Two Statistical Approaches to Justify the Use of the Logistic Function in Binary Logistic Regression," Math. Probl. Eng., vol. 2023, no. 1, Jan. 2023, doi: 10.1155/2023/5525675.

[42] A. Rizwan, N. Iqbal, R. Ahmad, and D.-H. Kim, "WR-SVM Model Based on the Margin Radius Approach for Solving the Minimum Enclosing Ball Problem in Support Vector Machine Classification," Applied Sciences, vol. 11, no. 10, p. 4657, May 2021, doi: 10.3390/app11104657.

[43] M. Azad, T. H. Nehal, and M. Moshkov, "A novel ensemble learning method using majority based voting of multiple selective decision trees," Computing, vol. 107, no. 1, p. 42, Jan. 2025, doi: 10.1007/s00607-024-01394-8.

[44] Y. Resti, C. Irsan, J. F. Latif, I. Yani, and N. R. Dewi, "A Bootstrap-Aggregating in Random Forest Model for Classification of Corn Plant Diseases and Pests," Science and Technology Indonesia, vol. 8, no. 2, pp. 288–297, Apr. 2023, doi: 10.26554/sti.2023.8.2.288-297.

[45] S. Gaïffas, I. Merad, and Y. Yu, "WildWood: A New Random Forest Algorithm," IEEE Trans. Inf. Theory, vol. 69, no. 10, pp. 6586–6604, Oct. 2023, doi: 10.1109/TIT.2023.3287432.

ISSN 2085-4579

# ERP Human Resource Management Readiness Framework Integrating Prioritization, Roadmaps, Risk Mitigation, McKinsey 7S

Safires Atalla Zaraski[1], Raymond Sunardi Oetama[2]
[1,2] Department of Information Systems. Universitas Multimedia Nusantara, Tangerang, Indonesia
safires.atalla@student.umn.ac.id
raymond@umn.ac.id

Accepted on June 6th, 2026
Approved on June 29th, 2026

***Abstract*—Enterprise Resource Planning (ERP) for Human Resource Management (HRM) is widely adopted to improve operational efficiency, data accuracy, and decision-making. However, successful implementation depends on organizational readiness. Previous studies have examined ERP readiness using various frameworks; however, quantitative ERP HRM readiness assessments remain limited, few studies prioritize organizational readiness factors, and integrated pre-implementation frameworks combining readiness assessment, implementation recommendations, roadmaps, and risk mitigation are still lacking. Therefore, this study proposes an integrated ERP HRM readiness assessment framework based on the McKinsey 7S Framework that quantitatively evaluates organizational readiness, prioritizes areas for improvement, and translates the assessment results into implementation recommendations, a phased ERP implementation roadmap, and risk mitigation strategies. A mixed-methods approach was employed through questionnaires and interviews. Data were collected from 30 employees using 22 indicators measured on a five-point Likert scale. Validity testing confirmed that all indicators were valid, and the overall Cronbach's alpha of 0.942 indicated excellent reliability. The readiness assessment showed that all seven dimensions were classified as Ready, with average scores ranging from 4.51 to 4.62. Structure achieved the highest readiness score, 4.62, while Skills achieved the lowest, 4.51. Interview findings identified challenges related to manual attendance recording, payroll processing, and fragmented employee data management. The results indicate that the organization is ready for ERP HRM implementation. This study contributes by proposing a structured ERP HRM readiness assessment framework, introducing a readiness-factor prioritization approach across the McKinsey 7S dimensions, and providing an integrated set of implementation recommendations, a roadmap, and risk-mitigation strategies.**

***Index Terms*—ERP; Human Resource Management; McKinsey 7S; Organizational Readiness.**

## I. INTRODUCTION

Enterprise Resource Planning (ERP) for Human Resource Management has become an important solution for organizations seeking to manage employee information, attendance, payroll, and other HR activities through a single integrated system. By streamlining administrative processes and providing accurate, real-time information, ERP can improve operational efficiency and support better decision-making [1]. However, successful implementation depends not only on the technology itself but also on the organization's readiness to adapt to new processes and ways of working [2]. Therefore, assessing ERP readiness is essential for identifying organizational strengths, potential challenges, and areas for improvement before implementation begins.

Although ERP offers significant benefits for Human Resource Management, many organizations still face challenges. ERP implementation is not easy to carry out. Around 70% of ERP implementations fail to achieve the expected results due to various factors such as a lack of organizational readiness, insufficient training, and resistance to change [3]. Organizational readiness is an important factor that must be considered before ERP implementation, because, without adequate readiness, the implemented system will not function optimally [4].

Furthermore, the organization currently faces several challenges in managing its human resources. Employee attendance is still recorded manually, increasing the risk of data inaccuracies and administrative inefficiencies [5]. Payroll calculations rely on manual attendance recaps, making them time-consuming and error-prone. In addition, employee information is not fully integrated into a centralized system [6], limiting data accessibility and complicating monitoring and reporting. These issues can reduce operational efficiency and hinder timely decision-making. Therefore, an integrated ERP solution is needed to streamline HR processes,

improve data accuracy, and enhance overall organizational performance.

Previous studies have widely examined ERP implementation readiness using various frameworks, such as the McKinsey 7S [4] or Critical Success Factors [7]. However, many studies focus on ERP implementation in general business functions and place greater emphasis on technological and operational aspects. Several studies have specifically investigated organizational readiness for ERP implementation within the Human Resource Management function. A Turkish study primarily examines the relationship between organizational readiness and technology acceptance variables, and does not assess ERP readiness from a holistic organizational perspective using the McKinsey 7S Framework [8]. While the PT HLM study also explored ERP HRM readiness using qualitative focus group discussions and identified key organizational factors that support implementation, it did not quantitatively measure the readiness level of each dimension or determine which factors required the greatest improvement [4].

Therefore, three research gaps remain. First, quantitative organizational readiness assessments specifically for ERP HRM implementation remain limited. Second, previous studies provide limited support for prioritizing organizational readiness factors that require managerial attention before ERP implementation. Third, existing studies rarely integrate readiness assessment results with implementation recommendations, deployment planning, and risk mitigation into a unified pre-implementation framework.

To address these gaps, this study proposes a structured ERP HRM readiness assessment framework based on the McKinsey 7S model. Unlike previous studies, the proposed framework not only quantitatively evaluates organizational readiness but also prioritizes areas for improvement across the seven organizational dimensions. Furthermore, it translates the assessment results into implementation recommendations, a phased ERP implementation roadmap, and risk mitigation strategies within a unified decision-support framework. The novelty of this study lies in integrating quantitative readiness assessment, readiness factor prioritization, implementation planning, and risk mitigation into a single framework to support ERP HRM implementation.

Accordingly, this study aims to assess organizational readiness for ERP implementation in the Human Resource Management function using the McKinsey 7S Framework. By evaluating the alignment of strategy, structure, systems, style, staff, skills, and shared values, the study identifies organizational strengths, prioritizes areas for improvement, and provides implementation recommendations to support successful ERP HRM adoption.

