

E-ISSN : 2614-8404
P-ISSN : 2776-3234



JISA

(JURNAL INFORMATIKA
dan SAINS)



VOL 6 No 2
December

Managed by :

Program Studi Teknik Informatika
UNIVERSITAS TRILOGI

2023

JISA

(Jurnal Informatika dan Sains)

Volume 6, Edition 2, December 2023

Interactive Multimedia as Information Media Parasitic Infection with Multimedia Development Life Cycle Method

Dasril Aldo, Miftahul Ilmi, Sapta Eka Putra, Widya Lelisa Army

Penetration Testing Tangerang City Web Application With Implementing OWASP Top 10 Web Security Risks Framework

Yoel Armando, Rosalina Rosalina

Sentiment Analysis of 2024 Presidential Candidates Election Using SVM Algorithm

Michael Alfonso, Dionisia Bhisetya Rarasati

DevOps, Continuous Integration and Continuous Deployment Methods for Software Deployment Automation

Mochamad Hanif Rifa'i Istifarulah, Rizka Tiaharyadini

Bayah Dome Geopark Information System Based On Website Using Prototype Method

Dentik Karyaningsih, Donny Fernando, Reza Pramudita, Fauzan Wahyu Ramadhan

DEVELOPMENT OF ARTS LEARNING MEDIA USING AUGMENTED REALITY WITH MARKERLESS BASED TRACKING METHOD AT SMKN 47

Ghali Hanan, Budi Arifitama, Silvester Dian Handy Permana

Implementation of The Resident's Dues Applications (SIUMAS) Using Waterfall Method in RT X Cinere Village

Febrianto Febrianto, Ikhwana Fakhir Nur Aziz, Mochammad Alvian Kosim, Muhammad Darwis, Retno Hendrowati

Implementation of Android-Based Tailoring Service Ordering Application with Geolocation Integration

Muhamad Agung Santoso, Joko Sutopo

Forecasting Blood Demand Using the Support Vektor Regression Method (Case Study: Blood Transfusion Unit-PMI Central Lombok)

Yati Apriati, Wafiah Murniati, Saikin Saikin, Sofiansyah Fadli, Hairul Fahmi

Improving Firebase BaaS Service Security in Counseling Chat Applications: AES-256 and CBC Approach for End-to-End Encryption

Mogar Nurhandhi, Agus Suhendar

Comparison of ANN Backpropagation Algorithm and Random Forest Regression in Predicting the Number of New Students

Padmavati Darma Tanuwijaya, Jhonatan Laurensius Tjahjadi, Yosefina Finsensia Riti

Mobile Based Application Design of the Catholic Church of the Redeemer of Sumbawa Using the Waterfall Method

Claudio Orlando De Araujo, Muhammad Zakariyah

Detection of Hate-Speech Tweets Based on Deep Learning: A Review

Ara Zozan Miran, Adnan Mohsin Abdulazeez

Managed by:
Program Studi Teknik Informatika
Universitas Trilogi

JISA	Vol : 6	Ed :2	Page : 096-188	Jakarta, Dec 2023	e-ISSN: 2614-8404	p-ISSN: 2776-3234
-------------	----------------	--------------	---------------------------------	------------------------------------	------------------------------------	------------------------------------

JISA

(Jurnal Informatika dan Sains)

Volume 6, Edition 2, December 2023

Advisor

Yodfiatfinda.,P.hD

Editor in Chief

Assoc.Prof.Budi Arifitama, S.T., MMSI

Editorial Board

Ade Syahputra, S.T., M.Inf.Comm.Tech.Mgmt.	: Universitas Trilogi
Ir.Yaddarabullah, M.Kom.IPM., ASEAN.Eng	: Universitas Trilogi
Maya Cendana, S.T., M.Cs.	: Universitas Bunda Mulia
Assoc.Prof.Silvester Dian Handy Permana, S.T., M.T.I.	: Universitas Trilogi
Ketut Bayu Yogha. B, S.Kom., M.Cs	: Universitas Trilogi
Ninuk Wiliani.,S.Si.,M.Kom	: Institut Teknologi dan Bisnis BRI
Dwi Pebrianti,Ph.D	: Universiti Malaysia Pahang, Malaysia
Dr.Wahyu Caesarendra	: Universiti Brunei Darussalam
Erneza Dewi Krishnasari.,S.Ds.,M.Ds	: Universitas Trilogi
Muhammad Lahandi Baskoro	: Conventry University, UK

Reviewers

Azida Zainol.,Ph.D	: Sheffield Hallam University, UK
Prof. Ir. Suyoto, M.Sc. Ph.D	: Universitas Atma Jaya Yogyakarta
Dr. Ir. Albertus Joko Santoso, M.T.	: Universitas Atma Jaya Yogyakarta
Setiawan Assegaff, ST, MMSI, Ph.D	: STIKOM Dinamika Bangsa, Jambi
Michael Marchenko, Ph.D	: Universitas Trilogi, Jakarta
Prof.Dr.Hoga Saragih.,ST.,MT	: Universitas Bakrie
Isham Shah Hassan.,Ph.D	: Port Dickson Polytechnic Malaysia
Prof.Dr Abdul Talib Bon	: Universiti Tun Hussein Onn, Malaysia
Wiwin Armoldo Oktaviani, S.T, M.Sc	: Universitas Muhammadiyah Palembang
Yosi Apriani, S.T, M.T	: Universitas Muhammadiyah Palembang
Dr. Gundung Triyono.,M.Kom	: Universitas Budi Luhur
Ir. Lukito Edi Nugroho, M.Sc., Ph.D	: Universitas Gadjah Mada
Dr. Soetam Rizky Wicaksono	: Universitas Ma chung,

Secretariat

Asih Wulandini

Editorial Address

Ruang Dosen Fakultas Sains Teknik dan Desain Lantai 3
Jalan Taman Makam Pahlawan No. 1, Kalibata, Pancoran, RT.4/RW.4, Duren Tiga, Pancoran, Kota
Jakarta Selatan, Daerah Khusus Ibukota Jakarta 12760Telp :(021) 798001

Managed by:
Program Studi Teknik Informatika
Universitas Trilogi

JISA	Vol : 6	Ed :2	Page : 096-188	Jakarta, Dec 2023	e-ISSN: 2614-8404	p-ISSN: 2776-3234
-------------	----------------	--------------	---------------------------------	------------------------------------	------------------------------------	------------------------------------

Table of Content

Interactive Multimedia as Information Media Parasitic Infection with Multimedia Development Life Cycle Method	96-104
<i>Dasril Aldo, Miftahul Ilmi, Sapta Eka Putra, Widya Lelisa Army</i>	
Penetration Testing Tangerang City Web Application With Implementing OWASP Top 10 Web Security Risks Framework	105-109
<i>Yoel Armando, Rosalina Rosalina</i>	
Sentiment Analysis of 2024 Presidential Candidates Election Using SVM Algorithm	110-115
<i>Michael Alfonso, Dionisia Bhisetya Rarasati</i>	
DevOps, Continuous Integration and Continuous Deployment Methods for Software Deployment Automation	116-123
<i>Mochamad Hanif Rifa'i Istifarulah, Rizka Tiaharyadini</i>	
Bayah Dome Geopark Information System Based On Website Using Prototype Method	124-129
<i>Dentik Karyaningsih, Donny Fernando, Reza Pramudita, Fauzan Wahyu Ramadhan</i>	
DEVELOPMENT OF ARTS LEARNING MEDIA USING AUGMENTED REALITY WITH MARKERLESS BASED TRACKING METHOD AT SMKN 47	130-136
<i>Ghali Hanan, Budi Arifitama, Silvester Dian Handy Permana</i>	
Implementation of The Resident's Dues Applications (SIUMAS) Using Waterfall Method in RT X Cinere Village	137-142
<i>Febrianto Febrianto, Ikhwana Fakhir Nur Aziz, Mochammad Alvian Kosim, Muhammad Darwis, Retno Hendrowati</i>	
Implementation of Android-Based Tailoring Service Ordering Application with Geolocation Integration	143-146
<i>Muhamad Agung Santoso, Joko Sutopo</i>	
Forecasting Blood Demand Using the Support Vektor Regression Method (Case Study: Blood Transfusion Unit-PMI Central Lombok)	147-152
<i>Yati Apriati, Wafiah Murniati, Saikin Saikin, Sofiansyah Fadli, Hairul Fahmi</i>	
Improving Firebase BaaS Service Security in Counseling Chat Applications: AES-256 and CBC Approach for End-to-End Encryption	153-160
<i>Mogar Nurhandhi, Agus Suhendar</i>	
Comparison of ANN Backpropagation Algorithm and Random Forest Regression in Predicting the Number of New Students	161-166
<i>Padmavati Darma Tanuwijaya, Jhonatan Laurensius Tjahjadi, Yosefina Finsensia Riti</i>	

JISA	Vol : 6	Ed.2	Page : 096-188	Jakarta, Dec 2023	e-ISSN: 2614-8404	p-ISSN: 2776-3234
------	---------	------	-------------------	----------------------	----------------------	----------------------

Mobile Based Application Design of the Catholic Church of the Redeemer of Sumbawa Using the Waterfall Method <i>Claudio Orlando De Araujo, Muhammad Zakariyah</i>	167-173
Detection of Hate-Speech Tweets Based on Deep Learning: A Review <i>Ara Zozan Miran, Adnan Mohsin Abdulazeez</i>	174-188

JISA	Vol : 6	Ed.2	Page : 096-188	Jakarta, Dec 2023	e-ISSN: 2614-8404	p-ISSN: 2776-3234
-------------	----------------	-------------	---------------------------	------------------------------	------------------------------	------------------------------

Interactive Multimedia as Information Media Parasitic Infection with Multimedia Development Life Cycle Method

Dasril Aldo^{1*)}, Miftahul Ilmi², Sapta Eka Putra³, Widya Lelisa Army⁴

¹Informatics Engineering Study Program, Faculty of Informatics, Institut Teknologi Telkom Purwokerto

²Information Systems, Institut Indobaru Nasional

³Retail management, Universitas Tamansiswa Padang

⁴Sistem Informasi, Universitas Pertiwi

Email: ¹dasril@ittelkom-pwt.ac.id, ²miftahulilmi12@gmail.com, ³saptaeka54putra@gmail.com,
⁴widya.lelisa@pertiwi.ac.id

Abstract – This study focused on the problem of lack of interactive multimedia-based educational information media that is interesting as a provider of information about parasitic infections in humans. This parasitic infection is very dangerous, especially for children who still lack knowledge about it. The aim of this research is to provide educational information media to the wider community, especially at the Seberang Padang Health Center, about parasitic infections in a way that is as interesting as possible and easy to use. Interactive multimedia is a solution to this problem. This research uses the Multimedia Development Life Cycle Method as a guide in building interactive multimedia with the stages of Concept, Design, Material Collecting, Assembly, Testing and Distribution. The results of the alpha test showed that all features of this interactive multimedia can function properly. A beta test was conducted on 30 respondents from the public who accompanied patients to the Seberang Padang Health Center. As a result, this interactive multimedia received a rating of 4.54 out of 5. Therefore, this interactive multimedia is very good and suitable for use as an educational medium for information about parasitic infections. With this interactive multimedia-based educational information media, it is hoped that the wider community can better understand and be aware of parasitic infections and preventive measures that can be taken to avoid these infections.

Keywords – *Multimedia Interaktif, Multimedia Development Life Cycle, Infeksi Parasit*

I. INTRODUCTION

The problem discussed is that there is still a lack of interactive multimedia-based educational information media that is interesting as a provider of information about parasitic infections in humans. This parasitic infection is very dangerous, especially for children who still lack knowledge about it. Many dangers are posed by parasite infection because it can cause various diseases in humans. Some types of parasites that often attack humans are worms, fleas, mites, and flies. Here are some dangers of parasitic infections to watch out for:

1. Indigestion: Some parasites can invade the human digestive system, such as tapeworms and hookworms. This infection can cause diarrhea, nausea, vomiting, and abdominal pain.
2. Nervous system disorders: Parasitic infections such as *toxoplasma gondii* can cause disorders of the human nervous system. Symptoms include headaches, convulsions, and impaired coordination.
3. Anemia: Parasitic infections such as hookworms can cause anemia in humans. This happens because the parasite feeds on human blood.
4. Skin diseases: Some parasites such as scabies mites and head lice can cause skin diseases in humans. Symptoms that appear include itching, rashes, and wounds on the skin.

Respiratory system disorders: Parasitic infections such as *fasciolopsis buski* can cause disorders of the human respiratory system. Symptoms that appear include shortness of breath and coughing.

The development of information technology has provided many benefits to everyone. The use of information technology can facilitate many jobs, and continues to emerge various innovations in various fields such as online learning, interactive [1]games, interactive tutorials, [2]simulations, interactive product catalogs, interactive [3] animated videos [4] [5] , [6]interactive 3d presentations and [7] health education [8] . One form of technological development is an interactive multimedia-based system, which can be used as a medium of educational information as in this study which focused on parasitic infections in humans.

Interactive multimedia is one way to display information to be very interesting. There are still many systems that do not use interactive multimedia as a tool to display information because its manufacture takes time and also costs. Previous research on interactive multimedia, including by Wulandari in 2022 with the resultsof the use of interactive multimedia-based UNO card media, is feasible to be used to help improve early childhood cognitive abilities in recognizing and counting numbers[9]. Further research by Munawaroh, et al in 2022 with the results of interactive multimedia use has a significant influence on regional language learning in children. Interactive multimedia has an appeal that can grab



children's attention and is proven to improve their understanding associated with regional language learning at an early age[10]. Furthermore, by Wirantini, et al in 2022 with the results that interactive multimedia-based learning media cycle topics are feasible to use, interactive multimedia-based learning can help students in learning, especially in science material on water cycle topics.[11] Furthermore, by Yuliana, et al in 2022 with the results of interactive multimedia used as an addition to learning, so that children can access material anytime and anywhere without time and space restrictions. This product has been tested by experts and students, and is considered suitable for use as an effective learning media, as [12] well as many other studies such as h utan ebation animation[13], ornamental fish disease information [14] and plant disease information[15]. The application of interactive multimedia in this study is to provide information to the public about parasitic infections that can attack humans.

If there is no interesting information media to present this information, then most people, especially children, will feel less interested in reading it, even though this information is very important, knowledge regarding parasite information is still inadequate. This is caused by a lack of research on parasites, especially parasites that attack humans [16]. In developing interactive multimedia-based information media, there are several methods that can be used, such as the Rapid prototyping method, User-centered design, Agile development, Storyboarding, Object, Gamification and many other methods. In this research, the Multimedia Development Life Cycle method was used.

The MDLC method is a multimedia development consisting of six stages, namely the stages of *concept*, *design*, *material collecting*, *assembly*, *testing*, and *distribution* [17]. This method was chosen because it has complete stages in the development of an interactive multimedia system. This can be seen in the many multimedia-based application developments that use it, such as virtual reality as a medium for campus introduction by Citra, et al in 2022, then by Fatimah, et al in 2022 as a method for [18]learning jaipong dance.[19]

The purpose of this study is to create a medium of educational information to the wider public about parasitic infections. Through this media, it is hoped that the public can obtain information that is as interesting as possible and easy to use about the types of parasites, symptoms that appear, and how to prevent and treat appropriately. With sufficient knowledge, it is hoped that people can take better preventive measures and reduce the risk of parasitic infections.

II. RESEARCH METHODOLOGY

2.1 Research Stages

In a study, a research method is a series of procedures that must be carried out. The goal is that the author can design the research appropriately and not deviate from the subject matter. In addition, the systematization of steps in research also aims to be easily understood and can be used as a clear guideline in solving existing problems. In this study, the author used the Multimedia Development Life Cycle (MDLC) method depicted in Figure 1. This method was chosen because it

has clear and systematic stages in designing and developing multimedia. .

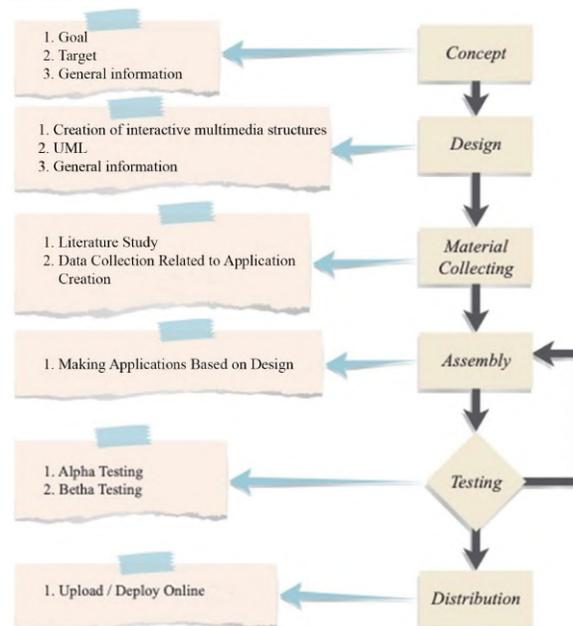


Fig 1. MDLC stage

There are six stages to the MDLC method, from Figure 1. can be described the process of MDLC, which is as follows:

1. Concept Stage

This stage is the purpose of making the system and who is the user of the designed system and also analyzing the needs of the system.

2. Design

The *design* stage is the stage of making program specifications such as material requirements or materials for the program, program display and program architecture. The design phase includes program structure, UML design and program design.

3. Material Collecting

The stage of collecting the materials to be used first. Then the material that has been collected will be continued at the next stage, namely *assembly*.

4. Assembly

The *assembly* stage is the stage of making all multimedia materials. Program creation is based on the design stage, program structure, UML design and program design.

5. Testing

This stage is carried out after completing the *manufacturing* stage (assembly). This stage is carried out by testing by the maker (alpha testing) in *black box* testing and *beta testing*.

6. Distribution

At this stage of distribution, the development team had to determine the right place to distribute interactive multimedia about parasitic infections. Some places that may be chosen include websites, mobile applications, or physical storage media such as CDs or flash drives. The selection of distribution sites must be adjusted to the target users and the purpose of the interactive multimedia. In addition, the development team must

also ensure that the interactive multimedia can be accessed easily and has good quality in order to provide optimal benefits for users.

2.1 Research Supporting Theories

There are several theories that will support this research process that are useful for the basis of research related to existing theories. The theory is as follows:

1. Multimedia

Multimedia is one of the learning media that combines several media elements presented in computer media. Interactive multimedia is a solution in making it easier for students to learn material compared to monotonous textbooks/*e-books*. In research Farida & Rahayu (2017) showed a significant difference in learning outcomes in the use of interactive multimedia rather than using textbooks [20]. The use of multimedia in learning will be able to produce more interesting learning and can increase students' understanding of the material provided[21]

2. Interactive Multimedia

Interactive Multimedia is a tool equipped with a controller that can be used by users to choose something expected, such as a gaming system and others. Interactive multimedia is a collection of several media such as technology, images, audio, *video*, and animation that are interactive in nature that are used to convey information[22]. Multimedia in learning allows everyone to process information visually and verbally, and utilize the brain's potential in learning. Making interactive multimedia can use a variety of systems or *software*[23].

III. RESULTS AND DISCUSSION

The results that can be obtained from the implementation of the MDLC method in building a Parasitic Infection information system based on interactive multimedia packaged in the form of interactive media in the form of educational *videos* include several menu options that support the system. Details of the stages that have been carried out in this study are as follows:

1. Concept

At this stage, researchers design the flow concept of a system that is built concisely and easily understood, so that this system can help educate about Parasitic Infections to the wider community, especially for children more easily. The concept flow built on this Parasitic Infection education system is as follows:

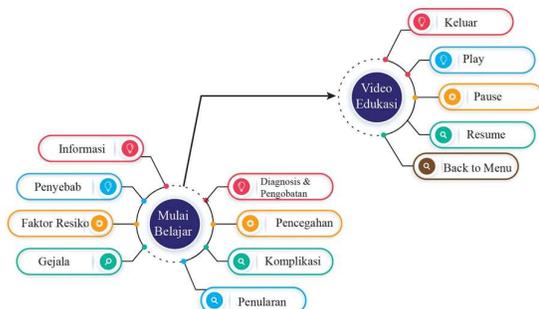


Fig 2. Interactive Multimedia Concept of Parasite Infection

a. Design

This design stage produces a research design using the program structure, UML and system design as an overview of the system that will be implemented later. The design made is made as attractive as possible so that people, especially children, are interested in using this interactive multimedia. For program structure design, UML and System design are as follows:

1. Program Structure

The following is the program structure of the parasitic infection interactive multimedia system:

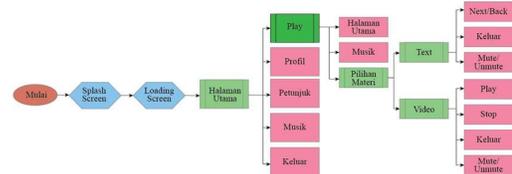


Fig 3. Structure of Parasitic Infection Interactive Multimedia System

2. UML Design

In making interactive multimedia Parasitic Infection, using UML which serves to provide visual modeling language or images to users of various kinds of programming and general engineering processes. The diagram used is as follows:

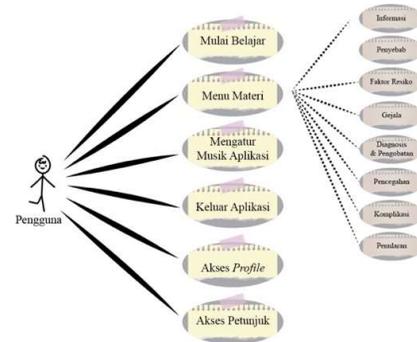
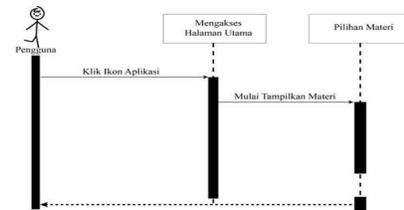


Fig 4. Use Case Interactive Multimedia Diagram

From Figure 4, a use case diagram is displayed from the interactive multimedia that is built, where the *use case diagram* will show how the interaction between the user and the system. In this interactive multimedia, there are one actor and eight main use cases and three use case include in the *material menu use case*. The next UML diagram is a *sequence diagram* as shown in figure 5.



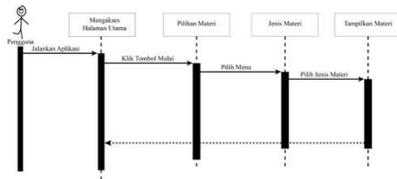


Fig 5. Interactive Multimedia Sequence Diagram

Sequence Diagram serves to explain and display the interaction between objects in a system in detail. In this interactive multimedia, there are two sequence diagrams, where the first sequence diagram shows the user when he wants to start the application. The second sequence diagram, when the user runs the application. The third diagram is the Activity Diagram.

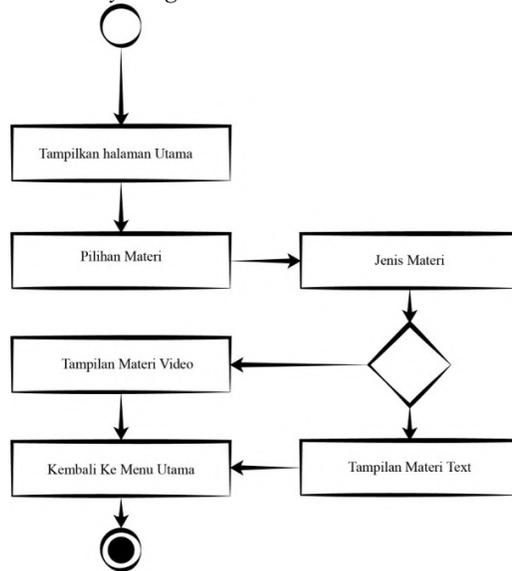


Fig 6. Interactive Multimedia Activity Diagram

The Activity Diagram will illustrate system activities not what actors do, so activities that can be performed by the system. In the interactive multimedia that will be built there are six activities that can be done by users, from displaying the main page, selecting material, choosing the type of material, displaying material in the form of text or video and returning to the main menu.

3. Design Program

Program design is one of the important things in building an interactive multimedia, from this design will be a guide to how the appearance and layout of menus and content will be displayed on interactive multimedia to be built. Here is the program design:



Fig 7. Interactive Multimedia Initial Display Design

The initial screen will be designed with five buttons consisting of profiles and instructions on the upper left side, mute and exit buttons on the upper right side, start button at the bottom. In the middle of the screen will be given text in the form of the title of the multimedia to be built. The next view is the profile page.

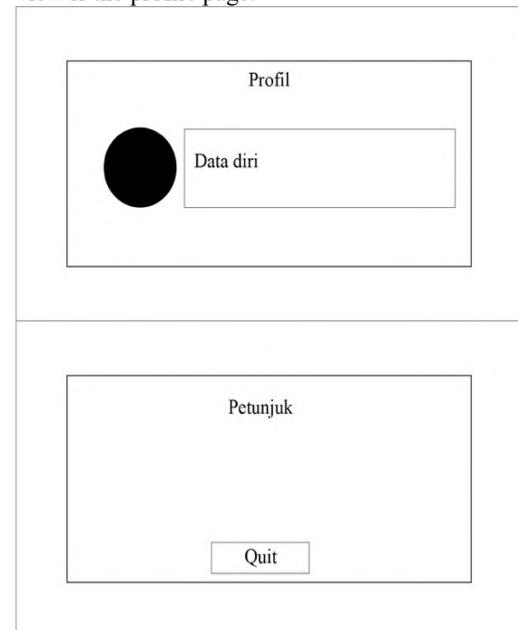


Fig 8. Profile Display Design & How-To Page

In the design of the profile page, it will be given a display in the form of photos from interactive multimedia creators and also personal biodata. On the top right, an exit feature will be given to exit the profile page and return to the main page. The next design is the instructions page.

On the instructions page will be designed with a display in the form of a page title in the upper center, then text will be given in the form of instructions at the bottom of the instruction title and on the upper right side will be given an exit button to return to the main page. Next page design type of Parasitic Infection

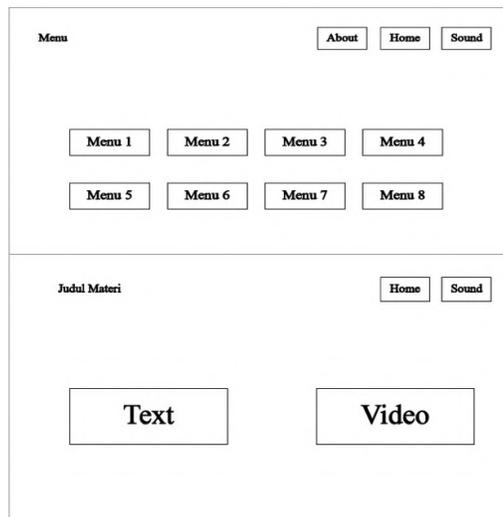


Fig 9. Menu Page Design Select Parasite Information Type & Material Selection Page

In the menu design, there is a choice in the form of a selection of materials regarding parasitic infections. If the user clicks on one of these menus, interactive multimedia will display a selection of material in the form of text material or *video* material, depending on the user's choice. On the upper left side is also displayed a button to return to the *home* page and a button to turn off music. Next, design the material selection page. On the design of the page the material appears, the title of the type of parasite information selected will be given, then in the middle will be given an option button in the form of text or *video* material, later the system will display material based on the choice of the user. At the top right will be given three additional buttons in the form of *home*, instructions and *mute*. The next design is a page of material text.

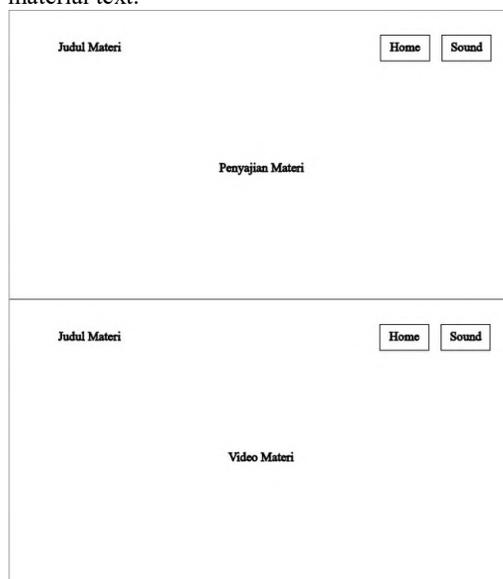


Fig 10. Material Text & Video Page Design

On the material text page, later the type of

parasitic infection will be displayed, then under the title will be given material in the form of text and on the left and right sides of the middle there will be navigation buttons for *next* and *previous* material. At the top right of the page there are three buttons, namely *home*, instructions and *mute*. The next design is the *Video* Material page. On the *video* material page, later the selected type of parasite information will be displayed, then under the title will be given material in the form of a *Video* that can be played by users and on the top right isis of the page there are three buttons, namely *home*, instructions and *mute*.

4. Material Collecting

At this stage, researchers will collect mater materials and assets to make interactive multimedia. At this stage the author uses adobe animate and adobe illustrator.



Fig 11. Process Asset Collection with Adobe Photoshop & Asset Collection with Adobe Animate

The collection process using Adobe Illustrator by designing and creating the necessary assets according to the function of each asset for interactive multimedia Parasitic Infection. The next collection of assets using Adobe Animate.

This process combines assets that have been created in Adobe Illustrator. At this stage, the process of selecting animations, transitions, *placing frames*, and so on is also carried out so that when used by interactive multimedia users it can run as attractive as possible.

5. Assembly

At this stage, researchers build interactive multimedia as interesting as possible that can meet information to users about Parasitic Infections, along with the results of the *assembly* process:



Fig 12. Home Interface

Setelah user open interactive multimedia, than the main page will appear, on the main page displayed several buttons that can be used by users to operate this Parasitic Infection interactive multimedia. These buttons are the *profile* button, hint button, music button, exit button and *play* button. The next interface is the *profile* page. 3



Fig 13. Interface Profile & Instructions

The profile page will appear when the user clicks the profile button on the main page. On the profile page is displayed the profile of the developer of interactive multimedia Parasitic Infection. The next page is the instructions page. The instructions page will be active if the user clicks on the hint button. On the page will appear instructions from *Video* material and text. Next, a Parasitic Infection drop-down menu page will be displayed that will be known information.



Fig 14. Menu Interface: Select Parasite Information Type &

Material Presentation Type

On this page, users will be presented with eight menus containing information on Parasitic Infections. If the user clicks on one of the Parasitic Infection options, the system will display information in the form of *Video* and text options. Next, the interface of the choice page type of information display. When the user clicks on one of the options menu for the type of Parasitic Infection, the system will display two options for presenting information, namely text and video. Next will be displayed the page of text material.

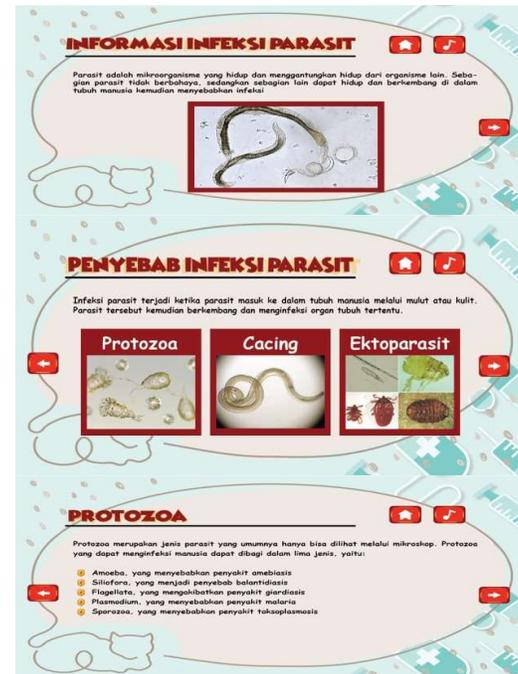


Fig 15. Display of Text Material

In the text material, information will be displayed in the form of text containing an explanation of the Parasitic Infection selected by the user. The text material is also provided with illustrative images so that users know the form of the Parasitic Infection in question. Next will be displayed *Video* material.



Fig 16. Video Materials

On the Video materi, information about the selected parasitic infection will be displayed, the information provided is also the same as the text but in the form of a *Video*, this is intended so that

users who cannot read the text can watch and listen so that they can better understand the information from the Parasitic Infection.

6. Testing

At this stage, researchers will carry out two testing processes, namely Alpha testing tested by researchers themselves and Beta testing tested

directly to users.

a) Alpha Testing

In this test, researchers will test the appearance and function of each button in the interactive multimedia Parasitic Infection as in Table 1.

Table 1. Alpha Testing

No	Components tested	Test Scenarios	Testing	Information
1	Start Page	Select the "Tap to Play" button	BlackBox	✓ Succeed
2	Home	Selecting the "Profile" button	BlackBox	✓ Succeed
		Selecting the "Hint" button	BlackBox	✓ Succeed
		Selecting the "Music" button	BlackBox	✓ Succeed
		Selecting the "Exit" button	BlackBox	✓ Succeed
		Selecting the "Play" button	BlackBox	✓ Succeed
3	Menu Options	Selecting the "Music" button	BlackBox	✓ Succeed
		Selecting the "Home" Button	BlackBox	✓ Succeed
		Selecting the "Insects" button	BlackBox	✓ Succeed
		Selecting the "Fish" button	BlackBox	✓ Succeed
		Selecting the "Reptile" button	BlackBox	✓ Succeed
4	Material Selection	Selecting the "Home" Button	BlackBox	✓ Succeed
		Selecting the "Hint" button	BlackBox	✓ Succeed
		Selecting the "Music" button	BlackBox	✓ Succeed
		Selecting the "Text" button	BlackBox	✓ Succeed
		Selecting the "Video" button	BlackBox	✓ Succeed
5	Material Page	Selecting the "Home" Button	BlackBox	✓ Succeed
		Selecting the "Hint" button	BlackBox	✓ Succeed
		Selecting the "Music" button	BlackBox	✓ Succeed
		Selecting the "Next" button	BlackBox	✓ Succeed
		Selecting the "Previous" button	BlackBox	✓ Succeed
6	Video	Selecting the "Play" button	BlackBox	✓ Succeed
		Selecting the "Pause" button	BlackBox	✓ Succeed
		Selecting the "Home" Button	BlackBox	✓ Succeed
		Selecting the "Hint" button	BlackBox	✓ Succeed
		Selecting the "Music" button	BlackBox	✓ Succeed
	Profile	Selecting the "Exit" button	BlackBox	✓ Succeed
	Instructions	Selecting the "Exit" button	BlackBox	✓ Succeed

From Table 1. Eight tests were carried out on the features of interactive multimedia, each component can run properly without any errors.

b) Beta Testing

In beta testing, it will involve as many as 30 respondents or users to see the feasibility of interactive multimedia built. There are 14 assessment indicators and five values, namely: 1 = very not good, 2 = not good, 3 = not good, 4 = good enough, 5 = very good. For example, with the indicator "The material presented in interactive learning media is in accordance with the learning objectives"

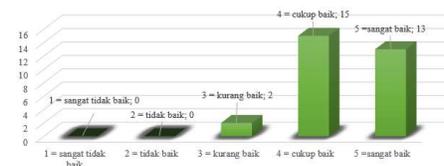


Fig 17. Measurement of Indicator 1

In Figure 17. Showing measurements of assessment indicator 1 with results, 13 respondents gave very good marks, 15 respondents gave quite good marks, 2 respondents gave poor grades, 0 respondents gave bad and very bad grades. So if calculated by rating, it is as follows:

$$\begin{aligned}
 1 = \text{very bad} &= 0 * 1 = 0 \\
 2 = \text{not good} &= 0 * 2 = 0 \\
 3 = \text{not good} &= 2 * 3 = 6 \\
 4 = \text{good enough} &= 15 * 4 = 60 \\
 5 = \text{very good} &= 13 * 5 = 65 \\
 \text{Total} &= 30 = 131 \\
 \text{Rating} &= 131 / 30 = 4.37
 \end{aligned}$$

The calculation of ratings on other indicators



is also carried out with the following results:

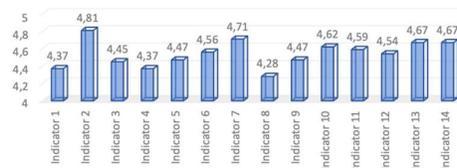


Fig 18. Rating Results

From the results obtained in Figure 18 is 4.54 out of a scale of 5, so with this value interactive multimedia that is built is very feasible to be used as a medium of information about Parasitic Infections.

7. Distribution

The distribution process through websites and playstores, this is because at present, the most appropriate interactive multimedia distribution place for the general public is through websites and mobile applications. This is because the use of the internet is increasingly widespread and easily accessible to the public. In addition, interactive multimedia published through websites and mobile applications can be accessed anytime and anywhere by users.

IV. CONCLUSION

Interactive multimedia on parasitic infections is an effective educational information medium for improving public knowledge of parasitic infections. This interactive multimedia can be used by various stakeholders, such as health centers, hospitals, and public health organizations. It can be used to educate the public about parasitic infections, thus preventing the spread of the disease. This multimedia was developed using the MDLC method, which consists of six phases. Beta testing of 30 community respondents who accompanied patients for treatment at the Seberang Padang Health Center resulted in a highly positive evaluation. The interactive multimedia was rated as engaging and interactive, containing accurate and up-to-date information, and being easily accessible.

REFERENCES

- [1] S. Prihatiningtyas and M. W. Tijanuddarori, "Media Interaktif E-Modul Biologi Materi Virus Sebagai Pendukung Pembelajaran Daring Di Man 3 Jombang," *Journal of Educatio*, vol. 4, no. 3, pp. 1–8, 2021.
- [2] O. Olisna, M. Zannah, A. Sukma, and A. N. Aeni, "Pengembangan Game Interaktif Wordwall untuk Meningkatkan Akhlak Terpuji Siswa Sekolah Dasar," *Jurnal Basicedu*, vol. 6, no. 3, pp. 4133–4143, Apr. 2022, doi: 10.31004/basicedu.v6i3.2737.
- [3] Y. Abdillah, S. Susilaningih, and A. Wedi, "Pengembangan Multimedia Tutorial Materi Tata Surya Untuk Membantu Siswa Belajar Di Rumah," *JKTP: Jurnal Kajian Teknologi Pendidikan*, vol. 4, no. 1, pp. 98–107, Feb. 2021, doi: 10.17977/um038v4i12021p098.
- [4] S. : A. Swandi *et al.*, "Pengembangan Simulasi Interaktif Gerak Parabola Menggunakan VBA Excel Dalam Pembelajaran Fisika," *Jurnal Riset Pendidikan Fisika*, vol. 6, no. 1, 2021, [Online]. Available: <http://journal2.um.ac.id/index.php/jrpf/>
- [5] Novia Risdianti and Bramanta Octa Danu Putra, "Perancangan Katalog Digital Interaktif untuk Membantu Remaja Putri Memilih Produk Makeup Emina Secara Online," *Jurnal Desain Komunikasi Visual*, vol. 2, no. 1, pp. 22–27, 2022.
- [6] E. D. Hardiningrum, T. Djudin, and A. Nurdini, "Pengembangan Media Video Animasi Berbasis Aplikasi Telegram Pada Pembelajaran Ipa Materi Tata Surya," *JPPK*, vol. 12, no. 3, pp. 998–1005, 2023, doi: 10.26418/jppk.v12i3.63834.
- [7] Y. Dela Carolina, "Augmented Reality sebagai Media Pembelajaran Interaktif 3D untuk Meningkatkan Motivasi Belajar Siswa Digital Native," *Ideguru: Jurnal Karya Ilmiah Guru*, vol. 8, no. 1, pp. 10–16, Oct. 2022, doi: 10.51169/ideguru.v8i1.448.
- [8] S. Hasta Mulyani and M. Julian Jodi, "PERANCANGAN AUGMENTED REALITY UNTUK LITERASI KESEHATAN GIGI PADA ANAK," *SINTaKS*, vol. 7, no. 1, pp. 1–7, 2022.
- [9] Abimanyu, E. A. Wijaya, Endro Muhammad Akbar Wijiantoro, F. R. Kamaludin, J. L. Ramadhan, and D. Aldo, "Sistem Informasi Mengenalkan Makanan Khas Jawa Tengah Menggunakan Multimedia Interaktif," *JURSIMA*, vol. 10, no. 1, pp. 269–274, 2022.
- [10] H. Munawaroh *et al.*, "Pembelajaran Bahasa Daerah melalui Multimedia Interaktif pada Anak Usia Dini," *Jurnal Obsesi: Jurnal Pendidikan Anak Usia Dini*, vol. 6, no. 5, pp. 4057–4066, Apr. 2022, doi: 10.31004/obsesi.v6i5.1600.
- [11] N. Putu, N. Wirantini, G. Astawan, and I. Gede Margunayasa, "Media Pembelajaran berbasis Multimedia Interaktif pada Topik Siklus Air," *Jurnal Edutech Undiksha*, vol. 10, no. 1, pp. 42–51, 2022, doi: 10.23887/jeu.v10i1.46558.
- [12] F. D. Yuliana, Susilaningih, and Z. Abidin, "PENGEMBANGAN MULTIMEDIA INTERAKTIF BERBASIS MOBILE PADA BAHASA INGGRIS," *JKTP: Jurnal Kajian Teknologi Pendidikan*, vol. 5, no. 1, pp. 11–21, Feb. 2022, doi: 10.17977/um038v5i12022p011.
- [13] M. B. Putra *et al.*, "Media Pembelajaran Animasi Reboisasi Hutan untuk Siswa kelas VIII SMA Telkom Puwokerto," vol. 10, no. 1, 2022.
- [14] K. N. Sinurat1 *et al.*, "SISTEM INFORMASI PENYAKIT PADA IKAN HIAS BERBASIS MULTIMEDIA INTERAKTIF," vol. 10, no. 1, 2022.
- [15] D. Larasae, F. Y. Pratama, N. D. Hammam, I. Y. Hutama, D. Prasetyo, and D. Aldo, "SISTEM INFORMASI PENYAKIT PADA TUMBUHAN BERBASIS MULTIMEDIA INTERAKTIF,"



- JURSIMA (Jurnal Sistem Informasi dan Manajemen)*, vol. 10, no. 1, 2022.
- [16] M. Ristić *et al.*, “Soil and sand contamination with canine intestinal parasite eggs as a risk factor for human health in public parks in Niš (Serbia),” *Helminthologia*, vol. 57, pp. 109–119, 2020, doi: 10.2478/helm20200018.
- [17] Mustika, “RANCANG BANGUN APLIKASI SUMSEL MUSEUM BERBASIS MOBILE MENGGUNAKAN METODE PENGEMBANGAN MULTIMEDIA DEVELOPMENT LIFE CYCLE (MDLC),” *Mikrotik*, vol. 8, no. 1, pp. 1–14, 2018.
- [18] Y. Purnama Citra, D. Zaliluddin, D. Abdurahman, P. Informatika, and U. Majalengka, “APLIKASI VIRTUAL REALITY SEBAGAI MEDIA PENGENALAN KAMPUS BERBASIS ANDROID DENGAN MENGGUNAKAN METODE MULTIMEDIA DEVELOPMENT LIFE CYCLE,” *SNETIK*, pp. 165–170, 2022, doi: 10.31284/p.snestik.2022.2711.
- [19] D. Destiani, S. Fatimah, R. Erwin, G. Rahayu, and M. D. Jaelani, “Media Pembelajaran Tarian Jaipong Berbasis Android Menggunakan Multimedia Development Life Cycle,” *Algoritma*, vol. 19, no. 2, pp. 759–769, 2022, [Online]. Available: <https://jurnal.itg.ac.id/>
- [20] F. Armansyah, Sulton, and Sulthoni, “MULTIMEDIA INTERAKTIF SEBAGAI MEDIA VISUALISASI DASAR-DASAR ANIMASI,” *JKTP Jurnal Kajian Teknologi Pendidikan*, vol. 2, no. 3, pp. 224–229, 2019.
- [21] N. Azlina, “Implementasi Media Pembelajaran Interaktif Mata Kuliah Pengantar Multimedia Menggunakan App Inventor Berbasis Android,” *CIRCUIT*, vol. 5, no. 2, pp. 155–165, 2021.
- [22] Irwanto and Program, “Pengembangan Bahan Ajar Berbasis Multimedia Interaktif untuk Mata Kuliah Elektronika Daya,” *JURNAL PENDIDIKAN*, vol. 30, no. 2, pp. 353–368, 2021.
- [23] S. Hidayat and A. A. Syahputa, “Sistem imun tubuh pada manusia,” *VISUAL HERITAGE*, vol. 2, no. 03, pp. 144–149, 2020.



Penetration Testing Tangerang City Web Application With Implementing OWASP Top 10 Web Security Risks Framework

Yoel Armando¹, Rosalina^{2*}

^{1,2}Study Program Informatics, Faculty of Computing, President University, Indonesia

*Email: rosalina@president.ac.id

Abstract – The speed of technological development has made it possible for all people to be connected to one another. The creation of web-based information systems that help in all areas, including government, health, and education, is one of the forces behind the development of technology. With these technological advancements, websites are susceptible to cybercrimes that could end in the theft of crucial data. Top 10 Web Application Security Risks is the most effective prevention process for decrease company information leaks. On the website tangerangkota.go.id, the researcher will conduct a test using the Top 10 Web Application Security Risks technique. Top 10 Web Application Security Risks consist of Broken Access Control, Cryptographic Failures, Injection, Insecure Design, Security Misconfiguration, Vulnerable and Outdated Components, Identification and Authentication Failures, Software and Data Integrity Failures, Security Logging and Monitoring Failures, Server-Side Request Forgery. The penetration testing results found on the Tangerang City website which are 4 injections, 2 broken access controls, 1 security misconfiguration.

