ISSN 2085-4552

ULTIMATICS

Jurnal Teknik Informatika

Engineering | Vol. 14 | No. 1 | Page 1 - 50 | June 2022 | ISSN 2085-4552

UMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA

ISSN 2085-4552

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 : ultimatics@umn.ac.id

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 Computer Science and Informatics, as well as the latest theoretical and practical issues, including Analysis and Design of Algorithm, Software Engineering, System and Network security, 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. Ultimatics : Jurnal Teknik Informatika is published regularly twice a year (June and December) and is managed by the Informatics Study Program at Universitas Multimedia Nusantara.

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 2085-4552

FOREWORD

ULTIMA Greetings!

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 Computer Science and Informatics, as well as the latest theoretical and practical issues, including Analysis and Design of Algorithm, Software Engineering, System and Network Security, 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. Ultimatics : Jurnal Teknik Informatika is published regularly twice a year (June and December) and is published by the Faculty of Engineering and Informatics at Universitas Multimedia Nusantara.

In this June 2022 edition, Ultimatics enters the 1st Edition of Volume 14. In this edition there are seven scientific papers from researchers, academics and practitioners in the fields of Computer Science and Informatics. Some of the topics raised in this journal are: Sentiment Analysis of An Internet Provider Company Based on Twitter Using Support Vector Machine and Naïve Bayes Method, The Influence of KMS Trello on the Intern Performance CDC UIN Jakarta, The Development of UMS Building Catalogue Information System, Feature Extraction using Lexicon on the Emotion Recognition Dataset of Indonesian Text, Elicitation of Needs Using User Personas to Improve Software User Experience, E-Business Software Product Line Methodology Based on SMEs Characteristics and Teakwood Grade Identification with GLCM and K-NN with Adaboost Optimization.

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 ultimatics@umn.ac.id and the web page of our Journal sini.

Finally, we would like to thank all contributors to this June 2022 Edition of Ultimatics. 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 2022,

M.B.Nugraha, S.T., M.T.
Ketua Dewan Redaksi

ISSN 2085-4552

TABLE OF CONTENT

ISSN 2085-4552

Sentiment Analysis of An Internet Provider Company Based on Twitter Using Support Vector Machine and Naïve Bayes Method

Farhan Hashfi[1], Dedy Sugiarto[2], Is Mardianto[3]
[1,3] Program Studi Teknik Informatika, Fakultas Teknologi Industri, Universitas Trisakti
[2] Program Studi Sistem Informasi, Fakultas Teknologi Industri, Universitas Trisakti
[1]farhan0640017000008@trisakti.ac.id, [2]dedy@trisakti.ac.id, [3]mardianto@trisakti.ac.id

Accepted 12 December 2021
Approved 20 February 2022

***Abstract*— Tweets from users in the form of opinions about a product can be used as a company evaluation of the product. To obtain this evaluation, the method that can be used is sentiment analysis to divide opinions into positive and negative opinions. This study uses 1000 data from Twitter related to an internet service provider company where the data is divided into two classes, namely 692 positive classes and 308 negative classes. In the Tweet there are still many words that are not standard. Therefore, previously carried out the initial process or preprocessing to filter out non-standard words. Before doing the classification, the data needs to be divided into training data and test data with a ratio of 90:10, then processed using the Support Vector Machine and Naïve Bayes techniques to get the results of the classification of positive opinions and negative opinions. The level of accuracy in the classification using the Support Vector Machine is 84% and using Naïve Bayes is 82%.**

***Index Terms*— Internet Provider; Naïve Bayes; Sentiment Analysis; Support Vector Machine; Twitter**.

I. INTRODUCTION

The internet has developed very rapidly to date in influencing media and communication. One of the factors supporting the success of the Internet in Indonesia is that infrastructure development has reached remote areas in Indonesia [1]. This can be proven by the increasing use of social media. Social Media is an Internet service most commonly used by Indonesian citizens. One of them is Twitter.

Twitter is used for various things such as sharing personal things, using it to sell, to reporting an opinion to a brand or company. Information shared on Twitter is typically 140 characters long [2]. In general, a company uses social media to gather information about the goods or services they offer. The most common use of social media by companies is to use social media for marketing activities and social media for customer service [3]. Therefore, the opinion group of Twitter users will be influenced by the emotions (emotions) that are classified in order to determine their polarization, namely positive opinions or negative opinions.

Sentiment analysis is the process of using text analysis to derive various data sources from the Internet and various other social media platforms. One of the purposes of sentiment analysis is to get someone's opinion on a company service and then classify that opinion into positive opinions and negative sentiments [4]. In conducting sentiment analysis, the technique used to retrieve data from Twitter uses the Crawling technique which requires API permission from the platform itself. Furthermore, the techniques for classifying the tweet data are Support Vector Machine, KNearest Neighbor, and Naïve Bayes [5]. By using this method, it produces a classification between two categories, namely positive opinions and negative opinions, where the results can be useful for observers of the company's data to determine the next marketing strategy.

II. THEORY

A. Sentiment Analysis

Sentiment analysis is one of the methods used to identify an opinion or sentiment expressed using a text or document and how that opinion is categorized as positive opinion and negative opinion. Basically, sentiment analysis tries to assess a different aspect of the standard language in order to help an agency or company to get positive opinions as well as negative opinions about the products they offer [6]. Sentiment itself can be interpreted as an emerging concept in which everyone's different emotions are determined from the content of the text, so that it can be processed to extract the opinions and sentiments of many people.

In sentiment analysis, there are 3 opinions that can be a reference for agencies or companies to obtain information on the quality of the products offered, namely positive opinions, negative opinions, and neutral opinions [7]. Sentiment refers to several topics, opinions on certain topics have different meanings from other opinions that are the same on other subjects. Analyst sentiment is usually used to determine the quality of services or the quality of a product from an

agency or company that wants to develop the services or products offered [8].

Sentiment analysis is a new part of research in Natural Language Processing (NLP) which aims to find subjectivity in texts or documents to classify opinions or sentiments. In the sentiment classification procedure, there are three techniques that can be used, namely Machine Learning, Lexicon Based, and Hybrid Approach. At this time, sentiment analysis is widely used using the Machine Learning method because this method is closer to the prediction of sentiment polarity based on the data that has been prepared [6].

B. *Data Crawling*

Crawling is a method of collecting data from a source to be analyzed or processed. Crawling is the first stage that is usually used to analyze the sentiments of social media users towards a product or service. Crawling can also be interpreted as a method of quickly gathering large numbers of web pages into a local storage and indexing them based on specified keywords [8]. Sentiment analysis usually uses the Crawling method on Twitter social media by utilizing the features provided by Twitter, namely Application Programming Interface Systems (APIs) [9].

Twitter provides features APIs that can be utilized using the crawling method to generate a collection of text data based on the desired keywords. Crawling tweet data on twitter is a method to get tweet data from twitter that gets data from the twitter server by utilizing these APIs features in the form of username data and tweet data as needed [9].

The data crawling method on Twitter is mostly done using the python programming language by utilizing the tweepy library provided by the python. That way, twitter data collection will be easier to obtain using the help of the python library and the APIs feature of twitter [10].

C. *Preprocessing*

Preprocessing is the stage of sentiment analysis to make documents more structured so that they are ready for analysis [7]. The steps taken for the preprocessing of this research are Case Folding, Cleansing, Tokenizing, Normalization, and Stopword Removal. In processing text mining data, it is necessary to carry out this stage [7].

1. *Case Folding*

   One of the steps of the text pre-processing stage which serves to convert all letters in the document to lowercase. This step is done to make the search easier.

2. *Cleansing*

   Cleansing is one of the steps used to remove links, mentions, hashtags, URLs, punctuation marks, numbers, and one letter. The purpose of this stage is to make the data used tidier for the next stage of text preprocessing.

3. *Tokenizing*

   This stage is used to make sentences or documents into words or tokens. This stage needs to be carried out for the next stage, namely changing all words that have wrong writing and are abbreviated. This stage is also carried out to carry out the word weighting stage.

4. *Normalization*

   After carrying out the Tokenizing stage, the next stage to do is the normalization stage, where this stage is used to change all words that have incorrect writing and are shortened to the words they should be. This stage is used so that there is no misunderstanding of the meaning of each word.

5. *Stopword Removal*

   The next step used for the text preprocessing process in this research is stopword removal. This stage is used to delete words that do not have important meaning.

D. *Term Frequency – Inverse Document Frequency (TF-IDF)*

Term Frequency-Inverse Document Frequency (TF-IDF) is a process for analyzing the interaction between phrases or sentences with multiple documents. The term frequency-reverse document frequency (TF-IDF) is the number of words available in a document. At the same time, the TF-IDF metric is statistical data that shows the importance of words in a data set or document [8]. Term frequency (TF) is an aspect that decides the weight of a word in a document based on the number of its presence in the document. When assigning weights to a word, the value of the number of occurrences of the word (term frequency) must be considered. The greater the weight given to the word, the greater the weight on a document, or provides a greater application value.

Inverse document frequency (IDF) is one aspect that reduces the excess of terms that often appear in documents. This is needed because terms that often appear in documents can be considered as general terms, so these terms are considered unimportant [5]. On the other hand, the term "scarcity factor in document collection" should be considered when determining the weights. TF-IDF can be disclosed namely:

$$TF - IDFt, d = TFt, d \times IDFt \quad (1)$$

Where IDFt = log N/DFt

t = counted words,
d = sentence weight (d),
TF-IDFtd = sentence weight (d) against word (t),

Tftd = Term Frequency,
IDFt = Inverse Document Frequency,
N = number of sentences,
DFt = number of words repeated

E. Support Vector Machine

A method used for linear classification is the Support Vector Machine (SVM), which can approximate the best ignition path, such as the separator between two categories. The supporting vector machine uses a linear function hypothesis space in the operation of high-dimensional features. The basic principle is linear classification, then it was developed to solve nonlinear problems by adding the concept of kernel techniques to high-dimensional workspaces [11].

$$a(B) = c\mathcal{x}\emptyset(B)+d \quad (2)$$

That is, B is a feature vector, c is a vector of different weights, is a function of non-linear mapping, and attribute d is a vector.

F. Naïve Bayes

Naïve Bayes is a classification method which holds that each term occurs as independent. In general, Naïve Bayes approach to probability or probability. The Naïve Bayes algorithm basically predicts future probabilities from past experiences [12].

$$p(C_k \mid x) = \frac{p(C_k)p(x|C_k)}{p(x)} \quad (3)$$

The above formula explains where P(x|C_k) is a conditional probability of the word x occurring in documents of class (C_k), P(C_k) is the previous probability of the dataset that has occurred in class C_k. P(x|ck) and P(ck) are estimated from the available data.

G. Evaluation

Evaluation was conducted to determine the accuracy of the modeling that has been applied to both methods. Then compare the results of two different data sets by applying the confusion matrix method. Confusion Matrix is a process that is generally used in data mining to calculate the level of accuracy. A classification system produces classification results and information will be loaded about the classification that has been correctly predicted by the Confusion Matrix [12]. Accuracy, Precision, and Recall are parameters to test the performance of calculations that have been done. Precision (P) is a parameter to find out how many results of processing that are relevant to the information you want to search for or with other bonds, namely the positive classification that is true (true positive) and the overall data which is predicted to be i-positive class. Precision can be obtained using equation [5].

$$P = \frac{tp}{tp+fp} \quad (4)$$

Recall (R) is how many irrelevant documents in the collection generated by the system or with other bindings are the number of documents which have a true classification positive (true positive) and all documents including true negative (true positive ones). Recall can be obtained by using equation [5].

$$R = \frac{tp}{tp+fn} \quad (5)$$

Furthermore, Accuracy (A) is the number of documents that are classified correctly, either true positive or true negative. Accuracy can be obtained by using equation [12].

$$A = \frac{(tp+fn)}{(tp+fp+tn+fn)} \quad (6)$$

Variables such as TN, TP, FN, and FP are derived from the confusion matrix [5].

TN = True Negative
TP = True Positive
FN = False Negative
FP = False Positive

III. METHOD

In this analysis, there are steps that are followed, namely generating tweet data from twitter using the twitter data crawling method, then the data is processed into more structured data using the preprocessing method, then the data classification uses the Naïve Bayes classification algorithm and the Support Vector Machine.

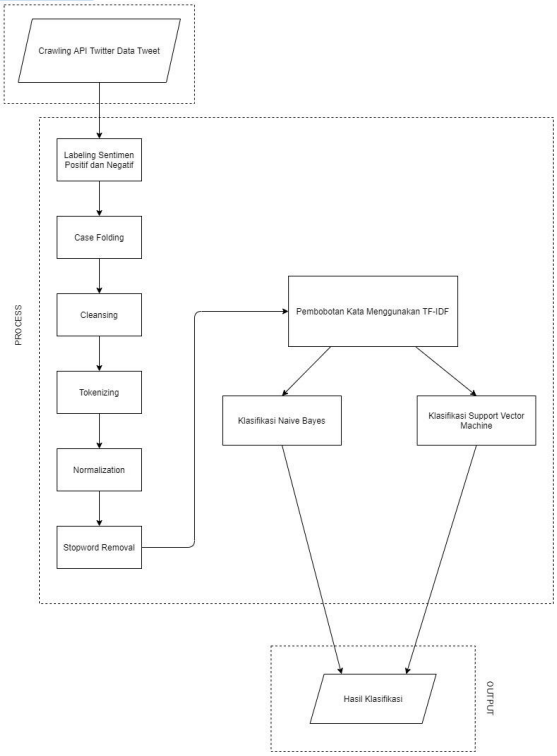

Fig 1. Flow of the Great Research Process

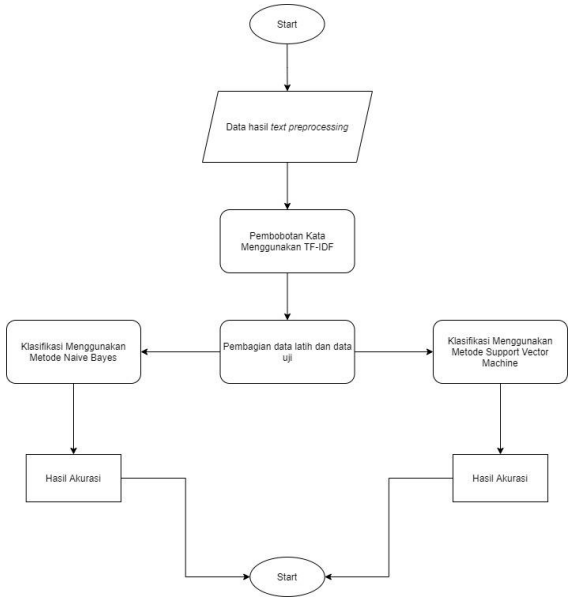

Fig 2. Classification Process Flow

Figure 1 shows the large process flow that was carried out to complete this research. Figure 2 shows the classification flow after the data has been structured through the text preprocessing process, involving two classification algorithms Naïve Bayes and Support Vector Machine [13].

IV. RESULT

The analysis process is carried out by testing and analyzing based on the results of sentiment analysis of Indonesian-speaking Indihome users on the Twitter platform using the Naive Bayes classification algorithm and Support Vector Machine [14].

A. Data

The tests carried out in this analysis used 1000 tweet data from Twitter with the keyword 'indihome' in Indonesian. This test uses 2 classes, namely positive and negative classes. The data collection is carried out using the Crawling technique that utilizes the tweepy library from the Python programming language. After that, labeling positive and negative sentiments on the dataset is done manually. For manual labeling of positive and negative sentiments, there should be the help of linguists in determining the positive or negative sentiments of an opinion. The author only uses two classes of sentiment, namely positive and negative because the author wants a more conical conclusion between the two classes.

TABLE I. TOTAL DATASET LABELS

Total Dataset Labels	
Positive	692
Negative	308
Total	1000

B. Text Preprocessing

After getting the dataset from the Crawling results and labeling positive and negative sentiments, the dataset needs to be changed into more neat and structured data to be processed using the Naïve Bayes algorithm and Support Vector Machine. Example of Text Preprocessing output based on the example of Table 2.

TABLE II. EXAMPLE OF TEXT PREPROCESSING

Before	After
indihome knp LOS dari tadi pagi?? ku mau drakoran help @IndiHomeCare	indihome kenapa los dari tadi pagi aku mau drakoran help

The Text Preprocessing step is needed to prepare the data so that the dataset becomes neater and more structured. Text Preprocessing has several stages in it, namely Case Folding, Cleansing, Tokenizing, Normalization, and Stopword Removal.

C. TF-IDF

After the dataset goes through the text preprocessing stage, then the dataset through the TF-IDF method functions for word weighting. Word weighting is needed to process datasets using the Naïve Bayes algorithm and Support Vector Machine. Example of term weighting based on Table 3.

TABLE III. TF-IDF WORD WEIGHTING EXAMPLE

Term	Weight
indihome	0.087980
banget	0.028923
rumah	0.026522
lambat	0.026136
kasih	0.022233

Word weighting utilizes the Scikit-learn library found in the Python programming language. The example in Table 3 shows 5 words that often appear in the dataset [15].

D. Classification Method

In this analysis, the classification process is carried out using two Naïve Bayes classification algorithms and the Support Vector Machine. From the results of the two classification methods have different accuracy values. The classification process is carried out using the Scikit-learn library contained in the Python programming language using the Naïve Bayes algorithm and the Support Vector Machine.

Before carrying out the classification method, the dataset needs to be divided into two parts including training data and test data. The ratio used in this study

uses 90:10 because according to research that has been done previously, it can be concluded that the higher the value of the training data, the better the level of accuracy obtained in the method used. The training data serves to train the model to recognize the existing patterns in the dataset, while the test data is used as test data for the classification method. In this study, the training data contained 90% of the dataset and the test data contained as much as 10% of the dataset. The results of testing the Naïve Bayes classification algorithm and Support Vector Machine are shown in Figure 3 and Figure 4

```
Multinomial Naive Bayes Acuracy: 82.0
Multinomial Naive Bayes Precision: 80.51948051948052
Multinomial Naive Bayes Recall: 95.38461538461539
Multinomial Naive Bayes f1_score: 87.32394366197184
==================================================
```

	precision	recall	f1-score	support
Negatif	0.81	0.95	0.87	65
Positif	0.87	0.57	0.69	35
accuracy			0.82	100
macro avg	0.84	0.76	0.78	100
weighted avg	0.83	0.82	0.81	100

Figure 3. Results of the Naïve Bayes method

```
Support Vector Machine Acuracy: 84.0
Support Vector Machine Precision: 82.66666666666667
Support Vector Machine Recall: 95.38461538461539
Support Vector Machine f1_score: 88.57142857142857
==================================================
```

	precision	recall	f1-score	support
Negatif	0.83	0.95	0.89	65
Positif	0.88	0.63	0.73	35
accuracy			0.84	100
macro avg	0.85	0.79	0.81	100
weighted avg	0.85	0.84	0.83	100

Figure 4. Results of the Support Vector Machine Method

The figure shows the results of Accuracy, Precision, Recall, and F1 Score from the two different classification methods, it can be seen that the accuracy generated in the Support Vector Machine calculation is higher than Naïve Bayes, namely the Accuracy value in the Support Vector Machine method is 84% while the accuracy value in the Support Vector Machine method is 84%. Naïve Bayes method is 82%.

E. Model Evaluation

The evaluation process uses a Confucian Matrix table to evaluate the results of the Naïve Bayes classification algorithm and Support Vector Machine. The results of the model evaluation can be seen based on Figure 5 and Figure 6

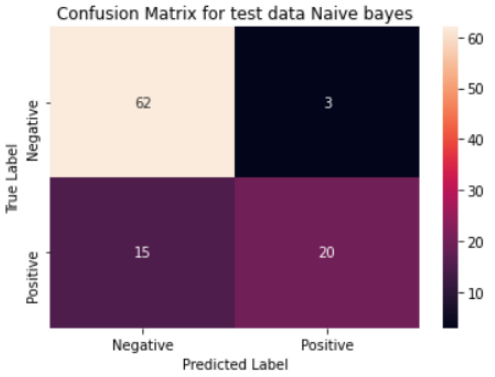

Fig 5. Evaluation of Naïve Bayes

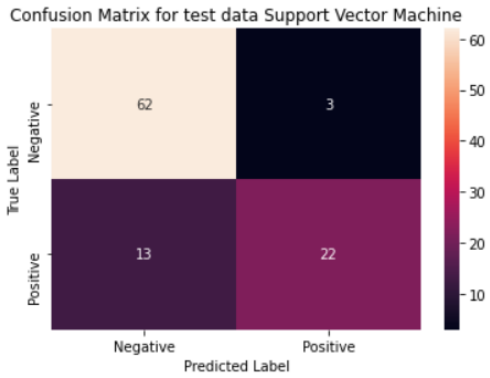

Fig 6. Evaluation of Support Vector Machine

From the evaluation value of the image above, it shows that the two classification methods carried out on this research dataset still have errors in classifying positive and negative sentiments, where both classification methods still place the positive sentiment class on negative sentiment, and vice versa.

V. CONCLUSIONS

In the results of research calculations that have been completed related to Indihome customer sentiment on Indihome services using the Naïve Bayes classification algorithm and Support Vector Machine to get the accuracy value, namely the accuracy of the Support Vector Machine algorithm is greater than the Naïve Bayes classification method. For this reason, in this study using 1000 Indihome customer datasets on the Twitter social media platform, the Support Vector Machine method is a better method than the Naïve Bayes method. Data is collected for 3 months starting from February 2021 to April 2021.

However, this research still has several shortcomings, namely the process of labeling positive and negative sentiments is done manually which produces more negative sentiments than positive sentiments. There are differences from the data labeling that is applied manually to test the model using the class prediction results from the model classification results. In addition, this study only uses 1000 datasets. The

accuracy of the Naive Bayes method is 82% while the Support Vector Machine is 84%.

REFERENCES

[1] F. Amalia, R. T. Sulistyo, and A. H. Brata, "Analisis Tingkat Penerimaan E-Learning Sebagai Alternatif Media Pembelajaran Pada Siswa SMK," *Smatika J.*, vol. 10, no. 02, pp. 41–47, 2020, doi: 10.32664/smatika.v10i02.450.

[2] A. W. Attabi, L. Muflikhah, and M. A. Fauzi, "Penerapan Analisis Sentimen untuk Menilai Suatu Produk pada Twitter Berbahasa Indonesia dengan Metode Naïve Bayes Classifier dan Information Gain," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 11, pp. 4548–4554, 2018.

[3] T. Yulianita, T. W. Utami, and M. Al Haris, "Analisis sentimen dalam penanganan covid-19 di indonesia menggunakan naive bayes classifier," *Semin. Nas. Variansi*, pp. 235–243, 2020.

[4] U. Rofiqoh, R. S. Perdana, and M. A. Fauzi, "Analisis Sentimen Tingkat Kepuasan Pengguna Penyedia Layanan Telekomunikasi Seluler Indonesia Pada Twitter Dengan Metode Support Vector Machine dan Lexion Based Feature," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 1, no. 12, pp. 1725–1732, 2017, [Online]. Available: http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/628.

[5] B. Gunawan, H. S. Pratiwi, and E. E. Pratama, "Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes," *J. Edukasi dan Penelit. Inform.*, vol. 4, no. 2, p. 113, 2018, doi: 10.26418/jp.v4i2.27526.

[6] H. Tuhuteru and A. Iriani, "Analisis Sentimen Perusahaan Listrik Negara Cabang Ambon Menggunakan Metode Support Vector Machine dan Naive Bayes Classifier," *J. Inform. J. Pengemb. IT*, vol. 3, no. 3, pp. 394–401, 2018, doi: 10.30591/jpit.v3i3.977.

[7] M. S. Hadna, P. I. Santosa, and W. W. Winarno, "Studi Literatur Tentang Perbandingan Metode Untuk Proses Analisis Sentimen Di Twitter," *Semin. Nas. Teknol. Inf. dan Komun.*, vol. 2016, no. Sentika, pp. 57–64, 2016, [Online]. Available: https://fti.uajy.ac.id/sentika/publikasi/makalah/2016/95.pdf.

[8] J. A. Septian, T. M. Fahrudin, and A. Nugroho, "Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF - IDF dan K - Nearest Neighbor," *J. Intell. Syst. Comput.*, no. September, pp. 43–49, 2019.

[9] E. S. Negara, R. Andryani, and P. H. Saksono, "Analisis Data Twitter: Ekstraksi dan Analisis Data G eospasial," *J. INKOM*, vol. 10, no. 1, p. 27, 2016, doi: 10.14203/j.inkom.433.

[10] F. A. Suharno and L. Listiyoko, "Aplikasi Berbasis Web dengan Metode Crawling sebagai Cara Pengumpulan Data untuk Mengambil Keputusan," *Semin. Nas. Rekayasa Teknol. Inf.*, no. November, pp. 105–109, 2018.

[11] A. A. Lutfi, A. E. Permanasari, and S. Fauziati, "Corrigendum: Sentiment Analysis in the Sales Review of Indonesian Marketplace by Utilizing Support Vector Machine," *J. Inf. Syst. Eng. Bus. Intell.*, vol. 4, no. 2, p. 169, 2018, doi: 10.20473/jisebi.4.2.169.

[12] P. Antinasari, R. S. Perdana, and M. A. Fauzi, "Analisis Sentimen Tentang Opini Film Pada Dokumen Twitter Berbahasa Indonesia Menggunakan Naive Bayes Dengan Perbaikan Kata Tidak Baku," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 1, no. 12, pp. 1733–1741, 2017, [Online]. Available: http://j-ptiik.ub.ac.id.

[13] P. Arsi and R. Waluyo, "Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM)," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 1, p. 147, 2021, doi: 10.25126/jtiik.0813944.

[14] E. Fitri, "Analisis Sentimen Terhadap Aplikasi Ruangguru Menggunakan Algoritma Naive Bayes, Random Forest Dan Support Vector Machine," *J. Transform.*, vol. 18, no. 1, p. 71, 2020, doi: 10.26623/transformatika.v18i1.2317.

[15] A. Sari, F. V., & Wibowo, "Analisis Sentimen Pelanggan Toko Online Jd. Id Menggunakan Metode Naïve Bayes Classifier Berbasis Konversi Ikon Emosi," *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 2, no. 2, pp. 681–686, 2019.

