ISSN 2355-0082

IJNMT

International Journal of New Media Technology

VOLUME
10
No. 1

UMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA

ISSN 2355-0082

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 : ultimacomputing@umn.ac.id

IJNMT (International Journal of New Media Technology) 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 regularly twice a year (June and December) by Faculty of Engineering and Informatics, Universitas Multimedia Nusantara in cooperation with UMN Press.

Call for Papers

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

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

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

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

ISSN 2355-0082

FOREWORD

Greetings!

IJNMT (International Journal of New Media Technology) 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 regularly twice a year (June and December) by Faculty of Engineering and Informatics, Universitas Multimedia Nusantara in cooperation with UMN Press.

In this June 2023 edition, IJNMT enters the 1st Edition of Volume 10. In this edition there are six scientific papers from researchers, academics and practitioners in the fields covered by IJNMT. Some of the topics raised in this journal are: Designing Enterprise Architecture Using TOGAF Framework (Case Study: PT. Indorama), Business Process Reengineering at Mulyoagung Village Community Service Office, Integrated System Design of Sales Module Using SDLC RAD Method, DistilBERT with Adam Optimizer Tuning for Text-based Emotion Detection, Developing HIV/AIDS Patient Profile Model Using K-Means Clustering Method, Intrusion Detection System on Nowaday's Attack using Ensemble Learning.

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 ultimaijnmt@umn.ac.id and the web page of our Journal here.

Finally, we would like to thank all contributors to this June 2023 Edition of IJNMT. 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 2023,

Fenina Adline Twince Tobing, S.Kom., M.Kom.
Editor-in-Chief

ISSN 2355-0082

TABLE OF CONTENT

ISSN 2355-0082

Designing Enterprise Architecture Using TOGAF Framework (Case Study: PT. Indorama)

Feby Janiar Husein[1], Melissa Indah Fianty[2]
Information Systems Program, Multimedia Nusantara University, Tangerang, Indonesia
[1]melissa.indah@umn.ac.id

Accepted 01 December 2022
Approved 06 January 2023

Abstract*— PT Indorama is an organization engaged in manufacturing. In business process activities, IT has been proven capable of increasing effectiveness and efficiency in planning business strategies and can establish relationships with customers that can create increased buying and selling transactions in a competitive manner. However, in the process of implementing IT into the company, there are many challenges with developments that are not under customer needs, which can be grouped into four main problems: lack of ownership of business strategy by customers, low level of alignment of IT development plans with strategy business, low capability to use IT as a competitive advantage, and reasonable standards for IT Operation services in business are not available. To obtain solutions that can be used to achieve its business goals, this study uses the Enterprise Architecture approach to identify the current and expected target architecture and perform a gap analysis. The gap is used as a recommendation to be fulfilled in answering the problems found. The methodology used is TOGAF ADM, which is process-based and provides flexibility in using artifacts according to the organization's specific conditions. The study produces business models and enterprise architecture on customer relationship management systems and services, recommendations for strengthening business areas, recommendations for aligning IT plans with business strategy, and recommendations for using IT solutions in the form of artifacts in the form of catalogs, analysis, and schematics.

Keywords*— Enterprise Architecture, Customer Relationship Management, TOGAF ADM

I. INTRODUCTION

Every company involved in manufacturing has a production process by which products and services are delivered and widely distributed from producers to consumers. One of the essential things a manufacturing company must face is maintaining customer satisfaction, so customer relationship management is needed so the company can develop as a whole [1]. Customer Relationship Management is an integrated concept in information technology and business with the primary goal of building a long relationship between the organization and its customers. Organizations invest heavily in CRM projects to better understand customers and respond quickly to their requests and needs [2].

In addition to improving the company's relationship with customers with customer relationship management to achieve a competitive advantage, companies also need the role of Enterprise Architecture (EA) related to business strategy planning. Enterprise architecture is an essential tool for organizational success. In practice, enterprise architecture has many methodologies used by organizations for IS / IT development, one of the most popular methodologies today is The Open Group Architecture Framework (TOGAF) [3].

Due to its complete architectural process, the TOGAF framework is widely used by most companies [4]. One of the companies that will be discussed in this research is PT. Indorama. PT. Indorama is one of the manufacturing companies in Indonesia. In the process of its business activities, PT. Indorama has tried to work on a strategic plan, but there are obstacles in the planning process. In other words, some plans are not following the objectives, so they cannot be carried out properly, such as one of the facilities' shortcomings in supporting PT's business activities. Indorama does not yet have an enterprise architecture system design plan.

There are several elements of the problem in different fields, including in the field of business, companies that do not optimize information technology in their business which makes companies feel unable to compete with their business competitors because, in the absence of specific needs, customers and prospective

customers who want to do business in the company will also retreat. In the field of data, some current business processes, such as data management and reporting, require a long processing time because they are still being executed manually [4]. These problems make the sales team often late in conveying information to the leadership in taking needs and handling complaints or complaints from customers. In the application field, integration between application systems is needed to accommodate business needs effectively and quickly to improve company performance in targeting or capturing all the needs that customers need. In the field of technology, from an element of technology, several applications are the same and also provide the same services but different infrastructures. Companies need help to make the right decisions [5].

Based on the problem above description above, the company's current IT architecture needs to be identified in depth to obtain a precisely targeted IT architecture planning solution to achieve the expected IT strategy objectives. The TOGAF ADM framework is adapted to the manufacturing department, has complete steps, and has a systematic structure. This design allows the company to create a corporate architectural design that produces detailed blueprints for developing information systems, especially in purchasing, manufacturing, and selling raw materials and customer relationship management systems at PT. Indorama can realize the goals that exist in the company [6].

The enterprise architecture implementation plan is designed to solve problems and failures that arise, and this is one of the best business strategic planning solutions needed by manufacturing companies such as PT. Indorama, so that further business activities can be integrated, planned, centralized, and efficient.

II. METHOD

a. *Customer Relationship Management*

Customer Relationship Management is how a business or other organization manages customer interactions, using data analysis to study large amounts of information [7]. Customer relationship management focuses on acquiring and retaining customers by enhancing customer relationships with the company [8]. Customer relationship management (CRM) refers to all activities or marketing activities carried out to stabilize, develop and exchange good customer relationships [9].

b. *Enterprise Architecture*

Enterprise Architecture is a management and technology practice aimed at improving company performance by seeing the company as a whole and integrated following the view of strategic direction, business practices, information flow, and technological resources [10].

Enterprise architecture consists of documents such as drawings, diagrams, textual documents, standards or models, and business methods that explain what kind of information system the company needs. Enterprise architecture will be used as a reference for developing information systems because developing a system without having a good architecture will be challenging in achieving maximum results [11]. The Open Group Architecture Framework (TOGAF).

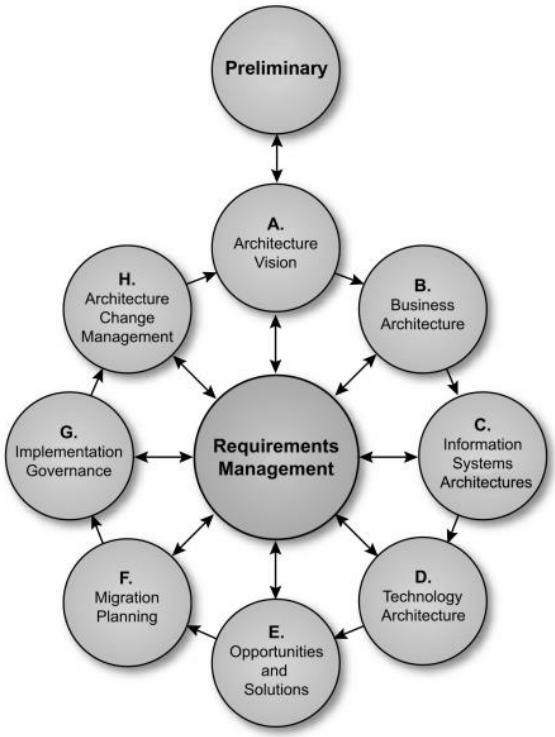


Figure 1. Architecture Development Method [8]

The stages of TOGAF ADM can be briefly explained as follows [12]:

- Preliminary Phase, explain the preparatory and initiatory activities needed to prepare to meet business objectives for the new company architecture, including an explanation of the organization-specific architecture framework and definition of principles.
- Phase A: Architectural Vision describes the initial phase of the architecture development process. This includes information about defining spatial boundaries, identifying stakeholders, creating an architectural vision, and getting approval.
- Phase B: Business architecture, explain business architecture development to support the agreed architecture vision.
- Phase C: Information systems architectures, describes the development of information systems architecture for architectural projects, including data architecture and applications development.

- Phase D: Technology architecture, describe the development of architectural technology for architectural projects.
- Phase E: Opportunities and solutions, planning for implementation, and identifying delivery vehicles for the architecture specified in the previous phase.
- Phase F: Migration planning, discuss the preparation of a series of transitional architecture sequences in detail with a supporting implementation and migration plan.
- Phase G: Implementation Governance provides architectural oversight of implementation.
- Phase H: Architecture Change Management, establish procedures for managing changes to the new architecture.

c. *Theoretical Framework*

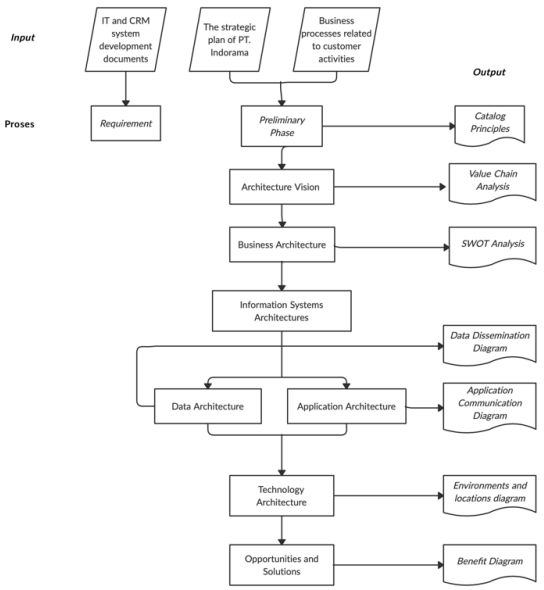


Figure 2. Theoretical Framework

The theoretical framework will consist of the following: [13]

- Requirement
  Input: documents related to IT and CRM system development.
  Process: it takes several documents from the company to support the process needs, such as reading documents related to the development of IT and CRM systems in the company and understanding the background of the CRM system services currently underway in the company. From this background, the researcher identified the problems faced by companies in providing CRM system services.
  Output: can describe the functionality or services of the system.
- Enterprise Architectural Design
  Input: a strategic plan of PT. Indorama and business processes related to customer activities.
  Process: strategic plans and business processes are analyzed into 5 phases of TOGAF ADM. There are three stages in order to fulfill the strategic plan related to the new CRM system to be built, namely:
  – Determine the purpose and scope of the research, while the scope of this research refers to the CRM system, such as customer visits, claims, and complaints at PT. Indorama by using TOGAF as its framework.
  – Literature study where researchers look for sources of literature on CRM system services carried out by previous researchers. Researchers conduct research comparisons to distinguish the research conducted by previous research.
  – Conduct interviews with resource persons with roles and experience in developing CRM systems in the IT field.
  Output: the results of the analysis of the 5 phases of TOGAF ADM, namely the preliminary phase, the vision architecture phase, the business architecture phase, and the business architecture phase, which is divided into two parts, namely the data and application architecture phase, the technology architecture phase, and the opportunity and solution phase.
- Enterprise Architecture Design Results
  Input: the results of the analysis of the 5 phases of TOGAF ADM.
  Process: drafting a new CRM system design based on the TOGAF framework. The data obtained from the ADM stages are analyzed using qualitative methods to be used as a basis for determining the condition of the running system or the baseline to determine the target of the new CRM system.

Output: from the diagram, it can be an artifact that will later be used as the final result of the enterprise architecture design

III. Result and Discussion

a. Preliminary Phase

The artifacts from the preliminary stage are the principles catalog. The Principles catalog is used to capture the principles of architectural solutions. The catalog principle is used to review and approve a result for a defined architectural decision. In determining the

Principle catalog, an interview was conducted with the IT manager at PT. Indorama

Table 1. Principle Catalog

No.	Principle	Category
1	Facilitate the management of customer relationship management services	Business principle
2	Add value to customer relationship management services	Business principle
3	Integrated data	Information systems principle
4	Ease of data access	Information systems principle
5	Data security	Information systems principle
6	Easy to use	Information systems principle
7	Stability	Information systems principle
8	System speed	Technology principle
9	Service backup plan	Technology principle

Table 1 above shows the capabilities required by the company in the enterprise architecture being developed. Nine required skills cover business principles, information systems principles, and technology principles [14].

b. Requirement Management

Requirement management is a dynamic set of requirements. The artifact of the requirement management stage is the architecture requirements specification. The following are the architecture requirement specified in this research:

- Architectural Vision
  - Make it easy for customers to contact the company
  - Add value to the company's customer relationship management service
- Business Architecture
  customer relationship management service improvement
- Information System Architecture
  - Easy to use user interface
  - Integrated apps and data
  - Automatic reporting app
- Technology Architecture
  - High availability server availability
  - Availability of bandwidth management for CRM system bandwidth management.
- Opportunities and Solutions
  A description of the benefits of designing customer relationship management services.

c. Architecture Vision

This phase is also the initial phase in the TOGAF ADM. This phase has the objective of defining the vision of the architecture. Furthermore, this phase includes defining the scope and identifying stakeholders. One of the artifacts explained in this phase is Architecture Vision, as in this phase. The Architecture Vision is as follows:

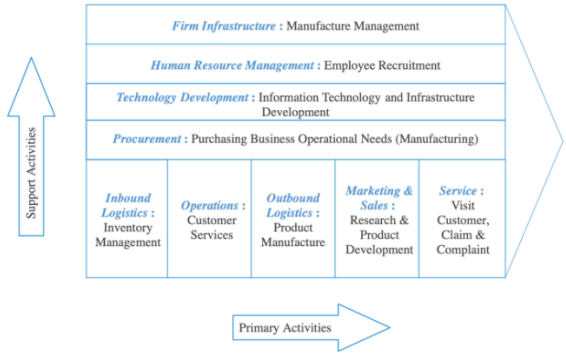


Figure 3. Value Chain Diagram

Based on Figure 3 above, the CRM service of PT Indorama has four leading activity roles: inventory management, customer services, product manufacturing, research and product development, customer visits, claims, and complaints. Meanwhile, the supporting activities are manufacturing management, employee recruitment, infrastructure, information technology development, and purchasing activities for operational business needs (for manufacturing)—Customer Relationship Management System at PT. Indorama is the main activity in serving customers. The team from Customer Relationship Management at PT. Indorama is involved in various activities ranging from service activities, scheduling visits (visiting customers), and serving any complaints from customers (claims and complaints).

d. Business Architecture

This phase provides the architecture of the business of customer relationship management. One of the artifacts used in this section is the SWOT analysis carried out to see companies' strengths, weaknesses, opportunities, and threats in developing customer relationship management services [15]. PT. Indorama has an experienced team, besides the availability of funds and management support in developing customer relationship management services, PT. Indorama. These things are the strength of PT. Indorama. Limitations in customer relationship management services are the limitations of the

information technology team needed for system development. Opportunities and development of customer relationship management services are to reduce telephone usage costs and reach markets and potential customers through the customer order application. The threat that arises is losing customers due to customer dissatisfaction with current customer service.

Table 2. SWOT Analysis

Strength (S)	Weakness (W)
Company management	Limitations of information technology team
Product quality	
Have an experienced team	
Availability of funds for service development	
management support in service development	
Opportunities (O)	**Threats (T)**
Reduce phone usage costs	Company competition in the same category
Marketing trend is increasing	Lost customers due to customer service dissatisfaction
Customer loyalty	

From the results of the SWOT analysis in table 2 above, there is a threat of losing customers due to dissatisfaction with customer service. The opportunity to reduce telephone usage costs because every incoming customer call is charged to PT. Indorama. Likewise, with the opportunity to reach a wider audience. From the results of the SWOT analysis, seeing the opportunities and threats that arise, the development of customer relationship management service designs relates to the attention of stakeholders and the requirements of business capabilities mentioned in the architecture vision stage.

The following is a gap analysis at the business architecture stage, which can be seen in table 3 below:

Table 3. Gap Analysis Business Architecture

No.	Present condition	Future Conditions
1.	It only has one customer service.	Has two customer services, namely by phone and customer order application.
2.	The filling in data is not practical because it is done more once.	The system is integrated with emails that are automatically distributed to stakeholders.
3.	Reporting from marketing on claims & complaints is not real-time.	Reporting from marketing on claims & complaints can be real-time.
4.	Customers can access no system.	There is already a system or application that customers can access to provide feedback, claim tickets, and complain directly.
5.	The web-based system cannot be accessed from anywhere (real-time).	The system is web-based and can be accessed in real-time.

e. Information System Architecture

The Information System Architecture phase explains the data and application architecture. The diagram will describe the data dissemination diagram that explains the relationship between the logical application and the data entity with the common objectives of the company [16].

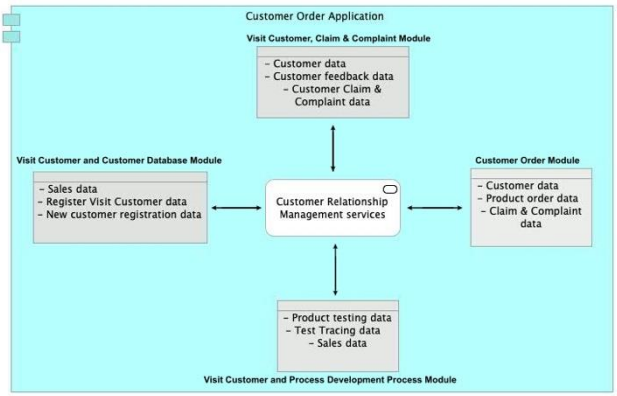


Figure 4. Data Dissemination Diagram

In figure 4. describes the relationship of customer relationship management services with customer order applications and data. In the customer visit, claim & complaint module, there are customer data, customer feedback data, and customer claim & complaint data. The customer order module has customer data, product order data, and claim & complaint data. The customer visit and process development process module include product testing data, test tracing data, and sales data. The customer visit & customer database module includes sales data, customer visit register data, and new customer register data.

Future / Current	Customer Relationship Management System	Visit Customer, Claim & Complain System	Visit Customer and Process Development Process System	Visit customer & customer database System	Customer order System
Customer Relationship Management System	Retain				
Visit Customer, Claim & Complain System		Retain			
Visit Customer and Process Development Process System			Retain		
Visit customer & customer database System				Retain	
Customer order System					New

Figure 5. Gap Analysis Information Systems Architecture

In figure 5 above, a new application is used, namely the customer order application. Four applications are maintained, namely the customer relationship management application, customer visits application, claims & complaints, customer and process development process application, and customer & customer order application.

f. Technology Architecture

Technology architecture identifies the usage of technology to enable applications within the company. Hence business performance could be improved [17]. In this phase, the relevant technology architecture will be developed using previously developed application architecture. The environment and location diagram is one of the artifacts produced in this architecture. The environment and location diagram shows the identified and proposed technology to support the application and data requirement.

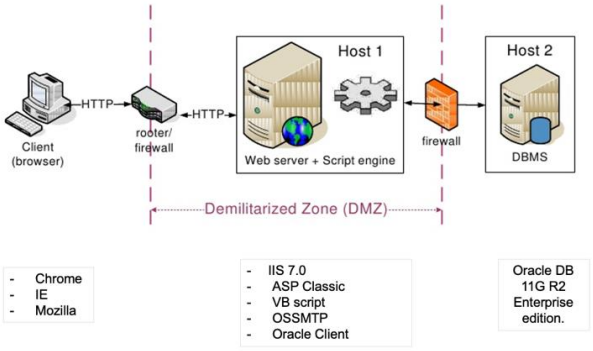


Figure 6. Existing Network Topology

The current network topology in figure 6 relies on a single router as a single-point service. The router currently functions as an internet gateway, user gateway, DHCP server, and bandwidth management. Using a router as a single-point service results in high CPU and memory usage, so the router's performance becomes unstable. Today's switches are still unmanageable, so it is impossible to do more in-depth configurations. Network and internet connections are needed to support business operations and customer relationship management services because data exchange is carried out through the DBMS system, customer service management uses software as a service, and voice-over IP connections to providers using SIP trunks are connected via the internet. The critical need for stable network and internet connections requires adjustments to network configurations to support business operations and customer relationship management systems.

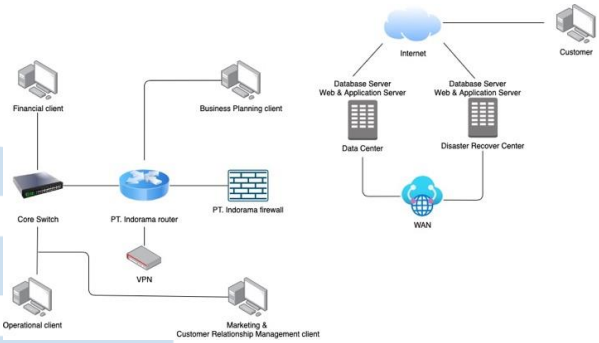


Figure 7. Environment and Location Diagram

In figure 7 above shows that the head office has business planning, customer service, operations departments, customer relationship management, finance, and IT divisions connected to the switching core and the router. IT departments use core switching and central office routers to unify and control internal networks such as data centers, disaster recovery centers, and customer sites so that the central office with customer clients can connect. Enterprises use private virtual or wide area networks using network methods. The customer headquarters and the customer must be able to connect to the data center and disaster recovery center so that the wide-area network uses the necessary VPN scheme. All customers can connect to the Internet to access the company's website and applications.

Future / Current	Router	Firewall	Switch	VPN	Core Switch	Wireless Lan Controller	Data Server
Router	Replace						
Firewall		Replace					
Switch			Replace				
N/A				New	New	New	New

Figure 8. Gap Analysis Topologi pada Technology Architecture

g. Opportunities and Solutions

Opportunities and Solutions is one of the TOGAF Architecture Development Method phases that aims to evaluate the design model of developed architecture [18]. This phase will then be the foundation or guidance for the implementation plan [19]. This phase helps the implementation plan to achieve its target implementation. One of the artifacts used in this application is the benefit diagram. This diagram aims to explain the relationship between the benefit of the opportunities identified and proposed [20].