Keywords – Penetration Testing, OWASP 10, Vulnerability Assessment, Exploitation.

I. INTRODUCTION

The advancement of information technology (IT), particularly the internet and information systems or applications that are web-based, has significantly altered human life. The security of users must be carried out in conjunction with this rapid technological advancement.

Internet users who browse the web need to be shielded. As a result, users should consider security as an important indicator. Due to the importance for protection, every field possesses distinct levels of security. Security in the public field (domain *.go.id) is distinct from those found in educational institutions (domain *.sch.id or.ac.id). There were 976.429.996 cyberattacks in 2022, with malware activities, web defacement, data breaches, human-operated ransomware, phishing, and advanced persistent threats continuing to cause the majority of traffic anomalies [1]. According to the report, the government agency sector was attacked by about 68% of respondents, followed by e-commerce by 17%, financial institutions by 11%, social media by 3%, and cryptocurrencies by 1%.

Tangerang in the province of Banten uses a web-based information system (with the domain * https://tangerangkota.go.id/) to disseminate information about official activities that is beneficial to residents in general and Tangerang residents in particular. Unfortunately, it cannot be denied that defacement attacks from outside sources led to user information being exposed within the Tangerang government website. One of the world's defacement archives, www.zone-h.org, allows users to view the attack's history. An archive website called Zone-H maintains a list of websites that have been affected.

Zone-h data indicates that cyberattacks caused by web defacement have occurred against the Tangerang website. Despite the fact that the attack took place in 2016, it may still occur due to a fresh revealing (vulnerability) that the manager of the Tangerang website has not yet discovered

and may be exploited again.

The aims of this study is to analyze web security testing with the domain tangerangkota.go.id from attacks from irresponsible outsiders that can damage Tangerang City, analyze applications under development for web security testing, implement a system tool for testing security gaps in web applications.

Muhammad Subagja is testing the application security system using Zero Entry Hacking. The results are that each web he tests has 41 vulnerabilities, and the average CVSS is medium level [2]. Ahmad Fikri Zulfi found an XSS gap on two subdomains of Jember University that can be exploited to change the web's appearance or insert links to enter other websites. The vulnerability cannot be exploited further [3]. The definition of vulnerability is a weak point where a system is vulnerable to attack [4]. Vulnerability is a weakness that threatens an asset's integrity, confidentiality, and availability [5].

Afif Zirwan uses Acunetix WVS tools and information system security principles to test the Institute of Technology Padang website, and have six vulnerabilities with the CVSS are medium and high levels [6]. Common Vulnerability Scoring System (CVSS) is an open framework used to communicate the characteristics and impact of an application vulnerability. A decrease in the vulnerability score after improvement indicates that the security evaluation carried out has been able to reduce the risks that previously existed, and it can be said that the website has become relatively safe [7].

Harry and M. Akbar analysis of binadarma.ac.id vulnerability in the outdated version application using SQL Injection attacks and found several vulnerabilities due to the outdated application [8].

Mira and Michael do security analysis using vulnerability assessment principles and NMAP to check for denial of service attacks, find XSS vulnerabilities in PHP files, and find vulnerabilities to SQL Injection attacks [9]. Nmap or Network Mapper is an open-source



network security exploration and auditing tool [10]. Nmap is an application used to scan Mikrotik routers to identify existing vulnerabilities [11]. Nmap is a tool used to determine the computer's services through port scanning [12].

Feri, Harjono, and Agung analyzed security gaps using the open vulnerability assessment system and Acunetix web method to compare the result of the two tools and the results they found weaknesses in 9 data on OpenVAS with the scanning time of 60 minutes. In contrast, Acunetix WVS found weaknesses in 166 data with scanning times of 954 minutes [13].

Vulnerability Assessment (VA) is a process carried out on a website to define, identify and classify possible security gaps in computer networks or communication infrastructure [14]. Vulnerability Assessment is a step to detect, identify and study a computer system's or network infrastructure's weaknesses [15].

II. RESEARCH METHODOLOGY

The steps taken in testing the security system can be seen in Figure. 1.



Fig1. Top 10 OWASP Framework

Based on the standards issued by the OWASP Framework, there are ten steps that can be taken to assess and test the security of a website:

1. **Broken Access Control**
These vulnerabilities allow attackers to exploit these weaknesses to access unauthorized data or functions.
2. **Cryptographic Failures**
This vulnerability arises due to the improper use of cryptographic technology in web applications, resulting in confidential data being accessible to attackers.
3. **Injection** This vulnerability is due to inadequate input methods in web applications, allowing attackers to insert malicious codes into systems that use inputs such as forms, URL parameters, or comment fields.
4. **Insecure Design**
This vulnerability arises when the manager neglects the proper design of the security system in the web application, such as not creating a One Time Password (OTP) or authentication feature on the login page. This allows attackers to access sensitive data and even steal user information.
5. **Security Misconfiguration**
This vulnerability occurs when unused features remain enabled or installed, the security system is not updated to the latest version, and default passwords are used. This allows

attackers to perform DDOS attacks or brute force attacks and obtain sensitive data, such as user information.

6. **Vulnerable and Outdated Components**
This vulnerability occurs in web applications that do not use the latest components or libraries and have vulnerabilities that can be exploited.
7. **Identification and Authentication Failures**
This vulnerability occurs in web applications when attackers have a list of valid usernames and passwords, or when weak default passwords such as admin/admin are used.
8. **Software and Data Integrity Failures**
This vulnerability occurs in web applications when attackers have a high probability of uploading updates to the system that they distribute, which can be downloaded or run by all installations.
9. **Security Logging and Monitoring Failures**
This vulnerability occurs in web applications when managers ignore the results contained in logs, such as login failures, transaction failures, and unclear error warnings. This allows attackers to exploit this to perform various injections into the security system.
10. **Server-Side Request Forgery**
This vulnerability occurs in web applications that allow attackers to manipulate HTTP requests made on the server side. This allows attackers to connect to the internal network or internet connection to the server to steal important data.

Black box testing is utilized during the process of penetration testing. The research process commences with planning and review, during which the object of research is selected and pertinent data is gathered. Subsequently, the scanning of the research object is carried out. Once exploitable information results are obtained from the scan results, the next step involves the use of tools or manual exploitation to exploit the gap information. For the purpose of penetration testing, researchers utilize a set of nine tools, which include NMAP, ZAP, SQLmap, BurpSuite, Wappalyzer, The Harvester, WHOIS, and CMD.

III. RESULTS AND DISCUSSION

Table 1. Project Summary

Type	Description	URL
Project/Object Scope	Web Dapodik Web sipdatapsm Web Sabakota Web silat	https://dapodik.tangerangkota.go.id/ https://sipdatapsm.tangerangkota.go.id/ https://sabakota.tangerangkota.go.id https://silat.tangerangkota.go.id/
Area	Production	
Methodology	Black-box Testing	
Date/Period	01 April 2023	Pentest #1



As showed in Table 1, the research object's scope is outlined, with the focus being on the production area that is currently utilized by the user. The methodology adopted by the researchers is black box testing. This approach signifies that the researchers undertake penetration testing without any prior knowledge of the research object.

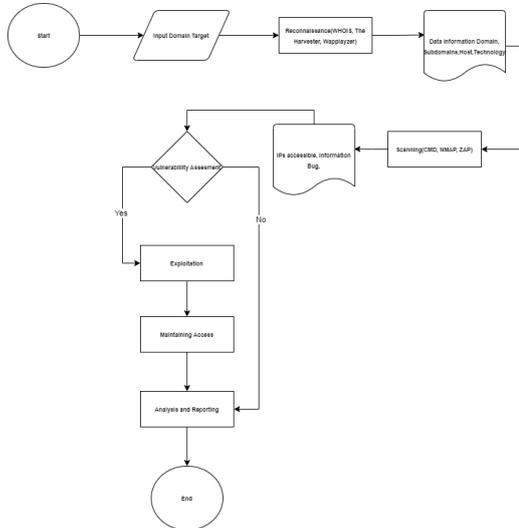


Fig 2. Step Penetration Testing

Figure 2 provides a detailed explanation of the steps executed by the researcher. There are several steps such as reconnaissance, scanning, exploitation, maintaining access, also analysis and reporting. In this context, reconnaissance serves to gather information about the domain, which will be highly beneficial for the scanning process. The results obtained from the reconnaissance process are as follows:

Table 2. Process of Reconnaissance

Tools Name	Data Obtained	Tools Function
Whois	Informasi Domain (Domain Created, Name hostmaster, Cellphone Number, Total Server)	Reconnaissance
The Harvester	Emails and Total Host	Reconnaissance
Wappalyzer	Built using PHP technology, nginx server, and javascript library	Reconnaissance

As showed in Table 2, the subsequent step involves scanning, which serves to identify bug information associated with the domain. WHOIS is also an exceptional method to identify the domain owner. However, they also possess their own applications to search for information in the WHOIS database. This implies that any information an individual seeks will be presented in detail there [16]. The Harvester functions as a comprehensive information-gathering

tool that is utilized by both ethical and non-ethical hackers to scrape emails, subdomains, hosts, employee names, open ports, and banners from various public sources, such as popular search engines [17]. The results obtained from the scanning process are as follows:

Table 3. Process of Scanning

Tools Name	Data Obtained	Tools Function
CMD	Accessible IP	Scanning
NMAP	Port, State, Service	Scanning
ZAP	Vulnerability Info, Risk Level	Scanning

As showed in Table 3, after obtaining information about the vulnerability contained in the domain, exploitation can be carried out based on the results of the scanning. Nmap is used to scan Mikrotik routers to identify existing vulnerabilities [18]. OWASP-ZAP is a vulnerability scanner application that is freely available (open source) and has been developed by the OWASP organization. It is extremely useful in assisting with the discovery of bug information [19]. The purpose of exploitation here is to verify whether the detected bug truly resides within the website and to identify any other potential bugs that have not yet been confirmed. The outcomes of the exploitation process are as follows:

Table 4. Process of Exploitation

Tools Name	Data Obtained	Tools Function
Burp Suite	Information Disclosure - Dapodik, Reflected XSS - Dapodik, Insecure Direct Object Reference (IDOR) – Tambah Dana LPJ Sabakota, Insecure Direct Object Reference (IDOR) – Cetak Profil Sabakota, Stored XSS - Sabakota	Exploitation
SQLmap	SQL Injection - Silat, SQL Injection - Sipdatapsm	Exploitation

As showed in Table 4, following the exploitation, it was found that there were several bugs in the Tangerang City subdomain. SQLMap is an open-source application or tool included in Kali Linux. This application is used to detect and exploit vulnerabilities in web applications. Researchers use this application during the exploitation phase [20]. The researchers then carried out maintaining access, which serves to re-exploit to reconfirm whether the bug has been fixed or remains unfixed. The results from the maintaining access process are as follows:



REFERENCES

Table 5. Process of Maintaing Access

No.	Vulnerability Name	Category	Risk
1	SQL Injection - Silat	Injection	A3
2	SQL Injection - Sipdatapsm	Injection	A3
3	Stored XSS - Sabakota	Injection	A3
4	Insecure Direct Object Reference (IDOR) – Cetak Profil Sabakota	Broken Access Control	A1
5	Insecure Direct Object Reference (IDOR) – Tambah Dana LPJ Sabakota	Broken Access Control	A1
6	Reflected XSS - Dapodik	Injection	A3
7	Information Disclosure - Dapodik	Security Misconfiguration	A5

As showed Table 5, in the final stage, the researcher conducted an analysis related to the bug findings on the Tangerang City website. The researchers categorized the vulnerabilities found according to the OWASP Top 10 Web Security Risks Framework and conducted a security vulnerability assessment using the Common Vulnerability Scoring System (CVSS). The results are as follows:

Table 6. Process of Analysis and Reporting

No.	Vulnerability Name	Severity	Status
1	SQL Injection - Sipdatapsm	Critical	Validate
2	SQL Injection - Silat	High	Validate
3	Stored XSS - Sabakota	High	Validate
4	Insecure Direct Object Reference (IDOR) – Cetak Profil Sabakota	High	Validate
5	Insecure Direct Object Reference (IDOR) – Tambah Dana LPJ Sabakota	High	Validate
6	Reflected XSS - Dapodik	Medium	Validate
7	Information Disclosure - Dapodik	Low	Validate

As showed Table 6, the researchers discovered seven vulnerabilities within the Tangerang city website. The vulnerabilities found include one in the critical category, four in the high category, one in the medium category, and one in the low category.

[1] M. Ayu, “BSSN Paparkan Serangan Keamanan Siber di Tahun 2022 Alami Penurunan Dibanding Tahun 2021,” *cloudcomputing.id*, Jan. 24, 2023. [Online]. Available: <https://www.cloudcomputing.id/berita/bssn-paparkan-serangan-siber-alami-penurunan>.

[2] M. Subagja, “Penetration Testing Terhadap Website Asosiasi Pekerja Professional Informasi Sekolah Indonesia (APISI),” 2019.

[3] Ahmad Fikri Zulfi, “Evaluation of Student Information System Application Security Using Vapt Framework (Case Study: Sister Universitas Jember),” 2017.

[4] A. Susanto and W. K. Raharja, “Simulation and Analysis of Network Security Performance Using Attack Vector Method for Public Wifi Communication,” *IJICS (International J. Informatics Comput. Sci.*, vol. 5, no. 1, 2021, pp. 7–15.

[5] D. C. Angir, Agustinus, Justinus, “Vulnerability Mapping Pada Jaringan Komputer Di Universitas X”, *Jurnal infra* vol. 3, no. 2, 2015.

[6] Afif Zirwan, “Pengujian Dan Analisis Keamanan Website Institut Teknologi Padang Menggunakan Acunetix Vulnerability Scanner,” 2022.

[7] A. Marta, D. Setiyadi, and Fata, “Keamanan Website Menggunakan Vulnerability Assessment,” *Information for Educators and Professionals*, vol. 2, no. 2, Juni 2018, pp. 171-180.

[8] Harry, M. Akbar, Andri, “Vulnerability Assessment Pada Web Server,” 2018.

[9] Mira and Michael, “Vulnerability Assesment Untuk Meningkatkan Kualitas Kemanan Web,” 2021.

[10] Brown and Nicholas, *Nmap 7: From Beginner to Pro*. USA: Independently Published, 2019.

[11] I. Kamilah and A. Hendri Hendrawan, “Analisis Keamanan Vulnerability pada Server Absensi Kehadiran Laboratorium di Program Studi Teknik Informatika,” *Pros. Semnastek*, vol. 16, no. 0, 2019, pp. 1–9.



- [12] M. Anis and Emah, “*Network Security Monitoring with Intrusion Detection System,*” JUTIF, vol. 3, no. 2, April 2022, pp. 249-253.
- [13] Feri, Harjono, and Agung, “*Uji Vulnerability pada Website Jurnal Ilmiah Universitas Muhammadiyah Purwokerto Menggunakan OpenVAS dan Acunetix WVS,*” 2019.
- [14] Mona, “*Analisis Celah Keamanan Website Sitasi Menggunakan Vulnerability Assessment,*” Jurnal Ilmiah Rekayasa dan Manajemen Sistem Informasi, vol. 1, no. 9, 2023, pp. 1-7.
- [15] Fadli and Sofyan, “*Implementasi Metode Vulnerability Dan Hardening Pada Sistem Keamanan Jaringan,*” Jurnal Teknologi Elektro, vol. 9, no. 1, 2018.
- [16] Timoteus and Jimmy, “*Analisis Yuridis Pelaksanaan Tugas Pokok Pengelola Domain Internet Indonesia,*” NJLO, vol. 1, no. 1, Juli 2020, pp. 53-63.
- [17] Thecybersecurityman, “*PenTest Edition: Using “theHarvester” to Gather Email accounts, Subdomains, Hosts, LinkedIn Users, Banner Information, and More!*” thecybersecurityman.com, Aug. 1, 2018. [Online]. Available:
<https://thecybersecurityman.com/2018/08/01/pentest-edition-using-theharvester-to-gather-e-mail-accounts-subdomains-hosts-linkedin-users-banner-information-and-more/>
- [18] I. Kamilah and A. Hendri Hendrawan, “*Analisis Keamanan Vulnerability pada Server Absensi Kehadiran Laboratorium di Program Studi Teknik Informatika,*” Pros. Semnastek, vol. 16, no. 0, 2019, pp. 1–9.
- [19] Dennis, Muhandi, and Warih, “*Analisis Resiko Keamanan Terhadap Website Dinas Penanaman Modal Dan Pelayanan Terpadu Satu Pintu Pemerintahan Xyz Menggunakan Standar Penetration Testing Execution Standard (Ptes),*” e-Proceeding of Engineering, vol. 7, no. 1, April 2020, pp. 2090.
- [20] Sudiharyanto, Roy, and Ihsan, “*Analisa Serangan Sql Injeksi Menggunakan Sqlmap,*” Jurnal Sistem dan Teknologi Informasi, vol. 4, no. 2, 2018, pp. 88-94.



Sentiment Analysis of 2024 Presidential Candidates Election Using SVM Algorithm

Michael Alfonso^{1*}, Dionisia Bhisetya Rarasati²

¹Program Studi Teknik Informatika, Fakultas Teknologi dan Desain, Universitas Bunda Mulia

²Program Studi Teknik Informatika, Fakultas Teknologi dan Desain, Universitas Bunda Mulia

Email: ¹s32190039@student.ubm.ac.id, ²l1783@lecturer.ubm.ac.id

Abstract – Elections for presidential candidates are held every 5 years with various candidates, especially on Twitter, arguments about political matters often occur that many Twitter users participate in discussions about the election for presidential candidate. Therefore, this study focuses on sentiment analysis to infer user responses to the presidential election and validate it by looking for a correlation between electability survey results and Twitter sentiment data using Pearson Correlation. In sentiment analysis model, the 10-Fold Cross Validation method is used to find the best model from a dataset with a division of training data and test data with 90:10 split. Then the alphabetic data will be converted into numeric data using the TF-IDF weighting method. To validate the best model, Confusion Matrix is used to get the best f1-score. The model is using Support vector machine algorithm with the Gaussian RBF (Radial Basis Function) kernel. The results of the analysis are compared with the results of the news portal electability survey which contains the 3 candidates using Pearson Correlation. This study produces the best fold for each data on each presidential candidate with the f1-score to find the best model for each fold. In the Pearson Correlation result, the higher positive sentiment of each presidential candidate, the higher electability survey data. For further research, research can be discuss about hyper tuning parameters and using other kernels on Support vector machine algorithm.

Keywords – NLP, Pearson Correlation, Sentiment analysis, SVM, TF-IDF

I. INTRODUCTION

Elections occur every 5 years with the number of voters increasing each period, with an increase of 28.90% from 2004 to 2019 [1]. Social media such as Facebook, Twitter, LinkedIn, YouTube have become part of human life, it is proven that most interactions can be done with social media, until it becomes self-identity to the general public [2]. On social media, especially Twitter, there is a lot of dissemination of information, entertainment, and political opinion. This is not spared from the election which is a political activity that will be held in 2024, referred to as a democratic party, this is certainly a lot of discussion on social media. One way to convey opinions and through social media is to use certain hashtags in tweets posted on Twitter. Twitter itself is used because of the ease in replying to tweets with likes, comments and even replies, so that topics that are currently being discussed can be easily spread [3]. Therefore, further analysis is needed to see how much Twitter users agree and disagree with the 2024 presidential election, then data will be obtained from Twitter user tweets related to this topic, namely the 2024 general election. Sentiment analysis or Opinion Mining is a study that learn about opinions, sentiments, evaluations, behavior and emotions towards an entity [2], [3]. Sentiment analysis is used to find out which tweets are posted containing positive, negative or neutral words [4]. In this study, the Support vector machine (SVM) method will be used, which is one of the classification methods in machine learning (supervised learning) that predicts classes based on patterns from training results. Classification is done by creating a hyperplane between positive and negative classes [2]. The results of this study are the sentiments of Twitter users towards the presidential election in 2024.

II. RESEARCH METHODOLOGY

This research consists of 6 processes in outline, namely data collection, preprocessing, Term Frequency-Inverse Document Frequency weighting, Support Vector Machine training using 10-Fold Cross Validation, Choosing the best fold to be used as a model, and testing as shown below.

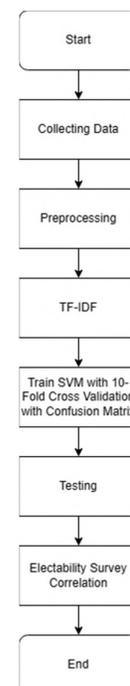


Fig 1. Research Flow

A. Collection Data

The first stage of this research is to scrape data sourced from Twitter with the keyword "(Name of Candidate) President" to narrow down the context of a Twitter user's Tweet. The data collection process is carried out with the help of the snsrape library by python. Data was collected from 1 January 2023 to 31 March 2023. This scraping process resulted in a total of 18,267 data with 7,810 data that were duplicate and irrelevant to the theme "2024 Presidential Candidate Election" resulting in a total of 10,457 clean data from 3 presidential candidates.

B. Preprocessing

The pre-processing stage functions to clean and transform data so that it can be processed by the system easily and accurately. Several stages of pre-processing in this study include:

a. Lowercasing

Lowercasing is a step to change all letters to lowercase so that the word being processed becomes the same form. Table 1 shows an example of the lowercasing process from some of Anies Baswedan's tweets:

Table 1. Lowercasing Result

No.	Before Process	After Process
1	PKS Mendukung Anies Baswedan Menjadi Calon Presiden 2024	pks mendukung anies baswedan menjadi calon presiden 2024
2	Anies Baswedan Presiden Indonesia	anies baswedan presiden indonesia
3	Walau sekarang didukung Demokrat dan PKS, ada empat alasan Anies Baswedan gak bakal bisa menang jadi presiden !!	walau sekarang didukung demokrat dan pks, ada empat alasan anies baswedan gak bakal bisa menang jadi presiden !!
4	Anies Baswedan lh yg pantas jadi presiden	anies baswedan lh yg pantas jadi presiden
5	Pak Anies Baswedan calon presiden 2024	pak anies baswedan calon presiden 2024

b. Remove Special Character and Number

This stage functions to produce sentences without punctuation, emoticons, and numbers. Table 2 shows an example of the Remove Special Character and Number process:

Table 2. Remove Special Character and Number Result

No.	Before Process	After Process
1	pks mendukung anies baswedan menjadi calon presiden 2024	pks mendukung anies baswedan menjadi calon presiden
2	anies baswedan presiden indonesia	anies baswedan presiden indonesia
3	walau sekarang didukung demokrat dan pks, ada empat alasan Anies Baswedan gak bakal bisa menang jadi presiden !!	walau sekarang didukung demokrat dan pks ada empat alasan anies baswedan gak bakal bisa menang jadi presiden
4	anies baswedan lh yg pantas jadi presiden	anies baswedan lh yg pantas jadi presiden
5	pak anies baswedan calon presiden 2024	pak anies baswedan calon presiden

c. Stemming

At this stage, all affixed words will be processed to become basic words, as shown in table 3 below:

Table 3. Stemming Result

No.	Before Process	After Process
1	pks mendukung anies baswedan menjadi calon presiden	pks dukung anies baswedan jadi calon presiden
2	anies baswedan presiden indonesia	anies baswedan presiden indonesia
3	walau sekarang didukung demokrat dan pks ada empat alasan Anies Baswedan gak bakal bisa menang jadi presiden	walau sekarang dukung demokrat dan pks ada empat alasan anies baswedan gak bakal bisa menang jadi presiden
4	anies baswedan lh yg pantas jadi presiden	anies baswedan lh yg pantas jadi presiden
5	pak anies baswedan calon presiden	pak anies baswedan calon presiden

d. Tokenize

This stage break sentences into tokens per word. This process is carried out with the aim that the next process can be applied at the word level. The process is shown in table 4 below:

Table 4. Tokenize Result

No.	Before Process	After Process
1	pks dukung anies baswedan jadi calon presiden	[[pks] [dukung] [anies] [baswedan] [jadi] [calon] [presiden]]
2	anies baswedan presiden indonesia	[[anies] [baswedan] [presiden] [Indonesia]]
3	walau sekarang didukung demokrat dan pks ada empat alasan Anies Baswedan gak bakal bisa menang jadi presiden	[[walau] [sekarang] [dukung] [demokrat] [dan] [pks] [ada] [empat] [alasan] [anies] [baswedan] [gak] [bakal] [bisa] [menang] [jadi] [presiden]]
4	anies baswedan lh yg pantas jadi presiden	[[anies] [baswedan] [lh] [yg] [pantas] [jadi] [presiden]]
5	pak anies baswedan calon presiden	[[pak] [anies] [baswedan] [calon] [presiden]]

e. Remove Stopwords

Stopwords are for removing words that don't have an impact on sentiment, but often appear in sentences. Table 5 shows the Remove Stopwords process as follows:

Table 5. Remove Stopwords Result

No.	Before Process	After Process
1	[[pks] [dukung] [anies] [baswedan] [jadi] [calon] [presiden]]	[[pks] [dukung] [anies] [baswedan] [calon] [presiden]]
2	[[anies] [baswedan] [presiden] [Indonesia]]	[[anies] [baswedan] [presiden] [Indonesia]]
3	[[walau] [sekarang] [dukung] [demokrat] [dan] [pks] [ada] [empat] [alasan] [anies] [baswedan] [gak] [bakal] [bisa] [menang] [jadi] [presiden]]	[[sekarang] [dukung] [demokrat] [pks] [empat] [alasan] [anies] [baswedan] [gak] [bakal] [menang] [presiden]]
4	[[anies] [baswedan] [lh] [yg] [pantas] [jadi] [presiden]]	[[anies] [baswedan] [lh] [pantas] [presiden]]
5	[[pak] [anies] [baswedan] [calon] [presiden]]	[[pak] [anies] [baswedan] [calon] [presiden]]



f. *Slangwords Filtering*

This stage removes slang words into standard words. The slang dictionary is taken from the github repository posted by [louisowen6](#) which is shown in Table 6 as follows:

Table 6. Slangwords Result

No.	Before Process	After Process
1	[[pks] [dukung] [anies] [baswedan] [calon] [presiden]]	[[pks] [dukung] [anies] [baswedan] [calon] [presiden]]
2	[[anies] [baswedan] [presiden] [Indonesia]]	[[anies] [baswedan] [presiden] [Indonesia]]
3	[[sekarang] [dukung] [demokrat] [pks] [empat] [alasan] [anies] [baswedan] [gak] [bakal] [menang] [presiden]]	[[sekarang] [dukung] [demokrat] [pks] [empat] [alasan] [anies] [baswedan] [tidak] [akan] [menang] [presiden]]
4	[[anies] [baswedan] [lh] [pantas] [presiden]]	[[anies] [baswedan] [pantas] [presiden]]
5	[[pak] [anies] [baswedan] [calon] [presiden]]	[[pak] [anies] [baswedan] [calon] [presiden]]

g. *Remove Synonim*

At this stage, the writer removes synonyms in the same sentence if any. Table 7 shows the Remove Synonim process as follows:

Table 7. Remove Synonim Result

No.	Before Process	After Process
1	[[pks] [dukung] [anies] [baswedan] [calon] [presiden]]	[[pks] [dukung] [anies] [baswedan] [calon] [presiden]]
2	[[anies] [baswedan] [presiden] [Indonesia]]	[[anies] [baswedan] [presiden] [Indonesia]]
3	[[sekarang] [dukung] [demokrat] [pks] [empat] [alasan] [anies] [baswedan] [tidak] [akan] [menang] [presiden]]	[[sekarang] [dukung] [demokrat] [pks] [empat] [alasan] [anies] [baswedan] [tidak] [akan] [menang] [presiden]]
4	[[alhamdulillah] [mudah] [mudah] [anies] [baswedan] [takdir] [allah] [swt] [presiden] [aamiin] [yra]	[[alhamdulillah] [mudah] [anies] [baswedan] [takdir] [allah] [swt] [presiden] [aamiin] [yra]
5	[[pak] [anies] [baswedan] [calon] [presiden]]	[[pak] [anies] [baswedan] [calon] [presiden]]

C. *TF-IDF*

TF-IDF weighting is a method of assessing the weight of each word in a document, TF-IDF is a strong feature weighting method that measures the frequency and distribution of words to determine their weight. TF-IDF consists of TF (Term Frequency), which is the frequency of occurrence of words in documents and IDF (Inverse Document Frequency), which is the opposite of the frequency of documents, the TF-IDF value can be obtained by using the following equation:

- a. The value of Term Frequency (TF) is obtained by the value of the frequency of appearance of the t feature in document d[5]

$$TF_t = (t, d) \quad (1)$$

- b. The Inverse Document Frequency (IDF) value is obtained from the logarithm of the number of documents n divided by df documents containing t features[6]

$$IDF_t = \log \frac{n}{df(t)} + 1 \quad (2)$$

- c. The value of the Term Frequency Inverse Document Frequency (TF-IDF) is obtained by multiplying the TF and IDF values[7]

$$W_t = TF_t \cdot IDF_t \quad (3)$$

TF-IDF functions to convert data into numerical form so that the features that have been processed can be loaded into the equation. So this method was chosen to extract the Twitter data.

D. *Support Vector Machine*

The algorithm chosen is the Support vector machine because the Support vector machine is a text mining algorithm that has high accuracy [8]. As seen in this journal, the Support vector machine algorithm has higher accuracy, precision, and recall compared to the Naïve Bayes algorithm, then for journal [9] it is known that the accuracy, precision, and recall of the Support vector machine algorithm has higher accuracy than the K -Nearest Neighbor (K-NN) algorithm and Naïve Bayes. Support Vector Mahine was created by Boser, Guyon, and Vapnik [10] which is a classification method for finding the best hyperplane to separate 2 classes [11]. This algorithm will produce a hyperplane that separates a plane into 2 parts in each class which can be seen in figure 2.

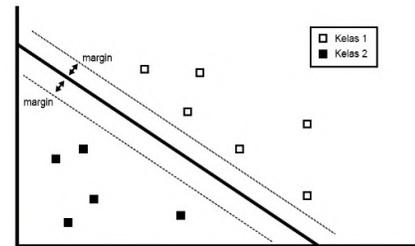


Fig 2. Hyperplane Support Vector Machine Example

Several kernels that can be applied to the SVM algorithm:

- a. Linear Kernel is a kernel function that separates 2 classes linearly with the following equation:

$$[(w^t \cdot x_i) + b] \geq 1 \text{ for } y_i = 1 \text{ dan } [(w^t \cdot x_i) + b] \leq -1 \text{ for } y_i = -1 \quad (4)$$

Description of the following equation is x_i as the train data, followed by i as ammount of data (ex: 1,2, ..., n), then y_i as the label of train data [10].

- b. Polynomial Kernel is a kernel function that is not linearly separated, the equation is as follows:

$$K(x_i, x_j) = (x_i, x_j) + c)^d \quad (5)$$

Description of the following equation is x_i, x_j as the train data, followed by c and d as constant [10].



- c. Radial Basis Function (RBF) The kernel is a function for analyzing data that is not linearly separated and requires 2 types of parameters (Gamma and Cost).

$$K(x_i, x_j) = \exp\left(\frac{-\|x_i - x_j\|^2}{2\sigma^2}\right) \quad (6)$$

Description of the following equation is $\|x_i - x_j\|$ as euclidean distance and 2σ as independent parameter which define subtraction level of RBF toward 0 [10].

To use the Support vector machine algorithm, it is necessary to find the best \vec{w} to separate each class. Normally with data that is not spread out, the hyperplane is either linear or polynomial. But if it comes to the author's case, the data distribution is very wide and the feature dimensions are unlimited, then it is necessary to use a kernel trick, the implementation that the writer has done to overcome this problem is to use a Gaussian RBF (Radial Basis Function) which is able to overcome the problem above.

E. 10-Fold Cross Validation

K-Fold Cross Validation is a form of testing that serves to assess the performance of the algorithm by dividing the sample data randomly and then dividing the data as much as the K value of the total fold. Furthermore, one of the K groups will be used as data testing and the rest of the group will be used as training data [12].

After the SVM machine is trained, the best model is obtained by finding the best f1-score value from the SVM model, 10-Fold Cross Validation method divides a dataset into 10 parts with a data ratio 10:90.

Next, it will be seen from the 10 data divisions, which division has the best f1-score to take the model from the fold

F. Testing

The Confusion Matrix is a table for evaluating classification models, the f1-score value is the value used to determine the best model [13]. So I apply evaluation to research using the Confusion Matrix method. The application of the Confusion Matrix can be seen in the following table:

Table 8. Confusion Matrix Table

	Prediksi Positif	Prediksi Netral	Prediksi Positif
Label Positif	True Positif	False Netral	False Negatif
Label Netral	False Positif	True Netral	False Negatif
Label Negatif	False Positif	False Netral	True Negatif

After that, look for precision, recall, and f1-score values obtained by using the equation:

$$Precision_{(A)} = \frac{TP_A}{TP_A + FP_A} \quad (7)$$

$$Recall_{(A)} = \frac{TP_A}{TP_A + FN_A} \quad (8)$$

$$f1_{(A)} = \frac{2 * Precision_{(A)} * Recall_{(A)}}{Precision_{(A)} + Recall_{(A)}} \quad (9)$$

Description of the following equation is TP_A as the True Positive data, FP_A as the False Positive data, and FN_A as False Negative.

G. Electability Correlation

Pearson Correlation is a method that aims to find a linear relationship between 2 or more variables. The results of the Pearson Correlation are correlation coefficients ranging from numbers 0 to 1 which apply to negative numbers [10]. The Pearson Product Moment Correlation is shown in equation as follows:

$$r_{x,y} = \frac{n \sum xy - (\sum x)(\sum y)}{\sqrt{\{n \sum x^2 - (\sum x)^2\} \{n \sum y^2 - (\sum y)^2\}}} \quad (10)$$

Description of the following equation is n as the amount of data, $\sum xy$ as sum of pair X and Y, $\sum x$ as sum of X variable, $\sum y$ as sum of Y variable, $\sum x^2$ as sum of square X, and $\sum y^2$ as sum of square Y.

III. RESULTS AND DISCUSSION

The implementation of the SVM algorithm is carried out exactly as previously attached, the development of the SVM system begins with the data cleaning stage, then transforms the data into numeric form, then conducts model training by dividing the data 10:90 using the 10-Fold Cross Validation method, and testing the best model by taking f1-score value with the Confusion Matrix method. The best fold is determined based on the f1-score when the False Negative (FN) and False Positive (FP) data are not very close [14]. Also, to find the F1-score value in models with more than 2 labels, I use the macro f1, where the value taken is the average of the f1-score values for all classes, the authors use the macro value because the values will be more even for datasets that are not balanced [15].

Next the results are obtained of the training process by finding the model with the best fold, followed by an analysis of the linkages of the predicted model results using the Pearson Moment Product Correlation. The test results of the 3 presidential candidates are shown in the following table:

Table 9. Best Fold Test Results for Every Presidential Candidate

Candidate Name	Best Fold	F1-score
Anies Baswedan	8	66,36%
Ganjar Pranowo	5	71,72%
Prabowo Subianto	4	77,99%

Then there are the results of electability survey data every month for each candidate in the following table:

Table 10. Electability Survey and Sentiment

Candidate Name	Month	Sentiment		Electability Score
		Positive	Negative	
	January	92,25%	7,75%	15,00%



Candidate Name	Month	Sentiment		Electability Score
		Positive	Negative	
Anies Baswedan	February	92,51%	7,49%	16,95%
	March	94,92%	5,08%	23,50%
Ganjar Pranowo	January	97,98%	2,02%	27,58%
	February	98,40%	1,60%	23,65%
	March	95,79%	4,21%	28,83%
Prabowo Subianto	January	99,36%	0,64%	21,68%
	February	99,92%	0,08%	24,62%
	March	100,00%	0,00%	23,12%

Result of Pearson Product Moment Correlation of data above shown in the following table:

Candidate Name	Correlation	
	Positive	Negative
Anies Baswedan	0,991	-0,991
Ganjar Pranowo	-0,789	0,789
Prabowo Subianto	0,801	-0,801

It can be concluded that there is a relationship between sentiment and news portal electability surveys, from the 3 presidential candidates. The result obtained by the correlation has a strong unidirectional relationship for candidate Anies Baswedan and Prabowo Subianto it means that the higher the positive sentiment, the higher the electability of the candidate. While in candidate Ganjar Pranowo has a strong non-unidirectional relationship it means that that the lower the positive sentiment, the higher the electability of the candidate.

IV. CONCLUSION

Based on the results of the research that has been done, it can be concluded that the implementation of the Sentiment Analysis Engine with the Support Vector Machine algorithm using the Gaussian RBF kernel found the best fold for each presidential candidate, namely the 8th fold with an f1-score of 0.66 for Anies Baswedan's candidate with a total 2,554 training data and 283 testing data, the 5th fold with an f1-score of 0.72 for the Ganjar Pranowo candidate with a total of 3,330 training data and 370 testing data, and the 4th fold with an f1-score of 0.78 for the Prabowo Subianto candidate with a total of 3487 training data and 387 testing data.

Also, the correlation between the electability survey and the results of the positive sentiment analysis of presidential candidate Anies Baswedan is 0.991 and the negative sentiment is -0.991. So that the conclusion obtained by the correlation has a strong relationship in the same direction. Then for the correlation between the electability survey and the results of the analysis of positive sentiment for presidential candidate Ganjar Pranowo is -0.789 and negative sentiment is 0.789. So that the conclusion obtained by the correlation has a strong relationship in the different direction. Furthermore, the correlation between the electability survey and the results of positive sentiment analysis for presidential candidate

Prabowo Subianto is 0.801 and negative sentiment is -0.801. So that the conclusion obtained by the correlation has a strong relationship in the same direction. It can be concluded that there is a link between sentiment and news portal electability surveys.

REFERENCES

- [1] BPS, "Hasil Perhitungan Suara Sah Pemilu Presiden dan Wakil Presiden Menurut Provinsi Tahun 2004, 2009, 2014, 2019." <https://www.bps.go.id/statistictable/2009/03/04/1574/hasil-perhitungan-suara-sah-pemilu-presiden-dan-wakil-presiden-menurut-provinsi-tahun-2004-2009-2014-2019.html> (accessed Jul. 25, 2023).
- [2] B. W. Sari and F. F. Haranto, "IMPLEMENTASI SUPPORT VECTOR MACHINE UNTUK ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP PELAYANAN TELKOM DAN BIZNET," *Jurnal Pilar Nusa Mandiri*, vol. 15, no. 2, pp. 171–176, Sep. 2019, doi: 10.33480/pilar.v15i2.699.
- [3] Fatihah Rahmadayana and Yuliant Sibaroni, "Sentiment Analysis of Work from Home Activity using SVM with Randomized Search Optimization," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 5, pp. 936–942, Oct. 2021, doi: 10.29207/resti.v5i5.3457.
- [4] S. Fendyputra Pratama, R. Andrean, and A. Nugroho, "Analisis Sentimen Twitter Debat Calon Presiden Indonesia Menggunakan Metode Fined-Grained Sentiment Analysis," *JOINTECS (Journal of Information Technology and Computer Science)*, vol. 4, no. 2, pp. 2541–3619, 2019, doi: 10.31328/jo.
- [5] D. W. Seno and A. Wibowo, "Analisis Sentimen Data Twitter Tentang Pasangan Capres-Cawapres Pemilu 2019 Dengan Metode Lexicon Based Dan Support Vector Machine," *Jurnal Ilmiah FIFO*, vol. 11, no. 2, p. 144, Nov. 2019, doi: 10.22441/fifo.2019.v11i2.004.
- [6] D. Darwis, E. Shintya Pratiwi, A. Ferico, and O. Pasaribu, "PENERAPAN ALGORITMA SVM UNTUK ANALISIS SENTIMEN PADA DATA TWITTER KOMISI PEMBERANTASAN KORUPSI REPUBLIK INDONESIA," 2020.
- [7] A. S. Arief, "SENTIMENTANALYSIS REVIEW APLIKASI MENGGUNAKAN ALGORITMA SVM PADA APLIKASI MYPERTAMINA," 2023.
- [8] S. Diantika, W. Gata, H. Nalatissifa, and M. Lase, "Komparasi Algoritma SVM Dan Naive Bayes Untuk Klasifikasi Kestabilan Jaringan Listrik," *JURNAL ILMIAH ELEKTRONIKA DAN KOMPUTER*, vol. 14, no. 1, pp. 10–15, 2021.
- [9] R. Risnantoyo, A. Nugroho, and K. Mandara, "Sentiment Analysis on Corona Virus Pandemic Using Machine Learning Algorithm," *JOURNAL OF INFORMATICS AND TELECOMMUNICATION ENGINEERING*, vol. 4, no. 1, pp. 86–96, Jul. 2020, doi: 10.31289/jite.v4i1.3798.



- [10] D. B. Rarasati and J. C. A. Putra, "Correlation Between Twitter Sentiment Analysis with Three Kernels Using Algorithm Support Vector Machine (SVM) Governor Candidate Electability Level," *COIESE*, pp. 249–256, 2021.
- [11] F. Rahutomo, P. Y. Saputra, and M. A. Fidyawan, "IMPLEMENTASI TWITTER SENTIMENT ANALYSIS UNTUK REVIEW FILM MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE," *Jurnal Informatika Polinema*, vol. 4, no. 2, pp. 93–100, 2018.
- [12] A. Hutapea and M. Tanzil Furqon, "Penerapan Algoritme Modified K-Nearest Neighbour Pada Pengklasifikasian Penyakit Kejiwaan Skizofrenia," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 2, no. 10, pp. 3957–3961, 2018, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [13] F. Istighfarizkya, N. A. S. ER, I. M. Widiartha, L. G. Astutia, I. G. N. A. C. Putra, and I. K. G. Suhartana, "Klasifikasi Jurnal menggunakan Metode KNN dengan Mengimplementasikan Perbandingan Seleksi Fitur," *Jurnal Elektronik Ilmu Komputer Udayana*, vol. 11, pp. 167–176, 2022, [Online]. Available: <https://scholar.google.com>
- [14] F. Satria, Zamhariri, and M. A. Syaripudin, "Prediksi Ketepatan Waktu Lulus Mahasiswa Menggunakan Algoritma C4.5 Pada Fakultas Dakwah Dan Ilmu Komunikasi UIN Raden Intan Lampung," *Jurnal Ilmiah MATRIK*, vol. 22, pp. 28–35, 2020.
- [15] O. H. Anidjar, A. Barak, B. Ben-Moshe, E. Hagai, and S. Tuvyahu, "A Stethoscope for Drones: Transformers Based Methods for UAVs Acoustic Anomaly Detection," *IEEE Access*, 2023, doi: 10.1109/ACCESS.2023.3262702.



DevOps, Continuous Integration and Continuous Deployment Methods for Software Deployment Automation

Mochamad Hanif Rifa'i Istifarulah^{1*}, Rizka Tiaharyadini²

¹Program Studi Magister Ilmu Komputer, Fakultas Teknologi Informasi, Universitas Budi Luhur

²Jl. Raya Ciledug, Petukangan Utara, Kebayoran Lama, Jakarta Selatan 12260

Email: ¹istifarulah@gmail.com, ²rizka.tiaharyadini@budiluhur.ac.id

Abstract - In the fast-paced landscape of software development, the need for efficient, reliable, and rapid deployment processes has become paramount. Manual deployment processes often lead to inefficiencies, errors, and delays, impacting the overall agility and reliability of software delivery. DevOps, as a cultural and collaborative approach, plays a central role in orchestrating the synergy between development and operations teams, fostering a shared responsibility for the entire software delivery lifecycle. Continuous Integration is a fundamental DevOps practice that involves regularly integrating code changes into a shared repository, triggering automated builds and tests. Continuous Deployment complements Continuous Integration by automating the release and deployment of validated code changes into production environments. The purpose of this research is to create a software deployment automation system to make it easier and reliable for organizations to deploy software. In conclusion, the results of this research show that by adopting DevOps, Continuous Integration, and Continuous Deployment, organizations can achieve enhanced collaboration, shortened release cycles, increased deployment frequency, consistent deployment, and improved overall software quality.

Keywords: *DevOps, Continuous Integration, Continuous Deployment, Software Development, CI/CD*

I. INTRODUCTION

PT EOA Teknologi Internasional with the branding name EOA Tech is an entity of the EOA Group whose business is mostly as a gold producer with the gold brand issued being EOA Gold. EOA itself is an abbreviation of Emas Optimasi Abadi. PT EOA Teknologi Internasional was founded in July 2020 as a software development company engaged in software development or Independent Software Vendor (ISV). PT EOA Teknologi Internasional has been heavily involved in helping clients solve problems with software creation solutions and providing services for hosting these applications on the company's servers. For cloud service providers, companies use services from Alibaba Cloud.

With the increasingly rapid development of software development and business demands that require feature releases and software improvements to be carried out quickly accompanied by increasing frequency, consistent and automated methods for software integration and deployment are needed to minimize human error. To support this, at the organizational level, the Product Development and Operation divisions at PT EOA Teknologi Internasional are required to work together and be able to coordinate and collaborate effectively. However, work bottlenecks often occur, communication problems and a lack of synergy between the Product Development and Operations divisions which cause feature releases and software improvements to be

hampered and take longer. There are other problems during deployment, namely inconsistent deployment results between the development environment, staging environment and production environments, in the development environment all features work well and have been checked by Quality Assurance, but when deployed to the production environment the results of testing in the development environment which show all features are running well are often different when deployed in the production environment, there are several features that do not work well in the production environment, this causes inconsistencies in software deployment. Another problem that causes delays in software release or deployment is that making changes to deployed software will often require both testing and planning, as well as coordination between the different involved departments [1], the deployment process carried out by the Operations division is required to be processed by the relevant team manually, it is indeed a time consuming task to develop, assure quality of and it is not desirable to add even more costly overhead to this process [2]. Apart from that, there are cases where the software deployment process must be carried out at night so it will not cause disturbance of applications that are running and used by users during the day or during working hours. Because software deployment requires intervention from the Operations division, the Product Development division often had difficulties when



deploying software at night due to the availability of the Operations division.