ISSN 2085-4552

The Influence of KMS Trello on the Intern Performance CDC UIN Jakarta

Jihan Fadhilah[1], Evy Nurmiati[2]
[1, 2] Faculty of Science and Technology, UIN Syarif Hidayatullah Jakarta, Tangerang, Indonesia
[1]jihan.fadhilah6@gmail.com, [2]evy.nurmiati@uinjkt.ac.id

Accepted 06 January 2022
Approved 01 March 2022

***Abstract*— Trello is software used for knowledge and project management. The ease of appearance and integration with Google Drive makes the UIN Jakarta Career Development Center interested in using Trello as a workspace for their intern. The problem of this research stems from the number of interns with non-technological backgrounds and this is the first time using this tool. So the purpose of this study is to see the effect of Trello software on the intern performance and to see the difference in influence between technology and non-technological interns. The research was conducted using a qualitative method, namely conducting interviews with 5 interns from each existing division and also making direct observations of their work through the WhatsApp Group within a period of 2 months.**

***Index Terms*— Career Development Center UIN Jakarta; Intern Performance; KMS; Trello**.

I. INTRODUCTION

In the current era of globalization, competition has begun to fade and is slowly being replaced by collaborating with each other. Collaboration is the action of the parties to produce a "win-to-win" decision so that it can explain and implement the practices of two or more parties with the aim of achieving the common goals that have been agreed at the beginning [1]. The collaboration process will create more innovations and benefits for the collaborating parties.

There is a lot of software that can be used as a collaboration medium at this time and also functions as a Knowledge Management System (KMS), one of the applications that is widely used is Trello [2]. Trello offers ease of use and a user-friendly interface that makes it very easy to use [3]. Trello users are dominated by students and workers because this application is widely used as a work collaboration application between teams so that they can exchange ideas and collaborate with one another.

UIN Jakarta Career Development Center is a self-development and career service unit that accommodates students and alumni of UIN Syarif Hidayatullah Jakarta, as well as the general public to develop themselves and prepare excellent careers for them. UIN Jakarta Career Development Center also helps disseminate information related to job vacancies, internships, as well as training or seminars related to self and career development [4].

Almost all workers who help in UIN Jakarta Career Development Center events and activities are students who are Season I and Season II Interns of UIN Jakarta Career Center. The collaboration software that is used to monitor the performance of the intern and as their workspace is Trello. Interns can add their jobs and documents via the card according to the program being run by the UIN Jakarta Career Center. All of their work can be directly documented into Google Drive if the Trello account is linked to the board owner's Google Drive account.

The application of knowledge management is considered important to improve the performance of employees [5][6]. In addition to knowledge management, human resource management also affects employee performance[7][8]. Some of the benchmarks that researchers use to measure the performance of the UIN Jakarta Career Development Center interns, including personal knowledge, job procedures, and technology literacy [9]. The results of this internal performance report are used as routine evaluation material for the UIN Jakarta Career Development Center in maintaining human resources who are the driving force for all forms of activities and events run by this organization.

KMS (Knowledge Management System) is a technology that is useful for improving employee performance and productivity. The ease with which interns feel when they want to document or ask for help from other colleagues can affect their work performance and productivity. The arrival of the new interns in Season II makes the adaptation of using Trello start at the starting point again.

The problem that is the focus of this research stems from the number of interns who do not come from a technology education background so they are not familiar with these tools and there are even some of them who are using Trello tools for the first time as their collaborative workspace.

On this basis, this study aims to determine how influential the Trello software which is used as a Knowledge Management System on the performance of

the UIN Jakarta Career Development Center interns and to see the difference in the influence of understanding the use of Trello software between technology and non-technological interns so that they can continuous improvement is carried out to increase the value of the organization. In addition, the results of this research can be used as further evaluation material for the UIN Jakarta Career Development Center to maintain Trello as their collaboration workspace and KMS or replace it with other software.

II. LITERATURE REVIEW

A. *Knowledge Management System*

Knowledge management can be interpreted as a scientific discipline that is useful for an organization or company in encouraging the creation, sharing, and development of knowledge in the organization or company [10]. The focus in Knowledge management is on the process, creation, acquisition, refinement, storage, transfer, sharing, and utilization of knowledge [11]. The process of implementing knowledge management requires 3 main factors, namely humans, technology, and processes.

To support optimizing the application of knowledge management, an organization can use a Knowledge Management System (KMS) as a fulfillment of technological factors for the collaboration process in the company to build a sharing habit in the company [12]. KMS is useful for sharing new information and knowledge obtained by employees and company officials, making it easier for employees to learn deeply about the company, and creating with this knowledge new innovations can be created to develop the products offered by the company [13].

B. *Trello Software*

Trello is a website and application-based software. Trello is referred to as a work management tool where teams can plan, collaborate on projects, plan workflows, and track each other's work progress so that it can be used as software to share knowledge together [3]. In Trello, there are many features that will help in project and knowledge management, including a calendar that can be used to create a scheduling schedule.

For knowledge storage, all files in the form of PDF files, documents, videos, and photos can be stored on Trello. Then the file can also be saved to the organization's Google Drive, making it very easy to share knowledge and store data.

C. *Internship Program*

according to[14], the internship program is intended as a program to adjust the existing industry needs with the material that has been studied during college students so that there is a real implementation of the theory that has been given. The internship program is regulated in Indonesian law, namely in Permenaker No. 6 Year 2020 [15], where it is stated that apprenticeship is part of the job training system by working directly under the guidance of competent workers in order to master certain skills or expertise. For some universities, the Internship program is one of the graduation requirements which is usually referred to as PKL (Field Work Training).

D. *Performance assessment*

Performance appraisal is used as a measuring tool that determines whether workers in an organization have provided maximum work results and carry out work according to applicable standards and carry out assessments that measure the strengths and weaknesses of these workers for further improvement [16]. according to [17], there are 5 work indicators, including:

1. Quality of work, related to work results, such as work tidiness, skills, conformity with deadlines, and others.
2. Quantity of work, related to targets such as the ability to carry out instructions, the ability to complete several jobs, and others.
3. Attendance, namely the presence of workers according to working time at events and daily at the office.
4. Cooperation, namely the ability to do work together
5. Initiative, namely the ability to decide something without having to be given a command, and do what should be done well.

III. METHOD

According to [18] Research method is a scientific method used to obtain data, facts, and information needed for certain purposes and uses.

A. *Research methods*

In this study, the researcher used a qualitative approach with a descriptive method. Qualitative approach according to [13] is a research approach method with a theoretical explanation of the data that emphasizes the perspective, definition, and interpretation of the informant with the direction and purpose of leading to an understanding of a problem based on their views.

While the descriptive method is a method that describes, explains, and answers the problem of a phenomenon or the relationship between variables and the phenomenon [19].

The use of the descriptive qualitative method was chosen because the researchers wanted to know more about their opinions from the perspective of each intern from various divisions regarding the Trello software used as a KMS and their workspace during their

internship at the UIN Jakarta Career Center. Researchers also want to know their opinion on how influential this software is on their daily performance so that it can be a material for continuous improvement for the UIN Jakarta Career Development Center in the future regarding this KMS.

B. *Data collection technique*

The researcher used a qualitative approach with several data collection techniques, namely by conducting observations, interviews with several informants, and looking for references related to research topics from several literature studies.

1. *Observation*

In general, observation is making observations on the behavior of others, and noting interesting things that are found [20]. In this study, observations were made by observing the work processes that occurred at the UIN Jakarta Career Development Center Intern. Interaction between Interns is carried out through WhatsApp Groups and observing work progress at Trello for 2 months since October 2021

2. *Interview*

Interview is a data collection technique by asking several questions related to a topic directly to the informant [20]. Interviews were conducted online via WhatsApp calls directly to the specified informants. Interviews were conducted using the In-Depth Interview technique [21] where questions are submitted spontaneously and informally between the researcher and one person.

There were 5 informants who were interviewed regarding the topic of this research. Researchers used purposive sampling method in selecting informants from each division at UIN Jakarta Career Center

3. *Literature review*

Literature study is a data collection technique that is carried out indirectly, but through other media in accordance with the research being researched [22]. To support the research and complete the information, the researchers used additional references from various books, journal articles, websites, and from previous research.

IV. Result and Discussion

A. *Characteristics of Informants*

In qualitative research, people who are used as sources of research data are referred to as informants. Because this study focuses on the performance of the UIN Jakarta Career Development Center Interns, the informants in this study are 5 season I and Season II UIN Jakarta Career Development Center interns, with characteristics as shown in Table 1 below.

The selection of informants in this study used a purposive sampling method, where the researcher selected samples based on needs and also on the consideration that the selected samples could represent certain groups or groups [23]. The five selected informants are representatives from each division at the UIN Jakarta Career Center. In addition, informants were also taken from Season I and Season II interns so that they can see the differences of opinion that will be conveyed by them later.

TABLE I. Information Characteristics

Initials	Profile	Position
SA	Lecturer at UIN Jakarta	Head of UIN Jakarta Career Center
SH	Islamic Communication and Broadcasting, 2018	Administration Intern, Season I
KD	Islamic Banking, 2018	Event Management Intern, Season I
MI	Information Systems, 2019	Data Specialist Intern, Season II
AL	Information Systems, 2019	Social Media Intern, Season I

The researcher met SA as the first informant to ask permission to conduct research because the results of this study can be used as material*continuous improvement*for the UIN Jakarta Career Development Center in the future. SA gave permission for the researcher to conduct interviews with SA and other informants (SH, KD, MI, and AL). The interview process was carried out online using a WhatsApp call on December 10, 2021 at different times. Interviews were conducted semi-formally so that the informants could comfortably convey their answers in detail and depth (depth interview). The interview time for each informant is different, with a time span of 10 – 20 minutes per informant.

B. *Result Category*

Based on the results of interviews and observations made, the researchers selected relevant data between the subject matter and the topic under study. Then several categories were generated related to the application of knowledge management, KMS Trello, and the internal performance of the UIN Jakarta Career Center. After the data is reduced, 3 categories of results are obtained, namely:

1. *Human Resources*

The success of implementing knowledge management cannot be separated from the role of human resources in the organization. The human

resources owned by the UIN Jakarta Career Development Center are almost entirely students of UIN Jakarta who do volunteer internships at the UIN Jakarta Career Center.

The HR selection process is also carried out in stages and through a rigorous selection process. The total number of interns at the UIN Jakarta Career Development Center is 40 people (10 Season I interns and 30 Season II interns). Qualified human resources and the right technology will support the increase in productivity and value of the UIN Jakarta Career Center. The following is an excerpt from an interview with Informant I:

- SA: “Me and the coordinators held a closed workshop with the theme of offline public speaking in October 2021, so that they could have knowledge for themselves and disseminate it to other internal friends who were unable to attend at that time through KMS Trello. I also often share information on webinars for them to participate in in order to add value to themselves and they can apply it at the UIN Jakarta Career Development Center later.”

The coordinator supports HR management and knowledge transfer through workshops and webinars in order to gain knowledge for internal staff. And the documentation is stored in KMS Trello. Other informants described the activities they participated in during their time as interns, including:

- SH: “I really like things related to journalism and was assigned as a publication review team too, so that some time ago an online workshop was held to write, and I attended the workshop.”
- KD: "I often follow the webinars shared by SA's mother because I like to learn new things"
- MI: "Because I just entered as part of the internal Career Development Center of UIN Jakarta, I am very active in participating in all activities to increase my knowledge and I will share it with my friends in my division"
- AL: “I was the first intern here so I gained a lot of knowledge from attending meetings with Mrs. SA, attending webinars, video conference meetings with partners, and of course workshops held by the UIN Jakarta Career Center”

Based on some excerpts from these interviews, it can be concluded that HR management at the UIN Jakarta Career Development Center related to knowledge sharing is carried out through seminars and workshops which are then distributed to other internal friends who do not attend these activities through KMS Trello.

2. Knowledge management

Knowledge management It is closely related to two types of knowledge, namely explicit knowledge (documented knowledge) and tacit knowledge (internal knowledge). Knowledge management at the UIN Jakarta Career Development Center according to several informants is as follows:

- SA: “Knowledge is a valuable thing for us. Moreover, the status of friends here is as an intern for a period of 6 months. So we manage their knowledge very well. We hold monthly reports to share knowledge and gather knowledge in one month then the admin team will save the minutes as documentation”
- SH: "As the admin team, I am in charge of writing meeting minutes as a way to store the knowledge gained during the meeting."
- KD: "The knowledge that I get from various webinars, I usually share directly with event friends so that it can become new insights for the next event, and is immediately greeted by group discussions by my usual friends"
- MI: “Because I am a data team, so we manage the existing knowledge in our database and then it will be visualized in the next monthly report, or in some activities that require visualization of the collected data”
- AL: “I often share the knowledge I get here on my personal social media, or use it as content feeds on the UIN Jakarta Career Development Center Instagram which I dropped first on Trello to get feedback”

Knowledge management for tacit knowledge, it is done by disseminating information between individuals and between divisions and also visualizing the knowledge that has been obtained every month in monthly reports so that knowledge can be distributed to other individuals and divisions. For explicit knowledge that is applied, namely storing the results of the minutes, division monthly reports, and also content posted on Instagram on KMS Trello.

3. Sharing and Storage the Knowledge

Trello is a KMS used by the UIN Jakarta Career Development Center as a tool for sharing and storing the knowledges.

- SA: “We use Trello because it's free and very easy to use. Since Season I, our workspace has been here and we convey the progress of all projects and we save the documents to Google Drive which is connected to Trello”

- SH: “At first I was not very familiar with this tool, but it was very easy for me, especially in terms of administrative document management because it can be directly connected to our Google Drive files”
- KD: “Even though I'm not from the technology department, I've used this tool before. The event team is very helpful because the progress of the project can be seen more clearly”
- MI: “This is the first time I've used this tool, but so far I haven't encountered any difficulties because it's easy to use and I can retrieve data and do the requested tasks quickly because each division has its own card”
- AL: “The social media team is very helpful, especially when requesting a post review or saving the results of posts that have been approved”

So it can be concluded that Trello greatly facilitates the work of internal employees and increases their productivity because it is easy to see the tasks that must be done by each division. The process of storing documents that are directly connected by Google Drive is also very easy in terms of finding the documents needed, thus speeding up their work process.

V. CONCLUSIONS

The following are some conclusions related to the research that has been carried out, including:

1. From the existing problems, it is concluded that the adaptation process can be carried out quickly by the intern with non-technological backgrounds and those who are using Trello KMS as their workspace for the first time.
2. UIN Jakarta Career Development Center supports the performance of its intern employees with KMS Trello so that their performance is considered to be faster, more effective, and efficient because of the easy-to-use Trello display and Trello storage media that really helps their work.
3. There is no significant effect related to the use of Trello for intern with non-technological backgrounds and for intern with technological backgrounds.
4. The results of the study stated that the knowledge management applied was by holding workshops, webinars, discussions and monthly reports for the management of tacit knowledge. While the explicit knowledge is storing the results of the minutes, reports per division during monthly reports, and storing ideas in Trello. So Trello greatly improves the performance of the intern because of its convenience and facilities.

For further research, the researcher suggests that an overall study be conducted using quantitative methods so that a more valid percentage of numbers can be obtained to measure the performance of the intern UIN Jakarta Career Center.

REFERENCES

[1] AB Pranogyo, "Implementation of Knowledge Management to Achieve Vision and Mission at PT PERTAMINA," *J. Accounts. and Manaj.*, vol. 13, no. 01, pp. 1–20, 2017.

[2] EO Choiri, “Trello, The Best Job Collaboration App Today - Qwords,” *qwords.com*, Jan. 2020. https://qwords.com/blog/trello-application-kolaborasi-kerja-terbaik-saat-ini/ (accessed Feb. 25, 2022).

[3] Trello, “About Us: Trello History, Logos & Customers.” https://trello.com/about (accessed Dec. 08, 2021).

[4] UIN Syarif Hidayatullah Career Development Center Jakarta, “News | CAREER DEVELOPMENT CENTER.” https://karier.uinjkt.ac.id/content/view?id=9&t=about-us (accessed Dec. 09, 2021).

[5] S. Syafarina, SH Senen, and S. Sumiyati, "Overview of the Effect of Knowledge Management Implementation on Employee Performance at the Regional Revenue Management Agency of Bandung City," *J. Bus. Manag. Educ.*, vol. 3, no. 3, pp. 123–135, 2018, doi:10.17509/jbme.v3i3.14316.

[6] GFA Bangalino, N. Nurwidianto, and LS Bopeng, “INFLUENCE OF KNOWLEDGE MANAGEMENT, TECHNOLOGY AND WORK DISCIPLINE ON EMPLOYEE PERFORMANCE (Case Study of Manokwari District Environmental Office Employees),” *Manag Horizon. Buses. J.*, vol. 2, no. 1, p. 292, 2019, doi:10.30862/cm-bj.v2i1.20.

[7] Y. Safarni and T. Gadeng, "The Influence of Knowledge and Talent Management on the Performance of Bappeda Employees (Regional Development Planning Agency) in Southwest Aceh Regency," *J. Ilm. manaj. Muhammadiyah Aceh*, vol. 8, no. 1, 2018, doi:10.37598/jimma.v8i1.962.

[8] A. Wise, *The influence of organizational culture and knowledge management on employee performance with employee job satisfaction the intervening variable for office employees…*. repository.um.ac.id, 2019.

[9] R. Anggraini and GM Rahmah, “The Influence of the Knowledge Management System Implementation Process on Employee Performance in the Marketing Division of the Pt. Xyz," *J. Teknol. and Manaj.*, vol. 19, no. 1, pp. 47–54, 2021, doi:10.52330/jtm.v19i1.17.

[10] IT Kapanraja, “The Influence of Knowledge Management Implementation and Human Resource Development on Employee Performance of PT Telkom Tbk. (Explanatory Survey Study on Employees of PT Telkom Tbk's Human Capital Management Unit.) *J. App. manaj.*, vol. 14, no. 1, pp. 140–146, 2016, doi:10.18202/hour23026332.14.1.15.

[11] R. Retnoningsih, "Analysis of Higher Education Knowledge Management System Implementation Using Most Admired Knowledge Enterprise," *Ferris wheel Inform.*, vol. 2, no. 2, pp. 1–10, 2015.

[12] F. Adzima and H. Sjahruddin, *The Effect of Knowledge Management on Employee Performance*. osf.io, 2019.

[13] I. Mutia, - Herlinda, and - Atikah, "Designing a 2013 Curriculum Knowledge Management System (KMS) Using the Tiwana and Zack Model," *J. Education and Researchers. information.*, vol. 3, no. 2, p.

152, 2017, doi:10.26418/jp.v3i2.22686.

[14] H. Hayati, "... IN GOVERNMENT INSTRUCTIONS WITH THE WORK-AT-HOME METHOD (Qualitative Studies as Preparation for Independent Interns to Learn Independent Campus),"*J. Psycho. educator. And …*, vol. 5, no. 1, pp. 47–55, 2021, [Online]. Available: https://ejournal.borobudur.ac.id/index.php/psikologi/article/download/860/799.

[15] Permenaker,*Implementation of Apprenticeships in the Country*. Ministry of Home Affairs, 2020.

[16] Moeheriono, *Key Performance Indicators (KPI): Planning, Application and Development*. PT RajaGrafindo PERsada, 2014.

[17] Suprayogi, *Guidelines for Implementing Internships/Work Practices in the Islamic Economics and Finance Sector in Supporting Independent Learning: Merdeka Campus*, Pert Edition. Jakarta: National Committee for Islamic Economy and Finance (KNEKS), 2021.

[18] Sugiyono, *Research Methods, Quantitative, Qualitative and R&D*. Bandung Bookstore, 2011.

[19] A. Zainal, "Action Research: New Methods and Paradigms," p. 41, 2014, Accessed: Dec. 08, 2021. [Online]. Available: https://scholar.google.com/scholar?hl=id&as_sdt=0%2C5&q=Zainal+Arifin.+2014.+Research+Education+Method+Dan+Paradigm+New.+Bandung%3A+PT+Teenagers+Rosda+Karya.&btnG=.

[20] M. Desita, "Application of Knowledge Management in Improving the Performance of Librarians: A Case Study at the Unika Santo Thomas Library, Medan," 2017.

[21] R. Kawasati, "Qualitative Method Data Collection Techniques."

[22] L. Emalia, "Employee Performance Appraisal Decision Support System Using the Electre–Topsis Method," *J. Computing. Business*, vol. 2013, no. 1, pp. 37–44, 2017, [Online]. Available: http://jurnal.lpkia.ac.id/index.php/jkb/article/view/11.

[23] F. Sudargo, "5 SAMPLING Meeting (Sampling)."

ISSN 2085-4552

The Development of UMS Building Catalogue Information System

Ganno Tribuana Kurniaji[1], Mohammad Faqih Eza 'Ammar[2], Yusuf Sulistyo Nugroho[3]
[1,2,3]Teknik Informatika, Universitas Muhammadiyah Surakarta, Surakarta, Indonesia
[1] l200184092@student.ums.ac.id, [2] l200183178@student.ums.ac.id, [3] yusuf.nugroho@ums.ac.id

Accepted 23 March 2022
Approved 10 May 2022

***Abstract*— Universitas Muhammadiyah Surakarta (UMS) has a large number of new students per year. However, most new students often get confused to identify and even do not know the location of the buildings that handle various activities in the early period of lectures, such as student orientation, English pre-tests, etc. This is due to lack of information provided by the university to the new enrolled students. To overcome this problem, this work aims to develop a web-based building catalogue by implementing a framework, namely Ionic. This system is built to help the new students getting the building-related information in UMS and find out the location of the building. By following the waterfall model to build the system, the results show that we have successfully developed the building catalogue information system. The acceptance testing result shows that the system is acceptable to implement, supported by the SUS score of 76.17. The system is helpful for new students to get information of all buildings in UMS.**

***Index Terms*— building; catalogue; information system; UMS**.

I. INTRODUCTION

Universitas Muhammadiyah Surakarta (UMS) is one of the largest private university in Indonesia. It has 12 faculties of Undergraduate Programs, Graduate Schools, Professional Programs, and Vocational Schools consisting of various disciplines, such as advancements in science, technology and social relations, and arts and culture. To support the business processes, UMS has 5 campuses that are integrated with IT systems, it has a land area of more than 40 hectares, and supported by an Edutorium that covers an area of 6.5 hectares which is considered as the largest convention center in Central Java. The implementation of the lectures in UMS is supported by very adequate facilities such as representative study rooms, complete laboratories, IT-based libraries, and other supporting facilities.

Each year, UMS receives around 8,000 new students that are distributed to 60 study programs. At the beginning of the lecture period, the new students are required to complete various activities such as student orientation activities, English pre-tests, etc. However, The students often get confused and even they only know the name of the buildings without knowing their locations where the activities will be held [1]. This is due to lack of information provided by the university for new students regarding the list of buildings in UMS.

Many studies have been carried out to develop similar applications. A research [2] explains that the virtual tour application provides early information as an earlier experience for students in terms of introducing the area of the new environment. Other study [3] developed a virtual tour as an effective means of introducing new academic environment to the wider community and prospective students. A web-based virtual tour application has also been developed by prior study [4] to display the campus atmosphere and provide information on the location of the building.

Due to the benefit of the application to overcome buildings introduction problems faced by UMS, we developed a similar web-based catalogue building app to help the new enrolled students in identifying the buildings in the university [5]. Catalogue is a medium that is able to provide optimal and easy-to-understand information, contains more detailed and complete information than other media packaged in the form of photographs and attractive layouts [6]. To build the catalogue information system, we implement the Ionic framework [7]. Ionic is a platform to design and build specific applications. It has its own development tools but is very time consuming and costly [8]. On the other hand, Ionic is an independent platform that uses HTML5, CSS and Ionic's JavaScript. The framework is built with AngularJS, which is arguably the most tested and widely used JavaScript framework out there [9].

We developed a web-based building catalogue to help the students in getting the information of UMS buildings [10]. With this building catalogue, the university can easily provide information and locations of the buildings in UMS to the new enrolled students. The design of this application makes it easier for new students to recognize the buildings in UMS.

II. METHODOLOGY

In designing the UMS building catalogue information system, we follow the Software Development Life Cycle (SDLC) approach, specifically the waterfall model [11]. Waterfall is the

oldest and most well-known model in SDLC. A special feature of this model is its sequential steps. It goes down through the phases of requirements analysis, design, coding, testing, and maintenance [12]. This model works well for projects where quality control is a major concern due to its intensive documentation and planning [13]. Each stage in this model is not overlapping the other stages, that means the process in the waterfall model starts and ends in one stage before starting the next stage [14].

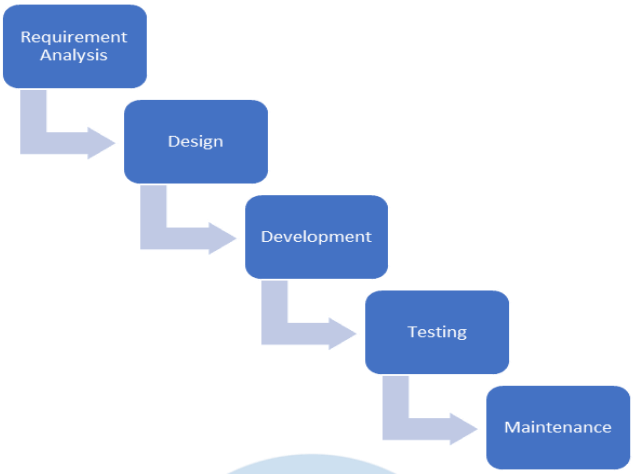

Fig. 1. The process flow of waterfall model in SDLC

Figure 1 shows that there are five stages in the waterfall model that we follow to develop the UMS building catalogue information system [15]. The details of each stage are as follows:

A. *Requirement Analysis*

Requirement analysis is a description of the behavior of the system to be developed [16]. Usually, it provides the information of requirements from clients. Hence, it establishes an agreement between the client and the developer for the specifications and features of the software. In short, requirements are collected, analyzed, and the proper documentation is prepared to help the development process.

The initial discussion was carried out to extract the main problems faced by the clients. The results of the discussion indicate that a web-based building catalogue information system is necessary. In the discussion, we also observe the requirements analysis, which is categorized into two types, namely (1) functional, and (2) non-functional requirements [17].

The functional requirement analysis describes that the building catalogue system is developed online. The system is able to provide detailed and complete information of the building. The catalogue system does not require a "user sign in" to access the information. The users of the system are classified into two types, that is administrator and common users [18]. An administrator is authenticated to have the whole access to the system, so that it can perform the CRUD (Create, Read, Update, and Delete) transaction in the system [19]. Common users do not have rights to update the content in the catalogue. They can only search and read the provided information.

For the non-functional requirements, we analyze the performance and scalability, portability and compatibility, reliability, availability, maintainability, and the usability, as described as follows:

1. *Performance and Scalability*

One of the steps of the text pre-processing stage which serves to convert all letters in the document to lowercase. This step is done to make the search easier.

2. *Portability and Compatibility*

The system should be able to run on a computer device that runs on Windows 7 version or newer, 1.8 GHz or faster processor, 2 GB of ram or higher, and video card that supports a minimum display resolution of 720p.

3. *Reliability, Availability, Maintainability*

There are 75% chances that the component can be fixed in 6 hours for the maintainability of this system. The web dashboard must be available to users 98% of the time every month during business hours.