## II. METHODS

### A. *Proposed Framework*

The proposed ERP HRM Readiness Assessment Framework is shown in Fig. 1. The framework consists of four sequential stages: (1) Input, (2) Readiness Evaluation Model, (3) Readiness Factor Prioritization, and (4) Practical Implications. It provides a structured approach for assessing organizational readiness, identifying improvement priorities, and generating implementation recommendations, an ERP roadmap, and risk mitigation strategies to support ERP HRM implementation.

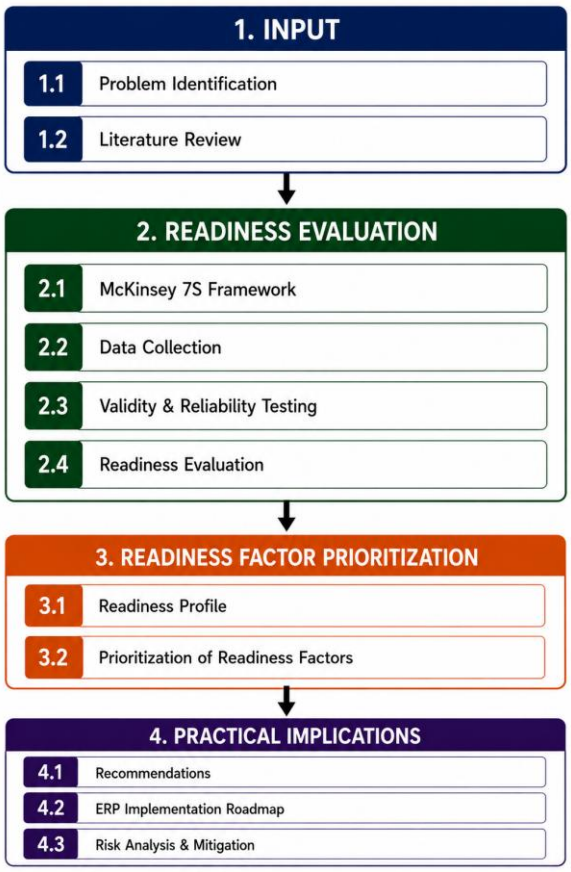


Fig. 1. Proposed ERP HRM Readiness Assessment Framework

### B. *Input*

The research is conducted at RP Company, an Indonesian Food and Beverage Company. Current HRM processes rely heavily on manual procedures for attendance recording, payroll calculation, and employee data management, resulting in inefficiencies, increased risk of errors, and limited data integration. These challenges highlight the need for an ERP system and motivate the assessment of organizational readiness before implementation. The

literature review stage involves reviewing prior studies and theories on ERP, Human Resource Management, organizational readiness, and the McKinsey 7S Framework. The purpose is to establish a theoretical foundation, identify research gaps, and develop indicators for assessing organizational readiness for ERP implementation.

### C. *Readiness Evaluation Model*

In this stage, the research instrument was developed based on the seven dimensions of the McKinsey 7S Framework, as shown in Fig. 2. The Framework is a model used to assess whether an organization is ready for change or improvement initiatives, such as ERP implementation [9]. It consists of seven interconnected elements: Strategy (the organization's plan to achieve its goals), Structure (how roles and responsibilities are organized), Systems (the processes and technologies used in daily operations), Style (leadership and management approach), Staff (employees within the organization), Skills (the knowledge and capabilities of employees), and Shared Values (the core beliefs and culture that guide organizational behavior). By evaluating these seven elements together, organizations can identify their strengths, weaknesses, and areas that need improvement to ensure a successful ERP implementation.

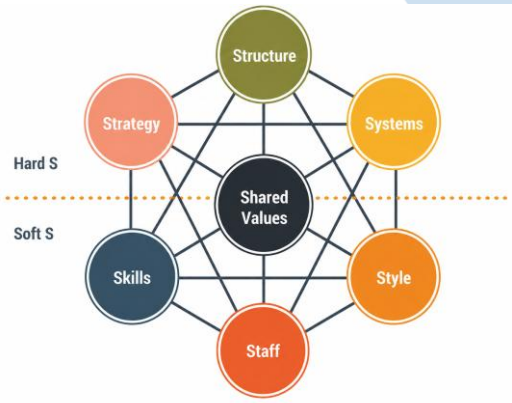


Fig. 2. McKinsey 7S Framework

The McKinsey 7S Framework was selected because it provides a comprehensive assessment of organizational readiness by evaluating seven interconnected elements: Strategy, Structure, Systems, Style, Staff, Skills, and Shared Values [10]. Unlike frameworks that focus primarily on technology or project success factors, McKinsey 7S examines both hard elements (strategy, structure, and systems) and soft elements (leadership, people, competencies, and organizational culture), which are critical for successful ERP implementation [2]. Since ERP adoption involves significant organizational change, this framework is well-suited to identify strengths, weaknesses, and alignment among organizational components, thereby providing a holistic view of readiness before implementation.

"Three steps were taken to ensure the final questionnaire was based on theoretical explication and applicable in our study. In Step 1, ERP and Organizational Readiness indicators were selected from validated studies identified in the literature. In Step 2, each indicator was mapped to one of the seven dimensions of the McKinsey® 7S model (Strategy, Structure, Systems, Style, Staff, Skills, and Shared Values) to achieve a comprehensive representation of organizational readiness constructs and to eliminate redundant or unrelated indicators. In Step Three, selected indicators were re-rendered in terms of HRM processes used in our case organization (i.e., time, attendance, payroll, employee information) and ERP adoption. In Step Three, any conceptually similar indicators were merged into a single item, and redundant questionnaire items were eliminated. As a result of this process, we developed an instrument comprising 22 questionnaire items, as shown in Table I, that provides an adequate representation of all 7 dimensions of the McKinsey 7S Models and contains relevant, clear, and appropriate content for assessing ERP HRM readiness. As the questionnaire was taken from previously validated tools, no formal expert opinion was provided about its adaptation. Nonetheless, the tool used in the readiness evaluation underwent Pearson Product-Moment Correlation testing to assess construct validity and Cronbach's Alpha to assess internal consistency before its implementation. The questionnaire applies a 5-point Likert scale [11].