DevOps combines a number of approaches that bring together developers and operations staff to create software and services rapidly, reliably, and of higher quality [3]. Based on this background, the research objective is to apply DevOps, Continuous Integration and Continuous Deployment methods to automate software deployment, make it easier for companies to deploy software, avoid work bottlenecks between the Product Development and Operations divisions and minimize human error.

II. RESEARCH METHODOLOGY

2.1. Software Development Life Cycle (SDLC)

The systems development life cycle (SDLC) is a methodology for designing, building, developing and maintaining information and process systems [4]

By applying existing SDLC rules, you can provide an understanding of how an Information System (IS) can support business needs, design and build a system according to user needs [5]. The SDLC project is a directive project, namely organized and planned. SDLC project standards require release phases and in each phase there will be a build and deploy process. The deploy stage indicates that code development is complete. Then the code will be built and deployed in the Testing environment. If a bug appears in the Testing environment, the code must be fixed and then rebuilt and used in the Pre-production or Production environment.

Agile methodologies are the standard practices for modern-day industries, Agile software development has a big focus on collaboration and the self-organizing team [6]. Topics that are the center of attention of Agile include its ability to reduce costs, increase speed and quality, and provide motivation to empower employees to support the development and success of a company. Scrum is the most popular of the Agile frameworks [7]. In the Scrum framework there are Sprints and teams that will work in them. After the Sprint is complete, the Demo is complete and the Definition of Done (DoD) is met according to the user story, then each team will integrate their code and build for use in a staging or test environment. Configuration management tools will be used to integrate the code. This is where the real problem arises. Fixing this issue will take a lot of time and cause the release time to be off schedule. Problems in the build and deploy phases are very common and occur frequently in Agile work environments. This also happens in application development at PT EOA Teknologi Internasional. DevOps is expected to be a solution to this.

Agile



DevOps



Fig 1. Differences in Design, Coding, Testing and Deploy Iteration in Agile and DevOps

2.2. DevOps

DevOps (development and operations) is a conceptual study of the development and delivery of software to infrastructure by taking a collaborative and integrative approach between developers (Dev) and software operations (Ops) [8]. DevOps is an organizational approach with the aim of creating collaboration, interaction and empathy across functions and divisions [9]. The adoption of DevOps is one example of a step to strengthen collaboration between IT teams, which is very much needed in software development and maintenance [10]

The DevOps method is proven to be able to reduce several development stages that existed in the old method [11]. DevOps is able to shorten the time between software development and operation without reducing the quality of the software itself [12]. DevOps reduces the gap between the development team, operations team and application users allowing to detect problems early. DevOps adoption can implement continuous development and more frequent application releases to users [13]. Not only that, DevOps also changes software engineering processes and practices to be faster which also increases the reliability, stability, resilience and security of the production environment [14].

2.3. Continuous Integration (CI)

[15] Continuous Integration is a method in software development, where code from each developer is combined and built periodically to detect errors as early as possible. In general, the stages in a CI system are as follows (triggered by push or commit commands):

1. Integrated
This is the stage where each developer integrates the codes or results of their work.
2. Compiled
Code is compiled into packages or executables
3. Tested
Test executables manually or automatically
4. Archived
Archive executable files, test results, and logs during the process.
5. Deployed
Install or share build results.



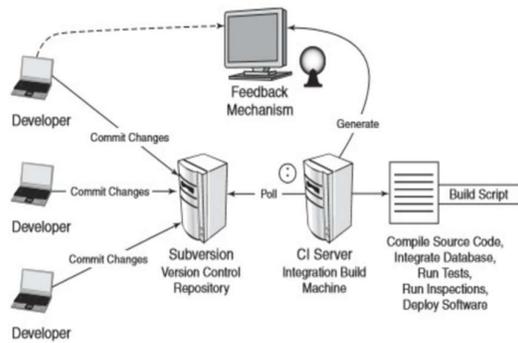


Fig 2. Continuous Integration Architecture System

2.4. Continuous Deployment (CD)

[16] Continuous Deployment (CD) is a software release process that uses automated testing to validate whether changes to the code base are correct and stable for immediate autonomous deployment to a production environment.

The difference between continuous deployment and continuous delivery can be confusing due to the nomenclature. Both are abbreviated as CD and have very similar responsibilities. Delivery is the beginning of deployment. In delivery, there is a final manual approval step before production release.

In the delivery stage, developers will review and merge code changes which are then packaged into artifacts. This package is then moved to the production environment where it awaits approval to be opened for deployment. In the deployment phase, packages are opened and reviewed with an automated inspection system. If the check fails the package is rejected. When the checks pass, the package is automatically deployed to production.

Continuous Deployment can be a powerful tool for modern organizations. Deployment is the final step in the entire continuous pipeline consisting of integration, delivery and deployment. The true experience of continuous deployment is automation to the level where code is deployed to production, tested, and automatically reverted when there are errors, or accepted if there are no errors.

2.5. Docker Container

[17] A Docker container image is a lightweight, standalone, executable package of software that includes everything needed to run an application (code, runtime, system tools, system libraries and settings).

Containers separate software from its surroundings, for example the difference between development and staging, and are also a solution to

conflicts between teams using different software on the same infrastructure.

2.6. Docker Swarm

Docker Swarm is an orchestration tool developed by Docker itself. Docker Swarm clusters allow one to add an unlimited number of nodes, and Docker allows one to run unlimited containers on nodes, this provides full-scale testing that is close to real conditions [18].

Docker Swarm provides clustering and an orchestration mechanism and thus can deploy several containers across different host machines. Docker Swarm also provides a fault tolerance mechanism not only by detecting failed containers on a host machine, but also redeploying the same container on another host machine [19].

2.7. Gitlab

Gitlab is a web-based Git, repository manager with a wiki, issue tracker and CI/CD pipeline. Gitlab uses an open source license developed by Gitlab Inc. In the release of Gitlab 10.0, Gitlab took a big step beyond just code management, but expanded into deployment and monitoring. Gitlab organizes and modifies people's permissions according to their roles and can provide issue tracking access without granting permissions to large pieces of code, which is great for teams and large companies with role-based contributions. Gitlab supports CI for free and is very useful for teams, besides that Gitlab also supports CI/CD automatically without human intervention [20]

III. RESULTS AND DISCUSSION

The software deployment automation system for the Product Development division and Operation division of PT EOA Teknologi Internasional consists of Gitlab as version control for developers to store source code, Continuous Integration (CI) Server and Continuous Deployment (CD) Server using the infrastructure provided by Gitlab, then Container Registry as a place to accommodate the Docker Image output from the Continuous Integration (CI) Server using services from Alibaba Cloud, and Elastic Compute Service (ECS) is the destination server for Continuous Deployment (CD). The server carries out software deployment after pulling the Docker Image from the Container Registry. Continuous Integration (CI) Server and Continuous Deployment (CD) Server when carrying out the integration or deployment process will provide output in the form of a running process log which can be seen on the account dashboard on the Gitlab site and provide notification via email if the Continuous Integration process or Continuous Deployment process has failed.



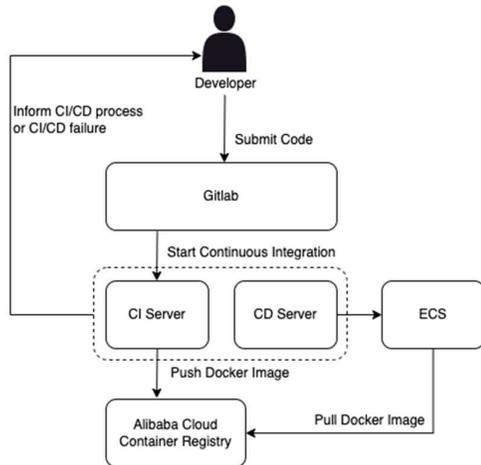


Fig 3. Software Deployment Automation System Design

Before the software deployment automation process can be carried out, the ECS for the software deployment destination must have Docker Engine and Docker Swarm installed, as well as the reverse proxy which in this study uses Traefik, must be configured and ready to use.

The entire Continuous Integration and Continuous Deployment process refers to scripts that have been prepared by the Operations division and have been submitted to the repository on Gitlab or version control. These scripts are in the form of .gitlab-ci.yml to manage the Continuous Integration and Continuous Deployment flow on the Gitlab CI/CD infrastructure. Dockerfile, stack.yml, and other supporting scripts required by the programming language in each application or repository.

The prototype of this software deployment automation system was built using the YAML (Yet Another Markup Language) programming language which is usually used to write configuration files. YAML is a popular programming language because it is human-readable and easy to understand. Meanwhile, the application used for software deployment automation testing uses the PHP programming language and the Laravel framework, because PHP and Laravel are popular programming languages and frameworks, that are widely used by developers.

The display and snippet of the script from the software deployment automation system are as follows:

1. Software Deployment Automation System Prototype File Structure

The following is the file structure of the software deployment automation system used to deploy software in the PHP programming language using the Laravel framework.

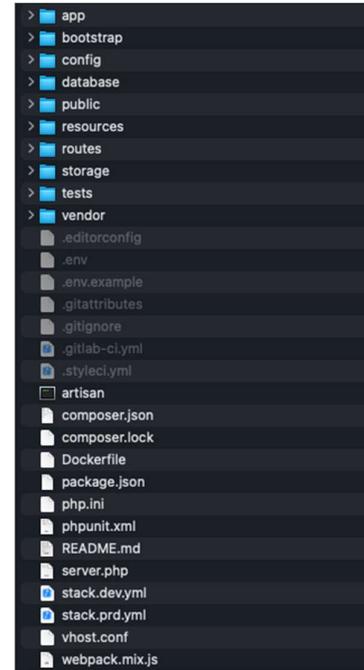


Fig 4 Software Deployment Automation System File Structure in PHP Laravel

2. Submit Code using Git CLI (Command Line Interface)

In the following submit code, there are several commands that are executed on the macOS terminal, namely looking at what branches are available in the repository, then pushing it to the selected branch.

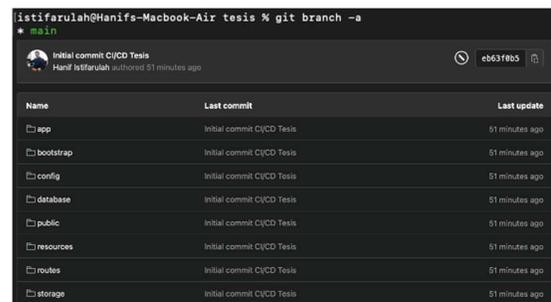


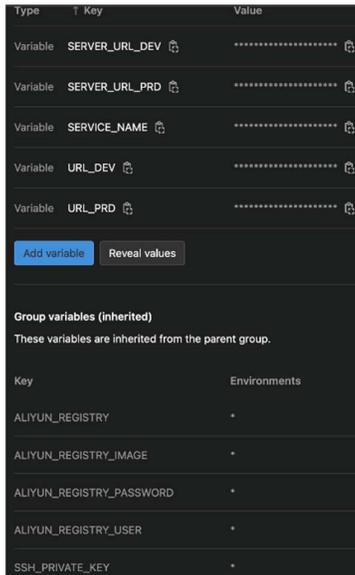
Fig 5, Submit Code Using Git CLI



Fig 6. Result of Git Push

3. CI/CD Variable in Gitlab

In the following Gitlab CI/CD variable, these are variables that will be used in the software deployment automation process. The contents of these variables cannot be seen by developers or can only be seen by



developers who have maintainer access or highest access to the repository. With this, system confidentiality is maintained even though many developers leave the repository due to resigning or moving to other projects.



Fig 7. CI/CD Variable in Gitlab

4. Runner View of the Continuous Integration Process

The following is the runner display when the continuous integration process is running, Gitlab provides real time logs regarding the running process, and if a failure occurs, the relevant developer will automatically receive a notification email.



Fig 8. Continuous Integration Runner

5. Runner View in Continuous Deployment Process

The following is a runner display when the continuous deployment process is running. Gitlab provides real time logs regarding the running process, and if a failure occurs, the relevant developer will automatically receive a notification email.

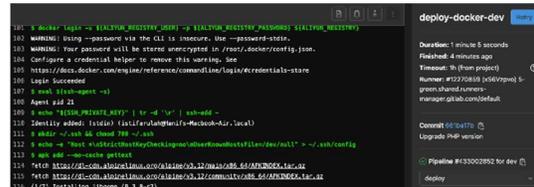


Fig 9. Continuous Deployment Runner

6. Application Display After Successful Deployment Process

The following is the appearance of the application in a web browser, which is used for testing the software deployment automation system after successful software deployment. This application can be accessed by users using a web browser.



Fig 10. Application Display After the CI/CD Process is Successful

7. Snippet of .gitlab-ci.yml Script

The following is a snippet from the .gitlab-ci.yml script, to set up Continuous Integration and Continuous Deployment flows on Gitlab CI/CD infrastructure.

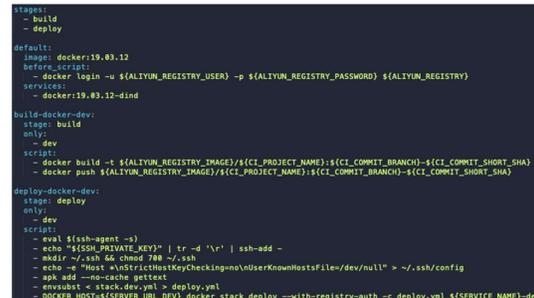


Fig 11 Snippet of .gitlab-ci.yml Script

8. Snippet of Dockerfile Script

The following is a snippet from the Dockerfile script, as a basic for building the output in the form of a Docker Image.

```
FROM composer:2.1 AS builder
WORKDIR /app
COPY composer.json ./
COPY composer.lock ./
RUN composer install \
    --ignore-platform-reqs \
    --no-interaction \
    --no-plugins \
    --no-scripts \
    --prefer-dist

FROM php:8.1.1-apache
RUN apt-get update -y && apt-get install -y \
    cron \
    libzip-dev \
    zip \
    && docker-php-ext-install -j$(nproc) \
    bcmath \
    opcache \
    zip
RUN a2enmod rewrite
COPY vhost.conf /etc/apache2/sites-available/000-default.conf
COPY php.ini ${PHP_INI_DIR}/php.ini
COPY --from=builder /app/vendor /var/www/html/vendor/
COPY . /var/www/html/
CMD apache2-foreground
```

Fig12. Snippet of Dockerfile Script

9. Snippet of stack.yml Script

The following is a snippet from the stack.yml script, as a basic for configuring a reverse proxy so that applications that have been deployed can be accessed by users. This script also arranges for copying environment files which contain secrets related to database access or other secrets for the application which is deployed. This script also regulates which Docker Image must be pulled for use in the software deployment process.

Figure 13 Snippet of stack.yml Script



IV. CONCLUSION

Software deployment that has been carried out using a software deployment automation system makes the results of software deployment in the development environment, staging environment and production environment consistent.

Because the software deployment process has been automated, there is no longer a work bottleneck in the Operations division because the software deployment process no longer requires human intervention after the software deployment automation system is configured at the start of creating the repository/project.

By building a software deployment automation system, human errors caused by the software deployment process being carried out manually can be minimized.

With the software deployment automation system, the software deployment process can be carried out at any time by the Product Development division, without having to ask the Operations division to stand by and intervene. This leads to shortened release cycle, increased deployment frequency thus improve software quality.

Suggestions from the author for the development of this system in the future:

1. In the future, this software deployment automation system can be developed using Kubernetes container orchestration with auto scaling.
2. Implement system monitoring using Grafana, Prometheus and other supporting applications.
3. This software deployment automation system can be developed to create a High Availability System.
4. There needs to be good cooperation between the Product Development division and the Operations division regarding the use of DevOps, Continuous Integration and Continuous Deployment methods.

REFERENCES

- [1] Bibi, S., Katsaros, D., & Bozaris, P. (2012). Business Application Acquisition: On-Premise or SaaS-Based Solutions? *IEEE Software*, 29, 86–93. <https://doi.org/10.1109/MS.2011.119>
- [2] Touma, Y. (2019). *An investigation of Automating Software Deployment Using Continuous Delivery Tools : A cost-benefit study in the case of multiple system instances.* <https://api.semanticscholar.org/CorpusID:196201845>
- [3] Battina, D. S. (2021). The Challenges and Mitigation Strategies of Using DevOps during Software Development. *International Journal of Creative Research Thoughts (IJCRT)*, 9(1), 4760–4765.
- [4] Alhamidi. (2017). Membangun Sistem Aplikasi untuk Seleksi Calon Mahasiswa Undangan pada Tingkat Sekolah Menengah Atas. *Jurnal J-Click; Vol 3 No 2 (2016): J-Click.* <http://ejournal.jayanusa.ac.id/index.php/J-Click/article/view/26>
- [5] Syamsiyah, N., & Sesunan, M. F. (2018). Penerapan Metode System Life Cycle Development Dan Project Management Body of Knowledge Pada Pengembangan Sistem. *Ikraith-Informatika*, 2(2).
- [6] Trivedi, D. (2021). Agile Methodologies. *International Journal of Computer Science & Communication*, 12(2), 91–100.
- [7] Naik, N., & Jenkins, P. (2019). Relax, it's a game: Utilising gamification in learning agile scrum software development. *IEEE Conference on Computational Intelligence and Games, CIG, 2019-Augus.* <https://doi.org/10.1109/CIG.2019.8848104>
- [8] Jha, P., & Khan, R. (2018). A Review Paper on DevOps: Beginning and More To Know. *International Journal of Computer Applications*, 180(48), 16–20. <https://doi.org/10.5120/ijca2018917253>
- [9] Dyck, A., Penners, R., & Lichter, H. (2015). *Towards Definitions for Release Engineering and DevOps.* <https://doi.org/10.1109/RELENG.2015.10>



- [10] CaTechnology. (2013). *TechInsights Report : What Smart Businesses Know About DevOps. September*, 300.
- [11] Taryana, A., Fadli, A., & Nurshiami, S. R. (2020). Merancang Perangkat Lunak Sistem Penjaminan Mutu Internal (SPMI) Perguruan Tinggi yang Memiliki Daya Adaptasi Terhadap Perubahan Kebutuhan Pengguna secara Cepat dan Sering. *Jurnal Al-Azhar Indonesia Seri Sains Dan Teknologi*, 5(3), 121. <https://doi.org/10.36722/sst.v5i3.372>
- [12] Erich, F., Amrit, C., & Daneva, M. (2017). A Qualitative Study of DevOps Usage in Practice. *Journal of Software: Evolution and Process*, 00. <https://doi.org/10.1002/smr.1885>
- [13] Erich, F., Amrit, C., & Daneva, M. (2014). Report: DevOps Literature Review. https://www.researchgate.net/publication/267330992_Report_DevOps_Literature_Review, October, 1–27. <https://doi.org/10.13140/2.1.5125.1201>
- [14] Mohamed, S. I. (2015). DevOps Shifting Software Engineering Strategy Value Based Perspective. *IOSR Journal of Computer Engineering Ver. IV*, 17(2), 2278–2661. <https://doi.org/10.9790/0661-17245157>
- [15] Prieria, J. M., & Ganefi, R. T. (2017). Automatic Deployment System Dengan Menggunakan Metode Continuous Integration Di Kakatu. *Jurnal Ilmiah Komputer Dan Informatika*.
- [16] Pittet, S. (2021). *Continuous Deployment*. Atlassian. <https://www.atlassian.com/continuous-delivery/continuous-deployment>
- [17] Docker. (2018). *What is a Container?* <https://www.docker.com/resources/what-container>
- [18] Shichkina, Y. A., Kupriyanov, M. S., & Moldachev, S. O. (2018). Application of Docker Swarm cluster for testing programs, developed for system of devices within paradigm of Internet of things. *Journal of Physics: Conference Series*, 1015, 32129. <https://doi.org/10.1088/1742-6596/1015/3/032129>
- [19] Ismail, B. I., Goortani, E. M., Karim, M. B. A., Tat, W. M., Setapa, S., Luke, J. Y., & Hoe, O. H. (2015). Evaluation of Docker as Edge computing platform. *2015 IEEE Conference on Open Systems (ICOS)*, 130–135. <https://doi.org/10.1109/ICOS.2015.7377291>
- [20] Peham, T. (2017). *GitLab vs GitHub: What are the key differences? The Ultimate Guide*. <https://usersnap.com/blog/gitlab-github/>



Bayah Dome Geopark Information System Based On Website Using Prototype Method

Dentik Karyaningsih^{1*)}, Donny Fernando², Reza Pramudita³, Fauzan Wahyu Ramadhan⁴

^{1,3,4}Informatics Engineering Study Program, Faculty of Information Technology, Serang Raya University

²Information Systems Study Program, Faculty of Information Technology, Serang Raya University

Email: ¹karya.tiek@gmail.com, ²mr.donny2008@gmail.com, ³hanyqueen14@gmail.com,

⁴fauzanwahyu785@gmail.com

Abstract – To realize a national Geopark and a World Geopark, it is necessary to socialize and disseminate information that can be accessed by the people of Lebak Regency in particular and the wider community in general to introduce all the geological heritage that has been determined by the Minister of Energy and Mineral Resources. Nowadays, digital information system technology has become an undeniable need, and has even become a source of information that can be accessed by all people in the world. The Bayah Dome Geopark management body in Lebak Regency does not yet have a website-based information system, so it is necessary to create a website-based Bayah Dome Geopark information system so that the Bayah Dome Geopark is known nationally and internationally. This research uses a Research and Development approach. Research procedures include: (a) analysis, (b) design, (c) implementation and (d) evaluation. Stage (a) analysis consists of two steps, namely literature study and field survey. The literature study examines concepts and theories relating to Geoparks and the Bayah Dome Geopark, while a field survey was carried out to collect data relating to the Bayah Dome Geopark. Stage (b) design is carried out by designing a website with UML (Unified Modeling Language) which is a program design based on the results obtained in the analysis stage which includes: use case diagrams, activity diagrams using the prototype system development method. In stage (c) implementation, the results of the design are outlined and implemented into a website.

Keywords – *Information Systems, Bayah Dome Geopark, Website, Prototype*

I. INTRODUCTION

Based on Presidential Regulation Number 9 of 2019 concerning the Development of a Geopark, it is a single or combined geographical area that has geological heritage sites (Geosites) and valuable natural landscapes, related to aspects of Geological Heritage, Geological Diversity, Biodiversity and Cultural Diversity, then every region throughout Indonesia that has geological heritage and natural landscapes along with other diversity must be designated as a Geopark. Geoparks which have aspects in the field of education as knowledge in the field of earth sciences regarding the uniqueness and diversity of earth heritage and economic aspects of the role of society in managing the area as geotourism.(Darsiharjo et al., 2016)Geoparks are managed for conventional purposes, education and sustainable community economic development with active involvement from the community and local government, so that they can be used to foster community understanding and concern for the earth and its surrounding environment. A Geopark is a single or combined geographic area, which has Geological Heritage Sites (Geosites) and valuable natural landscapes, related to aspects of Geological Heritage (Geoheritage), Geological Diversity (Geodiversity), Biodiversity (Biodiversity) and Cultural Diversity (Cultural Diversity), and managed for conventional purposes, education and sustainable community economic development with active involvement from the community and regional government,(Haryanto & Pirgana, 2021)

Lebak Regency has a lot of geological heritage as well as biodiversity and culture which has the potential to be developed into a national or even international class Geopark. Geoparks are a tool for sustainable development for people's welfare, which combines three diversities,

namely geological, biological and cultural diversity so that with the Minister of Energy and Mineral Resources Regulation Number 1 of 2020 concerning Guidelines for Determining Geological Heritage (Geoheritage), Lebak Regency, Banten Province has Geological Heritage (Geoheritage) that meets the criteria to be protected, preserved and utilized as an object of research, earth education and geotourism.

The Bayah Dome Geopark is still an aspiring Geopark towards becoming a national Geopark. It is necessary to introduce the Bayah Dome Geopark to the wider community with information system technology. Bayah Dome Geopark is geotourism so the information system technology currently developing is expected to help disseminate information and promote geotourism widely both nationally and internationally. Not many people around and the wider community are familiar with the Bayah Dome Geopark, therefore there is a need for outreach or disseminating information to the community by utilizing an information system that can be accessed anywhere and at any time. The natural landscape of the Bayah area consists of hills and valleys, very interesting and unique. The whole thing forms an area known as the dome and is called the Bayah Dome. Three large rivers flow in this dome area, namely Ci Bareno on the east wing. Ci Madur is in the middle, and Ci Peucangeuri is in the western part of the dome. From the center to the south coast, this dome stretches the karst area.(Permadi, 2015)Therefore, it is necessary to build a Website-based Bayah Dome Geopark information system that can be accessed online. Currently, technology is widely used as a means of promotion and information, especially in the field of websites which are currently an information medium that offers various conveniences in presenting information. Speed and convenience are positive values of the internet.



Websites are more easily accessed by people in various regions just by using the internet. (Ismail, 2018) Geotourism destinations should be equipped with a clear and easy to understand information system. With a good information system, for example about the history of the formation of geological formations, it is hoped that tourists will understand the natural processes that occur. With information, people are also expected to be aware not to damage the beauty of the environment around geotourism. (Hermawan, 2018) The Bayah Dome Geopark management body in Lebak Regency does not yet have a website-based information system, so it is necessary to create a website-based information system for the Bayah Dome Geopark so that the Bayah Dome Geopark is known nationally and internationally.

This is also supported by the Strategic Plan and Research Master Plan of Serang Raya University which carries the theme "Development of Science and Technology, Economy and Culture of a Competitive Nation towards a Smart City" that the industrial era 4.0 encourages the use of the Internet of Things (IoT), the use of big data (big data), and quality engineering. It is hoped that Lebak Regency will become a smart city area in particular and the Banten Province region in general.

II. RESEARCH METHODOLOGY

This research uses a research and development (R&D) approach. Software development in the form of an Online Website Based Information System through an engineering approach with stages of analysis, design, implementation and evaluation. The following are the research stages:



Fig 1. Research stages

The analysis stage is the first step in development, this stage consists of two steps, namely literature study, interviews and observations at the Bayah Dome Geopark Management Agency, Lebak Banten Regency. The literature study examines concepts and theories relating to the Information System and the Bayah Dome Geopark, while a field survey was carried out to collect data relating to the Bayah Dome Geopark data as well as information from the Bayah Dome Geopark Management Agency which is the object of the Information System.

The design stage is carried out with a program design based on the results obtained in the analysis stage, program design using UML (Unified Modeling Language). UML is a modeling language for systems or software with an 'object oriented' paradigm. Consists of: Class Diagram, Object Diagram, Use case diagram, Sequence diagram, Collaboration diagram, Statechart diagram, Activity diagram, Component diagram and Deployment diagram. (Wati & Kusumo, 2016) The program design in this research is based on the results obtained in the analysis stage which include: use case diagrams, activity diagrams, class diagrams. This website-based information system was

built using the PHP, Javascript, CSS programming languages. Laravel Bootstrap Framework and MySQL database.

The system development method used in this research is the Prototype method. The Prototype method is a software development method, which is in the form of a physical model of the system's work and functions as an initial version of the system. (GILANG RAMDHANI PUTRA, 2020) The prototyping model focuses on presenting the aspects of the software that will be visible to the user. The stages of this prototyping development model are depicted in Figure 3 below.

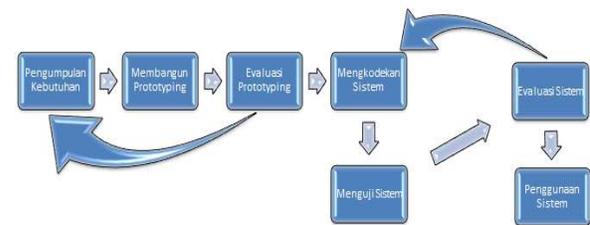


Fig 2. Prototyping Development Model

In the implementation stage, the results of the design are expressed through a program, while the material components are implemented with content in the Bayah Dome Geopark Information System. then the system testing stage uses the System Usability Scale (SUS). Usability testing can be carried out on small or large software and on multiplatforms such as web-based, desktop or mobile. The focus of usability testing is that users find it easy to use the software. (Sembodo et al., 2021) The results of the program evaluation are used as material to carry out revisions both in terms of design and implementation.

III. RESULTS AND DISCUSSION

The results of this research include:

A. System Design

1. Use Case Diagrams

Use case diagram is a model to describe the relationships that occur between actors and activities contained in the system. A use case is a modeling of the behavior of the information system that will be created. A use case describes an interaction between one or more actors and the information system to be created. (Rahma et al., 2019) The use case diagram for this application can be seen in the following image.

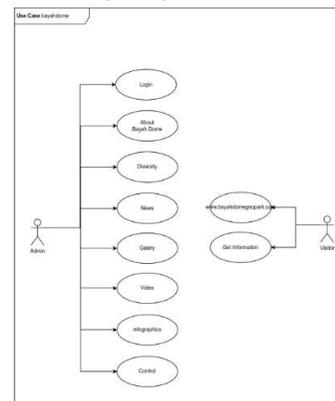


Fig 3. Use Case Diagram



In the application Use Case Diagram there are 2 actors, namely admin and visitor. Visitors have the role of receiving information from the website www.geoparkbayahdome.com. Meanwhile, the admin is the one who manages the entire contents of the website.

2. Activity Diagram

a. Visitor Activity Diagram

User activity diagrams is a diagram that describes the work flow or visitor activities of a system. From the use case in Figure 4, the Visitor Activity Diagram will be explained, including:

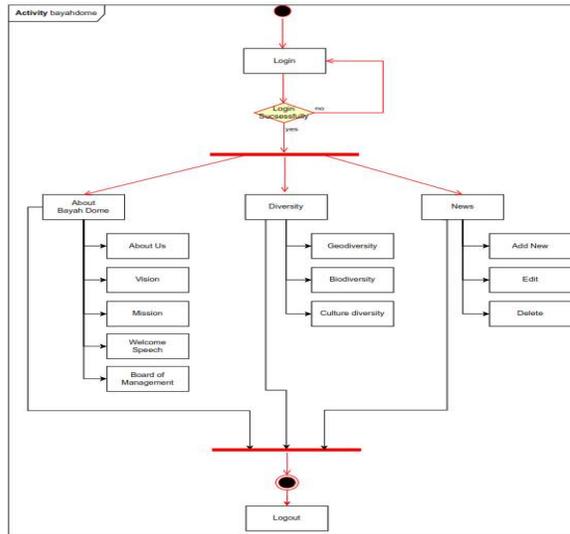


Fig 4. Visitor Activity Diagram

Activity diagram above is user activity using the Bayah Dome Geopark Information System. The system will display all information about the Bayah Dome Geopark including news, Geodiversity, Cultural Diversity, Biodiversity, Management Agency, photos and videos, and others.

b. Activity Diagram admin

Activity Diagrams describes the various activity flows in a system that is being designed. (Kirana & Wahdaniyah, 2018) An admin activity diagram is a diagram that describes the work flow or admin activities of a system. From the use case in Figure 5, the admin Activity Diagram will be explained, including:

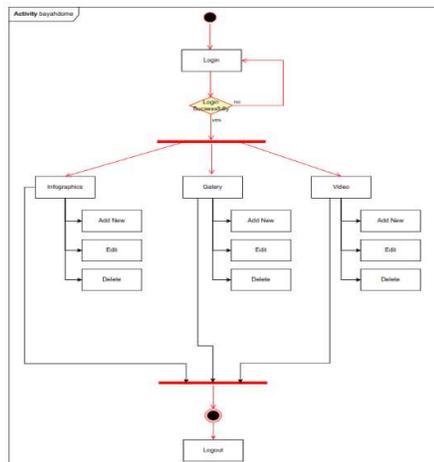


Fig 5. Activity Diagram admin

The activity diagram above is admin activity in managing the Bayah Dome Geopark Information System. Admin manages all contents of the Information System about the Bayah Dome Geopark including news, Geodiversity, Cultural Diversity, Biodiversity, Management Agency, photos and videos, and others by adding, editing and deleting.

c. Class Diagrams

The next system design plan is a class diagram. A class diagram is a diagram used to describe the objects in the system and the relationships between these objects. Apart from that, class diagrams function as descriptions of the types of objects that exist in the system and the various types of static relationships that occur. (Karyaningsih & Siswanto, 2020) The following is a class diagram design for the application being built. The class diagram contains 14 classes including user, setting, about, vision and mission, opening remarks, board, slides, geodiversity, cultural diversity, biodiversity, infographics, news, gallery and video.

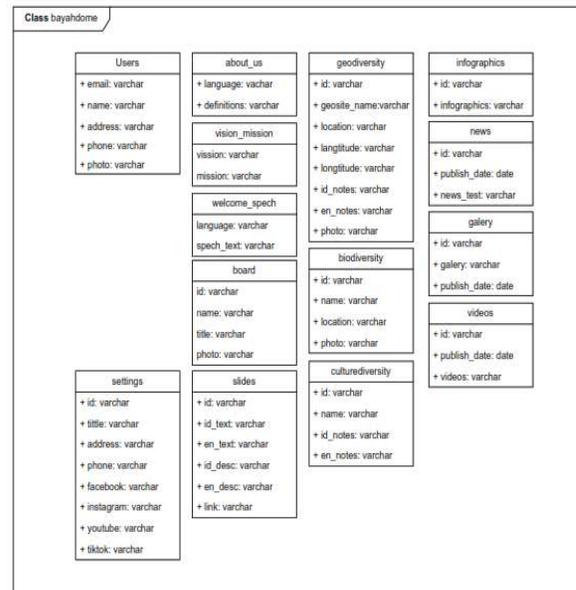


Fig 6. Class Diagram

B. Bayah Dome Geopark Information System Website

1. Main page

Website main page is the page that first appears when opened in a web browser by opening www.geoparkbayahdome.com amely displaying menu information on the website as well as displaying remarks from the Director of the Bayah Dome Geopark Management Agency, news, as well as infographics on website visits, sponsorships and connected applications.



Fig 7. Main page

2. Menu About Us

This menu is a menu page to explain about the Geopark and the Bayah Dome Geopark Management Agency, including about the geopark, management and vision and mission. By pressing/clicking the About Us menu section.



Fig 8. Menu About Us

3. Geodiversity Menu

This page is a menu page that displays Geodiversity information contained in the Bayah Dome Geopark. Complete with maps and Geosite information for 32 Geosites.

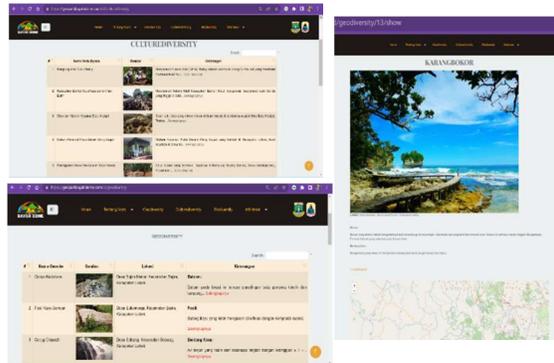


Fig 9. Geodiversity menu

4. Cultural Diversity Menu

This page is a menu page that displays information on Cultural Diversity found in the Bayah Dome Geopark.



Fig 10. Cultural Diversity Menu

5. Biodiversity Menu

This page is a menu page that displays Biodiversity information contained in the Bayah Dome Geopark.

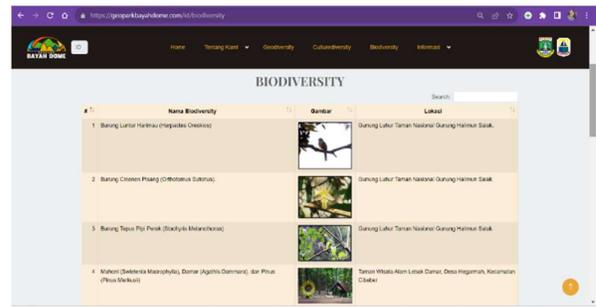


Fig 11. Biodiversity menu

6. Information Menu

The Information Menu is a page that displays information from the Bayah Dome Geopark Management Agency including news, infographics, videos and photo galleries.

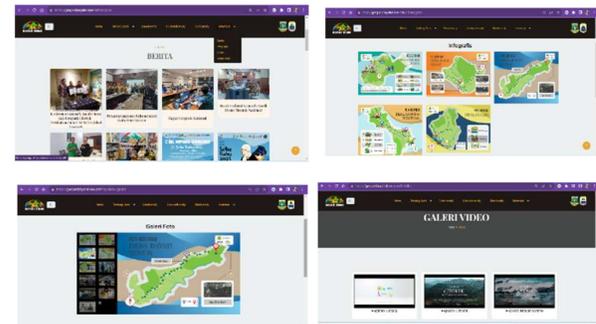


Fig 12. Information menu

7. Admin Page

The page on the admin website is for managing all contents of the Bayah Dome Geopark Information System website which is website-based with website pages www.geoparkbayahdome.com. By logging in admin via the login option with an account/email that is registered as admin/manager.

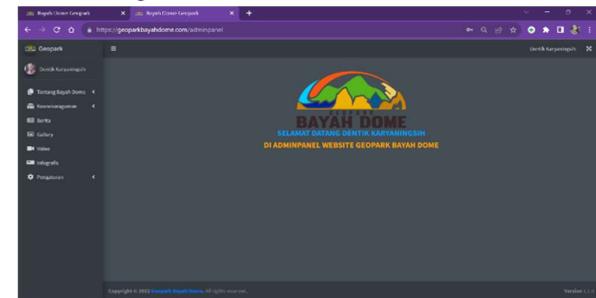


Fig 13. Website-based admin page

C. System Testing

Testing on the Geopark Bayah Dome Information System website at www.geoparkbayahdome.com uses the System Usability Scale (SUS) testing system. The System Usability Scale (SUS) is a questionnaire that can be used to measure the usability of a computer system according to the user's subjective point of view, consisting of 10 question items. This test was carried out on correspondents with various work and educational backgrounds.

The following are questions from the System Usability Scale (SUS) system testing:



Table 1. SUS Questions

NO	Question
1	I think I will use this system again.
2	I find this system complicated to use.
3	I find this system easy to use.
4	I need help from other people or technicians in using this system.
5	I feel that the system features work as they should.
6	I feel there are many things that are inconsistent (not harmonious in this system).
7	I feel like others will figure out how to use this system quickly.
8	I find this system confusing.
9	I feel there are no obstacles in using this system.
10	I need to get used to it first before using this system.

In its assessment, this test uses the following assessment scale:

Table 2. Rating scale

Mark	Scoring scale
1	STS: Strongly Disagree
2	T.S: Don't agree
3	RG: Doubtful
4	ST: Agree
5	SS: Strongly agree

The following is the System Usability Scale (SUS) test which was carried out online via Google form for 33 respondents.

Table 3. User Assessment Results

No	Name	Question									
		1	2	3	4	5	6	7	8	9	10
1	A	3	2	4	2	4	2	4	2	4	2
2	B	3	3	3	3	3	3	3	3	3	2
3	C	2	1	4	4	4	1	4	3	4	4
4	D	4	3	3	3	2	4	4	3	3	5
5	E	4	2	4	3	4	2	4	2	4	3
6	F	4	2	5	1	4	1	5	1	5	2
7	G	3	3	3	3	3	4	3	3	3	3
8	H	4	2	4	2	5	2	4	2	4	2
9	I	4	4	5	4	4	4	4	4	5	5
10	J	4	1	5	4	5	1	4	3	3	4
11	K	5	1	5	1	5	1	5	1	5	1
12	L	5	5	1	5	5	5	5	5	5	5
13	m	4	1	5	2	5	1	5	1	5	1
14	N	5	2	5	2	5	2	4	2	4	4
15	O	3	4	3	3	3	3	3	3	3	3
16	P	4	2	4	2	4	1	5	1	4	2
17	Q	4	4	5	5	4	4	5	4	4	4
18	R	4	1	4	1	4	2	4	1	4	2
19	S	5	1	5	1	5	1	5	1	5	1
20	Q	4	1	4	1	4	1	4	1	4	4
21	U	1	2	1	2	1	3	1	2	1	3
22	V	4	2	4	4	4	1	4	2	4	4
23	W	5	5	5	5	5	5	4	4	5	1
24	X	4	2	4	1	3	3	4	2	4	3
25	Y	4	2	4	4	4	3	4	2	3	3
26	Z	5	1	5	1	5	5	5	5	5	5
27	AA	1	3	3	3	3	3	3	4	3	4
28	Ab	4	1	4	1	4	1	4	1	4	1
29	AC	4	3	3	5	4	3	4	3	4	4
30	Ad	4	3	3	5	3	3	5	4	1	4
31	Ae	5	1	5	3	5	1	5	1	5	3

32 Af 1 4 1 3 3 1 2 3 2 2
33 Ag 4 2 4 1 5 2 4 1 5 5

From the assessment results, calculations are carried out based on the System Usability Scale (SUS) test. Following are the calculation results:

Table 4. SUS calculation.

Calculated Score Results (Data C example)										Amount	Mark (Amount x 2.5)
Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10		
2	3	3	3	3	3	3	3	3	3	29	73
2	2	2	2	2	2	2	2	2	3	21	53
1	4	3	1	3	4	3	2	3	1	25	63
3	2	2	2	1	1	3	2	2	0	18	45
3	3	3	2	3	3	3	3	3	2	28	70
3	3	4	4	3	4	4	4	4	3	36	90
2	2	2	2	2	1	2	2	2	2	19	48
3	3	3	3	4	3	3	3	3	3	31	78
3	1	4	1	3	1	3	1	4	0	21	53
3	4	4	1	4	4	3	2	2	1	28	70
4	4	4	4	4	4	4	4	4	4	40	100
4	0	0	0	4	0	4	0	4	0	16	40
3	4	4	3	4	4	4	4	4	4	38	95
4	3	4	3	4	3	3	3	3	1	31	78
2	1	2	2	2	2	2	2	2	2	19	48
3	3	3	3	3	4	4	4	3	3	33	83
3	1	4	0	3	1	4	1	3	1	21	53
3	4	3	4	3	3	3	4	3	3	33	83
4	4	4	4	4	4	4	4	4	4	40	100
3	4	3	4	3	4	3	4	3	1	32	80
0	3	0	3	0	2	0	3	0	2	13	33
3	3	3	1	3	4	3	3	3	1	27	68
4	0	4	0	4	0	3	1	4	4	24	60
3	3	3	4	2	2	3	3	3	2	28	70
3	3	3	1	3	2	3	3	2	2	25	63
4	4	4	4	0	0	4	0	4	0	24	60
0	2	2	2	2	2	2	1	2	1	16	40
3	4	3	4	3	4	3	4	3	4	35	88
3	2	2	0	3	2	3	2	3	1	21	53
3	2	2	0	2	2	4	1	0	1	17	43
4	4	4	2	4	4	4	4	4	2	36	90
0	1	0	2	2	4	1	2	1	3	16	40
3	3	3	4	4	3	3	4	4	0	31	78
Average Score (Final Result)											70

Following are the Scores for SUS testing:

Table 5. SUS Assessment Range

No	Value Range	Mark	Information
1	>81	A	Excellent
2	68-81	B	Good
3	68	C	OK/Fair
4	51-67	D	Poor
5	<51	E	Worst



System Usability Scale (SUS) testing in the study with a average score of 70 fell into the 68-81 category with a score of B Good.

Evaluasi Usability Website Shopee Menggunakan System Usability Scale (SUS). *Journal of Applied Informatics and Computing*, 5(2), 146–150.
<https://doi.org/10.30871/jaic.v5i2.3293>

IV. CONCLUSION

Based on the results of research conducted regarding the Website-Based Bayah Dome Geopark Information System Using the Prototype Method, it can be concluded that this research produces a website-based information system with a prototype method that can be used, and in testing the System Usability Scale (SUS) in the research with an average score 70 falls into the 68-81 value range category with a B Good value meaning the system is well used by the user.

V. ACKNOWLEDGEMENT

I would like to thank Serang Raya University for funding this research using an Internal Grant scheme with contract number: 009/HIBAHINTERNAL/UNSERAI/2023 On 26 Januari 2023, and I would also like to thank the Bayah Dome Geopark Management Agency, Lebak Regency, which supported this research until it was completed.