4. *Localization*

The system is intended for the new enrolled students since they are not familiar with the buildings in UMS. The language of the system is provided in Bahasa Indonesia, so that it is easy to understand for the students.

5. *Usability*

The building catalogue system is simple and easy to use. So that the users can quickly access the catalogue to get such information. In this system, the users do not need to login to access the catalogue system. Although

the main target of the system users is the new enrolled students in UMS, however, all academic staff can also access the catalogue system.

B. Design

The information gathered from the previous phase is evaluated and the proper implementation is formulated. The design describes the key components and the interface of the components. It focuses on how to deliver the required functionality (must-haves and wish-list items) to the system [20].

In this step, we initially design the use case diagram of the system, as shown in Figure 2. Use case diagram is a methodology used in system analysis to identify, clarify, and organise system requirements. Use case diagrams are employed in UML (Unified Modelling Language), a standard notation for the modelling of real-world objects and systems [21]. It shows how a system interacts with external entities. So, it is relatively sparse about the details of how the system behaves internally and how the external environment is arranged.

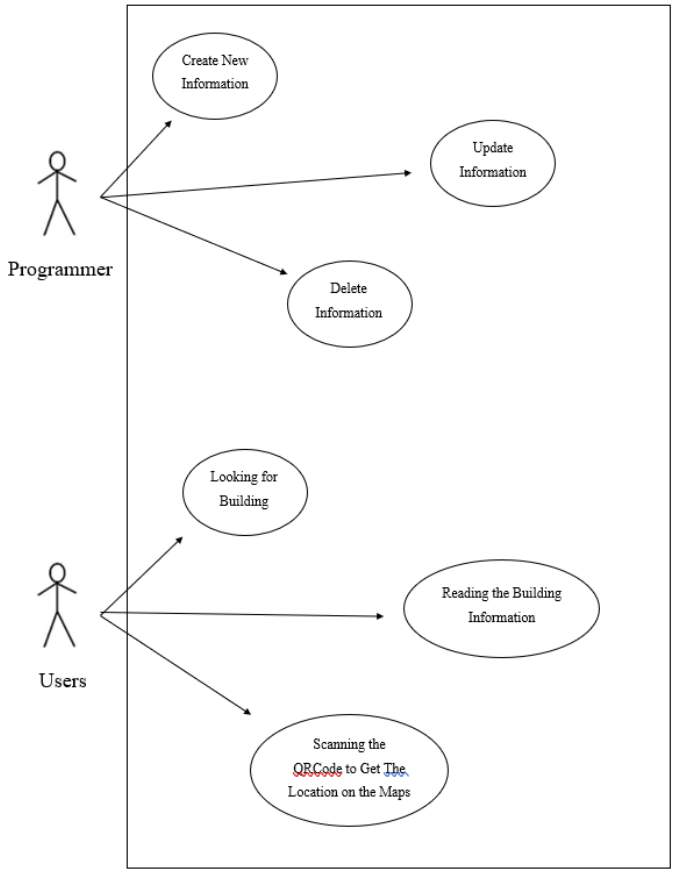

Fig. 2. Use case diagram of building catalogue information system

Figure 2 explains that there are 2 actors who can access the system, namely Administrator and Users. The administrator can change or add new information. On the users side, users can only look for buildings, reading the building information, and directed to the location of the building on the map.

The activity performed by the actors indicated in Figure 2, is then illustrated using an Activity Diagram, as shown in Figure 3. Activity Diagrams are behavioural diagrams which depict the internal behaviour of different operations of a program with the help of nodes and edges [22]. Activity Diagram is used to represent different activities, sub activities, transitions, decision points, guard conditions, concurrent activities, branch, merge, swim lanes, join forks, and etc. An activity diagram starts with one start activity and ends at one final activity.

In addition, this building catalogue system was not developed by implementing a database. This is because the information needed to be displayed on the system is simple and does not require data storage. Therefore, we do not include sequence diagrams and class diagrams in this discussion.

Figure 3 shows that the administrator can create, add, or update the detailed information in the catalogue system which will be sent to the server. The users who access the system via a browser will see the content of the system and be able to select the building based on the provided list. Once the users select one of the building lists, the server will send the detailed information and location of the building to the browser.

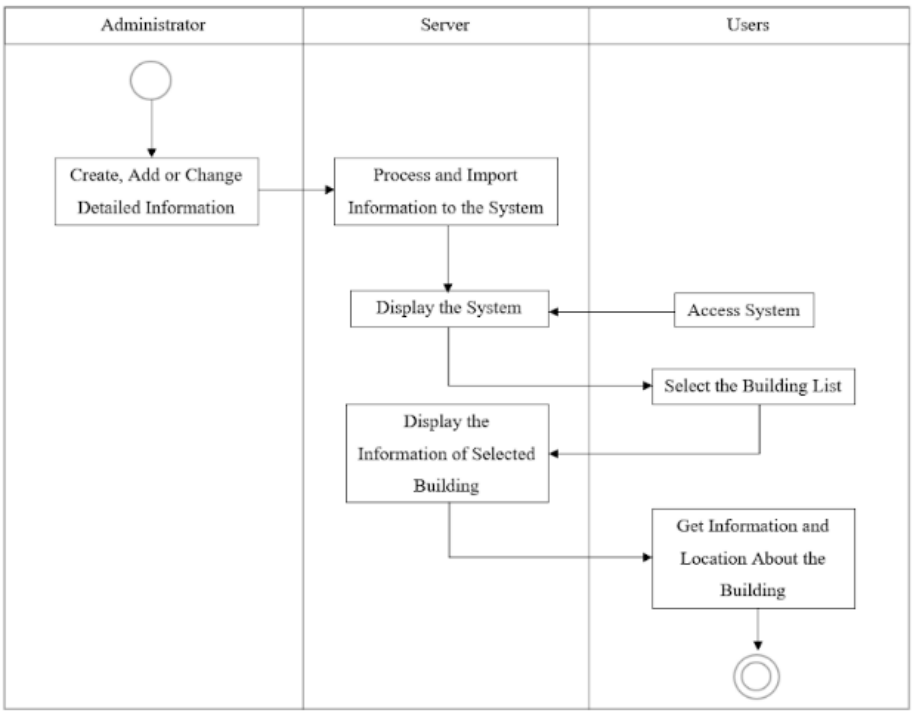

Fig. 3. Activity diagram of building catalogue information system

C. Development

In this phase, the whole requirements will be converted to the production environment. We designed the UMS building catalogue system by applying the Ionic framework. The Ionic framework can be run by installing the Node.js. Node.js is a platform built on top of the Chrome JavaScript runtime to easily build fast and scalable network applications. Node.js uses a non-blocking I/O model that makes it lightweight and efficient, perfect for real-time data-intensive applications running across distributed devices [23]. We are also required to install Ionic CLI using "npm". The Ionic CLI is used to scaffold, develop, and run Ionic applications [24].

D. Testing

This phase deals with the actual testing and inspection of the software solution that has been developed to meet the original requirements. Also, this is the phase where system bugs and glitches are discovered, fixed and refined [25]. In this step, we perform a black-box testing to validate the system and ensure that the system is running as expected. Black-box testing is a functional testing technique that designs test cases based on the information from the specification. Black box testing is not concerned with the internal mechanisms of a system. These are focused solely on the outputs generated in response to selected inputs and execution conditions [26].

Through the black-box testing, we test the building catalogue system by involving one of the representatives from the UMS Information Technology Bureau to ensure that there are no errors in the system and satisfy all the requirements needed [27]. Various checks on the main menu page are performed, starting from checking the image of the building, sorting the name of the building, and checking all features in the systems. On the building menu page, we validate all information listed, starting from the name of the faculty that occupies the building, the list of study programs in the building, information of the student organisation units in the building, and all related information. In addition, we check the general information of the building, such as address, telephone number, and email, as well as the barcode of the building's location.

To evaluate the system acceptance, we apply the testing using System Usability Scale (SUS). SUS is a Likert scale that is simply one based on forced choice questions, where a statement is made and the respondent then indicates a degree of agreement or disagreement with the statement on a 5 (or 7) point scale. The statements with which the respondent expresses agreement and disagreement should be chosen carefully [29].

As shown in Figure 4, it can be seen that the 10 statements in SUS testing actually covers various aspects of system usability, such as the need for support, training, and complexity, and thus has a high degree of facial validity for measuring the usability of a system.

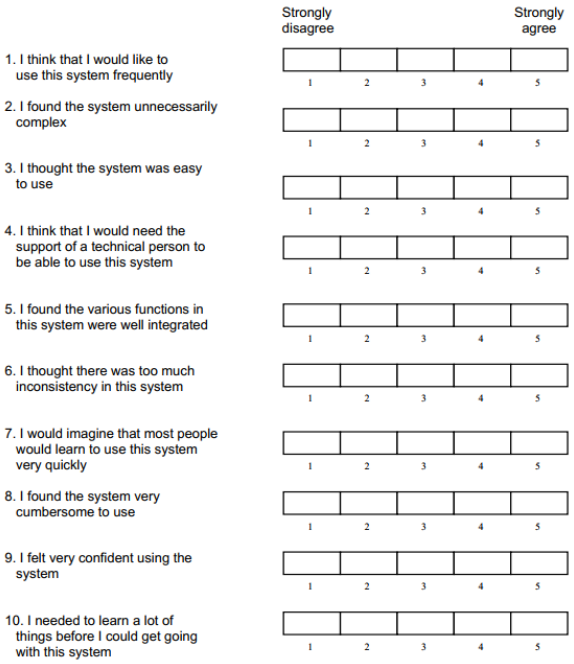

Fig. 4. The System Usability Scale chosen statement

To calculate the SUS score, we first add up the contribution of each item's score. The contribution score of each item will range from 0 to 4. For items 1,3,5,7, and 9 the contribution score is the scale position minus 1. For items 2,4,6,8 and 10 the contribution is 5 minus the scale position in this case we get a total score of 27,420 from 360 respondents. We take data as many as 360 respondents based on the calculation of a representative sample from around 8,000 new students using a survey calculator with the confidence level of 95% and the interval of 5. The calculation of the sample size produces 360 respondents.

E. Maintenance

Maintenance is the process of taking care of such concerns [24]. In this step, we release the building catalogue system. Although it has been implemented, it is possible that the system needs updates, improvements, errors fixing, and refinement depending on the user's requirements in the future.

III. Result and Discussion

The web-based building catalogue system has been successfully implemented. It contains 2 menus, namely Main Menu Page and Building Menu Page. The main menu page functions as the starting page, which contains a list of the names of the buildings accompanied by pictures. on this main menu page, the user first needs to select which building he wants to look for to view detailed information in it. On the building menu page, the function is to display all available information regarding the selected building on the main menu page. On the building menu page there are various kinds of complete information and also accompanied by complementary images and barcodes indicating the location of the building.

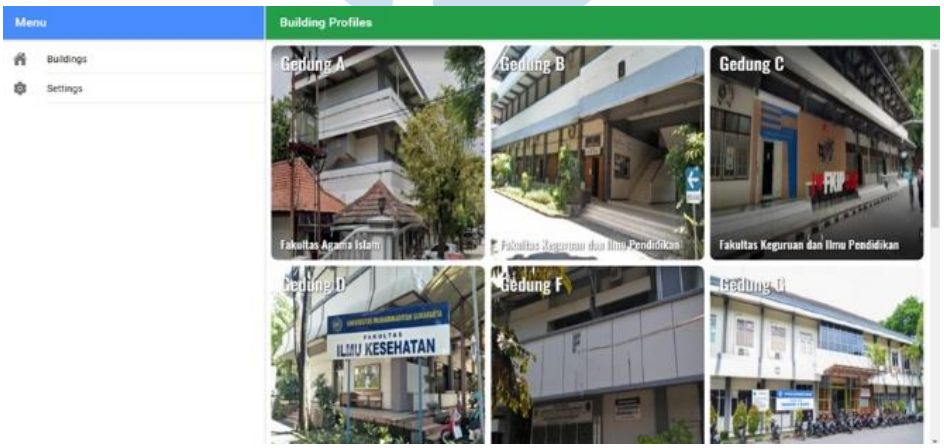

Fig. 5. The main page of the building catalogue information system

In Figure 5, The main page menu displays a list of buildings that are within the scope of UMS, starting from building A to building L. On the main page, there are also pictures of each building. In the bottom left corner of the image, the name of the faculty where the building is located is listed to clarify the description of the building. Each column in the list of buildings on the main page can be clicked to see the more detailed information.

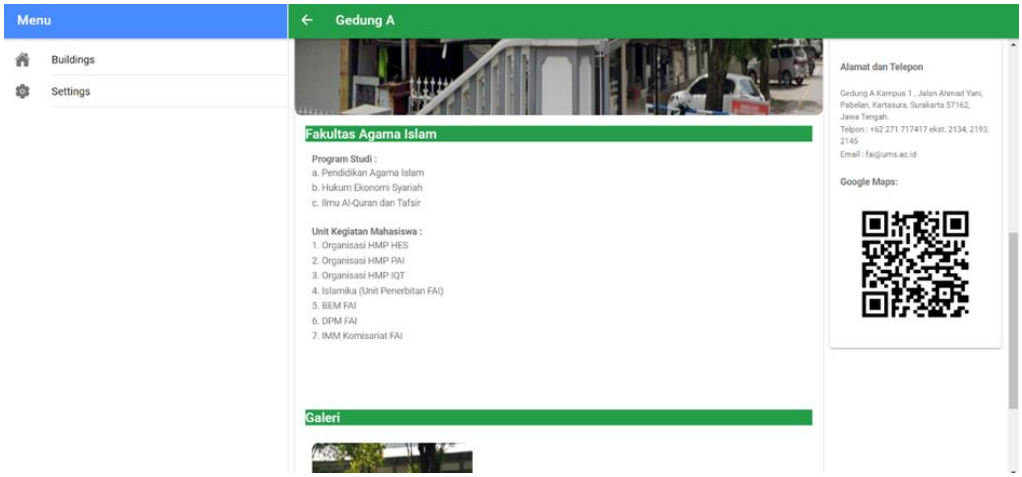

Fig. 6. Detail information of the building menu page

On the building menu, as shown in Figure 6, there is a lot of detailed information about the building such as the photos of the buildings, the study programs located in the buildings, the student organization units located in the buildings, the building addresses, telephone numbers, emails, and barcodes. In general, each building has more than one various study programs. Thus, it is necessary to mention the study programs in the building on the page. It also provides the information of student organization units that occupy the buildings to help the new students be more familiar with the description of each organization.

On the same page, the system displays a gallery of the building photos to show the actual image of the building. We also include detailed information of the building, such as the address where the building is located, the phone number of the faculty, and the email of the faculty. This information can be used for students to contact the associated staff to get further information that is not available in the system. A QR code is added to the page in the system to store additional information of the buildings, such as the location guidance that leads to the building. If the QR code is scanned, it will automatically be redirected to the Google Maps application and will show the exact location.

The building catalogue system is tested using a Blackbox technique [28] to ensure that there are no errors in the system and the System Usability Scale (SUS) to ensure that the system is well developed and easy to use. For the Blackbox technique, we test the system through 8 scenarios, as described in Table 1.

TABLE I. BLACKBOX TEST SCENARIOS

No	Test Scenario	Expected results	Test result
1	Check whether the image of the building matches the name of the building	The image of the building matches the name of the building	Valid
2	Check whether the name of the building is in order	The name of the building is in order	Valid
3	Check if there are a typo in the building name	There are no typo in the building name	Valid
4	Check the name of the Faculty that occupies the building whether is correct or not	The name of the faculty is correct	Valid
5	Check the list of Study Programs in the building whether it is complete and correct	The list of Study Programs in the building is complete and correct	Valid
6	Check the list of Student Organisation Units in the building whether it is complete and correct	The list of Student Organisation Units in the building is complete and correct	Valid
7	Check whether address, telephone number, and email are correct	The address, telephone number, and email are correct	Valid
8	Check the barcode whether it gives correct location to the building or not	The barcode gives correct location to the building	Valid

Table 1 describes that all scenarios in the Blackbox testing are running well. We perform various checks on the main menu page, including the image of the building that matches the name of the building, sorting the name of the building, and checking the spelling of the building names. On the building menu page, we check all the information listed, including the name of the faculty that is occupied in the building, the list of the study programs in the building, information of the student organization units in the building, and the related information.

Furthermore, the implementation of SUS testing produced a total score of 27,420 from all respondents. We then divided the total score by 360 of total respondents to get the overall system usability score. The calculation yields a SUS score of 76.17 which means that the application is acceptable and usable for the new students to get familiar with the new environment in UMS.

IV. CONCLUSION

Based on the results of the study, it can be concluded that we have succeeded in developing a web-based building catalog to assist students in obtaining information about UMS buildings. With this building catalog system, the University can easily provide information and locations about buildings at UMS to new students. Based on the SUS score, the system can be implemented to introduce the buildings in UMS to the new enrolled students in the future. This study also may open an opportunity to develop a mobile-based application in the future.

ACKNOWLEDGMENT

This work has been successfully conducted with the help of the Information Technology Bureau of Universitas Muhammadiyah Surakarta.

REFERENCES

[1] R. B. N. Budi and M. B. Wenas, "Perancangan Virtual Tour Kampus I UKSW sebagai Media Informasi Mahasiswa Baru," *ANDHARUPA J. Desain Komun. Vis. Multimed.*, vol. 4, no. 01, pp. 51–65, 2018.

[2] F. Annafi, "Pengembangan Aplikasi Virtual Tour Berbantuan Video sebagai Media Informasi Wilayah Fakultas Teknik Universitas Negeri Yogyakarta," *Yogyakarta Univ. Negeri Yogyakarta*, 2014.

[3] A. Z. Z. Falani, H. H. A. S. Ramadan, and E. S. Setiawan, "Implementasi Sistem Virtual Tour Berbasis E-Panorama Untuk Pengenalan Kampus Universitas Narotama Surabaya," *Insa. Comtech Inf. Sci. Comput. Technol. J.*, vol. 1, no. 1, 2016.

[4] A. Choiron, "Aplikasi Virtual Tour Dinamis Pada Universitas Dr. Soetomo Surabaya Berbasis Web," *Inf. J. Ilm. Bid. Teknol. Inf. dan Komun.*, vol. 2, no. 1, 2017.

[5] T. Bayer, D. P. Ames, and T. G. Cleveland, "Design and development of a web-based EPANET model catalogue and execution environment," *Ann. GIS*, vol. 27, no. 3, pp. 247–260, 2021.

[6] H. A. REVIL HUDRIANI and M. S. D. Ariusmedi, "Perancangan Katalog Panduan Wisata Seribu Rumah Gadang di Kabupaten Solok Selatan," *DEKAVE J. Desain Komun. Vis.*, vol. 8, no. 3, 2019.

[7] S. Alian, J. Li, and V. Pandey, "A personalized recommendation system to support diabetes self-management for American Indians," *IEEE Access*, vol. 6, pp. 73041–73051, 2018.

[8] S. Yusuf, *Ionic Framework By Example*. Packt Publishing Ltd, 2016.

[9] J. Waranashiwar and M. Ukey, "Ionic Framework with Angular for Hybrid App Development," *Int. J. New Technol. Res.*, vol. 4, no. 5, p. 263068, 2018.

[10] B. Safuan, R. K. Arief, and E. Y. T. Adesta, "Web based e-catalog implementation at TPK-KOJA: Case study of stock inventory division," *Int. J. Adv. Sci. Technol.*, vol. 29, no. 7, pp. 481–488, 2020.

[11] W. W. Royce, "Managing the development of large software systems: concepts and techniques," in *Proceedings of the 9th international conference on Software Engineering*, 1987, pp. 328–338.

[12] S. Balaji and M. S. Murugaiyan, "Waterfall vs. V-Model vs. Agile: A comparative study on SDLC," *Int. J. Inf. Technol. Bus. Manag.*, vol. 2, no. 1, pp. 26–30, 2012.

[13] Y. Bassil, "A simulation model for the waterfall software development life cycle," *arXiv Prepr. arXiv1205.6904*, 2012.

[14] A. Alshamrani and A. Bahattab, "A comparison between three SDLC models waterfall model, spiral model, and Incremental/Iterative model," *Int. J. Comput. Sci. Issues*, vol. 12, no. 1, p. 106, 2015.

[15] W. Hardyanto, A. Purwinarko, F. Sujito, and D. Alighiri, "Applying an MVC framework for the system development life cycle with waterfall model extended," in *Journal of Physics: Conference Series*, 2017, vol. 824, no. 1, p. 12007.

[16] U. Udi, "Penerapan Metode SDLC Waterfall Dalam Pembuatan Sistem Informasi Akademik Berbasis Web Studi Kasus Pondok Pesantren Al-Habib Sholeh Kabupaten Kubu Raya, Kalimantan Barat," *J. Teknol. Dan Manaj. Inform.*, vol. 4, no. 1, 2018.

[17] P. Becker, G. Tebes, D. Peppino, and L. Olsina, "Applying an Improving Strategy that embeds Functional and Non-Functional Requirements Concepts," *J. Comput. Sci. Technol.*, vol. 19, no. 2, pp. e15–e15, 2019.

[18] N. Farooqi, S. Alshehri, S. Nollily, L. Najmi, G. Alqurashi, and A. Alrashedi, "UParking: Developing a smart parking management system using the internet of things," in *2019 Sixth HCT Information Technology Trends (ITT)*, 2019, pp. 214–218.

[19] K. Košťál, P. Helebrandt, M. Belluš, M. Ries, and I. Kotuliak, "Management and monitoring of IoT devices using blockchain," *Sensors*, vol. 19, no. 4, p. 856, 2019.

[20] M. Kramer, "Best practices in systems development lifecycle: An analyses based on the waterfall model," *Rev. Bus. Financ. Stud.*, vol. 9, no. 1, pp. 77–84, 2018.

[21] A. Y. Aleryani, "Comparative study between data flow diagram and use case diagram," *Int. J. Sci. Res. Publ.*, vol. 6, no. 3, pp. 124–126, 2016.

[22] M. Touseef, N. A. Butt, A. Hussain, and A. Nadeem, "Testing from UML design using activity diagram: a comparison of techniques," *Int. J. Comput. Appl.*, vol. 975, p. 8887, 2015.

[23] M. Cantelon, M. Harter, T. J. Holowaychuk, and N. Rajlich, *Node. js in Action*. Manning Greenwich, 2014.

[24] A. Ravulavaru, *Learning ionic*. Packt Publishing Ltd, 2015.

[25] H. Nur, "Penggunaan Metode Waterfall Dalam Rancang Bangun Sistem Informasi Penjualan," *Gener. J.*, vol. 3, no. 1, pp. 1–10, 2019.

[26] S. Nidhra and J. Dondeti, "Black box and white box testing techniques-a literature review," *Int. J. Embed. Syst. Appl.*, vol. 2, no. 2, pp. 29–50, 2012.

[27] I. R. Munthe, B. H. Rambe, R. Pane, D. Irmayani, and M. Nasution, "UML Modeling and Black Box Testing Methods in the School Payment Information System," *J. Mantik*, vol. 4, no. 3, pp. 1634–1640, 2020.

[28] B. Beizer, *Black-box testing: techniques for functional testing of software and systems*. John Wiley & Sons, Inc., 1995.

[29] J. Brooke, "Sus: a "quick and dirty'usability," *Usability Eval. Ind.*, vol. 189, no. 3, 1996.

ISSN 2085-4552

Feature Extraction using Lexicon on the Emotion Recognition Dataset of Indonesian Text

Aprilia Nurkasanah[1], Mardhiya Hayaty[2]*
[1,2] Faculty of Computer Science, Universitas Amikom Yogyakarta, Yogyakarta, Indonesia
[1] aprilia.nurkasanah@students.amikom.ac.id, [2] mardhiya_hayati@amikom.ac.id
*corresponding author

Accepted 06 April 2022
Approved 24 May 2022

***Abstract*— Text Mining is a part of Neural Language Processing (NLP), also known as text analytics. Text mining includes sentiment analysis and emotion analysis which are often used to analyse social media, news, or other media in written form. The emotional breakdown is a level of sentiment analysis that categorises text into negative, neutral, and positive sentiments. Emotion is organised into several classes. This study categorized emotion into anger, fear, happiness, love, and sadness. This study proposed feature extraction using Lexicon and TF-IDF on the emotion recognition dataset of Indonesian texts. InSet Lexicon Dictionary is used as the corpus in performing the feature extraction. Therefore, InSet Lexicon was chosen as the dictionary to perform feature extraction in this study. The results show that InSet Lexicon has poor performance in feature extraction by showing an accuracy of 30%, while TF-IDF is 62%.**

***Index Terms*— Emotion Recognition Text, Feature Extraction, InSet Lexicon, Lexicon, Random Forest**.

I. INTRODUCTION

Communication between individuals is generally done verbally and non-verbally. Verbal communication begins and is obtained through words, sentences, or paragraphs. We often do not catch the emotion implied in the written words, sentences, and sections in verbal communication. As time goes by, the way of communication between individuals has changed. When technology is developing, the exchange of information flows spreads and develops rapidly. Social media is an oral communication tool that is often accessed and has become a trend. In 2021, Indonesia will have active social media users, as much as 61.8% of the existing population[1]. The number increased by 6.3% from the last year. Not only as a trend, but social media has also become a necessity. It can be seen from the statistical data of the social media users that it continuously grows every year. Various video, voice, and text data are added every minute. The flow of data information carried by social media is enormous. Thus, it is necessary to have an accurate processing model so that the information on social media can be processed correctly. Social media has become a part of life in modern society. Through social media, people often express themselves regarding existing changes. Not only do they give objective opinions on an event, but social media users also often express their subjective views. Social media is considered a critical communication medium [2]. Not only opinions or criticisms about an event, but social media users also often criticize products, services, and even government decisions currently hotly discussed. Social media can be used as a medium for monitoring public opinion. This social media user shares their emotions in posts that many people from various platforms watch. Emotion influences humans to express hundreds of words. Therefore, it is considered a crucial thing in communication.

Natural Language Processing (NLP) in computational linguistics is a modern study of linguistics using computer science tools. NLP is also known as a subfield of Artificial Intelligence (AI). [3] NLP extract meaningful information from various texts, including text from social media. Social media has big data containing essential informations. These data give valuable information about a particular phenomenon, government policies, or reviews of using specific products. [4] NLP allows researchers to easily extract beneficial insights in textual datasets while avoiding burdensome computational work. These responses, comments, and reviews are then assessed based on sentiment class, commonly known as sentiment analysis.