Based on figure 9 below, it is stated that there are two goals in customer relationship management services, namely the first goal: is to find out the comparison of customer relationship management systems using the TOGAF ADM framework, and the second goal: to produce an architectural model that can be used as a reference in the development of new infrastructure. The first goal is to provide a solution to provide tactical, strategic, and forecasting analysis applications, provide better decision-making results for customer relationship management services and provide an improved service decision measurement. The second goal is to provide a single solution, which is to provide integrated applications for managing strategy, policy, and process management, to provide faster and more integrated strategy, policy, and process management results, and to provide measurements with a higher chance of adapting strategies, policies, and new processes. Then from combining the two goals and their solutions, results, and measurements, two benefits or benefits from each goal are obtained, namely higher customer service and leading to higher revenue for the company and policies and processes that are faster in customer service and can generate revenue and better customer service. From the combination of these two benefits, the final result is in the form of benefits, namely better customer satisfaction overall and can generate better income.

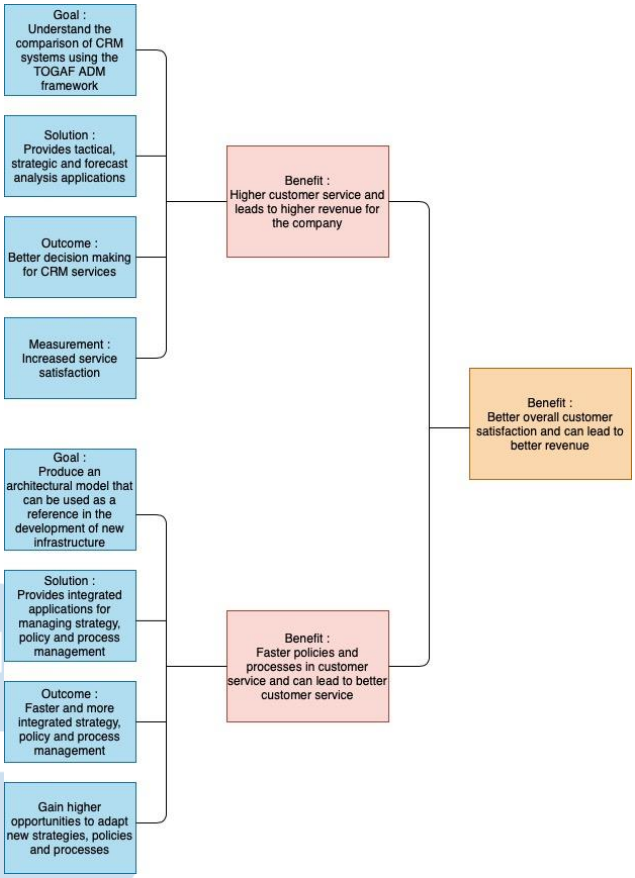


Figure 9. Benefit Diagram

IV. CONCLUSION

Based on the results of planning and designing a business model and analyzing the company's architectural system and customer relationship management services at PT. Indorama can be concluded as follows:

1. Design business models and enterprise architecture on customer relationship management systems and services at PT. Indorama was only carried out from the preliminary phase to the opportunities and solution phase. This adapts to the company's needs and time constraints so that it cannot produce an application that can be executed and passed on to the next step.
2. Planning and design of the application architecture stage have a target design that adds customer order applications in each function that is useful for managing applications for each existing function.
3. In the design phase of the technology architecture, technology adjustments are made to the application that will be used. Previously, the

technology was simple, using only one router and a standalone wireless network.

4. In this research, enterprise architecture design produces various artifacts at each stage. These artifacts can be in the form of catalogs, analyses, and schematics used in designs that can help achieve business strategies.

REFERENCES

[1] V. Soraya and W. S. Sari, "Perancangan Enterprise Architecture Sistem Informasi dengan Menggunakan Framework TOGAF ADM pada CV. Garam Cemerlang," *JOINS (Journal Inf. Syst.*, vol. 4, no. 2, pp. 148–156, 2019, doi: 10.33633/joins.v4i2.3054.

[2] M. S. Farhan, A. H. Abed, and M. A. Ellatif, "A systematic review for the determination and classification of the CRM critical success factors supporting with their metrics," *Futur. Comput. Informatics J.*, vol. 3, no. 2, pp. 398–416, 2018, doi: 10.1016/j.fcij.2018.11.003.

[3] F. E. Gunawan, J. F. Andry, H. Tannady, and R. Meylovsky, "Designing enterprise architecture using togaf framework in meteorological, climatological, and geophysical agency," *J. Theor. Appl. Inf. Technol.*, vol. 97, no. 20, pp. 2376–2385, 2019.

[4] S. Kotusev, "TOGAF-based Enterprise Architecture Practice: An Exploratory Case Study," Communications of the Association for Information Systems, vol. 43, p. 321 – 359, 2018.

[5] Qurratuaini, "Designing enterprise architecture based on TOGAF 9.1 framework," IOP Conference Series: Materials Science and Engineering, vol. 402, no. 1, 2018.

[6] D. Y. Ratnasari and D. A. O. Turang, "Perancangan Enterprise Architecture Pada Perusahaan Bidang Jasa Menggunakan The Open Group Architecture Framework (TOGAF)," *Semin. Nas. Inform.*, vol. 2018, no. November, pp. 31–42, 2018, [Online]. Available: http://103.23.20.161/index.php/semnasif/article/view/2614.

[7] I. Puspitasari, "Stakeholder's expected value of Enterprise Architecture: An Enterprise Architecture solution based on stakeholder perspective," in In 2016 IEEE 14th International Conference on Software Engineering Research, Management and Applications (SERA), Baltimore, 2016.

[8] A. A. C. Dewi and H. Semuel, "Pengaruh Customer Relationship Management (Crm) Terhadap Customer Satisfaction Dan Customer Loyalty Pada Pelanggan Sushi Tei Surabaya," vol. 3, no. 1, pp. 1–9, 2015.

[9] A. Iriandini, "Pengaruh Customer Relationship Management (Crm) Terhadap Kepuasan Pelanggan Dan Loyalitas Pelanggan (Survey pada Pelanggan PT. Gemilang Libra Logistics, Kota Surabaya)," *J. Adm. Bisnis S1 Univ. Brawijaya*, vol. 23, no. 2, p. 85998, 2015.

[10] R. E. Riwanto and J. F. Andry, "Enterprise Architectures Enable of Business Strategy and IS/IT Alignment in Manufacturing using TOGAF ADM Framework," *Int. J. Inf. Technol. Bus.*, vol. 1, no. 2, pp. 1–2, 2019.

[11] M. Fadhil, F. R. Industri, and U. Telkom, "Perancangan Enterprise Architecture Fungsi Sumber Daya Manusia Unit Operasional Mengunakan Framework Togaf Adm Pada Pt Albasia Nusa Karya Design of Enterprise Architecture Based Togaf Adm Case Study on the Function of Operational in Human Resources Manage," vol. 5, no. 2, pp. 3385–3390, 2018.

[12] D. Kwek, D. Maulana, E. R. Kaburuan, and N. Legowo, "Enterprise architecture planning information system based on cloud computing using togaf (case study: Pandi. Id registry)," *Int. J. Sci. Technol. Res.*, vol. 8, no. 9, pp. 1167–1178, 2019.

[13] L. Amerta, Y. A. Prasetyo, and B. Rahmad, "Anaysis and Design of Enterprise Architecture Using Togaf," vol. 4, no. 3, pp. 4591–4598, 2017.

[14] D. Minoli, Enterprise Architecture A to Z: Frameworks, Business Process Modeling, SOA, and Infrastructure Technology (Second Edi), United States of America: Auerbach, 2018.

[15] M. Lankhorst, Enterprise architecture at work: Modelling, communication and analysis, fourth edition, Springer, 2017.

[16] F. F. Purba and B. Rahmad, "Perancangan Enterprise Architecture Unit Iii Dengan Menggunakan," pp. 38–45, 2016.

[17] A. C. Rizkyanur, I. Darmawan, and B. Rahmad, "Perancangan Enterprise Architecture E-Commerce Pada Bagian Manajemen Hubungan Pelanggan Di Pt Xyz Menggunakan Framework Togaf Adm Designing E-Commerce Enterprise Architecture on Customer Relationship Management Section in Pt Xyz Using Togaf Adm," vol. 2, no. 2, pp. 5136–5143, 2015.

[18] A. Fauzi and Y. Handoko, "Analisa dan Perancangan Model Umum Enterprise Architecture untuk E-Business Usaha Mikro Kecil dan Menengah (UMKM) dengan Menggunakan Framework TOGAF ADM," *J. Tata Kelola dan Kerangka Kerja Teknol. Inf.*, vol. 4, no. 1, pp. 1–8, 2018, doi: 10.34010/jtk3ti.v4i1.1392.

[19] S. Eviana, "Perancangan Enterprise Architecture Sistem Penjualan Dengan Metode Togaf Adm Pada Marino Collection | Eviana | Prociding Kmsi," pp. 106–113, 2018, [Online]. Available:

http://ojs.stmikpringsewu.ac.id/index.php/procidingkmsi/article/view/627/560.

[20] R. S. M. v. S. P. F. H. v. V. Martin van den Berg, "How enterprise architecture improves the quality of IT investment decisions," Journal of Systems and Software, vol. 152, pp. 134-150, 2019.

ISSN 2355-0082

Business Process Reengineering at Mulyoagung Village Community Service Office

Ahmad Raihan Djamarullah[1], Budiman Hamsyurah[2], Ilyas Nuryasin[3]
[1,2,3] Informatics, University of Muhammadiyah Malang, Malang, Indonesia
[1] raihandj@webmail.umm.ac.id
[2]budimanhamsyurah@webmail.umm.ac.id
[3]ilyas@umm.ac.id

Accepted 06 December 2022
Approved 04 January 2023

***Abstract*— The Mulyoagung Village Office is a community service office in Mulyoagung Village. The service for making identity cards currently implemented at the village office uses the SiPeduli Desa website, but the results of the analysis carried out show that the throughput efficiency obtained is still quite low. There are several processes that require a long waiting time and also manual processing processes so that this condition is still less efficient. Therefore, changes in business processes need to be made to improve the old system that is currently running into a more optimal system. This significant change in business processes is called Business Process Reengineering. The concept of reengineering was born to solve the shortcomings of the old system or legacy system in a business process. Completion of a business process with Business Process Reengineering is done by eliminating processes that do not provide added value and changing manual processes into automated processes by utilizing information technology. The results of this study are business process designs that increase throughput efficiency from 30.85% in the previous process to 89.61% in processes that have been carried out by Business Process Reengineering. The application of Business Process Reengineering can help provide the design of a new business recommendation model that is obtained after an analysis of old business processes and analysis of redesign alternatives is carried out. Business Process Reengineering is also able to improve old processes both in terms of service and speed.**

***Keywords*— Business Process Reengineering; Kantor Desa Mulyoagung; Reengineering**

I. INTRODUCTION

The business world is now growing, plus transaction processes are increasingly complex and the scope of business is getting wider. Therefore, it has become a necessity for a company to further improve its strategies and methods to succeed in the business world[1][2]. The Changers in business processes are sometimes needed to replace the old system that is currently running with a new system that is more optimal. This significant change in business processes is called Business Process Reengineering[3].

The term reengineering itself first appeared in the late 90s in the field of information technology (IT) by Michael Hammer who published an article in Harvard Business Review, which explains the importance of fundamental changes in an organization/company due to global changes in the economy, increasingly fierce competition, and changes in customer demand[4].

The concept of reengineering was born to solve the shortcomings of the old system or legacy system in a business process[5], A business process is a series of related activities carried out to achieve business results that are in line with the overall business vision and mission [6][7].

Business Process Reengineering is part of reengineering which is a concept of updating business processes by evaluating the shortcomings that exist in business processes[8][9], The concept is applied to optimize operations, reduce costs, accelerate business processes and improve services provided to clients to be more efficient and competitive[10][11].

Business Process Reengineering is one of the critical solutions for improving all business processes and performance measures. However, the use of Business Process Reengineering can also fail when the process tends to be focused without paying attention to the surrounding environment and company knowledge [12][13]. Therefore, the implementation of Business Process Reengineering must start with aligning the vision and mission within a company[13][14].

Several studies have shown that the use of the concept of Business Process Reengineering can provide significant changes and can improve company performance. The following are some examples of success from the implementation of Business Process Reengineering, the first of which was implemented at the Mojokerto PDAM, increasing throughput efficiency to 94.46%. Then the implementation of Business Process Reengineering at the Batu District Attorney's Office resulted in a throughput efficiency of 85.77% [4]. And lastly, the implementation of Business Process Reengineering at Indonesian A&A companies resulted in improvements in cost, time, quality, and flexibility factors[15].

Business Process Reengineering will be carried out in this research at the Mulyoagung Village Office. The business process that will be used is related to the community service work program, namely the process of making an ID card at the village office.

One of the existing business processes at the Mulyoagung village office is the system for making ID cards at the village office using the SiPeduli Desa website. The use of the website helps in sending data to the Department of Population and Civil Registration, but after analysis, there are several business processes that require a long waiting time and also manual processing processes so this condition is still less efficient.

Therefore, Business Process Reengineering was carried out on the business process of making ID cards at the village office. this is done to increase the level of efficiency during the process of making ID cards.

II. METHOD

The method used to support this research is the Business Process Reengineering method which is shown in Figure 1 below.

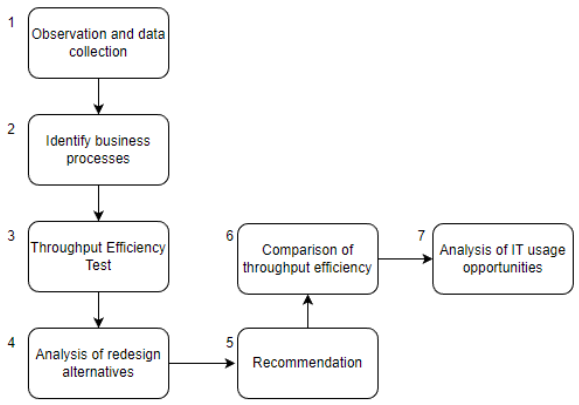


Fig. 1. Business Process Reengineering Method

Figure 1 shows the stages of the applied Business Process Reengineering methodology. Starting from the data collection stage to the solution stage.

1. Observation and Data Collection

At this stage, data collection and data search are carried out to meet the needs needed. Data were obtained from observations and interviews with the Mulyoagung Village Office as well as conducting a literature study related to Business Process ReengineeringAt this stage, data collection and data search are carried out to meet the needs needed. Data were obtained from observations and interviews with the Mulyoagung Village Office as well as conducting a literature study related to Business Process Reengineering.

2. Identify Business Process

At this stage, the results of data collection carried out will be identified to determine the business processes applied to the process of making ID cards at the Mulyoagung village office. Then the mapping was carried out using the ASME standard map.

3. Throughput Efficiency Test

In the throughput efficiency test phase, Equation 1 will be used below.

$$\text{Throughput Efficiency} = \frac{\text{processing time is not a delay}}{\text{total time in the system}} \text{x100\%} \quad (1)$$

The value used in the throughput efficiency test is based on the time result of business process performance using ASME (American Society of Mechanical Engineers) standards. Furthermore, a comparison will be made between the throughput efficiency results from the initial business process and the throughput efficiency results from the new business process reengineering.

4. Analysis of Redesign Alternatives

At this stage, an analysis of the business processes that have deficiencies is carried out and improvements to the business process design are carried out. The improvement process is carried out by simplifying the process, reducing processing time, eliminating errors in the process, standardizing and automating the process.

5. Recommendations

At this stage, a business redesign is carried out based on the results of alternative analysis which is formed into a new, more efficient business process. Then mapping using ASME standard maps and testing the efficiency of throughput.

6. Comparison of Throughput Efficiency

At this stage a comparison of the overall service time is carried out on the initial business process and the redesigned business process. Comparisons were made based on the results of the percentage of overall service time using the ASME standard map.

7. Analysis of IT Usage Opportunities

At this stage an analysis is carried out to find out the opportunities for using Information Technology (IT) that can support the redesign of business process designs

III. RESULT AND DISCUSSION

1. Observation and Data Collection

From the results of observations and interviews on October 10 and 13, 2022 at the Mulyoagung Village Office, the system for making ID cards at the village office uses the SiPeduli Desa website. The use of the website helps in sending data to the Department of Population and Civil Registration, but after analyzing there are several business processes that require a long waiting time and also manual processing processes so that this condition is still less efficient.

The hypothesis obtained from the results of observations and interviews related to the business process of making ID cards at the Mulyoagung village office that needs to be reengineered where the process of submitting an ID card is already using the website, but for the process of submitting the manufacture is still constrained by queuing problems and also the number of document processing processes that cause sufficient queuing time long. In addition, the process of verification and approval of data requires a process that is quite time consuming.

2. Identify Business Process

From the results of data collection conducted at the Mulyoagung Village Office, it was found that the process of making ID cards at the village office was identified as having a weak point of speed. In the business process, there are several parts involved in this business process, namely: applicants, village office employees, sub-district office employees, service operators, and heads of services. The flow of the business process for making ID cards at the Mulyoagung village office is shown in Figure 2.

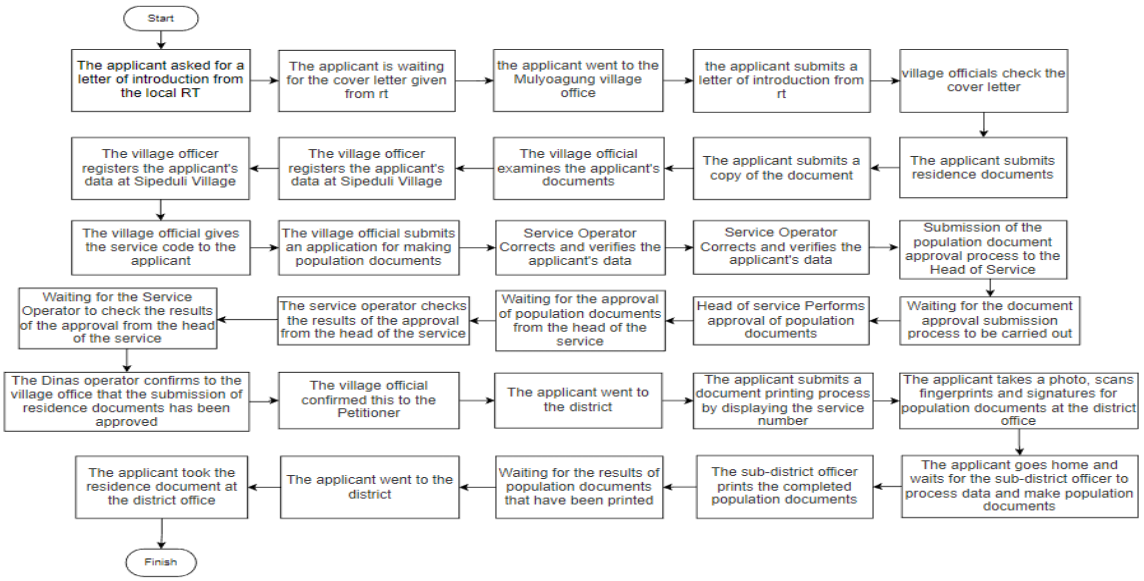


Fig 2. Business Process Flow for Making ID Cards at the Mulyoagung Village Office

3. Throughput Efficiency Test

At this stage, the business process for making ID cards is tested at the Mulyoagung village office using a throughput efficiency test. This test is mapped with ASME (American Society of Mechanical Engineers) standard maps. This test is carried out to measure the percentage of overall service time, the test results will be compared with the business process recommendations in terms of models and results.

The following is a map of the ASME standard and the throughput efficiency test of the business process for making ID cards at the Mulyoagung village office.

TABLE I. IMPROVING THE BUSINESS PROCESS FOR MAKING ID CARDS AT THE VILLAGE OFFICE

No	Business Process	○	□	⇨	D	▽	◯	Time / Minute	Process Owner
1.	Applicant request a cover letter from the local RT	●						120	Applicant

2.	Applicant is waiting for the cover letter given from RT							1320	Applicant
3.	Applicant goes to the mulyoagung village office							20	Applicant
4.	Applicant submits a cover letter from rt							5	Applicant
5.	Village officer checks cover letter							10	Village Officer
6.	Applicant submits a residence document							5	Applicant
7.	Submit a copy of the document							10	Applicant
8.	Village officials check the applicant's documents							30	Village Officer
9.	The officer scans the applicant's documents							15	Village Officer
10.	Village officers register applicant data at Sipeduli Desa							20	Village Officer
11.	The village officer gives the service code to the applicant							10	Village Officer
12.	Village officials apply for residence documents							10	Village Officer
13.	Service Operators Correcting and verifying applicant data							480	Service Operator
14.	Waiting for correction and verification of applicant data							960	Service Operator
15.	Service Operators Submit for the population document approval process to the Head of Service							480	Service Operator
16	Waiting for the document approval process to be done							960	Service Operator
17.	Head of department Approve the residence document to be published							480	Head of Department
18.	Waiting for the approval of the residence document							960	Head of Department

	from the head of the service								
19.	Service Operators check the results of approval from the head of service							480	Service Operator
20.	Waiting for the Service Operator to check the results of the approval from the head of the service							960	Service Operator
21.	The Operator confirms to the village office that the application for residence documents has been approved							30	Service Operator
22.	Village officer confirms to applicant							30	Village Officer
23.	The Applicant goes to the district							20	Applicant
24.	The Applicant submits the document printing process by displaying the service code							10	Applicant
25.	The Applicant takes photos, scans fingerprints and signs for residence documents at the sub-district office							120	Applicant
26.	Applicant go home and wait for sub-district officers to process data and create population documents							480	Applicant
27.	The sub-district officer prints the finished residence document							480	Petugas Kecamatan
28.	Waiting for the results of the residence documents that have been printed							960	Petugas Kecamatan
29.	The Applicant goes to the district							20	Applicant
30.	The Applicant takes the residence document at the sub-district office							60	Applicant
	Number of Stages	17	3	3	7	0	0		
		2,365	520	60	6,600	0	0	9,545	

Table 1 shows the ASME standard map of the process of making ID cards at the Mulyoagung village office. Table 1 contains process stages, process symbols, processing time in minutes, and process owners.