REFERENCES

- [1] Darsiharjo, Supriatna, U., & Saputra, I. M. (2016). Pengembangan Geopark Ciletuh Berbasis. *Jurnal Manajemen Resort Dan Leisure*, 13(1), 55–60. File:///C:/Users/User/Appdata/Local/Temp/2036-3625-1-Sm.Pdf
- [2] Gilang Ramdhani Putra. (2020). *Rancang Bangun Sistem Informasi Geografis Untuk Pemetaan Lokasi Wisata Alam Dengan Menggunakan Metode Prototyping Berbasis Mobile (Studi Kasus: Pulau Lombok) Rancang Bangun Sistem Informasi Geografis Untuk Pemetaan Lokasi Wisata Alam Dengan Menggunakan Me* [Universitas Mataram]. [Http://Begawe.Unram.Ac.Id/Index.Php/Ta/Article/View/282](http://Begawe.Unram.Ac.Id/Index.Php/Ta/Article/View/282)
- [3] Haryanto, D., & Pargana, W. (2021). Pengembangan User Interface Pada Website Geopark Belitong. *Jurnal Ilmiah Informatika Global*, 12(1), 53–58.
- [4] Hermawan, H. (2018). *Geowisata Sebagai Model Pemanfaatan*.
- [5] Ismai. (2018). Perancangan Website Sebagai Media Promosi Dan Informasi. *Journal Of Informatic Pelita Nusantara*, 3(1), 82–86.
- [6] Karyaningsih, D., & Siswanto, P. (2020). Rancang Bangun E-Learning Pronunciation Bahasa Inggris STKIP Setiabudhi Rangkasbitung Berbasis Web. *Jurnal Sisfokom (Sistem Informasi Dan Komputer)*, 9(2), 236–241. <https://doi.org/10.32736/sisfokom.v9i2.907>
- [7] Kirana, C., & Wahdaniyah, R. (2018). Implementasi Aplikasi Alumni Berbasis Mobile Application. *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*, 4(2), 179. <https://doi.org/10.26418/jp.v4i2.25752>
- [8] Permadi, R. (2015). Merintis Geowisata di Gunungsewu. *Geomagz*, Vol.5 No.1, 5(1), 32–33.
- [9] Rahma, P., Nasir, M., & Putra, A. (2019). Implementasi Metode Weighted Product dalam Sistem Pendukung Keputusan Kelayakan Pemberian Tunjangan Profesi Guru Lanjutan pada Dinas Pendidikan Kabupaten Ogan Komering Ilir (OKI) Bina Darma Conference on Computer Science 2019. *Fakultas Ilmu Komputer Universitas Bina Darma*, 263–269.
- [10] Sembodo, F. G., Fitriana, G. F., & Prasetyo, N. A. (2021).



DEVELOPMENT OF ARTS LEARNING MEDIA USING AUGMENTED REALITY WITH MARKERLESS BASED TRACKING METHOD AT SMKN 47

Ghali Hanan¹, Budi Arifitama², Silvester Dian Handy Permana³

^{1,2,3}Program Teknik Informatika, Fakultas Sains, Teknik Desain, Universitas Trilogi

Email: ¹ghalihan@trilogi.ac.id, ²budiarif@trilogi.ac.id, ³handy@trilogi.ac.id,

Abstract – During the transitional time between Covid-19 and endemic, all Indonesian schools, and SMKN 47 in particular, had restricted educational opportunities. Due to challenges encountered by the instructor and 36 students majoring in accounting in grade 12, they were unable to carry out teaching and learning activities at school; art lesson formats with three-dimensional learning modules must be visualized in 3D. However, with hybrid teaching and learning utilizing presentations, displaying the learned 3D forms is less successful. In order to visualize three dimensions in fine arts classes using augmented reality technology, we require a new learning medium. The application of the Markerless Based Tracking approach in this work enables the presentation of a tracked 3D model in the surrounding environment in real-time by merging the actual and virtual worlds as if their boundaries did not exist. The augmented reality system scans flat surfaces utilizing points, as opposed to markers or other auxiliary media. Points used as pedestals or containers to elevate three-dimensional items. There are three sculptures and three traditional dwellings on show. Online examination outcomes average resulted a 78.2% that the AR Fine Arts application by instructors and students are therefore consistent and well accepted.

Keywords – *Augmented Reality, Arts Learning, Markerless Based Tracking*

I. INTRODUCTION

Learning in all Indonesian schools is carried out online based on the decision of the Ministry of Education and Culture of the Republic of Indonesia with Circular Letter Number 4 of 2020 concerning the implementation of education during the Corona virus Disease (COVID-19) emergency period. SMKN 47, including schools in the red zone in the West Pejaten area, South Jakarta, is recommended to carry out teaching and learning activities online due to the pandemic. Based on interviews conducted with fine arts teachers at SMKN 47, 90% of the 36 grade 12 students majoring in accounting had difficulty understanding fine arts lessons, because the learning activity method was less interesting and difficult for students to understand. Therefore, we need a new visualization learning media to help teachers and students create interesting visualizations in teaching and learning even if they don't visit museums or galleries during this pandemic [1],[2].

In this research, the SMKN 47 school conducted online teaching and learning in accordance with the provisions of the Ministry of Education and Culture of the Republic of Indonesia, fine arts lessons are one of the lessons that require visualization aids in online fine arts subjects[3], [4],[5], because of the limited physical teaching aids which make teaching and learning activities ineffective, because during this pandemic Augmented Reality visualization can help in teaching and learning at SMKN 47 To overcome these problems, solutions are needed for teachers and students in the online learning process using Augmented Reality [6],[7]. With the help of Augmented Reality,

teachers can convey lessons to students more easily because of the new visualizations that will be provided, and students will understand better and not feel bored during class time [8]. Augmented Reality is a combination of the real world and the virtual world where the boundary between the two seems to be non-existent. The aim of Augmented Reality is to combine virtual and reality objects into one scope to interact with each other [9],[10].

The method used to overcome this problem is markerless based tracking with the help of the Vuforia SDK. The implementation in this research is to display 3D objects in the real world without using markers using real-time processing[11],[12]. The advantage of markerless based is that it only requires a flat surface such as a floor or table to scan to display an object. Previous research that also discussed markerless tracking was research from [13] where the research carried out was to utilize markerless tracking to learn about installing car engine spare parts where the results of this research stated that the use of markerless tracking could be used well. Other research conducted by [14] used markerless-tracking in game-based augmented reality applications and building space applications [15].

The results of this research that will be developed here are in the form of visualization learning media for fine arts lessons in the form of an Augmented Reality application with a total of 6 objects, namely 3 sculptures and 3 traditional houses using markerless based tracking to overcome problems in visualizing three-dimensional fine arts by utilizing Augmented Reality at SMKN 47 school using the markerless based tracking method.



II. RESEARCH METHODOLOGY

markerless based tracking is one method Augmented Reality which does not require special knowledge about the environment for the user to display virtual objects at a certain point [16], [17]. Camera on smartphone will scan the surface to collect visual data from the real world, when you want to display a 3D model, it will process the scan without a special marker, the 3D model will be taken from the process mapping by taking data from update, and scan the position of the 3D model on the floor to display the 3D model on the screen smartphone [18], [19].

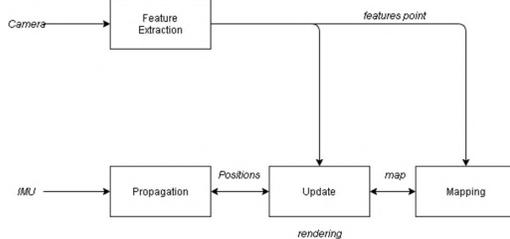


Figure 1. FlowMarkerless Based Tracking

Figure 1 shows the flow of stages in the method process markerless based tracking which has four stages, namely feature extraction, mapping, update, propagation. The following is an explanation of the stages contained in markerless based tracking among others:

Feature Extraction

At this stage what is captured by the camera scans the surface to retrieve data from the real world from the system providing points as a fulcrum segi curves, edges. After providing points for the basic pattern that will process displaying an object, the system will process creating the object directly Update and Mapping perform scanning in unknown environments without using markers or special printing media.

Stages Mapping

In level mapping, camera captured by feature extraction will be processed once when looking for points as a fulcrum. If points are detected by the image, it is possible that the 3D model displayed is unstable standing on the surface. Then the feature extraction process will carry out the process again in the same way and match the same points, and in the mapping process it will resemble the shape of the space surrounding the object that has been scanned by feature extraction.

Propagation

Task on propagation is if smartphone generating a new camera capture position or scanning to another location, will automatically process again update and mapping to produce a new position in the form of position data smartphone. IMU is a sensor used to measure three dimensions of an object. The IMU has three gyroscope axes (axes pitch, roll, and yaw) and three accelerometer axes (x, y and z axes) which have their respective uses, including gyroscopes for measuring tilt angles, measuring changes in each axis tilt angle and accelerometers for measuring and identifying the correct direction of a smartphone, [20].

Update Stages

Features point that have been detected in the process feature extraction will be updated by comparing features point which is on the map. Update carried out to improve the mapping and propagation results with the newly detected feature points. Next comes the registration stage which uses tracking to connect virtual content with the real world. This section will continue to process because there will be unstable movement of the gyroscope, therefore this will occur frequently rendering in section positioning or mapping. The process starts from virtual content to the real world, then by finding the best camera angle position, the virtual content will process the image first. Once the image is ready from the system, it will be scanned in this process to see if it has been fulfilled by the system in terms of feature extraction, and mapping. If you can interact with the application and scan with the camera smartphone, in this pose tracking process the user can interact by moving the camera to see around the 3D model. With the IMU sensor data, the image will be stable because the movement of the pose is not realized smartphone unstable when taking pictures of objects that have been scanned on a table or other flat place.

Mechanism Augmented Reality

A. Input

Users run applications that are ready to use on an Android smartphone. After going through the menu display, the user will select the 3D art menu, the application model will immediately switch to the camera Augmented Reality. Next, the user directs the camera smartphone to the floor surface until there is a camera indicator mark. In the camera Augmented Reality, there is a choice of 3D fine art models ranging from traditional houses, and sculptures from various regions, these objects are chosen by the user to display. Next, press the icon button for the desired 3D model to display it in the specified place.

B. Process

On camera smartphone directed at the floor surface, the system will identify the position smartphone with Vuforia SLAM (Egodagamage and Tuceryan, 2018), (Marchand et al., 2016), (Cortes et al., 2018) to display camera indicators, and determine the position of the object you want to display in the real world. By using the method markerless based tracking will process the mechanism or flow markerless based tracking in figure 6. After processing the mechanism markerless The user scans a flat surface to locate a 3D object. Next, the system will display the appropriate 3D model selected by the user with the specified object placement.

C. Output

The result of the general architecture is a visual art visualization information in the form of a virtual object selected by the user, with the help of smartphone. The object will display on the screen smartphone. The 3D model displayed on the floor surface is a form of visualization of images that have been processed so that it helps teachers



and students with teaching aids Augmented Reality desired fine art. The output is provided with the object process that has been directed, the user can process the dimensional shape and rotate the object.

III. RESULTS AND DISCUSSION

The implementation of the interface built for fine art AR applications is as follows:



Figure 2. Main Menu Display

Figure 2 shows the main menu display in the AR Seni Rupa application. In the display above there are 3 button features, namely Start, Guide, and Exit. When entering the AR Camera display, the user must point to a flat surface or floor so that it displays a square or so-called plane indicator, then the user must select one of the 3d model buttons and be ready to display a 3d art model object.

Guide Feature Display

The guide feature display contains information on how to use the application to display an object and feature function. Here's a look at a guide application in Figure 8.



Figure 3. Guide Feature Display

In Figure 3 is a display of the guide feature in the application. The feature guide display explains the meaning of the feature function and how to use the application steps. Feature guide provide information along with pictures.

Display of AR Application Art Objects

Below is an image of the appearance of a 3D object that has implemented markerless-based tracking which can be seen in Figure 4-6:



Figure 4. Toraja House Object





Figure 5. Lampung Traditional House



Picture 6. Catur Muka

Validation Testing

Validation testing was carried out in this research to ensure whether it can be accepted by teachers and students so that learning becomes interactive and increases students' knowledge in fine arts lessons by using Augmented Reality, the demographic profile of the respondents in this study are teachers and students from SMKN 47, where the profiles of the 38 respondents who have been collected are dominated by 17-19 year olds with 36 students and the remaining 2

respondents aged 25-35 with the teaching profession, a total of 38 Respondents consisted of 36 students and 2 teachers.

Online Testing Questionnaire Results

The following is a questionnaire that has been filled in by 70 respondents from AKL 1 (Institutional Accounting and Finance) and BDP 1 (Online Business and Marketing) classes at SMKN 47: Teachers and students experience difficulties during teaching and learning activities with textbooks. The first point to note asked whether teachers and students experienced difficulties during learning activities with textbooks, below in Figure 14 are the percentage results from the first point of the questionnaire.

70 jawaban

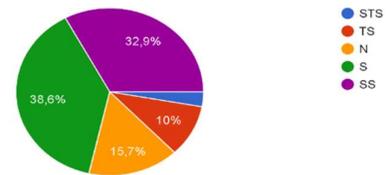


Figure 7. Percentage of results from questionnaire one

Figure 7 shows that 32.9% of 23 respondents strongly agree, 38.6% of 27 respondents agree, 15.7% of 11 respondents are neutral, 10% of 7 respondents disagree and 2.9% of 2 respondents strongly disagree. The average score for the first point is 78%. This result shows that the first score point is that teachers and students accept the first point statement, namely agreeing to the difficulties in teaching and learning with textbooks.

The second point asked whether teachers and students experienced difficulties during online learning activities. Below in Figure 8 are the percentage results from the second point of the questionnaire.

70 jawaban

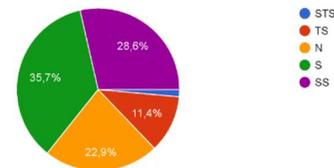


Figure 8. Percentage of second questionnaire results

Figure 8 shows that 28.6% of 20 respondents strongly agree, 35.7% of 25 respondents agree, 22.9% of 16 respondents are neutral, 11.4% of 8 respondents disagree and 1.4% of 1 respondent strongly don't agree. The average score for the second point is 76%. This result shows that the score points for both teachers and students accept the second point statement, namely agreeing with the difficulties in teaching and learning online.

The third point asked whether the AR Fine Arts Application is easy to understand and can operate well, below in Figure 9 is the percentage result of the third point questionnaire.

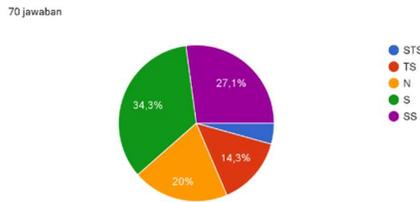


Figure 9. Percentage of third questionnaire results

Figure 9 shows that 27.1% of 19 respondents strongly agree, 34.3% of 24 respondents agree, 20% of 14 respondents are neutral, 14.3% of 10 respondents disagree and 4.3% of 3 respondents strongly disagree. The average score for the third point is 75%. That these results show that the third score point is that teachers and students accept the third point statement, namely agreeing that the AR Fine Arts Application is easy to understand and can operate well.

The fourth point asked whether teachers and students were really helped by the AR fine arts application, below in Figure 10 is the percentage result of the third point of the questionnaire.

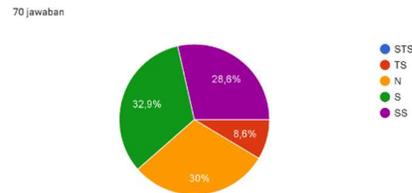


Figure 11. Percentage of results from the fourth questionnaire

Figure 11 shows that 28.6% of 20 respondents strongly agree, 32.9% of 23 respondents agree, 30% of 21 respondents are neutral, and 8.6% of 6 respondents disagree.

The average score for the fourth point is 76%. This result shows that the score of the fourth point is that teachers and students accept the statement on the fourth point, namely that they agree that teachers and students are greatly helped by the AR Fine Arts application.

The fifth point asked whether the AR fine art application could be a new innovation in teaching and learning activities, below in Figure 4.22 is the percentage result of the third point of the questionnaire.

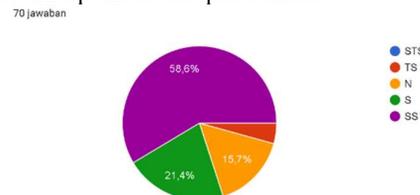


Figure 12. Percentage of Fifth Questionnaire Results

Figure 12 shows that 58.6% of 41 respondents strongly agree, 21.4% of 15 respondents agree, 15.7% of 11 respondents are neutral, and 4.3% of 3 respondents disagree. The average score for the fifth point is 87%.

These results show that the fifth score point is that teachers and students accept the fifth point statement, namely that they strongly agree that the AR Fine Arts Application can be a new innovation in teaching and learning activities. After the average of each point has been obtained, then calculate the average percentage of the total points from user testing. The overall average percentage point calculation is as follows:

Calculation results of the average percentage of overall questionnaire points from testing were 78.2%. So it can be said in the results of the assessment interval that 78.2% of teachers and students agreed and were well received.

IV. CONCLUSION

Based on the results obtained from this research, it can be concluded that the method markerless based tracking was successful in detecting the surface when implemented even though the manufacturing process was still imperfect due to the camera's lack of clarity in searching for surface points to display the object. The results of online testing regarding the AR Fine Art Application produced an average of 78.2%. So it can be said that the results of testing the AR Fine Arts application by teachers and students agreed and were well received.

REFERENCES

- [1] A. Pauls and A. Karsakov, "The concept of using augmented reality technology to present interactive calligraphic objects," in *Procedia Computer Science*, 2021. doi: 10.1016/j.procs.2021.10.042.
- [2] J. D. Hincapié-Ramos, K. Özacar, P. P. Irani, and Y. Kitamura, "GyroWand: An approach to IMU-based raycasting for augmented reality," *IEEE Comput. Graph. Appl.*, 2016, doi: 10.1109/MCG.2016.21.
- [3] F. Arici, P. Yildirim, Ş. Caliklar, and R. M. Yilmaz, "Research trends in the use of augmented reality in science education: Content and bibliometric mapping analysis," *Comput. Educ.*, 2019, doi: 10.1016/j.compedu.2019.103647.
- [4] S. C. Chang and G. J. Hwang, "Impacts of an augmented reality-based flipped learning guiding approach on students' scientific project performance and perceptions," *Comput. Educ.*, 2018, doi: 10.1016/j.compedu.2018.06.007.
- [5] A. J. Moreno-Guerrero, A. M. Rodríguez García, M. R. Navas-Parejo, and C. R. Jiménez, "Digital



- literacy and the use of augmented reality in teaching science in Secondary Education,” *Rev. Fuentes*, 2021, doi: 10.12795/REVISTAFUENTES.2021.V23.I1.12050.
- [6] S. Ceder and S. Ceder, “Augmented reality technology,” in *Towards a Posthuman Theory of Educational Relationality*, 2019. doi: 10.4324/9781351044196-11.
- [7] G. V. G. Putri, A. Syahputra, and S. D. H. Permana, “The Implementation of Augmented Reality Hairstyles at Beauty Salons Using the Viola-Jones Method (Case Study: Eka Salon),” *JISA (Jurnal Inform. dan Sains)*, vol. 3, no. 2, 2020, doi: <https://doi.org/10.31326/jisa.v3i2.847>.
- [8] R. Mauludin, A. S. Sukamto, and H. Muhandi, “Penerapan Augmented Reality Sebagai Media Pembelajaran Sistem Pencernaan pada Manusia dalam Mata Pelajaran Biologi,” *J. Edukasi dan Penelit. Inform.*, vol. 3, no. 2, p. 117, 2017, doi: 10.26418/jp.v3i2.22676.
- [9] W. Ramadhan, B. Arifitama, and S. D. H. Permana, “Mobile augmented reality for furniture visualization using Simultaneous Localization and Mapping (SLAM),” *IOP Conf. Ser. Mater. Sci. Eng.*, 2021, doi: 10.1088/1757-899x/1098/6/062008.
- [10] D. Rohendi and Y. Wihardi, “Learning Three-Dimensional Shapes in Geometry Using Mobile-Based Augmented Reality,” *Int. J. Interact. Mob. Technol.*, vol. 14, no. 09, p. 48, 2020, doi: 10.3991/ijim.v14i09.13035.
- [11] G.-J. Hwang, P.-H. Wu, C.-C. Chen, and N.-T. Tu, “Effects of an augmented reality-based educational game on students’ learning achievements and attitudes in real-world observations,” *J. Interact. Learn. Environment*, vol. 24, no. 8, pp. 1895–1906, 2016, doi: <https://doi.org/10.1080/10494820.2015.1057747>.
- [12] A. A. M. Akbar, “Sistem Pemandu Wisata Berbasis Augmented Reality Markerless Memanfaatkan GPS Based Tracking Dengan Platform Android (Studi Kasus : Lokasi Wisata di Kota Makassar),” Universitas Islam Negeri Alauddin Makassar, 2017.
- [13] H. Hermawan, R. Waluyo, and M. Ichsan, “Pengembangan Media Pembelajaran Mesin Menggunakan Teknologi Augmented Reality,” *J. Innov. Inf. Technol. Appl.*, 2019, doi: 10.35970/jinita.v1i01.88.
- [14] M. Reyhan, H. Tolle, and L. Fanani, “Pembangunan Aplikasi Entertainment Terminal Berbasis Sistem Kendali Head Movement Control System pada Perangkat IOS,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 2, no. 9, pp. 3258–3263, 2018.
- [15] B. Satria and P. Prihandoko, “IMPLEMENTASI METODE MARKER BASED TRACKING PADA APLIKASI BANGUN RUANG BERBASIS AUGMENTED REALITY,” *Sebatik*, 2018, doi: 10.46984/sebatik.v19i1.88.
- [16] I. M. O. Widyantara, D. M. Wiharta, and P. Widiadnyana, “Implementasi Aplikasi Mobile Augmented Reality Untuk Pengenalan Materi Bangun Ruang,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 9, no. 2, pp. 313–322, 2022, doi: 10.25126/jtiik.202295032.
- [17] M. A. Muhammad *et al.*, “Augmented Reality Pelacak Lokasi Pustaka,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 1, pp. 77–86, 2020, doi: 10.25126/jtiik.202071343.
- [18] A. K. Dash, S. K. Behera, D. P. Dogra, and P. P. Roy, “Designing of marker-based augmented reality learning environment for kids using convolutional neural network architecture,” *Displays*, 2018, doi: 10.1016/j.displa.2018.10.003.
- [19] J. C. Piao and S. D. Kim, “Adaptive monocular visual-inertial SLAM for real-time augmented reality applications in mobile devices,” *Sensors (Switzerland)*, 2017, doi: 10.3390/s17112567.
- [20] B. Arifitama, A. Syahputra, S. D. H. Permana, and K. B. Y. Bintoro, “Mobile Augmented Reality for



Learning Traditional Culture Using Marker Based Tracking,” in *IOP Conference Series: Materials Science and Engineering*, 2019. doi: 10.1088/1757-899X/662/2/022038.



Implementation of The Resident's Dues Applications (SIUMAS) Using Waterfall Method in RT X Cinere Village

Febrianto^{1*}, Ikhwana Fakhir Nur Aziz², Mochammad Alvian Kosim³, Muhammad Darwis⁴, Retno Hendrowati⁵

^{1,2,3,4,5}Program Studi Teknik Informatika, Fakultas Ilmu Rekayasa, Universitas Paramadina
Email: 1febrianto@students.paramadina.ac.id, 2ikhwana.aziz@students.paramadina.ac.id,
3mochammad.kosim@students.paramadina.ac.id, 4muhammad.darwis@paramadina.ac.id,
5retno.hendrowati@paramadina.ac.id

Abstract – Payment of dues is one of the obligations that must be fulfilled by every resident in a neighborhood association area. These dues are used for common purposes such as cleanliness and security. The collection and expenditure of dues are usually handled by the management committee. However, there are still difficulties in managing residents' dues using conventional methods. As a result, dues cards are often lost and the dues collection are not transparent. Therefore, the author developed The Resident's Dues Application (SIUMAS) as a solution used to manage residents' dues. The purpose of developing this application is to simplify the management tasks of the committee in handling the collection and expenditure of dues, as well as to facilitate residents in tracking the utilization of the dues. This application was developed by applying the waterfall method which consists of several stages, starting from analysis to the maintenance process. In addition, this application was built using the PHP programming language and utilizes the MySQL database. The development of this application adopts an object-based application design method. The results of the SIUMAS application testing using Black Box Testing shown that all functionalities of the SIUMAS application are functioning well and meet the expectations.

Keywords – Resident Dues Application, Web Programming, Waterfall Method, Object-Based Design, Black Box Testing

I. INTRODUCTION

The development of information and communication technology has a significant impact on society's social life. People need services that are functional, fast, transparent, and accurate. The role of technology is very important in providing good public services, starting from the RT/RW level. Ease of access and use of information technology in various fields, such as education, health, business, and government administration, provides various benefits. The public can quickly get information at a lower cost than traditional methods, while the government can manage public services more easily and transparently.

In RT areas, environmental safety and cleanliness are very important for the comfort of residents. To achieve this goal, residents formed a security and cleaning team and collected fees. However, the process of recording and publishing dues collections is still done manually and is less effective. To increase efficiency and transparency, an application is needed that can assist in recording and publishing dues collections. At RT X in Cinere Village for example, they feel that having application technology that can be accessed transparently can help citizens monitor contribution funds. This application is used to help manage fees, provide transparency to residents, and ultimately increase the effectiveness of security and cleanliness programs in the RT area.

To support this community program, the author developed an application called Resident's Dues Application (SIUMAS). SIUMAS aims to assist RT administrators in recording and reporting the use of dues

funds transparently to the community in the RT area or areas that use this application. With SIUMAS, the public can easily check and monitor the number of dues funds they have paid and how these funds are used. This will create transparency and accountability in the management of dues funds so that it can create a clean RT environment and good government [1].

Prior to this research, there had been previous research conducted by A.P. Sutrisno et al. (2022) which discusses the design and development of waste dues applications using the Android-based Rapid Application Development method [2]. In this research, the development of an Android-based waste fee application was carried out due to the problems of residents who had difficulty managing waste fees manually. As a result, after the application was developed and tested, the application could run well and meet the needs of residents.

Then next, there is research from N.G. Salsabila et al. (2021) regarding the creation of a series of citizen application modules for managing citizen contribution wallets [3]. This application was created to make it easier for administrators and residents to manage and supervise the collection and use of contribution funds. Then, there is also research from F. Wahyudi (2020) regarding the design of a web-based committee dues application. In this research, the application was designed based on the problems caused by the previous conventional contribution collection method. This application helps the process of paying committee dues [4]. Apart from that, several other studies which also have the same focus and objectives include Rahman and Aryani regarding the application of



citizen fees using Java [5], Alviana and Kurniawan's research on the Citizen Fee Information System [6], Siregar et al. research on the Security Fee System [7] and Sari and Kholil's research on Citizen Applications [8].

The author developed the SUIMAS application by proposing an object-oriented method, which focuses on design and implementation. The application was developed by applying the waterfall method which consists of several stages, starting from analysis to the maintenance process [9]. By using waterfall method, the author can implement the SUIMAS application quickly and purposefully. To support the development process, the author utilizes the Unified Modeling Language (UML) which is very suitable for the method used [10]. Some of the tools that the author also uses include MySQL for the database [11] and the PHP programming language for implementing the coding [12].

Finally, this research produced a resident's dues application that can be used by the RT X Cinere Village in the dues payment process. The reliability of the application has been tested using the black box method where the results show that all functions can run properly [13]. Of course, this application can also be adopted by other parties who have a similar process.

II. RESEARCH METHODS

The research method used to design the Resident's Dues Application (SIUMAS) is object-based design analysis. In this method, research will focus on analysis and application design based on object-oriented programming principles. This involves identifying and modeling the objects involved in the system, as well as the relationships and interactions between these objects. This object-based design analysis method aims to create good and effective designs in meeting the needs of application users and ensuring the sustainability and scalability of the application.

In this waterfall methodology, research is focused on understanding information about user needs for software. The information collection method is determined through discussion, observation, surveys and interviews. This information is then processed and analyzed to obtain complete data or information regarding user perceptions. and software to be developed [14].

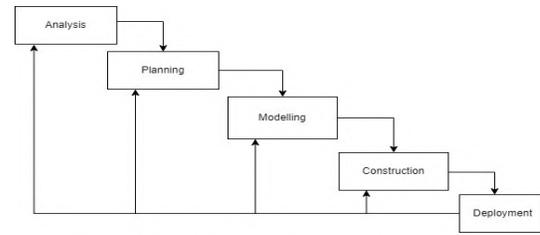


Fig 1. SUIMAS Research Waterfall Methodology

1. Analysis

This step involves analyzing software requirements and collecting data through meetings with customers as well as collecting additional information from journals, articles, and the internet.

2. Planning

The planning process is a continuation of needs analysis (communication). This stage produces user requirements documents or data related to user desires in creating software, including implementation plans.

3. Modeling

The modeling process translates requirements into a predictable software design before code generation. The focus is on data structure design, software architecture, interface representation, and procedural details (algorithms). This stage produces documents called software requirements.

4. Construction

The construction phase involves coding. Coding or coding translates a design into a language that can be recognized by a computer. The programmer translates the transactions requested by the user. This stage is the real stage in software creation, where computer use is maximized. After coding is complete, the system that has been created will be tested to find errors which can then be corrected.

5. Deployment

This stage can be considered the final stage in creating software or systems. After analyzing, designing and coding, the finished system will be used by users. Software that has been created must undergo regular maintenance after implementation.

Object-oriented design is defining all types of objects that are important for communicating with humans and equipment in the system and showing how objects interact with each other to complete certain work and improve the function of each type of object so that it can be implemented with a special language or environment [15].

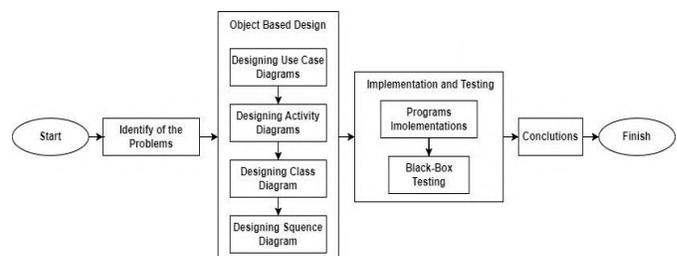


Fig 2. SIUMAS research methodology

Based on figure 2 steps of the research is:



1. Identification of problems

At this stage, the author collects information to analyze the dues collection methods that are being implemented in RT X. Then, the author identified any problems that occurred using the dues collection method that was being implemented at that time and then looked for solutions.

2. Object Based Design

At this design stage, the application design is described SIUMAS object based. This design includes designing use case diagrams, activity diagrams, class diagrams, and sequence diagrams.

3. Program Implementation and Testing

At this stage, the application code is written so that the application can be built and the results tested using the black box testing method [16]. The technology used to build SIUMAS includes the programming language and database used.

The programming language used to build SIUMAS is PHP. PHP (Hypertext Preprocessor) is a server-side scripting language, a programming language used to build static websites, dynamic websites, or web applications [17]. The database used for the application SIUMAS is MySQL. MySQL is a server that serves databases. To create and process a database, you are required to learn special programming called SQL queries [18].

In an application testing session SIUMAS, Black box testing was chosen to test the course of the application process. Black Box Testing is a type of testing that involves software without paying attention to how its internals work. In this testing, testers see the software as a "black box" entity that does not need to understand the internal details, but still carries out the testing process on the external part [19].

4. Conclusion

At this stage, the conclusions of the research conducted by the author are written.

III. RESULTS AND DISCUSSION

Citizen dues activities at RT X Cinere Village still using the manual method with printed dues cards which will be checked every month by officers. Residents who are visited by officers will pay their fees, then their printed dues cards will be marked as proof of payment. Printed fee cards have disadvantages such as being easily lost, so officers must carry replacement printed fee cards. In addition, contribution data could potentially be lost. Then, the amount of fees collected is also not transparent to residents because the total is on their respective printed contribution cards.

Based on the problem analysis above, the problems faced by RT X Cinere Village is the lack of transparency in the number of dues. Apart from that, it is also inefficient to use printed dues cards. So, the author created a Resident's Dues Application (SIUMAS) which can be used to help record and publish resident's dues.

As explained in the research methodology chapter, to assist the author in designing and developing the SIUMAS application, the author utilized the UML tool. Some of the diagrams that the author put forward are:

A. Use Case Diagram Design

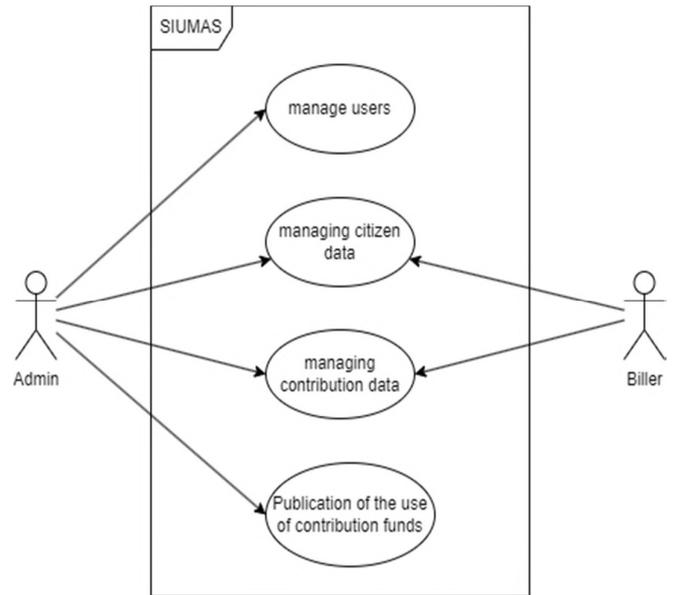


Fig 3. Use case diagram of SIUMAS

In Figure 3, it can be seen that the SIUMAS use case diagram has two main actors. The first actor is the biller whose job is to manage citizen data and manage contribution data. Then, the second actor is the admin whose job is to manage all the data in the application.

B. Activity Diagram Design

In the SIUMAS activity diagram design, there are two main activity flows, namely activity admin and biller as shown in Figure 4.

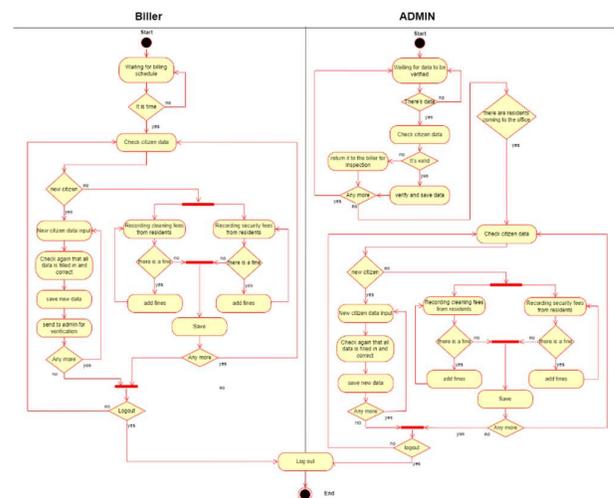


Fig 4. Activity diagram of SIUMAS

C. Class Diagrams Design

In the SIUMAS class diagram design, as can be seen in Figure 4, there are 5 main class entities, namely Billing



Users, Admin Users, Biller, Dues, Family, and Expenditures.

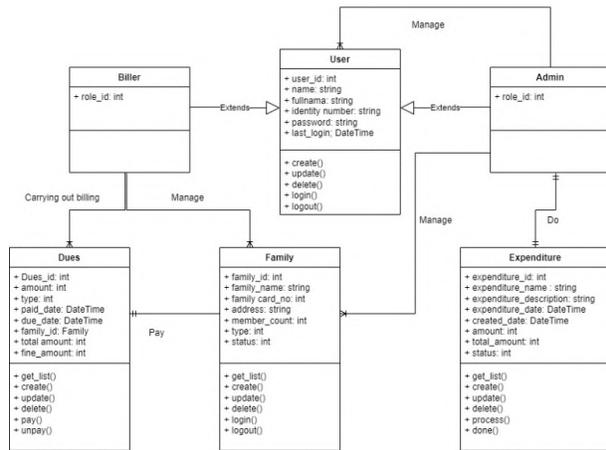


Fig 5. Class diagram of SIUMAS

D. Sequence Diagram Design

The author also designed a sequence diagram for the SIUMAS application to find out in more detail the processes running in the application. The results are as in Figure 6.

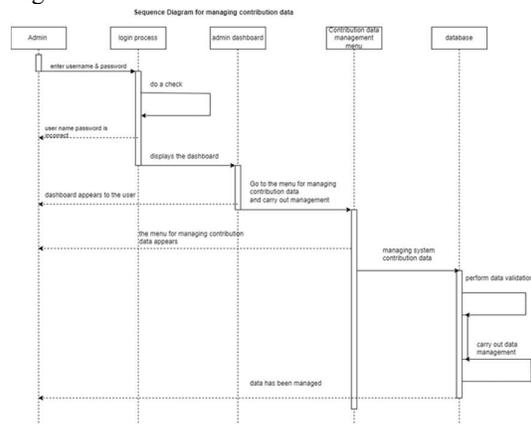


Fig 6. Sequence diagram of SIUMAS

E. Program Implementation

At the program implementation stage, the SIUMAS application began to be implemented in the form of programming code. Based on the object-based application design method, application creation follows Use Case Diagram, Activity Diagram, Class Diagram and Sequence Diagram.

1. Home Page SIUMAS

This page is the display of the main SIUMAS page when opened by residents. There are several menus such as View Reports, About Us, Terms & Conditions, and FAQ. There is also a Login button here to enter the data management menu.



Fig 7. Home page of SIUMAS

2. Reports Page SIUMAS

This page is a display of the SIUMAS report page which contains comprehensive income and expenditure data from SIUMAS as shown in Figure 7.

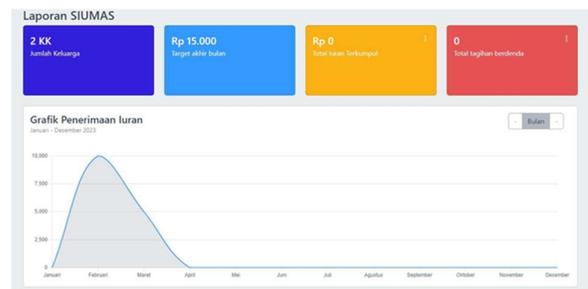


Fig 8. Report page of SIUMAS

3. Dues Management Page

This page displays the SIUMAS dues management page when opened by the biller or admin. On this page, dues and expense data is managed by the admin and biller as shown in Figure 8.

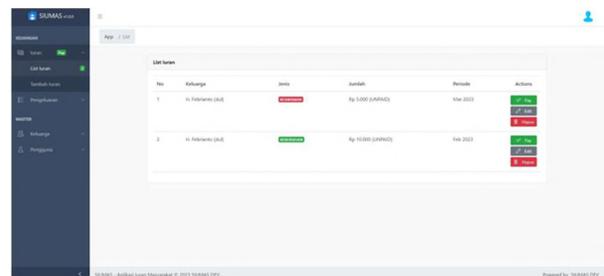


Fig 9. Dues management page of SIUMAS

F. Application Testing

To test the reliability of the SIUMAS application, the author uses the black box testing method. In this test, all test cases have passed all stages of testing successfully. In this test, there are several stages and processes in the SIUMAS application that are tested as in Table 1



Table 1. Black Box Testing results

Test ID	Black Box Testing Result			
	Description	Expected	Test Result	Remarks
T1	Admin Login	Admin will successfully logged into the application	Succeed login to the application	Success
T1	Admin Manages data	Admin will successfully manage data	Successfully manage expense data	Success
T2	expenses The biller logs in	The expense biller has successfully logged into the application	Successfully logged in to the application	Success
T2	The biller carries out the collection	The biller enters billing data	Successfully entered billing data	Success
T3	Open the web and see home page	The web is open and can see the home page	Home Page web open	Success

Description: T1 is Admin, T2 is Biller, T3 is Community.

IV. CONCLUSION

Based on research conducted by the author, the SUIMAS application can help RT administrators and residents to manage dues easily and transparently. Users can use the application to manage dues income and expenditure data optimally. By implementing this application thoroughly, cleanliness and security can be well maintained and will improve the quality of the environment and make people feel comfortable and safe.

During the application development process, the object-based design method was used for the author to complete this SIUMAS application. The application produced in this research has been tested using the black box method and it is known that the entire application function runs well. This shows that the SUIMAS application is ready to be used by the public to manage their contribution data.

As a recommendation from the author for further research, RT administrators who will use the SIUMAS application can be given socialization on the use of the application. Apart from that, after the socialization is carried out, a survey can also be given to find out the level of public acceptance of the SIUMAS application.

REFERENCES

- [1] "Portal Berita Resmi Pemerintah Kota Depok - Pemkot Depok Raih WTP ke-11 Kali, Transparansi dan Akuntabilitas Jadi Kunci." Accessed: Jun. 08, 2023. [Online]. Available: <https://berita.depok.go.id/pemkot-depok-raih-wtp-ke-11-kali-transparansi-dan-akuntabilitas-jadi-kunci>
- [2] A. P. Sutrisno, A. C. Wardhana, and A. Burhanuddin, "Rancang Bangun Aplikasi Aplikasi Iuran Sampah RT/RW Menggunakan Metode Rapid Application Development Berbasis Android," *JURIKOM (Jurnal Riset Komputer)*, vol. 9, no. 4, pp. 1073–1086, Aug. 2022, doi: 10.30865/JURIKOM.V9I4.4185.
- [3] N. G. Salsabila, S. Suryatiningsih, and D. R. Wijaya, "Aplikasi Warga Berseri Modul Pengelolaan Dompot Iuran Warga (studi Kasus: Perumahan Permata Buah Batu)," *eProceedings of Applied Science*, vol. 7, no. 6, Dec. 2021, Accessed: Dec. 09, 2023. [Online]. Available: <https://openlibrarypublications.telkomuniversity.ac.id/index.php/appliedscience/article/view/17318/17029>
- [4] F. Wahyudi, "RANCANG BANGUN APLIKASI PEMBAYARAN IURAN KOMITE BERBASIS WEB DI SMK TAMAN ILMU KROMENGAN," *G-Tech: Jurnal Teknologi Terapan*, vol. 3, no. 2, pp. 214–220, Aug. 2020, doi: 10.33379/GTECH.V3I2.376.
- [5] W. A. Rahman and L. Ariyani, "Rancang Bangun Sistem Informasi Pembayaran Iuran Warga RT 05 RW 002 Berbasis Java," *Jurnal Riset dan Aplikasi Mahasiswa Informatika (JRAMI)*, vol. 2, no. 04, pp. 657–662, Oct. 2021, doi: 10.30998/JRAMI.V2I04.1637.
- [6] S. Alviana and B. Kurniawan, "Penerapan Sistem Informasi Iuran Warga Griya Pataruman Asri Berbasis Website," *Jurnal Pengabdian Masyarakat Indonesia*, vol. 1, no. 6, pp. 343–350, May 2021, doi: 10.52436/1.JPMI.41.
- [7] W. Siregar, I. Irvan, and E. Rahayu, "SISTEM INFORMASI PEMBAYARAN IURAN KEAMANAN DAN KEBERSIHAN PADA PERUMAHAN BERBASIS WEBSITE MENGGUNAKAN METODE DESIGN THINKING," *JiTEKH*, vol. 8, no. 2, pp. 50–58, Sep. 2020, doi: 10.35447/JITEKH.V8I2.204.
- [8] A. O. Sari and I. Kholil, "Prototype Aplikasi SI-Warga Sebagai Penunjang Administrasi Surat Pengantar dan Iuran Warga pada Rukun Tetangga," *Infotek: Jurnal Informatika dan Teknologi*, vol. 5, no. 2, pp. 307–315, Jul. 2022, doi: 10.29408/JIT.V5I2.5897.
- [9] E. Septiawan *et al.*, "PENERAPAN METODE WATERFALL PADA PENGEMBANGAN SISTEM INFORMASI MONITORING PROSES BIMBINGAN SKRIPSI DI JURUSAN ILMU KOMPUTER UNIVERSITAS LAMPUNG," *Jurnal Pepadun*, vol. 3, no. 1, pp. 74–87, Apr. 2022, doi: 10.23960/PEPADUN.V3I1.102.
- [10] A. Shaikh and U. K. Wil, "Overview of Slicing and Feedback Techniques for Efficient Verification of UML/OCL Class Diagrams," *IEEE Access*, vol. 6, pp. 23864–23882, Jan. 2018, doi: 10.1109/ACCESS.2018.2797695.
- [11] "MySQL :: Why MySQL?" Accessed: Jun. 09, 2023. [Online]. Available: <https://www.mysql.com/why-mysql/>
- [12] "PHP: Hypertext Preprocessor." Accessed: Jun. 09, 2023. [Online]. Available: <https://www.php.net/>
- [13] T. Snadhika Jaya, P. Studi Manajemen Informatika, J. Ekonomi dan Bisnis, and P. Negeri Lampung JlnSoekarno, "Pengujian Aplikasi dengan Metode Blackbox Testing Boundary Value Analysis (Studi Kasus: Kantor Digital Politeknik



- Negeri Lampung),” *Jurnal Informatika Politeknik Harapan Bersama*, vol. 3, no. 1, pp. 45–48, Jan. 2018, doi: 10.30591/JPIT.V3I1.647.
- [14] D. S. Purnia, A. Rifai, and S. Rahmatullah, “Penerapan Metode Waterfall dalam Perancangan Sistem Informasi Aplikasi Bantuan Sosial Berbasis Android,” *Prosiding Semnastek*, Nov. 2019, Accessed: Dec. 09, 2023. [Online]. Available: <https://jurnal.umj.ac.id/index.php/semnastek/article/view/5238>
- [15] J. W. Satzinger, R. B. Jackson, and S. D. Burd, *Systems Analysis and Design in a Changing World*. Cengage Learning, 2011.
- [16] S. Kukolj, V. Marinković, M. Popović, and S. Bognár, “Selection and prioritization of test cases by combining white-box and black-box testing methods,” *Proceedings - 2013 IEEE 3rd Eastern European Regional Conference on the Engineering of Computer Based Systems, ECBS-EERC 2013*, pp. 153–156, 2013, doi: 10.1109/ECBS-EERC.2013.28.
- [17] “Pengertian PHP, Fungsi, Syntax & Kelebihannya.” Accessed: Jun. 09, 2023. [Online]. Available: <https://www.jagoanhosting.com/blog/pengertian-php/>
- [18] J. Enterprise, *HTML, PHP, dan MySQL untuk Pemula*. Elex Media Komputindo, 2018.
- [19] S. R. Wicaksono, *Rekayasa Perangkat Lunak*. Seribu Bintang, 2011.