Starting from the field of sentiment analysis, which analyzes opinions or messages into negative, positive, and neutral sentiment, emotion recognition is a sub-field of more detailed sentiment analysis that recognizes happy, sad, and angry emotions. Emotional interaction has been a part of a psychological phenomenon in daily life. It can be easily found in everyday interaction and encourages the existence of emotion recognition. Emotion recognition becomes the basis of successful human interaction, communication, and decision-making. The development of big data

influences it to become a significant issue both academically and industrially. [5] Text data produced in emotional communication is often used to understand the human's emotional state. A person's emotional state is often associated with feelings influenced by the individual's interaction with their environment [6]. According to psychology, basic emotions are classified into six classes: joy, sadness, fear, disgust, surprise, and anger [7].

Emotion recognition is currently done through facial expressions and utterances and written text. In applying human-computer interaction, recognizing human emotion helps improve the impression in executing the system. Emotion recognition in text form can be analyzed by several approaches: keyword-based, rule-based, classical learning-based, deep learning-based, and hybrid systems [8]. The methods usually used for emotion classification are Naïve Bayes, J48, K-Nearest Neighbor (KNN), and Support Vector Machine-Sequential Minimal Optimization (SVM-SMO) [9]. [10] compared automated text classification methods and showed that Naïve Bayes and Random Forest provide high accuracy. Naïve Bayes is often used in classifying emotions based on conditional probabilities to organize data in predefined classes [11].

However, this research has not achieved maximum accuracy because there are sentences that do not represent the actual emotions; for example, a sentence belongs to the fear class but is indicated to be in the happy class [11]. Machine learning is a method for allowing machines to learn from empirical data that experts have confirmed. Although the machine learning approach is used in many studies, our research takes a new approach to the problem by using the lexical approach to overcome data processing. Lexicon is also evaluated as a feature extraction tool in this study. As a result, this study utilizes a keyword-based approach to the Lexicon method. [12] Lexicon proved capable of providing high accuracy in sentiment analysis. This method is expected to perform feature extraction in emotion recognition in Indonesian texts more accurately.

II. Related Work

Many studies on emotion mining have been done, including in Indonesian text. The [13] study included sentiment analysis and emotion recognition of readers. With the available approaches and features, texts from various sources such as news and social media posts can be used as the dataset in the study of emotion mining [14], [15].

The researchers of Affective Computing proposed many rule-based approaches for extracting the text emotion automatically. [16] combined Lexicon-based, Bag-of-Words, Words embedding, orthography, and Part-of-Speech (POS) tag, and got an F1-Score accuracy of 69.7%. The research classifies emotions into 5 types: anger, fear, happiness, love, and sadness. Another model classifies emotions into six basic specifications of emotion: anger, disgust, fear, joy, sadness, and surprise [17]. Research [18] made a model for classifying emotions into five classes using an Indonesian language dataset and showed that the Maximum Entropy (ME) algorithm has better accuracy than SVM by 72%. Another feature extraction used in text extraction is TF-IDF [19]. TF-IDF works by extracting keywords based on the frequency of words that appear. In research [20], emotions were classified into 6 classes and showed that TF-IDF could provide high accuracy in text emotion analysis of the text. [21] showed that TF-IDF is a feature extraction based on a weighting method that can outperform feature extraction with N-Gram.

Research [22] combined the Random Forest Classifier with Unigram and SentiWordNet to produce the highest accuracy on sentiment analysis datasets in Malayam. [23] combined Lemmatization, TF-IDF, and Random Forest and showed a good result in speech emotion recognition models. The research used a Logistic Regression algorithm, Support Vector Machine, and Random Forest to calculate the accuracy of feature extraction. [24] used Random Forest on sentiment analysis with a deep learning approach. [25] The Random Forest classifier is often used in predictive modelling to reduce the number of variables required so that the burden of data collection is reduced and efficiency is increased. Research [26] shows that Random Forest worked based on various predictor variables without any assumption about the response variable. [27] Random Forest integrates trees, where a classification is made from several decision trees and produces an output based on the number of votes from all the tree outputs. [28] used domain-specific emotion lexicons (DSLs) and general-purpose emotion lexicons (GPELs) to study feature extraction emotional problems. [29] used the Lexicon approach to generate emotional weights or values.

Before a specified model recognizes the emotion in a text, the text must have a particular stage so the machine can understand the series of words or sentences. The stage is called pre-processing. [30] Pre-processing is a stage of converting unstructured data into structured data based on the needs. In extracting Lexicon features, research [31] carried out several pre-processing stages: case-folding, punctuation, conversation word, stopword removal, stemming, and tokenization. [32] Text pre-processing methods improve the predictive accuracy of the generated models for sentiment classification. Research [33] used NRC Affect Intensity Lexicon and SentiStrength techniques to extract and analyze the characteristics of Twitter's users towards sentiment and behaviors signaling the "suicide" sign. [34] showed that the Lexicon-based approach worked well for sentiment analysis. The Semantic Orientation Calculator (SO-CAL) assigns a positive or negative label to a particular

text, capturing the text's opinion toward the main topic. [35] detected Lexicon emotion by relying on the relationship between words and emotions in WordNet. Research [36] showed that Lexicon feature extraction is more commonly found in handling medical sentiment than SentiWordNet (SWN) in the drug review dataset. [37] showed that Lexicon significantly surpassed the BoW feature in emotion classification. [38] used Lexicon to perform calculations by representing each dataset with a binary vector.

The use of Lexicon in the classification of emotions in English datasets has been widely applied. However, the use of Lexicon in emotion analysis in Indonesian text is still rarely used. Another study [39] of emotion analysis used the Term-Weighting Scheme in the approach. In this study, the InSet Lexicon dictionary [40], which contains a weighting of 3,609 positive words and 6,609 negative words in Indonesian, is used. The value weights in the InSet Lexicon are obtained from manual weighting by Indonesian language experts with weights between -5 to +5 as done by Affected Lexicon (AFFIN) [41]. InSet Lexicon is proved to be the best among translated SentiWordNet, translated Liu Lexicon, translated AFINN Lexicon, and Vania Lexicon in analyzing the sentiment for Indonesian. This dictionary is based on Twitter data, so it contains common words and non-standard words. It uses Lexicon to perform feature extraction from the emotion recognition dataset.

III. METHODOLOGY

A. *Dataset*

This dataset is taken from [16] applying the collection method by using the Twitter Streaming API for two weeks starting from June 1 2018, to June 14, 2018, and setting the geolocation filters in Indonesia. The datasets contains 4,403 Indonesian-language tweets with a 0.917 score of annotation, which has 5 (five) emotion classes, namely love, anger, sadness, joy, and fear. Fig 1. shows the balanced number of emotions in the emotions of anger, joy, and sadness. However, emotions with the categories of love and fear have a limited number of tweets.

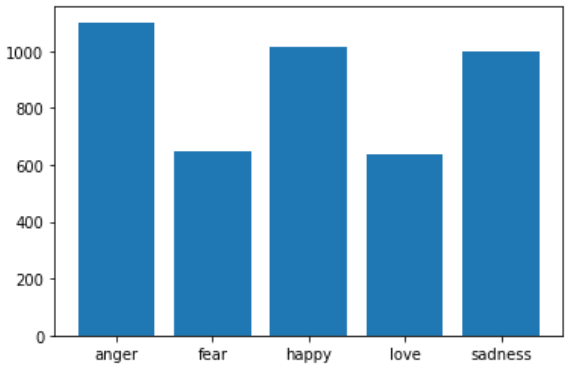

Fig. 1. Class Distribution on Dataset

B. *Preprocessing*

The pre-processing in this research is divided into six stages as follows:

- Case-folding: changing all letters to lowercase.
- Punctuation: deleting unnecessary symbols or characters.
- Convert Word: converting abbreviated words into standard language using the dictionary of abbreviations provided in the emotion dataset [16].
- Stopwords: removing words that are not needed in the used dataset.
- Stemming: finding the root word of a word that has an affix. The Indonesian stemming library, namely Sastrawi, is used in this process. In the Indonesian dataset, stemming is an important step. It is because Indonesian has many affixes. They are suffixes (addition at the end of a word), prefixes (addition at the beginning of a comment), and confixes (addition at the beginning and the end of a word).
- Tokenization: the last step of pre-processing is breaking sentences into text based on delimiters (spaces).

TABLE I. RESULT OF PREPROCESSING DATASET

Label	Text Clean	Text Processed
Anger	soal jalan jatibarupolisi tidak bisa gertak gubernur	['jalan', 'jatibarupolisi', 'gertak', 'gubernur']
Fear	pak ketua di temani sekretaris tapi selaku penanggung jawab	['ketua', 'teman', 'sekretaris', 'tanggung']
Happy	kepingin gudeg mbarek bu hj amad	['kepingin', 'gudeg', 'mbarek', 'bu', 'hj', 'amad']
Love	pasangan seringkali diibaratkan sebagai rumah karena rasa nyaman	['pasang', 'seringkali', 'rumah', 'nyaman']
Sadness	melihatmu sedih kebahagianku pudarseperti saat kala itu	['lihat', 'sedih', 'kebahagianku', 'pudarseperti']

Table 1 shows the example of the emotion recognition dataset of Indonesian text that passed the pre-processing phase.

C. *Feature Extraction*

Feature extraction is a part of feature engineering. Feature Engineering is divided into two, feature extraction and feature selection. After the pre-processing phase, the next step is feature extraction. Feature Extraction pulls out words from text data to be converted into features used by the classifier [42]. Feature Extraction helps get the best features by combining variables into components to reduce the

amount of data [19]. Feature extraction is the data transformation process in which raw data is changed into numerical features. These numerical features are processed by maintaining the information from the raw data set. Feature extraction is believed to give better results than using machine learning directly toward the raw data. It has a significant impact on the calculation results of the classifier. Through feature extraction, data is processed so that it can be read and used as input by the specified classifier. Fig. 2 shows the workflow of the Lexicon feature extraction used in this study.

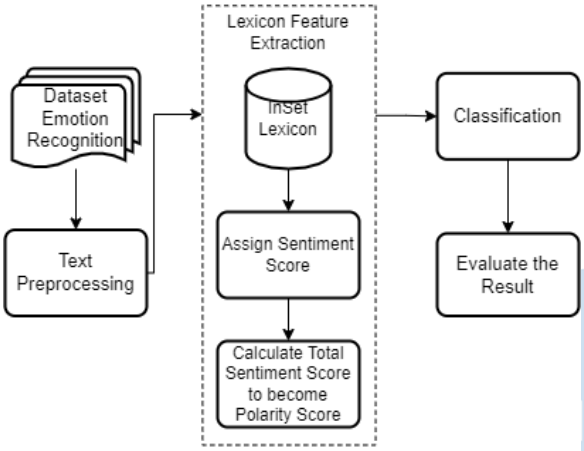

Fig. 2. Flowchart Diagram

The Lexicon-based approach relies on the emotions in the dictionary. This study uses the InSet Lexicon [37], an Indonesian sentiment dictionary accompanied by a weighting for each word to determine the polarity score. This dictionary is divided into two categories positive and negative. The following is the distribution of the Lexicon InSet dataset based on its type:

TABLE II. INSET LEXICON POSITIVE

Words	Weight
hai	3
merekam	2
ekstensif	3
paripurna	1
detail	2
pernik	3
belas	2
welas	4
kabung	1
rahayu	4

TABLE III. INSET LEXICON NEGATIVE

Words	Weight
putus tali gantung	-2
gelebah	-2
gobar hati	-2
tersentuh (perasaan)	-1
isak	-5
larat hati	-3
nelangsa	-3
remuk redam	-5
tidak segan	-2
gemar	-1

The dataset that passes the pre-processing stage is assigned a polarity score. This step analyzes sentiment words in the emotion recognition dataset of Indonesian texts and then determines the polarity score based on Lexicon [43].

TABLE IV. RESULT OF ASSIGN POLARITY SCORE

Assign Polarity Score				
Sentiment word	=	semangat	puasa	turun
Polarity Score	=	4+(-1)	0+(-2)	1+(-3)

From Table IV, it is known that the words 'semangat', 'puasa' and 'turun' are words that exist in the emotion recognition dataset of Indonesian texts. Among these three words, 'semangat' and 'turun' are categorized as positive and negative sentiments. At the same time, the word 'puasa' belongs to negative sentiments.

The sentiment score in the InSet Lexicon is the value used to determine the sentiment of the existing dataset. In this study, the polarity score was obtained from the sum of the sentiment scores in each row of the emotion recognition dataset of Indonesian texts.

$$Polarity\ Score = \sum_{i=1}^{n} Overall\ Sentiment\ Score(i) \quad (1)$$

In this step, after the polarity value is determined, the next step is labeling or categorizing the sentiment class of each data. The sentiments are classified according to the following rules:

$$\begin{aligned} Polarity\ Score > 0 &\rightarrow Positive \\ Polarity\ Score = 0 &\rightarrow Neutral \\ Polarity\ Score < 0 &\rightarrow Negative \end{aligned} \quad (2)$$

The sum of the polarity score can be seen in Table V.

TABLE V. RESULT OF POLARITY SCORE

Assign Polarity Score				
Sentiment word	=	semangat	puasa	turun
Polarity Score	=	3	-2	-2

From the calculation in Table V above, it can be seen that the emotion recognition data in Indonesian texts have negative sentiment or polarity with a score of = -1. This step will continue until the last row of data of the existing dataset. Fig. 3 displays the results of each data's polarity and polarity values.

	label	text_clean	text_preprocessed	polarity_score	polarity
0	anger	soal jalan jatibarupolisi tidak bisa gertak g...	['jalan', 'jatibarupolisi', 'gertak', 'gubernu...	-3	negative
1	anger	sesama cewek lho kayaknya harusnya bisa lebih...	['cewek', 'lho', 'kayak', 'rasain', 'sibuk', '...	-34	negative
2	happy	kepingin gudeg mbarek bu hj amad foto dari goo...	['kepingin', 'gudeg', 'mbarek', 'bu', 'hj', 'a...	0	neutral
3	anger	jalan jatibarubagian dari wilayah tn abangpen...	['jalan', 'jatibarubagian', 'wilayah', 'tn', '...	0	neutral
4	happy	sharing pengalaman saja kemarin jam batalin t...	['sharing', 'alam', 'kemarin', 'jam', 'batalin...	2	positive

Fig. 3. Polarity Score

The existing polarity values are divided into negative, neutral, and positive sentiments. Fig 4 shows the sentiment percentage in the dataset after calculating the polarity score based on the InSet Lexicon.

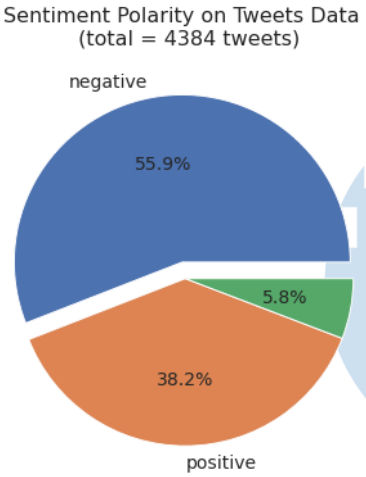

Fig. 4. Polarity Percentage

The following is the polarity value based on the emotion label with positive polarity in the dataset.

	text_clean	polarity_score	label
1	terimakasih papap terimakasih selalu ada terimakasih selalu siaga terimakasih selalu mencukupi terimakasih selalu memaklumi terimakasih selalu bersabar terimakasih selalu	36	happy
2	kegiatan ibadah sholat dzuhur uils koja penerima manfaat bersama petugas pendamping melakukan ibadah sholat dzuhur berjamaah indahnya kebersamaan url	31	happy
3	username bismillah happy bday kak wuatb gbu semoga selalu dilimpahi keberkahan rezeki kebahagiaan umur yang bermanfaat dan sukses dunia akhirat semoga citanya tercapai dan bisa membanggakan orangtua nusa negara dan agama amin btw kita sedom kak	30	happy
4	congratulation selamat ya adik cewek ke atas kelahiran putrinya yang ke saya sebagai kakak turut suka cita bersyukur senang sudah diberikan keselamatan allah swt dalam prosesing persalinan putri ke nya semalam kakak berdoa semoga putri ke nya menjadi putri yang shalihah url	30	happy
5	untukumu wanita yang selalu saya fikrikan bahagialah dengan bertambahnya usia mu sayang semoga kamu sehat semoga terjaga iman dan islammujaga bahagiamu kekasihku maka saya akan merasa tenang karena mencintaimu dalam diamsekali lagi selamat bertambah usia saya mencintaimu sampai akhir hdpku	30	love

Fig. 5. Positive Polarity

The following is the polarity value based on the emotion label with negative polarity in the dataset.

	text_clean	polarity_score	label
1	wartawan di bunuh guru di bunuh aktivis di bunuh sesorang di bunuh belasan orang di bunuh puluhan orang di bunuh ratusan orang di bunuh ribuan orang di bunuh puluan ribu orang di bunuh ratusan ribu orang di bunuh jutaan orang di bunuh	-52	anger
2	seandainya jokowi kalah sudah dipastikan hati bangsat bangkit teroris bangsat bangkit radikal bangsat bangkit semua politik balas dendam bangkit yang minoritas ditekan mending memisahkan diri dari nkri untuk itu harus tetapjokowi	-46	anger
3	namanya orangtua ya memang wajib kan bayarin anaknya sekolah jadi tidak ada itu yang namanya buang duit semua yang sudah dilakuin sama kita baik benar apa tidak itu gabakal sia sia tidak usah takut gagal tidak usah mikir buruk justru yang sia sia itu dengan kalian ngejudge orglain kayak gini	-41	anger
4	jujur saja cukup yah kegoblokan kegilaan saya harus sadar saya ingin segera keluar dari lingkaran setan hidup disini banyak saudara dengki n jahat sama saya tidak kuat saya n terutama sepupu sialan cln penghuni neraka si lina n lakinya n si lita juga najiiiis kamu sialan	-40	anger
5	itu namanya fitnah mbak tidak ada bukti ko fitnah orang lain korupsi ingat mbk dunia ini berlaku sistem tabur tuai mgkin anda menutupi keburukan anda dengan memfitnah orang lain saya tahu semua kok korupsi n pekerjaan terselubung anda	-39	anger

Fig. 6. Negative Polarity

Besides using Lexicon-based feature extraction, the researcher also uses TF-IDF to compare the accuracy. [44] TF-IDF Term Weighing, that is often used, is a result of an integration between Term Frequency and Inverse Document Frequency. This weighing assumes that infrequently appearing terms hold the highest importance. TF-IDF can be calculated using the formula as follows:

$$TF \cdot IDF(t_i, d_j) = TF(t_i, d_j) \times \log \frac{D}{d(t_i)} \quad (3)$$

Based on equation (3), Where is TF of the term t_i in document d_j, while $\frac{D}{d(t_i)}$ is IDF of term t_i.

D. Feature Extraction

The experiment in this study applies a programming language, namely Phyton 3.7.13, and Graphics Processing Units (GPU) based cloud service from Google called Google Collaboratory, also known as Collab. We utilise several libraries such as Pandas, Matplotlip, Sastrawi, and Sklearn. The investigation was started by importing the libraries used into the Google Collaboratory. Then, the pre-processing was done to process the dataset used in the study.

After the dataset was processed by adding the mass or value of each tweet, it was used as the input for the classifications of Machine Learning. This study uses 75% *data training* and 25% *data testing*. The researcher develops a supervised learning approach to analyse and detect emotion in a tweet text and classify them automatically. A Random Forest algorithm is applied to organise the data from the values generated by the extraction feature into emotional labels for each text. However, it is vital to highlight the focus of this research is the discussion of the performance of applying InSet-lexicon than the performance of the classification feature algorithm. There is no additional hyperparameter aiming to make a difference or improve the performance of the Random Forest algorithm in the classification.

Accuracy, precision, recall, and F-measure are applied to measure the performance with a confusion matrix. It can be seen in the following formula below:

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (4)$$

$$Precision = \frac{TP\ or\ TN}{TP+FN\ or\ TN+FP} \quad (5)$$

$$Recall = \frac{TP\ or\ TN}{TP+FP\ or\ TN+FN} \quad (6)$$

$$F1-Score = 2 \times \frac{(Precision \times Recall)}{(Precision+Recall)} \quad (4)$$

Explanation:

- TP (True Positive): Number of the sample that is correctly labelled positive.
- FP (False Positive): Number of the negative samples that are incorrectly labelled as positive.
- FN (False Negative): Number of the positive samples that are incorrectly labelled as negative.
- TN (True Negative): Number of the sample that is correctly labelled negative.

The accuracy is obtained by comparing the prediction ratio with the total of the existing data. Precision is a comparison between True Positive (TP) with the amount of data that is predicted to be positive. Meanwhile, Recall compares True Positive (TP) with the number of data that belongs to the positive class. In the F1 score, precision and recall are combined into a single metric. In short, the F1 score is a weighted average between precision and recall. The best value of the F1 score is indicated at one while the worst value is shown at 0.

IV. Result and Discussion

This research aims to test the Lexicon-based feature extraction in the form of a dictionary using the InSet Lexicon dictionary for the Indonesian language emotion recognition dataset. Feature extraction allows the simplification of the classification of the text data. The feature extraction process eliminates text dimensionality by removing unrelated features from the text data. InSet Lexicon itself provides high accuracy in analyzing sentiment datasets in Indonesian [12].

TABLE VI. Experiment Result On Random Forest Classifier

Classifier	Accuracy
InSet Lexicon	30%
TF-IDF	62%

The tests carried out in this analysis used 1000 tweet data from Twitter with the keyword 'indihome' in Indonesian. This test uses 2 classes, namely positive and negative classes. The data collection is carried out using the Crawling technique that utilizes the tweepy library from the Python programming language. After that, labeling positive and negative sentiments on the dataset is done manually. For manual labeling of positive and negative sentiments, there should be the help of linguists in determining the positive or negative sentiments of an opinion. The author only uses two classes of sentiment, namely positive and negative because the author wants a more conical conclusion between the two classes.

Table VI compares the accuracy between InSet Lexicon and TF-IDF in the study. The TF-IDF feature extraction gets better accuracy results than Lexicon using the Lexicon InSet dictionary. The polarity value obtained by adding each word in the emotion recognition dataset based on the InSet Lexicon is not influenced by the emotion labels in the dataset.

However, the results of the Lexicon InSet in this study were higher than the study [16] that has the same dataset. The previous study provided an accuracy of 24.92% in the Random Forest classification. Meanwhile, the accuracy value in this study increased by 5.08% from the prior research. The increase in accuracy value is influenced by the pre-processing. In the previous research, the pre-processing stage that was carried out was data normalization, which includes changing the letters in the dataset into lowercase form, deleting usernames and hyperlinks, and doing stopwords. After performing data normalization, Part-of-Speech (POS) Tagging was completed and continued with stemming. Meanwhile, in this study, the pre-processing is case folding, punctuation (removing symbols, characters, links, and words that are not needed), converting words (changing abbreviated words into common words), stopwords, stemming, and tokenization.

TABLE VII. Evaluation Of Each Emotion Class In The Dataset

Emotion Class	Feature Extraction	
	InSet	*TF-IDF*
Anger	38%	63%
Sadness	8%	49%
Happy	40%	60%
Fear	14%	76%
Love	23%	73%

This experiment used the F1-Score to determine the calculation results on the metric accuracy of each existing emotion class because the emotion recognition dataset of Indonesian texts had imbalanced data. The happy and anger emotion classes have the highest F1-Score value than other emotion classes in feature extraction using InSet Lexicon. The emotion class "Sadness" shows the lowest F1 score of the other emotion classes, 8%.

The emotion class "Sadness", with an 8% score, provides the lowest F1-Score value among the other emotion classes from Lexicon feature extraction. Meanwhile, this feature extraction also shows that the emotion classes "Anger" and "Fear" provide higher

accuracy than other classes. Fig 7 shows the confusion matrix of the Lexicon feature extraction in the Random Forest classifier.

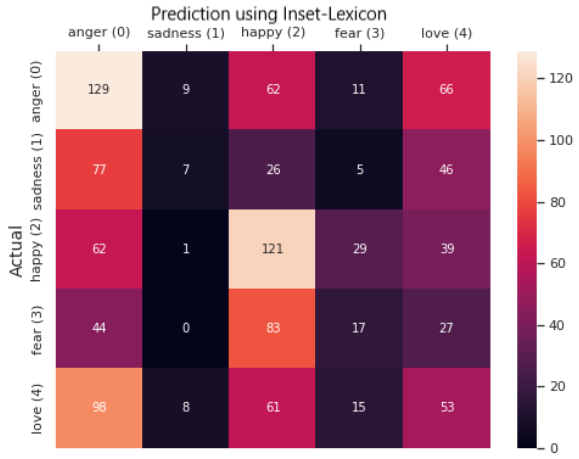

Fig. 7. Confusion Matrix of InSet Lexicon using Random Forest

The emotion class "Sadness" has a small proportion of false-positive values with moderate false negatives, which causes the "Sadness" class to have a small recall value even though it reaches a high precession value. Fig 8 shows the TF-IDF feature extraction confusion matrix in the Random Forest classifier.

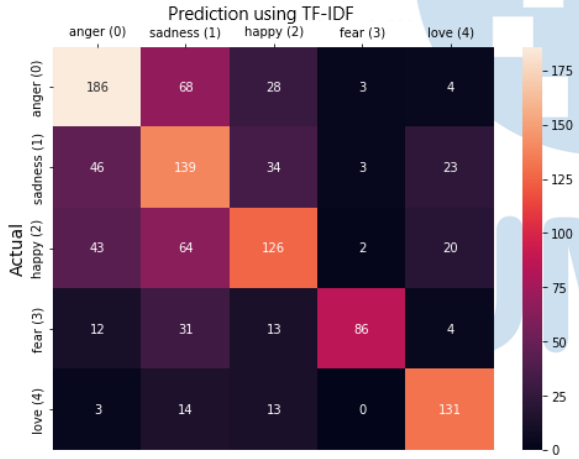

Fig. 8. Confusion Matrix of TF-IDF using Random Forest

Feature Extraction with TF-IDF provides better accuracy results on InSet Lexicon. Each emotion class can be predicted with a high F1-Score. The same with the Random Forest classification in InSet Lexicon feature extraction, the emotion class "Sadness", with a 49% score, has the lowest F1-Score value among other emotion classes.

V. CONCLUSIONS

This study examines the Lexicon based feature extraction on the emotion recognition dataset of Indonesian texts in the form of a corpus or dictionary called InSet Lexicon. The results in this study show an accuracy of 30%. The accuracy of this result is higher than the previous study that had the same dataset and classification. The difference in accuracy is influenced by the pre-processing stages carried out in both studies. However, it has less accuracy than feature extraction using TF-IDF which has an accuracy of 62% on the Random Forest classifier. In this study, each emotion class (Anger, Sadness, Happy, Fear, and Love) can be detected with the used classifier. The low accuracy value in the InSet lexicon is caused by the polarity value that is not influenced by the emotion labels in the dataset. Lexicon feature extraction using the Lexicon InSet, which usually provides high accuracy when used to analyze Indonesian sentiment, produces low accuracy in the emotion recognition dataset of Indonesian texts.