Furthermore, throughput efficiency testing was carried out to measure the overall service time performance from the results of the ASME standard mapping of the new installation process as follows.

$$throughput\ efficiency = \frac{2{,}945}{9{,}545}\ x\ 100\% = 30.85\%$$

The results of testing the throughput efficiency of the ASME standard mapping on the process of making ID cards at the Mulyoagung village office. The value of 2,945 is the processing time without delay, while the value of 9.545 is all processing time including the delay. The results of the throughput efficiency test obtained are 30.85% and the remaining time is 69.15% in the process is not running.

4. Analysis of Redesign Alternatives

At this stage, the process design is refined by simplifying the process, reducing processing time, eliminating errors in the process, standardizing and automating the process.

TABLE II. IMPROVING THE BUSINESS PROCESS FOR MAKING ID CARDS AT THE VILLAGE OFFICE

No	Process Stage	Completion Step
1	Applicant asks for a cover letter from the local RT	Elimination
2	Applicant is waiting for the cover letter given from rt	Elimination
3	Applicant submits a cover letter from rt	Elimination
4	Village officer checks cover letter	Elimination
5	Service Operators Correcting and verifying applicant data	Automate
6	Waiting for correction and verification of applicant data	Elimination
7	Service Operators Submit for the population document approval process to the Head of Service	Automate
8	Waiting for the document approval process to be done	Elimination
9	Head of department Approve the residence document to be published	Automate
10	Waiting for the approval of the residence document from the head of the service	Elimination
11	Service Operators check the results of approval from the head of service	Elimination
12	Waiting for the Service Operator to check the results of the approval from the head of the service	Elimination
13	The Operator confirms to the village office that the application for residence documents has been approved	Elimination
14	The Applicant goes to the district	Elimination

15	Applicant go home and wait for sub-district officers to process data and create population documents	Elimination
16	The sub-district officer prints the finished residence document	Automate
17	Waiting for the results of the residence documents that have been printed	Elimination
18	The Applicant goes to the district	Elimination
19	The Applicant takes the residence document at the sub-district office	Elimination

Table 2 is the result of improving the business process design for making ID cards at the Village Office by eliminating several processes that are considered lacking or can be changed into a new process. Then automate the process to make it easier to do using IT.

TABLE III. ALTERNATIVE BUSINESS PROCESS FOR MAKING ID CARDS AT THE VILLAGE OFFICE

No	Task Name	Alternative
1	Registration process via village account	This can provide 2 village accounts, a computer, and a scanner to speed up the process of submitting population documents.
2	The Service Operator verifies the applicant's data and submits the residence document approval to the Head of the Service	Automate data verification as well as apply for approval when data has been verified
3	Head of Service Approve the population documents to be issued	Automate the approval process and send confirmation messages to the village or related applicants via e-mail or phone number.
4	The process of printing the finished residence document	This can be done at the village office by providing a tool for printing residence document cards

Table 3 above is an alternative process offered so as to reduce the waiting time that occurs in several processes.

5. Recommendations

New business processes are designed following an analysis of redesign alternatives that eliminate and automate some business processes. Adjustments were then made based on the results of the analysis of opportunities for using IT for the business process of recommendations for making a new Mulyoagung village office ID card.

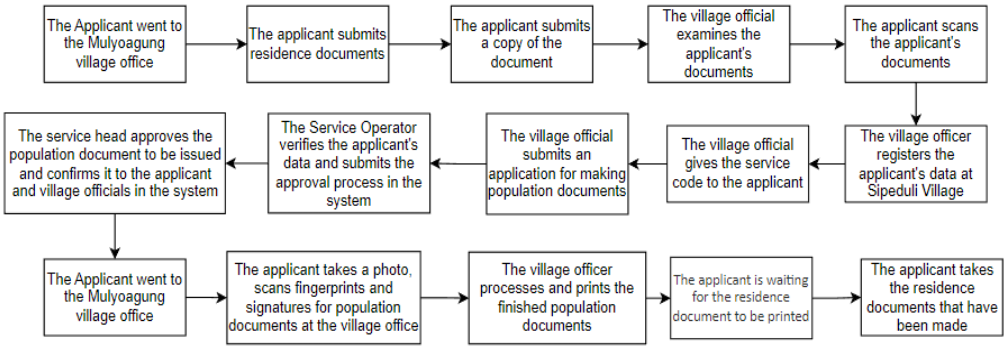


Fig 3. Business Process Recommendations for Making KTPs at the Mulyoagung Village Office

Figure 3 is a business process recommendation for making a new Mulyoagung village office ID card. Where in the business process, the applicant's recommendation does not need to ask for a cover letter from the RT and the process of making an ID card can be done at the Mulyoagung village office.

From the recommendation business process that has been designed, mapping is carried out using ASME standards and throughput efficiency testing to test the overall service time.

TABLE IV. ASME STANDARD MAP BUSINESS PROCESS RECOMMENDATIONS FOR MAKING KTP AT THE VILLAGE OFFICE

No	Business Process	○	□	⇨	D	▽	▣	Time /Minute	Process Owner
1.	The applicant goes to the mulyoagung village office			●				20	Applicant
2.	The applicant submits a residence document	●						5	Applicant
3.	Applicant Submit a copy of the document	●						10	Applicant
4.	Village officials check the applicant's documents			●				30	Village Officer
5.	The applicant scans the applicant's documents	●						15	Applicant
6.	Village officers help register applicant data at Sipeduli Desa	●						20	Village Officer
7.	The village officer gives the service code to the applicant	●						10	Village Officer
8.	Village officials apply for residence documents	●						10	Village Officer

9.	The Service Operator verifies the applicant's data and submits the approval process in the system							480	Service Operator
10.	Head of service Approval of population documents to be issued and confirm to the applicant and village officials in the system							480	Head of Department
11.	the applicant goes to the mulyoagung village office							20	Applicant
12.	The applicant takes a photo, scans fingerprints and signs for residence documents at the village office							120	Applicant
13.	Village officials process and print ready-made residence documents							120	Village Officer
14.	The applicant waits for the residence document to finish printing							180	Applicant
15.	The applicant takes the completed residence document							20	Applicant
	Number of Stages	9	3	2	1	0	0		
		330	990	40	180	0	0	1,540	

Table 6 is the result of mapping the business process recommendations for making KTPs at the Mulyoagung village office using the ASME standard map. Furthermore, throughput efficiency testing is carried out based on the results of the recommendation business process mapping.

$$efisiensi\ throughput = \frac{1.380}{1.540}\ x\ 100\% = 89{,}61\%$$

The results of the throughput efficiency test carried out on the business process recommendations for making ID cards at the Mulyoagung village office gave very good results with a high percentage of 89.61% and 10.39% of service times that were not running. This percentage increase is due to several processes that are less efficient and require a long time to be eliminated and automated so that they can help the process to be more efficient.

6. Comparison of Throughput Efficiency

At this stage, a comparison of the overall service time on the initial business process and the recommendation

business process is carried out based on the mapping results with ASME standards and the value obtained from the throughput efficiency test results

TABLE V. COMPARISON OF BUSINESS PROCESSES FOR MAKING ID CARDS AT THE VILLAGE OFFICE

No	Business process	Initial Throughput Efficiency	Recommended Throughput Efficiency	Initial Processing Speed	Recommended Processing Speed
1.	The Business Process of Making an ID Card at the Village Office	30,85%	89,61%	9545 Minutes	1540 Minutes

Table 7 is the result of the comparison of the initial business processes and the recommended business processes. The results obtained are the throughput efficiency value increased by 58.76% from 30.85% to 89.61% and also the processing speed increased from 9,545 minutes to 1,540 minutes.

7. Analysis of IT Usage Opportunities

Utilization of Information Technology can increase the efficiency of a business process. Therefore, at this stage an analysis is carried out to find out the opportunities for using information technology to support the redesign of the business process for making ID cards at the Mulyoagung village office

TABLE VI. HARDWARE REQUIREMENTS ANALYSIS RESULTS

No	Hardware	Number of Devices	Unit price
1.	Computer/PC	1	Rp 7.000.000,00
2.	Scanner	2	Rp 5.000.000,00
3.	Card Printing Tool	1	Rp 20.000.000,00
4.	Camera	1	Rp 4.500.000,00
5.	Wifi	1	Rp 500.000,00
	Total Price		Rp 42.000.000,00

TABLE VII. SOFTWARE REQUIREMENTS ANALYSIS RESULTS

No	Software
1.	Microsoft Office
2.	Browser
3.	Website SiPeduli Desa

Table 4 and Table 5 components of information technology that can be used in the business process of making ID cards at the Mulyoagung village office. By utilizing this technology, the waiting time in the queue for the KTP making process at the Muryoagung village office can be shortened, and the KTP issuance process can be simplified so that the KTP can be issued directly at the Muryoagung village office.

IV. CONCLUSION

The application of Business Process Reengineering can help change old business processes to new business processes, so as to be able to get more efficient processes. Completion of a business process is done by eliminating processes that do not provide added value and changing manual processes into automated processes with the help of information technology.

Business Process Reengineering is also able to find the difference between the old process and the new process. Both in terms of service, speed, and cost. In the business process of making ID cards, throughput efficiency tests were carried out with the results of the old business process being 30.85% and the new business recommendation process being 89.61%. In the new business recommendation process, it is superior because the ID card creation time is 89.61% running well with a speed of 1,540 minutes.

Business Process Reengineering can provide new business recommendation model designs that are obtained after analysis of old business processes and analysis of redesign alternatives.

REFERENCES

[1] Z. Zaini and A. Saad, "Business process reengineering as the current best methodology for improving the business process," Journal Of ICT In Education, vol. 6, pp. 66–85, Jun. 2019, doi: 10.37134/jictie.vol6.7.2019., in press.

[2] T. Alhawamdeh, "The impact of Business process reengineering on cost reduction of international business operating in the middle east," ALHAWAMDEH / Journal of Asian Finance, vol. 8, no. 10, 2021, doi: 10.13106/jafeb.2021.vol8.no10.0087., in press.

[3] F. F. Rozaqi, W. Suharso, and I. Nuryasin, "Business process reengineering at PDAM companies in Mojokerto district to improve company business performance," "Business process reengineering pada perusahaan PDAM kabupaten Mojokerto Untuk Meningkatkan Kinerja Bisnis Perusahaan," REPOSITOR, vol. 2, no. 5, pp. 635–648, 2020., in press.

[4] D. Arya and W. Suharso, "Business process reengineering at the batu state prosecutor," "Business process reengineering pada kejaksaan negeri batu," vol. 1, no. 2, pp. 159–170, 2019. , in press. state prosecutor

[5] Majthoub Manar, Qutqut Mahmoud H, and Odeh Yousra, "Software re-engineering: an overview," 2018 8th International Conference on Computer Science and Information Technology (CSIT), 2018, doi: 10.1109/CSIT.2018.8486173., in press.

[6] K. C. Dewi and N. W. D. Ayuni, "Business process re-enginering of tourism e-marketplace by engaging government, small medium enterprises and tourists," Bulletin of Electrical Engineering and Informatics, vol. 10, no. 5, pp. 2866–2874, Oct. 2021, doi: 10.11591/eei.v10i5.3159., in press.

[7] M. N. Waluyo, E. Suhendar, and H. A. Suprapto, "Rancang ulang proses bisnis dengan metode business process

reengineering Pada TLS cargo," CSRID (Computer Science Research and Its Development Journal), vol. 12, no. 3, p. 161, Mar. 2021, doi: 10.22303/csrid.12.3.2020.161-169. , in press.

[8] M. Dachyar and Z. A. H. Sanjiwo, "Business process re-engineering of Engineering Procurement Construction (EPC) project in oil and gas industry in Indonesia," Indian J Sci Technol, vol. 11, no. 9, pp. 1–8, Mar. 2018, doi: 10.17485/ijst/2018/v11i9/92741., in press.

[9] S. Srinivas, R. P. Nazareth, and M. Shoriat Ullah, "Modeling and analysis of business process reengineering strategies for improving emergency department efficiency," Simulation, vol. 97, no. 1, pp. 3–18, Jan. 2021, doi: 10.1177/0037549720957722., in press.

[10] M. Sunil Kumar and D. Harshitha, "Process innovation methods on business process reengineering," International Journal of Innovative Technology and Exploring Engineering, vol. 8, no. 11, pp. 2766–2768, Sep. 2019, doi: 10.35940/ijitee.K2244.0981119., in press.

[11] A. Harika, M. Sunil Kumar, V. Anantha Natarajan, and S. Kallam, "Business process reengineering: issues and challenges," 2021, pp. 363–382. doi: 10.1007/978-981-15-6707-0_35., in press.

[12] A. Mohammad Khashman, "The effect of business process re-engineering on organizational performance: the mediating role of information and communications technology," International Journal of Business and Management, vol. 14, no. 9, p. 132, Aug. 2019, doi: 10.5539/ijbm.v14n9p132., in press.

[13] S. Nargesi and G. Ali Bazaee, "A study on the effect of business process reengineering on the information technology management improvement in knowledge-based IT organizations," Int. J. Nonlinear Anal. Appl, vol. 13, pp. 2008–6822, 2022, doi: 10.22075/ijnaa.2022.26867.3434., in press.

[14] E. Maraizia and J. Swartz, "Challenges to the implementation of business process re-engineering of the recruitment process in the ministry of fisheries and marine resources, Namibia," 2018. [Online]. Available: https://scholar.sun.ac.za.

[15] H. Dinata, "Business process reengineering: the role of information technology as a determinant of success for improving performanc," Inform : Jurnal Ilmiah Bidang Teknologi Informasi dan Komunikasi, vol. 5, no. 1, pp. 25–31, Feb. 2020, doi: 10.25139/inform.v5i1.2255. in press.

ISSN 2355-0082

Integrated System Design Of Sales And Production Module Using RAD Method (Case Study: PT Shafira Putri Kreatif)

Alvin Dzaki Hernindyaputra[1], Monika Evelin Johan[2], David Tjahjana[3]
[1,2,3] Faculty of Informatics, Department of Information Systems, Universitas Multimedia Nusantara, Tangerang, Indonesia
[1]alvin.hernindyaputra@student.umn.ac.id
[2]monika.evelin@umn.ac.id
[3]david.tjahjana@lecturer.umn.ac.id

Accepted 25 January 2023
Approved 09 June 2023

***Abstract*— PT Shafira Putri Kreatif is a company engaged in the field of fashion and apparel for men and women. In business processes, communication is carried out manually in the form of face-to-face or in-office meetings between subsidiaries. This results in inappropriate or forgotten information on production and sales. Therefore, implementing the ERP module is important to help point problems faced by companies to better control its financial performance. The SDLC Rapid Application Development (RAD) method was used. System design is done by making DFD and ERD. The system design is coded using the PHP programming language with the Laravel framework and SQL database. The result of the research is a website-based application that is integrated with the sales module which can assist in viewing and managing transaction data from subsidiaries.**

Keywords*— *ERP; Sales Module; Design and Build; SDLC RAD; Website

I. INTRODUCTION

ERP or Enterprise resource planning is a method used by the industry in carrying out business processes more efficiently by sharing information for business processes and running on a system that is integrated with each other in carrying out company operations, production or distribution [1].

PT Shafira Putri Kreatif is a company engaged in the field of fashion and apparel for men and women. PT Shafira Putri Kreatif has several subsidiaries that carry out certain fashion manufacturing fields, these companies consist of: Ina Butik (engaged in perfecting dresses, kebaya and hijab), Lumonggasari (making clothes with minimalist and elegant fashion styles) and Ads Collection (producing various uniforms for work, school and other general fashion clothing).

PT Shafira Putri Kreatif in conducting communication between subsidiaries is done manually, in the form of face-to-face or in-office meetings. From the existing business processes, the information submitted is not appropriate and the production business processes at PT Shafira Putri Kreatif are not fully running properly or according to needs. From the existing system, a new system is needed to help provide production and sales information at PT Shafira Putri Kreatif by creating an ERP system that can be used as a guide in carrying out business processes by implementing best practices so as to increase productivity, reduce waste and improve product quality and create data standards. and information through uniform reports [1].

Based on the description that has been explained, in this study carried out ERP development using the RAD (Rapid Application Development). The RAD method is a linear sequential software development process model that emphasizes a very short development cycle (range of 60 to 90 days). The development of the RAD model will shorten the development time in the software development cycle between system design and implementation [2]. In the RAD method there are Requirements Planning, Design Workshop (Design Process) and Implementation (Implementation) stages [3]. In conducting research, this method is needed to design several important problem points faced by the company, as well as the solutions offered in the modernization of the PT Shafira Putri Kreatif system.

II. METHOD

In this study, a reference or comparison of previous research was carried out in finding material and avoiding material similarities.

Research that carried out ERP implementation was carried out by Fakung Rahman on report presentation. PT Surya Citra Television has succeeded in implementing SAP by creating a design that applies ERP concepts and financial reports. The system uses SAP R/3 which can present financial reports quickly, accurately, facts and objective [4].

Subsequent research using the Rapid Application Development (RAD) method was conducted by Jansen Wiratama, Hari Santoso and Sobiyanto. In this research, the dashboarding management system executive monitoring project progress determines project feasibility using the forecasting method approach (Case Study: PT Rajawali Mas Mandiri). In this study, it was successful in making dashboards for project monitoring using the RAD software development method. The research results are used as a support in project monitoring activities for corporate executives at PT Rajawali Mas Mandiri [5].

Research conducted by Ririn Ikana Desanti, Carolyn Feiby Supit, Andree E. Widjaja made an employee recruitment and appraisal application using the RAD software development method with a prototype approach. The results of the research resulted in a web-based employee recruitment and appraisal application at PT. XYZ which was developed using the RAD method. This application can help (support) the HRD department [6].

In this study using the RAD software development method in the design of an integrated sales module system. The RAD method was chosen because the RAD method is suitable for projects that require a short time, namely the range of 60 to 90 days and produces a system that meets the immediate needs of the customer [1]. The RAD method has the stages described in Figure 3.1 for carrying out the design of an integrated sales module system at PT Putri Shafira Kreatif.

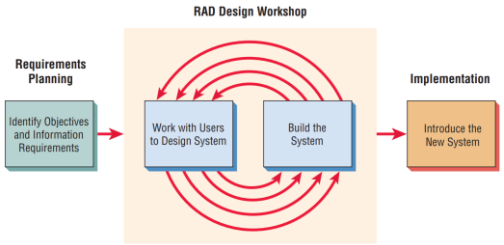


Fig 1. RAD method

The RAD design for designing the ERP system required for PT Shafira Putri Kreatif has stages which can be explained as follows:

A. Requirements Planning

At this stage it is carried out to identify the needs, limitations and objectivity of the system to be built by collecting data at PT Shafira Putri Kreatif. This stage is carried out through an interview process and distributing questionnaires that aim to obtain information about ongoing business processes and problems that occur at PT Shafira Putri Kreatif. After collecting data, a solution was found, namely the design of an integrated sales module system with the RAD method at PT Shafira Putri Kreatif.

B. RAD Design Workshop

a) Design System

At this stage, modeling is carried out based on the results of data collection in the previous stage. In this study the modeling that was carried out was DFD design and database design by making the ERD system to be built.

b) Build the System

At this stage, the implementation of the system is carried out into the coding program and database. In this study the database used is SQL and the PHP programming language with the Laravel framework

C. Implementation

At this stage, implementing a system that has been approved by PT Shafira Putri Kreatif. At this stage it is customary to provide feedback on the system that has been made and to obtain approval for the system.

III. RESULT AND DISCUSSION

In conducting this research using the RAD software development method. In system design, DFD (Data Flow Diagram) and ERD (Entity Relationship Diagram) design is made. The coding is done using the PHP programming language with the Laravel framework and SQL database.

A. Requirements Planning

Requirements planning is a stage for defining what system requirements are needed in building a system that is integrated with the sales module at PT Shafira Putri Kreatif using the RAD method. In this study, the needs analysis consists of functional requirements and non-functional requirements.

Based on the results of interviews and data analysis through distributing questionnaires, the functional requirements of the sales module integrated system at PT Shafira Putri Kreatif using the RAD method are obtained, namely:

1) Admin

Admin is an actor assigned to the center who can manage transaction data from 3 subsidiary branches of PT Shafira Putri Kreatif. Following are

the functional requirements of the admin actor namely:

1. Log in
2. Manage supplier data
3. Manage customer data
4. Manage branch data
5. Manage unit data
6. Manage category data
7. Manage product data
8. Print purchase data
9. Print sales data

2) Branch Admin

The branch admin is an actor in charge of managing product data, in this case, a subsidiary of PT Shafira Putri Kreatif. The following are the functional requirements of the branch admin actor, namely:

1. Log in
2. Manage product data
3. Manage sales data
4. Manage purchasing data
5. Print purchase data
6. Print sales data

Non-functional requirements are stages for defining what devices are needed in building a system, both software and hardware. The following are the non-functional requirements of the system to be built, namely:

1) Hardware Requirements

Hardware is a device in physical form with specifications to run the system. The following hardware is used as follows:

a. Computer or Laptop
b. Intel Core i3 processor
b. 3.4GB of RAM
c. 500GB hard drive
d. Mouse and Keyboard

2) Software requirements

Software requirements is an application that is needed to be able to build the system to be made. The software used is:

a. Operating System: Windows 10
b. Databases: SQL
c. Web Server: XAMPP
d. Web Browsers: Mozilla, Chrome
e. Text Editor: Visual Studio Code
f. Draw.io

B. RAD Design Workshop

The design process is the stage for creating a design based on the results of the needs analysis. In this study, the design for the integrated system design of the sales module at PT Shafira Putri Kreatif was made using the RAD method, namely making a system design in the form of data flow diagrams (DFD) and entity relationship diagrams (ERD).

Data flow diagrams or context diagrams are diagrams that describe the entire process in the system in the sales module integrated system design at PT Shafira Putri Kreatif with the RAD method which is depicted in Fig 2.