TABLE I. QUESTIONNAIRE INDICATORS

| No | Code | Questionnaire Item | Source |
|---|---|---|---|
| 1 | STG1 | The organization has a clear HR management strategy that supports its business objectives. | [4] |
| 2 | STG2 | Current HR strategies effectively support employee performance. | [4] |
| 3 | STG3 | Organizational strategies are aligned with operational requirements. | [4] |
| 4 | STG4 | The organization has a plan to develop systems supporting ERP implementation. | [12] |
| 5 | STR1 | Roles and responsibilities within the organization are clearly defined. | [4] |
| 6 | STR2 | Coordination and communication among departments are effective. | [4] |
| 7 | STR3 | The organizational structure supports ERP implementation. | [4] |
| 8 | SYS1 | HR procedures and workflows are well documented. | [13] |
| 9 | SYS2 | An integrated ERP system is perceived to improve work effectiveness. | [13] |
| 10 | SYS3 | Technology is utilized | [13] |

| No | Code | Questionnaire Item | Source |
|---|---|---|---|
| | | effectively in attendance and payroll processes. | |
| 11 | STY1 | The organization promotes a supportive work culture. | [14] |
| 12 | STY2 | Management is open to employee feedback and suggestions. | [14] |
| 13 | STY3 | Leadership actively supports organizational change and ERP adoption. | [14] |
| 14 | STA1 | The organization has sufficient human resources to support operations. | [15] |
| 15 | STA2 | Employees perform their responsibilities effectively. | [15] |
| 16 | STA3 | Employees are ready to adapt to ERP implementation. | [15] |
| 17 | SKI1 | Employee competencies match job requirements. | [14] |
| 18 | SKI2 | Employees understand their roles and responsibilities well. | [14] |
| 19 | SKI3 | Employees are capable of learning and using ERP technology. | [14] |
| 20 | SV1 | Employees understand the organization's core values. | [4] |
| 21 | SV2 | Organizational values encourage innovation and change. | [4] |
| 22 | SV3 | Employees are committed to achieving organizational goals. | [13] |

This study employed purposive sampling, a non-probability sampling technique in which respondents were selected based on specific criteria relevant to the research objectives. The sample consisted of 30 employees directly involved in operational and Human Resource Management (HRM) activities and expected to be affected by the future ERP implementation. This technique was chosen because these employees possess the knowledge and experience necessary to evaluate organizational readiness, existing HR processes, and potential challenges associated with ERP adoption. Therefore, purposive sampling ensured that the collected data were relevant and capable of providing meaningful insights into the organization's readiness for ERP implementation. Since ERP HRM implementation will directly affect employees' daily activities, respondents from various operational functions were included to assess their perceptions of organizational readiness, support for change, and preparedness for ERP implementation. Recent methodological studies emphasize that there is no universal minimum sample size for survey research. Instead, the required sample size depends on the study design, population size, statistical precision, and analytical method [15]. Because the organization consisted of only 32 employees, this study prioritized maximizing population coverage and obtained responses from 30 employees (93.75% of the total population), resulting in near-census coverage and minimizing the potential for non-response bias.

Instrument quality was evaluated using Pearson Product-Moment Correlation and Cronbach's Alpha, as the objective of this study was an organizational readiness assessment rather than theory testing or model development [16]. The product-moment correlation formula (1) is used for validity testing, in which an item is considered valid if its correlation coefficient (r-count) exceeds the critical value (r-table) [17].

$$r_{xy} = \frac{N\Sigma xy_{-(\sum x)}(\sum y)}{\sqrt{(N\Sigma x^2 - (\sum x)^2(N\Sigma y^2 - (\Sigma y)^2)}} \quad (1)$$

Reliability testing was conducted to evaluate the internal consistency of the questionnaire items using (2). In this study, a Cronbach's Alpha coefficient greater than 0.60 was considered to indicate acceptable reliability [18].

$$r_{11} = \left(\frac{n}{n-1}\right)\left(1 - \frac{\sum \sigma_t^2}{\sigma_t^2}\right) \quad (2)$$

The mean score calculated using (3) from the dimension assessment determines the readiness level status. A score of 1.00–2.00 indicates the organization is not ready; 2.01–3.00 indicates lower readiness; 3.01–4.00 indicates moderate readiness; and 4.01–5.00 indicates the organization is ready for ERP implementation. These alignment rules provide a practical basis for measuring an organization's consistency in guiding ERP system implementation. The readiness-level classification in this study is similar to that of Firdaus et al. [18], with minor 203 overlapping scoring intervals and enhanced understanding of the readiness levels.

$$\bar{X}_d = \frac{\sum_{i-1}^{m_d} X_{di}}{m_d} \quad (3)$$

*D. Readiness Factor Prioritization*

A radar chart displays the mean scores for organizational readiness across the McKinsey 7S dimensions. According to the 7S framework, organizational change will be successful only if all these dimensions are aligned [19]. In this research, the alignment was measured by the Readiness Classification Criteria. If the dimensions were classified as Ready, this would demonstrate strong alignment throughout the organization. Moderate alignment exists when more than half of the dimensions are classified as Ready, and the rest are classified as Moderate Readiness. Weak alignment exists when one or more dimensions are classified as Lower Readiness or Not Ready, and is reflected in lower readiness scores for those dimensions [20].

In this study, organizational alignment is assessed using Readiness factor prioritization, which ranks the readiness scores of the seven McKinsey 7S dimensions to determine improvement priorities prior to ERP HRM implementation. The priority ranking was established by ordering the mean readiness scores for each of the McKinsey 7S dimensions from lowest to highest [21]. Lower scores indicate higher implementation priority, while higher scores indicate lower priority due to greater organizational readiness.

The readiness score is calculated by taking the mean readiness score for each of the 7 McKinsey 7S Dimensions and ranking them from lowest to highest. The first three dimensions (lowest scores) were considered the Improvement Priorities, as they will require the greatest management attention before the ERP (HRM Implementation) begins. The medium-ranked dimension (midpoint) is considered to be in the Maintain Category and has a satisfactory level of readiness for ERP implementation; thus, it should continue to be maintained and monitored. The three upper-ranked dimensions (highest scores) were considered to be the Organizational Strength dimensions and represent the strengths of the organization that should be maintained during the implementation of the ERP.

### *E. Practical Implication*

Recommendations were developed by systematically translating the readiness assessment results into improvement actions [22], as shown in Table II. The recommendation is based on an analysis of the key findings during data collection.

TABLE II. READINESS CLASSIFICATION AND RECOMMENDATION TABLE

| Score | Recommendation |
|---|---|
| 1.00–2.00 | ERP implementation should be postponed. |
| 2.01–3.00 | **Correction**: The organization should address critical gaps identified in the readiness assessment. |
| 3.01–4.00 | **Improvement**: The organization may begin ERP preparation while implementing targeted improvements in dimensions with lower readiness scores. |
| 4.01–5.00 | **Action**: The organization is ready to proceed with ERP implementation. Efforts should focus on executing a phased implementation roadmap. |

The ERP Implementation Road Map was created by combining the readiness assessment findings with the ERP implementation life cycle [23]. First, we analyzed the readiness and prioritization data to identify which organizational capabilities required more management attention. Second, we matched our prioritized readiness factors to each phase of the ERP implementation process to determine when each organizational capability should be addressed. Thirdly, based on our readiness assessment findings, we broke down all implementation activities into five phases (i.e., project preparation, system analysis and design, development and data preparation, testing and training, deployment, and finally, post-implementation evaluation).

Risk analysis was conducted by identifying potential implementation risks from the organizational readiness assessment results, interview findings, and existing organizational conditions [24] The identified risks were subsequently grouped into six categories: human resources, skills, data, technology, management, and operational risks [3],[13],[15]. Each risk was then evaluated for its potential impact on ERP HRM implementation, and appropriate mitigation strategies were formulated by drawing on ERP implementation best practices reported in previous studies.