We will examine and improve the current results for further research to achieve better performance. The researcher also suggests future studies to investigate the effect of class imbalance on the dataset for each text weighting scheme. The imbalance plays a significant role in creating a presupposition toward selecting the majority class in the emotion recognition dataset of Indonesian texts.

REFERENCES

[1] S. Kemp, "DIGITAL 2021: THE LATEST INSIGHTS INTO THE 'STATE OF DIGITAL,'" *We Are Social*, 2021.

[2] K. M. Carley, M. Malik, M. Kowalchuck, J. Pfeffer, and P. Landwehr, "Twitter Usage in Indonesia," *SSRN Electron. J.*, 2018, doi: 10.2139/ssrn.2720332.

[3] Y. Kang, Z. Cai, C. W. Tan, Q. Huang, and H. Liu, "Natural language processing (NLP) in management research: A literature review," *J. Manag. Anal.*, vol. 7, no. 2, pp. 139–172, 2020, doi: 10.1080/23270012.2020.1756939.

[4] K. Li, C. Zhou, X. (Robert) Luo, J. Benitez, and Q. Liao, "Impact of information timeliness and richness on public engagement on social media during COVID-19 pandemic: An empirical investigation based on NLP and machine learning," *Decis. Support Syst.*, no. June 2021, p. 113752, 2022, doi: 10.1016/j.dss.2022.113752.

[5] J. Deng and F. Ren, "A Survey of Textual Emotion Recognition and Its Challenges," *IEEE Trans. Affect. Comput.*, vol. 3045, no. c, pp. 1–20, 2021, doi: 10.1109/TAFFC.2021.3053275.

[6] F. A. Acheampong, C. Wenyu, and H. Nunoo-Mensah, "Text-based emotion detection: Advances, challenges, and opportunities," *Eng. Reports*, vol. 2, no. 7, pp. 1–24, 2020, doi: 10.1002/eng2.12189.

[7] T. M. Li, H. C. Chao, and J. Zhang, "Emotion classification based on brain wave: a survey," *Human-centric Comput. Inf. Sci.*, vol. 9, no. 1, 2019, doi: 10.1186/s13673-019-0201-x.

[8] N. Alswaidan and M. E. B. Menai, "A survey of state-of-the-art approaches for emotion recognition in text," *Knowledge and Information Systems (2020) 62*, 2020.

[9] Muljono, A. S. Winarsih, and C. Supriyanto, "Evaluation of classification methods for Indonesian text emotion detection," *Proc. - 2016 Int. Semin. Appl. Technol. Inf. Commun. ISEMANTIC 2016*, pp. 130–133, 2016, doi: 10.1109/ISEMANTIC.2016.7873824.

[10] J. Hartmann, J. Huppertz, C. Schamp, and M. Heitmann, "Comparing automated text classification methods," *Int. J. Res. Mark.*, vol. 36, no. 1, pp. 20–38, 2019, doi: 10.1016/j.ijresmar.2018.09.009.

[11] R. C. W. Fera Fanesya Indriati, "Deteksi Emosi pada Twitter Menggunakan Metode Naive Bayes dan Kombinasi Fitur," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 7, p. 3,

2019.

[12] Y. Fauziah, B. Yuwono, and A. S. Aribowo, "Lexicon Based Sentiment Analysis in Indonesia Languages : A Systematic Literature Review," *RSF Conf. Ser. Eng. Technol.*, vol. 1, no. 1, pp. 363–367, 2021, doi: 10.31098/cset.v1i1.397.

[13] J. Bata, "# AkuGalau : Korpus Bahasa Indonesia untuk Deteksi Emosi dari Teks," *J. Elektro*, vol. 12, no. 2, pp. 103–110, 2019.

[14] M. A. Azim and M. H. Bhuiyan, "Text to emotion extraction using supervised machine learning techniques," *Telkomnika (Telecommunication Comput. Electron. Control.*, vol. 16, no. 3, pp. 1394–1401, 2018, doi: 10.12928/TELKOMNIKA.v16i3.8387.

[15] D. William and D. Suhartono, "Text-based Depression Detection on Social Media Posts: A Systematic Literature Review," *Procedia Comput. Sci.*, vol. 179, no. 2019, pp. 582–589, 2021, doi: 10.1016/j.procs.2021.01.043.

[16] M. S. Saputri, R. Mahendra, and M. Adriani, "Emotion Classification on Indonesian Twitter Dataset," *Proc. 2018 Int. Conf. Asian Lang. Process. IALP 2018*, no. November, pp. 90–95, 2019, doi: 10.1109/IALP.2018.8629262.

[17] P. Ekman, "An Argument for Basic Emotions," *Cogn. Emot.*, vol. 6, no. 3–4, pp. 169–200, 1992, doi: 10.1080/02699939208411068.

[18] J. E. The, A. F. Wicaksono, and M. Adriani, "A two-stage emotion detection on Indonesian tweets," *ICACSIS 2015 - 2015 Int. Conf. Adv. Comput. Sci. Inf. Syst. Proc.*, pp. 143–146, 2016, doi: 10.1109/ICACSIS.2015.7415174.

[19] H. Liang, X. Sun, Y. Sun, and Y. Gao, "Text feature extraction based on deep learning: a review," *Eurasip J. Wirel. Commun. Netw.*, vol. 2017, no. 1, pp. 1–12, 2017, doi: 10.1186/s13638-017-0993-1.

[20] V. Sundaram, S. Ahmed, S. A. Muqtadeer, and R. Ravinder Reddy, "Emotion Analysis in Text using TF-IDF," *Proc. Conflu. 2021 11th Int. Conf. Cloud Comput. Data Sci. Eng.*, pp. 292–297, 2021, doi: 10.1109/Confluence51648.2021.9377159.

[21] M. Das, S. Kamalanathan, and P. Alphonse, "A Comparative Study on TF-IDF feature weighting method and its analysis using unstructured dataset," *CEUR Workshop Proc.*, vol. 2870, pp. 98–107, 2021.

[22] S. S. and P. K.V., "Sentiment analysis of malayalam tweets using machine learning techniques," *ICT Express*, vol. 6, no. 4, pp. 300–305, 2020, doi: 10.1016/j.icte.2020.04.003.

[23] N. Kholodna, V. Vysotska, and S. Albota, "A machine learning model for automatic emotion detection from speech," *CEUR Workshop Proc.*, vol. 2917, pp. 699–713, 2021.

[24] A. Onan, "Sentiment analysis on massive open online course evaluations: A text mining and deep learning approach," *Comput. Appl. Eng. Educ.*, vol. 29, no. 3, pp. 572–589, 2021, doi: 10.1002/cae.22253.

[25] J. L. Speiser, M. E. Miller, J. Tooze, and E. Ip, "A comparison of random forest variable selection methods for classification prediction modeling," *Expert Syst. Appl.*, vol. 134, pp. 93–101, 2019, doi: 10.1016/j.eswa.2019.05.028.

[26] P. Probst, M. N. Wright, and A. L. Boulesteix, "Hyperparameters and tuning strategies for random forest," *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 9, no. 3, pp. 1–15, 2019, doi: 10.1002/widm.1301.

[27] H. Luzern, *Cyber Security Cyber Security*, no. March. 2017. doi: 10.1007/978-981-16-9229-1.

[28] A. Bandhakavi, N. Wiratunga, D. Padmanabhan, and S. Massie, "Lexicon based feature extraction for emotion text classification," *Pattern Recognit. Lett.*, vol. 93, pp. 133–142, 2017, doi: 10.1016/j.patrec.2016.12.009.

[29] J. Akram and A. Tahir, "Lexicon and Heuristics Based Approach for Identification of Emotion in Text," in *2018 International Conference on Frontiers of Information Technology (FIT)*, Dec. 2018, pp. 293–297. doi: 10.1109/FIT.2018.00058.

[30] D. D. Palmer, "Text pre-processing," *Handb. Nat. Lang. Process. Second Ed.*, pp. 9–30, 2010, doi: 10.4018/978-1-5225-4990-1.ch006.

[31] Y. Nooryuda Prasetya and D. Winarso, "Penerapan Lexicon Based Untuk Analisis Sentimen Pada Twiter Terhadap Isu Covid-19," *J. Fasilkom*, vol. 11, no. 2, pp. 97–103, 2021.

[32] V. V Nhlabano and P. E. N. Lutu, "Impact of Text Pre-Processing on the Performance of Sentiment Analysis Models for Social Media Data," *2018 Int. Conf. Adv. Big Data, Comput. Data Commun. Syst. icABCD 2018*, pp. 1–6, 2018, doi: 10.1109/ICABCD.2018.8465135.

[33] S. M. Sarsam, H. Al-Samarraie, A. I. Alzahrani, W. Alnumay, and A. P. Smith, "A lexicon-based approach to detecting suicide-related messages on Twitter," *Biomed. Signal Process. Control*, vol. 65, no. November 2020, 2021, doi: 10.1016/j.bspc.2020.102355.

[34] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, "Lexicon-Based Methods for Sentiment Analysis DRAFT DRAFT DRAFT!," *Comput. Linguist.*, vol. 37, no. 2, pp. 267–307, 2011.

[35] F. Calefato, F. Lanubile, and N. Novielli, "EmoTxt: A toolkit for emotion recognition from text," *2017 7th Int. Conf. Affect. Comput. Intell. Interact. Work. Demos, ACIIW 2017*, vol. 2018-Janua, no. July, pp. 79–80, 2018, doi: 10.1109/ACIIW.2017.8272591.

[36] S. Liu and I. Lee, "Extracting features with medical sentiment lexicon and position encoding for drug reviews," *Heal. Inf. Sci. Syst.*, vol. 7, no. 1, 2019, doi: 10.1007/s13755-019-0072-6.

[37] A. Bandhakavi, N. Wiratunga, P. Deepak, and S. Massie, "Generating a word-emotion lexicon from #emotional tweets," *Proc. 3rd Jt. Conf. Lex. Comput. Semant. *SEM 2014*, pp. 12–21, 2014, doi: 10.3115/v1/s14-1002.

[38] A. Gholipour Shahraki and O. R. Zaïane, "Lexical and Learning-based Emotion Mining from Text," *Univ. Alberta, Canada*, pp. 24–55, 2017.

[39] K. S. Nugroho, "Detecting Emotion in Indonesian Tweets : A Term-Weighting Scheme Study," vol. 8, no. 1, pp. 61–70, 2022, [Online]. Available: https://www.e-journal.unair.ac.id/JISEBI/article/view/33017

[40] F. Koto and G. Y. Rahmaningtyas, "Inset lexicon: Evaluation of a word list for Indonesian sentiment analysis in microblogs," in *Proceedings of the 2017 International Conference on Asian Language Processing, IALP 2017*, 2018, vol. 2018-Janua, no. December, pp. 391–394. doi: 10.1109/IALP.2017.8300625.

[41] F. Å. Nielsen, "A new ANEW: Evaluation of a word list for sentiment analysis in microblogs," in *CEUR Workshop Proceedings*, 2011, vol. 718, no. March, pp. 93–98.

[42] R. N. Waykole and A. D. Thakare, "International Journal of Advance Engineering and Research A REVIEW OF FEATURE EXTRACTION METHODS FOR TEXT," pp. 351–354, 2018.

[43] Q. Rajput, S. Haider, and S. Ghani, "Lexicon-Based Sentiment Analysis of Teachers' Evaluation," *Appl. Comput. Intell. Soft Comput.*, vol. 2016, pp. 1–12, 2016, doi: 10.1155/2016/2385429.

[44] F. A. Bachtiar, W. Paulina, and A. N. Rusydi, "Text Mining for Aspect Based Sentiment Analysis on Customer Review : a Case Study in the Hotel Industry," *5th Int. Work. Innov. Inf. Commun. Sci. Technol.*, no. March, 2020.

ISSN 2085-4552

Elicitation of Needs Using User Personas to Improve Software User Experience

Ahmad Raihan Djamarullah[1], Wahyu Andhyka Kusuma[2]
[1,2] Dept. of Informatics, University of Muhammadiyah Malang, Malang, Indonesia
[1] raihandj@webmail.umm.ac.id, [2] wahyukusuma@umm.ac.id

Accepted 28 May 2022
Approved 29 June 2022

***Abstract*— Tweets from users in the form of opinions about a product can be used as a company evaluation of the product. To obtain this evaluation, the method that can be used is sentiment analysis to divide opinions into positive and negative opinions. This study uses 1000 data from Twitter related to an internet service provider company where the data is divided into two classes, namely 692 positive classes and 308 negative classes. In the Tweet there are still many words that are not standard. Therefore, previously carried out the initial process or preprocessing to filter out non-standard words. Before doing the classification, the data needs to be divided into training data and test data with a ratio of 90:10, then processed using the Support Vector Machine and Naïve Bayes techniques to get the results of the classification of positive opinions and negative opinions. The level of accuracy in the classification using the Support Vector Machine is 84% and using Naïve Bayes is 82%.**

***Index Terms*— Elicitation; HCI Technique; User Persona**.

I. INTRODUCTION

The development of technology is currently very fast and will continue to grow, it has a huge impact on various aspects that exist. Many industrial fields are finally competing to develop technology in these fields[1]. This is because it is very helpful to increase effectiveness and efficiency in work processes within a company or organization[2]. Thus, it can be concluded that today's information technology is very important in the success of an organization or agency management[3]. In order to create an effectiveness and efficiency, proper governance of information technology is needed[4]. Thus, "organizational and agency information technology is related to how top management of organizations and agencies gain confidence that the Information System Manager (Chief Information Officer) and IT organizations can provide returns or returns in the form of value for the organization or agency[5]." In other words, information technology can provide support to organizations/agencies in achieving the desired goals[6].

Information system analysis is very important to do, it is needed in order to produce a system that is easy, effective, efficient, and also appropriate for the users of the system. One model of information systems analysis by applying the concept of Human Computer Interaction (HCI)[7].

Techniques in software development are growing with the times, one of which is widely used, namely the Human Computer Interaction (HCI) approach. Producing a usable and safe system in the sense that the system can function properly is the goal of HCI, while also placing more emphasis on evaluation, design and implementation aspects. In several previous studies on HCI the HCI model can help analyze how users can interact with computers, so that they can describe behavioral models from various points of view[8].

The software development process is a phase that is often researched, including the excavation and requirements analysis phase. The purpose of this stage is to obtain specific requirements, both functional and non-functional. Several approaches have been carried out by several previous researchers. Requirements analysis is a process to get user requirements and constraints clearly and precisely[9], software requirements will be needed through intensive communication to get user needs[10], in order to eliminate ambiguity that often occurs due to the inability to explore these needs. or also because stakeholders (stakeholders) cannot properly express the desired needs.

Several approaches to exploring needs include group methods and individual methods, one of which is to group customer needs into 5 segments through adscore[11]. The group method is often used for several similar stakeholders and usually in large numbers, while the individual method focuses more on person to person within these stakeholders. Previous research conducted individual interviews from several experts with the aim of obtaining model requirements[12].

When developing an information system, the developer of the system must pay attention to the human-computer interaction factor. Human and Computer Interaction (HCI) is a scientific discipline that examines communication or interaction between users and systems. The main role of HCI is to produce

a system that is useful, safe, productive, effective, efficient and functional[13].

HCI focuses more on system design on the user or User Center Design (UCD), by paying attention to the users who will use the system, the system can be created according to and appropriate for the user. A system that is right for the user will provide comfort to the user when using the system so that the purpose of implementing the system will be achieved and will not fail[13].

In this case study, the needs analysis was obtained from stakeholders related to or related to the practicum module uploading system through infotech which is within the academic scope of the University of Muhammadiyah Malang. In this study, it involved stakeholders from students majoring in informatics class 2019 who took part in practicum activities, lecturers who were in charge of courses who had practicums, and laboratory assistants who supervised the practicum. The current practicum module uploading system is considered less effective due to the sometimes late uploading of practicum modules so that students or practitioners lack time to study and complete the given practicum assignments.

In this study, we will focus on exploring existing needs through the human-computer interaction (HCI) approach, then collecting data on characteristics from stakeholders related to the existing practicum module uploading system. Persona taken later will be in the form of characteristics and objectives based on the results of interviews and observations[14]. Quantitative and qualitative data will later become the basis for conducting analysis and synthetics to be collected as persona designs[15]. This method will produce an approach in various aspects of psychology, objectivity, motivation and existing habits, based on demographics and social.

II. Related Research

TABLE I. Literature Review and Hypothesis Development

Aspect	Acuna et al (2012)	Idoughi et al. (2012)	Aoyama (2007)
Persona Identification	X	X	X
Usage context/Environment	X	_	_
Psychological details used	X	X	X
Application /service	X	_	X
Need/Expectations	X	X	_
Skills/Previous Experiences	X	X	X
Goals	X	_	X
Segments	X	X	X
Special Needs/Accessibility	X	X	X
Relations/Interact ions with other People Problems Existing Solutions	X	X	X

In the table above, it can be seen that these characteristics are considered techniques for creating personas. Most of the given characteristics relate to the behavioral characteristics of the personas (Persona Identification, Psychological details used, Need/Expectations, Special Needs/Accessibility, and Relations/Interactions with other People Problems Existing Solutions), which describes the personal life and psychological details of each person. personas. These characteristics are useful for understanding users, but cannot identify possible application requirements.

Furthermore, for the characteristics of Application / service, Skills / Previous Experiences, and Goals can help identify possible application requirements. However, these characteristics are so general that they may present information that is not part of the application domain and may not result in the requirements. Attitudes and behaviors are important to describe in personas, they are needed to understand and identify possible requirements that are useful for application development[16].

Therefore, it can be identified that some of the limitations in the techniques listed in the table are:

- The technique focuses on understanding rather than identifying possible requirements
- Characteristics that help identify behavior are not focused specifically in the application domain, as these characteristics are described in general

So it is necessary to propose specifically the techniques used to guide the description of personas that focus on identifying possible requirements related to the application domain to be developed.

A. Data Crawling

Crawling is a method of collecting data from a source to be analyzed or processed. Crawling is the first stage that is usually used to analyze the sentiments of social media users towards a product or service. Crawling can also be interpreted as a method of quickly gathering large numbers of web pages into a local storage and indexing them based on specified keywords [8]. Sentiment analysis usually uses the Crawling method on Twitter social media by utilizing the features provided by Twitter, namely Application Programming Interface Systems (APIs) [9].

III. METHOD

A. Activity 1: State Hypotheses

State Hypotheses are formulated to identify the possibility of persona. This hypothesis helps to determine who will be the subject of the interview. This is the first stage in getting acquainted with users[17]. In addition, at this stage also carried out stages to design a formulation in the interview.

B. Activity 2: Identify Behavioral Variables

Identify a list of behavioral variables to characterize possible users taken from the results of the interview synthesis[17]. Which interview subjects will be easier to map if the range of values has been defined before mapping[16].

C. Activity 3: Identify Significant Behavior Patterns

Identify specific subject groups that appear in the range leading to the use of archetypes for personas and modeling users. The purpose of this activity will be to get a result in the form of two products, namely, the percentage grouping table and the arrangement of significant behavioral patterns, the percentage grouping table containing the percentage of interview subjects that have been mapped to each behavioral variable value[17].

D. Activity 4: Map Interview Subjects to Behavioral Variables

Represents how many subjects are grouped with respect to significant behavioral variables, which is the initial model of system users[17]. This activity aims to detect the grouping of certain behavior patterns with the existing range of values[17], so that they can be made significant and become the basic pattern of the persona.

E. Activity 5: Synthesize Characteristics and Relevant Goals

This activity generates a basic persona document. In short, the description of the relevant characteristics and goals reflects the personality of the created persona. This document is an end-user model document[17].

F. Activity 6: Check for Redundancy and Completeness

Examine the redundancy, completeness, mapping, characteristics and goals of personas to find out if there are important gaps[17]. To do this, we check that the essential identification and aspects are fully defined in the created personas. Looking for important information from the previous activity stages.

G. Activity 7: Expand the Description of Attributes and Behaviors

Extends Attribute and Behavior Descriptions in a narrative manner. Narrative is a clearer way to convey or communicate people's attitudes, needs and problems. Many aspects of the Persona Document are obscure, and useless for this task as Narrative[17]. The narrative can also be with additional data by interviewing stakeholders

H. Activity 8: Designate Persona Types

Determine the persona we want to create by determining the primary and secondary types[17]. After describing all the personas in a narrative, in the next activity we need to determine which persona is the main target of the system being built. We must define only one of all the personas whose needs and objectives can be met and represented in one interface without upsetting the other personas. To determine which persona will be the main target of the system, we must divide the persona into 2 types, namely primary and secondary personas. The primary persona is the persona that represents the main needs & objectives of the system development. While the secondary persona is a person who has additional needs that are not in the primary persona.

I. Activity 9: Build Use Cases

The activity that must be done after the persona that is the main target is clear is to create a use case. Use cases are made based on the persona in the basic document as well as in the narration as well as information from users obtained from previous activities[17]. Use cases are made in the form of diagrams like use case diagrams in general, which are made using annotations that show the relationship with primary and secondary personas. And also made a use case scenario or the flow of the prototype that we will make in the next stage.

J. Activity 10: Mockup

Implement and Evaluate Prototypes. We have used the use case developed in Activity 9 to implement the prototype. Creating mockups designed by stakeholders[17]. Mock Ups are made based on use cases and also various needs that have been recorded in basic documents in previous activities. Then from the mock-up that is made, we evaluate the environment in which the system will be implemented, because there we will find potential users of the system being built.

IV. RESULT AND DISCUSSION

A. Activity 1: State Hypothesis

The purpose of this activity is to identify the variables that can be created, to find out the differences between users based on their needs and behavior.

TABLE II. STATE HYPOTHESIS

Hypothesis	Personas	Explanation
H0	College student	Requires sufficient time to complete and understand

		assignments and practicum materials
H1	Lab Assistant	Difficulty getting lecturer approval regarding modules
H2	Lecturer	Helping students to easily understand the material that has been given

From the table above shows a list of hypotheses for each persona, from the results obtained, it can be identified and made personas synthesize between students, lab assistants and lecturers to assemble the necessary needs.

B. Activity 2: Identify Behavioral Variables

The results of the interviews will be given a range with values that have been defined before the mapping is made.

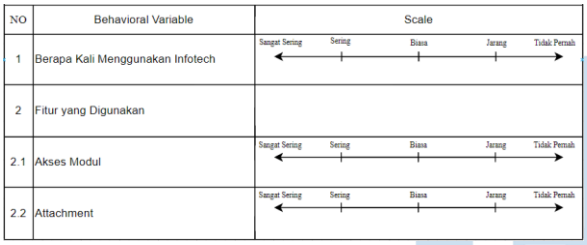

Fig. 1. Identify Behavioral Variables

From the table of behavioral variables above, it can be identified various characteristics of each persona. Which is included in the range of values that have been set in an infotech system access activity, which is in accordance with the existing problem.

C. Activity 3: Identify Significant Behavior Patterns

At this stage, each variable from each range of values will be grouped. From the results above, the percentage will be made to get a significant grouping of each behavior pattern of the existing respondents. To get the percentage, the results of the values obtained from each subject will be calculated in comparison with the total number of existing subjects.

TABLE III. IDENTIFY SIGNIFICANT BEHAVIOR PATTERNS (STUDENT)

Observed behavioural variable	Scale	Result	Percentage
How many times to use Infotech	Very Often	5 People	100 %
	Often	0 People	0 %
	Normal	0 People	0 %
	Infrequently	0 People	0 %
	Never	0 People	0 %
Features Used			
Access Module	Very Often	0 People	0 %
	Often	5 Orang	100 %
	Normal	0 People	0 %
	Infrequently	0 People	0 %
	Never	0 People	0 %
Attachment	Very Often	0 People	0 %
	Often	0 People	0 %
	Normal	5 People	100 %
	Infrequently	0 People	0 %
	Never	0 People	0 %

For the results of respondents' answers taken from students, in the first variable the higher percentage is very often, then for the second variable the higher percentage is often, and for the last variable the higher percentage is normal.

TABLE IV. IDENTIFY SIGNIFICANT BEHAVIOR PATTERNS (LECTURER)

Observed behavioural variable	Scale	Result	Percentage
How many times to use Infotech	Very Often	0 People	0 %
	Often	0 People	0 %
	Normal	0 People	0 %
	Infrequently	5 People	100 %
	Never	0 People	0 %
Features Used			
Access Module	Very Often	0 People	0 %
	Often	0 People	0 %
	Normal	2 People	40 %
	Infrequently	3 People	60 %
	Never	0 People	0 %
Attachment	Very Often	0 People	0 %
	Often	0 People	0 %
	Normal	0 People	0 %
	Infrequently	0 People	0 %
	Never	5 People	100 %
Evaluation	Very Often	0 People	0 %
	Often	0 People	0 %
	Normal	0 People	0 %
	Infrequently	0 People	0 %
	Never	5 People	100 %

For the results of respondents' answers taken from students, in the first variable the higher percentage is Infrequently, then for the second variable the higher percentage is Infrequently, then for the third variable the higher percentage is never, and for the last variable the higher percentage is.

TABLE V. IDENTIFY SIGNIFICANT BEHAVIOR PATTERNS (LECTURER)

Observed behavioural variable	Scale	Result	Percentage
How many times to use Infotech	Very Often	0 People	0 %
	Often	0 People	0 %
	Normal	0 People	0 %
	Infrequently	5 People	100 %
	Never	0 People	0 %
Features Used			
Access Module	Very Often	0 People	0 %
	Often	0 People	0 %
	Normal	2 People	40 %
	Infrequently	3 People	60 %
	Never	0 People	0 %
Attachment	Very Often	0 People	0 %
	Often	0 People	0 %
	Normal	0 People	0 %
	Infrequently	0 People	0 %
	Never	5 People	100 %
Evaluation	Very Often	0 People	0 %
	Often	0 People	0 %
	Normal	0 People	0 %
	Infrequently	0 People	0 %
	Never	5 People	100 %

For the results of respondents' answers taken from students, in the first variable the higher percentage is often, then for the second variable the higher percentage is normal, then for the third variable the higher percentage is never, and for the last variable the higher percentage is Often.

D. Activity 4: Map Interview Subjects to Behavioral Variables

At this stage, fragment mapping will be carried out from the answers given from stakeholders based on the behavioral identification that has been made. Of the cases that have been identified, there are 5 respondents each from each stakeholder who have provided answers.