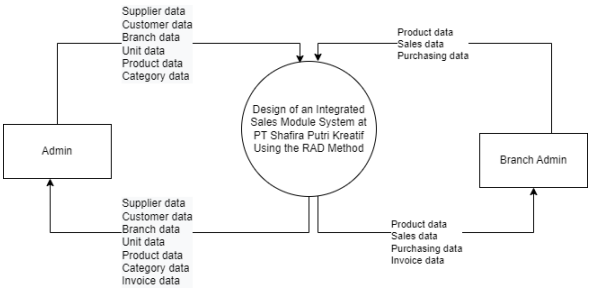


Fig 2. DFD Level 0 Website Sales Module of PT Shafira Putri Kreatif

From Fig 2. DFD Level 0 it can be explained as follows:

a. The admin inputs data on the sales module integrated system design application at PT Shafira Putri Kreatif. The data input is supplier data, customer data, branch data, unit data, product data and category data. Furthermore, the admin can see the results of the data that has been input.
b. The branch admin inputs product data, sales data and purchase data into the system. From this process, the results of product data, sales data and purchase data are obtained.
c. The sales module integrated system collects data on invoices that have been carried out by the branch admin and displays the results to the admin and branch admin.

From DFD Level 0 or context diagram then DFD Level 1 is described as a continuation of DFD Level 0. DFD Level 1 in this study is depicted in Fig 3.

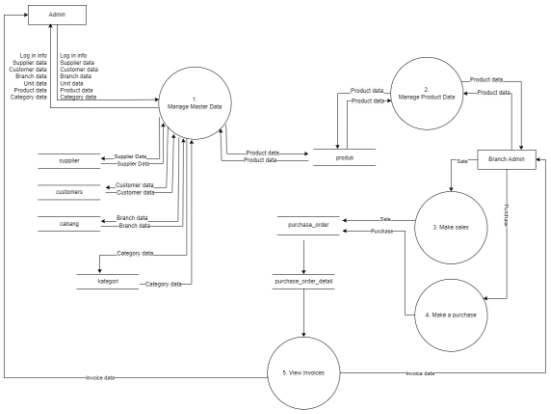


Fig 3. DFD Level 1 Website Sales Module of PT Shafira Putri Kreatif

Explanation of Fig 3. DFD Level 1 will be explained as follows:

a. DFD Level 1 for process number 1 is used to manage master data. In process number 1 it involves an external entity admin. Admin can perform the login process, input supplier data, customer data, branch data, category data and product data on the system which is stored in data store suppliers, customers, branches, categories and products.
b. DFD Level 1 for process number 2 is used to manage product data. In process number 2 it involves an external entity admin and branch admin. Admin and branch admin input product data on the system which is stored in the product data store.
c. DFD Level 1 for process number 3 is used to make sales. In process number 3 it involves an external entity admin branch. The branch admin inputs sales data on the system which is stored in the purchase_order and purchase_order_detail data stores.
d. DFD Level 1 for process number 4 is used to make purchases. In process number 4 it involves an external entity admin branch. Branch admins who want to make purchases by inputting purchase data into the system, then the data is stored in the purchase_order and purchase_order_detail data stores.
e. DFD Level 1 for process number 5 is used to view invoices. Invoice data is generated in the system based on purchase_order and purchase_order_detail store data. Invoice data can be seen by external entity admins and branch admins.

Database design is done by making ERD used to describe entities and relationships between entities contained in the system. Each entity in the ERD has attributes, the ERD used is depicted in Fig 4.

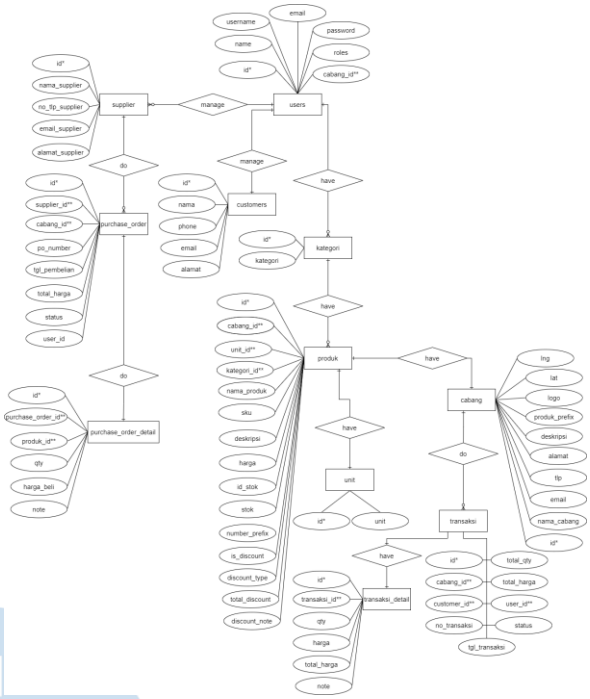


Fig 4. ERD Website Sales Module of PT Shafira Putri Kreatif

C. Implementation

1) Admin

When a user logs in as an admin actor, he has access rights/functions, namely managing user data, suppliers, customers, branches, units, categories, products, purchase orders (PO) and sales and logout.

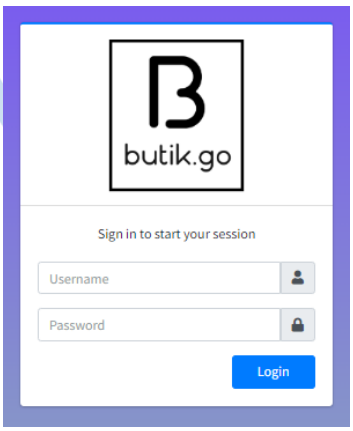


Fig 5. Login Page

When the admin has successfully logged in, they will be directed to the dashboard page as shown in Fig 6.

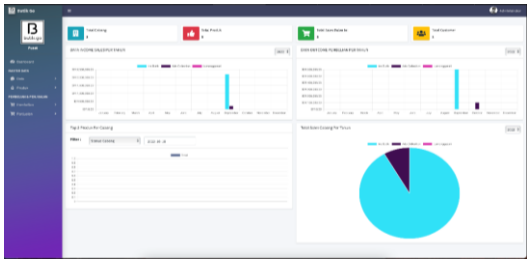


Fig 6. Dashboard

Admin can manage user data such as add, edit and delete user data. display of user data page can be seen in Fig 7.

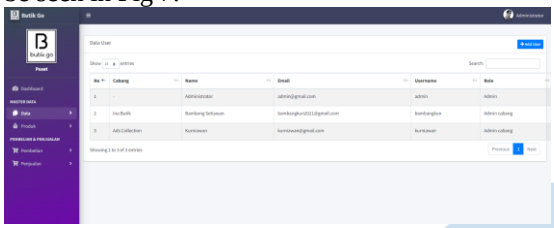


Fig 7. User Data Page

Views for managing category data can be seen in Fig 8. On the category data page you can add, edit and delete category data.

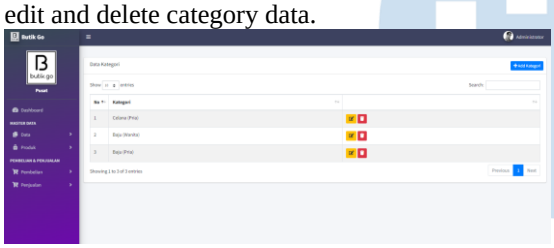


Fig 8. Category Data Page

The product page is used by the admin to view product data and manage product data such as adding, editing and deleting product data as shown in Fig 9.

Fig 9. Product Data Page

The purchase order data page is a page that displays information on purchases made by the company's branches which can be seen in Fig 10. On this page you can manage purchase order data.

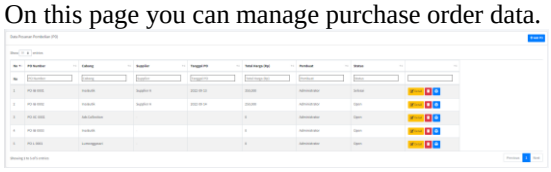


Fig 10. Purchase Order Data Page

The sales data page is a page that displays sales information made by the company's branches for products available at PT Shafira Putri Kreatif which can be seen in Fig 11. On this page you can add sales data and view sales details.

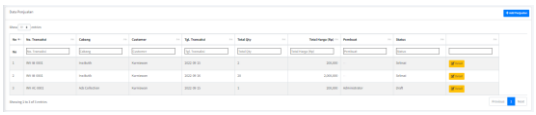


Fig 11. Sales Data Page

Print purchase data is the action used by the admin to print the selected purchase data by pressing the "Print" button. Then the system will display the selected purchase data for printing which can be seen in Fig 12.

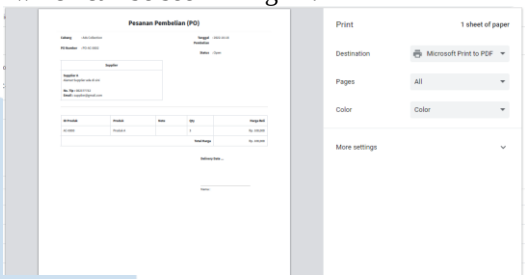


Fig 12. Purchase Data Print Page

2) Branch Admin

When a user enters as a branch admin actor, he has access rights/functions, namely managing customer data, products, purchase orders (PO) and sales and logout.

On the login page, the branch admin is used to enter the system by entering the username and password in Fig 13.

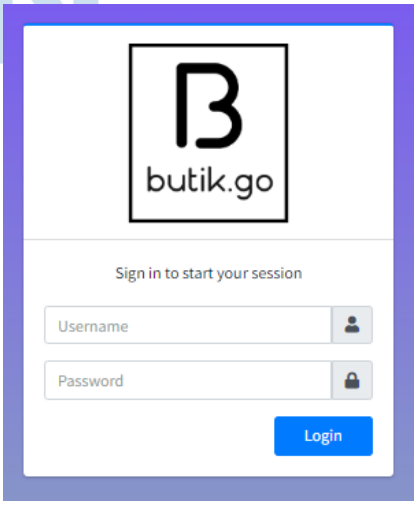


Fig 13 Login Page

The dashboard page shown in Fig 14 is a page that displays brief information regarding the number of branches, products, total sales for this month, total customers, income data graphs, sales outcome data per year and the top 5 products per branch.

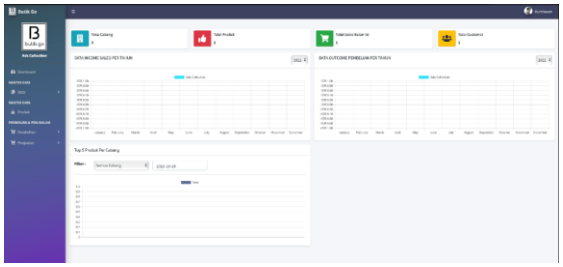


Fig 14. Dashboard Branch Admin

The customer data page is a page that displays customer data information consisting of name, telephone number, email and address. On the customer data page, the branch admin can manage customer data, namely adding, editing and deleting customer data. Implementation of customer data pages can be seen in Fig 15.

Fig 15. Customer Data Page

The product data page is a page that displays available product information. On the product data page, the branch admin can manage product data, namely add, edit and delete product data. Product data page implementation can be seen in Fig 16.

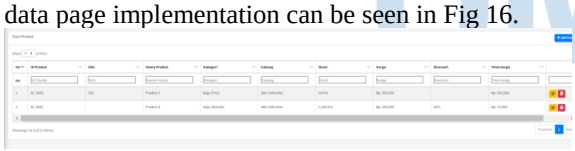


Fig 16. Product Data Page

The purchase data page is a page that displays information on purchase data made by the branch admin for PT Shafira Putri Kreatif. Implementation of purchasing data pages can be seen in Fig 17.

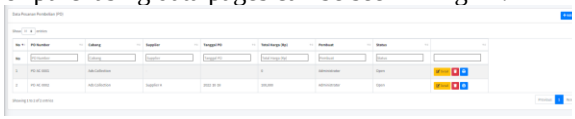


Fig 17. Purchase Order Data Page

Add purchase data is a page that is used by the branch admin to add purchase data made to PT Shafira Putri Kreatif's products. Add purchase data by pressing the "Add PO" button. Then the system displays a pop-up select a branch and selects the Ads Collection branch which can be seen in Fig 18.

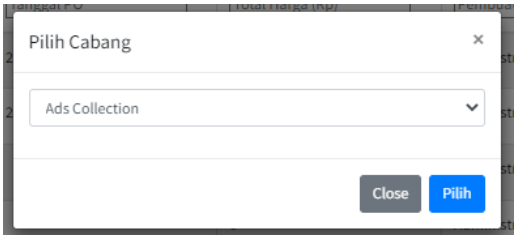


Fig 18. Add Purchase Data page

In Fig 19 it is used to add purchasing data such as products purchased, date of purchase, supplier and purchase status.

Fig 19. Successful Add Purchase Data Page

Branch admins can add purchased product data by pressing the "Add Product" button. Then the system displays the add product form and press the "Add Product" button to save the product data. Implementation of the added product page on purchases can be seen in Fig 20.

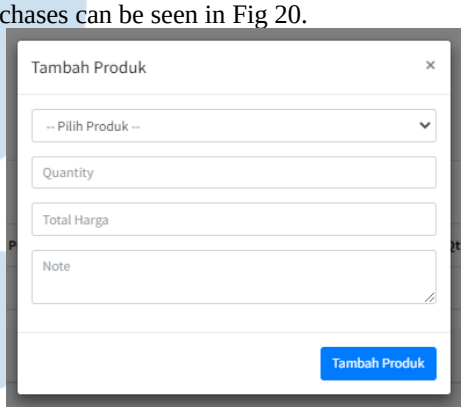


Fig 20. Add Product Data Page on Purchase

When successfully adding product data to a purchase it will appear in Fig 21.

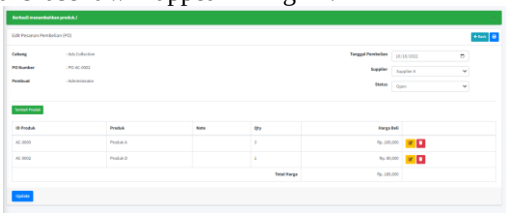


Fig 21. Successful Add Product Page On Purchase

The sales data page is a page that displays information on sales data at PT Shafira Putri Kreatif which consists of transaction no, branch, customer, transaction date, total qty, total price, maker and status. On the sales data page, the branch admin can manage data, namely adding sales data

and viewing detailed sales data. Implementation of sales data pages can be seen in Fig 22.

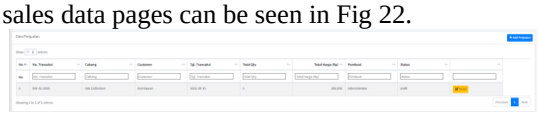


Fig 22. Sales Data Page

Add sales data is a page used by the branch admin to add sales data by pressing the "Add Sales" button. Then the system displays an added sales pop-up which can be seen in Fig 23.

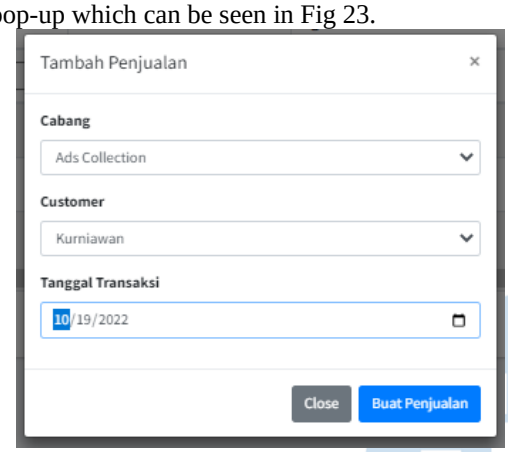


Fig 23. Add Sales Data Page

The successful page for adding sales can be seen in Fig 24. On this page you can add products, set status and print sales data.

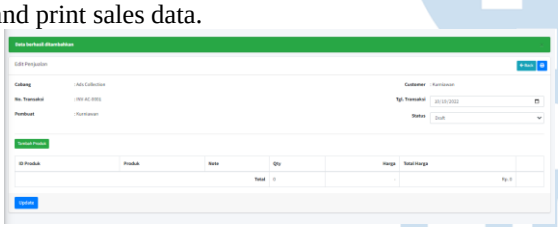


Fig 24. Successful Add Sales Data Page

When the product data has been added, it will be displayed on the sales data product list which can be seen in Fig 25.

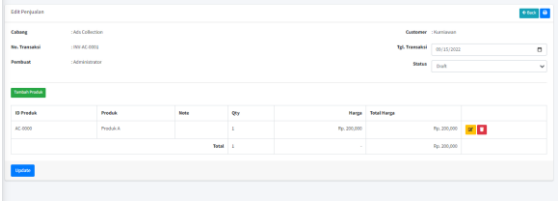


Fig 25. Sales Data Detail Page

On the sales data detail page, the branch admin can print sales transactions by pressing the "Print" button and displaying the printed page as shown in Fig 26.

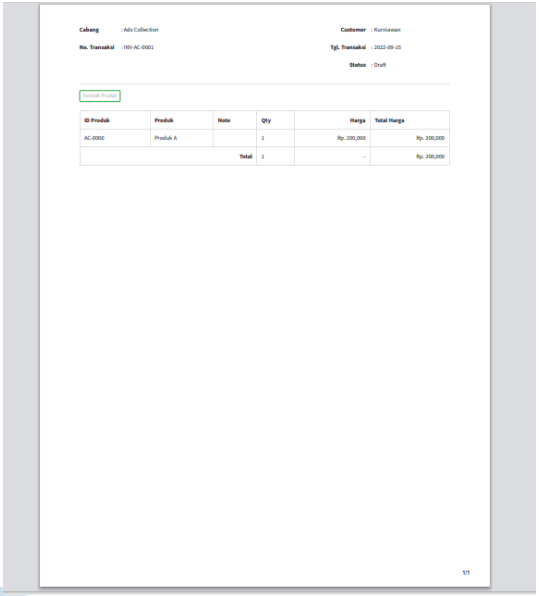


Fig 26. Sales Data Print Page

D. System Testing

System testing is carried out using the black box method which tests the functionality of the system that has been made. From the results of the tests that have been carried out, the results show that all functions can run properly.

Subsequent tests were carried out on application users, namely testing the user acceptance test (UAT) by making a questionnaire related to this application. Respondents used amounted to 9 people.

Calculation of UAT on the results of the questionnaire is carried out as follows:

1. Calculate the answer score
   Calculation of the total score of the respondent's answers was carried out on all the answers chosen by the respondent and then added up.
   The results of calculating the total score are as follows:
   SS = 19 X 5 = 95
   S = 46 X 4 = 184
   N = 8 X 3 = 24
   TS = 7 X 2 = 14
   STS = 10 X 1 = 10
   Total = 327
2. Calculating the highest score (X)
   To calculate the highest score is done as follows.

$$X = Highest\ score \times (number\ of\ questions \times number\ of\ respondents) \quad (1)$$

= 5 x (10 x 9)
= 450

3. Calculate the lowest score (Y)
   To calculate the lowest score is done as follows

   $Y = Lowest\ score \times (jnumber\ of\ questions \times number\ of\ respondents)$
   (2)
   = 1 x (10 x 9)
   = 90
4. Calculation of the percentage of UAT
   After obtaining the number of respondents' scores, the UAT percentage calculation is carried out using the following.

$$P = \frac{f}{n} \times 100\ \% \quad (3)$$

Information:
P = Percentage
F = Response Frequency
n = Number of Respondents

$$P = \frac{327}{450} \times 100\ \%$$
$$P = 72.6\ \%$$

From the results obtained it was concluded that the total score obtained was 90 (72.6%) of the highest score of 450 (100%). This shows that the implementation of an integrated sales module system at PT Shafira Putri Kreatif is included in the Feasible category because it is in the range of 60 – 80%.

IV. CONCLUSION

From the results of the research on the integrated system design of the sales module at PT Shafira Putri Kreatif using the RAD method, it can be concluded that:

1. The system can record and store reports on goods, sales, suppliers and customers at PT Shafira Putri Kreatif, namely with an admin actor who can control data with the feature of managing data on suppliers, customers, branches, units, categories, products and invoices from 3 subsidiary companies PT Shafira Putri Kreatif.
2. How the system works in connecting entry, exit, and the state of stock of goods in each subsidiary, that is, in the system there is a branch admin actor who can manage product data, sales and purchases which can be displayed on the admin dashboard.
3. Prioritized ERP implementation in this study, namely production and sales at PT Shafira Putri Kreatif in the form of a website that can establish communication between subsidiaries at PT Shafira Putri Kreatif. The results of system testing using the blackbox method show that all features can be carried out according to their expected functions with minimal errors/bugs and a fairly effective and easy-to-use appearance for novice users. Furthermore, a user acceptance test (UAT) test was carried out which obtained a percentage of 72.6% which was included in the "Decent" feasibility category.

Based on the above conclusions, suggestions are made for further research, namely:

1. For further research, it is hoped that it will be able to deepen the application of Enterprise Resource Planning (ERP) in business processes carried out in a company such as considering the production raw materials used, the production processes undertaken.
2. For further system development, it is necessary to evaluate existing/implemented ERP modules, then carry out an analysis to find out the needs of further users.

REFERENCES

[1] N. R. Dewanti, "SEBUAH PEMAHAMAN SYSTEM APPLICATION PRODUCT PEMBELIAN MELALUI FOCUS GROUP DISCUSSIONPADA PT. X SURABAYA," *Jurnal Ilmu dan Riset Akuntansi*, vol. 11, 2022.

[2] T. Pricillia and Z. , "Survey Paper: Perbandingan Metode Pengembangan Perangkat Lunak (Waterfall, Prototype, RAD)," *Bangkit Indonesia*, vol. X, 2021.

[3] D. Hariyanto, R. Sastra and F. E. Putri, "Implementasi Metode Rapid Application Development Pada Sistem Informasi Perpustakaan," *Jurnal JUPITER*, vol. XIII, pp. 110 - 117, 2021.

[4] F. Rahman, "EVALUASI PENERAPAN ENTERPRISE RESOURCES PLANNING (ERP) TERHADAP PENYAJIAN LAPORAN KEUANGAN (STUDI KASUS DI PT. SURYA CITRA TELEVISI)," *Jurnal KREATIF : Pemasaran, Sumberdaya Manusia dan Keuangan*, vol. 6, pp. 109-126, 2018.

[5] J. Wiratama, H. Santoso and S. , "DASHBOARDING MANAGEMENT SISTEM EKSEKUTIF MONITORING PROGRESS PROYEK MENENTUKAN KELAYAKAN PROJECT MENGGUNAKAN PENDEKATAN METODE FORCASTING (Studi Kasus: PT Rajawali Mas Mandiri)," *Jurnal Komputer dan Informatika*, vol. 15, pp. 297-307, 2020.