Implementation of Android-Based Tailoring Service Ordering Application with Geolocation Integration

Muhamad Agung Santoso^{1*}, Joko Sutopo²

^{1,2}Informatics Study Program, Faculty of Science and Technology, Universitas Teknologi Yogyakarta
Email: ¹muhamad.5200411549@student.uty.ac.id, ²jksutopo@uty.ac.id

Abstract – Basic needs are an aspect that cannot be separated from human life, and one of the important aspects of these basic needs is clothing which not only functions to protect the body but also plays a role in the world of fashion. To meet these fashion needs, tailoring services have a key role in providing clothes that suit consumer desires. While this need is important, customers often rely on conventional methods, such as word-of-mouth recommendations or seeking out a familiar tailor. Some tailors may have a sign in front of their house, but it is often difficult for potential customers to find and connect with them. Therefore, this research aims to implement a tailor service ordering application that is integrated with geolocation features and can make things easier for tailors and consumers. This application development uses the Java programming language and Android Studio as the framework and Firebase as the database. The result of this research is an application that allows customers to order tailor services via smartphone anywhere and anytime.

Keywords – *Android Studio, Firebase, Geolocation, Java, Tailor*

I. INTRODUCTION

Clothing or clothing needs are one of the main aspects of human life [1]. In the past, clothing was used primarily as a means of protection for the body. However, over time, clothing has transformed, and one of its important roles today is to provide comfort in various situations, such as when dealing with important moments in life. In the history of technological development, humans have developed various tools to produce clothing, ranging from simple sewing machines to sophisticated modern sewing machines [2].

The increasingly rapid development of technology has brought convenience to various aspects of people's lives. One way is through an online goods and services ordering application [3]. With the help of smartphones, people can now order a variety of products and services without having to leave the comfort of their homes.

The tailoring business sector is very popular among the public, attracting the interest of many individuals to start a business in this sector, especially because of its promising profit potential [4]. Sewing courses are also increasingly in demand, with many people taking them to develop skills that can later be used in setting up their own tailoring business [5].

Even though many people have opened tailoring services, many consumers still face difficulties when ordering this service directly at the tailoring business location [6]. This process is considered quite complicated because consumers must measure their body size and choose their clothing materials. In addition, difficulties arise if consumers cannot meet with the tailor due to schedule inconsistencies. This condition can reduce efficiency because consumers are forced to come repeatedly to the tailoring service.

Apart from that, another additional problem is the schedule conflict that often arises between the schedule submitted by the customer and the tailor's schedule.

Sometimes, the schedule agreed upon by both parties does not always match the initial agreement, resulting in a situation where the customer must return without success and wait another day.

To overcome several challenges mentioned above, this research aims to develop a mobile application designed with a feature for ordering tailoring services using geolocation integration [7]. The app will allow users to book tailoring services and confirm their location and order details easily. Within the framework of this research, we will design and analyze this mobile application system as an innovative solution for the needs of ordering tailoring services.

Previous research has focused on developing various applications and platforms for ordering tailoring services. Reference [8], introduces an application that allows consumers to create and view their orders while providing the ability to enter tailor data. Reference [9] notes the development of a platform in the form of a website for ordering tailoring services online. The analysis process in this research adopts the prototyping method, which involves creating a prototype model for testing before creating the final product. The platform involves three types of users, each having a unique role. Admin is responsible for confirming payments made by buyers, while buyers can easily order tailoring services online. Tailors themselves have access to add and update their tailoring services through the website.

II. RESEARCH METHODOLOGY

Previously, the result of this research was the development of an application that allows users to order sewing services via mobile smartphones. During the development of this application, Location Based Service (LBS) was integrated as the main method [10]. The application structure and design can be illustrated in detail in Figure 1.



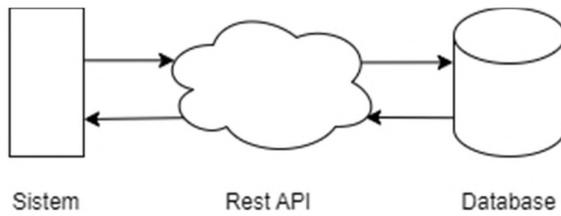


Fig1. Application Architecture

In Figure 1, this application uses the REST API to access the database. RESTful API is an interface used to interact with web databases. This interface allows users to easily access the database from various applications [11]. Sometimes a RESTful API is also called a gateway because it acts as an intermediary between the system and the database.

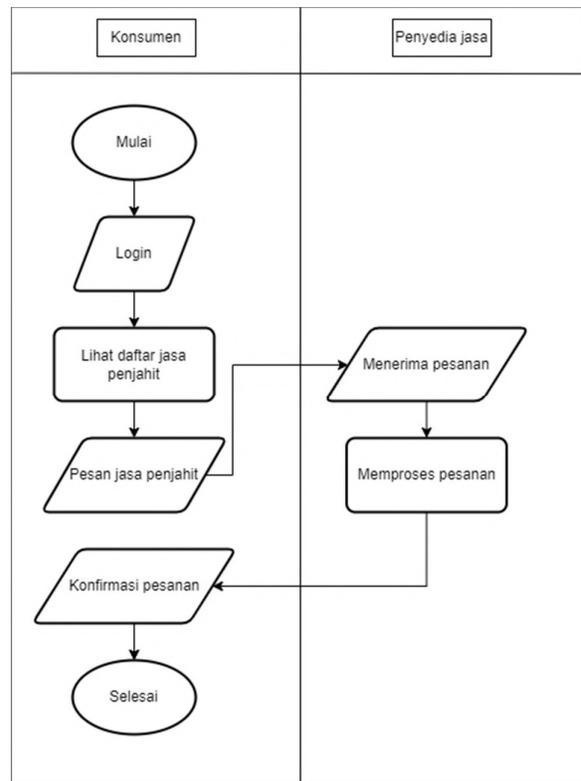


Figure 2 above illustrates the application flow, which

Fig 1. Application Flow

can be seen in the illustration. In this application, users can log in and access the catalog of available tailoring services. After finding the service they want, users can place an order directly. The order will be received by the service provider, who will immediately process it and send it to the user. The user is also asked to confirm the order that has been received from the service provider.

III. RESULTS AND DISCUSSION

In the initial stages of developing this application, the design was prepared using the popular online graphic design application Figma. Figma is a highly flexible and internet-accessible design tool that allows designers to create various types of designs including web designs, mobile applications, desktop applications, as well as presentations. The decision to use Figma as the main design tool was influenced by its ease of use and the availability of online access which makes it easy to create attractive designs [3].



Fig 3. Log in Page Design

Figure 3 shows the interface of the login page, including a login form with fields for filling in the username and password. Next to it is also a button that takes the user to the registration page.



Fig 2. Registration Page Design



The registration page shown in Figure 4 contains various information to be filled in by the user. It contains a form that asks the user to enter their email, username, and password. In addition, the user is asked to identify their role as a customer or a tailoring service provider.



Fig 5. Home Page Design

In Figure 5, the home page contains various information about the customization services available for consumer orders. On this page, users will find various navigation buttons that allow them to move to different sections, including buttons that lead to relevant information pages.

The next process involves the application coding stage. For application development, Android Studio is used as an integrated development environment (IDE) specifically designed by Google and JetBrains to create Android applications [12]. Android Studio is currently capable of supporting two programming languages used in application development: Java and Kotlin. In this research, application development was carried out using the Java programming language.

The database system used in this project is Firebase which is an application development solution designed by Google [13]. Firebase offers various related features, including the use of a non-SQL database capable of storing various data types such as strings, integers, and booleans [14]. Firebase is often used in the development of Android-based applications. The output of this non-SQL database is usually in JSON or XML format, which can then be processed by the application as needed [15].

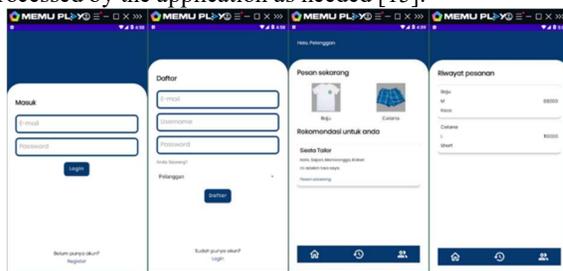


Fig 6. Application Page View

In the figure above, you can see the visualization of the developed application. The first image (from left to right) shows the initial interface of the application which serves as the login page. On this page, the user has the option to switch to the registration view if they do not yet have an account. The second image shows the registration view that allows users to create a new account on the application.

When the user successfully accesses the app, the third image displays the customer information panel that contains information about the services available in the app. Here customers can place an order by pressing a button according to their choice, such as clothes or pants.

Finally, the fourth image displays the order history page which shows a complete list of orders that the user has placed along with the status of each order.

Table 1. Testing Table

No.	Testing Scenario	What to expect	Conclusions
1.	Fill in the login form and click the login button	The system denies login access	Successful
2.	Clear the "username" field on the login form	The system redirects to the login form with "username cannot be empty" and the login fails.	Successful
3.	Clear the "password" field on the login form	The system redirects to the login form with "password cannot be empty" and the login fails.	Successful
4.	Fill in the username and password incorrectly	The system will redirect to the login form with "invalid username and password" and the login failed.	Successful
5.	Fill in the username with "customer"	The system will redirect to the customer's homepage	Successful
6.	Mengisi username dengan "penjahit"	Fill in the username with "tailor"	Successful

Based on the test results in Table 1 above, there are no problems when users log in to the application. However, if



the user enters the wrong username or password, the system will issue a warning. After the user enters the correct login details, they will be directed directly to the dashboard based on their user role.

IV. CONCLUSION

Based on the results of the research conducted, it is evident that Android applications have great potential to improve location-based on-demand ordering services. Using location-based services and advanced integration into Android programming, the app can effectively identify the location of customers ordering tailoring services. This helps the service provider or tailor deliver the order more easily and quickly.

REFERENCES

- [1] A. H. Suryana, M. Djafar, R. Sefia, and R. Djatalov, "Implementasi Metode Prototype pada Perancangan Aplikasi Pemesanan Jasa Jahit Berbasis WEB pada Korita Tailor," vol. 1, no. 1, pp. 9–13, 2023.
- [2] E. B. Siregar, "Aplikasi Layanan Jahit menggunakan Framework React Native," Batam, 2021. [Online]. Available: <http://repository.upbatam.ac.id/1271/>
- [3] E. C. Narendra, S. A. W. Wati, A. S. Fitri, M. A. Priyanto, and D. A. Putri, "Analisis Desain Aplikasi Jahit Pakaian Custom Online Berbasis Online," *Jitet (Jurnal Inform. dan Tek. Elektro Ter.*, vol. 11, no. 1, pp. 86–95, 2023.
- [4] K. Sabri, Afrijal, and R. Karno, "Sistem Informasi Inventory Penjahit di Kecamatan Rambah," *Riau J. Comput. Sci.*, vol. 8, no. 2, pp. 116–122, 2022.
- [5] C. E. Atiaini, Mustaqiem, and Minarni, "Pengembangan Teknologi E-Marketplace Jasa Penjahit di KOTIM dengan Metode SMART dan Location Based Services," *Tin Terap. Inform. Nusant.*, vol. 2, no. 8, pp. 501–505, 2022, doi: 10.47065/tin.v2i8.1149.
- [6] Susandri, H. Spitri, Lusiana, and K. Harianto, "Aplikasi Jasa Jahit Pakaian Berbasis Mobile dengan Teknologi Location Based Services dan Metode SMART," *J. Inovtek Polbeng*, vol. 5, no. 1, pp. 128–138, 2020.
- [7] A. Setianni and R. Syahputri, "Pesona : Sistem Informasi Pemesanan Salon Online Berbasis Location Based Service," *Pros. Semin. Nas. 2019 Sinergitas Multidisiplin Ilmu Pengetah. dan Teknol.*, vol. 2, no. 2011, pp. 1–7, 2019, [Online]. Available: <https://jurnal.yapri.ac.id/index.php/semnassmipt/article/view/62>
- [8] B. A. Pramukti, F. D. Astuti, and M. A. Nugroho, "Implementasi Aplikasi Pemesanan Jasa Penjahit Berbasis Android," *Juti (Jurnal Teknol. Informasi)*, vol. 1, no. 1, pp. 22–29, 2022.
- [9] J. P. B. Saputra, T. Bratakusuma, and Z. Rifai, "Aplikasi Untuk Order Jahit Secara Online," *Naratif (Jurnal Ilm. Nas. Ris. Apl. dan Tek. Inform.*, vol. 3, no. 01, pp. 25–31, 2021, doi: 10.53580/naratif.v3i01.115.
- [10] J. S. Sinaga and M. Rahman, "Rancang Bangun Aplikasi Pemesanan Jasa Jahit Menggunakan Metode Location Based Service (LBS) Di Kota Medan," *J. JUREKSI (Jurnal Rekeyasa Sist.*, vol. 1, no. 1, pp. 147–158, 2023, [Online]. Available: <http://kti.potensi-utama.ac.id/index.php/JUREKSI/index>
- [11] A. W. Syahroni and Slamet, "Rancang Bangun Aplikasi Jasa Online Berbasis Mobile," *J. Teknol. Inf.*, vol. 15, no. 3, pp. 102–108, 2020.
- [12] A. Studio, "Mengenal Android Studio," Android Studio. [Online]. Available: <https://developer.android.com/studio/intro?hl=id>
- [13] A. Sonita and R. F. Fardianitama, "Aplikasi E-Order Menggunakan Firebase dan Algoritme Knuth Morris Pratt Berbasis Android/," *J. Pseudocode*, vol. 5, no. 2, pp. 270–271, 2018, doi: 10.1007/978-1-4842-8745-3_10.
- [14] R. Andrianto and M. H. Munandar, "Aplikasi E-Commerce Penjualan Pakaian Berbasis Android Menggunakan Firebase Realtime Database," *J. Comput. Sci. Inf. Technol.*, vol. 3, no. 1, pp. 20–29, 2022, [Online]. Available: <https://jurnal.ulb.ac.id/index.php/JCoInT/article/view/2478>
- [15] F. A. Oktavirahani and R. Maharesi, "Implementasi Algoritma Decision Tree Cart Untuk Merekomendasikan Ukuran Baju," *JURIKOM (Jurnal Ris. Komputer)*, vol. 9, no. 1, p. 138, 2022, doi: 10.30865/jurikom.v9i1.3838.



Forecasting Blood Demand Using the Support Vector Regression Method (Case Study: Blood Transfusion Unit-PMI Central Lombok)

Yati Apriati¹, Wafiah Murniati², Saikin³, Sofiansyah Fadli⁴, Hairul Fahmi⁵

^{1,2,4,5}Teknik Informatika, STMIK Lombok, Praya, Indonesia

³Sistem Informasi, STMIK Lombok, Praya, Indonesia

Email: ¹apriatiyati750@gmail.com, ²wafiah.mr@gmail.com, ³eken.apache@gmail.com, ⁴sofiansyah182@gmail.com, ⁵iroel.ami@gmail.com

Abstract – Blood is an important component produced by the human body. Blood is also a very vital part of human survival. When blood levels in the human body are less than they should be, the way to overcome this is by donating blood or blood transfusion. The health facilities that organize blood donations, provide blood and distribute blood are called Blood Transfusion Units (UTD). UTD in carrying out its duties encountered several obstacles, such as blood only having a shelf life of 35 days from donation. If it has passed the expiration date, it cannot be used anymore for blood transfusions. Meanwhile, regarding the demand for blood, the need for blood is greater than those donating. Making it difficult for UTD if the demand occurs when the existing blood stock is not sufficient. And if the stock in UTD experiences an excess, it can cause losses because the blood is wasted due to expiration. Apart from that. The problem is that in everyday life, many people's need for blood is reduced. Many of their families intervened directly to find available donors. They even search on social networks or social media such as WhatsApp, Facebook, Instagram and others. And this shows that many of them lack donors. To anticipate these problems. So it is necessary to carry out research on forecasting blood demand using the Support Vektor Regression method at UTD PMI Central Lombok. The aim of this research is to forecast or predict the demand for blood at UTD PMI Central Lombok in the coming period. To reduce the impact of lack or excess blood. SVR is the application of Support Vektor Machine (SVM) in the case of regression to find the best dividing line in the regression function. The advantage of the SVR model is that it can handle overfitting problems in the data. The tests used to measure the best model are Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE) and Coefficient of Determination (R²). The results of this research shows that the best model is Support Vektor Regression (SVR) with a polynomial kernel and based on the tuning results, the parameters used are C=10, degree=1, epsilon=1. The SVR model using a polynomial kernel produces a MAPE value of 18.7502% and RMSE value of 0.6919, which means the model has very good predictive ability. Prediction accuracy was achieved with an R² value of 0.9936 or 99.36% and an MSE value of 0.4787, which means that the prediction of blood demand data at UTD PMI Central Lombok using SVR with a polynomial kernel function had very good prediction accuracy. With predicted result in january for blood type A it was 1654, B was 920, O was 2205 and AB was 1104.

Keywords: blood, blood transfusion units (UTD), Support Vektor Regression (SVR), MAPE, RMSE, MSE, R²

I. INTRODUCTION

Blood is an important component produced by the human body. Blood is also the important part of human survival. In general, the function of blood is as a means of transportation in the body such as O₂, CO₂, nutrition, metabolism, hormones, heat and body immunity [1]. When blood level in the human body are less than they should be, the way to overcome this is by donating blood or blood transfusion [2]. Blood transfusion is the process of donating blood voluntarily by someone, which will later be used to help people in need [3][4].

In Indonesian, the need for blood continues to increase every year, this happens because blood is needed for accidents and planned operations. Blood donation is a very important thing but is often forgotten. The Indonesian Red Cross an one of the blood providers is currently experiencing limitations due to a lack of blood supply for patients, because according to WHO (World Health Organization), ideally blood availability is 2% of the population, meaning that Indonesian currently needs 4.6

million bags of blood but PMI only received less than 0.5% of this need in 2005 [5][6]. In accordance with Government Regulation Number 7 of 2011 concerning Blood Transfusion Services, Minister of Health Regulation Number 83 of 2014 concerning Blood Transfusion Units and Minister of Health Regulation Number 14 of 2021 concerning Related Standards relating to commercial activities and products. In implementing risk-based commercial licensing. In the health sector, the health service facility that organizes blood donations, provides blood and distributes blood is the Blood Transfusion Unit (UTD) [[7][8]. UTD in organizing blood donations, providing and distributing blood, there are several obstacles such as the nature of blood which has a limited shelf life [9]. Or is easily damaged [10]. Quantity requested and amount of blood available at UTD. For example, blood has as a shelf life of 35 days from donation. If it has passed the expiration date, it cannot be used anymore for blood transfusion [11]. Meanwhile, regarding the demand for blood, the need for blood is greater than those donating, making it difficult for UTD if the demand occurs when the



existing blood stock is not sufficient. And if the stock in UTD experiences an excess, it can cause losses because the blood is wasted due to expiration [12]. Apart from that, the problem is that in everyday life, many of their families intervened directly to find willing donors. They even search on social networks or social media such as WhatsApp, Facebook, Instagram and others. And this shows that many of them have not had their blood needs met.

To anticipate losses if there is an excess or shortage of blood, forecasting is necessary. Forecasting is the activity of estimating, predicting future events to provide blood reserves to meet blood needs in the future [13]. The method used is Support Vektor Regression (SVR). The application of Support Vektor Machines (SVM) in regression cases with output in the form of real numbers or sequential data is called SVR [14]. SVR is an algorithm that can overcome the problem of overfitting where a good model is produced from training data and not test data. A data that at the training stage produces almost perfect prediction accuracy values is called Overfitting [15].

Research conducted by [16] with the title Support Vektor Regression for Forecasting Blood Demand: Case Study of UTD Branch-PMI Malang City to Determine Optimal Norms for SVR Parameters in predicting Blood Demand in PMI Malang City. Based on the results of blood data simulations for the 2010-2014 period, a minimum MAPE value of 3,899% was obtained with parameter values

lambda=10, sigma=0.5, cLR= 0.01, C=0.1, epsilon=0.01, number of data features=4 and the number of iterations is 5000, from 12 test data used. The resulting MAPE value is <10% and can be said to be good for predicting the amount of blood needed. Meanwhile, research was also carried out regarding forecasting blood demand which was carried out by [17] by taking a case study in Malang City. In his research, it was concluded that based on the experimental result obtained, the accuracy value produced by the system in the Fuzzy Time Series Interval Optimization research using Particle Swarm Optimization in predicting blood needs was 92.49670% with an error rate (MAPE) of 7.50330% which was obtained from the error calculation results from actual data with forecasting results on 12 test data. In addition, research conducted by [18] used the SVR method to predict monthly rainfall. The SVR prediction results obtained show that the SVR prediction method is very effective in predicting rainfall with forecast accuracy obtained by testing the RMSE value against the SVR model parameter values. The best RMSE is 0.038800637 with gamma of 0.0005, C of 0.0001, and epsilon of 1. Based on the problems presented and research references, the author proposes research entitled “forecasting blood demand using the Support Vektor Regression method (Case Study: Blood Transfusion Unit – PMI Central Lombok)”. Hopefully the results of this research can help blood transfusion units adjust blood stocks.

support the research.

2.2 Data Collection

This data collection was carried out by conducted a survey directly at the research site or field, namely UTD PMI Central Lombok. The data used in this research was incoming and outgoing blood data from March 2021 to December 2022 at UTD PMI Central Lombok.

No	Desa	Nama	Usia	Gol	Tanggal_Masuk	Jumlah_Masuk	Total	
0	1	KOPANG	NaN	NaN	A	2021-03-01	3	4
1	2	KOPANG	NaN	NaN	B	2021-03-01	3	4
2	3	KOPANG	NaN	NaN	O	2021-03-01	7	10
3	4	KOPANG	NaN	NaN	AB	2021-03-01	1	2
4	5	TIBU SISOK	NaN	NaN	A	2021-03-03	7	8
...
2700	2701	KOPANG	NaN	NaN	AB	2022-12-31	3	4
2701	2702	DD GEDUNG	NaN	NaN	A	2022-12-02	47	51
2702	2703	DD GEDUNG	NaN	NaN	B	2022-12-05	44	45
2703	2704	DD GEDUNG	NaN	NaN	O	2022-12-10	129	130
2704	2705	DD GEDUNG	NaN	NaN	AB	2022-12-08	6	7

2705 rows x 8 columns

Fig 2 Incoming Blood Data

2.3 Data Preprocessing

Data preprocessing is the stage for carrying out an initial process in data processing. Data Preprocessing aims to convert raw data into quality data so that the data is suitable for processing at the next stage. The stages in data preprocessing are [19]:

a. Data Integration

Data integration is the process of combining data from various databases to become one new database. The data needed in the data mining process does not only come from one database but can also come from several databases.

II. RESEARCH METHODOLOGY

The steps taken in this research is began with a literature rivev, collecting data, pre-processing data, modeling, evaluating, implementing, and drawing conclusions. These processes shown in the diagram below in figure 1.

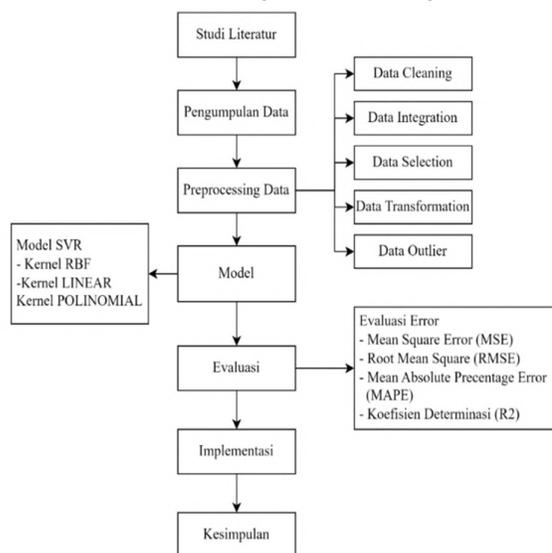


Fig 1. Research Steps

2.1 Literature Review

at this stage, a literature study is carried out or collecting information based on references from books, journals, and some research related to the study, a national and international research. Literature studies also discuss theories that are relevant to the research and



```
[33]: df=pd.read_csv('masuk.csv',sep=',')
      df1=pd.read_csv('keluar.csv',sep=',')

      data= pd.merge(df,df1, on='No')
      data.shape
```

[33]: (2785, 15)

Fig 3. Integration Data Source Code

b. Data Cleaning

Cleaning is the process of cleaning the data that will be used from unnecessary characters and even words. Data cleaning is carried out to clean unvalid data. unvalid data is data that in a dataframe has an empty value and the data has a Nan value. In the incoming and outgoing data, there are some blanks and some values, such as the name feature, age feature, aptave date feature and pocket number feature, have empty data. As in Figure 4 below, it shows missing data.

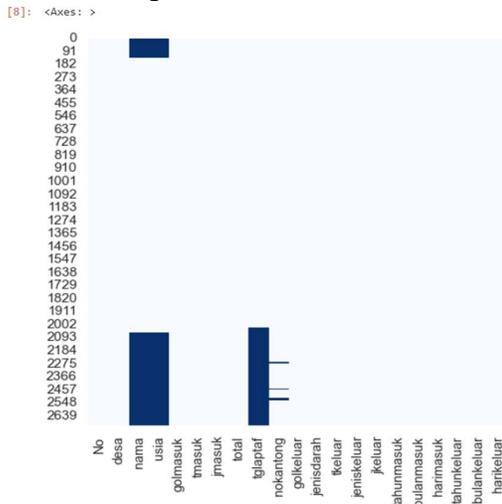


Fig 4. Data Missing

There are 2 steps used to clean a missing data, namely deleting missing features and unimportant features and inputting the average or mean value for features with few missing values. Here the features that are deleted such as village, aptaf date and pocket number because apart from having missing values, these features are also not very important. and the age feature is input based on the average value.

c. Data Selection

Data selection is data from a dataset which is be processed later. This data selection aims to select attributes that are considered to be attributes that influence blood data classification.

```
[31]: data= data[['usia','golmasuk','bulanmasuk','jmasuk','golkeluar','bulankeluar','jenisdarah','jkeluar','total']]
      data.head()
```

```
[31]:
```

	usia	golmasuk	bulanmasuk	jmasuk	golkeluar	bulankeluar	jenisdarah	jkeluar	total
0	36.376691	A	3	3	A	3	PRC	1	4
1	36.376691	B	3	3	B	3	PRC	1	4
2	36.376691	O	3	7	O	3	PRC	3	10
3	36.376691	AB	3	1	B	3	PRC	1	2
4	36.376691	A	3	7	O	3	PRC	1	8

Fig 5. Selection Data

d. Data Outlier

Outliers are data that deviate too far from other data in a data series (outliers). The existence of outliers will bias the analysis of a series of data.

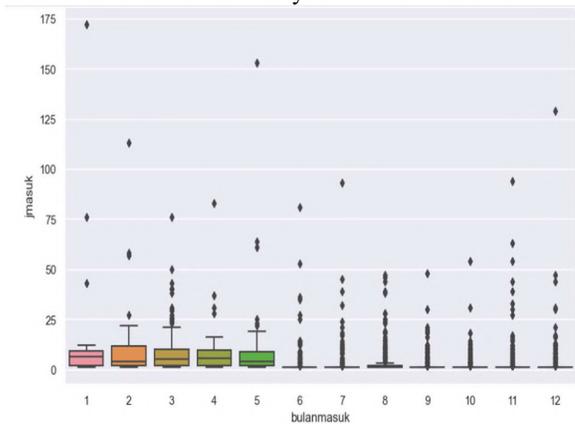


Fig 6. Outlier Data

e. Data Transformation

Transformation is the stage of making the data to be processed according to the model or algorithm that you want to use in the data processing stage. Data that past the data selection and data cleaning will be transformed, aiming to the distribution of the data so that the data becomes normal data and is easier to group. Data transformation here means changing categorical values into numerical values by utilizing the source code provided by Sklearn, such as the label encoder. In Figure 7, the columns that are transformed are the goal-in column, the goal-out column, the blood type column and the age column. The age column is binned first, after which it continues to the transformation stage.

```
from sklearn.preprocessing import LabelEncoder
from sklearn.compose import ColumnTransformer

le= LabelEncoder()
data['golmasuk'] = le.fit_transform(data['golmasuk'])
data['golkeluar'] = le.fit_transform(data['golkeluar'])
data['jenisdarah'] = le.fit_transform(data['jenisdarah'])
data['usia'] = le.fit_transform(data['usia'])
data.head()
```

Fig 7. Transformation Data Source Code

2.4 SVR Models



In this SVR model there are three kernels, namely the linear kernel, polynomial kernel and radial basis function (RBF) kernel. These three kernels will be tested using train data and good results will be selected for forecasting. a support vector regression diagram can be seen below:

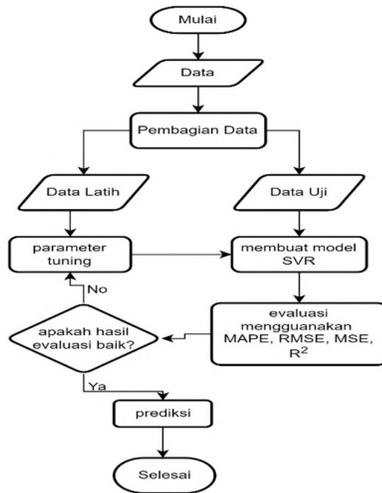


Fig 8. Support Vector Regression Steps

2.5 Evaluation

after getting good results from all three kernels. Then, an evaluation stage was carried out using MAPE, MSE, RMSE and d2. After the evaluation results for the SVR model and the three evaluation methods come out, the next stage is to make predictions. If the combination of kernel functions, parameters, and kernel evaluation used gives good results, then predictions can be made using the test data that was provided at the time of data sharing.

2.6 Implementation

This stage is carried out based on the results of tests that have been carried out previously. This implementation includes algorithm implementation and interface implementation using the Python programming language with the Jupyter notebook editor.

2.7 Conclusion

At this stage, conclusions are drawn after passing through the initial stages of literature study to system testing and analysis which the result in conclusions based on the results of the research carried out by the author. Drawing conclusions is used to answer the problem formulation that has been formulated before. And provide suggestions for further research.

III.RESULTS AND DISCUSSION

3.1 SVR Models

The data used in this research was 2075. The data was divided into 2, namely 80% training data or 2164 data, while 20% was test data amounting to 541 data. Then the training data is used to find the value of each kernel, including:

a. Linear Kernels

The linear kernel is to look for the x_{train} , y_{train} scores and x_{test} , y_{test} scores. The scores obtained

are for the x_{train} , y_{train} scores of 0.3703 and for the x_{test} , y_{test} scores of 0.4006. as in figure 9.

```
svr_linear=SVR(kernel='linear')
svr_linear.fit(x_train, y_train)

svr_linear.score(x_train, y_train),svr_linear.score(x_test,y_test)

(0.37036642225181704, 0.40062565856545373)
```

Fig 9. Source Code Linear Kernels

b. Rbf Kernels

Then, the formation of SVR uses Radial Basis Function (RBF). This is a way to look for x_{train} , y_{train} scores and x_{test} , y_{test} scores. The scores obtained are the scores for x_{train} , y_{train} of 0.2662 and the scores for x_{test} , y_{test} of 0.3210. as in figure 10.

```
svr_rbf=SVR(kernel='rbf')
svr_rbf.fit(x_train, y_train)
svr_rbf.score(x_train, y_train),svr_rbf.score(x_test,y_test)

(0.2662399274200602, 0.32104012447481844)
```

Fig 10. Source Code RBF Kernels

c. Polynomial Kernels

The next thing is to look for the x_{train} , y_{train} scores and x_{test} , y_{test} scores. The scores are the scores for x_{train} , y_{train} of 0.8131 and for x_{test} , y_{test} of 0.6856. As seen in Figure 11 below:

```
svr_poly=SVR(kernel='poly' )
svr_poly.fit(x_train, y_train)
svr_poly.score(x_train, y_train),svr_poly.score(x_test,y_test)

(0.8131136206308167, 0.6856683787743065)
```

Fig 11. Source Code Polynomial Kernels

To make it clearer, you can see in table:

Table 1 Comparison of SVR Models

SVR Models	X_train, y_train	X_test, y_test	Information
Kernel linear	0.3703	0.4006	Based on the score obtained, the linear kernel results are not good.
Kernel RBF	0.2662	0.3210	For the RBF kernel the results obtained are not good.
Kernel Poly	0.8131	0.6856	Meanwhile for poly kernels the results are much better than linear kernels and RBF kernels.

Based on table 1, it can be concluded that, based on the three kernels. The kernel used to make predictions is a polynomial kernel. After getting the best SVR model, proceed with improving the model using tuning parameters. Tuning parameters are used to increase the



score of x_{train} , y_{train} and x_{test} , y_{test} and find the best parameters. Based on the tuning results, it can be seen that $C=10$, $degree=1$, $epsilon=1$ and cross validation(cv)= 4 with a score of x_{train} , y_{train} of 0.9959 and for the score of x_{test} , y_{test} of 0.9946 with a best_score of 0.9944 . So it can be concluded that the score has increased very well.

3.2 error evaluation

In this research, the calculation of prediction accuracy uses R^2 , while the prediction error calculation uses MSE, RMSE and MAPE.

Table 2 Error Evaluation

SVR Models	C	MSE	RMS E	MAPE	R2
Kernel	1	0.406	0.637	22.1498	99.4620
polynomial	0	4	5	%	%

Based on table 2, it can be explained that in this study the error calculation was that the MSE value was 0.4064 , the RMSE value was 0.6375 and the MAPE value was 22.1498% . and the calculation of accuracy or precision of predictions using R^2 is 99.46% . It means that the SVR model with a polynomial kernel is more accurate in predicting blood demand at UTD PMI Central Lombok. After calculating the error evaluation, the next step is to predict blood demand for the next period. This is the prediction of results to be obtained for the next month:

```
model.predict(x_pred)
array([1654.51470459, 920.3394106, 2205.14617507, 1103.8832341 ])
```

Fig 12. Prediction Results

Based on the pictures above about the prediction results for the next month. Or more details can be seen below:

	usia	golmasuk	bulanmasuk	jmasuk	golkeluar	bulankeluar	jkeluar	jenisdarah	total
Januari	Remaja	A	Januari	9	A	Januari	9	PRC	1655.0
Januari	Dewasa	B	Januari	5	B	Januari	5	PRC	920.0
Januari	Remaja	O	Januari	12	O	Januari	12	PRC	2205.0
Januari	Dewasa	AB	Januari	6	AB	Januari	6	PRC	1104.0

Fig 12. DataFrame Prediction Results

Based on the pictures above, it can be explained that the prediction results are in January. For incoming blood group A of 9 and outgoing blood group A of 9, the prediction result obtained is 1655. Next, for incoming blood group B of 5 and outgoing blood group of 5, the prediction result obtained is 920. Furthermore, for the incoming blood group O of 12 and the outgoing blood group of 12, the prediction result is 2205. Meanwhile, for incoming blood group AB of 6 and outgoing blood group of AB of 6, the predicted result is 1104.

IV. CONCLUSION

Based on the discussion above, it can be concluded that the SVR model used is a poly kernel because the results shows was very good. With the parameters used, namely $C=10$, $degree=1$, $epsilon=1$, this means that the poly kernel is able to overcome the overfitting problem in predicting blood demand at UTD PMI Central Lombok. forecasting blood demand at UTD PMI Central Lombok using support vector regression with a polynomial kernel. then the MSE obtained is 0.4787 , the RMSE is 0.6919 and the MAPE value is 18.7502% , which means that the forecasting accuracy is very good or the model has very good forecasting capabilities. The R^2 value is 0.9936 or 99.36% , which means that the prediction of blood demand data at UTD PMI Central Lombok using SVR with a polynomial kernel function has very good accuracy or prediction accuracy results. With predicted results in January for blood type A of 1654, B of 920, O of 2205 and AB of 1104.

REFERENCES

- [1] Y. Heryanita *et al.*, "4. The Value of Erythrocytes, Hemoglobin, and Hematocrit of Mice (Mus musculus) Exposed to Cigarette Smoke and Given Red Watermelon Extract (Citrullus vulgaris)," *J. Med. Vet.*, vol. 12, no. 1, pp. 24–31, 2018, doi: 10.21157/j.med.vet.v12i1.4106.
- [2] G. C. Heriyanto, U. K. Petra, J. Siwalankerto, S. Ji, S. Surabaya, and J. Siwalankerto, "Prediksi Kebutuhan Darah Menggunakan Metode ARIMA Dengan Mempertimbangkan Faktor Deterioration," *J. Infra*, 2022, [Online]. Available: <https://publication.petra.ac.id/index.php/teknik-informatika/article/view/12734%0Ahttps://publication.petra.ac.id/index.php/teknik-informatika/article/viewFile/12734/11036>.
- [3] H. H. Rohan, Y. Amalia, P. Ayu, and D. Reswari, "Journal of Community Engagement in Health Kegiatan Donor Darah Di Fakultas Ilmu Kesehatan Universitas Dr. Soetomo Surabaya Tahun 2018," vol. 4, no. 2, pp. 475–480, 2021, [Online]. Available: <http://jceh.orghttps://doi.org/10.30994/jceh.v4i2.272>.
- [4] A. C. Cagliano, S. Grimaldi, C. Rafele, and C. Campanale, "An enhanced framework for blood supply chain risk management," *Sustain. Futur.*, vol. 4, no. July, p. 100091, 2022, doi: 10.1016/j.sfr.2022.100091.
- [5] B. Apriliasari, N. S. Utami, L. Lidia, and A. Abdullah, "AKSI UNIASIAH (Blood Transfusion Unit) Based on Android With Waterfall," *J. Transform.*, vol. 17, no. 1, p. 57, 2019, doi: 10.26623/transformatika.v17i1.1410.
- [6] Kemenkes RI, *PROFIL KESEHATAN INDONESIA 2018*. 2018.
- [7] K. W. D. Nugraha, *Profil Kesehatan Indonesia 2021*. 2022.
- [8] A. Datumaya, W. Sumari, A. K. Febrianto, and Y. Pramitarini, "Sistem Prediksi Permintaan Darah



- Menggunakan Metode Regresi Linier (Studi Kasus Pada UTD PMI Kabupaten Bojonegoro),” *J. Inform. Polinema*, pp. 85–90, 2021, doi: 10.33795/jip.v7i2.495.
- [9] A. Budipriyanto and F. Anggraini, “Manajemen Persediaan Darah Untuk Meminimalkan Shortage dan Wastage pada Bank Darah PMI Jakarta,” *Pros. Senantias*, vol. 1, no. 1, pp. 1235–1244, 2020.
- [10] O. S. S. Filho, W. Cezarino, and G. R. Salviano, *A Decision-making Tool for Demand Forecasting of Blood Components*, vol. 45, no. 6. IFAC, 2023.
- [11] C. Rutherford, S. Y. Cheng, and K. Bailie, “Evidence of bullwhip in the blood supply chain,” *23rd EurOMA Conf. Proc.*, no. June, pp. 1–10, 2016.
- [12] M. Meneses, D. Santos, and A. Barbosa-póvoa, “Modelling the Blood Supply Chain,” *Eur. J. Oper. Res.*, vol. 307, no. 2, pp. 499–518, 2023, doi: 10.1016/j.ejor.2022.06.005.
- [13] N. Prastyanda, P. Pratama, T. Sukmono, B. I. Putra, and W. Sulistiyowati, “Forecasting the Amount of Blood Storage Using the Support Vector Machine (Svm) Method Peramalan Jumlah Permintaan Darah Menggunakan Metode Support Vector Machine (Svm),” vol. 3, no. 07, 2022.
- [14] A. D. Dubey, “Gold price prediction using support vector regression and ANFIS models,” *Int. Conf. Comput. Commun. Informatics, ICCCI*, no. January 2018, 2018, doi: 10.1109/ICCCI.2016.7479929.
- [15] R. Melati, T. W. Purboyo, and M. Kallista, “Prediksi Penderita Tuberkulosis Menggunakan Algoritma Support Vector Regression (SVR) Prediction Of Tuberculosis Patients Using The Support Vector Regression (SVR),” vol. 10, no. 1, pp. 736–741, 2023.
- [16] M. R. Rifqi, B. D. Setiawan, and F. A. Bachtiar, “Support Vector Regression Untuk Peramalan Permintaan Darah : Studi Kasus Unit Transfusi Darah Cabang – PMI Kota Malang,” vol. 2, no. 10, pp. 3332–3342, 2018.
- [17] A. D. A. Rifandi, B. D. Setiawan, and Tibyani, “Optimasi Interval Fuzzy Time Series Menggunakan Particle Swarm Optimization pada Peramalan Permintaan Darah : Studi Kasus Unit Transfusi Darah Cabang - PMI Kota Malang,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 2, no. 7, pp. 2770–2779, 2018, [Online]. Available: <http://j-ptiik.ub.ac.id>.
- [18] N. A. Siregar, “Peramalan Curah Hujan Di Kota Medan menggunakan Metode Support Vector Regression,” vol. 1, no. 1, pp. 7–9, 2022.
- [19] E. Virantika, K. Kusnawi, and J. Ipmawati, “Evaluasi Hasil Pengujian Tingkat Clusterisasi Penerapan Metode K-Means Dalam Menentukan Tingkat Penyebaran Covid-19 di Indonesia,” *J. Media Inform. Budidarma*, vol. 6, no. 3, p. 1657, 2022, doi: 10.30865/mib.v6i3.4325.



Improving Firebase BaaS Service Security in Counseling Chat Applications: AES-256 and CBC Approach for End-to-End Encryption

Mogar Nurhandhi^{1*}, Agus Suhendar²

¹Program Studi Informatika, Fakultas Sains dan Teknologi, Universitas Teknologi Yogyakarta

²Program Studi Informatika, Fakultas Sains dan Teknologi, Universitas Teknologi Yogyakarta

Email: ¹mogarnurhandhi@gmail.com, ²agus.suhendar@staff.uty.ac.id

Abstract – The activities of using the internet and exchanging information and sending messages have used a lot of internet media, one of which is chat message media, but over time data and information security problems in chat messages that are often encountered are active and passive wiretapping. In this research, the author wants to create a counseling chat message security application using the AES (Advanced Encryption Standard) algorithm cryptographic method combined with the CBC (Cipher Block Chaining) technique which is an advanced development of the ECB (Electronic Code Book) technique. AES basically uses a block cipher with a length of 128 bits as the default operation, and the key length size varies from 128, 192 and 256 bits. so AES uses a 4x4 matrix equation with each section having a size of 1 byte. From these problems, research will be conducted to develop an application to accommodate the counseling process using a chat application that has the main focus of securing messages with image types and stored in the Firebase database service (Backend as a Service). As well as using the End-to-End service principle so that users do not need to do the encryption or decryption process directly because the process has been carried out by the system, this will also provide more security aspects in terms of confidentiality of key data and initialization vectors. So that the process of exchanging information using the media chat counseling application is secured and avoids tapping by irresponsible parties.

Keywords – Counseling, Advanced Encryption Standard, Cipher Block Chaining, Cryptography, Chatting

I. INTRODUCTION

The need for counselling for school students is one of the means that must be available in every school, this is not just because of the formality in the subjects and curriculum accompanied by the school. But counselling has a position in building the personality and mentality of student participants in experiencing various problems such as career paths or further learning [1]. The counselling process can be tried by teachers with various procedures that have existed and tried so far, but with the rapid growth of technology and data, the teaching procedures related to counselling must be able to explore this growth and one of them is by using internet-based communication technology. Based on the explanation of the case, the author wants to create a practical message application or chat by implementing the AES algorithm, which is not only able to accommodate the counselling process but also takes into account the security aspects and confidentiality of student information as a form of privacy that must be protected. The procedure that can be used in securing chat messages is cryptography, which is a scientific discipline that focuses on methods of hiding messages so that they are not known to unauthorised parties. But in the modern definition of cryptography, it is closer to the basic science in the field of informatics with the aim of securing data, protecting the integrity of information or authenticating an entity so that modern cryptography has a more complex purpose [2].

In its application, cryptographic procedures can use methods of securing messages by changing messages and hiding the meaning of messages by encryption and decryption processes. In this procedure, something is known as encryption which can be interpreted as a code or chipper. Cryptographic procedures have many encoding methods and types that can be applied in various aspects of

pc, data, or network security, not only that this procedure also allows the combination of some algorithms in it so that the system built by this procedure has a strong level of security [3]. Basically, cryptography can be categorized into two parts, symmetric cryptography and asymmetric cryptography, both of which have their own algorithmic derivatives [4]. In this research, the symmetric key cryptography method will be applied using the AES encryption algorithm. AES (Advanced Encryption Standard) is a continuation of the development of the DES (Data Encryption System) algorithm then the AES algorithm slowly acts as a replacement for DES because the DES algorithm has weaknesses in the hardware sector, namely the problem of using keys that are too long. The AES algorithm has advantages over its predecessor, namely a high security system and has several types of levels with different key lengths to support security. The AES algorithm has another advantage, namely relatively fast and efficient computing speed with 128-bit keys but also provides 192 and 256 keys for complex computations [5].