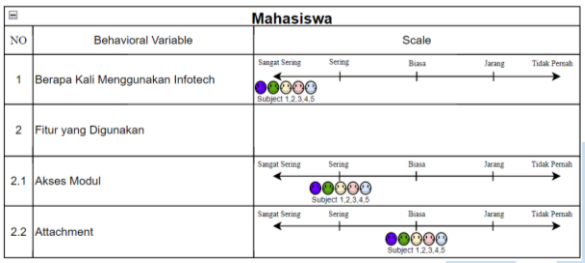

Fig. 2. Student Mapping Fragments

In the mapping fragment image above, the results of each variable choice for student respondents are obtained. For the first question, how many times have you used infotech, the answer is very often, because all UMM informatics students in each semester must take at least 1 course that has a practicum. So they need to access infotech for practical needs such as accessing modules to study and do assignments from the given practicum, this is because the modules that are distributed sometimes have not been uploaded so some students need to check whether the modules have been uploaded or not. There is also the use of the file attachment feature which is commonly used by students to collect practicum assignments.

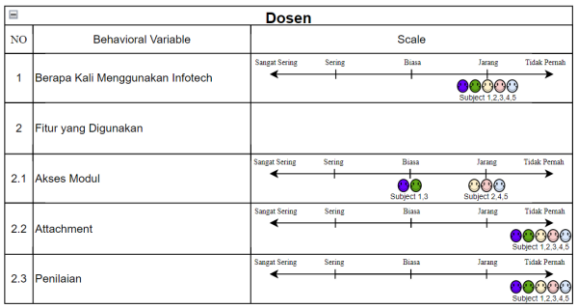

Fig. 3. Lecturer Mapping Fragment

Furthermore, for respondents from lecturers. For the first question, namely how many times have used infotech, the answer is infrequently or only occasionally. Furthermore, for the features used, namely module access which is sometimes used to view modules so that they can be taught a little to students. Then the attachment feature is never used by lecturers because they do not collect practicum assignments and upload modules. And the last one is the assessment feature where every lecturer never uses this feature because the practicum assessment is submitted to the lab assistant who will later pay the grade to the lecturer.

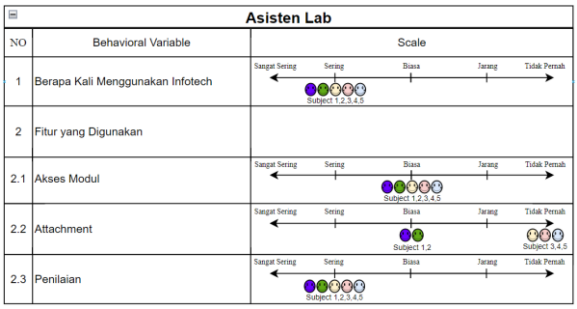

Fig. 4. Lab Assistant Mapping Fragment

The last one is the lab assistant respondent. For the first question, namely how many times have you used infotech, the answer is often, because the lab assistant is responsible for carrying out the practicum of each practitioner. Furthermore, the features used are module access which is often used to view the modules that will be used by the practitioner when doing practicum. Then for the use of the attachment feature, the lab assistant uses the attachment feature to upload the module, but sometimes only the CO of the lab assistant from each class uploads the module. And the last is the assessment feature, this feature is often used by every lab assistant when conducting practicums to assess practitioners.

E. Activity 5: Synthesize Characteristics and Relevant Goals

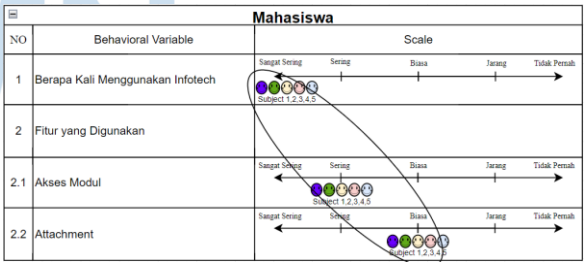

Fig. 5. Synthesize Characteristics and Relevant Goals (Student)

In the picture above, the results obtained from each choice of variables for student respondents. For the first question, how many times have you used infotech, the answer is very often, because all UMM informatics students in each semester must take at least 1 course that has a practicum. So they need to access infotech for practical needs such as accessing modules to study and do assignments from the given practicum, this is because the modules that are distributed sometimes have not been uploaded so some students need to check whether the modules have been uploaded or not. There is also the use of the file attachment feature which is commonly used by students to collect practicum assignments.

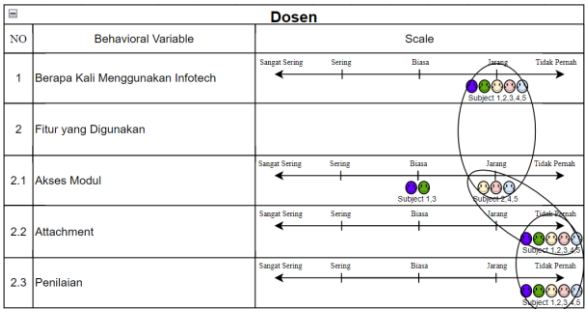

Fig. 6. Synthesize Characteristics and Relevant Goals (Lecturer)

Furthermore, for respondents from lecturers. For the first question, namely how many times have used infotech, the answer is infrequently or only occasionally. Furthermore, for the features used, namely module access which is sometimes used to view modules so that they can be taught a little to students. Then the attachment feature is never used by lecturers because they do not collect practicum assignments and upload modules. And the last one is the assessment feature where every lecturer never uses this feature because the practicum assessment is submitted to the lab assistant who will later pay the grade to the lecturer. From the four questions that have been given, it can be seen that there is one fragment that shows the significance of the choice of all respondents. For the existing fragments, the significant respondents chose the seldom choice and never chose the option.

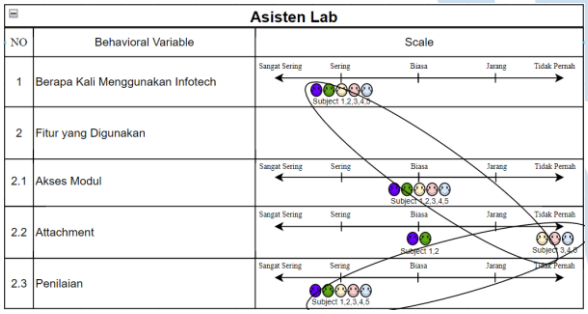

Fig. 7. Synthesize Characteristics and Relevant Goals (Lab Assistant)

The last one is the lab assistant respondent. For the first question, namely how many times have you used infotech, the answer is often, because the lab assistant is responsible for carrying out the practicum of each practitioner. Furthermore, the features used are module access which is often used to view the modules that will be used by the practitioner when doing practicum. Then for the use of the attachment feature, the lab assistant uses the attachment feature to upload the module, but sometimes only the CO of the lab assistant from each class uploads the module. And the last is the assessment feature, this feature is often used by every lab assistant when conducting practicums to assess practitioners. From the four questions that have been given, it can be seen that there are two fragments that indicate the significance of the choices of all respondents. For the existing fragments, the respondents consistently chose the frequent choice, the usual choice and the never choice.

F. Activity 6: Check for Redundancy and Completeness

At this stage, a check will be carried out regarding the behavioral variables that have been made by adjusting the results of the behavioral variables that have been carried out with the behavioral variables from the new respondents. This is done to find out whether the needs analysis carried out is appropriate or still requires additional additions.

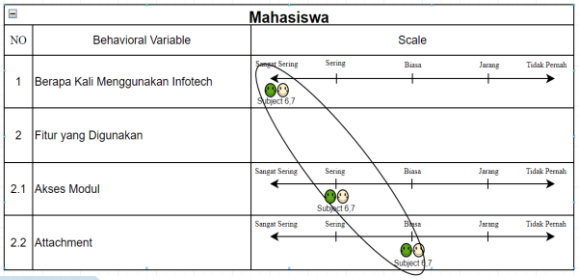

Fig. 8. Behavioral Variables to Check For Redundancy And Completeness (Student)

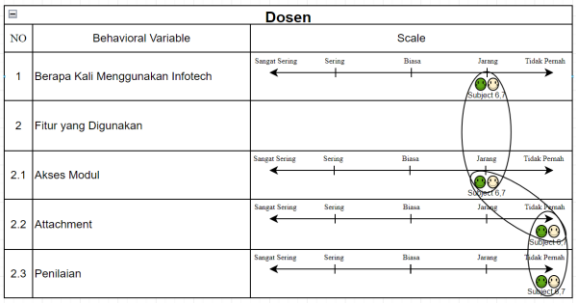

Fig. 9. Behavioral Variables to Check For Redundancy And Completeness (Lecturer)

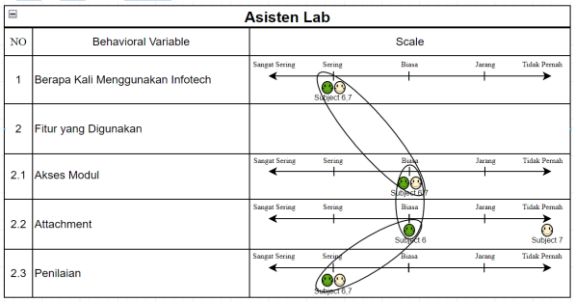

Fig. 10. Behavioral Variables to Check For Redundancy And Completeness (Lab Assistant)

In the picture above, there is a table of Behavioral Variables to complete Redundancy And Completeness, there are 3 types of respondents, namely Students, Lecturers and also Lab Assistants. From the three types of respondents, 2 people were taken each to be respondents.

G. Activity 7: Expand the Description of Attributes and Behaviors

PERSONAS FOUNDATION DOCUMENT

1. PERSONA IDENTIFICATION
   - Name : Ibnu Jahsy
   - Age : 21
2. ROLES & TASKS
   - Ibnu Jahsy adalah seorang mahasiswa informatika angkatan 2019 di Universitas Muhammadiyah Malang
   - Kegiatan regularnya adalah bermain game dan belajar membuat game
3. OBJECTIVES
   - Untuk objectives jangka pendek dan panjang adalah menjadi game developer
4. SEGMENT
   - Dia sekarang tinggal di malang
5. SKILL & KNOWLEDGE
   - Ibnu Jahsy menggunakan laptop untuk kegiatan seperti kuliah dan membuat game
   - Ibnu Jahsy menggunkan aplikasi Unity dan Unreal Engine
6. CONTEXT & ENVIRONMENT
   - Motivasi yang dimiliki adalah ingin membuat game yang bagus, dan ingin menjadi game developer
   - Ibnu Jahsy suka bermain Game
7. PERSONAL AND PSYCHOLOGICAL
   - Ibnu Jahsy adalah seorang yang sulit menghadapi kegiatan atau tugas yang tiba – tiba.

Fig. 11. Stakeholder : Student

PERSONAS FOUNDATION DOCUMENT

1. PERSONA IDENTIFICATION
   - Name : Ilyas Nuryasin
2. ROLES & TASKS
   - Ilyas Nuryasin adalah seorang dosen yang mengajar di Universitas Muhammadiyah Malang
3. OBJECTIVES
   - Untuk objectives jangka pendek dan panjang adalah menjadi pengajar yang baik dan dapat memberikan pembelajaran yang berkualitas
4. SEGMENT
   - Ilyas Nuryasin sekarang tinggal di malang
5. SKILL & KNOWLEDGE
   - Ilyas Nuryasin menggunakan laptop untuk kegiatan seperti mengajar
   - Ilyas Nuryasin menggunkan aplikasi zoom untuk mengajar ketika kelas online
6. CONTEXT & ENVIRONMENT
   - Motivasi yang dimiliki adalah ingin menjadi pengajar yang baik dan dapat memberikan pembelajaran yang berkualitas
7. PERSONAL AND PSYCHOLOGICAL

   Ilyas Nuryasin, S.Kom., M.Kom adalah seorang dosen yang baik, menerapkan cara mengajara dengan melibatkan mahasiswa dalam setiap pertemuannya.

Fig. 12. Stakeholder : Lecturer

H. Activity 8: Designate Persona Types

TABLE VI. USER PERSONA

	Ibnu Jahsy • 21 • Malang • Student • Universitas Muhammadiyah Malang Ibnu Jahsy is an informatics student from the class of 2019 at the University of Muhammadiyah Malang, his daily habit is making and playing games. For the short and long term objectives are to become a game developer, the motivation is to want to make good games, and to become a game developer. In doing this, respondents use laptops to open the Unity application so that they can make games, respondents often use their laptops to play and make games. For fear and the feared challenge is the sudden assignment of.

I. Activity 9: Build Use Cases

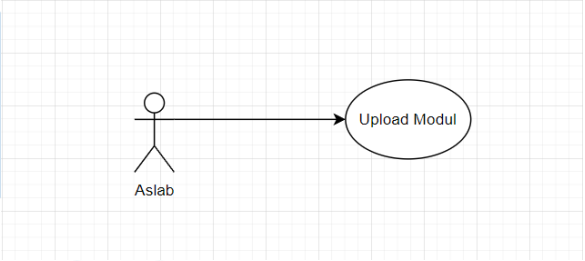

Fig. 13. Use Case

There is a document that describes details of the activities carried out for each use case. Specifications are described in the form of use case descriptor that provides a description more about use case, result of use case descriptors are described by Table VII.

TABLE VII. IDENTIFY SIGNIFICANT BEHAVIOR PATTERNS (LAB ASSISTANT)

USE CASE: Aslab Menguplоad Modul

Principal Actor:

- Lab Assistant (Aslab)

Stakeholders and Goals:

- Lab Assistant: Want to Easily Upload Modules

Principal Success Scenario:

1.	Masuk ke website infotech
2.	Tekan menu login
3.	Login dengan menggunakan nim dan password
4.	Pilih menu student
5.	Pilih kelas praktikum
6.	Pilih menu upload modul
7.	Input modul yang ingin diupload
8.	Masukkan waktu publish dari modul yang diupload
9.	Tekan tombol save
10.	Modul di publish sesuai waktu

J. *Activity 10: Mockup*

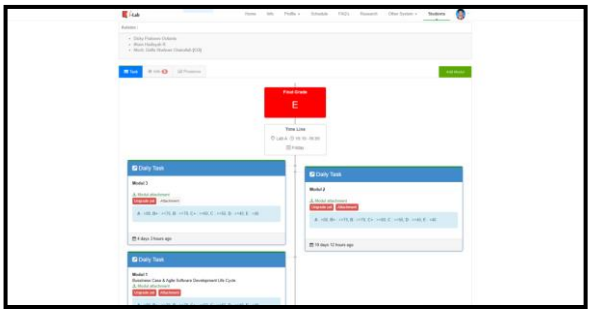

Fig. 14. Homepage Mockup

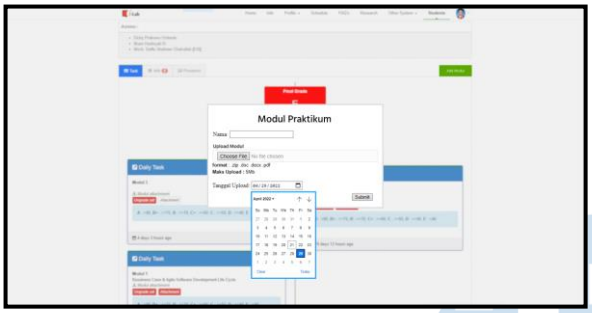

Fig. 15. Module Upload Mockup

The final results at the research stage that have been carried out are made in the form of a mockup according to the evaluation carried out with the related persona. In Figure 12 is the start page where there is a button to upload the practicum module. Then in the 13th picture, after the button to upload the module is pressed, a menu for uploading the practicum module will appear and also a menu for setting the schedule for uploading the practicum module so that the practicum module will be uploaded according to a predetermined schedule.

V. CONCLUSIONS

In the research conducted, it was found problems related to the module uploading system through infotech, namely the upload of the practicum module which is sometimes late and also the absence of a notification feature when the practicum module has been uploaded. practice has been uploaded.

To overcome this, a solution that might help is needed is the addition of a scheduled upload feature so that later the practicum module uploader can be done automatically according to a predetermined schedule and also the addition of a notification feature when the practicum module has been uploaded.

REFERENCES

[1] Y. I. Maulana, P. Studi, and M. Informatika, "Perancangan Perangkat Lunak Sistem Informasi Pendataan Guru Dan Sekolah (SINDARU) Pada Dinas Pendidikan Kota Tangerang Selatan," J. Pilar Nusa Mandiri, vol. 13, no. 1, pp. 21–27, 2017.

[2] A.. M. Djohan, "Penilaian Tingkat Kematangan Strategic Alignment Bisnis Dan Ti Menggunakan Cobit 4.1: Studi Kasus Universitas Xyz," Semin. Nas. Sist. Inf. Indones., no. September, pp. 1–8, 2014.

[3] V. Sahfitri, "Perguruan Tinggi Swasta Di Kota Palembang," pp. 1–7, 2014.

[4] A. S. Hidayat, W. Ubleeuw, A. Fauzi, and P. M. Akhirianto, "Sistem Pengolahan Data Nilai Berbasis Web Pada Sekolah Menengah Pertama (Smp) Karel Sadsuitubun Langgur," J. Teknol. Inform. dan Komput., vol. 5, no. 2, pp. 13–23, 2019.

[5] L. A. Purwanto and T. Dirgahayu, "Pengukuran Tingkat Kematangan Tata Kelola Pengelolaan Permasalahan Sistem Informasi Menggunakan Kerangka Kerja COBIT 4.1 (Studi Kasus: Sistem Informasi Akademik Universitas Muhammadiyah Purwokerto)," JUITA J. Inform., vol. 5, no. 2, p. 103, 2018, doi: 10.30595/juita.v5i2.1629.

[6] R. Carneiro, Discovering the treasure of learning. 2011.

[7] Haryoko, S. (2012). Penerapan Sistem Human Computer Interaction (HCI) Pada Website E-Learning. Inspiration: Jurnal Teknologi Informasi Dan Komunikasi, 2(2). doi:10.35585/inspir.v2i2.20

[8] Z. He, T. Chang, S. Lu, H. Ai, D. Wang, and Q. Zhou, "Research on Humancomputer Interaction Technology of Wearable Devices Such as Augmented Reality Supporting Grid Work," Procedia Comput. Sci., vol. 107, no. Icict, pp. 170–175, 2017, doi: 10.1016/j.procs.2017.03.074.

[9] A. Safwat and M. B. Senousy, "Addressing Challenges of Ultra Large Scale System on Requirements Engineering," Procedia Comput. Sci., vol. 65, no. Iccmit, pp. 442–449, 2015, doi: 10.1016/j.procs.2015.09.116.

[10] T. Alsanoosy, M. Spichkova, and J. Harland, "The influence of power distance on requirements engineering activities," Procedia Comput. Sci., vol. 159, pp. 2394–2403, 2019, doi: 10.1016/j.procs.2019.09.414

[11] R. Chen, Q. Wang, and W. Xu, "Mining user requirements to facilitate mobile app quality upgrades with big data," Electron. Commer. Res. Appl., vol. 38, no. August, p. 100889, 2019, doi: 10.1016/j.elerap.2019.100889

[12] H. Scherer, A. Albers, and N. Bursac, "Model Based Requirements SINTECH Journal | 99 Engineering for the Development of Modular Kits," Procedia CIRP, vol. 60, pp. 145–150, 2017, doi: 10.1016/j.procir.2017.01.032

[13] Prihati, P., Mustafid, M., & Suhartono, S. (2011). Penerapan Model Human Computer Interaction (HCI) Dalam Analisis Sistem Informasi. JSINBIS (Jurnal Sistem Informasi Bisnis), 1(1), 01-08. https://doi.org/10.21456/vol1iss1pp01-08

[14] B. Ferreira, G. Santos, and T. Conte, "Identifying Possible Requirements using Personas - A Qualitative Study," *Proceedings of the 19th International Conference on Enterprise Information Systems*, vol. 2, no. Iceis, pp. 978–989, 2017.

[15] Å. Blomquist and M. Arvola, "Personas in action," Proceedings of the second Nordic conference on Human-computer interaction - NordiCHI 02, 2002.

[16] A. Cooper, R. Reimann. "About Face 2.0: The Essentials of Interaction Design, Wiley Publishing, Indianapolis", 2004.

[17] S. T. Acuña, J. W. Castro, and N. Juristo, "A HCI technique for improving requirements elicitation," in Information and Software Technology, vol. 54, no. 12, pp. 1357–1375, 2012.

ISSN 2085-4552

E-Business Software Product Line Methodology Based on SMEs Characteristics

Ni Made Satvika Iswari[1,3], Eko Kuswardono Budiardjo[1], Zainal Arifin Hasibuan[2]
[1] Faculty of Computer Science, Universitas Indonesia, Depok, Indonesia
[2] Faculty of Computer Science, Universitas Dian Nuswantoro, Semarang, Indonesia
[3] Faculty of Engineering and Informatics, Universitas Multimedia Nusantara, Tangerang, Indonesia
[1] satvika@umn.ac.id, [2] eko@cs.ui.ac.id, [3] zhasibua@dsn.dinus.ac.id

Accepted 02 June 2022
Approved 10 July 2022

***Abstract*— The growth of Small and Medium Enterprises (SMEs) will certainly have a positive influence on economic development in a country. However, many SMEs struggle to survive, grow, and show limited productivity. It was recorded that there were around 63.9% of SMEs in Indonesia whose turnover had decreased by more than 30% during the COVID-19 pandemic. Among the factors that cause the low survival rate of SMEs, a factor that is considered critical is the lack of success in the use of e-business. The use of e-business by SMEs is not a one-size-fits-all solution, because SMEs have various characteristics. In this study, an E-Business Software Product Line Methodology based on SMEs characteristics was proposed using Software Product Line Engineering (SPLE) approach. In general, the proposed methodology was sufficient to describe the aspects needed in building a software development methodology. Aspects that had not been described were those related to software project management. The proposed methodology is useful for building an e-business application platform that can be customized based on SMEs characteristics.**

***Index Terms*— Customization; E-Business; Platform; SMEs; SPLE**.

I. INTRODUCTION

Despite being a key contributor to the economy in several countries, SMEs experience several challenges caused by several factors, such as social, economic, geographical, cultural, or other factors that are beyond their control. Among the factors that cause the low survival rate of SMEs, a factor that is considered critical is the lack of successful adoption and use of Information Technology [1], or what is known as e-business. The usage of e-business can help to make business procedures more efficient, which can help SMEs flourish. Several studies have stated that IT adoption in SMEs dependent on the size of the company [2]. However, this is contrary to [3], which states that the size of the organization does not guarantee the success of IT adoption. According to the survey, firms that have made IT a habit in their operations will find it easier to adopt IT than traditional organizations.. In addition, the field of services also affects the adoption of IT. Based on the very diverse characteristics of SMEs, an e-business solution is not a one-sized-fits-all solution. The proposed solution needs to consider two main stakeholders in the system to be built, namely SME actors and e-business application providers. In addition to these two stakeholders, a platform provider that connects SME actors and e-business application providers is also needed.

For SME actors, an e-business system is needed, which consists of various kinds of e-business applications according to the characteristics of the organization. Organizational characteristics can be viewed from several main factors, namely Environment, Organizational, and Technology[4]. Based on the organizational characteristics, SMEs can obtain recommendations for e-business applications that can support their organization's business processes. Through the main platform proposed in this study, SMEs can easily get customized e-business applications according to their organizational characteristics. For e-business application providers, providing different e-business applications for each SME is a challenge. If building different e-business applications always starts from scratch, then of course it will consume a lot of resources, both time, cost, and effort. Through the main platform proposed in this study, e-business application providers can customize applications that are built without having to start from scratch. This study is a further development of the Conceptual Model of the Use of E-Business by SMEs [5], using a complementary research model between Design Science Research (DSR) and Behavioral Science Research (BSR) [6].

II. SOFTWARE PRODUCT LINE ENGINEERING

SPLE stands for Software Product Line Engineering, and it is a methodology for developing software applications (software intensive systems and software products) that uses platforms and mass customization. The motivation of SPLE is to reduce product line development costs, improve quality, reduce production time, reduce maintenance efforts, overcome product line evolution,

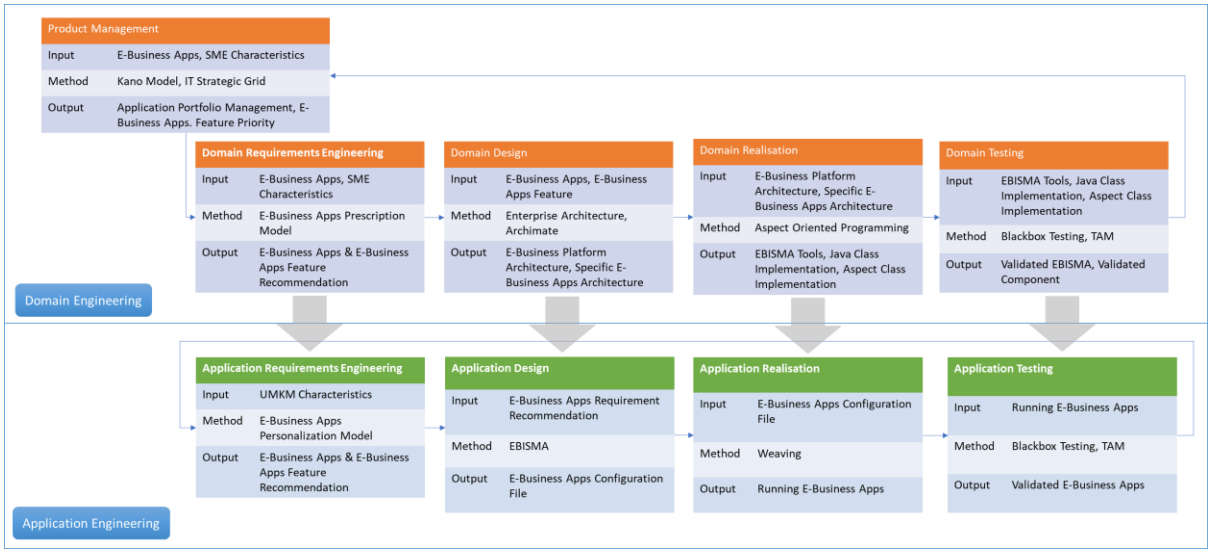

Fig. 1. The Proposed E-Business Software Product Line Methodology based on The Characteristics of SMEs

overcome complexity, increase cost estimates, and increase benefits for customers [7].

There are two development processes in SPLE, namely domain engineering and application engineering[7], which will be explained as follows:

1. Domain engineering process consists of five key subprocesses, namely product management, requirements engineering domain, design domain, realization domain, and testing domain. This process results in a platform that embraces the commonality of application and variability to support mass customization.
2. The application engineering process consists of subprocesses, namely application requirements engineering, application design, application realization, and application testing.

III. Design of E-Business Software Product Line Methodology

In this study, an e-business software product line methodology based on the SMEs characteristics was proposed using Software Product Line Engineering (SPLE) approach. SPLE has been proven as a methodology for building a wide range of software products and software-intensive systems at lower costs, in less time, and with better quality than other methodologies. The methodology proposed in this study is shown in Figure 1. The proposed methodology adopts the existing process stages in SPLE [8]. In general, there are two processes that exist in this methodology, namely domain engineering and application engineering. The proposed methodology is used to build an e-business application platform that can be customized based on the SMEs characteristics.