[6] R. I. Desanti, C. F. Supit and A. E. Widjaja, "Aplikasi Perekrutan dan Penilaian Karyawan

Berbasis Web Pada PT. XYZ," *ULTIMA InfoSys*, vol. 8, pp. 74-80, 2017.

[7] H. Purwanto, A. Z. Hananto, F. Maulana and G. Pratama, "PENERAPAN ENTERPRISE RESOURCE PLANNING (ERP) MODUL SALES UNTUKPENINGKATAN PENJUALAN LITTLE INK'S BANDUNG," *Jurnal Ilmiah Teknologi Informasi Terapan*, vol. 7, pp. 205-210, 2021.

[8] C. Tristianto, "PENGGUNAAN METODE WATERFALL UNTUK PENGEMBANGAN SISTEM MONITORING DAN EVALUASI PEMBANGUNAN PEDESAAN," *Jurnal Teknologi Informasi ESIT*, 2018.

[9] Y. W. S. Putra and M. F. Adhim, "Sistem Informasi Presensi Online Menggunakan Teknologi Face Recognition dan GPS," *Jurnal TEKNO KOMPAK*, vol. 16, pp. 149-161, 2022.

[10] D. Irawan, "Pengembangan Sistem Informasi Penagihan Piutang Premi Asuransi Menggunakan Metode RAD," *JURNAL CYBERAREA*, pp. 2(6), 1-9, 2022.

[11] K. E. Kendall and J. E. Kendall, System Analysis and Design 8thed, New Jersey: Prentice Hall, 2010.

[12] A. Prayogo, O. A. Putri and D. M. Kusumawardani, "IMPLEMENTASI ENTERPRISE RESOURCE PLANNING MODUL SALES DENGAN MENGGUNAKAN ODOO PADA PT XXX," *PROSIDING SEMINAR NASIONAL SAINS DAN TEKNOLOGI FAKULTAS TEKNIK UNIVERSITAS WAHID HASYIM*, 2021.

[13] F. N. Hasanah and R. S. Untari, "BUKU AJAR REKAYASA PERANGKAT LUNAK," 15 August 2020. [Online]. Available: https://press.umsida.ac.id/index.php/umsidapress/article/view/978-623-6833-89-6. [Accessed 9 Januari 2023].

[14] D. P. Sari and R. Wijanarko, "Implementasi Framework Laravel pada Sistem Informasi Penyewaan Kamera (Studi Kasus di Rumah Kamera Semarang)," *Jurnal INFORMATIKA dan Rekayasa Perangkat Lunak*, vol. 2, pp. 32-36, 2020.

[15] A. Z. Muchtar and S. Munir, "PERANCANGAN WEB E-COMMERCE UMKM RESTORAN BAKSOAREMAMENGGUNAKAN FRAMEWORK LARAVEL," *Jurnal Teknologi Terpadu*, vol. 5, pp. 26-33, 2019.

[16] M. S. Novendri, A. Saputra and C. E. Firman, "APLIKASI INVENTARIS BARANG PADA MTS NURUL ISLAM DUMAI MENGGUNAKAN PHP DAN MYSQL," *Lentera Dumai*, vol. 10, pp. 46-57, 2019.

[17] H. D. Lumbanraja, "PERANCANGAN SISTEM INFORMASI AKADEMIK ONLINEMENGGUNAKAN BLACK BOX TESTINGPADA SEKOLAH TINGGI ILMU EKONOMI SURYA NUSANTARA," *Jurnal TeIKa*, vol. 8, pp. 9-18, 2018.

[18] U. D. Mariyani, W. Setiyaningsih and R. Agustina, "Pengembangan Sistem Koreksi Jawaban Esai Otomatis Menggunakan Naive Bayes Dan Pengujian Menggunakan User Acceptance Test (UAT)," *Jurnal Terapan Sains &Teknologi*, pp. 61-73, 2022.

[19] D. Azzahra and S. Ramadhani, "PENGEMBANGAN APLIKASI ONLINE PUBLIC ACCESS CATALOG(OPAC) PERPUSTAKAAN BERBASIS WEBPADA STAI AULIAURRASYIDDINTEMBILAHAN," *Jurnal Teknologi Dan Sistem Informasi Bisnis* , 2020.

[20] H. Hasanah, R. Fatullah and M. R. Abdullah, "RANCANG BANGUN APLIKASI PELAYANAN KARYAWAN BERBASIS WEB DI PT ASIA CHEMICAL INDUSTRI," *Jurnal Innovation And Future Technology*, vol. 4, pp. 1-10, 2022.

[21] I. G. N. Suryantara and J. F. Andry, "Development of Medical Record With Extreme Programming SDLC," *IJNMT*, vol. 5, pp. 47-53, 2018.

[22] R. Delima, H. B. Santoso, G. H. Aditya, J. Purwadi and A. Wibowo, "Development of Sales Modules for Agricultural E-Commerce Using Dynamic System Development Method," *IJNMT*, vol. 5, pp. 95-103, 2018.

ISSN 2355-0082

DistilBERT with Adam Optimizer Tuning for Text-based Emotion Detection

Farica Perdana Putri[1]
[1]Department of Informatics, Universitas Multimedia Nusantara, Tangerang, Indonesia
farica@umn.ac.id

Accepted 31 May 2023
Approved 11 July 2023

***Abstract*— Emotion detection (ED) refers to identifying individual emotions or feelings, such as happiness, sadness, disappointment, fear, etc. The classic machine learning technique still relies on feature engineering, which makes it difficult to convey the meaning of words. Deep learning-based algorithms have recently been shown to be beneficial for emotion detection because they require only a simple feature creation process. Transfer learning is an approach that uses data similarities, data distribution, models, tasks, and other factors to apply knowledge learned in one domain to a new domain. This study is to shed light on the fine-tuned models' efficacy in detecting emotions from the International Survey on Emotion Antecedents and Reactions (ISEAR) dataset. In order to optimize the model, we conducted the hyperparameters tuning on the Adam optimizer in DistilBERT. The experiment examined the moment estimators and learning rate of the Adam optimizer. The effect of the parameters on training and validation accuracy was presented and analyzed. Adam optimization first-moment estimators provide more robust convergence to the model during the training process as their value approaches one. The testing results of emotion detection is 97.14%..**

Keywords*— *emotion detection; transfer learning; Adam optimizer; DistilBERT.

I. INTRODUCTION

The advancement of internet technology results in a rising amount of multimedia data. Multimedia data includes text, speech, images, and video and is produced and contains a vast amount of information. One type of digital content that is shared online is textual data. Social media generates millions of messages per day, necessitating immediate unstructured data processing [1]–[3]. Opinion mining has become essential in recent studies as a result of this occurrence [4]. Users can express or share their actions, thoughts, views, and emotions on social media [5]. As a result, a lot of businesses use this opportunity to analyze user-generated data as a source of analysis for internal decision-making [6], [7].

Emotion detection (ED) refers to identifying individual emotions or feelings, such as happiness, sadness, disappointment, fear, etc. Emotions can be detected in a person's voice, facial expressions, movements, and writing. This is based on the opinion that when someone is sad, he will use words that are not encouraging. Emotional acceptance and detection are used in various fields, such as healthcare, education, and advertising [8]. In healthcare, emotions or feelings can affect a person's health. Depressed emotions or depression can drain mental energy. Therefore, it harms the body and causes a decrease in health. Meanwhile, a joyful atmosphere can help restore one's health. In education, emotions play an essential role in the quality and quantity of student learning. Brain activity will increase when children experience positive emotions so that they can concentrate better. Emotional marketing is currently in great demand by companies because of its effectiveness in attracting consumer interest. Emotional marketing is a marketing and advertising effort that uses emotions to make audiences pay more attention to, remember, and make purchases. Many researchers are interested in improving the relationship between humans and computers using emotional extraction on social media, such as Twitter, Instagram, YouTube, Facebook, etc. [9]. Text-based emotion detection can be applied using NLP techniques to analyze the semantic meaning of a text.

Different from emotion detection research from faces [8] and voices [10], text detection is more challenging because it requires careful modeling of text since words associate with different emotions in different contexts with varying levels of magnitude making the identification of words for document representation more challenging [11]. Online social media content is short, informal, and unstructured text comprising incomplete and misspelled words, abbreviations, acronyms, and special characters. Moreover, like most NLP, we face problems such as negative words and anaphora which can change the meaning of sentences. An example of a negative word:

'I am not sad'. Sad is a word that shows implicit emotion, but the presence of the word not before sad makes the meaning of the sentence change to be similar to happy. The classification could be wrong if the computer cannot grasp this meaning. Anaphora refers to using a grammatical substitute like a pro-noun or pro-verb to denote a preceding word or a group of words.

Previous studies recognized human emotion from text using a rule-based and feature representation technique. The rules of emotion are then derived utilizing statistics, linguistics, and computation techniques. The best rules are chosen later. The rules are then applied to emotion datasets to determine emotion labels. Dibyendu et al. [12] proposed semantic rules to identify emotion at the sentence level. The method examines emotion words and phrasal verbs, as well as negation words, and performs better than previous approaches. presents a rule-based technique for detecting the emotion or mood of a tweet and categorizing it in the proper emotional category. The drawbacks of rule-based systems are a lack of contextual meaning and a lexicon with insufficient words.

The feature-based method relies heavily on machine learning techniques. It has improved classification accuracy by using appropriate models and characteristics to categorize text, but its efficiency is often lower than that based on the emotion dictionary [5]. The classic machine learning technique still relies on feature engineering [13], which makes it difficult to convey the meaning of words. Singh et al. [14] proposed the feature combination using semantic and statistical of words during the selection of significant features to construct the word vector. The framework consists of two stages: semantic-based to extract the meaningful words using POS tagger and statistical-based to remove the weak semantic feature using the Chi-square method.

Pang et al. [15] construct the topic-level feature space by grouping semantically related words. They develop the weighted labeled topic model (WLTM) and X-term emotion-topic model (XETM) to detect emotions toward certain topics. WLTM defines one-to-many mapping between emotions and multiple topics. XETM uses emotion distributions of labeled documents to constrain the topic probability of each feature during the training process. J.Guo [16] utilized deep learning and natural language processing to detect humans' emotions in text. They combined the questionnaire-based approach and the text-analysis-based approach features as feature vectors in the prior classifier. The method produced a human detection emotion rate of 97.22% and a classification emotion rate of 98.02%. Anzum and Gavrilova [4] introduced a novel approach of feature representation from Twitter based on Genetic Algorithm (GA). It is composed of stylistic, sentiment, and linguistic features extracted from tweets data. They used an ensemble classifier with weights optimized by GA to increase the detection accuracy.

Deep learning-based algorithms have recently been shown to be beneficial for emotion detection because they require only a simple feature creation process. One of the primary benefits of deep learning-based text mining approaches is efficient feature engineering. Transfer learning is an approach that uses data similarities, data distribution, models, tasks, and other factors to apply knowledge learnt in one domain to a new domain [17]. The labeled data can be utilized to construct the model and employed in the target domain data to enhance the annotation of the target data via transfer learning. Transfer learning starts with a pre-trained model, and fine-tuning involves further training the pre-trained model on the new task by updating its weights. Thus, the purpose of this study is to shed light on the fine-tuned models' efficacy in detecting emotions from the International Survey on Emotion Antecedents and Reactions (ISEAR) dataset [18].

II. METHOD

1. *Data Acquisition*

The ISEAR project was coordinated by Klaus R. Scherer and Harald Wallbott during the 1990s. Students, both psychologists and non-psychologists, were asked to describe scenarios in which they had felt all seven major emotions (joy, fear, anger, sadness, disgust, shame, and guilt). Thus, the final data set included reports on seven emotions from nearly 3000 respondents from 37 countries across all five continents. As we can see in Figure 1, the distribution of classes in the dataset is balanced.

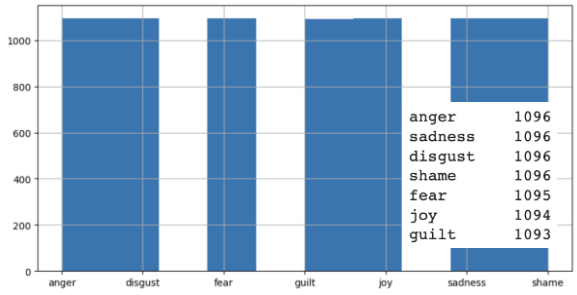


Fig. 1. Distribution of classes in the ISEAR dataset

2. *Data Preprocessing*

In this preprocessing step, the null values are checked and encode the classes from categorical to numerical values used in the fine-tuning process. The training and testing data is 80% and 20%, respectively. We used 20% of training data as validation data. The details of the number of training and testing data each class are described in Table 1.

TABLE I. NUMBER OF TRAINING AND TESTING DATA PER CLASS

	Number of training data	6132	**Number of testing data**	1534
Classes	Joy	875	Joy	219
	Fear	876	Fear	219
	Anger	877	Anger	219
	Sadness	877	Sadness	219
	Disgust	877	Disgust	219
	Shame	876	Shame	220
	Guilt	874	Guilt	219

3. *Fine-tuning Model*

In this work, human emotion detection is developed using fine-tuned DistilBERT. BERT [19] is a bidirectional transformer that was trained on a huge corpus of the Toronto Book Corpus and Wikipedia using a combination of masked language modeling objectives and next sentence prediction. BERT, in contrast to current language representation models, is intended to pre-train deep bidirectional representations from unlabeled text by conditioning on both left and right context in all layers.

DistillBERT is a lightweight, simple, inexpensive transformer model based on the BERT architecture. DistillBERT reduces 40% of the size of the BERT during the training phase using the knowledge distillation technique while 60% faster at inference time. Knowledge distillation is a technique used to transfer knowledge from a larger model, called the teacher, to a smaller model, called the student. DistilBERT retains 97% of the language understanding capabilities [20].

4. *Adam Optimizer*

In order to optimize the model, we conducted the hyperparameters tuning on the Adam optimizer. Adam [21] is an adaptive learning rate optimization algorithm that's been designed specifically for training deep neural networks. However, the hyperparameters have intuitive interpretations and typically require little tuning. The Adam update rule is as follows:

$$m_t = \beta_1 m_{t-1} + (1-\beta_1) g_t \quad (1)$$

$$v_t = \beta_2 v_{t-1} + (1-\beta_2) g_t^2 \quad (2)$$

where m and v are moving averages, g is gradient on the current batch, t is the number of iterations and β_1 and β_2 are hyper-parameters of the algorithm. β_1 is the exponential decay rate of the first momentum estimate and β_2 is the exponential decay rate of the second momentum estimate. This value should be set close to 1.0 on problems with a sparse gradient (e.g. NLP and computer vision problems).

Moving averages are set to zero at the beginning of each iteration. By evaluating the predicted values of moving averages, one can deduce moving averages' correlation with moments. As a result, the correction step removes the bias from the first and second moments caused by the optimizer's bias toward zero due to our zero initialization. Following estimator bias adjustment, the expected value becomes the desired value. Eqs. (3) and (4) provide bias-corrected estimators for the first and second moments.

$$\widehat{m}_t = \frac{m_t}{1-\beta_1^t} \quad (3)$$

$$\widehat{v}_t = \frac{v_t}{1-\beta_2^t} \quad (4)$$

The algorithm's last step is to employ moving averages to scale the learning rate independently for each parameter. To apply this step, compute the weight update using Eq. (5).

$$w_t = w_{t-1} - \alpha \frac{\widehat{m}_t}{\sqrt{\widehat{v}_t} + \epsilon} \quad (5)$$

where w is model weight, α is the learning rate or step size, t is the number of iterations, and ϵ is to prevent division by zero and the default value is 10^{-8}.

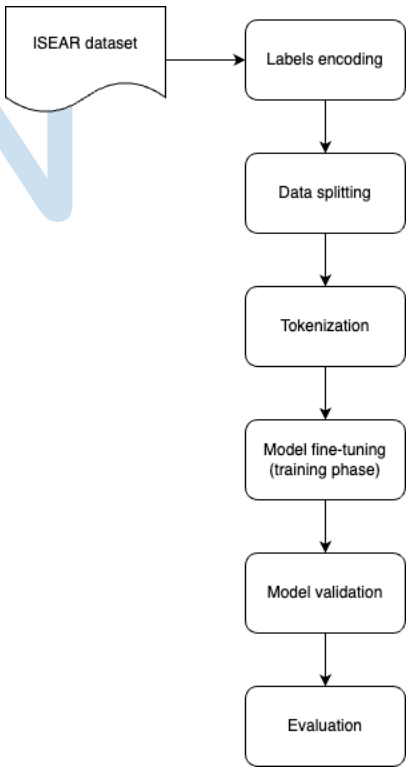


Fig. 2. Pipeline of the proposed system

III. Result and Discussion

The overall stages are explained for the proposed emotion detection system in the diagram in Figure 2. Firstly, the labels of the ISEAR dataset are transformed into numerical values ranges from 0 to 6 as depicted in Figure 3. There are seven classes include: joy, fear, sadness, shame, disgust, guilt and anger. Total number of samples is 7666 and divided into training, validation, and testing data based on the defined percentage.

Words in the sentence are tokenized using pre-trained uncased DistilBERT tokenizer specific for emotion detection. Tokenization is used in natural language processing to split paragraphs and sentences into smaller units that can be more easily assigned meaning. The results then forwarded to the model to perform the fine-tuning process. We created the customized model, by adding a drop out and a dense layer on top of DistilBERT to get the final output for the model.

	Emotion	Encode_Emot
0	joy	0
1	fear	1
2	anger	2
3	sadness	3
4	disgust	4
5	shame	5
6	guilt	6

Fig. 3. Encoded labels of the ISEAR dataset

During the validation stage, we pass the unseen data from the validation dataset to the model. This step determines how well the model performs on the unseen data before we evaluate it. During the validation stage, the weights of the model are not updated. Only the final output is compared to the actual value. This comparison is then used to calculate the accuracy of the model. Finally, the testing data is predicted using the best parameters.

The model parameters are described in Table 2. The number of batch sizes and epochs were determined by training the model with a learning rate of 0.00001 and default values of β_1 and β_2, which are 0.9 and 0.999, respectively. Different batch sizes (4, 8, and 16) and epochs (5, 10, and 15) have been tested. The results show that the best training and validation accuracy obtained from batch size is equal to 4 and epoch of 10, as shown in Table 3. Since the first and second moment estimators of the Adam optimizer can be evaluated for different learning rates, the objective is to compare the combination of different hyper-parameters in each training process. The accuracy of the training and validation processes for each possible hyperparameter is shown in Table 4.

Different learning rate parameters are taken into account while modifying the first and second moment estimations, which are compared among themselves. Option 1 is the best hyperparameters selection compared to other hyperparameters pairs. The overall training and validation accuracy become smaller while the learning rate is increased. A low learning rate would result in slower model training, requiring many parameter updates to reach the point of minimum. A high learning rate, on the other hand, would imply huge steps or abrupt modifications to the parameters, which frequently leads to divergence rather than convergence. Selected hyperparameters in Options 9-12, with a learning rate of 0.001, are not suitable for the network. It means the model get stuck in local optima because the learning parameter is too high. In most cases, the training accuracy is directly proportional to validation accuracy when β_1 is closer to 1. The increase in β_1 reduces the effect of gradients on the moving average of variance and provides more stable convergence. The testing accuracy using the best hyperparameters selection is 97.14%.

TABLE II. The Proposed DistilBERT Parameters

Parameter	Value
Batch size	[4, 8, 16]
Epochs	[5, 10, 15]
Tokenizer	Distilbert-base-uncased-emotion
Dropout	0.3
Optimizer	Adam
Loss function	Cross-entropy

TABLE III. The Results for each batch size and epoch selection

Epoch	Batch Size	Training Accuracy	Validation Accuracy
5	4	93.65	98.68
5	8	93.37	98.84
5	16	85.57	88.68
10	**4**	**97.80**	**97.61**
10	8	97.73	97.68
10	16	97.39	97.35
15	4	98.10	96.60
15	8	98.03	96.77
15	16	97.86	95.60

TABLE IV. THE RESULTS FOR EACH HYPER-PARAMETER SELECTION

Option	Learning Rate	β_1	β_2	Training Accuracy	Validation Accuracy
1	**1e-05**	**0**	**0.99**	**98.57**	**96.39**
2	1e-05	0	0.999	97.57	95.33
3	1e-05	0.9	0.99	98.51	94.76
4	1e-05	0.9	0.999	97.79	97.21
5	1e-04	0	0.99	94.49	87.67
6	1e-04	0	0.999	92.03	78.37
7	1e-04	0.9	0.99	96.82	83.43
8	1e-04	0.9	0.999	95.48	80.90
9	1e-03	0	0.99	13.73	13.70
10	1e-03	0	0.999	13.98	13.87
11	1e-03	0.9	0.99	13.90	13.88
12	1e-03	0.9	0.999	14.14	13.70

.

IV. CONCLUSION

In this study, various hyperparameters of the Adam optimizer are tested on the ISEAR dataset to detect human emotion from text. According to the conducted experiments, the best hyperparameter selection was Option 1. We conclude that training and validation accuracy are lower as the learning rate increases. Adam optimization first-moment estimators provide more stable convergence to the model during the training process as their value is close to 1. For the future works.

REFERENCES

[1] A. R. Murthy and K. M. Anil Kumar, "A Review of Different Approaches for Detecting Emotion from Text," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1110, no. 1, p. 012009, Mar. 2021, doi: 10.1088/1757-899X/1110/1/012009.

[2] P. Nandwani and R. Verma, "A review on sentiment analysis and emotion detection from text," *Soc. Netw. Anal. Min.*, vol. 11, no. 1, p. 81, Dec. 2021, doi: 10.1007/s13278-021-00776-6.

[3] K. Shrivastava, S. Kumar, and D. K. Jain, "An effective approach for emotion detection in multimedia text data using sequence based convolutional neural network," *Multimed. Tools Appl.*, vol. 78, no. 20, pp. 29607–29639, Oct. 2019, doi: 10.1007/s11042-019-07813-9.

[4] F. Anzum and M. L. Gavrilova, "Emotion Detection From Micro-Blogs Using Novel Input Representation," *IEEE Access*, vol. 11, pp. 19512–19522, 2023, doi: 10.1109/ACCESS.2023.3248506.