In the AES algorithm based on the shortest key length of 128-bit, the AES algorithm is resistant to exhaustive key search attacks and the AES algorithm is also proven to be able to withstand square attacks where this attack utilises the byte approach structure. If the length of AES is 128-bit then it allows this attack to accelerate iterations 6 times faster than AES, but this algorithm is able to survive with more cycles so that the resulting resilience by this type of attack [6]. Then the source of the literature review used by the author based on previous research with the same research family, among others, is the AES algorithm as a block algorithm based on the classic logic operation (XOR) which again operates on both binary states 0 and 1. This research proposes a modification by replacing the XOR operation between binary (0,1) with (#) on a 256 state table



that works with a specified 8-bit. Although it provides more complexity, this modification succeeds in providing efficient encryption and decryption processing time and provides more security against differential attacks [7]. The process of securing data in email messages with the AES algorithm method is possible using the java programming language and the netbeans platform. The encryption process mechanism is based on the transformation of SubByte, ShiftRows, MixColumns and AddAroundKey while the decryption will produce an inverse cipher of the encryption process transformation [8]. Data security in cloud computing is also an important issue with the encryption data authentication method through the conversion of cipher blocks into cipher streams by separating text into several AES-256 encryption blocks [9]. In the process of sharing and transferring data on cloud computing technology, there is a risk of data theft, therefore the application of encryption using AES in the data transfer process is a worthy preventive measure as a security system [10]. Data in the form of audio can be encrypted using the AES-128 method and allows using analogue signals in audio communication. With an FPGA (Field Programmable Gate Array) device, namely Artix-7 (xc7a200tffg1156-2L) and Kintex-7 (xc7k160tffg676-2L) as well as the VIVADO tool application. further simulation and delay configuration by reducing the column matrix operation in the AES algorithm so as to provide real-time peer to peer communication time [11]. The combination of the AES algorithm with the PRNG (Pseudorandom Number Generator) chaos system provides advantages in terms of the size of data that can be processed and the best key acquisition in securing medical image data. FPGA (Field Programmable Gate Array) devices are an option because circuitry is possible for specific algorithms as well as being dynamically programmable [12].

The AES algorithm is available in several programming language libraries and one of them is java which will be implemented in this research. In addition, the development does not only use the pure AES cryptography method, but also combined using the CBC (Cipher Block Chaining) block cipher technique, this will strengthen the system with the use of two security key parameters and initialization vector. So in this research cryptography will be applied using the java programming language with the output of android mobile applications with the AES-256 encryption algorithm.

II. RESEARCH METHODOLOGY

In completing this research, the author uses a method to design systematic system development, while the method used is the waterfall model. The waterfall model emphasises a system development scheme in an orderly and sequential manner at each stage, starting from initial identification to the testing or maintenance stage [13]. The waterfall model used in the system development process in this study can be seen in Figure 1 below.

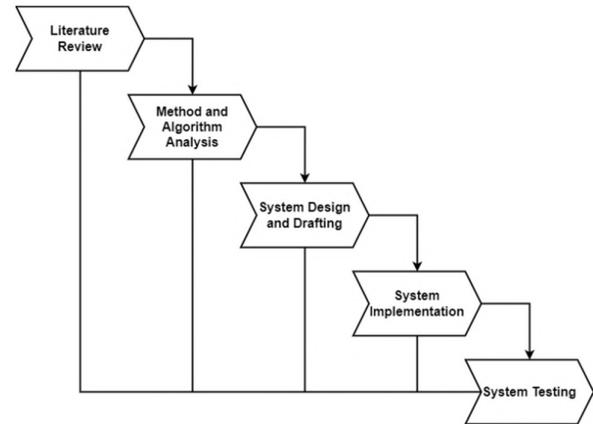


Fig. 1 research stages

A. Literature Review

The first stage in this research and its relationship to system development is a literature study, this stage is carried out by analysing and studying previously existing related research such as in journal media, proceedings, books, and other publication media. The results at this stage can be seen in the literature review in the previous chapter which explains the process and results achieved by previous researchers.

B. Method and Algorithm Analysis

In the previous chapter it was explained that cryptography methods can be broadly divided into two, namely based on symmetric and asymmetric keys. In this research, AES is included in symmetric cryptography so that both the same key will be used in the encryption process and the decryption process of a data. The AES algorithm is a type of symmetric block ciphertext algorithm in which it has a key length variation which is divided into 3 namely 128-bit, 192-bit, and 256-bit. The key length will determine the number of rounds that will be carried out in the encryption and decryption process, as for the key length and other information can be seen in table 1 below.

Table 1. AES Key Variations

AES Key	Parameter AES		
	Key Length	Block Size	Turnaround
AES-128	4	4	10
AES-192	6	4	12
AES-256	8	4	14

Before entering the main stage in the basic AES algorithm encryption process, there are several preparations that must be made, including the conversion of plaintext and cipher keys in the form of a 4x4 matrix, then both are converted in hexadecimal form and then converted again into binary. The results of the conversion mentioned earlier can be seen in some of the images below.

Plaintext				Cipher Key			
K	J	K	K	2	0	1	9
E	L	O	M	2	0	2	3
P	O	K	O	0	0	0	0
6	P	I	F	0	0	0	0

Fig 2. 4x4 Matrix Conversion



Plaintext				Cipher Key			
4B	4A	4B	4B	32	30	31	39
45	4C	4F	4D	32	30	32	33
50	4F	4B	30	30	30	30	30
36	50	49	46	30	30	30	30

Fig 3. Hexadecimal Conversion

Hexadecimal				Sub-bytes			
79	7A	7A	72	B6	DA	DA	40
77	7C	7D	7E	F5	10	FF	F3
60	7F	7B	00	D0	D2	21	63
06	60	79	76	6F	3C	B6	38

Fig 7. Sub-bytes Transformation

Plaintext				Cipher Key			
01001011	01001010	01001011	01001011	00110010	00110000	00110001	00111001
01000101	01001100	01001111	01001101	00110010	00110000	00110010	00110011
01010000	01001111	01001011	00110000	00110000	00110000	00110000	00110000
00110110	01010000	01001001	01000110	00110000	00110000	00110000	00110000

Fig 4. Binary Conversion

Next is the initial round where this stage is actually part of the add round key. XOR operation is performed between plaintext and cipher key, starting from 4B XOR 32, 45 XOR 32, 30 XOR 30, 36 XOR 30 so that the complete equation is as follows.

$$01001011 \text{ XOR } 00110010 = 01111001 \quad (1)$$

$$01000101 \text{ XOR } 00110010 = 01110111 \quad (2)$$

$$01010000 \text{ XOR } 00110000 = 01100000 \quad (3)$$

$$00110110 \text{ XOR } 00110000 = 00000110 \quad (4)$$

With the same calculation method, all binary columns of plaintext and cipher key will produce a new matrix and must be converted again into hexadecimal, as for the results can be seen in the following figure.

01111001	01111010	01111010	01110010
01110111	01111100	01111101	01111110
01100000	01111111	01111011	00000000
00000110	01100000	01111001	01110110

Fig 5. XOR Calculation Result

79	7A	7A	72
77	7C	7D	7E
60	7F	7B	00
06	60	79	76

Fig 6. Hexadecimal Conversion

After several processes have been carried out, the next stage is the main stage in the encryption process using the AES Algorithm, this stage is divided into four parts, namely sub bytes, shift rows, mix columns, and add around key.

a. Sub Bytes

Is a transformation process by utilising the S-Box (Substitution Box) table, this table is an indexing-based tool for the efficiency of the block cipher algorithm substitution process [14], as for the transformation results as follows.

5F	B9	4D	E0
CD	33	70	4A
AC	12	D6	CB
80	37	2E	90

Fig 9. Result Mix Column

b. Shift Rows

Is a row move stage where the first row is not moved, but the second to fourth row is moved consecutively between 1 to 3 bytes.. The shift result is as shown in the following image.

Sub-bytes				Shift-rows			
B6	DA	DA	40	B6	DA	DA	40
F5	10	FF	F3	10	FF	F3	F5
D0	D2	21	63	21	63	D0	D2
6F	3C	B6	38	38	6F	3C	B6

Fig 8. Shift-rows Transformation

c. Mix Columns

It is the process of multiplying the results obtained previously with a fixed polynomial matrix, as for example in row 1 column 1 in equation 5 and the calculation is in equation 6 below.

$$\begin{bmatrix} 02 & 03 & 01 & 01 \\ 01 & 02 & 03 & 01 \\ 01 & 01 & 02 & 03 \\ 03 & 01 & 01 & 02 \end{bmatrix} \times \begin{bmatrix} B6 \\ 10 \\ 21 \\ 38 \end{bmatrix} \quad (5)$$

$$02 \times B6 = 00000010 \text{ XOR } 10110110$$

$$= x \cdot x^7 + x^5 + x^4 + x^2 + x$$

$$= x^8 + x^6 + x^5 + x^3 + x^2 \text{ mod } x^8 + x^4 + x^3 + x + 1$$

$$= x^6 + x^5 + x^4 + x^2 + x + 1$$

$$= 01110111$$

$$03 \times 10 = 00000011 \text{ XOR } 00010000$$

$$= (x + 1) \cdot (x^4)$$

$$= x^5 + x^4 = 00110000$$

$$01 \times 21 = 00000001 \text{ XOR } 00100001$$

$$= 1 \cdot x^5 + 1$$

$$= 00100001$$

$$01 \times 38 = 00000001 \text{ XOR } 00111000$$

$$= 1 \cdot x^5 + x^4 + x^3$$

$$= 00111000$$

From the calculation process, the next step is to add up the results as in equation 7 below.

$$01110111 + 00110000 + 00100001 + 00111000 = 01011110 \text{ (5F)} \quad (7)$$

So with the same calculation method, all columns in the matrix will produce mix columns as follows.

where this process uses the XOR operation of the mix-columns result with the initial round as shown below.

Mix-Column				Initial-Round				
5F	B9	4D	E0	⊕	79	7A	7A	72
CD	33	70	4A		77	7C	7D	7E
AC	12	D6	CB		60	7F	7B	00
80	37	2E	90		06	60	79	76

Fig 10. XOR Operation

26	C3	37	92
BA	4F	0D	34
CC	6D	AD	CB
86	57	57	E6

Fig 11. XOR Result

The completion is adjusted to the round used based on the length of the AES key, because only as an example, the calculation only reaches the round 0 stage. To do the decryption process, it is done with the inverse transformation of all previous transformations.

e. CBC Method

The AES algorithm is one type of block cipher cryptography algorithm where in this type of cryptography there are several advanced methods or modes for security and one of them is CBC (Cipher Block Chaining). In its flow the CBC method performs the encryption process with the initialization vector parameter in its first block. Parameter IV operates the OR logic gate which will perform the XOR advanced logic operation on the next plaintext, the encryption data is generated using the process to produce the first ciphertext block. so that the new plaintext block only performs the XOR operation with the previous ciphertext output, the inverse function will be used in the next decryption stage, and the plaintext is generated. So that it can be said that this encryption mechanism is like a chain that is related to each other in each block. [15]. The use of the CBC method can provide advantages from the security aspect, where each IV will be different in each encryption session with the same key that will produce different ciphertexts., as for how the CBC method works as in the following figure 12.

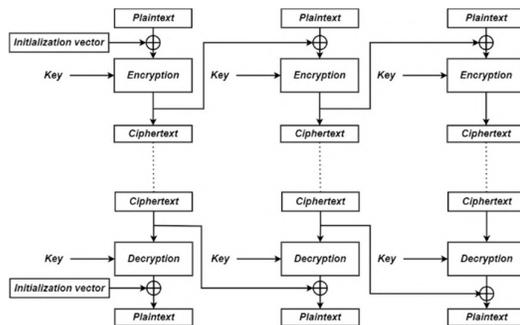


Fig 12. How the CBC Method Works

In this method originally developed in 1976 [16] known operations using XOR boolean logic as described earlier so that the basic mathematical calculation of the CBC method is as in equation 8 below.

$$C_1 = E_k(P_1 \oplus IV) , C_i = E_k(P_i \oplus C_{i-1}) , i = 2, n \quad (8)$$

C. System Design and Drafting

In the research conducted by the author in relation to the development of the application system, the author uses DFD (Data Flow Diagram) which in its implementation is displayed in two parts, namely the message encryption and message decryption processes.

a. DFD of Encryption Process

The sender enters the original image message or plaintext and goes through the encryption process into ciphertext which then continues the contact authentication process to find out the purpose of sending the chat message. Then for the process of sending messages and receiving them as ciphertext, decryption is carried out first before displaying, and more details can be seen in Figure 13 below.

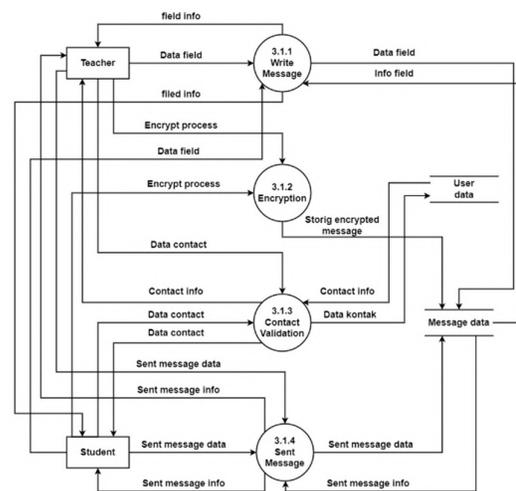


Fig 13. Encryption Process

b. DFD of Decryption Process

The following is a process that describes the reception of messages between teachers and students and vice versa up to the stage of message reading. It contains three processes that can be explained in more detail. Message reception begins. The message data sent as an encrypted message is decrypted by the system and displayed to the message recipient as a normal message, and you can see more details in Figure 14 below.

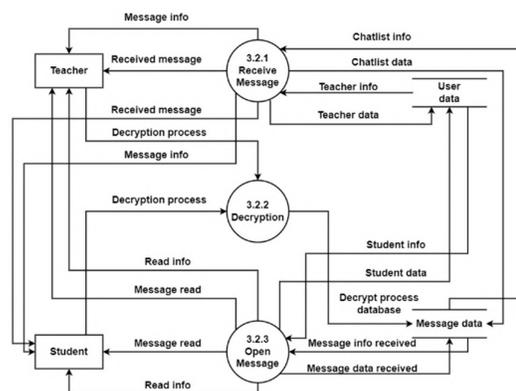


Fig 14. Decryption Process

D. Implementation System



In the implementation stage, a system will be developed to secure chat messages used in the counseling procession. The type or type of message that is the main focus in this research is the type of image or image, besides that the cryptographic method using the AES-256 algorithm is not only used independently but will also be combined using the CBC method. The developed application system produces an android mobile application output by utilizing the Backend As a Service service in the form of a Firebase database with three main types of storage, namely realtime and storage, with firebase authentication as an authentication API for every user who registers or enters the system. Thus the system will use the End-to-End User principle so that both sending or receiving parties only see plaintext data during counseling sessions. The flow of system implementation can be seen in Figure 15 below.

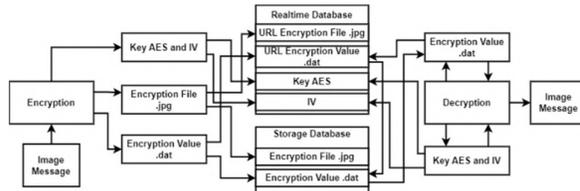


Fig 15. Schematic of Encryption and Decryption Process

E. System Testing

At this stage, testing of the application system that has been developed will be carried out, while the testing method that will be used is the blackbox testing method to test the functional structure of the encryption and decryption system. In addition to the blackbox approach, other tests will be carried out using brute force attacks to determine the resilience of the application system in the security domain. And especially to test the ability to hack the keys used in the encryption process.

III. RESULTS AND DISCUSSION

At this stage the process of securing counseling messages using the AES-256 algorithm and the CBC method will be explained in two stages, namely the encryption and decryption processes. Then proceed with the results of the implementation of the application that has been developed for the final stage tested with two test parameters.

A. Flow of Encryption Code

In the encryption process, the author uses three main functions in the script code logic, namely *generateSecretKeySpec()* to generate AES keys with a size of 32-byte randomly. Next is *encrypt()* whose role is to retrieve data, keys, and iv as a condition for performing AES encryption with a combination of CBC. The last function is *uploadTask(Bitmap bitmap)* whose role is to upload the encrypted image message data to the firebase BaaS service.

B. Flow of Decryption Code

In the decryption process, the author uses three main functions in the script code logic, namely *generateSecretKeySpec(byte[] aesKey)* to generate objects and hold the AES key. Next *decrypt(byte[] encryptedData, byte[] aesKey, byte[] iv)* whose role is to decrypt using the

AES key and iv used in the previous encryption process. And the last function *decryptAndDisplayImage(String encryptedDataURL, final ImageView imageView, final byte[] aesKey, final byte[] iv)* which has the role of downloading the encrypted image and decrypting it before displaying it again.

then for each encryption and decryption process code can be seen in the following figures 16 and 17.

```
private byte[] generateSecretKeySpec(){
    SecureRandom secureRandom = new SecureRandom();
    byte[] key = new byte[32];
    secureRandom.nextBytes(key);
    return key;
}

private byte[] encrypt(byte[] data, byte[] aesKey, byte[] iv) throws Exception {
    SecretKeySpec secretKeySpec = new SecretKeySpec(aesKey, "AES");
    Cipher cipher = Cipher.getInstance("AES/CBC/PKCS5Padding");
    IvParameterSpec ivSpec = new IvParameterSpec(iv);
    cipher.init(Cipher.ENCRYPT_MODE, secretKeySpec, ivSpec);
    return cipher.doFinal(data);
}

private void uploadTask(Bitmap bitmap){
    final String userId = intent.getStringExtra("userid");
    String folderPath = "image file";
    String filename = System.currentTimeMillis() + ".jpg";
    final StorageReference filereference = storageReference.child(folderPath).child(filename);
    ByteArrayOutputStream baos = new ByteArrayOutputStream();
    bitmap.compress(Bitmap.CompressFormat.JPEG, quality: 90, baos);
    byte[] imageData = baos.toByteArray();
    try {
        byte[] iv = new byte[16];
        SecureRandom secureRandom = new SecureRandom();
        secureRandom.nextBytes(iv);
        byte[] aesKey = generateSecretKeySpec();
        byte[] encryptedData = encrypt(imageData, aesKey, iv);
        uploadTask = filereference.putBytes(encryptedData);
        uploadTask.addOnSuccessListener(new OnSuccessListener<UploadTask.TaskSnapshot>() {
            @Override
            public void onSuccess(UploadTask.TaskSnapshot taskSnapshot) {
                filereference.getDownloadUrl().addOnSuccessListener(new OnSuccessListener<Uri>() {
                    @Override
                    public void onSuccess(Uri uri) {
                        myUri = uri.toString();
                        sendEncryptedImageMessage
                            (fuser.getUid(), userId, myUri, encryptedData, iv, aesKey);
                    }
                });
            }
        });
    }
}

private void sendEncryptedImageMessage(String sender, String receiver, String imageURL,
    byte[] encryptedData, byte[] iv, byte[] aesKey){
    final String userId = intent.getStringExtra("userid");
    final String folderPath = "encrypted_data_image";
    final String filename = System.currentTimeMillis() + ".dat";
    final StorageReference filereference = storageReference.child(folderPath).child(filename);
    UploadTask uploadTask = filereference.putBytes(encryptedData);
    uploadTask.addOnSuccessListener(new OnSuccessListener<UploadTask.TaskSnapshot>() {
        @Override
        public void onSuccess(UploadTask.TaskSnapshot taskSnapshot) {
            filereference.getDownloadUrl().addOnSuccessListener(new OnSuccessListener<Uri>() {
                @Override
                public void onSuccess(Uri uri) {
                    String encryptedDataURL = uri.toString();
                    DatabaseReference reference = FirebaseDatabase.getInstance().getReference();
                    HashMap<String, Object> hashMap = new HashMap<>();
                    hashMap.put("sender", sender);
                    hashMap.put("receiver", receiver);
                    hashMap.put("message", "");
                    hashMap.put("type", "image");
                    hashMap.put("imageUrl", imageUrl);
                    hashMap.put("encryptedDataURL", encryptedDataURL);
                    hashMap.put("isseen", false);
                    hashMap.put("dataKey", Base64.getEncoder().encodeToString(aesKey) + ":" +
                        Base64.getEncoder().encodeToString(iv));
                    reference.child(pathString: "Chats").push().setValue(hashMap);
                }
            });
        }
    });
}

private SecretKeySpec generateSecretKeySpec(byte[] aesKey) {
    return new SecretKeySpec(aesKey, "AES");
}

private byte[] decrypt(byte[] encryptedData, byte[] aesKey, byte[] iv) throws Exception {
    SecretKeySpec secretKeySpec = generateSecretKeySpec(aesKey);
    Cipher cipher = Cipher.getInstance("AES/CBC/PKCS5Padding");
    IvParameterSpec ivSpec = new IvParameterSpec(iv);
    cipher.init(Cipher.DECRYPT_MODE, secretKeySpec, ivSpec);
    return cipher.doFinal(encryptedData);
}
```

Fig 16. Encryption Code Process

```
private SecretKeySpec generateSecretKeySpec(byte[] aesKey) {
    return new SecretKeySpec(aesKey, "AES");
}

private byte[] decrypt(byte[] encryptedData, byte[] aesKey, byte[] iv) throws Exception {
    SecretKeySpec secretKeySpec = generateSecretKeySpec(aesKey);
    Cipher cipher = Cipher.getInstance("AES/CBC/PKCS5Padding");
    IvParameterSpec ivSpec = new IvParameterSpec(iv);
    cipher.init(Cipher.DECRYPT_MODE, secretKeySpec, ivSpec);
    return cipher.doFinal(encryptedData);
}
```



```

if (chat.getType().equals("image")) {
    holder.img_message.setVisibility(View.VISIBLE);
    String dataKey = chat.getDataKey();
    String[] keys = dataKey.split("regex:.*");
    byte[] aesKey = Base64.getDecoder().decode(keys[0]);
    byte[] iv = Base64.getDecoder().decode(keys[1]);
    decryptAndDisplayImage(chat.getEncryptedDataURL(), holder.img_message, aesKey, iv);
} else {
    holder.img_message.setVisibility(View.GONE);
}
private void decryptAndDisplayImage
(String encryptedDataURL,
 final ImageView imageView, final byte[] aesKey, final byte[] iv) {
    StorageReference storageReference =
        FirebaseStorage.getInstance().getReferenceFromUrl(encryptedDataURL);
    try {
        final File localFile = File.createTempFile( prefix "image", suffix ".dat");
        storageReference.getFile(localFile).addOnSuccessListener
            (new OnSuccessListener<FileDownloadTask.TaskSnapshot>() {
                @Override
                public void onSuccess(FileDownloadTask.TaskSnapshot taskSnapshot) {
                    byte[] encryptedData = readBytesFromFile(localFile);
                    try {
                        byte[] decryptedData = decrypt(encryptedData, aesKey, iv);
                        Bitmap bitmap =
                            BitmapFactory.decodeByteArray(decryptedData,
                                0, decryptedData.length);
                        Glide.with(mContext).load(bitmap).into(imageView);
                    } catch (Exception e) {
                        e.printStackTrace();
                    }
                }
            });
    }
}

```

Fig 17. Decryption Code Process

C. Results Of System Implementation

The results of the implementation of the developed system can be seen in the following interfaces starting from the initial registration login page to the main focus of sending counseling messages with the type of image or image, until at the end it will be shown the acquisition of encryption results in the form of encrypted file data and the encryption value of the data.

a. Login Interface

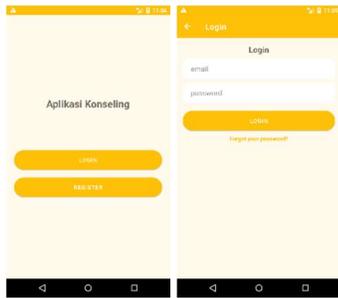


Fig 18. Interface Login Page

b. Registration Interface

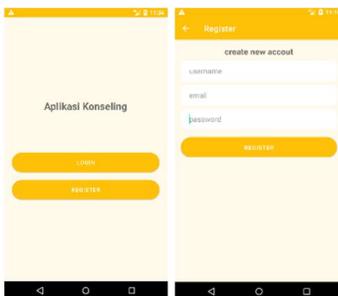


Fig 19. Interface Registration Page

c. Reset Password Interface

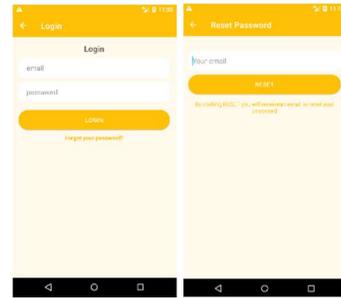


Fig 20. Interface Reset Password Page

d. Dashboard User Interface

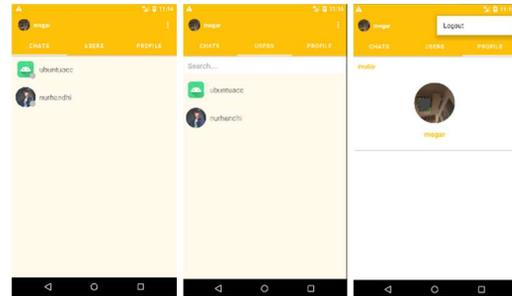


Fig 21. Interface Dashboard

e. Chat Room Interface

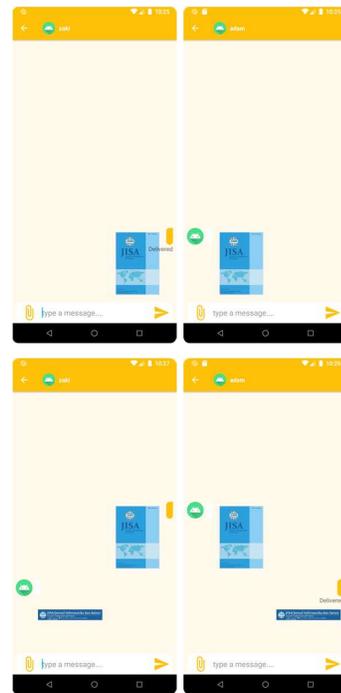


Fig 22. Interface Chat Room

f. Encryption Result

From the results of sending messages with the type of image or image above, the system automatically carries out the encryption and decryption process. This process produces encryption data stored in the firebase database, as for the encryption data as follows.





Fig 23. Realtime Database

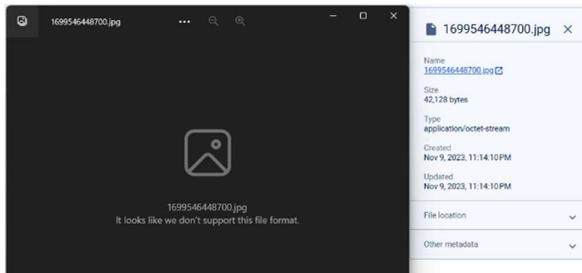


Fig 24. First Encrypted Image Message (.jpg)

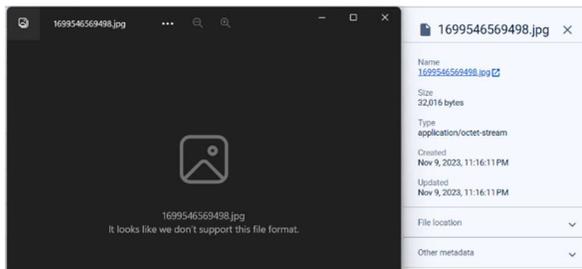


Fig 25. Second Encrypted Image Message (.jpg)

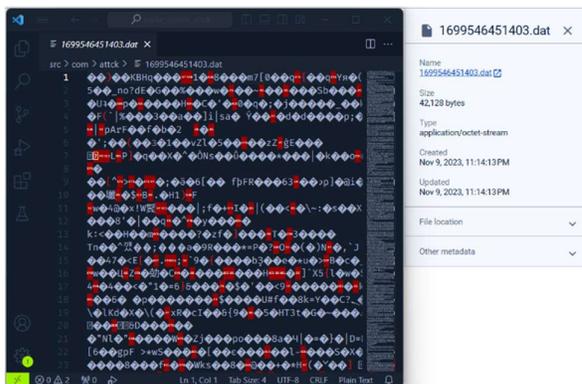


Fig 26. First Encryption Value File (.jpg)

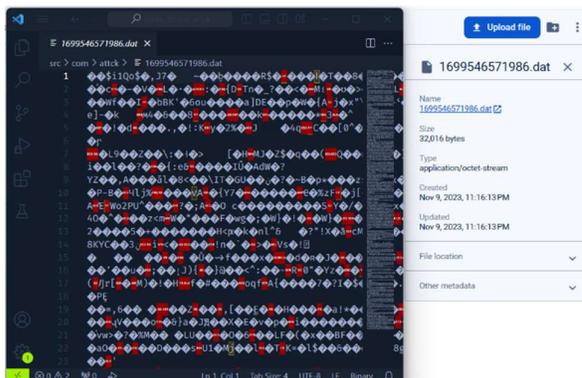


Fig 27. Second Encryption Value File (.jpg)

Apart from the encryption results in the form of encrypted image messages and encryption values above, the realtime database also stores the AES key and initialization vector used as parameters for the encryption and decryption process, as for the data can be seen in table 2 below.

Table 2. Key Acquisition and IV

No	AES Parameters	
	AES Key	Initialization Vector
1.	G70iFX/AANyfidg5yYncIIG8 7HKl7DVakMwPxgmryA= x/TQ16Y5/yDbak/2Am8T	9GKhuUWHj8 HjKfz1qq8SJw= T6b36chMVu
2.	9dzivgDrMI6UjtHgoKMUXZM=	GqtBZY9QV/zQ=

D. Results Of System Testing

In this testing stage, two methods will be carried out, namely blackbox and brute force attacks, the results of the testing method are as follows.

Table 3. Blackbox Testing

Testing Schema	Estimated Results	Results
Trying the registration process	The system displays the register form and registers the user	Correct
Trying the login process	The system displays the login form and authenticates the user.	Correct
Sorting users on the search tab	The system displays user search results based on keywords	Correct
Trying to personalize the profile	The system displays the profile tab and opens the file manager to update the profile.	Correct
Create a new message	The system opens the gallery before the image message is sent and encrypted	Correct
Attempting to encrypt a message	The system encrypts the image message using the key and IV	Correct
Storing message data in the database	The system stores the AES key and IV data in the realtime database, and stores the encrypted image and encryption result value in the storage database.	Correct
Send a message	The system sends a specific encrypted image message based on the registered user id.	Correct
Attempt the message decryption process	The system decrypts the encrypted image message using the key and IV used in the previous encryption process.	Correct
Display a message to the end-user	The system displays the decrypted image message again.	Correct
Send message notification	The system sends a notification of the encrypted image message according to the message destination.	Correct
Attempt message notification decryption	The system decrypts the image message notification using the same key and IV	Correct
Display a message notification to the end-user	The system displays a decrypted image message notification	Correct

Based on the tests in Table 3, the results obtained show that the tested system functionality can run well. Testing is



done with several scenarios on the main functionality of the application system developed.

Table 4. Brute Force Testing

AES Key	Key Size	Possible Number of Keys	Valid Key
128 bit	16 byte	3.402823669209385E38	0
192 bit	24 byte	6.277101735386681E57	0
256 bit	32 byte	1.157920892373162E77	0

Based on Table 4, the results of system robustness testing are obtained using the brute force attack test. The test case is done by trying all possible AES key lengths and counting the number of possible keys. because the number of possible keys is very large, it is written in scientific notation. Despite finding the number of possible keys, it failed to find a valid key for the encryption process.

IV. CONCLUSION

The conclusion obtained from this research is the application of a combination of the AES 256 algorithm and the CBC method on the BaaS service security system in the chat counseling application can run well. Combining these two methods produces stronger encrypted data because the use of initialization vectors and message authentication are additional security parameters. In addition, based on testing conducted with the blackbox method, it shows that the system functionality has worked properly and in accordance with their respective parameters. Testing the system against brute force attacks also shows strong resistance, so that valid keys cannot be guessed. Further research development can further include the use of double encryption on encrypted data to increase system data security and reduce the risk of data leakage.

REFERENCES

- [1] Alimzhanova, Z., Skublewska-Paszowska, M., & Nazarbayev, D. (2023). Periodicity Detection of the Substitution Box in the CBC Mode of Operation: Experiment and Study. *IEEE Access*, 11, 75686–75695. <https://doi.org/10.1109/ACCESS.2023.3295909>
- [2] Al-Mashhadani, M., & Shujaa, M. (2022). IoT security using AES encryption technology based ESP32 platform. *Int. Arab J. Inf. Technol.*, 19(2), 214–223.
- [3] Arianto, B., Kurniadi, H., & Kurniasari, I. (2023). IMPLEMENTASI PENGARSIPAN ELEKTRONIK MENGGUNAKAN ENKRIPSI DAN DEKRIPSI DENGAN METODE AES DI UNISKA. *JURNAL FASILKOM*, 13(02), 259–268. <https://doi.org/10.37859/jf.v13i02.5060>
- [4] Gupta, M., & Sinha, A. (2021). Enhanced-AES encryption mechanism with S-box splitting for wireless sensor networks. *International Journal of Information Technology*, 13(3), 933–941. <https://doi.org/10.1007/s41870-021-00626-w>
- [5] Hafsa, A., Gafsi, M., Malek, J., & Machhout, M. (2021). FPGA Implementation of Improved Security Approach for Medical Image Encryption and Decryption. *Scientific Programming*, 2021, 6610655. <https://doi.org/10.1155/2021/6610655>
- [6] Hidayat, A. (2022). Application of the AES Cryptographic Algorithm for E-mail Encryption and Description. *INFOKUM*, 10(5), 494–500. <http://infor.seaninstitute.org/index.php/infokum/article/view/1001>
- [7] Hidayat, T. (2019). ENCRYPTION SECURITY SHARING DATA CLOUD COMPUTING BY USING AES ALGORITHM: A SYSTEMATIC REVIEW. *TEKNOKOM*, 2(2), 11–16. <https://doi.org/10.31943/teknokom.v2i2.39>
- [8] K. R., R., Aithal, G., Shetty, S., & K., B. (2020). Image encryption scheme in public key cryptography based on cubic pells quadratic case. *Indonesian Journal of Electrical Engineering and Computer Science*, 20(1), 385. <https://doi.org/10.11591/ijeecs.v20.i1.pp385-394>
- [9] Kareem, S. M., & Rahma, A. M. S. (2021). New method for improving add round key in the advanced encryption standard algorithm. *Information Security Journal: A Global Perspective*, 30(6), 371–383. <https://doi.org/10.1080/19393555.2020.1859654>
- [10] Khaliq, K. F. (2021). Pengamanan Data Akta Dengan Metode Aes Berbasis Cloud Computing. *JURNAL TEKNOLOGI DAN ILMU KOMPUTER PRIMA (JUTIKOMP)*, 4(1), 509–512.
- [11] Kumar, K., Ramkumar, K. R., & Kaur, A. (2022). A lightweight AES algorithm implementation for encrypting voice messages using field programmable gate arrays. *Journal of King Saud University - Computer and Information Sciences*, 34(6), 3878–3885. <https://doi.org/10.1016/j.jksuci.2020.08.005>
- [12] Muttaqin, K., & Rahmadoni, J. (2020). Analysis And Design of File Security System AES (Advanced Encryption Standard) Cryptography Based. *Journal of Applied Engineering and Technological Science (JAETS)*, 1(2), 113–123. <https://doi.org/10.37385/jaets.v1i2.78>
- [13] Nida, K., & Usiono, U. (2023). Peranan Bimbingan dan Konseling Dalam Pembentukan Karakteristik Siswa. *Jurnal Pendidikan Dan Konseling (JPDK)*, 5(3), 64–72.
- [14] Panagiotou, P., Sklavos, N., Darra, E., & Zaharakis, I. D. (2020). Cryptographic system for data applications, in the context of internet of things. *Microprocessors and Microsystems*, 72, 102921. <https://doi.org/10.1016/j.micpro.2019.102921>
- [15] Rantellinggi, P. H., & Saputra, E. (2020). Algoritma Kriptografi Triple Des dan Steganografi LSB sebagai Metode Gabungan dalam Keamanan Data. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 7(4), 661. <https://doi.org/10.25126/jtiik.2020741838>
- [16] Rifki Sadikin. (2012). *Kriptografi untuk Keamanan Jaringan* (Th. Arie Prabawati, Ed.). Andi Publisher .



Comparison of ANN Backpropagation Algorithm and Random Forest Regression in Predicting the Number of New Students

Padmavati Darma Putri Tanuwijaya¹, Jhonatan Laurensius Tjahjadi², Yosefina Finsensia Riti^{3*}

¹Program Studi Ilmu Informatika, Fakultas Teknik, Universitas Katolik Darma Cendika Surabaya

²Program Studi Ilmu Informatika, Fakultas Teknik, Universitas Katolik Darma Cendika Surabaya

³Program Studi Ilmu Informatika, Fakultas Teknik, Universitas Katolik Darma Cendika Surabaya

Email: ¹padmavati.tanuwijaya@student.ukdc.ac.id, ²jhonatan.tjahjadi@student.ukdc.ac.id, ^{3*}yosefina.riti@ukdc.ac.id

Abstract – Higher education institutions are educational units located at a higher level after high school or vocational school. Catholic University Darma Cendika Surabaya (UKDC) faces challenges in managing the admission of new students due to variations in the number of prospective students applying to each department, which is also influenced by changing trends in interests and job needs in Indonesia. The use of Artificial Neural Network with Backpropagation and Random Forest Regression algorithms for comparing the prediction of new student admissions in the following year will be beneficial for the administration of Catholic University Darma Cendika Surabaya (UKDC) to gain a clearer understanding of the dynamics of admissions and to support decision making in the future development of the university. The predicted number of students joining Catholic University Darma Cendika Surabaya (UKDC) in the 2024 period using Artificial Neural Network is 219 students with a Mean Squared Error (MSE) of 0,1046 and Root Mean Square Error (RMSE) of 0,32.

Keywords – Artificial Neural Network, Backpropagation, Random Forest Regression, MSE, RMSE.

I. INTRODUCTION

In a university there are three important roles in its implementation, namely: carry out education, carry out research, and carry out community service [1]. The new student admission process is the starting point of the University in forming a diverse academic community and has the potential to develop their interests and ambitions, this is a key aspect of higher education administration that will affect the continuation and academic development of the University [2]. Darma Cendika Catholic University (UKDC) has a variety of variations in new student admissions each year such as the number of prospective new students who register in each department, this must be considered so that new student admissions can increase the following year. Darma Cendika Catholic University can prepare by predicting the number of students who will register in the following year by predicting it. In this study, researchers predicted new students in the following year using the Artificial Neural Network and Random Forest Regression algorithms using historical data on student admissions from 2017-2023.

Artificial Neural Network (ANN) is an information processing system that can produce predictive values from data with high accuracy using artificial neural networks that model complex relationships between inputs and outputs so that they resemble the characteristics of human nerves. ANN is also able to identify patterns in data when processing large datasets [3]. Meanwhile, Random Forest Regression is a regression learning algorithm used to produce predictive values more strongly and accurately than a single model. Random Forest Regression is suitable for solving various types of problems [4].

Previous research has illustrated the progress of the use of Artificial Neural Network (ANN) application in designing optimization of rice production prediction using multilayer [5], application in predicting nickel ore production [6], use in predicting rainfall [7], use in predicting student graduation predicate [3], use in predicting stock prices [8] and Random Forest Regression use in predicting the number of class participants for schedule planning [4], use in predicting house prices [9], use in predicting coffee quality [10], use in predicting Ciliwung river water quality [11], use in predicting cell phone prices [12] for different algorithm applications. However, each researcher has various tests, at this time there is still no one who uses Artificial Neural Network and Random Forest Regression models as a comparison to predict the number of new students at Darma Cendika Catholic University Surabaya which is a research factor with a different context.

In this research, we aim to achieve a more accurate and relevant prediction comparison with our university situation. The purpose of this research is to provide predictions using Artificial Neural Network and Random Forest Regression that are useful for the administration of Darma Cendika Catholic University Surabaya to provide a clearer view of the dynamics of enrollment in support of decision making on the future development of the university.

II. RESEARCH METHODOLOGY

In this research, the process began with the collection of a dataset. For gathering data on new students, we requested information from the admission office of Darma Cendika Catholic University in Surabaya. Subsequently, we



conducted data cleaning to remove unnecessary information, ensuring that the resulting data would yield optimal results. After the data cleaning process, we proceeded to implement the data into the prediction program we had developed. The process flow is depicted in Figure 1.

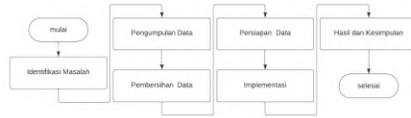


Fig 1. Flowchart of the Research Process.

1. Problem Identification

Problem identification is crucial for formulating the research focus. The identified issues include:

- Accessibility of data collection through the administration of New Student Admissions.
- Quality of the provided data.
- Variables influencing the data results.

2. Data Collection

Data collection involved obtaining the dataset for new student admissions from 2017 to 2023 through the UKDC administration.

Table 1. Raw Data for New Student Admissions in 2017

No.	No. Pendaftaran	NPM	Nama	Fakultas
1	004336	17126002	Melanny Natalia	Ekonomi
2	004321	17120003	Hanna Kristina	Ekonomi
3	004342	17120005	Bernadeta Dete	Ekonomi
4	004341	17126003	K. Mega Lestari	Ekonomi
5	004339	17120004	Trisha Claudia Prinzessa F	Ekonomi

Kelas	Gel	Jalur	Alamat
Sore	Gel 1	Umum	Perum. Taman Pondok Jati H/17
Pagi	Gel 1	JMLA	Jl. Mojoarum IX/15
Pagi	Gel 1	Umum	Waru Gunung RT. 05, RW. 03, Karang Pilang, Surabaya
Sore	Gel 1	JMLA	Jl. Kapas Baru 2 / 108, Surabaya
Pagi	Gel 1	JMLA	Deltasari Indah Blok AV-36, Sidoarjo

Kota	Propinsi	Pulau	Asal Sekolah
Sidoarjo	Jawa Timur	Jawa	SMAK ST. Yusuf Karang Pilang
Surabaya	Jawa Timur	Jawa	SMAK Kr. YBPK 1, Surabaya
Surabaya	Jawa Timur	Jawa	SMAK ST. Yusuf Karang Pilang
Surabaya	Jawa Timur	Jawa	SMK Adhikawacana, Surabaya
Sidoarjo	Jawa Timur	Jawa	SMAK Untung Surapati, Sidoarjo

3. Data Cleaning

Data cleaning involves making changes to the dataset, addressing issues such as handling missing data, ensuring data consistency, and other necessary adjustments. This step is crucial during the data preparation process, allowing the transformation of data into a model dataset ready for analysis and comprehensible to the researcher [6].

4. Data Preparation

In this study, a division of 80-20 is employed, with 80% as the training dataset and 20% as the test data. Subsequently, data normalization and scaling are performed to ensure that the data has a suitable range and characteristics, aiming to enhance accuracy and reduce the potential for data leakage [13].

5. Student Prediction Process

The prediction of the number of new students involves the application of Artificial Neural Network and Random Forest Regression through several stages of the process.



Fig. 2 Flowchart of Student Prediction Process.

6. Prediction

Prediction is an effort or action in which someone anticipates or estimates something in the future by utilizing relevant information from previous periods [4].

7. Artificial Neural Network

Artificial Neural Network is a data classification model that shares a concept almost similar to the human brain's neural system. The goal of an Artificial Neural Network is to enable a computer to have cognitive thinking abilities, mimicking the way the human brain operates to solve problems [3][13].

8. Backpropagation



Backpropagation, developed by Rumelhart, Hinton, and Williams around 1986, resulted in an iterative algorithm that is considered simple and easy to use [14][15]. Backpropagation is a method in artificial neural networks that uses supervised learning algorithms. It involves multiple perceptrons with many layers used to adjust weight values in hidden layers [15].

9. Random Forest Regression

Random Forest Regression is a machine learning technique that combines predictions from various algorithms to generate a random forest regression. It produces an average result from hundreds to thousands of decision trees consisting of decision nodes and leaf nodes used for sample evaluation through its test function [4].

10. MSE and RMSE

Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) are methods for measuring how far the predictions of a model are from the actual data. They provide an indication of the standard deviation of the difference between predicted and observed values, yielding optimal predictive recommendations [16][17].

$$MSE = \frac{1}{n} \sum_{k=1}^n (y_k - \hat{y}_k)^2 \tag{1}$$

$$RMSE = \sqrt{\frac{\sum_{k=1}^n (y_k - \hat{y}_k)^2}{n}} \tag{2}$$

where :

t_k = The actual data.

y_k = The predicted data.

n = Total number of observations.

11. Implementation

This research utilizes Google Colab as the text editor provided by Google. Google Colab is based on the Python programming language and is designed to be accessible online, allowing researchers to develop and build Artificial Neural Network (ANN) models for predictions more conveniently and efficiently.

The researcher implemented the Artificial Neural Network (ANN) model using the KERAS library from TensorFlow. The model consists of an input layer, 14 hidden layers, and an output layer. For the Random Forest Regressor model, the researcher used the Scikit-Learn library, which employs a tree-based algorithm suitable for regression data.

12. Measurement of Error Rate

After implementation, the process involves measuring the average error rate

using Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) [18]. These metrics provide insights into the accuracy of the prediction models, with lower values indicating better performance by minimizing the differences between predicted and actual values.

III. RESULTS AND DISCUSSION

3.1 Data Cleaning

The following Tables 2, 3, 4 present the results after the data cleaning process and grouping of the initial dataset.

Table 2. Total data of new student admissions

Prodi	2017	2018	2019	2020	2021	2022	2023
Total	173	200	282	204	173	226	211

In the results of the missing data check, as indicated in Table 6's missing data column, it is evident that the NPM column has a high number of missing values. The researcher performed data cleaning by removing several columns, such as NMP, and irrelevant tables, such as No and No. Pndftan. This process resulted in the cleaned data image.