A. Domain Engineering

The process in domain engineering is responsible for building a reusable platform, so in this process it is necessary to determine the commonality and variability of the product line being built. In this subprocess, the platform built consists of 2 types, namely the

1. **Information System Platform E-Business**. It is the main platform consisting of various integrated e-business applications. This platform connects e-business application users with e-business application providers.
2. **Application Platform E-Business Specific**, which are applications part of the E-Business Information System. The e-business applications designed can be customized according to user requirements.

Domain engineering has 5 subprocesses, namely product management, domain requirements engineering, domain design, domain realization, and domain testing. At this stage, an analysis related to e-business applications was carried out in accordance with the characteristics of SMEs based on their clusters. The process carried out was to analyze the requirements of e-business applications by SMEs using the Kano Model, as had been done in previous research [9].

1. **Product Management**. In this process, identification of common and variable requirements was conducted to develop portfolio management. The process of E-Business Application Portfolio used survey data that has been carried out at the commonality & variability application-level e-business. At this stage, the results of a survey of e-business by SMEs have been obtained which are evaluated using the Kano Model. The evaluation results of the Kano Model are based on 6 categories, namely attractive, one-dimensional, must-be,

indifferent, reverse, and questionable mapped into four quadrants of the combination of the Kano Model and the Strategic Grid.

2. **Domain Requirements Engineering**. The entire list of requirements that have been collected up to this stage will then be recommended to SMEs based on their organizational characteristics using the Advanced UCF algorithm. Meanwhile, the features of each e-business can be determined based on the preferences of the SMEs requirements. To get preferences for these features, SMEs will be given a pair of functional and dysfunctional questions based on the Kano Model. Features included in the Attractive, One-Dimensional, and Must-Be applications e-business. The features in the classification are features that are considered needed by SMEs and can increase the level of satisfaction of SMEs with the e-business used.

3. **Domain Design**. The E-Business Platform Architecture can be used for applications needed today or can be developed for the development of applications needed in the future. The main architecture is designed to be pluggable, where the implemented e-business can be used according to the characteristics of SMEs. Different service providers can create e-business apps based on their areas of competence. Platform providers, service providers, and SME players are the three stakeholders in this architecture. Platform providers provide as a link between small businesses and service suppliers. The platform provider's service is an e-business application framework that may be used to create e-business apps. Service providers are companies that develop e-business solutions based on their expertise and experience in specific industries. Service providers can create e-business in the proposed architectural design, which is divided into four categories: sales & marketing, manufacturing & production, finance & accounting, and human resources. SMEs, on the other hand, are the end customers of the e-business services offered. SMEs can utilize the profiling tool to get e-business recommendations that fit their characteristics.

4. **Domain Realization**. The architectural design that has been generated in the Domain Design is implemented using the Aspect Oriented Programming. Each SME will be given application recommendations that match the characteristics of its organization. For this reason, we need an application outside of the main e-business that functions to assess organizational characteristics and provide recommendations for e-business that are in accordance with these characteristics.

B. Application Engineering

Process on application engineering is responsible for obtaining product line applications from platforms built on the domain engineering. This is done by taking advantage of the variability of the product line and ensuring that the variability aggregation process is correct according to the specific requirements of the application. At this stage, SMEs can get recommendations for e-business applications according to their characteristics. The multi-label classification approach had been carried out in the previous study [10], used to link an SME profile with several appropriate e-business applications.

There are several sub-processes in Application Engineering which will be explained as follows:

1. **Application Requirements Engineering**. In this subprocess, a list of e-businesses apps for SMEs is generated using the Advanced UCF algorithm. In addition, SMEs can also choose features in each e-business using the Kano Model. By using the Kano Model, SMEs are recommended to choose features that are in the Attractive, Must-Be, and One-Dimensional because they will help the organization's business processes and increase user satisfaction.

2. **Application Design**. To get recommendations for e-business that are in accordance with organizational characteristics, SMEs use the EBISMA application (Figure 2) as a tool. The recommendation results obtained from the EBISMA application are in the form of a list of e-business that are able to help SMEs to carry out their business processes. In addition to the list of e-business, EBISMA will also generate file containing the configurations required to produce e-business based on the specified features.

3. **Application Realization**. To get an application that can run based on customized requirements, it is necessary to configure the main class and aspect with weaving technique[11]. Additional features on the website implemented in the aspect. If these features are not needed, they can be removed by changing file the XML configuration. For users who want to use the Online Store Application can determine the combination of features that suit their requirements. By using this approach, different e-business apps can be built from the same platform. These applications are called the Software Product Family.

IV. Implementation and Analysis

In this study, the proposed methodology was implemented to build an application called EBISMA, which stands for E-Business for Indonesian SMEs. This application serves to assess the characteristics of the organization and provide recommendations for e-business applications in accordance with these characteristics. In addition, EBISMA also functions as a configurator that will produce e-business applications that are ready to be used by SME actors. The main view of the EBISMA Application is shown in Figure 2.

Fig. 2. EBISMA Application Main Page

In using EBISMA, SMEs are asked to answer several questions to assess the characteristics of their organization. The questions are divided into two stages, namely questions related to the identity of SMEs and questions related to the readiness of SMEs to implement e-business. After answering these questions, the system will provide recommendations for appropriate e-business applications. ESBIMA is an application that will generate a configuration file, which contains main classes and aspect classes that will be used to build specific e-business applications. The EBISMA configuration design is carried out using an Aspect Oriented Programming approach, as shown in Figure 3. A more detailed description of the AOP approach for implementing the EBISMA configuration has been presented in previous research [11].

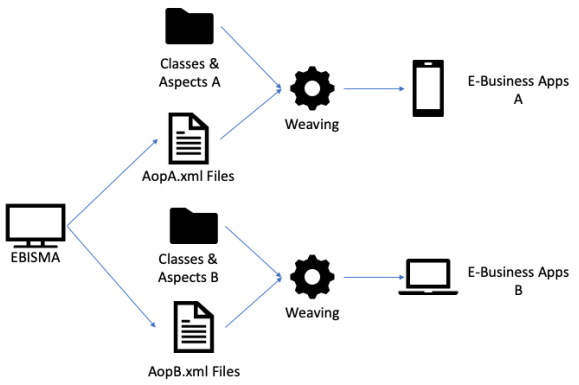

Fig. 3. E-Business Application Design with AOP [11]

V. Conclusions

At this stage, the proposed e-business software product line methodology was evaluated using a general framework for evaluating the software development methodology proposed by [12]. Implementing methodology evaluation framework can provide several advantages. Evaluation can help in understanding certain aspects of a methodology. Evaluation can also be used to compare methodologies, and thereby help users choose between several methodologies according to their desired goals. In addition, improvements can be accelerated using the evaluation results. In this section, the evaluation was carried out by assessing the criteria with a measurement matrix with a value of 0 to 6 based on [13].

In this section, an assessment of the existing process in the methodology is carried out. The proposed methodology does not focus on the management aspects of software development projects, which consist of risk management, people management, quality management, configuration management, and project scheduling. Another thing that is not focused on the methodology is the aspect of efficiency, both in terms of time and resources.

A. Modeling Language Features

A modeling technique is a set of models that describe a system at various levels of abstraction and explain its various aspects. In this section, an assessment is made regarding the modeling language used in the proposed methodology. Based on the results of the assessment, which are shown in **Error! Reference source not found.**, in general the proposed methodology covers most aspects of the assessment in terms of the modeling language. Things that have not been covered in the methodology are aspects of formalization. The proposed methodology uses UML and ArchiMate as its modeling language and the two modeling languages do not provide formalization aspects to symbolize semantics. Another aspect is related to handling inconsistencies that have not been provided by the methodology. The last aspect is related to the User Interface, which is not provided by the methodology.

TABLE I. ASSESSMENT OF MODELING LANGUAGE FEATURES

No.	Aspects	Criteria	Values	Basic
1	*Ease of*	*Understanding*	6	Modeling language using *Archimate* and UML which are easy to understand in general
2		*Usage*	6	Modeling language using *Archimate* and UML which are easy to use in general
3	*Precisenses*		6	Modeling language is considered appropriate and can avoid ambiguity
4	*Power of language*	*Formalism*	0	*Archimate* and UML do not provide formalization aspects to represent semantics
5		*Supported views*	6	*Archimate* includes *structural view*, *behavioral view*, and *functional view*
6		*Support for transformation logic*	6	Model provides design from analysis phase to implementation phase for programming code
7	*Complexity management model*	*Modularity*	6	The model is designed in a modular fashion, which consists of a *java* (*common*) and *aspect* (*variable*).
8		*Handling model inconsistencies*	3	The model is designed with a clear step flow to avoid *overlapping*. However, techniques for dealing with inconsistencies are not discussed
9	*Expressivenes*	*Static and dynamic aspects*	6	UML represents the structural (*static*) and *behavioral* (dynamic) aspects of the system
10		*Physical architecture of systems*	6	Architectural design is built using *Archimate*

B. Process Features

The process in the methodology has two roles, namely regulating and directing development from analysis to implementation, and enabling traceability of development by determining results and milestones. In this section, an assessment of the existing process in the methodology is carried out. Based on the results of the assessment, which are shown in **Error! Reference source not found.**, the proposed methodology does not focus on the management aspects of software development projects, which consist of risk management, people management, quality management, configuration management, and project scheduling. Another thing that is not focused on the methodology is the aspect of efficiency, both in terms of time and resources.

TABLE II. ASSESSMENT OF PROCESS FEATURES

No	Aspect	Sub-Aspect	Criteria	Value	Baseline
1	*Lifecycle*	*Work Unit*	*Coverage of generic phases*	6	The methodology addresses the generic phases of software development (requirements, analysis, design, implementation, testing)
2			*Smooth transition between phases*	6	transition between phases is clear and smooth
3			*Seamless transition between phases*	6	Artifacts produced between phases are quite clear to make a *seamless transition*
4			*Kind of lifecycle*	6	The type of SDLC used is *incremental*
5			*Workflow*	6	Each phase of the methodology consists of different activities and has clear guidelines
6		*Product*	*Adequacy*	6	Each phases of the methodology produce sufficient artifacts to be used in the next phase
7			*Consistency*	6	Each phase of the methodology produces artifacts that are consistent and complement other artifacts with minimum *overlapping*
8			*Supported view*	6	The artifacts from each phase of the resulting methodology consist of *structural*, *behavioral*, and *functional*
9			*Abstract ction levels*	6	Artifacts represent abstraction levels for systems, *packages*, *components*, and *objects*
10			*Tangibility/Testability/ Visibility*	6	Artifacts produced are *tangible*, tested, and understood
11			*Appropriate Documentation*	6	documentation can be generated from the methodology
12		*Role*	*User involvement*	6	Users are involved in all stages of the methodology clearly
13			*Roles specification*	3	*Roles* of users have not been specified in detail
14	*Management*	*Risk management*		0	The methodology does not discuss *risk management*
15		*People management*		0	The methodology does not discuss *people management*
16		*Quality management*	*Quality control*	0	The methodology does not discuss *quality control*

17			*Quality assurance*	0	The methodology does not address *quality assurance*
18		*Configuration management*		0	The methodology does not address *configuration management*
19		*Project scheduling*		0	The methodology does not address *project scheduling*
20	*Development context*	*Ease of*	*Understanding*	6	The development process is easy to understand
21			*Usage*	6	The development process easy to use in the *context* development *e-business software*
22		*Efficiency*	*Time*	0	Methodology does not address time efficiency
23			*Resources*	0	Methodology does not address resource efficiency
24		*Precision*	*Traceability*	6	The resulting artifacts have traceability to requirements
25			*Formalism*	0	The methodology does not address aspects of *formalism*
26			*Expressiveness*	6	Each phase of the methodology is clearly defined to avoid ambiguity
27			*Rationality*	6	Each phase of the methodology is defined rationally
28			*Completeness*	3	Each phase of the methodology is fully defined for all aspects, such as artifacts and processes. However, some aspects such as *roles* and resources have not been defined
29		*Maintainability*	*Modularity*	6	The model is designed in a modular manner, consisting of a *java* (*common*) and *aspect* (*variable*)
30			*Reusability*	6	The model is designed in a modular manner and can be reused
31			*Testability*	6	The methodology provides testing phases in *the engineering* and *application engineering*
32		*Complexity Management*	*Evolvability*	6	phase *Product Management* allows for product development
33			*Extensibility*	6	phase *Product Management* also allows for additional product features
34			*Distribution*	6	Methodology supports modeling and implementation of components that support distributed functionality
35			*Integrity*	6	Methodology provides an integrated architecture

C. Applicability Features

Another important aspect that needs to be evaluated is related to the application of the methodology in software development projects. In this section, an assessment is carried out on aspects of the application of the methodology, which includes aspects of pragmatic, marketability, and application limitations. Based on the results of the assessment, which are shown in Table 3, most aspects of the application of the methodology have been discussed in detail. Things that are not discussed in this methodology are aspects of team skills and the level of satisfaction of the development team. This is not the focus of the proposed methodology.

TABLE III. ASSESSMENT OF APPLICABILITY FEATURES

No	Aspect	Sub-Aspect	Criteria	Value	Basic
1	*Pragmatics*	*Adaptability*	*Size and complexity*	6	The methodology covers a fairly large and complex process
2			*Criticality*	6	The methodology has *impact* on the success of *E-Business Software Product Line*
3			*Scalability*	6	Methodology can be used on different project sizes
4		*Extant resources*	*Available information*	6	Information related to the methodology is clearly available
5			*Tools*	6	EBISMA *tools* are available to support the methodology
6		*Required resources*	*Team skills*	0	The methodology does not address aspects of *team skills*
7			*Platform suitability*	6	The methodology can be used on different platforms
8	*Marketability*	*User satisfiability*		6	Users are satisfied with the product resulting from the methodology
9		*Development team satisfiability*		0	The methodology does not address *development team satisfiability*
10	*Application constraints*	*Legal constraints*		6	The methodology does not limit legal aspects
11		*Technical constraints*		6	The methodology limits on the use of AOP (*Aspect Oriented Programming*) for design implementation
12		*Management constraints*		6	The methodology does not limit management aspects, such as management culture, etc.
13		*Environment constraints*		6	Methodology does not constrain the development environment

D. Type-specific Methodology Features

In this section, an assessment of the specific aspects provided by the proposed methodology is carried out. Based on the results of the assessment, as shown in Table 4, the proposed methodology meets the aspect of a type-specific, where the methodology uses the AOP concept to design and implement the model.

TABLE IV. TYPE-SPECIFIC METHODOLOGY FEATURES

No	Aspect	Value	Basic
1	*Concept-specific*	6	The methodology uses the AOP concept to design and implement the model
2	*Philosophy*	6	The methodology uses the AOP concept because AOP considers modularization, encapsulation, and *crosscutting concerns* as the main focus

Based on the results of all methodological assessments, in Table 5 summary of the results of the methodology assessment is presented. In general, the proposed methodology is sufficient to describe the aspects needed in building a software development methodology. Aspects that have not been described are those related to software project management. This aspect is a limitation in this research.

TABLE V. SUMMARY OF E-BUSINESS SOFTWARE PRODUCT LINE METHODOLOGY ASSESSMENT

Aspect			Assessment
Generic features	Modeling language		Partial coverage
	Process	Lifecycle	Partial coverage
		Management aspects	No coverage
		Development context	Partial coverage
	Applicability		Partial coverage
Specific features	Concept-specific		Acceptable coverage

VI. CONCLUSIONS

In this research, an e-business software product line methodology was produced. The research contribution refers to several issues for further development of the software product line engineering [7], which consists of domain specialization, software product line evolution, multiple product lines, and tool support. The applicative research contribution from this research is that SMEs can get recommendations for e-business applications that are in accordance with the characteristics of their organizations. In addition, the results of the analysis of commonality and variability can be a reference for e-business application developers to make the transition from single application development to product family development. E-business application developers can determine the priority of application development based on the commonality and variability analysis. E-business application developers also can take advantage of the e-business platform architecture to develop services that match their expertise.

The following will explain in more detail the research contributions which refer to several issues that are issues of further development of the Software Product Line Engineering[7].

A. *Domain Specialization*

The biggest challenge for research in Software Product Line Engineering (SPLE) is to specialize framework for an application domain. By specifying the SPLE framework for an application domain, a series of types can be generated to model domain-specific variation points, variants, variability dependencies, and constraint dependencies. To take full advantage of domain specialization, it is necessary to define a domain-specific modeling language to define software development artifacts [7].

In this study, the proposed SPLE methodology is devoted to building a Software Product Line for e-business domain for SMEs. To be able to support the use of e-business by SMEs, several e-business apps is needed that function to assist SME business processes for certain functional areas. Sales & marketing functional areas need e-business in the form of Online Store, E-marketplace, etc. Meanwhile, in the human resources functional areas, e-business applications are needed in the form of payroll, management personnel, etc. The use of the e-business will certainly help SME business processes. Therefore, in this study, a recommendation model for e-business applications was produced based on the characteristics of SMEs. Based on the results of these recommendations, SMEs can use e-business applications that are in accordance with their business characteristics.

B. *Evolution of the Software Product Line*

In the development of the Software Product Line, software developers are not only faced with the evolution of the Software Product Line over time, but also with the existence of different variants at the same time (variability in space). Managing the evolution of Software Product Line artifacts over time and ensuring the consistent integration of changes in all affected Software Product Lines is a key research challenge. Developing and validating comprehensive techniques that support the management of Software Product Line artifact evolution over time and the management of variability in artifacts is also a research issue open to further study [7].

In the Product Management subprocess of the proposed E-Business Software Product Line Development Methodology, a product roadmap has been generated, which is represented in the form of an application portfolio. In the application portfolio, it can be determined which e-business applications are useful for the present and which e-business applications will be useful in the future. The resulting application portfolio combines the McFarland matrix with the Kano Model. By combining these two theories, it was hoped that it will provide appropriate results from grouping e-business applications, both in terms of user preferences and in terms of their contribution to business processes. In addition, the results of the combined analysis of McFarlan and Kano will support the evolutionary management of the E-Business Software Product Line.

C. *Multiple Product Lines*

In some domains, there is a need to manage variability across product lines. However, solutions for defining and managing variability across product lines and across all software development artifacts remain unclear. Managing variability across product lines is even more challenging if the product lines are owned by different companies [7].

In the Product Management stage of the proposed E-Business Software Product Line Development Method, an e-business application portfolio has been obtained which functions as a product roadmap, which is used for application planning and management based on current and future contributions to the business. In addition to the e-business application portfolio, at the Product Management stage, priority features will be developed for e-business applications. The priority of feature building starts with basic features, then performance features, entertainment features, and ends with potential features.

D. *Tool Support*

High-quality software engineering in the industry requires adequate tool support. However, for most aspects of Software Product Line Engineering, adequate tool support does not exist. For example, support for tools used to manage variability across all development artifacts, or to manage linkages between product line applications and domain artifacts is very weak. Building seamless tool support to manage variability and development artifacts in process domains and application engineering is a challenge for future research [7].

In this study, a prototype was built named EBISMA, which stands for E-Business for Indonesian SMEs. This application serves to assess the characteristics of the organization and provide recommendations for e-business applications in accordance with these characteristics. In addition, EBISMA also functions as a configurator that will produce e-business applications that are ready to be used by SME actors. In EBISMA application, SMEs are asked to answer a number of questions to assess the characteristics of their organization. The questions are divided into two stages, namely questions related to the identity of SMEs and questions related to the readiness of SMEs to implement e-business. After answering these questions, the system will provide recommendations for appropriate e-business applications using the Advanced UCF algorithm.

REFERENCES

[1] M. Bernaert, G. Poels, and M. Snoeck, Information Systems for Small and Medium-sized Enterprises. 2014. doi: 10.1007/978-3-642-38244-4.

[2] V. A. Mabert, A. Soni, and M. A. Venkataramanan, "The impact of organization size on enterprise resource planning (ERP) implementations in the US manufacturing sector," Omega (Westport), vol. 31, no. 3, pp. 235–246, 2003, doi: 10.1016/S0305-0483(03)00022-7.

[3] P. O. H. Putra, H. B. Santoso, and Z. A. Hasibuan, "Antecedents and patterns of e-Business adoption among small and medium enterprises (SMEs)," 2017 Second International Conference on Informatics and Computing (ICIC), pp. 1–6, 2017, doi: 10.1109/IAC.2017.8280638.

[4] P. O. H. Putra, Z. A. Hasibuan, and P. O. A. Sunarya, "Depicting E-business application adoption for small and medium enterprises (SMEs): The e-business triangle model," 3rd International Conference on Computing, Engineering, and Design, ICCED 2017, vol. 2018-March, pp. 1–6, 2018, doi: 10.1109/CED.2017.8308135.

[5] P. O. H. Putra and H. B. Santoso, "Contextual factors and performance impact of e-business use in Indonesian small and medium enterprises (SMEs)," Heliyon, vol. 6, no. 3, p. e03568, 2020, doi: 10.1016/j.heliyon.2020.e03568.

[6] A. Hevner and S. Chateerjee, Design Research in Information Systems : Theory and Practive, vol. 28. 2012. doi: 10.1007/978-1-4419-6108-2.

[7] K. Pohl, G. Böckle, and F. van der Linden, Software Product Line Engineering. Foundations, Principles, and Techniques. New York: Springer-Verlag Berlin Heidelberg, 2005. doi: 10.1007/978-3-642-36583-6_1.

[8] K. Pohl, G. Böckle, and F. van der Linden, Software Product Line Engineering: Foundations, Principles, and Techniques. 2005.

[9] N. M. S. Iswari, E. K. Budiardjo, and Z. A. Hasibuan, "E-business applications recommendation for SMES using advanced user-based collaboration filtering," ICIC Express Letters, vol. 15, no. 5, pp. 517–526, 2021, doi: 10.24507/icicel.15.05.517.

[10] N. Made, S. Iswari, E. K. Budiardjo, H. B. Santoso, and Z. A. Hasibuan, "E-Business Application Recommendation for SMEs based on Organization Profile using Random Forest Classification," 2019 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), pp. 522–527, 2019.

[11] N. Made Satvika Iswari, E. K. Budiardjo, and Z. A. Hasibuan, "Aspect oriented programming approach for variability feature implementation in software product line engineering," 2020 5th International Conference on Informatics and Computing, ICIC 2020, pp. 4–8, 2020, doi: 10.1109/ICIC50835.2020.9288558.

[12] S. Hesari, H. Mashayekhi, and R. Ramsin, "Towards a general framework for evaluating software development methodologies," Proceedings - International Computer Software and Applications Conference, pp. 208–217, 2010, doi: 10.1109/COMPSAC.2010.69.

ISSN 2085-4552

Teakwood Grade Identification with GLCM and K-NN with Adaboost Optimization

(Case Study at KPH Cepu)

Nirma Ceisa Santi[1] , Hastie Audytra[2]
[1,2] University of Nahdlatul Ulama Sunan Giri, Bojonegoro, Indonesia
[1] nirmaceisa@unugiri.ac.id, [2] hastie@unugiri.ac.id

Accepted 24 June 2022
Approved 10 July 2022

***Abstract*— Teak is one type of tree that has many functions and uses. Teak wood have a very high quality to choose as resource for the manufacture of home furniture such as tables, chairs, cabinets, and others. But middle testers (Perhutani staff) who test the quality of wood grade have limitations if the classification uses the five senses of sight and there are still many furniture entrepreneurs who are often mistaken about teak wood quality assessment. This resulted in a lack of quality grade teak wood used as raw material for making home appliances or for furniture and commerce in Perhutani Corporation, especially KPH Cepu. The teak wood image data is then acquired through preprocessing data ready to be processed. By using GLCM as an image feature extraction both training data and testing data. After the image characteristics are obtained, the image is classified by the K-Nearest Neighbor method with adaboost optimization. The final result is obtained in the form of wood grade quality classification namely grade A, B, C and D according to the class.**

***Index Terms*— Classification; GLCM; K-nn; Teakwood.**.

I. INTRODUCTION

Teak is a type of woody tree that has high quality. A large tree, straight trunk and can grow to a height of 30-40 m straight towering. Teak has the biological name *Tectona grandis L.f.* In today's industrial bussines, teak wood is mostly made into veneers to coat the outer skin of expensive plywood, besides that, parquet is also made for floor coverings, even twigs that cannot be used as raw material for processed furniture can still be used as fuel because teak can produce high degrees of heat high enough so that in the past it could be used as fuel in steam locomotives. Not only domestic commodities, Indonesia also exports processed teak wood in the form of outdoor furniture [1]. The largest teak production in Central Java is in Perum Perhutani Kph Cepu which has an area of 33,017.29 Ha. The teak forest area is located in Blora Regency and Tuban Regency and Bojonegoro Regency. Production management in the forest area in the Cepu KPH area is divided into two Forest Management Sub Units (SKPH), namely SKPH north Cepu Utara and SKPH South Cepu. The demand for teakwood production at Kph Cepu is the highest to date. Kph Cepu carries out production activities in accordance with the legal basis of central regulations. The production function is more focused on logging and testing the quality of teak wood. Teak wood grouping serves to determine with grades A, B, C, and D. This test was carried out by the staff of Perum Perhutani Kph Cepu Middle examiner. So far, the intermediate examiners have been doing it manually based on the five senses of sight with teak logs that have been classified by the length and diameter by the loggers. The grouping or classification of the quality of teak wood types aims to facilitate the sale of teak on the Perum Perhutani trading portal at Tokoperhutani.com and facilitate the fulfillment of buyer needs.

Identification of teak wood quality based on the image of the vertical cross-section is still done manually and takes a long time so that the production of teak wood for sale is not optimal. Many studies have been carried out by completing the classification of data record numbers taken through measurement actions during felling [2]. In the processed image data, texture feature extraction is the most recommended considering the homogeneity of the data features that are still in image form. The steps in feature extraction can use Gray level co-occurrence matrix which is a statistical method used as texture feature extraction. After the feature extraction results are complete, the classification is not only processed. The classification process is carried out to group images that have been identified using GLCM into classes that have been determined by processing data information. One of the specific problems with texture images is that the image edges are considered as textures in the search for image features.

R. Qayyum, K. Kamal ,T. Zafar and S. Mathava uses the Particle Swarm Optimization algorithm to train a feed forward neural network used to classify defects using GLCM-based features. The proposed technique shows promising results for the classification of wood defects. Means found that the Square Error of the network for the training dataset was 0.3483, whereas, for the test dataset, the accuracy was found to be 78.26% [3].

While another study by Stefanus Santosa, R. A. Pramunendar, D. P. Prabowo, and Yonathan P. Santosa proved that GLCM displays the adjacency relationship between pixels in the image. The angles used are 0°, 45°, 90° and 135°. In the experiments carried out, it was produced 99.14% with the combination of BPNN GA [4].