## III. RESULTS AND DISCUSSION

### *A. Profile of Respondents*

Data were collected through interviews and the distribution of questionnaires. The questionnaires were distributed to 30 employees involved in operational and Human Resource Management (HRM) activities. At the same time, interviews were conducted with key personnel responsible for HRM and operations to gain a deeper understanding of the organization's current conditions and ERP implementation needs. Data were collected between March 9 and March 16, 2026. Table III presents the respondents' profiles. Most respondents were male (66.7%) and aged 18–30 years (70.0%). The largest job group was waiters (40.0%), followed by chefs and bartenders (20.0% each).

TABLE III. RESPONDENT'S DEMOGRAPHIC PROFILES

| Demographic Profiles | Composition | Frequency | Percentage (%) |
|---|---|---|---|
| Gender | Male | 20 | 66.7 |
| | Female | 10 | 33.3 |
| Age | 18–30 | 21 | 70.0 |
| | 31–40 | 5 | 16.7 |
| | 41–50 | 1 | 3.3 |
| | 51+ | 3 | 10.0 |
| Job Posisiton | Chef | 6 | 20.0 |
| | Bartender | 6 | 20.0 |
| | Cashier and Administration | 2 | 6.7 |
| | Waiter | 12 | 40.0 |
| | Cleaning Service | 4 | 13.3 |

### *B. Interview Findings*

The interview results with the HR Manager revealed several challenges in the current HRM processes. Employee attendance is still recorded manually because the previous fingerprint system is no longer functioning properly, leading to inefficiencies

and an increased risk of inaccurate records. Payroll processing is also performed manually from attendance recaps, making it time-consuming and error-prone. In addition, employee information is not managed in a centralized system, leading to difficulties with data retrieval, monitoring, reporting, and decision-making. These findings indicate that the existing HRM processes are not fully integrated and highlight the need for an ERP system to improve operational efficiency, data accuracy, and information accessibility.

### C. *Validity and Reliability Testing Results*

All questionnaire items were found to be valid, as shown in Table IV, because their r-count values exceeded the r-table value of 0.361 (degree of freedom = 28, n = 30, α = 0.05, two-tailed). This indicates that each indicator accurately measures the intended construct within the McKinsey 7S dimensions. Therefore, all 22 questionnaire items were deemed appropriate for assessing organizational readiness for ERP HRM implementation.

TABLE IV. VALIDITY TEST RESULTS

| Factor | Code | r-count | r-table | Result |
|---|---|---|---|---|
| Strategy | STG1 | 0.746 | 0.361 | Valid |
| | STG2 | 0.871 | 0.361 | Valid |
| | STG3 | 0.725 | 0.361 | Valid |
| | STG4 | 0.912 | 0.361 | Valid |
| Structure | STR1 | 0.687 | 0.361 | Valid |
| | STR2 | 0.835 | 0.361 | Valid |
| | STR3 | 0.926 | 0.361 | Valid |
| Systems | SYS1 | 0.852 | 0.361 | Valid |
| | SYS2 | 0.875 | 0.361 | Valid |
| | SYS3 | 0.651 | 0.361 | Valid |
| Style | STY1 | 0.865 | 0.361 | Valid |
| | STY2 | 0.933 | 0.361 | Valid |
| | STY3 | 0.597 | 0.361 | Valid |
| Staff | STA1 | 0.788 | 0.361 | Valid |
| | STA2 | 0.823 | 0.361 | Valid |
| | STA3 | 0.721 | 0.361 | Valid |
| Skills | SKI1 | 0.861 | 0.361 | Valid |
| | SKI2 | 0.595 | 0.361 | Valid |
| | SKI3 | 0.902 | 0.361 | Valid |
| Shared Values | SV1 | 0.870 | 0.361 | Valid |
| | SV2 | 0.870 | 0.361 | Valid |
| | SV3 | 0.805 | 0.361 | Valid |

All McKinsey 7S dimensions achieved Cronbach's Alpha values greater than 0.60 as shown in Table V, indicating that the questionnaire items are internally consistent and reliable. Therefore, the instrument can be considered dependable for measuring organizational readiness for ERP HRM implementation.

TABLE V. RELIABILITY TEST RESULTS

| No | Factor | Cronbach's Alpha | Reliability Level |
|---|---|---|---|
| 1 | Strategy | 0.809 | Reliable |
| 2 | Structure | 0.733 | Reliable |
| 3 | Systems | 0.703 | Reliable |
| 4 | Style | 0.729 | Reliable |
| 5 | Staff | 0.656 | Reliable |
| 6 | Skills | 0.684 | Reliable |
| 7 | Shared Values | 0.792 | Reliable |

The reliability test results for all questionnaire items using Cronbach's Alpha produced a value of 0.942 across 22 statement items. This value indicates that the questionnaire instrument has a very high level of internal consistency. In other words, each statement in the questionnaire provides stable, closely related measures of the constructs being assessed. Therefore, the questionnaire used in this study is highly reliable and suitable as the basis for subsequent data analysis.

### D. *Readiness Assessment Results*

Higher mean scores reflect greater organizational preparedness for successful ERP adoption. Table VI shows the results of the McKinsey 7S Strategic Readiness Survey with SD = Strongly Disagree, D = Disagree, N = Neutral, A = Agree, and SA = Strongly Agree. The results indicate that all McKinsey 7S factors achieved average scores ranging from 4.51 to 4.62, placing them in the Ready category for ERP implementation. The highest readiness score was observed in the Structure factor (4.62), followed by Staff (4.61), Style (4.59), Systems (4.58), Strategy and Shared Values (4.56), while Skills obtained the lowest score (4.51). Overall, the findings suggest that the organization possesses adequate organizational structure, management support, systems, employee readiness, competencies, and shared values to support successful ERP implementation.