Table 3. Missing Data from Dataset

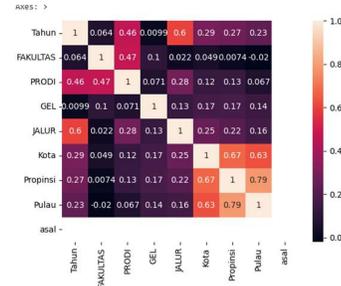
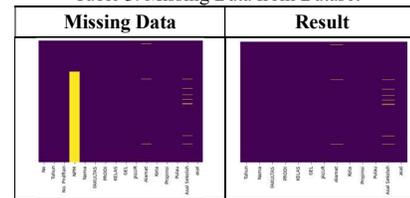
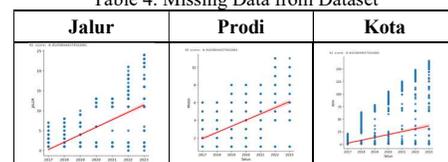
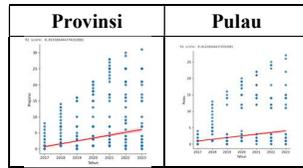


Fig 3. Correlation Matrix among Data.

Based on the correlation matrix above, linear regression is determined for the top 5 data, namely pathway, program, city, province, and island. The table provides a visualization of the linear regression for each correlation.

Table 4. Missing Data from Dataset





3.2 Results of CSV Data Transfer

In "Figure 4," the results of transferring data from an Excel file to a CSV format are presented.

```
tahun;jumlah
2017;173
2018;200
2019;282
2020;204
2021;173
2022;226
2023;211
```

Figure. 4 Result Transferring Data to a CSV file.

3.3 Implementation

Conducting a test by setting the parameter for 2023 as the prediction target involves creating a new CSV file by excluding data from the year 2023.

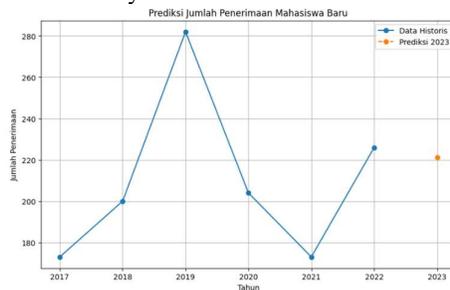


Fig 5. Prediction of New Students in 2023 using ANN.

Analysis of the predicted number of new student admissions for the year 2023 using ANN:

1. In the prediction section, the estimated number of new student admissions expected to join Darma Cendika Catholic University (UKDC) in 2023 is approximately 221.31 or 221 students.
2. The model produces an MSE value of around 0.116, indicating a low level of prediction error in the Artificial Neural Network model.
3. The RMSE value is approximately 0.34, signifying that the predictions are close to the actual data.

Comparing the predicted results with the actual data, there is a difference of around 4.74%, with the predicted value being slightly higher than the actual data (211 to 221 in the prediction results).

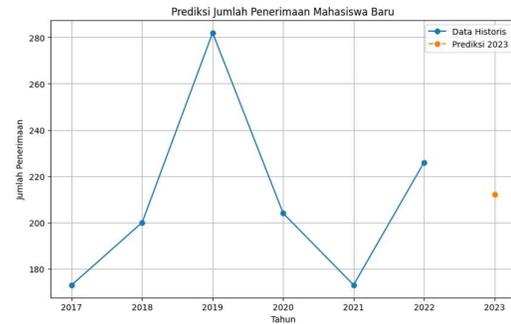


Fig 6. Prediction of New Students in 2023 using Random Forest Regressor.

Analysis Results of Predicting the Number of New Students in 2023 using Random Forest Regressor:

In the prediction section, the estimated number of new student admissions expected to join Darma Cendika Catholic University (UKDC) in 2023 is approximately 212.21 or 212 students.

1. The model produces an MSE value of around 0.3651, indicating a low level of prediction error in the Random Forest Regressor model.
2. The RMSE value is approximately 0.60, signifying that the predictions are close to the actual data.
3. Comparing the predicted results with the actual data, there is a difference of around 0.47%, with the predicted value being slightly higher than the actual data (211 to 212 in the prediction results).

Further testing will be conducted with datasets from 2017 to 2023, followed by predictions for new student admissions in 2024. Figure 4 represents the graphical results of predicting new student admissions in 2024.

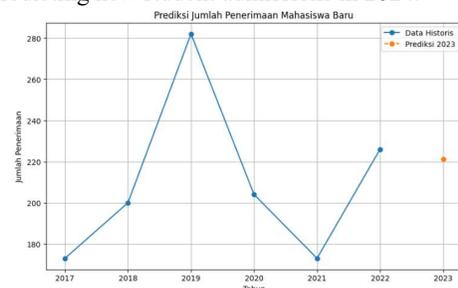


Fig 7 Prediction of New Students in 2024 using ANN.

Analysis Results of Predicting the Number of New Students in 2024 using ANN:

1. In the prediction section, the estimated number of new student admissions expected to join Darma Cendika Catholic University (UKDC) in 2024 is approximately 219.26 or 219 students.
2. The model produces an MSE value of around 0.1046, indicating a low level of prediction error in the Artificial Neural Network model.
3. The RMSE value is approximately 0.32, signifying that the predictions are close to the actual data.



These results suggest that the Artificial Neural Network model provides accurate predictions for the number of new student admissions in 2024.

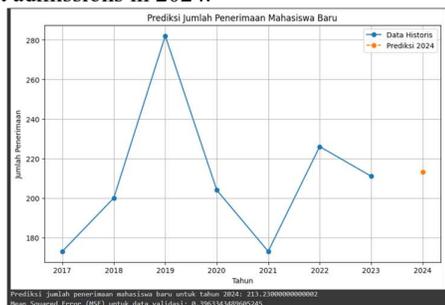


Fig 8. Prediction of New Students in 2024 using Random Forest Regressor.

Analysis Results of Predicting the Number of New Students in 2024 using Random Forest Regressor:

1. In the prediction section, the estimated number of new student admissions expected to join Darma Cendika Catholic University (UKDC) in 2024 is approximately 213.23 or 213 students.
2. The model produces an MSE value of around 0.3963, indicating a low level of prediction error in the Random Forest Regressor model.
3. The RMSE value is approximately 0.63, signifying that the predictions are close to the actual data.

These results suggest that the Random Forest Regressor model provides accurate predictions for the number of new student admissions in 2024.

IV. CONCLUSION

From the analysis results, it can be concluded that the Artificial Neural Network (ANN) model provides a more accurate prediction of the number of new student admissions for the year 2024, estimating 219 students compared to the Random Forest Regressor's estimate of 213 students. The ANN model exhibits a lower prediction error rate with an MSE value of approximately 0.1046 and better prediction accuracy with an RMSE value of around 0.32, compared to the Random Forest Regressor model. However, the difference in error values being higher in the ANN model compared to the Random Forest Regressor with a lower error rate can be attributed to several factors:

1. The dataset has limited data that influences the model's performance, especially the ANN model, which requires a substantial amount of data for effective training.
2. The ANN model is more complex than the Random Forest Regressor, potentially leading to overfitting, whereas the Random Forest Regressor demonstrates robustness due to its use of multiple trees from random subsets.

In this study, it can be concluded that the Artificial Neural Network (ANN) model provides a more accurate

prediction of the number of new student admissions for 2024, estimating around 219 students, compared to the Random Forest Regressor model, which predicts approximately 213 students. The ANN model has a lower prediction error rate, as indicated by an MSE value of around 0.1046, and better prediction accuracy, with an RMSE of around 0.32, compared to the Random Forest Regressor model. However, despite the ANN model providing more accurate predictions and having a lower error rate, the possibility of errors cannot be ruled out due to the relatively small amount of data. Thus, the results of this prediction study can provide support for university administration and development in the future.

REFERENCES

- [1] R. N. Sedyati, "Perguruan tinggi sebagai agen pendidikan dan agen pertumbuhan ekonomi," vol. 16, pp. 155–160, 2022, doi: 10.19184/jpe.v16i1.27957.
- [2] M. Metode, M. Carlo, K. K. Simulasi, and M. Carlo, "Simulasi prediksi jumlah mahasiswa baru universitas dehasen bengkulu menggunakan metode monte carlo," vol. VII, 2020, doi:10.33369/pseudocode.7.1.8-16.
- [3] A. R. Suleman and I. Palupi, "Penerapan Artificial Neural Network (ANN) untuk Prediksi Prestasi Akhir Mahasiswa Melalui Nilai Mata Kuliah Dasar Tingkat 1," vol. 10, no. 2, pp. 1849–1859, 2023.
- [4] R. Septiriana, A. Perwitasari, and R. Septiriana, "Jurnal Mantik Prediction of the Number of Course Participants Using Random Forest Regression Algorithm," vol. 6, no. 3, 2022, doi:10.35335/mantik.v6i3.3175.
- [5] H. Putra and N. Ulfa, "Jurnal Nasional Teknologi dan Sistem Informasi Penerapan Prediksi Produksi Padi Menggunakan Artificial Neural Network Algoritma Backpropagation," vol. 02, pp. 100–107, 2020, doi:10.25077/TEKNOSI.v6i2.2020.100-107.
- [6] M. E. Rosadi, D. Agustini, M. Farida, and D. D. Anjani, "Analisis Penerapan Neural Network dalam Memprediksi Produksi Bijih Nikel di Indonesia," vol. 4, no. 1, pp. 40–50, 2022, doi:doi.org/10.30645/brahmana.v4i1.108.



- [7] M. R. M, R. D. Putri, M. R. N. R, S. Amin, and M. Akli, "Pengaplikasian Artificial Neural Network (ANN) dalam Memprediksi Curah Hujan Menggunakan Python," pp. 369–373.
- [8] I. I. Ridho, C. F. Ramadhani, and A. P. Windarto, "Penerapan Artificial Neural Network dengan Metode Backpropagation Dalam Memprediksi Harga Saham (Kasus : PT . Bank BCA , Tbk) diantaranya oleh A . Santoso dan S . Hansun [1] dalam riset mereka yang sebenarnya . Dalam kesimpulannya , penerapan NN dalam memprediksi harga," vol. 8, pp. 295–303, 2023.
- [9] E. Fitri, "JOURNAL OF APPLIED COMPUTER SCIENCE AND TECHNOLOGY (JACOST) Analisis Perbandingan Metode Regresi Linier , Random Forest Regression dan Gradient Boosted Trees Regression Method untuk Prediksi Harga Rumah," vol. 4, no. 1, pp. 58–64, 2023, doi:10.52158/jacost.v4i1.491.
- [10] K. Ciptady, M. Harahap, and Y. Ndruru, "Prediksi Kualitas Kopi Dengan Algoritma Random Forest Melalui Pendekatan Data Science," vol. 2, no. 1, 2022, doi:10.47709/dsi.v2i1.1708.
- [11] S. Kasus, S. Ciliwung, C. River, and C. Study, "Jurnal Teknologi Lingkungan Prediksi Kualitas Air Sungai Menggunakan Metode Pembelajaran Mesin : Studi Kasus Sungai Ciliwung Prediction of River Water Quality Using Machine Learning Methods : Ciliwung River Case Study," vol. 24, no. 2, pp. 273–282, 2023.
- [12] C. Science, V. W. Siburian, and I. E. Mulyana, "Prediksi Harga Ponsel Menggunakan Metode Random Forest," vol. 4, no. 1, pp. 978–979, 2018.
- [13] N. A. Riani, R. Andreswari, R. Fauzi, and S. Informasi, "IMPLEMENTASI ALGORITMA ARTIFICIAL NEURAL NETWORK," vol. 4307, no. 3, pp. 241–247, 2021.
- [14] M. Adi, P. Hutabarat, M. Julham, A. Wanto, P. Algoritma, and M. Produksi, "Meychael Adi Putra Hutabarat* 1 , Muhammad Julham 2 , Anjar Wanto 3," vol. 4, no. 1, pp. 77–86, 2018.
- [15] J. R. Saragih, M. Billy, S. Saragih, and A. Wanto, "ANALISIS ALGORITMA BACKPROPAGATION DALAM PREDIKSI NILAI EKSPOR (JUTA USD)," vol. 15, no. 2, pp. 254–264, 2018, doi:10.23887/jptk-undiksha.v15i2.14362.
- [16] R. H. Dananjaya, "PENERAPAN ARTIFICIAL NEURAL NETWORK (ANN) DALAM MEMPREDIKSI," vol. 10, no. 4, pp. 419–426, 2022, doi:10.20961/mateksi.v10i4.65034.
- [17] R. Sistem, D. Kartini, F. Abadi, and T. H. Saragih, "Prediksi Tinggi Permukaan Air Waduk Menggunakan Artificial Neural," vol. 1, no. 10, pp. 39–44, 2021, doi:10.29207/resti.v5i1.2602.
- [18] G. I. Marthasari, S. A. Astiti, and Y. Azhar, "Prediksi Data Time-series menggunakan Jaringan Syaraf Tiruan Algoritma Backpropagation Pada Kasus Prediksi Permintaan Beras," vol. 6, no. 3, pp. 187–193, 2021, doi:https://dx.doi.org/10.30591/jpit.v6i3.2627.
- [19] C. E. Larsen, R. Trip and C. R. Johnson, "Methods for procedures related to the electrophysiology of the heart", U.S. Patent 5,529,067, (1995) June 25.

Mobile Based Application Design of the Catholic Church of the Redeemer of Sumbawa Using the Waterfall Method

Claudio Orlando De Araujo^{1*}, Muhammad Zakariyah²

¹Informatics Study Program, Faculty of Science and Technology, Yogyakarta University of Technology

²Informatics Study Program, Faculty of Science and Technology, Universitas Teknologi Yogyakarta

Email: claudio.5200411214@student.uty.ac.id ²muhammad.zakariyah@staff.uty.ac.id

Abstract - The church is a gathering place for Christians to carry out worship. The Sumbawa Redeemer Catholic Church is a church that originates from Sumbawa Besar Regency, West Nusa Tenggara Province. This Catholic Church is the only church in Sumbawa Regency. The purpose of this research is to make the Application of the Catholic Church of the Redeemer of Sumbawa using the waterfall design method and the ADDIE model. The data collection method was carried out using the Unified Modeling Language (UML) method, use cases and activity diagrams are used as analysis and design tools. The programming languages used include Kotlin and Java, the software used includes the IDE Android Studio 4.0, MS. This research yielded results, namely a program that can be used to make it easier for people to find information about the Sumbawa Redeemer Catholic Church.

Keywords - Catholic Church, Android Studio, Firebase, Java, Kotlin

I. INTRODUCTION

In the current era of globalization, technological developments are increasingly widespread in their use [1]. With so many technologies, it can make it easier for users to access internet information [2]. In everyday life, users always use the internet as one of the information media and can be accessed through websites or applications. Technology is currently able to provide better services in various circles of society so as to make it easier for people to access information media [3]. With the development of technology, there are problems in various circles of society, for example among the church[4]. There are some churches that have not used information media systems in the form of web or applications, so it can be difficult for church people to get information.

The Catholic Church is a whole and universal church, encompassing all those united in the bond of baptism[5]. The Catholic Church of the Redeemer of Sumbawa is one of the catholic churches in the middle of the city center. The Catholic Church of the Redeemer of Sumbawa is located on Jalan Diponegoro, West Nusa Tenggara Province, Sumbawa Regency, Sumbawa Besar District, Uma Sima Village[6]. The Catholic Church of the Redeemer of Sumbawa has been established since November 24, 1990 which has a population of 2311 people and some of its people come from overseas and students. The Catholic Church of the Redeemer of Sumbawa is still difficult in managing the information media system, because the information media system of the Catholic Church of the Redeemer of Sumbawa is still running smoothly. manual, which is the delivery of information through readings that can be obtained during the celebration of the Eucharist [7]. If there are some people who are not present at the Eucharist celebration on Sunday, they cannot get information[8]. The management of people's data is still done manually, so to find out these data, you have to find back the people's data files that

have been stored[9]. Some devotions are still done manually, that is , they are done by reading daily devotions at morning worship. Thus, a solution is needed to help the catholic church of the redeemer of Sumbawa in the information media system so that it cannot facilitate church people in use information media provided by the church[10].

From these problems, research was conducted to design a church information media system, so that it could help manage information data in church services [11]. The information media system that will be built has several features, namely church scheduling, church devotional, community data collection, and church information. The scheduling feature will discuss the schedule of the Eucharistic celebration and the tasks in charge of the Eucharistic celebration, the devotional feature will discuss several daily devotions each day, the church community data collection feature will discuss the data on the names of church members who have been registered in the Sumbawa redeemer catholic church and the information feature will discuss information important church. The information media system to be built is in the form of an application. The application to be built is based on Android and uses the Kotlin programming language. The application to be built uses a database in the form of firebase[12].

Android Studio is an Integrated Development Environment (IDE) software used to develop android applications[13]. Android Studio first appeared in 2013 and began to be released in June 20 14, the stable version was released in December 2014, while the stable version was released in May 2020 . Android Studio is the official software supported by Google. Android Studio provides an interface for creating and managing management apps. For programming language used is kotlin. Android Studio has the advantage of giving access to the Android Software Development Kit (SDK).

Website is a collection of digital pages that contain information in the form of text, images, videos, or a combination that has been connected to the internet [14]. The website can be seen or accessed by various circles of society. There are several web programming languages used namely html, php, css, and javascript. The website first appeared in the world in 1991 and was inaugurated on April 30, 1993 which can be used for free by the public. The website is very important in various circles of society, because that way, many information media can be accessed through the website.

Firebase is an app development platform to simplify the work of mobile developers[15]. With Firebase, developers can focus more on developing applications. Firebase was first established in 2011, the first platform to be developed is the realtime database. Realtime database is useful to help developers store and synchronize data to many users. There are several features that can be used by users such as firebase analytics, firebase cloud messaging, notifications, firebase authentications, firebase remote config, firebase realtime database, and firebase crash reporting.

Kotlin is a programming language that is often used by developers and is based on open source[16]. Kotlin development has been carried out since 2010, in 2012 Kotlin became open source. Kotlin is one of the programming languages that are in great demand by developers, because it can overcome null, concise, no runtime overhead, a large community, and is flexible.

II. RESEARCH METHODOLOGY

2.1 Metode SDLC (System Development Life Cycle)

In this study, researchers used the SDLC (Software Development Life Cycle) method. Software Development Life Cycle (SDLC) is a method used to design and develop a system. The design of this application uses the SDLC (Software Development Life Cycle) method with a waterfall model. The SDLC method consists of several stages, namely analysis, design, design, implementation, trial and management [17]. This SDLC method was chosen because it is in accordance with the application to be designed and can assist researchers in designing and developing applications. In figure 1 is the stage of the SDLC method with a *waterfall* model

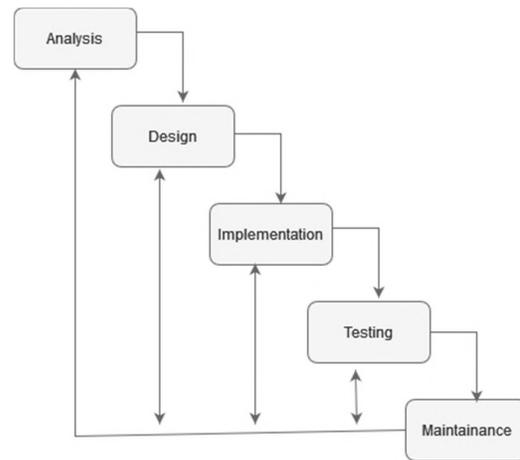


Fig1. Waterfall model stages

2.1.1 Analysis

The initial stage of the SDLC method is the analysis stage. The analysis stage is the stage of collecting data according to user needs. The analysis stage is carried out to find out what needs are needed by users for the software designed and developed [18]. Using analytics, it can be determined the purpose of the application to be shown to the user. There are several methods that can be used in data collection including observation, surveys, interviews and others.

2.2.2 Planning

At this stage is the stage of designing application design based on the analysis that has been done. The results of the analysis that have been carried out are implemented in the design of the software. This stage is carried out to provide an overall picture of the application to be designed and developed.

2.2.3 Implementation

The implementation stage is the stage of implementing designs that have previously been designed into programming codes. At this stage, testing and checking functionality is carried out so that it can assist researchers in designing and developing an application.

2.2.4 Testing

This stage is the next stage of the implementation stage. At this stage, testing is carried out throughout the system to find out whether the application designed is in accordance with its design and function [19]. The testing phase is very important to ensure the application is being designed and developed according to needs and functions.

2.2.5 Management

The last stage of the *waterfall* model is the management stage. System management serves to find out which systems are not detected during the testing phase. At this stage, improvements are also made to the software such as features and user needs.

III. RESULTS AND DISCUSSION

This section is a section that discusses the implementation of the SDLC method with a waterfall model in the Mobile-Based Sumbawa Redeemer Catholic Church Application Design Using the Waterfall Method.

3.1 Analysis

At this stage, the analysis and data collection used is using observation and interview methods. Information data can be found through books, journals, and literature on the internet. Data collection is carried out based on user needs and system needs. Data collection aims to identify what needs are used in application design and development. In this stage, there are software requirements specifications, including the design of *use case diagrams*, the design of *activity diagrams* and the design of input needs (*input*) and needs output (*output*).

The need for software design and development in the information system Application Design of the Catholic Church The Redeemer Sumbawa Mobile Based, as follows:

- a. Register
 - b. Log In
 - c. Access the Main Menu
 - d. Accessing Scheduling Features
 - e. Accessing the Devotional Feature
 - f. Accessing Information Features
 - f. Accessing the Logging Feature
1. Input Needs

This application has several user-specific inputs, namely users who can see church service information. Then another user is an admin who can add and manage data in the application. Here are inputs by different users.

- a. The input given by the user, to login via Android smartphone is username, email, for login purposes and for registration, the user fills out several forms, namely name, email, mobile number, and password.
- b. The input given by the admin, to login via Android smartphone is email, filled with the registered email and password, filled with the registered password.

2. Output Requirement (*Output*)

Output needs analysis is the process of designing and developing the Android-based Sumbawa Redeemer Catholic Church application using the Waterfall method. These needs include hardware needs and software needs. The required devices are Android Studio, Kotlin, Firebase, while for hardware needs, namely laptops, smartphones, and data cables.

3.2 Planning

Unified Modelling Language (UML) is a system design carried out to describe the flow of relationships between users and applications. The system design used to explain how the application works is the *Use case Diagram* and *Activity Diagram*.

a. Use Case Diagram

Use case Diagram is an interaction between the system and the user. In the picture below there is an admin who can manage data, then the admin can input data, edit data and delete data[20]. Then for users can access the features available in the system, following the system workflow *Use case Diagram*.

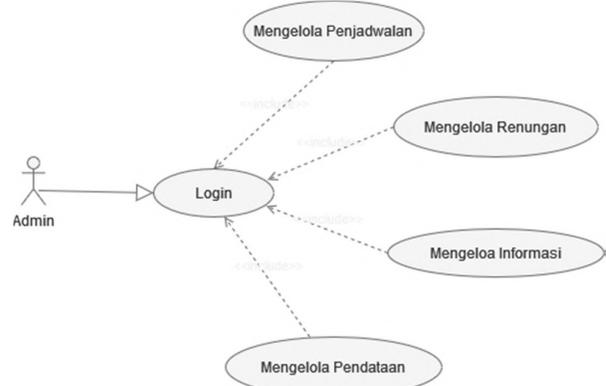


Fig 2. Use Case Admin Diagram

In figure 2, the *admin use case diagram*, the admin is the church administrator who serves as the administrator of the information system. First, the login admin enters the name and password, then goes to the admin page, on the admin page there are several features, namely scheduling features, devotional features, information features, and logging features. The admin has the task of managing people's data, information data, devotional data, scheduling data, uploading images, videos, news and others

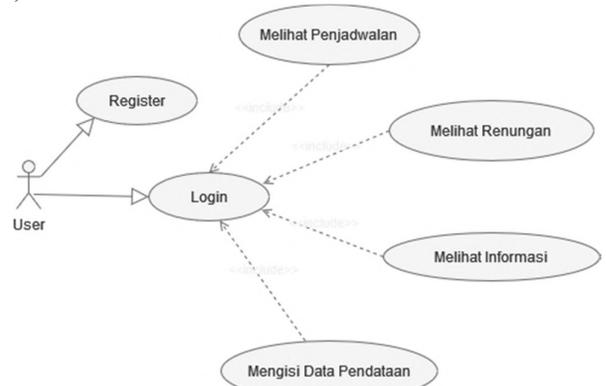


Fig 3. Use Case Diagram User

In figure 3, the *user diagram use case*, which is the user, namely the user or church members. First, the user registers, the user fills in data such as name, email, mobile number, and password. If the user has filled in the data in the register, the user can log in, the user fills in the email or mobile number and password that has been registered before. The user will be directed to the main menu, in the main menu there are several features, such as scheduling features, reflection features, information features, and logging features.

B. Activity Diagram



Activity Diagram is a depiction of the workflow of an application system that is interrelated on the system, along with an activity diagram image.

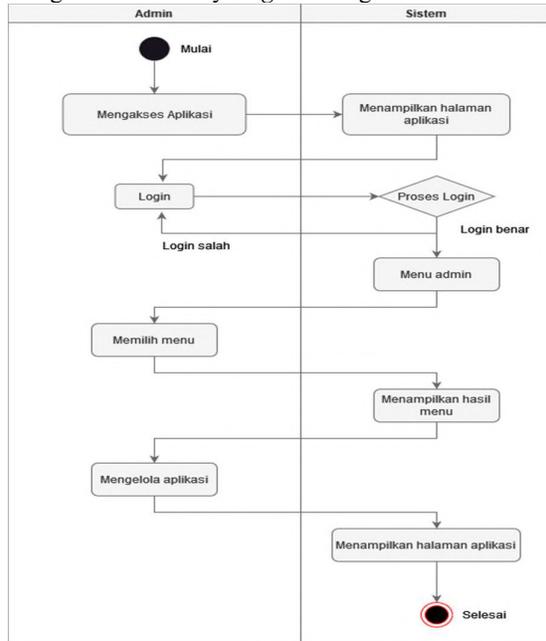


Fig 4. Activity Diagram Admin

Figure 4 is the activity diagram on the admin, first the admin accesses the application.church, then the system will display the login. The admin fills out the login form, then the system will manage the admin data. If it is correct, then the system will go to the admin menu page. Furthermore, the admin will manage the contents of the application and save data from the results of application management. Then the system will save the data previously done by the admin.

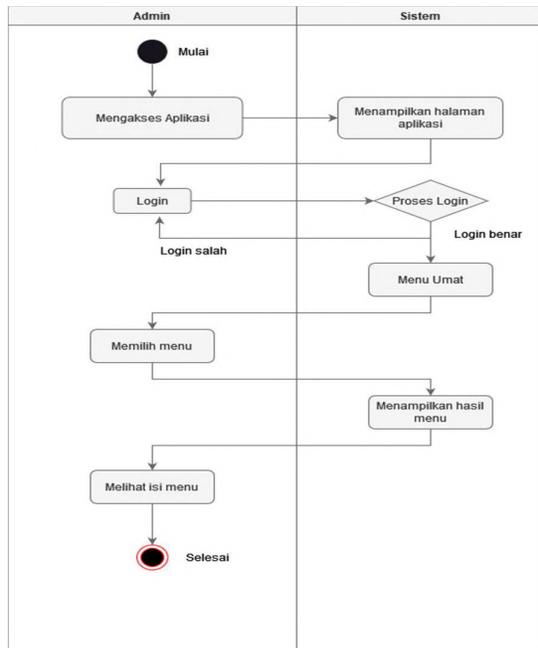


Fig 5. Activity Diagram User

Figure 5 is the activity diagram user, the user user is the church. Church members will access the application and the system will display a login form. Then the church members will fill out the login form that has been provided, then the system will manage the data from church members and will be sent to the database. If the login is incorrect, the system will display a message that is a login error. If the login is correct, the system will go to the church community menu page. Then the church members can choose the desired menu.

3.3 Implementation/Testing

a. Start page

On this page serves as the main page. On this page there is a login and register menu, before going to the login page, church members first register so that they can go to the menu page. On the register page there are several forms, namely name, email, mobile number, and password. If church members have registered, the system will send the church congregation data to the database. If the data is valid, church members can fill out the login form, on the login page there are several forms, namely email and password. If the church members fill in the data correctly, the system will go to the menu page.



Fig 6. Home

b. Menu Page

After going to the menu page, there are several features such as scheduling, planning, information, and data collection of people. On the scheduling menu, there is a Eucharistic celebration officer and a Eucharistic celebration schedule. Devotions contain daily devotions at morning, evening, or evening mass. Information includes church announcements such as pictures, videos, news, and other posts. The data collection of people contains data on people who have just joined the catholic church of the redeemer of Sumbawa, on the menu of the people's data collection contains the name, place of birth, parish origin, and origin of residence.



Fig 7. Menu Page

3.4 Testing

This test is carried out using two methods, namely by using *Black Box* testing and *User Acceptance*. *Black Box* testing is used as application testing to obtain system response to applications that have been designed and developed. Testing is conducted by respondents who are selected and pose as users. The results of testing with the *Black Box* method are shown in Table 1.

Table 1. Unit test results

Unit	System Testing	System Reaction	Test Results	Tested By
Register	Fill in name, email, phone, password	Data has been registered	Succeeded	User

Login	Fill in the registered email and password	Display the menu page	Succeeded	User
	Filling in inappropriate emails and passwords	Please fill in the data correctly	Succeeded	User
Scheduling Menu	View Eucharistic celebration schedules and Eucharistic celebration attendants	Displays data on Eucharistic celebration schedules and Eucharistic celebration officials	Succeeded	User
Devotional Menu	View daily devotionals each day	View daily devotional data	Succeeded	User
Information Menu	View a variety of church-related information such as pictures, videos, news and other posts.	Display data of various information	Succeeded	User
Logging Menu	Fill in personal data such as name, place, date of birth, parish origin and origin of residence	Display data of people who have been diagnosed	Succeeded	User

The next test is to use *User Acceptance* testing. This test is carried out to determine whether the application that has been designed is in accordance with user needs at the analysis stage. Testing was conducted with respondents and several questions regarding the application. The results of the answers from respondents are given weight to each assessment. The weighting starts from strongly disagree to strongly agree. Value weighting is shown in Table 2.



Table 2. Value weighting

Valuation	Weight
Strongly disagree	1
Disagree	2
Netral	3
Agree	4
Totally agree	5
Totally agree	

The results of the weighting that have been carried out are then classified and grouped based on the questions that have been asked. This test was conducted involving five respondents. The results of the questions asked with the break-inof predetermined values are shown in Table 4.

Table 3. User Acceptance test results

Respond	Question					Sum
	1	2	3	4	5	
User 1	4	4	3	4	3	18
User 2	4	4	3	5	5	21
User 3	3	5	4	4	5	21
User 4	5	4	4	4	5	22
User 5	4	4	3	5	4	20
						102
Total Score						2040

Based on Table 3 test results using *Black Box* testing, get good results. The entire system has been tested and no system errors were found. All units tested are in accordance with previous analyses. Overall the system can run well and display reactions according to the user's wishes.

Testing with the *User Acceptance* method gets a fairly high score. The results of this test will be calculated to get the final value of user satisfaction with the application that has been designed. The calculation is done using the following formula (1):

$$n = \frac{\text{Jumlah}}{\text{Ju}} \times 100$$

Description:

n = result

Ju = Number of Users

In table 3, the results of the calculation using the formula above get a total score of **102** on all users from 5 questions.

IV. CONCLUSION

Based on the results of research and discussion of the Sumbawa Redeemer Catholic Church Application Design, there are several problems, such as the delivery of information, and the management of church data. This makes it difficult for the church to manage and deliver information to church members, because the church information system is still done manually. Therefore, researchers designed a church application to make it easier for churches to manage and convey church information. Researchers conduct observations and interviews to obtain church data, making it easier for researchers to design and develop applications. Researchers designed six features to assist churches in and manage conveying information such as church scheduling features, church devotional features, church information features, and the church's community data collection feature. In the scheduling feature contains about the schedule and church officials at the Eucharistic celebration, the devotional feature contains daily devotions such as Bible verses, the feature Information contains images, videos, news and other posts, the Church Community Data Collection feature contains a data collection form for people who have just joined in parish of the catholic church of the redeemer of Sumbawa. Researchers also provide testing of users, so that users can make an assessment of the application. There are five users and have different ratings, and result in the number 102, this number divided by the number of users, and multiplied by 100 to produce 2040. By conducting this assessment, it can find out what percentage of user interest in the application designed. In this study, researchers have limitations in developing applications, so that It is hoped that the next researcher can develop ApplicationI such as adding several features and presenting more complete information on people's data.

REFERENCES

- [1]I. Christina, "Design of Mobile Web-Based Volunteer Management Interface at Mawar Sharon Church," *J. DKV Adiwarna*, vol. 1, no. 12, pp. 1–7, 2018.
- [2]J. Lasso and E. Wahyuningtyas, "Mobile-based Church and Vihara Place of Worship Search Application," *IT Literacy Inf. Technol. ...*, Vol. 2, No. 1, pp. 55–64, 2016, [Online]. Available: <http://melekit.if.uwks.ac.id/index.php/printmelekit/article/view/206>
- [3] R. Herawati, B. Sumboro, and A. Najib, "Development of Learning Media for Android-Based BLK Karanganyar Graphic Design Training," *Go Infotech J. Ilm. STMIK AUB*, vol. 28, no. 1, pp. 1–8, 2022, doi: 10.36309/goi.v28i1.161.
- [4]J. Technology *et al.*, "Web-Based Church Activities Management Information System in Peace-loving Catholic Churches," *J. Teknol. Wellness. Social Science.*, Vol. 4, No. 2, pp. 383–391, 2022, [Online]. Available: <http://e-journal.sari-mutiara.ac.id/index.php/tekesnos>
- [5]V. Darmin, *Catholic religious education and discretion*, vol. 4, no. volume 1. 2018.
- [6]M. Sagala, "Android-based Church News Information System (Case Study: HKBP Tanjung Sari Church Medan)," vol. 2, pp. 74–82, 2022.
- [7] Y. R. Asih, A. Priyanto, and D. A. Puryono, "Website-Based Church Congregation Service Information System Using



- PIECES Analysis," *J. Tek. Inform. and Sist. Inf.*, Vol. 8, No. 1, pp. 175–186, 2022, DOI: 10.28932/jutisi.v8i1.4406.
- [8]A. Widiyastuti and Daliman, "Web-based information system at St. Pius X Gisting Church," *JISN (Jurnal Inform. Softw. and Network)*, vol. 01, no. 01, pp. 1–8, 2020.
- [9]M. E. Dien, "Website Design and Information System of Santa Maria Bintang Laut Parish Catholic Church Ambon," *J. Simetrik*, vol. 12, no. 2, pp. 613–621, 2023, doi: 10.31959/js.v12i2.1119.
- [10] A. Alex, A. A. Pekuwali, and P. A. R. L. Ledo, "Implementation of Worship Scheduling Information System and Website-Based Financial Management (Case Study: Gks Jemaat Padadita) Implementation of Worship Scheduling Information System and Website-," vol. 01, no. 03, pp. 177–186, 2022.
- [11] Owen, B. Yudi Dwiandiyanta, and Suryanti Ch., "Development of Information Systems with Data Visualization at Santa Maria Church with Immaculate Origin of Tulungagung," *J. Inform. Atma Jogja*, vol. 3, no. 2, pp. 117–125, 2022, doi: 10.24002/jiaj.v3i2.6785.
- [12]J. S. Ananta and R. Somya, "Design and Development of Web-Based Gbcp Church Management Information System," *J. Komput. and Inform.*, Vol. 11, No. 1, pp. 44–53, 2023, DOI: 10.35508/jicon.v11i1.10101.
- [13]M. P. Sari and D. P. Mulya, "Development of Railway Boking Information System," vol. 1, no. 1, pp. 7–12, 2021.
- [14]P. S. Ganney, *Web Programming1*. 2022. doi: 10.1201/9781003316244-11.
- [15] Ilham Firman Maulana, "Implementation of Firebase Realtime Database on Android Mobile-based Smartphone E-Tilang Application," *J. RESTI (Engineering Sist. and Technology. Information)*, vol. 4, no. 5, pp. 854–863, 2020, doi: 10.29207/resti.v4i5.2232.
- [16]A. Febriandirza, "Online Time Attendance Application Design Using Kotlin Programming Language," *Pseudocode*, vol. 7, no. 2, pp. 123–133, 2020, doi: 10.33369/pseudocode.7.2.123-133.
- [17]M. M. Lucini, P. J. Van Leeuwen, and M. Pulido, "Model error estimation using the expectation maximization algorithm and a particle flow filter," *SIAM-ASA J. Uncertain. Quantif.*, vol. 9, no. 2, pp. 681–707, 2021, doi: 10.1137/19M1297300.
- [18]N. D. Nathasia and W. Winarsih, "Development of Video Gallery System Results of Upt MPR Coverage Using SDLC Method," *J I M P - J. Inform. Merdeka Pasuruan*, vol. 4, no. 3, pp. 30–35, 2020, doi: 10.37438/jimp.v4i3.229.
- [19]M. Stefanus and J. Fernandes Andry, "Development of Web-Based E-Learning Applications Using Waterfall Model at Smk Strada 2 Jakarta," *J. Fasilkom*, vol. 10, no. 1, pp. 1–10, 2020.
- [20]F. Juliyanto and Parjito, "Application Engineering of E-Filling Management of Letter Documents at PT ALP (Atoism Lampung Pelayaran)," *J. Teknol. and Sist. Inf.*, Vol. 2, No. 1, pp. 43–49, 2021.



Detection of Hate-Speech Tweets Based on Deep Learning: A Review

Ara Zozan Miran^{1*}, Adnan Mohsin Abdulazeez²,

¹ Duhok Polytechnic University, Duhok, Kurdistan Region, Iraq, Akre University for Applied Science- Technical College of Informatics- Akre- Department of Information Technology.

² Duhok Polytechnic University, Duhok, Kurdistan Region, Iraq.

Email: ¹ara.miran@auas.edu.krd, ²adnan.mohsin@dpu.edu.krd

Abstract – Cybercrime, cyberbullying, and hate speech have all increased in conjunction with the use of the internet and social media. The scope of hate speech knows no bounds or organizational or individual boundaries. This disorder affects many people in diverse ways. It can be harsh, offensive, or discriminating depending on the target's gender, race, political opinions, religious intolerance, nationality, human color, disability, ethnicity, sexual orientation, or status as an immigrant. Authorities and academics are investigating new methods for identifying hate speech on social media platforms like Facebook and Twitter. This study adds to the ongoing discussion about creating safer digital spaces while balancing limiting hate speech and protecting freedom of speech. Partnerships between researchers, platform developers, and communities are crucial in creating efficient and ethical content moderation systems on Twitter and other social media sites. For this reason, multiple methodologies, models, and algorithms are employed. This study presents a thorough analysis of hate speech in numerous research publications. Each article has been thoroughly examined, including evaluating the algorithms or methodologies used, databases, classification techniques, and the findings achieved. In addition, comprehensive discussions were held on all the examined papers, explicitly focusing on consuming deep learning techniques to detect hate speech.

Keywords – Twitter, Hate Speech, Toxic, Cyberbullying, Deep Learning.

I. INTRODUCTION

Online social media enables the spread of humanities to be associated. However, one disadvantage of these social media is the ability to publish and propagate malicious and harmful content [1]. Hate speech seeks to promote violence and incite hostility against individuals or groups based on characteristics—such as sexual orientation/gender identity, religion, disability, gender, age, veteran status, or age, gender, or disability. The increasing number of social media platforms has caused matters to be excessive [2]. Not all substances are pertinent; some may cause damage to individuals, which is a disgraceful indictment when those who use the media to spread hatred [3]. Hate speech affects everyone regardless of age, which might be too young, an adult, or an older adult [4]. Individuals use a lot of social media platforms daily to express their thoughts, emotions, and progress. Usually, the comments are more hate speech than positive comments, which leads to different kinds of problems; for that reason, it is essential to select a method to detect the words, analyze them, and then show the result of accuracy [5]. Deep learning utilizes networks that have multiple layers [6]. This allows them to understand patterns and representations in data by learning from examples [7]. The training process involves providing these networks with amounts of labeled data, enabling the system to adjust its parameters until it can accurately predict or classify information [8]. This ability makes deep learning especially effective for tasks that involve decision-making and pattern recognition. Deep understanding encompasses such architectures tailored for specific tasks, each designed to address the unique challenges of different data types [9]. Artificial neural networks, inspired by biological neurons, serve as the foundational components of deep learning, replicating the core elements of human intellect [10]. These networks, which have several layers of nodes, have led to many architectures, each specialized for distinct tasks and

applications. Convolutional Neural Networks (CNN) excel in analyzing videos and photos, Recurrent Neural Networks (RNN) are adept at addressing issues related to sequential data, and Generative Adversarial Networks (GAN) are at the forefront of generating synthetic data that closely mimics accurate data. [11], [12]. Hate speech detection in social media Twitter is detecting and defining all the cyberbullying an individual can receive on their regular posts [13]. The hate speech might be on racism, sexuality, child abuse, Politicians, and many others. Detecting hate speech is difficult because its nature and context-specific attributes frequently characterize it [14],[15]. Conventional rule-based or keyword-based methods may face problems in accurately identifying the subtleties of hate speech, as hate speech can take on different forms, such as indirect language, changing slang, or statements that depend on the context [16]. Hate speech detection utilizes deep learning, specifically natural language processing algorithms [17]. This paper aims to review other works on detecting hate speech in social media using different types of deep learning methods. The rest of this paper is organized to describe the background theory of deep learning and the algorithms to detect hate speech on Twitter platforms. After that, in section 2, different research on using CNN and GRU, CNN, and LSTM to detect hate speech on social media are described and compared in Table 1 and Table 2. Finally, section 3 describes a discussion session, a conclusion and limitations, and a list of references are shown.

II. RESEARCH METHODOLOGY

Detecting hate speech in social media requires datasets and an algorithm to learn and extract the comments to be analyzed [18]. For that reason, researchers are using deep learning algorithms to detect hate speech. Deep Learning, a subset of machine learning, aims to replicate the brain's function for analyzing and processing data [15]. At its core, Learning utilizes networks of intricate structures made up



of interconnected nodes to model and interpret complex patterns within large datasets. What makes deep learning understandable is its ability to learn representations of data automatically [19], enabling it to identify features and relationships that may pose challenges for conventional machine learning methods [20]. This technology has shown achievements in file recognition, speech processing, natural language understanding, and autonomous systems [21], [22]. The continuous progress in deep learning algorithms alone and the availability of computing resources have paved the way for advancements in artificial intelligence[23]. These advancements shape technology and impact domains like healthcare, finance, and self-driving vehicles. This technology has shown achievements in file recognition, speech processing, natural language understanding, and autonomous systems[24] [20]. The continuous progress in deep learning algorithms alone and the availability of computing resources have paved the way for advancements in artificial intelligence [25],[26]. These advancements shape technology and impact domains like healthcare, finance, and self-driving vehicles. The Advances in learning models, such as (CNNs) for image processing and (RNNs) for sequential data, have resulted in substantial changes in domains such as machine vision, natural language processing, and recognition of words[8]. Deep Learning is executed by neural networks, which consist of multiple layers, a concept that is not novel [27], [28]. Nevertheless, its popularity has surged recently, driven mainly by three factors: Firstly, there has been a significant enhancement in processing capabilities, such as video cards and graphical processors. Secondly, the availability of affordable computer hardware has played a crucial role. Lastly, recent advancements and breakthroughs in Deep Learning research have also contributed to its rise. Deep learning algorithms can be categorized into three groupings based on whether the algorithms are trained to produce achievable results. The subgroups are unsupervised, supervised, and hybrid [27]. The reliability of each deep learning algorithm is demonstrated in Figure. 1.

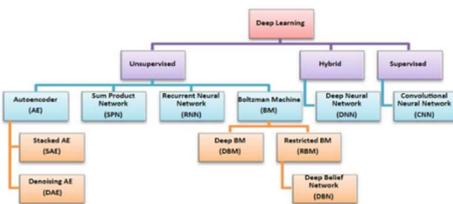


Fig 1. Classification of deep learning techniques,
<https://www.researchgate.net>

A. Convolutional Neural Networks (CNN) and Gated Recurrent Units (GRU) for Detecting Hate Speech:

Due to its potentially devastating effects on society, identifying instances of hate speech in digital communications is a significant difficulty [29]. One viable strategy for this problem is to use cutting-edge technologies like (CNNs). Combining (CNNs) and (GRUs) offers a powerful technique for detecting hate speech, harnessing the benefits of both architectures [14]. While CNNs are great at comprehending short-term dependencies in sequences, (GRUs) excel at understanding long-term

dependencies [30]. In this combined method, textual information is tokenized and processed, then supplied into an embedding layer, as shown in Figure 2 [31]. It covers identifying hate speech on social media, particularly Twitter and provides a convolution-GRU-based deep neural network solution [32]. The authors address hate speech identification features such as a bag of words, word and character n-grams, sentiment analysis, linguistic resources, and standard and deep learning methods [33]. The authors also noted a need for comparable evaluation and limited public hate speech datasets.