[5] Y. Qian, W. Liu, and J. Huang, "A Self-Attentive Convolutional Neural Networks for Emotion Classification on User-Generated Contents," *IEEE Access*, vol. 8, pp. 154198–154208, 2020, doi: 10.1109/ACCESS.2019.2938560.

[6] A. Rusli, A. Suryadibrata, S. B. Nusantara, and J. C. Young, "A Comparison of Traditional Machine Learning Approaches for Supervised Feedback Classification in Bahasa Indonesia," *IJNMT Int. J. New Media Technol.*, vol. 7, no. 1, pp. 28–32, Jul. 2020, doi: 10.31937/ijnmt.v1i1.1485.

[7] G. P. Wiratama and A. Rusli, "Sentiment Analysis of Application User Feedback in Bahasa Indonesia Using Multinomial Naive Bayes," in *2019 5th International Conference on New Media Studies (CONMEDIA)*, Bali, Indonesia: IEEE, Oct. 2019, pp. 223–227. doi: 10.1109/CONMEDIA46929.2019.8981850.

[8] H. Zhang, A. Jolfaei, and M. Alazab, "A Face Emotion Recognition Method Using Convolutional Neural Network and Image Edge Computing," *IEEE Access*, vol. 7, pp. 159081–159089, 2019, doi: 10.1109/ACCESS.2019.2949741.

[9] M. N. Meqdad, F. Abdali-Mohammadi, and S. Kadry, "Recognizing emotional state of user based on learning method and conceptual memories," *TELKOMNIKA Telecommun. Comput. Electron. Control*, vol. 18, no. 6, p. 3033, Dec. 2020, doi: 10.12928/telkomnika.v18i6.16756.

[10] M. Gokilavani, H. Katakam, S. A. Basheer, and P. Srinivas, "Ravdness, Crema-D, Tess Based Algorithm for Emotion Recognition Using Speech," in *2022 4th International Conference on Smart Systems and Inventive Technology (ICSSIT)*, Tirunelveli, India: IEEE, Jan. 2022, pp. 1625–1631. doi: 10.1109/ICSSIT53264.2022.9716313.

[11] 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, Jul. 2017, doi: 10.1016/j.patrec.2016.12.009.

[12] D. Seal, U. K. Roy, and R. Basak, "Sentence-Level Emotion Detection from Text Based on Semantic Rules," in *Information and Communication Technology for Sustainable Development*, M. Tuba, S. Akashe, and A. Joshi, Eds., in Advances in Intelligent Systems and Computing, vol. 933. Singapore: Springer Singapore, 2020, pp. 423–430. doi: 10.1007/978-981-13-7166-0_42.

[13] M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," *Science*, vol. 349, no. 6245, pp. 255–260, Jul. 2015, doi: 10.1126/science.aaa8415.

[14] L. Singh, S. Singh, and N. Aggarwal, "Two-Stage Text Feature Selection Method for Human Emotion Recognition," in *Proceedings of 2nd International Conference on Communication, Computing and Networking*, C. R. Krishna, M. Dutta, and R. Kumar, Eds., in Lecture Notes in Networks and Systems, vol. 46. Singapore: Springer Singapore, 2019, pp. 531–538. doi: 10.1007/978-981-13-1217-5_51.

[15] J. Pang *et al.*, "Fast Supervised Topic Models for Short Text Emotion Detection," *IEEE Trans. Cybern.*, vol. 51, no. 2, pp. 815–828, Feb. 2021, doi: 10.1109/TCYB.2019.2940520.

[16] J. Guo, "Deep learning approach to text analysis for human emotion detection from big data," *J. Intell. Syst.*, vol. 31, no. 1, pp. 113–126, Jan. 2022, doi: 10.1515/jisys-2022-0001.

[17] R. Liu, Y. Shi, C. Ji, and M. Jia, "A Survey of Sentiment Analysis Based on Transfer Learning," *IEEE Access*, vol. 7, pp. 85401–85412, 2019, doi: 10.1109/ACCESS.2019.2925059.

[18] K. R. Scherer and H. G. Wallbott, "Evidence for universality and cultural variation of differential emotion response patterning.," *J. Pers. Soc. Psychol.*, vol. 66, no. 2, pp. 310–328, 1994, doi: 10.1037/0022-3514.66.2.310.

[19] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," 2018, doi: 10.48550/ARXIV.1810.04805.

[20] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter." arXiv, Feb. 29, 2020. Accessed: Jan. 08, 2023. [Online]. Available: http://arxiv.org/abs/1910.01108

[21] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," 2014, doi: 10.48550/ARXIV.1412.6980.

ISSN 2355-0082

Developing HIV/AIDS Patient Profile Model Using K-Means Clustering Method

Rena Nainggolan[1], Fenina Adline Twince Tobing[2]
[1] Komputer Akuntansi, Universitas Methodist Indonesia, Medan , Indonesia
[2] Teknik Informatika, Universitas Multimedia Nusantara, Tangerang, Indonesia
[1]renanainggolan@methodist.ac.id

Accepted 08 June 2023
Approved 12 July 2023

Abstract*— RSUD Dr. Pirngadi is one of the Regional General Hospitals in North Sumatra that handles services for HIV/AIDS patients. Patients who actively take ARV therapy every month are low, namely around 20%. People infected with HIV become carriers and transmitters of the HIV throughout HIVes, HIV though they don't feel sick and look healthy, the sufferer still carries HIV. To be able to carry out more effective and efficient handling of the prevention and control from HIV/AIDS transmission, it is very important for the government and related parties, such as the Health Office, Social Services, and Hospital management, especially in VCT/CST (Voluntary Counseling and Testing/Care Support Treatment) to find out about understanding patient profiles, and prevent the development of HIV/AIDS disease transmission is very important. This knowledge can be used by the government to carry out programs that can prevent and break the chain of transmission of the HIV/AIDS virus as early as possible and help those involved in health services to become more familiar with the situation of their patients. This research takes the service area at RSUD Dr. Pirngadi Medan as one of the research domains in the field of data mining with data sources from RSUD Dr. Pirngadi Medan. This is done as information is known to the VCT/CST services at RSDU Dr. Pirngadi Medan. With data groupings like this, it is hoped that the Government or related agencies can create programs and implement them so that they can prevent and overcome the spread of HIV/AIDS in Indonesia. Obtained the number of patients in each cluster where for cluster 1 there were 7 patients and cluster 2 obtained as many as 3 clusters and cluster 3 obtained as many as 10 clusters.9 It is on this basis that the authors are interested in taking the title of the study regarding the formation of a patient profile model using the K-Mean clustering method.

***Keywords*— Clustering; K-Means Clustering; HIV/AIDS, Patient Profile**

I. INTRODUCTION

In Various fields of life today, a lot of data is generated but the data is hidden, to be able to find out the hidden information from the data, needed for to do data processing of the data. Especially in the current, it is very demanding for business people to use computer technology so they can complete. In the current era of globalization, which is called e-commerce, transactions in the Internet world are defined as e-commerce.

Data mining techniques are focused on building methods for disclosing knowledge stored in data and are used to uncover hidden information in data that is not visible on the surface but has the potential to be used. Data mining is a semi-automated process that applies statistical, learning, artificial intelligence, math, machine learning techniques to identify and extract .Until now, many algorithms have been developed to answer this problem, each of these algorithms is used based on the technique/function approach to processing input into the desired output. Broadly speaking, the techniques often used in data mining include association rule mining, clustering, classification, neural networks, and nearest country[1].

Cluster analysis plays an important role in classifying objects. Depending on the application, objects can be signals, customers, patients, news, plants, and others [2]. The algorithm that is often applied for clustering techniques because it makes estimates that are efficient and does not require many parameters is the K-Means AlgorithmK-Means uses predefined k groups (first k groups as centroids) [3].

The K-Means Clustering method is a method for grouping or classifying objects (data) into K-groups (clusters) based on certain criteria. Classification of

data is done by calculating the shortest distance between cluster center data. The main concept of this method is to compile K centroid values or the average (mean) of a group of data with N dimensions, where this method requires that the K value be determined first.. The K-means algorithm starts with the formation of a prototype cluster at the beginning and then iteratively repairs the prototype cluster so that a convergent condition is achieved, namely a condition where no significant changes were found in the prototype cluster [4]. This turnover is measured by applying the objective function D which is defined as the sum or average distance of each data item from its centroid group.

II. METHOD

A. Data Definition

In Webster's New World's Dictionary, it is written that a datum: is something known or assumed*. That is, a datum (a single form of data) is something that is known/assumed. Thus, data can provide an overview of a situation or problem. Meanwhile, data according to the Oxford Dictionary is The Facts. So, it can be concluded that data is something that is known or assumed to be used for analysis, discussion, scientific presentation, or statistical tests. The data can be divided into 4 parts, namely:

1. Types of Data Based on Their Nature.

Data type can be like divided according to their nature, according to their sources, according to how they are obtained, and according to when they are collected. The nature of the data can be divided into two types, namely qualitative data (non-metric) and quantitative data (metric). Then the type of qualitative data is further divided into two types, namely nominal data and ordinal data. Likewise, the type of quantitative data is divided into two types, namely interval data and ratio data.

2. Types of Data by Source

The division of data types according to their sources is based on the sources of data acquisition, namely internal data and external data. Data grouped by an organization to define organizational conditions or activities that are related and useful for daily needs and internal control. For example production data, sales data at the company, employment data, financial data, and so on are internal data

External data is data collected to describe a situation or activity outside the organization. Examples of external data such as population data and national income data were obtained from the local statistical center office. A company needs external data such as population to predict demand potential, while national income data to determine the level of people's purchasing which is useful for the basis of price level policy.

3. Types of Data According to How to Obtain It.

Based on how to get it, data is divided into two types, namely secondary data and primary data. Secondary data is data obtained in finished form and has been processed by other parties. Secondary data is usually in the form of publications while primary data is data that is processed and collected by individuals or organizations, which are taken directly from the object. For example, a company wants to obtain the average use of a product by residents in a place by conducting direct interviews with the community.

4. Types of Data According to Time of Collection.

Based on the time of collection, data can be divided into two types, namely cross-sectional data and periodic data (time series). Cross-section data is data collected in a certain period, usually describing the conditions or activities in that period. For example, the results of the 2014 population census describe the condition of Indonesia in 2014 according to age, gender, religion, level of education, and so on.

Periodic data (time series) is data collected from time to time. Its purpose is to describe the development of activity over time. For example, the development of production in a company over the last five years, the development of product sales over the last five years, and so on. This type of data is also often referred to as historical data.

B. Data Mining

Data mining is a data processing method to find hidden patterns in data. The results of data processing with this data mining method can be applied to decision making in the future.

Data mining is a method of processing large amounts of data, therefore data mining has an important role in the fields of finance, industry, science, technology and weather. Broadly speaking, data mining studies discuss methods such as classification, clustering, regression, market basket analysis and variable selection, and [4]. Data mining is divided into several groups based on the tasks that can be done, namely

1. Description

Sometimes analytical research just wants to find a way to explain the patterns and trends present in the data. For example, voting committees may not be able to get hold of evidence or facts if they are not professional enough to gain little support in the presidential election. Pattern and trend descriptions often provide possible explanations for a pattern or trend.

2. Estimation

Estimation is almost the same as classification, the difference is that the target variable is estimated more numerically than categorically. The model is built by implementing a complete record that gives the value of the target variable as a predictive value. After that, in the next review, an estimate of the value of the target variable is made based on the value of the predicted variable. For example, in hospital patients it will be based on the patient's age, weight index, gender and blood sodium level to predict systolic blood pressure. The estimation model is generated by the value of predictive variables in the learning process. For other new cases, it can be obtained from other estimation models.

3. Predictions

Prediction is almost the same as estimation and classification, except that it estimates the value of future results. Examples of predictions in business and research are:

A. Estimated percentage increase in traffic accidents in the coming year if the lower speed limit is increased. Several techniques and methods applied in classification and estimation can be used (under suitable conditions) for predictions
B. Estimated price of rice in the next three months

4. Classification (Classification)

In classification, there are target categorical variables. For example, income classification can be grouped into three types, namely low income, medium income and high income

5. Clustering

Clustering is the process of observing, by collecting records or observing and creating classes of similar objects. Clusters are groups of records that correspond to each other and differ from records in other clusters. Clustering has no target variable in it so it is different from classification. Clustering does not attempt to estimate, classify, or estimate the value of the target variable. However, the clustering algorithm attempts to include all data into equal groups, where the similarity of records in one group will be high, while the similarity with records in other groups will be small.

Examples of clustering in business and research are:

1. Clustering the expression of genes, to obtain similar behavior from a large number of genes.
2. For accounting audits, namely to separate financial behavior in good or suspicious conditions.

6. Association (Association)

The way associations work in data mining is to look for attributes that appear at one time. In the business world it is more often referred to as shopping cart analysis.

Examples of associations in business and research are:

A. Finding products in the supermarket that are never bought together and those that are bought together.
B. Menarahakan customer groups by targeting the marketing of a product for companies that do not have large marketing funds.
C. Examine the number of cellular telecommunications company subscribers.
D. which is expected to provide a positive response to the service upgrade offers provided.

One technique known in data mining is clustering. The definition of scientific clustering in data mining is grouping several objects or data into clusters (groups) so that each group has data that is very suitable and different from data in other groups. Until now, scientists are still making various records to improve the cluster model and calculate how many clusters are optimal so that the best clusters can be produced. There are two clustering methods that we are familiar with, namely partition and hierarchical clustering. hierarchical clustering method itself consists of complete link clustering whereas the partitioning method itself consists of k-means and fuzzy k-means, average link grouping, single link grouping, and central link grouping.

C. K-Means Clustering

K-Means is an algorithm for grouping n objects based on attributes into k partitions, where k < n. The K-means method is the most common and simplest clustering method. This is because K-means is able to classify large amounts of data in a relatively fast and efficient time. The following figure shows the k-means clustering algorithm in action, for the two-dimensional case. Randomly generated initial centers to show more detailed stages. The partition space background is only for illustration and is not generated by the k-means algorithm[5].

The K-Means algorithm begins with a random determination of K, where K is the number of clusters you want to form. After that, assign K values randomly, for a while this value is the center of the cluster or commonly called the mean, centroid or "means". Find the shortest distance from each data to

the centroid. By using the Euclidian formula to Classify each data based on its proximity to the centroid. Do this process so that the centroid value is fixed or does not change (stable). K-Means is an iterative clustering algorithm.

Stages of the K-Means Clustering Method [6]

1. The first stage determines the number of clusters.
2. The second stage is the determination of the cluster center. In this study, the determination of the cluster center was carried out randomly.
3. To determine the object/data to be placed in the cluster, calculations are performed based on the distance between the two objects, as well as the distance between the object and the center of the cluster. To calculate the distance of all data to the cluster center point.

Use Euclidean theory:[7]

$$D(i,j) = \sqrt{(X1i - X1j)^2 + (X2i - X2j)^2 + .. + (Xki - Xkj)^2}$$

Where:

D(i,j)= Distance of data i to the center of Cluster j
xki = Data to i on attribute data to k
Xkj = jth center point on the kth attribute

4. Recalculate the cluster center with the current cluster membership. The cluster center is the average of all data/objects in a particular cluster.
5. If the cluster center has not changed anymore, then the clustering process stops, or returns to step 3 until the cluster center remains.

D. HIV/AIDS

The Human Immunodeficiency Virus (HIV) is a virus belonging to the Ribonucleic Acid (RNA) class that specifically undermines the human body's immune system and causes Acquired Immunodeficiency Syndrome (AIDS) [6]. They are a potential source of infection for other people. People who have been infected with HIV and their bodies have formed antibodies (anti-bodies) against the virus are HIV positive.

AIDS (Acute Immunodeficiency Syndrome/SIDA) is a combination of clinical symptoms due to a weakening of the body's immune system that arises as a result of HIV infection. AIDS often manifests with the emergence of various opportunistic infectious diseases, malignancies, metabolic disorders, and others [8]

E. Patient Profile

The patient profile is influenced by the Adherence factor (Adherence/ Therapy or Medication Taking). Adherence or adherence to therapy is a condition where the patient adheres to his medication based on his awareness, not just because he obeys the doctor's orders. This is important because it is hoped that it will further increase the level of medication adherence. Compliance or compliance must always be considered and evaluated regularly at every consultation. Failure of ARV therapy or taking medication is often caused by patient disobedience in taking drugs or ARVs. [9] Patient characteristics include sociodemographic factors (gender, age, race/ethnicity, education, income, literacy/illiteracy/, health insurance, and group origin in society, for example (commercial sex workers or transgender) and psychosocial factors (drug use, mental health). Narcotics, Alcohol, Psychotropics, and other Addictive Substances) environment and social support, knowledge and behavior towards HIV and its therapy).

III. Result and Discussion

A. Data Transformation

For data to be managed by applying the k-mean clustering method, data of nominal data type such as the city of origin (address/city of origin/city of origin), risk factors, gender, occupation, and first the data must be initialized in the form of numbers. Process initialize the address/city of origin/city of origin/city of origin, the steps are as follows:

1. In the city of origin data, the region is first divided into several regions.
2. Then these areas are sorted Based on the number of patients coming from the area seen from the highest frequency of origin from the area.
3. The number of origins from the highest patient area will be given a number 1. After that the area with the second highest frequency will be given a number 2, and so on until the area with the lowest frequency.Apart from the city of origin, major, risk factors, ethnicity/ethnicity, education, and occupation are also included in the nominal data type, so it needs to be initialized as a number. As in the city of origin, department, risk factors, ethnicity/ethnicity, education occuand passion, they are also given an initialization based on the frequency of the number of HIV/AIDS patients present.

B. Development Techniques

This research will be carried out with the following steps.

1. Literature study and guidance consultation

At this stage, research materials are collected through various sources of literature, either in the form of books, journals, proceedings, magazines, and so on as supporting materials, and also carry out consultations with the thesis supervisor.

2. Field data collectionAt this stage field observations are carried out to collect the required data and problems that are often encountered.
3. Data initialization
   At this stage, data identification is carried out to determine the validity of the data and the variables that will be used. If the data is not valid then field observations are carried out again.
4. Preparation of test datasets After obtaining valid data, testing methods are now being developed so that the research objectives are fulfilled.
5. Data mining application design
   At this stage, the application is designed using Matlab.
6. Implementation of tests using applications and evaluation of results.
7. This stage is to test the data using the application program and analyze the test results and evaluate errors.
8. Compile the final assignment book This final stage is the documentation of supporting theories, application system design, test results, and analysis, as well as suggestions and conclusions.

Below is data on 20 patients for testing the Modified K-Mean Clustering Algorithm by applying the calculation of the Sum of Squared Error (SSE) value to determine the center of a cluster

Table 1. Preliminary Data

No	Initial	Daerah	Pekerjaan	Umur
0	A	Perjuangan Medan	Tenaga Kerja Indonesia	33
1	B	Selayang Medan	Pegawai Swasta	25
2	C	Belawan Medan	Pedagang	44
3	D	Perjuangan Medan	Tenaga Kerja Indonesia	31
4	E	Perjuangan Medan	Tenaga Kerja Indonesia	37
5	F	Selayang Medan	Pegawai Swasta	40
6	G	Kota Medan	Pegawai Swasta	24
7	H	Belawan Medan	Pegawai Negeri Sipil	42
8	I	Kota Medan	Pegawai Negeri Sipil	38
9	J	Kota Medan	Therapy	28
10	K	Selayang Medan	Tenaga Kerja Indonesia	27
11	L	Selayang Medan	Tenaga Kerja Indonesia	22
12	M	Perjuangan Medan	Tenaga Kerja Indonesia	27
13	N	Perjuangan Medan	Theraphys	30
14	O	Kota Medan	Pegawai Swasta	44
15	P	Belawan Medan	Pedagang	41
16	Q	Belawan Medan	Pedagang	24
17	R	Belawan Medan	Tenaga Kerja Indonesia	22
18	S	Belawan Medan	Pedagang	27
19	T	Selayang Medan	Pegawai Swasta	26

A. Data Transformation

In this study, to process the above data using the K-Means Clustering method, nominal data such as occupation and region must first be initialized in the form of numbers.

To initialize the region, sort from the highest must be based on the frequency of patients coming from that region. After that, an initial with the number 1 will be given to the area with the highest frequency, and an initial with the number 2 will be given to the area with the second largest frequency. The following can be seen in the initialization results table for regional categories.Apart from the region, occupation is also included in the nominal data type, so it needs to be initialized as a number. As with the region, the job is also given an initialization based on the frequency of the patient's work. The results of the initialization of the work can be seen in Table 2

Table 2. Job Data Initialization

No	Pekerjaan	Frekuensi	Inisialisasi
1	Tenaga Kerja Indonesia	7	1

2	Pegawai Swasta	5	2
3	Pedagang	4	3
4	PNS	2	4
5	Therapy	2	5

The data can be grouped using the K-Mean Clustering method. After processing all patient data is transformed to numbers. The next process needs to be grouped data into several clusters, namely the following stages:

1. Determine in advance the number of clusters. In this study, the existing data will be grouped into three clusters.
2. Then in each cluster determine the starting point. In this study, the initial center point was generated randomly. The cluster center on the initial solution can be seen in Table..

B. Test Results

The distance of each patient's data to the new cluster center in the 1st iteration

a. Patient data 1

The first process will calculate the distance from the patient data to the cluster center

$$D(1,1) = \sqrt{(2-2)^2 + (1-1)^2 + (33-37)^2}$$
$$= \sqrt{(0)^2 + (0)^2 + (-4)^2}$$
$$= \sqrt{0+0+16}$$
$$= \sqrt{16}$$
$$= 4$$

The result is that the distance between patient 1's data and the center of the first cluster is 4. From the calculation results above. Then the distance from the first patient data to the second cluster center will be calculated as below..

$$D(1,2) = \sqrt{(2-2)^2 + (1-1)^2 + (33-33)^2}$$
$$= \sqrt{(0)^2 + (0)^2 + (0)^2}$$
$$= \sqrt{0+0+0}$$
$$= \sqrt{0}$$
$$= 0$$

The results of the above calculation results show that the distance from the center of the second cluster to patient 1 is 0. Then the distance from the center of the third cluster to patient 1's data will be calculated as below

$$D(1,3) = \sqrt{(2-3)^2 + (1-1)^2 + (33-27)^2}$$
$$= \sqrt{(-1)^2 + (0)^2 + (6)^2}$$
$$= \sqrt{1+0+36}$$
$$= \sqrt{37}$$
$$= 6.083$$

The distance between the first patient data and the third cluster center was obtained at 6.083. So it can be concluded that patient 1's data is grouped into cluster 3 because the minimum distance is 6,043, this means that patient 1's data will become a member of cluster 3 (C3).