TABLE VI. MCKINSEY 7S STRATEGIC READINESS SURVEY

| Code | SD | D | N | A | SA | Mean | Std. Dev. |
|---|---|---|---|---|---|---|---|
| STG1 | 0 | 0 | 3 | 9 | 18 | 4.50 | 0.682 |
| STG2 | 0 | 0 | 1 | 8 | 21 | 4.67 | 0.547 |
| STG3 | 0 | 0 | 1 | 6 | 23 | 4.73 | 0.521 |
| STG4 | 0 | 3 | 2 | 7 | 18 | 4.33 | 0.994 |
| Average Score and Status | | | | | | 4.56 | Ready |
| **Code** | **SD** | **D** | **N** | **A** | **SA** | **Mean** | **Std. Dev.** |
| STR1 | 0 | 0 | 0 | 8 | 22 | 4.73 | 0.450 |
| STR2 | 0 | 0 | 3 | 9 | 18 | 4.50 | 0.682 |
| STR3 | 0 | 0 | 0 | 11 | 19 | 4.63 | 0.490 |
| Average Score and Status | | | | | | 4.62 | Ready |
| **Code** | **SD** | **D** | **N** | **A** | **SA** | **Mean** | **Std. Dev.** |
| SYS1 | 0 | 1 | 3 | 7 | 19 | 4.47 | 0.819 |
| SYS2 | 0 | 1 | 1 | 10 | 18 | 4.50 | 0.731 |
| SYS3 | 0 | 0 | 0 | 7 | 23 | 4.77 | 0.430 |
| Average Score and Status | | | | | | 4.58 | Ready |
| **Code** | **SD** | **D** | **N** | **A** | **SA** | **Mean** | **Std. Dev.** |
| STY1 | 0 | 0 | 3 | 7 | 20 | 4.57 | 0.679 |
| STY2 | 0 | 0 | 1 | 12 | 17 | 4.53 | 0.571 |
| STY3 | 0 | 0 | 0 | 10 | 20 | 4.67 | 0.479 |
| Average Score and Status | | | | | | 4.59 | Ready |
| **Code** | **SD** | **D** | **N** | **A** | **SA** | **Mean** | **Std. Dev.** |
| STA1 | 0 | 3 | 0 | 9 | 18 | 4.50 | 0.682 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| STA2 | 0 | 0 | 1 | 11 | 18 | 4.57 | 0.568 |
| STA3 | 0 | 0 | 0 | 7 | 23 | 4.77 | 0.430 |
| Average Score and Status | | | | | | 4.61 | Ready |
| **Code** | **SD** | **D** | **N** | **A** | **SA** | **Mean** | **Std. Dev.** |
| SKI1 | 0 | 0 | 2 | 11 | 17 | 4.50 | 0.630 |
| SKI2 | 0 | 0 | 0 | 9 | 21 | 4.70 | 0.466 |
| SKI3 | 0 | 3 | 2 | 7 | 18 | 4.33 | 0.994 |
| Average Score and Status | | | | | | 4.51 | Ready |
| **Code** | **SD** | **D** | **N** | **A** | **SA** | **Mean** | **Std. Dev.** |
| SV1 | 0 | 0 | 3 | 11 | 16 | 4.43 | 0.679 |
| SV2 | 0 | 0 | 0 | 11 | 19 | 4.63 | 0.490 |
| SV3 | 0 | 0 | 2 | 7 | 21 | 4.63 | 0.615 |
| Average Score and Status | | | | | | 4.56 | Ready |

### *E. Readiness Profile*

Fig. 3 shows a visualization of the Readiness Profile, where each of the seven McKinsey 7S dimensions stands with relatively high scores - all mean scores for all dimensions were between 4.51 and 4.62 on the readiness scale, indicating that the organizational elements are well aligned to support ERP HRM implementation.

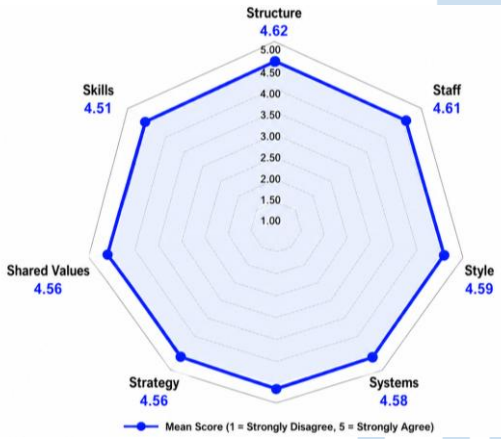


Fig. 3. McKinsey 7S Readiness Assessment Score Radar Chart

### *F. Prioritization of Readiness Factors*

The readiness factors were ranked in ascending order by their mean readiness scores. Lower scores reflect higher priority for improvement. Skills were the highest priority (4.51). The next priorities were Strategy and shared values (4.56), Systems (4.58), Style (4.59), and Staff (4.61), and the lowest priority was Structure (4.62). All six dimensions were classified as Ready. The results for the Prioritization of Readiness Factors are shown in Table VII. Although all dimensions were classified as Ready, the ranking represents a relative comparison rather than a measure of poor performance. Lower-ranked dimensions are prioritized because they have relatively lower readiness than the others, while higher-ranked dimensions represent strengths to be maintained. This enables management to focus improvement efforts without overlooking the organization's overall high level of readiness.

TABLE VII. PRIORITIZATION OF READINESS FACTORS

| Rank | Dimension | Score | Category |
|---|---|---|---|
| 1 | Skills | 4.51 | Improvement Priority |
| 2 | Strategy | 4.56 | Improvement Priority |
| 2 | Shared Values | 4.56 | Improvement Priority |
| 4 | Systems | 4.58 | Maintain |
| 5 | Style | 4.59 | Organizational Strength |
| 6 | Staff | 4.61 | Organizational Strength |
| 7 | Structure | 4.62 | Organizational Strength |

### *G. Recommendation*

Recommendations developed from the findings of this evaluation focused solely on the Improvement Priority and Maintain dimensions of the 7S model, as they offer the greatest potential to enhance the organisation's preparedness for the ERP HRMS implementation, as shown in Table VIII. While the results show that all sub-dimensions are in the "Ready" state, the lower-Importance sub-dimensions should receive more managerial time than the "Maintain" sub-dimensions, which should receive constant monitoring. Both Organisational Strengths should be maintained and used to support a successful ERP HRMS Implementation, and should not be improved further. ERP and IT training will equip employees with the skills needed to perform their jobs; phased implementation will enable organizations to align ERP with their organizational goals; cultural reinforcement will foster collaboration and adoption of new technology; and the development of an integrated HRM system will replace fragmented, manual processes. The goal of the recommendations provided here is to increase organizational readiness, reduce implementation risks, and support the successful implementation of ERP HRM.

TABLE VIII. RECOMMENDATIONS BASED ON PRIORITIZATION OF READINESS FACTORS

| **Factor** | **Key Findings** | **Recommendations** |
|---|---|---|
| Skills | ERP implementation requires employees with adequate ERP and IT competencies. | Improve employee capabilities through ERP training and strengthen basic information technology competencies. |
| Strategy | A clear ERP implementation strategy is required to support operational efficiency and future business growth. | Develop a phased ERP implementation plan based on operational needs. |
| Shared Values | Successful ERP implementation requires a culture that supports technology adoption, collaboration, and organizational commitment. | Build a work culture that reinforces discipline, responsibility, teamwork, and technology adoption. |
| Systems | ERP implementation requires integrating HR | Develop an integrated HRM system covering |

| | | |
|---|---|---|
| | processes to replace fragmented manual procedures. | attendance, employee data, payroll, and scheduling, supported by an adequate IT infrastructure. |

### *H. Roadmap*

The roadmap, as shown in Table IX, indicates when these competencies are developed throughout the life cycle of an ERP system. Strategic direction and organizational commitment are set during project preparation; business process alignment occurs during the analysis and design phase; System Configuration & Data Quality is achieved through development and Data Preparation; and User Competency is developed through testing and training. Deployment is focused on System Adoption, and Post Implementation Evaluation is focused on Continuous Improvement. Therefore, the road map translates the readiness assessment into an actionable implementation plan, helping organizations reduce implementation risks and increase their chances of successfully adopting ERP HRM. The proposed six-month timeline is intended as a practical reference rather than a fixed schedule. It represents a typical duration for completing the major ERP HRM implementation phases in a small- to medium-sized organization. The actual timeline may be adjusted according to organizational size, project scope, resource availability, and implementation complexity.