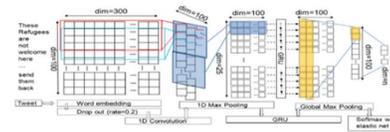


Fig. 2 CNN+GRU Architecture [31]

The CNN's convolutional layers then process the embedded sequences to pull out abstract patterns and features at various levels of granularity [34]. The (GRU) then uses these traits to learn to record sequential and contextual information, drawing connections between previously unrelated parts of the text [22]. The (GRU)'s gated structure allows for the storage and updating of data, which aids in the nuanced understanding of the text's context and the detection of hate speech [35]. Training a mixed model requires iteratively adjusting parameters to achieve a target loss function minimization. To better detect hate speech, the model combines (CNNs) for feature extraction with GRU for sequence modeling [10]. The model's performance and robustness in identifying hate speech across online content can be improved by fine-tuning, optimizing, and studying other architectures and hyperparameters. There are negative impacts on humans' mental and physical health as a direct result of the rapid growth of social media mediums and the constant use by their users. To guard against and avoid any mishaps that might take place behind the scenes on social media platforms, hate speech has become one of the most hazardous subjects to discuss. Researchers turned to deep learning strategies, which will be covered in more detail in the following paragraphs, to identify instances of hate speech on social media sites like Twitter [36], [37] proposed method utilizes a deep learning ensemble approach that combines multiple sub-models to improve classification performance. The paper also mentions the use of a publicly available embedding model. As for the number of datasets used, the article does not provide a specific number. However, the authors mention using the SemEval dataset to compare findings and a split set with an 85/15 ratio for testing purposes. Regarding the results, the paper reports a 5-point improvement in F-measure compared to previous work. The ensemble approach also showed a reduction in variance compared to individual models. [37]discusses a new method for detecting hate speech on Twitter using a Convolution-(GRU)-based deep neural network. The proposed method outperforms previous methods on 6 out of 7 datasets by between 1 and 13% in F1. The comparison was made against several baselines and state-of-the-art on the most extensive publicly available Twitter datasets. The paper also



discusses the implementation, parameter tuning, and evaluation metrics used in the study. [38] systematically analyze hate speech in 9 languages and offer a low-resource detection approach. They test CNN-GRU, BERT, and LASER embedding with logistic regression to detect hate speech in different languages under different scenarios. The authors also present a portfolio of effective models for each language based on data availability. The research sheds light on the difficulties of detecting hate speech in languages other than English and suggests a low-resource detection approach. In [39] they used CNN and GRU algorithms to detect hate speech. CNN is better for local features, and GRU is better for long-range context dependencies. LSTM, another RNN, is slower and more likely to overfit small datasets than GRU. Both models have pros and cons, and hyperparameter adjustment affects performance. [40] suggests utilizing deep neural networks to detect internet hate speech. According to the authors, automatic identification is needed due to digital communication and online hostility. They propose multiclass classification using fastText, BERT embeddings, SVM, and deep neural network classifiers. They find encouraging results using their method on numerous datasets. [41] Combining GRU, ELMo, BERT, and CNN data improves classification performance. The proposed approach has good accuracy and F1 score, making it a feasible option for social media hate speech. However, the approach's success may depend on the training data's quality and representativeness and the classifiers and fusion methods used. [9] discusses a model for detecting hate speech on Twitter using CNN and character-level representation. The authors present their approach to building and cleaning datasets, as well as the architecture of their model. They also discuss related work in sentiment analysis and hate speech detection. The results of their experiments were presented and discussed, with the model achieving satisfactory accuracy compared to the state-of-the-art. [42] highlight various hate speech detection research using these neural network topologies. Zimmerman et al. represented 50 tokens using CNN parameters, while Founta et al. offered a two-layer RNN model using GRU and metadata on people, networks, and content. Some hate speech detection investigations have used hybrid CNN and RNN models, which perform better than either architecture alone. The research thoroughly analyzes CNN and GRU in hate speech detection and shows their potential to improve model accuracy. [43] discusses a study that classified Bengali Facebook comments as Hate Speech, Communal Attacks, Incitement, Religious Hatred, Political Comments, and Religious Comments using (GRU). The study presented an annotated Bengali corpus of six-class comments. The experiment indicates that a GRU-based classification model can achieve 70.10% accuracy. The article found that the BiLSTM performed best with a weighted classification F1 score of 91%. In [36], to identify harmful remarks in virtual communities and social media platforms. The algorithm employs pre-trained word embeddings and many channels to forecast multilabel toxicity indicators precisely. The report additionally examines pertinent literature, outlines the structure of the suggested model, and presents the

findings of its evaluation. Moreover, in [37], the research paper discussed the use of deep neural network (DNN) architectures, including (CNN), (GRU), and Universal Language Model Fine-tuning (ULMFiT), for the classification of hate speech in tweets. It highlights the advantages of using DNN models for hate speech detection and emphasizes the superior performance of the ULMFiT model over other architectures. The ULMFiT model, based on the three-layered Average-SGD Weight-Dropped Long Short-Term Memory (AWD-LSTM) architecture, demonstrates substantial improvements in accuracy and F1 score. After that, in [46]. It outlines the method used in the study, which involves transforming speech signals from the time domain to the frequency domain, extracting features using mel-frequency cepstral coefficients, and using a CNN-GRU model to detect and classify dysarthria patients and healthy people. The article also mentions the prevalence of dysarthria in patients with neurological diseases and the potential applications of this technology in healthcare. [47] Investigates the frequency of hate speech on Twitter amidst the COVID-19 outbreak. The study uses machine learning algorithms to discern and examine the data to ascertain patterns and trends within hate speech. The results indicate that there was a significant presence of hate speech about COVID-19 on Twitter during the pandemic, with specific demographics being disproportionately targeted. The authors propose that their discoveries can be employed to counteract hate speech and foster inclusiveness on social media sites. The researcher in [29] Worked on A comprehensive analysis of scholarly research on the identification of hate speech in social media, particularly on Twitter, using the utilization of neural networks. The review comprises a subset of 20 studies from a pool of 565 distinct works, chosen according to specific inclusion and exclusion criteria and the study's primary and secondary inquiries. The review examines the models, algorithms, and methodologies employed in these investigations and the datasets, languages, and nations implicated. The review reveals that the predominant models used are based on Convolutional Neural Networks (CNNs). While these models offer certain advantages over recent models in detecting hate speech, they also exhibit limitations and deficiencies. These include challenges in automated hate speech detection, compatibility with specific languages only, and ambiguity in speech classification when applied to datasets in different languages. The review presents a thorough overview of the experimental particulars and findings of the incorporated investigations, encompassing F1 score, recall, precision, and accuracy. In [48], The authors address the issue of vocabulary mismatch in tweets by using feature expansion to build a corpus of similarity. The study uses four deep learning methods: Convolutional Neural Network (CNN), Gated Recurrent Unit (GRU), and a combination of the CNN-GRU and GRU-CNN classifications with Boolean representation as feature extraction. The experiment was carried out to find the best performance by comparing the value of the accuracy. The accuracy results were obtained from the average of the five tests in each scenario.



Table 1. Overview of the literature on detecting hate speech Tweets based on CNN and GRU.



Ref	Year	Datasets	Based Model	Accuracy	Advantages	Limitations
[31]	2018	Seven datasets	CNN+GRU	range from 0.72% to 0.94%	Capacity to catch word sequence and order in brief texts, high Twitter hate speech detection accuracy.	The lack of comparison judgments makes it hard to evaluate individual works. The existence of abstract ideas like 'sexism,' 'racism,' or 'hate' is difficult to discern in literary content, thus knowledge of social groupings.
[36]	2018	SemEval dataset	CNN	5%	Shows a significant improvement in F-measure compared to previous work. The authors also provide detailed experimental results and analysis, including confusion matrices, to aid reproducibility and comparison with other methods.	One limitation is using pre-trained word embeddings, which may not capture all relevant information for the specific task—also, there is a lack of explanation for the decisions made by the model, which may be necessary for applications. Additionally, the proposed approach may not be suitable for low-resource languages or domains with limited training data.
[37]	2018	Seven public datasets	CNN+GRU	13%	The proposed method uses only word-based features, which are more straightforward and interpretable than character-based features yet achieve better results.	The proposed method may have limitations and potential issues not discussed in the paper.
[9]	2019	Seven datasets	CNN	0.8893%	It addresses the challenge of multilingual hate speech detection, which is a significant advantage given the diverse nature of social media content. The model was found to be more than satisfactory when compared to the state-of-the-art model, indicating its effectiveness in dealing with the issue of hate speech on the internet.	Potential biases in the dataset, constraints in the model's generalizability, or challenges in real-time implementation.
[39]	2019	Million tweets	CNN+GRU	0.678%	Gives a summary of the essential works in social media hate speech automatic detection for scholars and practitioners.	Focuses on hate speech detection's technological components, such as machine learning techniques and feature engineering, rather than its ethical and societal ramifications.
[40]	2019	24,883 tweets	CNN, SVM, GRU, BERT	11.6%	Examine multiple models and find that BERT fine-tuning performed best, showing that deep neural networks can detect internet hate speech. Their approach is adaptable to additional languages and domains outside Twitter.	It is focused on English-language tweets and says the confusion between hate speech and offensive speech needs further study.



[41]	2020	Two datasets	CNN, BERT, GRU	0.764 to 0.787	It uses machine learning algorithms like ELMo, BERT, and CNN to detect hate speech on social media. Combining classifier results improves classification performance. The proposed approach has good accuracy and F1 score, making it a feasible option for social media hate speech.	Performance may depend on training data quality and representativeness and classifier and fusion algorithms used.
[38]	2020	16 datasets	CNN-GRU, BERT and LASER	0.93%	Employs pre-trained Google News Corpus word embeddings to capture semantic links between words and improve model performance. The model also uses convolutional and recurrent layers to collect local and global text information.	Requirements for massive training data, trouble capturing complicated syntactic patterns, and a potentially confusing architecture.
[44]	2021	223,549	CNN + BiGRU	75%	Achieves state-of-the-art performance on several evaluation metrics, including ROC_AUC score, precision, recall, and F1-score, compared to other existing models. Additionally, the proposed MCBiGRU model is designed to handle multilabel classification, which is more challenging than binary classification.	The dataset used for training and evaluation is limited; there is a lack of study regarding interpretability and extensive analysis of the computational resources needed for training and deployment.
[42]	2022	Numerous hate speech datasets	CNN + GRU	0.88%	Relevance of using current datasets and text context to improve hate speech detection models. The authors also emphasize the necessity to develop characteristics that may be applied to varied datasets and topics of interest to create more generalized and high-performance models.	Hate speech detection research is limited by a lack of unanimity on hate speech, bias in hate speech datasets, poor quality datasets, and the need for more generic model development.
[43]	2022	5,126 comments	BiLSTM, CNN, CNN-LSTM, GRU	70.10%	It gives unique insights into hate speech detection, its impact on society, and deep learning models' capacity to remedy this issue.	Language experts selected and labeled the dataset used to train and test deep learning models, which may not reflect popular opinion.
[46]	2022	One large dataset	CNN + GRU	78.57%	The ability to provide high accuracy.	The models worked on a small sample size, and further research is needed to validate the effectiveness of the CNN-GRU model on a larger scale.
[47]	2022	1180 tweets	CNN	0.615 %	Simple to use, not making assumptions about the data, and achieving high levels of accuracy. The authors conducted an F1-Score of 0.85 using the BERT algorithm, considered high accuracy.	The study was limited to a specific period and may not represent hate speech on Twitter. The study only focused on hate speech related to Asian individuals during the COVID-19 pandemic and did not consider other forms of hate speech or different periods.

[29]	2023	2400 tweets	CNN	94%	Advantages of some of the models in detecting abusive language in non-generalizable (unseen) problems and in terms of their speed and ability to load all the data quickly.	Handling different topical focuses and targets and the ability to load data quickly.
[48]	2023	183,472	CNN, CNN, and GRU	CNN= 87.94% CNN+GRU= 0.8733%	Used deep learning models and feature extraction techniques to achieve high accuracy in hate speech detection. The study also provides a detailed description of the methodology used, which can be helpful for researchers working on similar problems.	The study did not compare the performance of their proposed method with other state-of-the-art methods for hate speech detection. It did not analyze the computational resources required to train and test the models.

B. Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) for Detecting Hate Speech:

CNNs allow for fast processing of text data, extracting linguistic features and patterns that can be used to identify hate speech from other types of expression [49], [50]. Embeddings, convolution, pooling, and dense connections are just some of the layers that can train CNN models to recognize the linguistic and contextual cues indicative of hate speech. [51]addresses the difficulties of identifying hate speech on social media and the strategies that have proven successful in doing so. When evaluating the precision of various machine learning and deep learning approaches, the authors find that BiRNN (Bidirectional Recurrent Neural Network) excels. [52]Each time step, the LSTM neural network processes one word embedding from a multilayer perceptron (MLP) with two hidden layers, keeping word order. LSTM neural network output is processed via hyperbolic tangent activation. The number of terms must be specified before training the recurrent neural network. HaterNet detected at most 33 terms after preprocessing the tagged corpus tweets. All tweets under 33 terms have padding rows of 0s. In. After that, [53]benefits from identifying risky comments using LSTM neural networks. The authors cleaned data with NLTK and built their own stop words. The model had great precision, recall, and accuracy. The paper tries to filter hate speech and make social media safer. [54] Presents a comprehensive approach to detecting hate speech on Twitter using traditional machine and deep learning techniques. The authors explore shallow learning algorithms, including logistic regression, random forest, decision tree, naïve Bayes, K-NN, SVM, and deep learning methods such as LSTM, BiLSTM, and CNN. Further examines the difficulties linked to detecting hate speech on social media platforms. It provides insights into potential applications of the proposed approach for mitigating the impact of hate speech on individuals and communities. On the road to realizing their full potential, these models are trained extensively using labeled datasets and then fine-tuned [55]. Aside from helping with automatic identification, using this model for hate speech

detection allows for prompt interventions and mitigation methods, making the internet safer for everyone. CNNs and (LSTM) networks are crucial in identifying hate speech in textual data[56]. CNNs detect specific patterns and characteristics in the text using convolutional and pooling layers to extract important linguistic clues about hate speech. Meanwhile, (LSTM) models excel in comprehending the sequential structure of language, effectively capturing contextual interdependencies and enduring associations among words or phrases. By leveraging the capabilities of (CNNs) for extracting features and (LSTMs)[57], [58]. This collaboration allows for identifying hate speech content that is nuanced and embedded within its context. In [49],the study focuses on detecting hate speech diffusion on Twitter using graph-based methods. It finds that classification based on the sharing graph yields strong F1 scores for hate speech detection and highlights the vulnerability of existing textual hate speech detection methods to adversarial attacks. The study also considers the effects of automated bots in sharing hate speech content. The implications of the findings are relevant for addressing hate speech and online safety on social media platforms.[7]The DeepHate model utilizes multi-faceted text representations, including semantic, sentiment, and topical information, to improve hate speech detection. By combining pre-trained word embeddings, sentiment analysis, and Latent Dirichlet Allocation (LDA), the model achieves better performance compared to other configurations. Additionally, empirical studies provide insights into the salient features that aid in hate speech detection. [7]discusses the challenges of detecting hate speech in social media, particularly on platforms like Twitter. It explores deep learning approaches, such as various embeddings, to improve the detection of different types of hate speech. The experiments on publicly available datasets showed significant improvements in accuracy and F1-score, offering promising solutions to combat hate speech online. The authors [16] discussed cyberbullying detection using machine learning and deep learning approaches. Cyberbullying's potential dangers are highlighted, as are the benefits of effective detection methods. Methods for categorizing cyberbullying are



investigated in the research. The research examines machine learning and deep learning strategies for cyberbullying detection using two datasets: the Wikipedia Talk Corpus and the Twitter Hate Speech Corpus. The paper also addresses the role that technology plays in making cyberbullying more severe than in-person harassment. Overall, the research gives valuable insights into the detection of cyberbullying and the use of machine learning and deep learning technologies to solve this issue. [60] examines the challenges of detecting hate speech on social networking sites and the research in natural language processing (NLP) and machine learning (ML) to address this issue. The authors used an up-sampling method to balance the data. They implemented deep learning models like Long Short-Term Memory (LSTM) and Bi-directional Long Short-Term Memory (Bi-LSTM) for improved accuracy in detecting hate speech. LSTM was found to have better accuracy, precision, and F1 score, while Bi-LSTM had a higher recall. [61] Explores the problem of harmful language on social media platforms and suggests a technique for identifying it through the utilization of conventional machine learning models, as well as BERT and fastText embedding with deep neural networks. The authors merged the ALONE and HASOC'20 datasets to create a consolidated dataset for their research. The researchers pre-processed data and applied various machine learning and ensemble techniques, including TF-IDF, POS tagging, and trigrams. Among these approaches, LR and XGBoost yielded the most favorable results. In the second case, word embedding techniques such as fastText and BERT were employed to generate embeddings, which were subsequently utilized as inputs for DNN classifiers. The researchers used multiple deep neural network classifiers and found optimal performance was achieved by combining BERT embeddings with a convolutional neural network (CNN). [62] focused on hate speech detection methods using deep and shallow learning techniques. The study provides insights into the detection accuracy, computational efficiency, and practical implications of using pre-trained models and domain generalization. The paper presents a large-scale empirical evaluation of 14 shallow/deep classification-based hate speech detectors, evaluated on three large and publicly available hate speech detection benchmarks. [63] a study on detecting hate speech on Twitter using deep convolutional neural networks. The study uses machine learning-based classifiers such as Logistic Regression, Random Forest, Naive Bayes, Support Vector Machines, Decision Trees, and K-Nearest Neighbors to identify hate speech-related tweets on Twitter. The study also uses word embedding methods such as LSTM and Bi-LSTM models. The results show that the Deep Convolutional Neural Network architecture performs well in detecting hate speech on Twitter. Furthermore, in [54], the authors discussed detecting hate speech in social media using the LSTM algorithm. Results show that the capability to detect hate speech in online text automatically, high levels of accuracy, recall, and F1 score, and the potential to counteract hate speech on social media platforms are all features of this technology. In [55], BERT and Hate Speech Word Embedding with Deep Model detect hate speech in text

data. The paper describes feature and classifier approaches, reviews recent investigations, and includes datasets, embedding models, and experiment results. It successfully detects hate speech in English text data. [66] examines fuzzy categorization CNN-LSTM and Random Forest deep learning models. The research discusses Twitter hate speech detection issues and model efficacy. The discovery could be used in pattern recognition, machine learning, and AI beyond Twitter. [67] the research proposes an attentional multi-channel convolutional stacking Bidirectional LSTM network that uses word representation approaches to capture semantic relations at multiple windows. The model is compared to five state-of-the-art and five baseline models on three Twitter benchmark datasets. The presented model outperforms the others in most circumstances, and the absence of channels and attention mechanisms has the most significant influence, as proven in Figure 3. An online social network hate speech detection methodology can be utilized for data-driven cyber security.

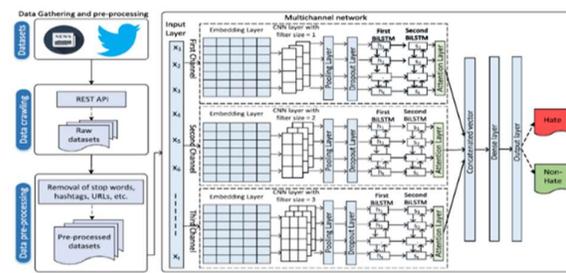


Fig. 1 CNN+LSTM proposed model [67]



Table 2. Overview of the literature on detecting hate speech Tweets based on CNN and LSTM.



Ref	Year	Datasets	Based Model	Accuracy	Advantages	Limitations
[51]	2019	Two datasets	LSTM	0.90%	Compares numerous Twitter hate speech detection techniques utilizing varied data ratios for a more complete evaluation. The article also evaluates machine learning and deep learning models for Twitter hate speech detection.	Only detects hate speech in tweets, not other social media or online platforms.
[52]	2019	16k tweets dataset	LSTM+MLP	0.895% and 0.892%	An integrated hate speech detection method using deep learning, feature engineering, and ensemble learning.	Due to the deep learning model's computational cost, the approach may not scale to massive datasets. The authors note that their method may not detect implicit or coded hate speech.
[53]	2020	Two datasets	LSTM	94.94%	The model classifies a statement as toxic or non-toxic and gives its proportion. High accuracy, precision, and recall scores on test data show that the model can distinguish harmful and non-toxic phrases.	It focuses on English comments exclusively; its performance in other languages is unknown. This may limit the model's use in multilingual online platforms.
[54]	2020	2 Wikipedia Talk Corpus and the Twitter Hate Speech Corpus	CNN+DBOW, CNN+DMM	CNN+DBOW=94% CNN+DMM=98.20%	CNN models are computationally efficient and can capture local features, while the DBOW and DMM models are more straightforward and faster to train.	DBOW and DMM models may not capture the order of words in a sentence, which can be important for hate speech classification.
[59]	2020	99,799	LSTM	F1 scores of 0.80% and recall scores of 0.93%	The study does highlight that graph convolutional networks yield robust F1 scores for the imbalanced classification task and high hate speech detection precision. Additionally, the study reproduces existing results showing how adversarial attacks can weaken text detection models for hate speech detection.	Adversarial attacks can easily fool current text-based detection methods, and simple manipulations of messages can impair the detection system's abilities.
[7]	2020	3 Public datasets	RNN+LSTM+Hybrid CNN	Not mentioned	Utilizes multi-faceted textual representations for automatic hate speech detection in social media. Empirical analyses of the Deep Hate model provide insights into the salient features that helped detect hate speech in social media.	Not mentioned any limitations.



[16]	2020	16K	CNN, LSTM, BiLSTM	CNN=90.47 % LSTM=90.55 %	The ability to detect various types of hate speech, even on platforms like Twitter, where contextual information is limited. These approaches significantly improve accuracy and F1 score, providing promising solutions to combat hate speech online.	The quality and representativeness of the training data, the generalization to new and diverse types of hate speech, and the potential biases in the training data.
[60], [61]	2021	51962 tweets	LSTM and BiLSTM	LSTM=0.9785% BiLSTM=0.9781%	LSTM excelled in accuracy, precision, and F1 score, whereas Bi-LSTM excelled in recall. The up-sampling technique adopted by the authors to achieve data balance ultimately led to better model results.	Without a large, labeled dataset, the models built using pure NLP concepts can be slower than those made using machine learning or deep learning models.
[61]	2021	25,000 tweets	CNN+BERT LSTM+GRU	13%	Combining two datasets increases the data's diversity and improves the results' generalizability. It also comprehensively analyzes various machine learning and deep learning techniques for hate speech detection on social media platforms.	This research is used only to detect English tweets and hate speech without categorizing them into different types of cyberbullying.
[62]	2022	25K	CNN+ BiLSTM	F1 score of 0.61% and a weighted average F1 score of 0.73 %. Weighted average model= 0.96-0.97%	Offers valuable insights into these methodologies' precision, effectiveness, and applicability. The authors also provide open access to their programs on GitHub, which might be helpful for scholars and practitioners seeking to reproduce or expand upon the work.	The study exclusively detects English-language hate speech; findings may not apply to other languages. The work only tests the models on three datasets; therefore, the results may not represent all hate speech scenarios. The results may be difficult to interpret because the publication needs to provide overall accuracy.
[63]	2023	3 Different datasets	LSTM, BiLSTM, and CNN	CNN=0.892 % LSTM=0.901 % BiLSTM=0.902%	We achieved higher accuracy than traditional machine learning algorithms, with BiLSTM outperforming other deep learning models.	A limited dataset, lack of generalizability, computational complexity, interpretability, and limitations of feature extraction techniques.
[64]	2023	Fixed partitioned datasets were used.	Bi-LSTM	Bi-LSTM=0.72 %	Ability to automatically detect hate speech in online text, high accuracy, recall, and F1 score, and its potential to combat hate speech on social media platforms.	It only addresses hate speech issues with text data and does not include images or videos. It also only uses tweets written in English, which limits its generalizability to other languages.
[65]	2023	Three datasets	CuDNNLSTM+BERT	0.96%	Pre-trained models can be fine-tuned for specific tasks, improving performance. Pre-trained models can also learn from enormous volumes of data to detect language nuances.	Code words or implicit language may prevent these models from detecting hate speech. Hate speech detection is subjective, and people may have various definitions.



[66]	2023	One dataset is divided into four categories	CNN-LSTM	75.35% to 88.69%	More complex and flexible categorization to catch hate speech and other damaging language. Decision boundaries can be more flexible, improving data acquisition, and classifications can account for uncertainty by enabling different confidence levels.	It is computationally intensive and may demand more resources than binary classification.
[67]	2023	Two datasets	BiLSTM	0.93%	On imbalanced datasets, the suggested model outperforms comparative models by 10%. Ablation analysis is used to evaluate model neural network components.	The dataset was restricted to English-language hate speech, ignored external influences, and did not analyze model interpretability.

III. RESULTS AND DISCUSSION

Hate speech detection on social media is a big concern in machine learning [68], with numerous researchers actively addressing hate speech originating from various sources [29]. CNN is considered one of the most effective algorithms for detecting hate speech, as it utilizes several methodologies [69], algorithms, and techniques [70]. Table 1 of this research presents an overview of 15 studies focusing on detecting hate speech using (CNN) and (GRU) and Table 2 gives an overview of 15 studies focusing on detecting hate speech using (CNN) and (LSTM). Moreover, the reviewed papers have been updated and published within the last five years. The CNN-based models achieve an accuracy ranging from 90% and higher in eighteen of the thirty works of literature—each of the previous research papers utilized separate datasets comprising several Twitter comments. The findings indicate that the quantity and nature of datasets have minimal impact on the precision of CNN-based models. In the end, the review articles prove that machine learning models are also used for detecting hate speech texts. Still, the accuracy and reliability of deep learning models such as CNN, LSTM, and GRU are higher among other algorithms. Half the researchers found that models designed for a single dataset may make mistakes. Another gap is that the models cannot detect other languages because they are trained on English tweets only.

IV. CONCLUSION

The application of techniques for "deep learning" to identify hate speech on Twitter is an exciting strategy for reducing the negative impacts of cyberbullying. Since language on social media constantly changes, rule-based approaches could be more effective; however, deep learning models are ideally adapted to this challenge. These models employ natural language processing and deep neural networks to analyze hate speech, and they show promise in picking up on nuanced contextual differences and shifting patterns of intolerance. However, there are still obstacles to overcome, such as the requirement for extensive and varied labeled datasets, the possibility of

biased training data, and the persistent development of language and online communication. Fine-tuning algorithms to reduce false positives is also essential because it is difficult to compromise between suppressing hate speech and protecting free speech. Incorporating deep learning models into content moderation systems on platforms like Twitter can make the internet safer as technology and research in this area evolve. Researchers, platform developers, and communities must continue to work together to improve models, combat biases, and cultivate a more welcoming digital space. This research reviews fifteen studies on detecting hate speech on Twitter using deep learning methods. Different models and other datasets were used, and the results show us that CNN is the most accurate, reliable, and easier to detect comments and then classify as offensive, sexual, racist...etc.

REFERENCES

- [1] W. Alorainy, P. Burnap, H. Liu, and M. Williams, "The Enemy Among Us: Detecting Hate Speech with Threats Based 'Othering' Language Embeddings," 2018, [Online]. Available: <http://arxiv.org/abs/1801.07495>
- [2] N. A. Kako and A. M. Abdulazeez, "Peripapillary Atrophy Segmentation and Classification Methodologies for Glaucoma Image Detection: A Review," *Current Medical Imaging Formerly Current Medical Imaging Reviews*, vol. 18, no. 11, pp. 1140–1159, Mar. 2022, doi: 10.2174/1573405618666220308112732.
- [3] L. Ketsbaia, B. Issac, and X. Chen, "Detection of hate tweets using machine learning and deep learning," *Proceedings - 2020 IEEE 19th International Conference on Trust, Security and Privacy in Computing and Communications, TrustCom 2020*, pp. 751–758, 2020, doi: 10.1109/TrustCom50675.2020.00103.
- [4] G. (Computer scientist) Wang, IEEE Computer Society, and Institute of Electrical and Electronics Engineers., *2020 IEEE 19th International Conference on Trust, Security and Privacy in Computing and Communications: proceedings :*



- 29 December 2020-1 January 2021, Guangzhou, China.
- [5] O. S. Kareem, A. M. Abdulazee, and D. Q. Zeebaree, "Skin Lesions Classification Using Deep Learning Techniques: Review," *Asian Journal of Research in Computer Science*, pp. 1–22, May 2021, doi: 10.9734/ajrcos/2021/v9i130210.
- [6] J. N. Saeed, A. M. Abdulazeez, and D. A. Ibrahim, "2D Facial Images Attractiveness Assessment Based on Transfer Learning of Deep Convolutional Neural Networks," in *ICOASE 2022 - 4th International Conference on Advanced Science and Engineering*, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 13–18. doi: 10.1109/ICOASE56293.2022.10075585.
- [7] R. Cao, R. K. W. Lee, and T. A. Hoang, "DeepHate: Hate Speech Detection via Multi-Faceted Text Representations," in *WebSci 2020 - Proceedings of the 12th ACM Conference on Web Science*, Association for Computing Machinery, Inc, Jul. 2020, pp. 11–20. doi: 10.1145/3394231.3397890.
- [8] K. Ismael Taher and A. Mohsin Abdulazeez, "Deep Learning Convolutional Neural Network for Speech Recognition: A Review," 2021, doi: 10.5281/zenodo.4475361.
- [9] A. Elouali, Z. Elberichi, and N. Elouali, "Hate speech detection on multilingual twitter using convolutional neural networks," *Revue d'Intelligence Artificielle*, vol. 34, no. 1, pp. 81–88, 2020, doi: 10.18280/ria.340111.
- [10] M. Jameel Barwary and A. Mohsin Abdulazeez, "Impact of Deep Learning on Transfer Learning : A Review IJSB Literature Review," 2021, doi: 10.5281/zenodo.4559668.
- [11] L. Alzubaidi *et al.*, "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," *J Big Data*, vol. 8, no. 1, Dec. 2021, doi: 10.1186/s40537-021-00444-8.
- [12] J. N. Saeed, A. M. Abdulazeez, and D. A. Ibrahim, "FIAC-Net: Facial Image Attractiveness Classification Based on Light Deep Convolutional Neural Network," in *2022 2nd International Conference on Computer Science, Engineering and Applications, ICCSEA 2022*, Institute of Electrical and Electronics Engineers Inc., 2022. doi: 10.1109/ICCSEA54677.2022.9936582.
- [13] A. Chaudhari, A. Parseja, and A. Patyal, "CNN based hate-o-meter: A hate speech detecting tool," in *Proceedings of the 3rd International Conference on Smart Systems and Inventive Technology, ICSSIT 2020*, Institute of Electrical and Electronics Engineers Inc., Aug. 2020, pp. 940–944. doi: 10.1109/ICSSIT48917.2020.9214247.
- [14] R. J. Hassan and A. Mohsin Abdulazeez, "Deep Learning Convolutional Neural Network for Face Recognition: A Review Literature Review," 2021, doi: 10.5281/zenodo.4471013.
- [15] J. N. Saeed and A. M. Abdulazeez, "Facial Beauty Prediction and Analysis based on Deep Convolutional Neural Network: A Review," *Journal of Soft Computing and Data Mining*, vol. 02, no. 01, Apr. 2021, doi: 10.30880/jscdm.2021.02.01.001.
- [16] P. Kapil, A. Ekbal, and D. Das, "Investigating Deep Learning Approaches for Hate Speech Detection in Social Media."
- [17] H. T. Sadeeq and A. M. Abdulazeez, "Metaheuristics: A Review of Algorithms," *International journal of online and biomedical engineering*, vol. 19, no. 9. International Association of Online Engineering, pp. 142–164, 2023. doi: 10.3991/ijoe.v19i09.39683.
- [18] J. N. Saeed, A. M. Abdulazeez, and D. A. Ibrahim, "Automatic Facial Aesthetic Prediction Based on Deep Learning with Loss Ensembles," *Applied Sciences (Switzerland)*, vol. 13, no. 17, Sep. 2023, doi: 10.3390/app13179728.
- [19] B. Charbuty and A. Abdulazeez, "Classification Based on Decision Tree Algorithm for Machine Learning," *Journal of Applied Science and Technology Trends*, vol. 2, no. 01, pp. 20–28, Mar. 2021, doi: 10.38094/jast20165.
- [20] L. Alzubaidi *et al.*, "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," *J Big Data*, vol. 8, no. 1, Dec. 2021, doi: 10.1186/s40537-021-00444-8.
- [21] H. T. Sadeeq and A. M. Abdulazeez, "Giant Trevally Optimizer (GTO): A Novel Metaheuristic Algorithm for Global Optimization and Challenging Engineering Problems," *IEEE Access*, vol. 10, pp. 121615–121640, 2022, doi: 10.1109/ACCESS.2022.3223388.
- [22] C. H. Salh and A. M. Ali, "Breast cancer recognition based on performance evaluation of machine learning algorithms," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 27, no. 2, pp. 980–989, Aug. 2022, doi: 10.11591/ijeecs.v27.i2.pp980-989.
- [23] H. Saeed Yahia and A. Mohsin Abdulazeez, "Medical Text Classification Based on Convolutional Neural Network: A Review," 2021, doi: 10.5281/zenodo.4483635.
- [24] K. Xia, J. Huang, and H. Wang, "LSTM-CNN Architecture for Human Activity Recognition," *IEEE Access*, vol. 8, pp. 56855–56866, 2020, doi: 10.1109/ACCESS.2020.2982225.
- [25] Z. A. Aziz and A. M. Abdulazeez, "Application of Machine Learning Approaches in Intrusion Detection System," *Journal of Soft Computing and Data Mining*, vol. 2, no. 2, Oct. 2021, doi: 10.30880/jscdm.2021.02.02.001.
- [26] Y. Sun, B. Xue, M. Zhang, G. G. Yen, and J. Lv, "Automatically Designing CNN Architectures Using the Genetic Algorithm for Image Classification," *IEEE Trans Cybern*, vol. 50, no. 9, pp. 3840–3854, Sep. 2020, doi: 10.1109/TCYB.2020.2983860.
- [27] E. Benavides, W. Fuertes, S. Sanchez, and M. Sanchez, "Classification of Phishing Attack



- Solutions by Employing Deep Learning Techniques: A Systematic Literature Review,” in *Smart Innovation, Systems and Technologies*, Springer Science and Business Media Deutschland GmbH, 2020, pp. 51–64. doi: 10.1007/978-981-13-9155-2_5.
- [28] H. Chen *et al.*, “A deep learning CNN architecture applied in smart near-infrared analysis of water pollution for agricultural irrigation resources,” *Agric Water Manag*, vol. 240, Oct. 2020, doi: 10.1016/j.agwat.2020.106303.
- [29] A. Z. Miran and H. S. Yahia, “Hate Speech Detection in Social Media (Twitter) Using Neural Network,” *Journal of Mobile Multimedia*, vol. 19, no. 3, pp. 765–798, 2023, doi: 10.13052/jmm1550-4646.1936.
- [30] C. H. Salh and A. M. Ali, “Comprehensive Study for Breast Cancer Using Deep Learning and Traditional Machine Learning”, doi: 10.21271/zjpas.
- [31] Z. Zhang, J. Tepper, and D. Robinson, “Detecting hate speech on Twitter using a convolution-GRU based deep neural network,” 2018. [Online]. Available: <https://www.researchgate.net/publication/323723283>
- [32] C. H. Salh and A. M. Ali, “Unveiling Breast Tumor Characteristics: A ResNet152V2 and Mask R-CNN Based Approach for Type and Size Recognition in Mammograms,” *Traitement du Signal*, vol. 40, no. 5, pp. 1821–1832, Oct. 2023, doi: 10.18280/ts.400504.
- [33] S. Modha, P. Majumder, T. Mandl, and C. Mandalia, “Detecting and visualizing hate speech in social media: A cyber Watchdog for surveillance,” *Expert Syst Appl*, vol. 161, Dec. 2020, doi: 10.1016/j.eswa.2020.113725.
- [34] D. Arya *et al.*, “Transfer Learning-based Road Damage Detection for Multiple Countries,” Aug. 2020, [Online]. Available: <http://arxiv.org/abs/2008.13101>
- [35] H. T. Sadeeq and A. M. Abdulazeez, “Car side impact design optimization problem using giant trevally optimizer,” *Structures*, vol. 55, pp. 39–45, Sep. 2023, doi: 10.1016/j.istruc.2023.06.016.
- [36] S. Zimmerman, C. Fox, and U. Kruschwitz, “Improving Hate Speech Detection with Deep Learning Ensembles.” [Online]. Available: <https://www.economist.com/news/europe/21734410->
- [37] Z. Zhang, D. Robinson, and J. Tepper, “Detecting Hate Speech on Twitter Using a Convolution-GRU Based Deep Neural Network,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Springer Verlag, 2018, pp. 745–760. doi: 10.1007/978-3-319-93417-4_48.
- [38] S. S. Aluru, B. Mathew, P. Saha, and A. Mukherjee, “Deep Learning Models for Multilingual Hate Speech Detection,” Apr. 2020, [Online]. Available: <http://arxiv.org/abs/2004.06465>
- [39] A. Al-Hassan and H. Al-Dossari, “DETECTION OF HATE SPEECH IN SOCIAL NETWORKS: A SURVEY ON MULTILINGUAL CORPUS,” Academy and Industry Research Collaboration Center (AIRCC), Feb. 2019, pp. 83–100. doi: 10.5121/csit.2019.90208.
- [40] AshwinGeetd’Sa, IrinaIllina, and DominiqueFohr, “Ashwin,” *Hal Open Science*, pp. 1–12, Jan. 2021.
- [41] Y. Zhou, Y. Yang, H. Liu, X. Liu, and N. Savage, “Deep Learning Based Fusion Approach for Hate Speech Detection,” *IEEE Access*, vol. 8, pp. 128923–128929, 2020, doi: 10.1109/ACCESS.2020.3009244.
- [42] F. Alkomah and X. Ma, “A Literature Review of Textual Hate Speech Detection Methods and Datasets,” *Information (Switzerland)*, vol. 13, no. 6. MDPI, Jun. 01, 2022. doi: 10.3390/info13060273.
- [43] G. O. Ganfure, “Comparative analysis of deep learning based Afaan Oromo hate speech detection,” *J Big Data*, vol. 9, no. 1, Dec. 2022, doi: 10.1186/s40537-022-00628-w.
- [44] A. K. J, A. S, T. E. Trueman, and E. Cambria, “Comment toxicity detection via a multichannel convolutional bidirectional gated recurrent unit,” *Neurocomputing*, vol. 441, pp. 272–278, Jun. 2021, doi: 10.1016/j.neucom.2021.02.023.
- [45] 2019 *International Conference on Intelligent Computing and Control Systems (ICCS)*. IEEE.
- [46] D. H. Shih, C. H. Liao, T. W. Wu, X. Y. Xu, and M. H. Shih, “Dysarthria Speech Detection Using Convolutional Neural Networks with Gated Recurrent Unit,” *Healthcare (Switzerland)*, vol. 10, no. 10, Oct. 2022, doi: 10.3390/healthcare10101956.
- [47] W. Zaghouni, J. Alberto Benítez-Andrades, U. de León, S. Mabrouka Besghaier, A. Abdelali, and A. Toliyat, “Asian hate speech detection on Twitter during COVID-19,” 2019.
- [48] K. U. Wijaya and E. B. Setiawan, “Hate Speech Detection Using Convolutional Neural Network and Gated Recurrent Unit with FastText Feature Expansion on Twitter,” *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI)*, vol. 9, no. 3, pp. 619–631, 2023, doi: 10.26555/jiteki.v9i3.26532.
- [49] A. Sharma, A. Zozan, and Z. R. Ahmed, “The 3D Facemask Recognition: Minimization for Spreading COVID-19 and Enhance Security.”
- [50] M. Umer, Z. Imtiaz, S. Ullah, A. Mehmood, G. S. Choi, and B. W. On, “Fake news stance detection using deep learning architecture (CNN-LSTM),” *IEEE Access*, vol. 8, pp. 156695–156706, 2020, doi: 10.1109/ACCESS.2020.3019735.
- [51] L. Jiang, K. Japan, and Y. Suzuki, “Detecting hate speech from tweets for sentiment analysis.” [Online]. Available: <https://www.kaggle.com/pandeyakshive97/hate-speech-dataset>.
- [52] J. C. Pereira-Kohatsu, L. Quijano-Sánchez, F. Liberatore, and M. Camacho-Collados, “Detecting



- and monitoring hate speech in twitter,” *Sensors (Switzerland)*, vol. 19, no. 21, Nov. 2019, doi: 10.3390/s19214654.
- [53] K. Dubey, R. Nair, M. U. Khan, and P. S. Shaikh, “Toxic Comment Detection using LSTM,” in *Proceedings of 2020 3rd International Conference on Advances in Electronics, Computers and Communications, ICAECC 2020*, Institute of Electrical and Electronics Engineers Inc., Dec. 2020. doi: 10.1109/ICAECC50550.2020.9339521.
- [54] L. Ketsbaia Northumbria University, B. Issac Northumbria University, and X. Chen Northumbria University, “Detection of Hate Tweets using Machine Learning and Deep Learning”, doi: 10.1109/TrustCom50675.2020.00103/20/\$31.00.
- [55] P. K. Sahoo, S. Mishra, R. Panigrahi, A. K. Bhoi, and P. Barsocchi, “An Improved Deep-Learning-Based Mask R-CNN Model for Laryngeal Cancer Detection Using CT Images,” *Sensors*, vol. 22, no. 22, Nov. 2022, doi: 10.3390/s22228834.
- [56] T. Van Huynh, V. D. Nguyen, K. Van Nguyen, N. L.-T. Nguyen, and A. G.-T. Nguyen, “Hate Speech Detection on Vietnamese Social Media Text using the Bi-GRU-LSTM-CNN Model,” Nov. 2019, [Online]. Available: <http://arxiv.org/abs/1911.03644>
- [57] F. Elmaz, R. Eyckerman, W. Casteels, S. Latré, and P. Hellinckx, “CNN-LSTM architecture for predictive indoor temperature modeling,” *Build Environ*, vol. 206, Dec. 2021, doi: 10.1016/j.buildenv.2021.108327.
- [58] S. Khan *et al.*, “BiCHAT: BiLSTM with deep CNN and hierarchical attention for hate speech detection,” *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 7, pp. 4335–4344, Jul. 2022, doi: 10.1016/j.jksuci.2022.05.006.
- [59] M. Beatty, “Graph-Based Methods to Detect Hate Speech Diffusion on Twitter,” in *Proceedings of the 2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2020*, Institute of Electrical and Electronics Engineers Inc., Dec. 2020, pp. 502–506. doi: 10.1109/ASONAM49781.2020.9381473.
- [60] C. Paul and P. Bora, “Detecting Hate Speech using Deep Learning Techniques.” [Online]. Available: www.ijacsa.thesai.org
- [61] P. Malik, A. Aggrawal, and D. K. Vishwakarma, “Toxic Speech Detection using Traditional Machine Learning Models and BERT and fastText Embedding with Deep Neural Networks,” in *Proceedings - 5th International Conference on Computing Methodologies and Communication, ICCMC 2021*, Institute of Electrical and Electronics Engineers Inc., Apr. 2021, pp. 1254–1259. doi: 10.1109/ICCMC51019.2021.9418395.
- [62] J. S. Malik, G. Pang, and A. van den Hengel, “Deep Learning for Hate Speech Detection: A Comparative Study,” Feb. 2022, [Online]. Available: <http://arxiv.org/abs/2202.09517>
- [63] A. Toktarova *et al.*, “Hate Speech Detection in Social Networks using Machine Learning and Deep Learning Methods.” [Online]. Available: www.ijacsa.thesai.org
- [64] S. Shekhar Pandey, I. Chhabra, R. Garg, and S. Sahu, “Hate Speech Detection,” *International Journal of Engineering and Management (IJAEM)*, vol. 5, p. 897, 2023, doi: 10.35629/5252-0504897903.
- [65] H. Saleh, A. Alhothali, and K. Moria, “Detection of Hate Speech using BERT and Hate Speech Word Embedding with Deep Model,” *Applied Artificial Intelligence*, vol. 37, no. 1, 2023, doi: 10.1080/08839514.2023.2166719.
- [66] A. Abraham, A. J. Kolanchery, A. A. Kanjookaran, B. T. Jose, and D. PM, “Hate Speech Detection in Twitter Using Different Models,” *ITM Web of Conferences*, vol. 56, p. 04007, 2023, doi: 10.1051/itmconf/20235604007.
- [67] M. Fazil, S. Khan, B. M. Albahlal, R. M. Alotaibi, T. Siddiqui, and M. A. Shah, “Attentional Multi-Channel Convolution With Bidirectional LSTM Cell Toward Hate Speech Prediction,” *IEEE Access*, vol. 11, pp. 16801–16811, 2023, doi: 10.1109/ACCESS.2023.3246388.
- [68] D. Marrugo, J. Carlos Martinez Santos, E. Puertas, D. Andres Marrugo-Tobón, and J. Carlos Martinez-Santos, “Natural Language Content Evaluation System For Multiclass Detection of Hate Speech in Tweets Using Transformers,” 2023. [Online]. Available: <https://github.com/EdwinPuertas>
- [69] G. M. Zebari, D. A. Zebari, D. Q. Zeebaree, H. Haron, A. M. Abdulazeez, and K. Yurtkan, “Efficient CNN Approach for Facial Expression Recognition,” in *Journal of Physics: Conference Series*, IOP Publishing Ltd, Dec. 2021. doi: 10.1088/1742-6596/2129/1/012083.
- [70] M. Jakubec, E. Lieskovská, B. Bučko, and K. Záborská, “Comparison of CNN-Based Models for Pothole Detection in Real-World Adverse Conditions: Overview and Evaluation,” *Applied Sciences (Switzerland)*, vol. 13, no. 9. MDPI, May 01, 2023. doi: 10.3390/app13095810.