Table 3. The Distance of Each Patient Data to the Centroid Point in the 1st Iteration

No	Initial	W	P	U	Jarak Ke			Min *Cluster*			
					C1	C2	C3	C1	C2	C3	
0	A	2	1	33	4.0	0.0	6.083		*		C2
1	B	3	2	25	12.083	8.124	2.236			*	C3
2	C	1	3	44	7.348	11.225	17.234	*			C1
3	D	2	1	31	6.0	2.0	4.123		*		C2
4	E	2	1	37	4.0	10.05	10.05	*			C1
5	F	3	2	40	3.317	7.141	13.038	*			C1
6	G	4	2	24	13.191	9.274	3.317			*	C3
7	H	1	4	42	5.916	9.539	15.427	*			C2
8	I	4	4	38	3.742	6.164	11.446	*			C2
9	J	4	5	28	10.05	6.708	4.234			*	C3
10	K	3	1	27	10.05	6.083	0.0			*	C3
11	L	3	1	22	15.033	11.045	5.0			*	C3
12	M	2	1	27	10.0	6.0	1.0			*	C3
13	N	2	5	30	8.062	5.0	5.099		*		C2
14	O	4	2	44	7.348	11.225	17.059	*			C1
15	P	1	3	41	4.583	8.307	14.283	*			C1
16	Q	1	3	24	13.191	9.274	4.123			*	C3
17	R	1	1	22	15.033	11.045	5.385			*	C3
18	S	1	3	27	10.247	6.403	2.828			*	C3

19	T	3	2	26	11.091	7.141	1.414			*	C3

Do the same process at the job of every patient.

The following will show a comparison of the number of patients in each cluster, where in cluster 1 the number of patients is 7 people, in cluster 2 there are 3 people and in cluster 3 there are 10 people.

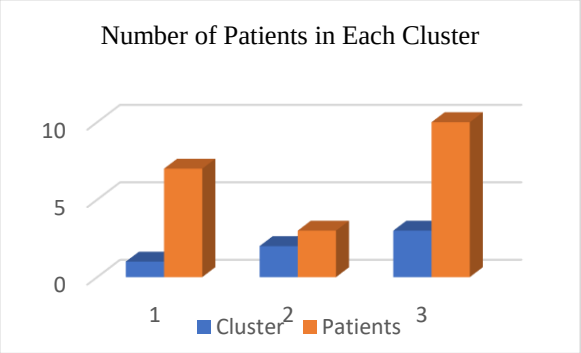


Fig 1. Number of Patients in Each Cluster

.

IV. CONCLUSION

Based on the test results and analysis of the results obtained, a conclusion can be drawn, namely:

1. From the results of testing and analysis we can see that the first cluster center point determines the number of iterations produced to obtain the final clustering.
2. In the K-Means Clustering Algorithm, data processing is very fast. This algorithm often experiences premature convergence so that its accuracy is not guaranteed. In this algorithm, the results look unsatisfactory, because it is not guaranteed that the distance between each centroid does not span so that if there are two or more groups with adjacent centroid points.
3. It cannot be concluded that the most optimum cluster center with the K-Mean Clustering method will produce the minimum iterations.
4. Obtained the number of patients in each cluster where for cluster 1 there were 7 patients and cluster 2 obtained as many as 3 clusters and cluster 3 obtained as many as 10 clusters

REFERENCES

[1] Yuli Mardi, "Data Mining : Klasifikasi Menggunakan Algoritma C4 . 5 Data mining merupakan bagian dari tahapan proses Knowledge Discovery in Database (KDD) . Jurnal Edik Informatika," *J. Edik Inform.*, vol. 2, no. 2, pp. 213–219, 2019.

[2] M. W. Talakua, Z. A. Leleury, and A. W. Talluta, "Analisis Cluster Dengan Menggunakan Metode Provinsi Maluku Berdasarkan Indikator Indeks Pembangunan Manusia Tahun 2014," *J. Ilmu Mat. dan Terap.*, vol. 11, no. 2, pp. 119–128, 2017.

[3] B. Harahap, "Penerapan Algoritma K-Means Untuk Menentukan Bahan Bangunan Laris (Studi Kasus Pada UD. Toko Bangunan YD Indarung)," *Reg. Dev. Ind. Heal. Sci. Technol. Art Life*, pp. 394–403, 2019, [Online]. Available: https://ptki.ac.id/jurnal/index.php/readystar/article/view/82

[4] W. M. P. Dhuhita, "Clustering Metode K-Means Untuk Menentukan Status Gizi Balita," *J. Inform.*, vol. 15, no. 2, pp. 160–174, 2015.

[5] N. Wakhidah, "Clustering Menggunakan K-Means Algorithm (K-Means Algorithm Clustering)," *Fak. Teknol. Inf.*, vol. 21, no. 1, pp. 70–80, 2014.

[6] A. Rauf, Sheeba, S. Mahfooz, S. Khusro, and H. Javed, "Enhanced K-mean clustering algorithm to reduce number of iterations and time complexity," *Middle East J. Sci. Res.*, vol. 12, no. 7, pp. 959–963, 2012, doi: 10.5829/idosi.mejsr.2012.12.7.1845.

[7] E. Z. Khulaidah and N. Irsalinda, "FCM using squared euclidean distance for e-commerce classification in Indonesia," *J. Phys. Conf. Ser.*, vol. 1613, no. 1, 2020, doi: 10.1088/1742-6596/1613/1/012071.

[8] H. Sur, "Hubungan Pengetahuan HIV / AIDS dengan Stigma terhadap Orang dengan HIV / AIDS di Kalangan Remaja 15-19 Tahun di Indonesia (Analisis Data SDKI Tahun 2012) Relationship HIV / AIDS Knowledge related Stigma towards People Living with HIV / AIDS among Adole," vol. 1, no. 2, pp. 35–43, 2017.

[9] Y. Marlinda and M. Azinar, "Jurnal of Health Education," vol. 2, no. 2, pp. 192–200, 2017.

ISSN 2355-0082

Intrusion Detection System on Nowaday's Attack using Ensemble Learning

Fajar Henri Erasmus Ndolu[1], Ruki Harwahyu[2]
[1,2] Dept. of Electrical Engineering, Faculty of Engineering, Universitas Indonesia, Depok, Indonesia
[1]fajar.henri@ui.ac.id
[2]ruki.h@ui.ac.id

Accepted 13 June 2023
Approved 14 July 2023

***Abstract*— Attacks on computer networks are becoming more and more widespread nowadays, making this an important issue that must be considered . These attacks can be detected with the Intrusion Detection System (IDS). However, at this time there are new attacks that have not been detected by IDS. Therefore, ensemble learning is used. This research we used Random Forest algorithm for attack detection as an increase in the ability of IDS to detect cyberattacks. The use of the CSE-CIC-IDS2018 dataset is used in this research as a current representative dataset for cyberattack detection. The results of this study we get a binary classification accuracy of 99.6856% and an f1-score of 99.5803% and a multiclass classification accuracy of 99.6944 and an f1-score of 97.8032% with a data ratio ratio dataset of 3:1 normal class to attack class.**

***Keywords*— IDS; random forest; undersampling; chi square; CSE-CIC-IDS2018.**

I. INTRODUCTION

The rapid development of technology makes cyberattacks more massive and more be attention to. According to Saxena et al and Morgan, financial losses are predicted to reach 10.5 trillion dollars by 2025 [1], [2] .

Detection of this cyberattack can be detected with a system developed called the Intrusion Detection System (IDS). However, IDS has not been able to accurately detect new attacks that have occurred and generates a high false alarm rate.

For this reason, in this research we used a dataset that is considered representative to reflect the current situation. The dataset used is CSE-CIC-IDS2018 [3] . The dataset used can be downloaded from Cloud Amazone Services (AWS) [4] with a total sample data of 16 million samples with 79 features with a benign class distribution of 83% with an attack class of 17% consisting of 14 attack classes.

There have been several studies on IDS such as the Support Vector Machine algorithm [5] carried out by Kotpaliwar et al, the k-Nearest Neighbor algorithm [6] , Gaussian Naïve Bayes [7] , various decision tree algorithms carried out by Hota et al [8] , Convolution Neural Network algorithm [9] . However, these studies still use an old dataset, that is the KDDCUP99 [10] dataset , which does not reflect the current state of the attack.

The use of the algorithm in this study is ensemble learning, because ensemble learning can be optimal for classes with unbalanced datasets. Ensemble learning is learning that combines several basic algorithms to get better predictive results based on the highest voting [11] . Ensemble learning is carried out by using the Random Forest algorithm which is a boosting approach from ensemble learning [12] . Random Forest is recognized as being quite good at overcoming class imbalances in datasets and providing fairly accurate results [13] .

Therefore, in this study we conducted research to detect attacks on computer networks using the Random Forest algorithm. This is expected to be able to detect attacks, especially today's types of attacks that cannot be detected with IDS.

II. METHOD

In this research, there are several research steps, as follows (Fig.1):

Fig. 1. Research procedure

A. *Data Exploration*

We used the CSE-CIC-IDS2018 dataset. The dataset consists of 10 CSV files and a total of 16 million samples with 83% benign class and 17% attack class.

B. *Data Preprocessing*

In this step, we do a number of things as shown in Fig. 2, including:

- Merging 10 files from dataset
- Remove duplicate header rows
- Convert timestamp to UnixTime
- The infinity value becomes NaN
- Remove features with a number of NaNs > 50%
- Delete row on feature number of NaN < 50%
- Remove any of the features that have a correlation coefficient equal to one

Balancing the normal class against the attack class by undersampling nearmiss-2 (ratio 1:1, 2:1, and 3:1)

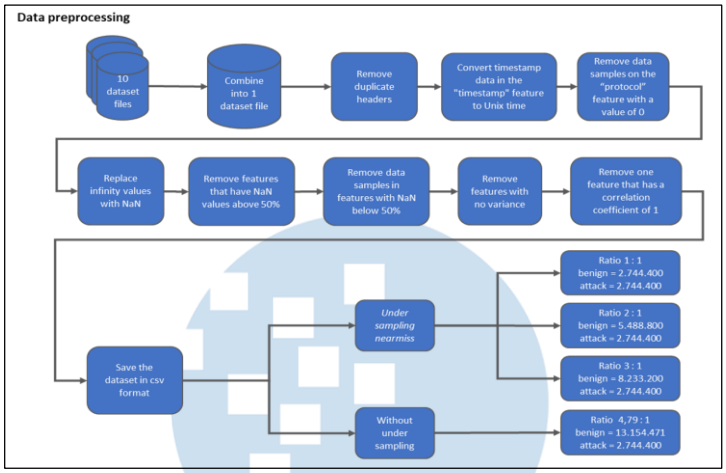


Fig. 2. Data Preprocessing

C. *Data sampling and normalization*

Data folding was carried out by fold out of 80% for data training and 20% of the data test on the balanced dataset and the original dataset without under sampling (ratios 1:1, 2:1, 3:1 and 4.79:1). Then normalize the data with a min-max scaler to re-scale the feature values to the value range [0,1].

D. *Feature selection*

Feature selection is carried out using the chi square method and binary or multiclass target vectors with a score percentage threshold of 99%, as shown in Fig. 3. so that there are 2 feature combinations in each dataset, a total of 8 feature combinations from the four datasets (ratio 1:1, 2:1, 3:1 and 4,79:1). Features with the remaining 1% score percentage will be deleted.

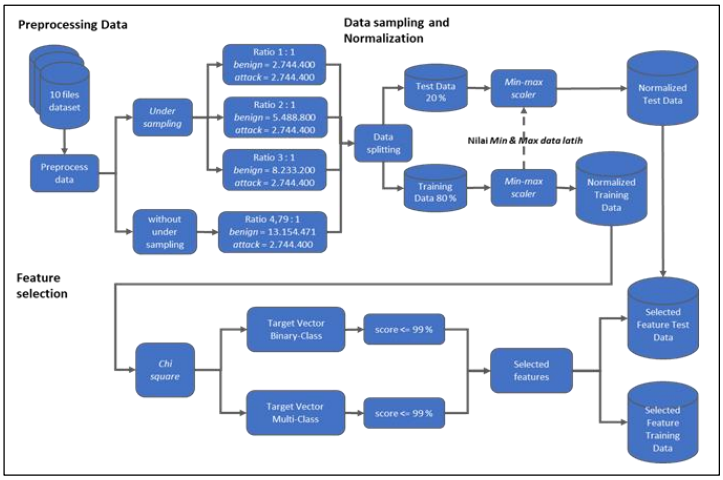


Fig. 3. Data sampling and feature selection

E. Hyperparameter tuning

In this research, to get optimal results, hyperparameter tuning was conducted in this research. Hyperparameter tuning is done using a random grid search technique. Random grid search randomly selects 15 predefined hyperparameter value combinations. Each combination of hyperparameter values is cross-validated by 5-fold cross validation. Then the combination of hyperparameter values is selected which produces the model with the highest average f1-score.

The choice of a combination of hyperparameter values is based on the highest f1-score value, because the dataset used is an unbalanced dataset so that a better measurement metric is the f1-score which is a harmonization between precision values and recall values. The hyperparameter values used in the tuning process include estimators, max features, max depth, min samples split, and min samples leaf, with the following hyperparameter value ranges, shown in Table 1:

TABLE I. RANDOM FOREST HYPERPARAMETERS

Hyperparameters	Value
Estimators	10,15,20,25,25,30,35,40,45,50
Max features	5,9,12,15,18
Max depth	None,5,10,15,20,25,30,35
Min samples split	2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17 ,18,19,20
Min samples leaf	2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17 ,18,19,20

F. Random Forest

At this stage, a model is built for each dataset (ratios 1:1, 2:1, 3:1 and 4.79:1) using the best hyperparameter values that have been obtained from the tuning stage. The Random Forest algorithm is classified binary and multi-class.

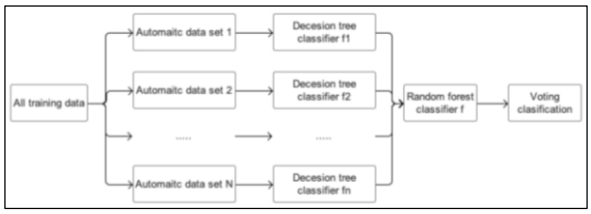


Fig. 4. Semantic diagram of the Random Forest algorithm [17]

Random Forest is used because it can prevent overfitting and is better at classifying minority classes in datasets. Its main advantage is that it can predict new data and cope with it class imbalance problem in the dataset [13], [14] . This is because the algorithm performs the learning process on a number of random decision trees that are generated from random subsamples and random feature subsets in the dataset. Thus, this algorithm can reduce the tendency to study irrelevant details and improve the generalization ability of new data [15] . In addition, the Random Forest algorithm is also resistant to measurement errors that occur during model development [16] . Therefore, the use of the Random Forest algorithm can help improve the accuracy and efficiency of the model. Fig. 7 shows the semantic diagram of the Random Forest algorithm.

G. Model evaluation

Evaluate the model, we are using the metrics of accuracy, precision, recall and f1 score. Accuracy is measured by calculating the percentage of normal and attack classes that are correctly predicted, or true positive and true negative, from the total dataset (see equation 1) [16] .

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (1)$$

To measure the classification of an attack as benign, recall is used, which refers to the number of correctly predicted true positives compared to the total actual positives in the dataset, i.e. true positives and false negatives (see equation 2) [16] .

$$recall = \frac{TP}{TP+FN} \times 100\% \quad (2)$$

Meanwhile, precision is used to calculate the number of true positives that are correctly predicted for all positive predictions, namely true positives and false positives (see equation 3) [16] .

$$Precision = \frac{TP}{TP+FP} \times 100\% \quad (3)$$

From measuring precision and gain, harmonization calculations are needed to overcome the trade-off between precision and gain. This measurement is called the f1-score (see equation 4) [16] .

$$F1 - score = 2 \times \frac{precision \times recall}{precision + recall} \times 100\% \quad (4)$$

III. RESULT AND DISCUSSION

In this research, model development and evaluation used the programming language Python 3.10, IDE Jupyter Lab 6.4.5, Pandas 1.3.4, NumPy 1.20.3, scikit-learn 1.0.2. The research procedure was carried out from pre-processing the dataset and ending with model evaluation.

A. Preprocessing data and sampling data

In the pre-processing stage, the first 10 datasets were merged into one dataset, with around 83% normal traffic data and 17% for attack datasets. Table 1 shows the class distribution of the dataset for this study.

TABLE II. DISTRIBUTION OF NORMAL AND ATTACKS CLASS

Traffic	Distribution(%)	Number of samples
Benign	83.070014	13,484,708
DDoS Attack HOIC	4.226048	686,012
DDoS Attacks LOIC HTTP	3.549517	576,191
Hulk's DoS attacks	2.845522	461,912
Bots	1.763026	286,191
Brute force FTP	1.191158	193,360
SSH Brute force	1.155607	187,589
Infiltration	0.997564	161,934
DoS attacks SlowHTTPTest	0.861766	139,890
DoS attacks GoldenEye	0.255702	41,508
Slowloris DoS attacks	0.067702	10,990
UDP LOIC DDoS attacks	0.010657	1730
Web brute force	0.003764	611
Brute Force XSS	0.001417	230
SQL Injections	0.000536	87
Total	**100**	**16,232,943**

Furthermore, data duplication was removed for 59 header rows with the same name. To make it easier to access the features in the dataset, the feature names in the dataset are changed to lowercase and change the symbol characters and spaces (white space) to underscores. Then the timestamp feature is converted to Unix time and the timestamp data type is converted from object to numeric (int64).

In this research, the protocol used is the TCP protocol with a value of 6 and the UDP protocol with a value of 17, apart from the removed TCP and UDP protocols. In the dataset there is a protocol value of "0". Therefore, samples on protocol features that have a value of "0" are removed so as not to cause bias in the built model.

Then delete samples and features based on the number of NaN. The four additional features in the fourth file " *Thursday-20-02-2018_TrafficForML_CICFlowMeter.csv* " namely flow_id, src_ip, src_port, and dst_ip are missing in the other nine files, resulting in a NaN when combined. The resulting number of NaNs reached 8,190,014 or 51.2% of the total sample in the dataset, so these features were deleted. Whereas in the flow_byts_s and flow_pkts_s features which have a total NaN of 95,759 or 0.59% of the total sample dataset, samples are deleted.

Feature deletion also applies to features that do not have variants, because these features do not contribute to the classification of the target class. Removed features include bwd_psh_flags, bwd_urg_flags, fwd_byts_b_avg, fwd_pkts_b_avg, fwd_blk_rate_avg, bwd_byts_b_avg, bwd_pkts_b_avg, and bwd_blk_rate_avg. In addition, feature deletion is also carried out on one of the two features that have the same value distribution. If the value of the correlation coefficient is equal to one, then one of the features is removed. There are 8 features removed, namely cwe_flag_count, subflow_fwd_byts, syn_flag_cnt, subflow_bwd_pkts, bwd_seg_size_avg, fwd_seg_size_avg, subflow_bwd_byts.

After several preprocessing stages, the remaining dataset has 63 features and 15.898.871 samples from the initial dataset which has 83 features and 16.232.943 samples. There was a deletion of 20 features and 334.072 samples.

Then, from the remaining datasets, class balancing was carried out. This was done with three different sampling ratios, namely 1:1, 2:1, and 3:1. The model development was also carried out with a dataset without under sampling with a normal class to attack class ratio of 4.79:1. The amount of data in each dataset can be seen in Table 3.

TABLE III. NORMAL CLASS TO ATTACK CLASS RATIO COMPARISON

Under sampling	Ratio	Number of Samples	
		normal	attacks
Nearmiss-2	1 : 1	2,744,000	2,744,400
Nearmiss-2	2 : 1	5,488,800	2,744,400
Nearmiss-2	3 : 1	8,233,200	2,744,400
None	4, 79 : 1	13,154,471	2,744,400

After the data pre-processing stage, then divide the data into 80% training data and 20% test data. Then normalize the data with the MinMax scaler to the value range [0,1] [13] .

B. *Feature selection*

Feature selection was performed using the Chi-square method with binary and multi-class target vectors. The calculation results are then sorted from the largest and summed up until the total score forms a percentage of ≤ 99%. Features with the remaining 1% percentage removed. This feature selection was carried out on 4 different datasets with ratios of 1:1, 2:1, 3:1 and 4.79:1 and each dataset produced 2 different selected feature combinations, so that the total in the four datasets produced 8 selected feature combinations as listed in Table 4.

C. Hyperparameter tuning

This research uses a random grid search technique with cross-validation (k = 5) to select a combination of hyperparameter values that produce the optimal model. From a total of 8 different selected feature combinations, a total of 8 combinations of hyperparameter values were generated from the tuning process. Then, a combination of hyperparameter values with the highest F1 is selected from each dataset, as shown in Table 5.