TABLE IX. ERP HRM IMPLEMENTATION ROADMAP

| Phase | Activities | Expected Output | Timeline |
|---|---|---|---|
| Preparation | Establish an ERP project team, define the project scope, allocate the budget, and communicate the project objectives. | Project charter and implementation plan. | Month 1 |
| System Analysis and Design | Analyze HR business processes, define system requirements, and configure ERP HRM modules. | Business process mapping and system design. | Month 2 |
| Development and Data Preparation | System customization, employee data cleansing, and data migration preparation. | Configured ERP system and validated master data. | Month 3 |
| Testing and Training | Conduct system testing, User Acceptance Testing (UAT), and employee training. | Tested the system and trained users. | Month 4 |
| Go-Live and Implementation | Deploy the ERP HRM system and monitor system performance. | Operational ERP HRM system. | Month 5 |
| Evaluation and Continuous Improvement | Evaluate implementation outcomes and address identified issues. | Performance evaluation report. | Month 6 |

As shown in Table X, the ERP implementation risk analysis identified six major risk categories: human resources, skills, data, technology, management, and operational risks. The most critical concerns include employee resistance, limited ERP knowledge, data migration issues, and system integration challenges. To minimize these risks, the company should implement change management initiatives, provide training and technical support, perform data validation and system testing, ensure continuous management commitment, and adopt a phased deployment strategy. These mitigation actions can improve the likelihood of a successful ERP HRM implementation.

Each identified risk will have at least one related organizational capability required for the successful implementation of the ERP HRM system. The risk of human resources and skills has been found to be critical, since your workforce's acceptance of the new system and their ability to use it are vital to ensuring the successful implementation and use of your organization's system. The data and technology risks stem from common technical problems organizations frequently encounter during data migration, system integration, and ERP deployment. The management risks highlight the importance of sustained leadership commitment throughout implementation, while the operational risks reflect the potential for temporary disruptions in business operations as organizations transition from manual processes to an ERP-based system. Finally, for each identified risk, we propose a risk mitigation strategy that organizations can use to reduce its potential negative effects and increase their chance. The purpose of this risk analysis is to identify potential implementation risks and propose mitigation strategies based on the readiness assessment results, rather than to quantitatively prioritize project risks using formal risk assessment techniques such as FMEA or Risk Matrix. The benefits of a successful ERP HRM implementation.

TABLE X. ERP IMPLEMENTATION RISK ANALYSIS

| Risk Category | Potential Risk | Impact | Mitigation Strategy |
|---|---|---|---|
| Human Resources | Employee resistance to new work processes. | Delayed adoption and reduced system utilization. | Conduct change management and user training programs. |

| Risk Category | Potential Risk | Impact | Mitigation Strategy |
|---|---|---|---|
| Skills | Limited ERP knowledge among employees. | User errors and inefficient system usage. | Provide ERP training and technical support. |
| Data | Incomplete or inaccurate employee data during migration. | Data inconsistency and reporting issues. | Perform data validation and data cleansing before migration. |
| Technology | System integration issues. | Operational disruption and data synchronization problems. | Conduct comprehensive system testing before go-live. |
| Management | Lack of continuous management support. | Reduced project commitment and implementation delays. | Establish regular project monitoring and executive sponsorship. |
| Operational | Temporary productivity decline during transition. | Reduced operational efficiency. | Implement phased deployment and provide user assistance. |

*I. Discussion*

Fig. 3 shows that all seven dimensions of the McKinsey 7S model were rated above 4.50. This means that the organization is in a strong state of readiness on all dimensions and is aligned to implement the ERP HRM system. Structure had the highest readiness score (4.62), indicating a clear delineation of roles and responsibilities, effective coordination among entities, and a sound organizational structure that serves as a solid foundation for ERP adoption. While the Skills dimension scored the lowest (4.51), employees are still rated as ready; they have the least competency in using the systems, and rate the use of information technology as the lowest priority among all dimensions. Therefore, the organization will benefit from implementing a competency-based ERP Training Program prior to ERP implementation to maximize the benefits of the strong Structural Readiness for ERP HRM implementation. The paper's findings align with prior studies on the organizational aspect of enterprise resource planning (ERP) readiness. It goes beyond the scope of prior readiness studies by providing a quantified measure of organizational ERP readiness through the readiness score, prioritized readiness factors with implementation recommendations, and a single decision framework for use as an implementation roadmap. The proposed research framework would assess an organization's readiness and provide management with guidance on which specific areas to prioritize for improvement before implementing an ERP system.

The comparison in Table XI shows how this research builds on the previous studies about McKinsey 7s as they relate to how ready the organization is for an ERP, because it includes the quantitative measurement of the readiness, the prioritization of the readiness factors, recommendations for how to implement, a roadmap for implementing an ERP, and risk analysis and mitigation, all in one framework. By having all of these within one framework, organizations can not only determine what their level of readiness is, but also know what their areas of higher priority are to improve, and they can convert the results of the readiness assessment into the actual steps to implement the ERP system.

TABLE XI. COMPARISON WITH OTHER FRAMEWORKS

| Capability | Masfi [25] | Cahayadi [4] | This Study |
|---|---|---|---|
| McKinsey 7S framework | ✓ | ✓ | ✓ |
| Quantitative readiness assessment | ✗ | ✗ | ✓ |
| Qualitative exploration | ✗ | ✓ | ✓ |
| Validity & reliability testing | ✓ | ✗ | ✓ |
| Readiness score for each 7S dimension | ✗ | ✗ | ✓ |
| Readiness prioritization/ranking | ✗ | ✗ | ✓ |
| Readiness profile | ✓ | ✓ | ✓ |
| ERP implementation recommendations | ✗ | ✓ | ✓ |
| ERP implementation roadmap | ✗ | ✗ | ✓ |
| ERP Implementation Risk Analysis and Mitigation | ✗ | ✗ | ✓ |

The proposed roadmap for ERP implementation and the risk mitigation strategies are important in two ways: they translate the results of the readiness assessment into actionable implementation items. The roadmap was systematically derived by mapping the prioritized readiness factors to the ERP implementation life cycle, while the risk analysis was developed by integrating the readiness assessment results, interview findings, and evidence from previous ERP implementation studies. Therefore, both outputs are directly grounded in the proposed readiness assessment framework rather than being developed independently. While the readiness assessment identifies an organization's current strengths and areas for development, the roadmap provides a systematic sequence of steps that will guide the organization through each implementation process. Risk mitigation strategies assist organizations in identifying potential risks, such as employee resistance, inadequate ERP competencies, poor data migration, and integration issues among systems, before these risks negatively

impact project success. Thus, the suggested Framework serves as an effective decision-making tool for managers, enabling them to identify and rank areas for improvement based on available readiness scores, allocate resources more efficiently, and prepare the organization for implementing the ERP HRM system.

## IV. CONCLUSION

This research used the McKinsey 7S model to determine if the company was prepared to implement an ERP Human Resource Management system. The results demonstrated that all seven elements of the model were deemed to be in a state of Readiness, indicating an excellent foundation for moving forward with the creation of an ERP system within the organization. Of these elements/constructs of the 7S Model, Structure had the highest Readiness score; therefore, an organization with well-defined roles, effective communication, and a clear organizational chart is in place to support the implementation of ERP HRM. Although Skills received the lowest Readiness score of all the elements of the 7S Model, they still met the criteria for being considered Ready. This indicates that additional training on the ERP system and user development programs to enhance staff skills beyond current levels is likely to contribute to the overall success of implementing the ERP system and achieving business objectives.