Previous research can be concluded if the use of Gray Level Co-occurrence Matrix (GLCM) and Euclidean Distance as a calculation of the distance between images that can produce accuracy accurately. In this study, texture feature extraction for teak wood that has the possibility to be used is Gray Level Co-occurrence Matrix (GLCM) with 5 features, namely Contrast, Energy, Correlation, Homogeneity, and Entropy. As for the classification used is the K-NN classifier because the use of training data to be tested is quite a lot so it is expected to have a sufficient and good level of accuracy. To cover the weakness of the k-nn algorithm, adaboost optimization is used to reduce the presence of weak data or data outliers so that all data can be processed optimally. The data classified is of the type of teak wood texture image data which does not have information that can be processed for classification and ith the K-nn algorithm method, there are still some missing values in the k-nn classification model so that the classification algorithm's performance is not optimal. Processing data classified as teak texture image data in order to have information that can be processed for classification as data information using Gray level co-occurance matrix for feature extraction on teak board texture images. The creation of a grade classification model for teak wood boards using k-nn with adaboost optimization for classification and improving the performance of the classification algorithm.

II. METHOD

A. *Extraction Texture Feature*

Sentiment analysis is one of the methods used to Rizky Andhika Surya, Abdul Fadlil, Anton Yudhana explained that the research he did using the image extraction method which was converted through the gray level coocurance matrix method could be used as raw material for batik image classification. This is done because it is felt that the results of GLCM can show differences between images with different texture [5].

B. *Gray Level Co-occurrence Matrix*

Gray Level Co-occurrence Matrix (GLCM) is a matrix whose elements are the number of pairs of pixels that have a certain brightness level, where the pixel pairs are separated by distance and angle. Angle orientation is formed based on four corner directions, namely, 0°, 45°, 90° and 135°, and the distance between pixels is 1 pixel [5]. The input value from GLCM is a matrix which is a representation of the grayscale image, the output from GLCM is a co-occurrence matrix which we can then extract characteristics based on second-order statistical feature parameters such as contrast, correlation, homogeneity and energy.

C. *Data Mining*

Data mining has an important function in this research. data mining processes teak wood chip data to be able to make the information needed and increase knowledge by users of this research. Basically, data mining has four main functions in processing the teak wood image that has been converted through GLCM. The main thing is to predict, including where the testing data has been taken, then to describe what information is in the processed data, then to classify the data based on predetermined groups and finally to see if the data is associated with data that is close to its characteristics.

Classification is the process of finding a model or function that describes and differentiates data into classes. Classification involves the process of examining the characteristics of an object and assigning the object to one of the predefined classes.

D. *Adaboost in KNN*

The K-Nearst Neighbor classification works based on an analogy, where test and training data are compared and conclusions are drawn based on the similarity of the data generated by the comparison [6]. The calculation is carried out based on the distance (closeness of the data) which is then known as Euclidean Distance

$$d_{ik} = \sqrt{\sum_j^m (Cij - Ckj)^2} \qquad (2.1)$$

Where :

Cij = training data

Ckij = testing data

one of the supervised algorithms for data mining with a classification model function is the AdaBoost Method. Initially this algorithm was applied to the regression model, along with the rapid development of computer technology, this method can also be applied to other statistical models. The adaBoost method is an ensemble technique using the loss function of the exponential function to improve the accuracy of the predictions made. Basically the boosting method (AdaBoost) can increase the accuracy in the classification and prediction process by generating a combination of a model, but the results of the classification or prediction chosen are the model that has the greatest weight value. The adaptive boosting method has been reported as a meta-technique to overcome the class imbalance problem.

Adaboost and its variants are very successful in application to domains because of their strong

theoretical basis, accurate predictions and great simplicity. In theory adaboost functions as a performance optimizer for the k-nn classification algorithm so that its performance can be maximized. In the case of an unbalanced data set, the application of boosting will not change the structure of the dataset, which means that the condition of the dataset remains in an unbalanced form.

III. Result and Discussion

A. Data Collection

The main data in this study is the image data of teak wood boards acquired by intermediate examiners (experts) in the amount of 20 wooden boards for Grade A, 20 wooden boards for grade B, 20 wooden boards for Grade C and 20 wooden boards for grade D, respectively. each taken from 4 angles, namely 0°, 45°,90°,135°. In total there are 316 data for training data and 20 wooden boards as testing data with 4 corners also producing 80 test data.

Fig. 1. Picture of Main Data

There are 4 classes in the research that will be tested, namely grade A which indicates the grade of wood that has the best durability, density, and humidity which must be valued the most expensively by the seller, in this case Perum Perhutani Kph Cepu, then Grade B shows that durability, density, and humidity good enough with no wood defects or cambium sources. Grade C is a wood plank grade that has sufficient durability, density, and moisture and is in the category of being able to be used as raw material for furniture or equipment. While the last grade is grade D which has durability, density, and moisture. It can still be used for raw materials but with the condition of the wood. which must be adjusted to the needs (the load to be borne by the strength of the wood). Data collection techniques are taken directly or commonly referred to as Image Acquisition. This process captures or scans an analog image using a tool to obtain a digital image that has been determined by PUSBANGHUT. When taking photos, the teak wood is placed parallel and 15 cm from the recording device (camera). After that the data is collected in the form of images [7].

B. Preprocessing

Eliminating the Background, the original image data that has been obtained with different backgrounds and less same angles will be cut off the background with the aim of simplifying calculations in the feature extraction process. Then equalize pixel size piece per piece. The image data resulting from the uniformity of the background will be equalized with the pixel size with the aim of simplifying calculations during the feature extraction process. The pixel size of the image data will be standardized to 400 x 406 pixels. After second step is converting image to grayscale. Converting an image to grayscale can be done by taking all the pixels in the image then the color of each pixel will be retrieved information about the 3 basic colors, namely red, blue and green (via the colortoRGB function), these three basic colors will be added up and then divided by three to get the value average. This average value will be used to color the image pixels so that the color becomes grayscale.

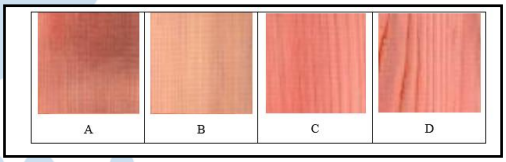

Fig. 2. Image Data After Preprocessing

C. Classify with Knn

K-nn calculation is an algorithm that functions to classify some learning data or training data. The training data in this study is in the form of images that have been extracted to produce image characteristics in the form of numbers. The numerical data from the feature extraction will be projected with some test data. The training data sample used in this manual calculation contains 19 training data and 1 data whose class/label is unknown and is called testing data. The training data used are 4 class A data, 5 class B data, 6 class C data and 4. class D data. Because the training data is even, it is recommended to use odd k. The criteria record is c90 means correlation on 90°angle etc.

TABLE I. Sample Training Data

C90	C135	L0	L45	L90	L135	Label
0,040414	0,076267	0,91083	0,88181	0,93734	0,88094	A
0,03165	0,051598	0,87263	0,82111	0,89269	0,82069	A

0,02443	0,062726	0,88888	0,87363	0,94998	0,87169	A
0,038466	0,088221	0,80603	0,77591	0,90258	0,77488	A
0,043501	0,084148	0,83501	0,80578	0,89541	0,79785	B
0,029763	0,046631	0,65786	0,51001	0,71078	0,52107	B
0,043065	0,15724	0,62025	0,59862	0,89714	0,59345	B
0,024143	0,043709	0,74565	0,68661	0,83128	0,68507	B
0,023148	0,047316	0,91926	0,90381	0,95439	0,90679	B
0,041837	0,10722	0,76995	0,74034	0,89948	0,74183	B
0,042542	0,10765	0,77484	0,7497	0,908	0,76786	C
0,007576	0,040417	0,63174	0,63349	0,92763	0,61665	C
0,005649	0,039971	0,26669	0,26773	0,8938	0,25417	C
0,046941	0,13563	0,53493	0,49927	0,82617	0,50017	C
0,023442	0,075833	0,85024	0,83771	0,95101	0,84152	C
0,016924	0,051342	0,57847	0,54986	0,85269	0,55322	D
0,034852	0,080501	0,79101	0,73985	0,9085	0,78819	D
0,024959	0,070892	0,83014	0,81902	0,93249	0,80603	D
0,03657	0,09994	0,83019	0,81694	0,93241	0,81452	D

TABLE II. SAMPLE TESTING DATA

0,14777	0,15522	0,043065	0,15724	0,62025	…	0,9322
0,088611	0,11898	0,048778	0,083889	0,86263	…	0,95806

1. Finding the Euclidean Distance

Calculating the Euclidean distance is calculating the root in the test data minus the training data and then squared on each training data. After calculating it produces the table below.

2. Sorting data by 5 nearest neighbors

The determination of the value of K is determined as k = 5, meaning that 5 data records will be taken with the closest distance value.

3. After ranking the 5 smallest data and then checking the class on the nearest neighbor.

TABLE III. KNN WITH K=5

ID	Label	D	RANK
2A	A	0,110198	3
4A	A	0,110253	4
162C	C	0,156476	5
166C	C	0,078499	1
320D	D	0,100962	2

From this data processing, the closest neighbor sequence is data with data id 166C which has a label C, in sequence 2 there is data with id 320D with label D. In the order of closest distances, sequences 3 and 4 with test data are data labeled A with data id 2A and 4A. In the last order, there is data 162C with label C. From the data processing, the test data has the majority of proximity to data labeled A and C.

D. Optimation with Adaboost

In the adaboost optimization process, the same sample data will be used with the data above in the knn process [8].

- The first stage is to determine the sample weight of the data. Sample weights are selected from 1 divided by the total number of datasets. There are 20 datasets used, so the weight on the sample records is 1/20.
- The next stage is to determine the total number of possible errors that can occur. And errors that may occur 1/20.
- The third step is to calculate the performance value.
- The last stage is normalization of weight.

E. Modeling Process

1. Experiment without Adaboost

The next stage is to apply the KNN algorithm in the workspace. First click and put the training data into the blankpage, test data and look for the KNN algorithm and apply the model and performance for the confusion matrix. Search can be done by typing in the toolbar. After everything is there and complete, connect the training data to the KNN algorithm as shown below.

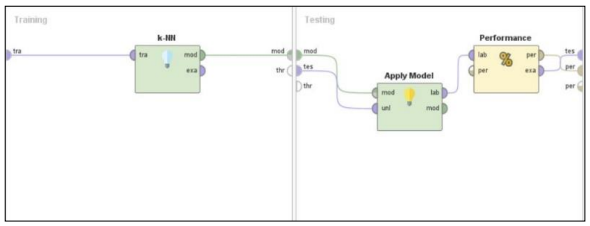

Fig. 3. Rapid Miner Knn

2. Experiment with Adaboost

In implementing this adaboost, special attention needs to be paid to the rapid miner step. The important thing is the implementation of adaboost for optimization of the training data rule attributes and the k-nn algorithm in the workspace.

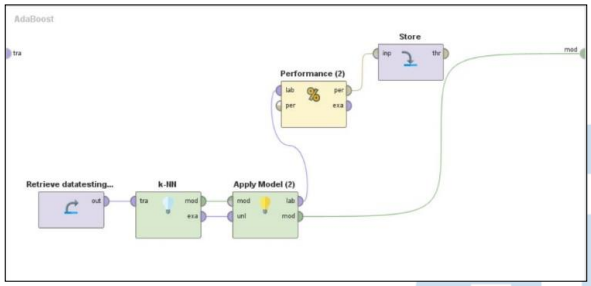

Fig. 4. Rapid Miner with Adaboost

Details of the experimental value of k on knn adaboost from numbers 5 – 30, the greatest accuracy value is obtained at the value of k = 5.

TABLE IV. ACCURACY COMPARISON KNN WITH KNN OPTIMIZATION ADABOOST

K	KNN	KNN & adaboost
5	80,31	91,56
7	67,50	91,56
9	69,06	83,44
11	60,94	76,56
13	62,19	66,88
15	59,69	64,06
17	59,69	66,25
19	59,19	64,06
21	61,25	65
23	60,62	64,06
25	58,75	63,44
27	58,13	65,31
29	55,93	64,69

IV. CONCLUSIONS

To improve the accuracy of the Knn algorithm in the process of classifying the quality of wood board grades at Kph Perhutani Cepu which will be forwarded to the District Research and Development Center. Semarang, data processing is carried out with the knn process optimized with adaboost. This is because Adaboost can improve accuracy in data processing with the Knn algorithm. The k-nearest neighbor algorithm produces 80.31% accuracy results, while adaboost optimization can increase the accuracy to almost 100% with an increase of about 19.9%.

TABLE V. ACCURACY WITH KNN ALGORITHM

KNN	80,31%					
	True. A	True. B	True. C	True. D		Class Precission
Pred. A	**76**	21	12	10		63,87%
Pred. B	0	**57**	5	6		83,82%
Pred. C	4	0	**63**	3		90,00%
Pred. D	0	2	0	**61**		96,83%
Class Recall	95,00%	71,25%	78,75%	76,25%		

TABLE VI. ACCURACY KNN WITH ADABOOST ALGORITHM

KNN + adaboost	100,00%					
	True. A	True. B	True. C	True. D		Class Precission
Pred. A	**80**	0	0	0		100,00%
Pred. B	0	**80**	0	0		100,00%
Pred. C	0	0	**80**	0		100,00%
Pred. D	0	0	0	**80**		100,00%
Class Recall	100,00%	100,00%	100,00%	100,00%		

From the data above, the boosting method, namely adaboost, has been proven to increase the accuracy of the KNN algorithm on wood sample data.

In the cross validation experiment, trials have been carried out with different k values, namely from numbers 10 – 30 with a value of k at knn k = 5 in the experiment. So that the difference in accuracy obtained is conveyed in the following chart.

So we can conclude that the addition of adaboost optimization in k-NN processing can increase accuracy so that the testing data input process can provide more precise results than using only the knn classification algorithm.

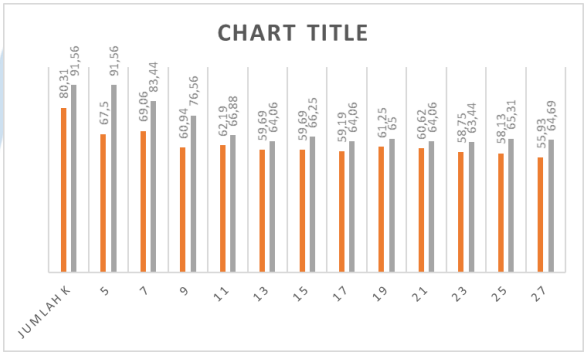

Fig. 5. Comparasion of Accuracy by Knn and Knn Adaboost

This research contributes ideas for the development of business processes in Perhutani, especially Kph Cepu, especially in the process of buying and selling wooden boards. A good and correct dataset plays an important role in the accuracy produced, so that in future research it is necessary to pay attention to taking datasets in both numbers and are expected to update the data set.

REFERENCES

[1] K. CEPU, Pengelolaan Lingkungan Perum Perhuani, Blora, 2014.

[2] S. R. Widodo, "Pemanfaatan Ciri Grey Level CoOccurance Matrix (GLCM) Citra Batik," Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, vol. Nov, pp. 5769 - 5776, 2018.

[3] K. T. Z. a. S. M. R.Qoyyum, "Wood defects Classification Using GLCM Based Feature and PSO Trained Neural Network," IEEE, 2016.

[4] R. P. D. P. S. Santosa, "Wood type classification using back Propagation," IAENG International Journal of Computer Science, vol. 2, p. 46, 2019.

[5] A. a. A. Y. R.A Surya, "Ekstraksi Ciri Metode GLCM dan Filter Gabor untuk Klasifikasi Citra Batik Peklaongan," Jurnal Informatika : Jurnal Pengembangan IT, vol. 02 no 02, no. ISSN : 2477-5124, 2017.

[6] Informatikalogi, "Algoritma K-Nearest Neighbor," 2017. [Online]. Available: https://informatikalogi.com/algoritma-k-nn.

[7] PERHUTANI, "TATA KELOLA PERUM PERHUTANI," KPH CEPU.

[8] K. W. a. Y. S.Y Mohamed, "Cost- Sensitive Boosting," pp. 3358 - 3378, 2007.

[9] A. K. a. s. ajib, "Sistem Pendukung keputusan produk olahan kayu jati menggunakan naive bayes".

[10] K. S. a. A.Remesh, "Efficient CBIR Technique Using Color, Texture and Histogram," Internationa; jurnal of engineerin technologi and engineering, Vols. 41 - 44, p. 2, 2015.

[11] E.Acar, "Extraction of texture features from local iris areas by GLCM and iris recognition system based on KNN," European Journal of technic, vol. 6, 2016.

[12] R. w. a. C. Fitriyah, "Perbandingan Akurasi Klasifikasi Citra kayu jati menggunakan Naive Bayes dan KNN," JEPIN, 2019.

[13] J. A. A. R. M.I Fatturahman, "Sistem Klasifikasi Kualitas Kayu Jati berdasarkan tekstur dengan JST menggunakan GLCM," e-Proceding engineering , vol. 6, p. 9537, 2019.

[14] K. S. a. M. Putra, "Klasifikasi Citra Mamogram menggunakan metode K-Means, GLCM, dan Support Vector Machine(SVM)," MERPATI, p. 2252, 2018.

[15] S.Sudrajat, "Deteksi Penyakit Gigi Dengan Menggunakan Metode Filter Gabor dan K-Nearest Neighbour Berdasarkan tekstur citra," 2018.

[16] s. l. a. K. w.song, "Hyperspectral image classification based on knn sparse," IGARSS, no. 978-1-5090-3332-4/16, 2016

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, Ultimatics : Jurnal Teknik Informatika 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: **ultimatics@umn.ac.id**

Paper Title

Subtitle (if needed)

Author 1 Name[1], Author 2 Name[2], Author 3 Name[2]
[1] Line 1 (of affiliation): dept. name of organization, organization name, City, Country
Line 2: e-mail address if desired
[2] Line 1 (of affiliation): dept. name of organization, organization name, City, Country
Line 2: e-mail address if desired

Accepted on mmmmm dd, yyyy
Approved on mmmmm dd, yyyy

***Abstract*—This electronic document is a "live" template which you can use on preparing your ULTIMATICS paper. Use this document as a template if you are using Microsoft Word 2007 or later. Otherwise, use this document as an instruction set. Do not use symbol, special characters, or Math in Paper Title and Abstract. Do not cite references in the abstract.**

***Index Terms*—enter key words or phrases in alphabetical order, separated by semicolon (;)**

I. INTRODUCTION

This template, modified in MS Word 2007 and saved as a Word 97-2003 document, provides authors with most of the formatting specifications needed for preparing electronic versions of their papers. Margins, column widths, line spacing, and type styles are built-in here. The authors must make sure that their paper has fulfilled all the formatting stated here.

Introduction commonly contains the background, purpose of the research, problem identification, and research methodology conducted by the authors which been describe implicitly. Except for Introduction and Conclusion, other chapter's title must be explicitly represent the content of the chapter.

II. EASE OF USE

A. Selecting a Template

First, confirm that you have the correct template for your paper size. This template is for ULTIMATICS. It has been tailored for output on the A4 paper size.

B. Maintaining the Integrity of the Specifications

The template is used to format your paper and style the text. All margins, column widths, line spaces, and text fonts are prescribed; please do not alter them.

III. PREPARE YOUR PAPER BEFORE STYLING

Before you begin to format your paper, first write and save the content as a separate text file. Keep your text and graphic files separate until after the text has been formatted and styled. Do not add any kind of pagination anywhere in the paper. Please take note of the following items when proofreading spelling and grammar.

A. Abbreviations and Acronyms

Define abbreviations and acronyms the first time they are used in the text, even after they have been defined in the abstract. Abbreviations such as IEEE, SI, MKS, CGS, sc, dc, and rms do not have to be defined. Abbreviations that incorporate periods should not have spaces: write "C.N.R.S.," not "C. N. R. S." Do not use abbreviations in the title or heads unless they are unavoidable.

B. Units

- Use either SI (MKS) or CGS as primary units (SI units are encouraged).
- Do not mix complete spellings and abbreviations of units: "Wb/m2" or "webers per square meter," not "webers/m2." Spell units when they appear in text: "...a few henries," not "...a few H."
- Use a zero before decimal points: "0.25," not ".25."

C. Equations

The equations are an exception to the prescribed specifications of this template. You will need to determine whether or not your equation should be typed using either the Times New Roman or the Symbol font (please no other font). To create multileveled equations, it may be necessary to treat the equation as a graphic and insert it into the text after your paper is styled.

Number the equations consecutively. Equation numbers, within parentheses, are to position flush right, as in (1), using a right tab stop.

$$\int_0^{r_2} F(r,\phi)\, dr\, d\phi = [\sigma r_2 / (2\mu_0)] \qquad (1)$$

Note that the equation is centered using a center tab stop. Be sure that the symbols in your equation have been defined before or immediately following the

equation. Use "(1)," not "Eq. (1)" or "equation (1)," except at the beginning of a sentence: "Equation (1) is ..."

D. Some Common Mistakes

- The word "data" is plural, not singular.
- The subscript for the permeability of vacuum μ_0, and other common scientific constants, is zero with subscript formatting, not a lowercase letter "o."
- In American English, commas, semi-/colons, periods, question and exclamation marks are located within quotation marks only when a complete thought or name is cited, such as a title or full quotation. When quotation marks are used, instead of a bold or italic typeface, to highlight a word or phrase, punctuation should appear outside of the quotation marks. A parenthetical phrase or statement at the end of a sentence is punctuated outside of the closing parenthesis (like this). (A parenthetical sentence is punctuated within the parentheses.)
- A graph within a graph is an "inset," not an "insert." The word alternatively is preferred to the word "alternately" (unless you really mean something that alternates).
- Do not use the word "essentially" to mean "approximately" or "effectively."
- In your paper title, if the words "that uses" can accurately replace the word using, capitalize the "u"; if not, keep using lower-cased.
- Be aware of the different meanings of the homophones "affect" and "effect," "complement" and "compliment," "discreet" and "discrete," "principal" and "principle."
- Do not confuse "imply" and "infer."
- The prefix "non" is not a word; it should be joined to the word it modifies, usually without a hyphen.
- There is no period after the "et" in the Latin abbreviation "et al."
- The abbreviation "i.e." means "that is," and the abbreviation "e.g." means "for example."

IV. Using the Template

After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention as below

ULTIMATICS_firstAuthorName_paperTitle.

In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper. Please take note on the following items.

A. Authors and Affiliations

The template is designed so that author affiliations are not repeated each time for multiple authors of the same affiliation. Please keep your affiliations as succinct as possible (for example, do not differentiate among departments of the same organization).

B. Identify the Headings

Headings, or heads, are organizational devices that guide the reader through your paper. There are two types: component heads and text heads.

Component heads identify the different components of your paper and are not topically subordinate to each other. Examples include ACKNOWLEDGMENTS and REFERENCES, and for these, the correct style to use is "Heading 5."

Text heads organize the topics on a relational, hierarchical basis. For example, the paper title is the primary text head because all subsequent material relates and elaborates on this one topic. If there are two or more sub-topics, the next level head (uppercase Roman numerals) should be used and, conversely, if there are not at least two sub-topics, then no subheads should be introduced. Styles, named "Heading 1," "Heading 2," "Heading 3," and "Heading 4", are prescribed.

C. Figures and Tables

Place figures and tables at the top and bottom of columns. Avoid placing them in the middle of columns. Large figures and tables may span across both columns. Figure captions should be below the figures; table heads should appear above the tables. Insert figures and tables after they are cited in the text. Use the abbreviation "Fig. 1," even at the beginning of a sentence.

TABLE I. TABLE STYLES

Table Head	**Table Column Head**		
	Table column subhead	***Subhead***	***Subhead***
copy	More table copy		

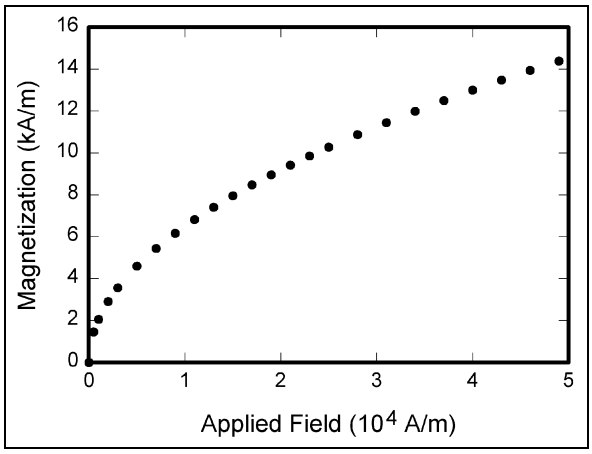

Fig. 1. Example of a figure caption

V. Conclusion

A conclusion section is not required. Although a conclusion may review the main points of the paper, do not replicate the abstract as the conclusion. A conclusion might elaborate on the importance of the work or suggest applications and extensions.

Appendix

Appendixes, if needed, appear before the acknowledgment.

ACKNOWLEDGMENT

The preferred spelling of the word "acknowledgment" in American English is without an "e" after the "g." Use the singular heading even if you have many acknowledgments. Avoid expressions such as "One of us (S.B.A.) would like to thank ... ." Instead, write "F. A. Author thanks ... ." You could also state the sponsor and financial support acknowledgments here.

References

The template will number citations consecutively within brackets [1]. The sentence punctuation follows the bracket [2]. Refer simply to the reference number, as in [3]—do not use "Ref. [3]" or "reference [3]" except at the beginning of a sentence: "Reference [3] was the first ..."

Number footnotes separately in superscripts. Place the actual footnote at the bottom of the column in which it was cited. Do not put footnotes in the reference list. Use letters for table footnotes.

Unless there are six authors or more give all authors' names; do not use "et al.". Papers that have not been published, even if they have been submitted for publication, should be cited as "unpublished" [4]. Papers that have been accepted for publication should be cited as "in press" [5]. Capitalize only the first word in a paper title, except for proper nouns and element symbols.

For papers published in translation journals, please give the English citation first, followed by the original foreign-language citation [6].

[1] G. Eason, B. Noble, and I.N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," Phil. Trans. Roy. Soc. London, vol. A247, pp. 529-551, April 1955. (*references*)

[2] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68-73.

[3] I.S. Jacobs and C.P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G.T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271-350.

[4] K. Elissa, "Title of paper if known," unpublished.

[5] R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.

[6] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," IEEE Transl. J. Magn. Japan, vol. 2, pp. 740-741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[7] M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

UMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA

ISSN 2085-4552

Universitas Multimedia Nusantara
Scientia Garden Jl. Boulevard Gading Serpong, Tangerang
Telp. (021) 5422 0808 | Fax. (021) 5422 0800