TABLE IV. SELECTED 8 FEATURES COMBINATION

Ratio	Under sampling	Feature selection - Target vector - Percentage (%) - features	Selected feature combinations
1 : 1	Nearmiss-2	Chi-square - Multi-class - 99 - 50	'dst_port', 'protocol', 'timestamp', 'flow_duration', 'tot_fwd_pkts', 'totlen_fwd_pkts', 'fwd_pkt_len_max', 'fwd_pkt_len_mean', 'fwd_pkt_len_std', 'bwd_pkt_len_min', 'bwd_pkt_len_ mean', 'bwd_pkt_len_std', 'flow_pkts_s ', 'flow_iat_mean', 'flow_iat_std', 'flow_iat_max', 'flow_iat_min', 'fwd_iat_tot', 'fwd_iat_mean', 'fwd_iat_max', 'fwd_iat_min', 'bwd_iat_tot', 'bwd_iat_mean', 'bwd_iat_std', 'bwd _iat_max', 'bwd_iat_min', 'fwd_psh_flags', 'fwd_header_len', 'fwd_pkts_s', 'bwd_pkts_s', 'pkt_len_mean', 'pkt_len_std', 'pkt_len_var', 'rst_flag_cnt', 'psh_flag_cnt', 'ack_flag_cnt', 'urg_flag_cnt', 'ece_flag_cnt ', 'pkt_size_avg', 'init_fwd_win_byts', 'init_bwd_win_byts', 'fwd_act_data_pkts', 'fwd_seg_size_min', 'active_mean', 'active_std', 'active_max', 'idle_mean', 'idle_std', 'idle_max', 'idle_min'
	Nearmiss-2	Chi-square - Binary class – 99 - 31	'dst_port', 'fwd_pkt_len_max', 'fwd_pkt_len_mean', 'fwd_pkt_len_std', 'bwd_pkt_len_max', 'bwd_pkt_len_mean', 'bwd_pkt_len_std', 'flow_pkts_s', 'flow_iat_mean', 'flow_iat_min', 'fwd_iat_mean', 'fwd_iat_std', 'fwd_iat_min ', 'fwd_pkts_s', 'bwd_pkts_s', 'pkt_len_max', 'pkt_len_mean', 'pkt_len_std', 'pkt_len_var', 'rst_flag_cnt', 'psh_flag_cnt', 'ack_flag_cnt', 'urg_flag_cnt', 'ece_flag_cnt' , 'pkt_size_avg', 'init_fwd_win_byts', 'init_bwd_win_byts', 'fwd_seg_size_min', 'idle_mean', 'idle_max', 'idle_min'
2 : 1	Nearmiss-2	Chi-square - Multi-class - 99 - 47	'dst_port', 'protocol', 'flow_duration', 'tot_fwd_pkts', 'totlen_fwd_pkts', 'fwd_pkt_len_max', 'fwd_pkt_len_mean', 'fwd_pkt_len_std', 'bwd_pkt_len_mean', 'bwd_pkt_len_std', ' flow_pkts_s', 'flow_iat_mean', 'flow_iat_std ', 'flow_iat_max', 'flow_iat_min', 'fwd_iat_tot', 'fwd_iat_mean', 'fwd_iat_max', 'fwd_iat_min', 'bwd_iat_tot', 'bwd_iat_mean', 'bwd_iat_std', 'bwd_iat_max', 'bwd_iat_min', 'fw d_psh_flags', 'fwd_header_len', 'fwd_pkts_s', 'bwd_pkts_s', 'pkt_len_mean', 'pkt_len_std', 'rst_flag_cnt', 'psh_flag_cnt', 'ack_flag_cnt', 'urg_flag_cnt', 'ece_flag_cnt', 'pkt_size_avg' , 'init_fwd_win_byts', 'init_bwd_win_byts ', 'fwd_act_data_pkts', 'fwd_seg_size_min', 'active_mean', 'active_std', 'active_max', 'idle_mean', 'idle_std', 'idle_max', 'idle_min'
	Nearmiss-2	Chi-square - Binary class - 99 - 37	'dst_port', 'protocol', 'flow_duration', 'fwd_pkt_len_max', 'fwd_pkt_len_min', 'fwd_pkt_len_mean', 'fwd_pkt_len_std', 'bwd_pkt_len_max', 'bwd_pkt_len_min', 'bwd_pkt_len_mean', 'bwd_ pkt_len_std', 'flow_pkts_s', 'flow_iat_mean ', 'flow_iat_max', 'flow_iat_min', 'fwd_iat_tot', 'fwd_iat_mean', 'fwd_iat_max', 'fwd_iat_min', 'fwd_psh_flags', 'fwd_pkts_s', 'bwd_pkts_s', 'pkt_len_min', 'pkt_len_max' , 'pkt_len_mean', 'pkt_len_std', 'pkt_len_var', 'psh_flag_cnt', 'ack_flag_cnt', 'urg_flag_cnt', 'pkt_size_avg', 'init_fwd_win_byts', 'init_bwd_win_byts', 'fwd_seg_size_min', 'idle_mean', 'idle_max', ' idle_min'
3 : 1	Nearmiss-2	Chi-square - Multi-class - 99 - 46	'dst_port', 'protocol', 'timestamp', 'flow_duration', 'tot_fwd_pkts', 'totlen_fwd_pkts', 'fwd_pkt_len_max', 'fwd_pkt_len_mean', 'fwd_pkt_len_std', 'bwd_pkt_len_std', 'flow_pkts_s ', 'flow_iat_mean', 'flow_iat_std ', 'flow_iat_max', 'flow_iat_min', 'fwd_iat_tot', 'fwd_iat_mean', 'fwd_iat_max', 'fwd_iat_min', 'bwd_iat_tot', 'bwd_iat_mean', 'bwd_iat_std', 'bwd_iat_max', 'bwd_iat_min', 'fw d_psh_flags', 'fwd_header_len', 'fwd_pkts_s', 'bwd_pkts_s', 'pkt_len_std', 'rst_flag_cnt', 'psh_flag_cnt', 'ack_flag_cnt', 'urg_flag_cnt', 'ece_flag_cnt', 'init_fwd_win_byts', 'init_bwd _win_byts', 'fwd_act_data_pkts', 'fwd_seg_size_min ', 'active_mean', 'active_std', 'active_max', 'active_min', 'idle_mean', 'idle_std', 'idle_max', 'idle_min'
	Nearmiss-2	Chi-square - Binary class - 99 - 36	'dst_port', 'protocol', 'timestamp', 'flow_duration', 'fwd_pkt_len_max', 'fwd_pkt_len_min', 'fwd_pkt_len_mean', 'fwd_pkt_len_std', 'bwd_pkt_len_min', 'bwd_pkt_len_mean', 'flow_pkts_s', 'flow_iat_mean', 'flow_iat_max ', 'flow_iat_min', 'fwd_iat_tot', 'fwd_iat_mean', 'fwd_iat_max', 'fwd_iat_min', 'fwd_psh_flags', 'fwd_pkts_s', 'bwd_pkts_s', 'pkt_len_min', 'pkt_len_max', 'pkt_len_me an', 'pkt_len_std', 'fin_flag_cnt', 'rst_flag_cnt', 'ack_flag_cnt', 'ece_flag_cnt', 'pkt_size_avg', 'init_fwd_win_byts', 'init_bwd_win_byts', 'fwd_seg_size_min', 'idle_mean', 'idle_max', 'idle_min'
4, 79 : 1	None	Chi-square – Multi-class - 99 - 37	'dst_port', 'protocol', 'timestamp', 'flow_duration', 'tot_fwd_pkts', 'totlen_fwd_pkts', 'bwd_pkt_len_min', 'flow_pkts_s', 'flow_iat_mean', 'flow_iat_std', 'flow_iat_max', 'flow_iat_min', 'fwd_ iat_tot ', 'fwd_iat_mean', 'fwd_iat_max', 'fwd_iat_min', 'bwd_iat_tot', 'bwd_iat_mean', 'bwd_iat_max', 'bwd_iat_min', 'fwd_psh_flags', 'fwd_header_len', 'fwd_pkts_s', 'bwd_pkts _s', 'rst_flag_cnt', 'psh_flag_cnt', 'ack_flag_cnt', 'urg_flag_cnt', 'ece_flag_cnt', 'init_fwd_win_byts', 'init_bwd_win_byts', 'fwd_act_data_pkts', 'fwd_seg_size_min', 'idle_mean', 'idle_std', 'idle_max' , 'idle_min'

Ratio	Under sampling	Feature selection - Target vector - Percentage (%) - features	Selected feature combinations
	None	Chi-square – Binary class - 99 - 35	'dst_port', 'protocol', 'timestamp', 'flow_duration', 'fwd_pkt_len_max', 'fwd_pkt_len_min', 'bwd_pkt_len_min', 'flow_pkts_s', 'flow_iat_std', 'flow_iat_max', 'fwd_iat_tot', 'fwd_iat_mean', ' fwd_iat_std ', 'fwd_iat_max', 'bwd_iat_tot', 'bwd_iat_mean', 'bwd_iat_std', 'bwd_iat_max', 'fwd_psh_flags', 'fwd_pkts_s', 'bwd_pkts_s', 'pkt_len_min', 'pkt_len_mean', 'fin_flag_cnt ', 'rst_flag_cnt', 'ack_flag_cnt', 'ece_flag_cnt', 'pkt_size_avg', 'init_fwd_win_byts', 'init_bwd_win_byts', 'fwd_act_data_pkts', 'fwd_seg_size_min', 'idle_mean', 'idle_max', 'idle_min'

D. Model development and model evaluation

Model development with the Random Forest (RF) algorithm is carried out on each different dataset with a ratio of 1:1, 2:1, 3:1, and 4.79:1. There are 4 models built. The four models were evaluated by binary and multi-class classification which were then compared.

The best model is obtained from a dataset with a ratio of 3:1 . Evaluation of binary classification produces an average accuracy of 99.6856%, precision of 99.6414%, recall of 99.5196%, and f1 of 99.5803%. While the results of the multi-class classification evaluation show that the model has an accuracy of 99.6944%, precession of 98.8319%, recall of 96.904%, and F1 of 97.8032% as shown in Table 6.

TABLE V. BEST COMBINATION OF HYPERPARAMETERS FROM HYPERPARAMETER TUNING

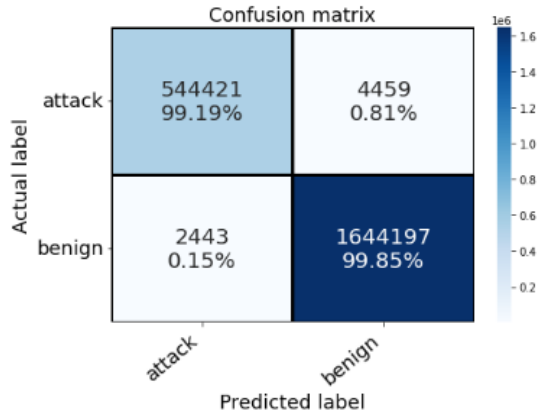
Ratio	Under Sampling	Feature selection - Target vector - Percentage of score (%) - features	Best Hyperparameters
1 : 1	Nearmiss-2	Chi-square - Multi class - 99% - 50	n estimators = 20, min samples split = 18, min samples leaf = 2, max features = 15, max depth = None
2 : 1	Nearmiss-2	Chi-square - Binary class - 99% - 37	n estimators = 20, min samples split = 2, min samples leaf = 6, max features = 18, max depth = 35
3 : 1	Nearmiss-2	Chi-square - Multi class - 99% - 46	n estimators = 15, min samples split = 17, min samples leaf = 2, max features = 15, max depth = 30
4, 79 : 1	None	Chi-square - Binary class - 99% - 35	n estimators = 35, min samples split = 3, min samples leaf = 2, max features = 15, max depth = 35

TABLE VI. EVALUATION RESULTS

Ratio	feature selection (features)	Multiclass classification (%)				Time (sec)	
		accuracy	Precision	recall	F1-score	Trains	test
1:1	Chi-square (50)	99.7564	98.1668	96.0803	96.9755	215.62	1.41
2:1	Chi-square (37)	98.2595	93.5775	89.9453	91.2306	564.62	3.22
3:1	Chi-square (46)	99.6944	98.8319	96,904	97.8032	333.13	2.29
4.79:1	Chi-square (35)	99.2835	96521	95.0857	95.7519	1166.25	7.43

Fig. 5 shows the binary classification confusion matrix . The confusion matrix shows a false alarm rate of 0.15 % (2443) and a false negative rate of 0.8 1 % (4 459).

Confusion matrix :
==========



Misclassification:
==========

	Misclassification	total	Percentage
attack	4459	548880	0.812382
benign	2443	1646640	0.148363

Fig. 5. Binary classification confusion matrix

From the results of the multi-class evaluation shown in Fig 6 , this model has an average accuracy value of above 99%, even 7 out of 15 classes have an F1 value of 100%. This shows that the hybrid model can classify these classes accurately.

Accuracy, Precision, Recall dan F1-score :
==========
Accuracy : 0.9969437764174318
Precision : 0.9883185413766322
Recall : 0.9690401872991661
F1-score : 0.9780318699424325

Evaluation metrics :
==========

	precision	recall	f1-score	support
benign	0.997276	0.998655	0.997965	1646640
bot	1.000000	1.000000	1.000000	57187
brute_force__web	0.974138	0.957627	0.965812	118
brute_force__xss	1.000000	0.956522	0.977778	46
ddos_attack_hoic	1.000000	1.000000	1.000000	137202
ddos_attack_loic_udp	1.000000	1.000000	1.000000	346
ddos_attacks_loic_http	0.999965	0.999965	0.999965	115238
dos_attacks_goldeneye	1.000000	1.000000	1.000000	8302
dos_attacks_hulk	1.000000	1.000000	1.000000	92382
dos_attacks_slowhttptest	1.000000	1.000000	1.000000	27978
dos_attacks_slowloris	1.000000	0.999545	0.999772	2198
ftp_bruteforce	0.999974	1.000000	0.999987	38671
infilteration	0.924853	0.858636	0.890515	31677
sql_injection	0.928571	0.764706	0.838710	17
ssh_bruteforce	1.000000	0.999947	0.999973	37518
accuracy			0.996944	2195520
macro avg	0.988319	0.969040	0.978032	2195520
weighted avg	0.996869	0.996944	0.996888	2195520

Fig . 6. Multi-class classification evaluation results

Misclassification:
==========

	Misclassification	total	Percentage
sql_injection	4	17	23.529412
infilteration	4478	31677	14.136440
brute_force__xss	2	46	4.347826
brute_force__web	5	118	4.237288
benign	2214	1646640	0.134456
dos_attacks_slowloris	1	2198	0.045496
ssh_bruteforce	2	37518	0.005331
ddos_attacks_loic_http	4	115238	0.003471
bot	0	57187	0.000000
ddos_attack_hoic	0	137202	0.000000
ddos_attack_loic_udp	0	346	0.000000
dos_attacks_goldeneye	0	8302	0.000000
dos_attacks_hulk	0	92382	0.000000
dos_attacks_slowhttptest	0	27978	0.000000
ftp_bruteforce	0	38671	0.000000

Fig. 7. Multi-class misclassification dataset ratio of 3:1

Based on the misclassification analysis in Fig.7 , SQL Injection is the type of attack with the lowest performance and the highest percentage of misclassification of 23.52%. This is due to the small size of the SQL Injection sample which only amounts to 87 samples or around 0.00054% of the entire dataset. This small sample size is not sufficient to represent the class data in this study, making it difficult to achieve an F1 score above 90%.

Furthermore, infiltration is a type of attack with the second largest misclassification, namely 14, 13 %. From the confusion matrix in Figure 8, it can be seen that 4478 Infiltration samples were incorrectly classified as benign class (normal class), and similarly, 2210 out of 2214 benign samples were incorrectly classified as Infiltration class. This indicates that several Infiltration classes and Benign classes have similar patterns, making it difficult for the model to distinguish between them.

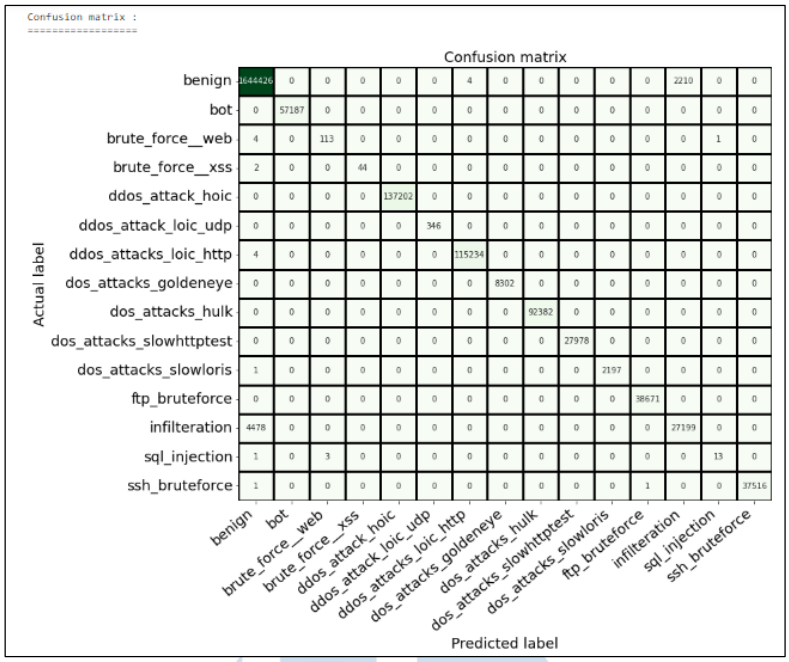


Fig . 8 integration matrix dataset ratio 3:1

IV. CONCLUSION

Based on the tests and analysis performed, it can be concluded that the random foreset algorithm with hyperparameter tuning provides increased performance in terms of accuracy, precision, recall, and F1, and has the potential to reduce false detection . This model gives the best results with a ratio of 3:1 (normal versus attack). The results of the binary classification evaluation show an average accuracy of 99.6856%, precision of 99.6414%, recall of 99.5196%, and f1 of 99.5803%. Meanwhile, the results of the multi-class classification evaluation showed an accuracy of 99.6944%, precession of 98.8319%, recall of 96.904%, and F1 of 97.8032%. Overall, the model has been successfully developed and tested, demonstrating improved performance on the 2018 CSE-CICIDS data set.

ACKNOWLEDGMENTS

Authors would like to thanks to Dept. Electrical Engineering Faculty of Engineering Universitas Indonesia, for supporting this research.

REFERENCES

[1] AK Saxena, S. Sinha, and P. Shukla, "General study of intrusion detection systems and surveys of agent based intrusion detection systems," *Proceeding - IEEE Int. Conf. Comput. commun. Autom. ICCCA 2017* , vol. 2017-January, pp. 417–421, 2017, doi: 10.1109/CCAA.2017.8229866.

[2] S. Morgan, "Cybercrime To Cost The World 8 Trillion Annually In 2023." https://cybersecurityventures.com/cybercrime-to-cost-the-world-8-trillion-annually-in-2023/ (accessed Feb. 03, 2023).

[3] "IDS 2018 | Datasets | Research | Canadian Institute for Cybersecurity | UNB." https://www.unb.ca/cic/datasets/ids-2018.html (accessed Feb. 03, 2022).

[4] "A Realistic Cyber Defense Dataset (CSE-CIC-IDS2018) - Registry of Open Data on AWS." https://registry.opendata.aws/cse-cic-ids2018/ (accessed Feb. 03, 2023).

[5] MV Kotpalliwar and R. Wajgi, "Classification of attacks using support vector machine (SVM) on KDDCUP'99 IDS database," in Proceedings - *2015 5th International Conference on Communication Systems and Network Technologies, CSNT 2015* , 2015, pp. 987–990, doi: 10.1109/CSNT.2015.185.

[6] Y. Liao and VR Vemuri, "Use of K-Nearest Neighbor classifier for intrusion detection," *Comput. Secur.* , vol. 21, no. 5, pp. 439–448, 2002, doi: 10.1016/S0167-4048(02)00514-X.

[7] BS Sharmila and R. Nagapadma, "Intrusion Detection System using Naive Bayes algorithm," in *2019 IEEE International WIE Conference on Electrical and Computer Engineering (WIECON-ECE)* , 2019, pp. 1–4, doi: 10.1109/WIECON-ECE48653.2019.9019921.

[8] HS Hota and AK Shrivas, "Decision Tree Techniques Applied on NSL-KDD Data and Its Comparison with Various Feature Selection Techniques," in *Advanced Computing, Networking and Informatics- Volume 1* , 2014, pp. 205–211, doi: 10.1007/978-3-319-07353-8_24.

[9] Y. Xiao, C. Xing, T. Zhang, and Z. Zhao, "An Intrusion Detection Model Based on Feature Reduction and Convolutional Neural Networks," *IEEE Access* , vol. 7, pp. 42210–42219, 2019, doi: 10.1109/ACCESS.2019.2904620.

[10] "KDD Cup 1999 Data." http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html.

[11] X. Gao, C. Shan, C. Hu, Z. Niu, and Z. Liu, "An Adaptive Ensemble Machine Learning Model for Intrusion Detection," *IEEE Access* , vol. 7, pp. 82512–82521, 2019, doi: 10.1109/ACCESS.2019.2923640.

[12] A. Saifudin and RS Wahono, "Application of Ensemble Techniques to Handle Class Imbalance in Prediction of

Software Defects," *J. Softw. Eng.* , vol. 1, no. 1, 2015.

[13] AS More and DP Rana, "Review of random forest classification techniques to resolve data imbalances," in *2017 1st International Conference on Intelligent Systems and Information Management (ICISIM)* , 2017, pp. 72–78, doi: 10.1109/ICISIM.2017.8122151.

[14] L. Breiman, "Random Forests," *Mach. Learn.* , vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.

[15] TB Laboratories, M. Avenue, and M. Hill, "Random Decision Forests Tin Kam Ho Perceptron training," 1995.

[16] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques* , 3rd ed. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2011.

[17] L. Yang, M. Cai, Y. Duan, and X. Yang, "Intrusion detection based on approximate information entropy for random forest classification," *ACM Int. Conf. Proceeding Ser.* , pp. 125–129, 2019, doi: 10.1145/3335484.3335488.

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.

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: **ultimaijnmt@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 IJNMT 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 IJNMT. 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

IJNMT_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		

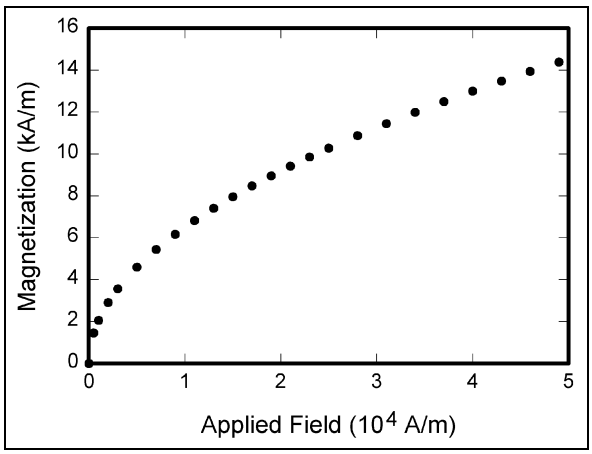


Fig. 1. Example of a figure caption

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 2355-0082

UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA
PRESS

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