Additionally, the results suggest that aligning all seven dimensions of the McKinsey 7S Model creates an optimal environment for successfully deploying an ERP HRM system within an organization. Interviews revealed inefficiencies in employee attendance tracking, payroll processing, and records management – all of which required a more integrated HRM system. Therefore, based on these findings, the organization is ready to implement the ERP HRM system with expected improvements in operational efficiency, data accuracy, interdepartmental collaboration, and decision-making capabilities. The study will also provide the organization with recommendations on implementing the ERP system, an ERP implementation roadmap, and ways to mitigate risks, ensuring the adoption process is structured for sustainable results.

The study presents three main contributions. The first contribution is that the study presents a comprehensive ERP HRM Readiness Assessment Model developed using the McKinsey 7S Framework to assist organizations in evaluating their Readiness prior to implementing an ERP System. The second contribution of the study introduces a method for prioritizing readiness factors that identifies organizational strengths and the same seven dimensions of the McKinsey 7S model, enabling more effective resource allocation. The third contribution of the study is that it uses the readiness assessment results to recommend implementation activities, a phased implementation plan, and risk mitigation strategies, thereby providing organizations with a pragmatic decision-support framework in the pre-implementation phase.

This study has some shortcomings. It is executed at a single organization, and the results have therefore been obtained from a fairly small sample of 30 individuals. As a result, there may be limitations on the generalizability of the results. In addition, the study relied exclusively on the McKinsey 7S Framework to assess readiness for an ERP implementation, but no post-implementation measurement was conducted. Future research should clearly involve both a larger population and multiple organizations to establish a much broader basis for generalization, to explore the use of the McKinsey 7S Framework in combination with other ERP readiness frameworks, and to use the proposed framework to measure the level of readiness for an ERP implementation during the process of implementing an ERP system across multiple types of organizations.

## ACKNOWLEDGMENT

The authors would like to thank Universitas Multimedia Nusantara for providing academic support and research facilities.

## REFERENCES

[1] J. Wiratama and V. Kuswanto, “An evaluation of integrating erp system to develop a strategy business,” in *2023 International Conference on Information Management and Technology (ICIMTech)*, IEEE, 2023, pp. 1–6.

[2] S. Shiri, A. Anvari, and H. Soltani, “Identifying and prioritizing of readiness factors for implementing ERP based on agility (extension of McKinsey 7S model),” *Eur. Online J. Nat. Soc. Sci. Proc.*, vol. 4, no. 1 (s), p. pp-56, 2015.

[3] S. D. Larasati, I. Eitiveni, and P. Mahardhika, “Analysis of ERP critical failure factors: A case study in an Indonesian mining company,” *J. Sist. Inf.*, vol. 19, no. 2, pp. 34–47, 2023.

[4] P. A. Cahayadi, A. Faza, and J. Wiratama, “Exploring organizational readiness for ERP HRM implementation using the McKinsey 7S Framework,” *AITI*, vol. 22, no. 2, pp. 236–250, 2025.

[5] A. Munshi, N. Aljojo, A. Zainol, R. Al-Saadi, and B. Babteen, “Employee attendance monitoring system by applying the concept of enterprise resource planning (ERP),” *Int. J. Educ. Manag. Eng.*, vol. 9, no. 5, pp. 1–9, 2019.

[6] M. Elian-Gabriel and L. MIHAI-VALENTIN-CĂTĂLIN, “The Importance Of Using Erp Systems In Human Capital Reporting,” *Ann. Ser.*, vol. 3, pp. 156–169, 2024.

[7] J. Wiratama and R. Sutomo, “Integrating Balanced Scorecard and Enterprise Management for ERP Readiness Assessment,” in *2025 8th International Conference on New Media Studies (CONMEDIA)*, IEEE, 2025, pp. 97–102. doi: https://doi.org/10.33379/gtech.v7i4.2488.

[8] M. Esen and G. K. Özbağ, “An investigation of the effects of organizational readiness on technology acceptance in e-HRM applications,” *Int. J. Hum. Resour. Stud.*, vol. 4, no. 1, p. 232, 2014.

[9] P. Hanafizadeh and A. Z. Ravasan, “A McKinsey 7S model-based framework for ERP readiness assessment,” *Int. J. Enterp. Inf. Syst.*, vol. 7, no. 4, pp. 23–63, 2011.

[10] W. H. Organization, “Managing change towards universal health coverage service provision in Africa,” 2023.

[11] B. Tanujaya, R. C. I. Prahmana, and J. Mumu, “Likert scale in social sciences research: Problems and difficulties,” *FWU J.*

*Soc. Sci.*, vol. 16, no. 4, pp. 89–101, 2022.
[12] R. Sutomo, "Developing a Pre-Implementation Model for the ERP Material Management Module in the PVC Supply Industry in Indonesia," *JOIV Int. J. Informatics Vis.*, vol. 9, no. 4, pp. 1571–1583, 2025.
[13] J. W. Santo Fernandi Wijaya and A. E. J. Egeten, "Modeling the readiness measurement for enterprise resource planning system implementation success," *J. Nas. Tek. elektro dan Teknol. Inf.*, vol. 12, no. 3, pp. 159–166, 2023.
[14] A. Errida, B. Lotfi, and Z. Chatibi, "Development of an assessment model of organizational change readiness by using fuzzy logic," *Procedia Comput. Sci.*, vol. 219, pp. 1909–1919, 2023.
[15] S. M. Jayadeva, R. R. Shikhare, and S. Verma, "Factors affecting the effectiveness of HRIS (Human Resource Information System)-an empirical study," *J. Posit. Sch. Psychol.*, vol. 6, no. 5, pp. 5795–5802, 2022.
[16] S. D. Alfian *et al.*, "Evaluation of usability and user feedback to guide telepharmacy application development in Indonesia: a mixed-methods study," *BMC Med. Inform. Decis. Mak.*, vol. 24, no. 1, p. 130, 2024.
[17] A. Anto, S. Sugiyanto, Y. Yuliati, and A. Kustanti, "Validity and reliability of the adoption questionnaire of agricultural mechanization in the food estate area of central Kalimantan, Indonesia," *Int. J. Sci. Technol. Manag.*, vol. 4, no. 4, pp. 736–741, 2023.
[18] J. Sigudla and J. E. Maritz, "Exploratory factor analysis of constructs used for investigating research uptake for public healthcare practice and policy in a resource-limited setting, South Africa," *BMC Health Serv. Res.*, vol. 23, no. 1, p. 1423, 2023.
[19] C. June, "The Impact of Change Management on Organizational Performance-a Case Study of a Medium-sized Organization in the Information Technology Sector in Sub-Saharan Africa Title: The Impact of Change Management on Organizational Performance-a Case Study of a Med," 2023.
[20] R. D. Wiwoho, F. T. Hastuti, H. Hernando, and H. F. Thousani, "The McKinsey Framework as an Analysis of Organizational Change Readiness: Public Service Agency State Polytechnic," *Int. J. Heal. Econ. Soc. Sci.*, vol. 7, no. 2, pp. 573–579, 2025.
[21] K. Sathish Kumar, R. Venkatesh Babu, K. P. Paranitharan, and A. Saravana Kumar, "Fifty criteria based integrated quality healthcare system readiness assessment model in organization using scoring approach," in *AIP Conference Proceedings*, AIP Publishing LLC, 2024, p. 20017.
[22] M. Ashar, "EERP IN PUBLIC SECTOR REFORM: A SYSTEMATIC LITERATURE REVIEW OF TECHNOLOGICAL, ORGANIZATIONAL, AND INSTITUTIONAL FACTORS," *J. Res. Soc. Sci. Econ. Manag.*, vol. 4, no. 10, 2025.
[23] A. Lamey, H. Abdelkader, A. Keshk, and S. Eletriby, "A realistic and practical guide for creating intelligent integrated solutions in higher education using enterprise architecture," *Sustainability*, vol. 15, no. 11, p. 8780, 2023.
[24] A. Efe, "Risk modelling of cyber threats against MIS and ERP applications," *Pamukkale Üniversitesi İşletme Araştırmaları Derg.*, vol. 11, no. 2, pp. 502–530, 2024.
[25] A. Masfi, T. Sukartini, and A. A. A. Hidayat, "Performance improvement model utilizing the McKinsey 7S approach for public health centers in Sampang Regency of Indonesia," *Int. J. Sci. Technol. Res.*, vol. 9, no. 3, p. 5073â, 2020.

# AUTHOR GUIDELINES

## 1. Manuscript criteria

- The article has never been published or in the submission process on other publications.
- Submitted articles could be original research articles or technical notes.
- The similarity score from plagiarism checker software such as Turnitin is 20% maximum.
- For December 2021 publication onwards, Ultima Infosys : Jurnal Ilmu Sistem Informasi will be receiving and publishing manuscripts written in English only.

## 2. Manuscript format

- Article been type in Microsoft Word version 2007 or later.
- Article been typed with 1 line spacing on an A4 paper size (21 cm x 29,7 cm), top-left margin are 3 cm and bottom-right margin are 2 cm, and Times New Roman's font type.
- Article should be prepared according to the following author guidelines in this template. Article contain of minimum 3500 words.
- References contain of minimum 15 references (primary references) from reputable journals/conferences

## 3. Organization of submitted article

The organization of the submitted article consists of Title, Abstract, Index Terms, Introduction, Method, Result and Discussion, Conclusion, Appendix (if any), Acknowledgment (if any), and References.

- Title
  The maximum words count on the title is 12 words (including the subtitle if available)
- Abstract
  Abstract consists of 150-250 words. The abstract should contain logical argumentation of the research taken, problem-solving methodology, research results, and a brief conclusion.
- Index terms
  A list in alphabetical order in between 4 to 6 words or short phrases separated by a semicolon ( ; ), excluding words used in the title and chosen carefully to reflect the precise content of the paper.
- Introduction
  Introduction commonly contains the background, purpose of the research, problem identification, research methodology, and state of the art conducted by the authors which describe implicitly.
- Method
  Include sufficient details for the work to be repeated. Where specific equipment and materials are named, the manufacturer's details (name, city and country) should be given so that readers can trace specifications by contacting the manufacturer. Where commercially available software has been used, details of the supplier should be given in brackets or the reference given in full in the reference list.
- Results and Discussion
  State the results of experimental or modeling work, drawing attention to important details in tables and figures, and discuss them intensively by comparing and/or citing other references.
- Conclusion
  Explicitly describes the research's results been taken. Future works or suggestion could be explained after it
- Appendix and acknowledgment, if available, could be placed after Conclusion.
- All citations in the article should be written on References consecutively based on its' appearance order in the article using Mendeley (recommendation). The typing format will be in the same format as the IEEE journals and transaction format.

## 4. Reviewing of Manuscripts

Every submitted paper is independently and blindly reviewed by at least two peer-reviewers. The decision for publication, amendment, or rejection is based upon their reports/recommendations. If two or more reviewers consider a manuscript unsuitable for publication in this journal, a statement explaining the basis for the decision will be sent to the authors within six months of the submission date.

## 5. Revision of Manuscripts

Manuscripts sent back to the authors for revision should be returned to the editor without delay (maximum of two weeks). Revised manuscripts can be sent to the editorial office through the same online system. Revised manuscripts returned later than one month will be considered as new submissions.

## 6. Editing References

- **Periodicals**
  J.K. Author, “Name of paper,” Abbrev. Title of Periodical, vol. x, no. x, pp. xxx-xxx, Sept. 2013.

- **Book**
  J.K. Author, “Title of chapter in the book,” in Title of His Published Book, xth ed. City of Publisher, Country or Nation: Abbrev. Of Publisher, year, ch. x, sec. x, pp xxx-xxx.

- **Report**
  J.K. Author, “Title of report,” Abbrev. Name of Co., City of Co., Abbrev. State, Rep. xxx, year.

- **Handbook**
  Name of Manual/ Handbook, x ed., Abbrev. Name of Co., City of Co., Abbrev. State, year, pp. xxx-xxx.

- **Published Conference Proceedings**
  J.K. Author, “Title of paper,” in Unabbreviated Name of Conf., City of Conf., Abbrev. State (if given), year, pp. xxx-xxx.

- **Papers Presented at Conferences**
  J.K. Author, “Title of paper,” presented at the Unabbrev. Name of Conf., City of Conf., Abbrev. State, year.

- **Patents**
  J.K. Author, “Title of patent,” US. Patent xxxxxxx, Abbrev. 01 January 2014.

- **Theses and Dissertations**
  J.K. Author, “Title of thesis,” M.Sc. thesis, Abbrev. Dept., Abbrev. Univ., City of Univ., Abbrev. State, year. J.K. Author, “Title of dissertation,” Ph.D. dissertation, Abbrev. Dept., Abbrev. Univ., City of Univ., Abbrev. State, year.

- **Unpublished**
  J.K. Author, “Title of paper,” unpublished. J.K. Author, “Title of paper,” Abbrev. Title of Journal, in press.

- **On-line Sources**
  J.K. Author. (year, month day). Title (edition) [Type of medium]. Available: http:// www.(URL) J.K. Author. (year, month). Title. Journal [Type of medium]. volume(issue), pp. if given. Available: http://www.(URL) Note: type of medium could be online media, CD-ROM, USB, etc.

## 7. Editorial Adress

Universitas Multimedia Nusantara
Jl. Scientia Boulevard, Gading Serpong
Tangerang, Banten, 15811
Email: **ultimainfosys@umn.ac.id**

UMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA

ISSN 2085-4579



Universitas Multimedia Nusantara
Scientia Garden Jl. Boulevard Gading Serpong, Tangerang
Telp. (021) 5422 0808 | Fax. (021) 5422 0800