VOLUME
9
No. 2

EDITORIAL ADDRESS

Call for Papers

International Journal of New Media Technology (IJNMT) is a scholarly open access, peer-reviewed, and interdisciplinary journal focusing on theories, methods and implementations of new media technology. Topics include, but not limited to digital technology for creative industry, infrastructure technology, computing communication and networking, signal and image processing, intelligent system, control and embedded system, mobile and web based system, and robotics. IJNMT is published twice a year by Faculty of Engineering and Informatics of Universitas Multimedia Nusantara in cooperation with UMN Press.
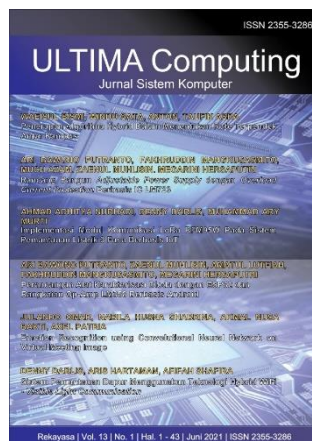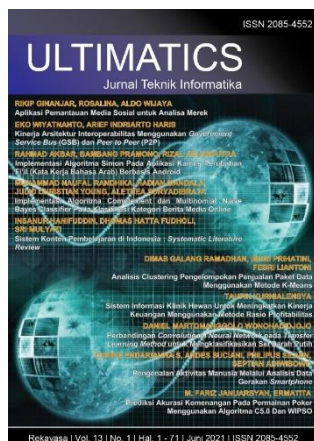
Ultimatics : Jurnal Teknik Informatika is the Journal of the Informatics Study Program at Universitas Multimedia Nusantara which presents scientific research articles in the fields of Analysis and Design of Algorithm, Software Engineering, System and Network security, as well as the latest theoretical and practical issues, including Ubiquitous and Mobile Computing, Artificial Intelligence and Machine Learning, Algorithm Theory, World Wide Web, Cryptography, as well as other topics in the field of Informatics.

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

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

FOREWORD

Greetings!

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

In this December 2022 edition, IJNMT enters the 1st Edition of Volume 9. In this edition there are six scientific papers from researchers, academics and practitioners in the fields covered by IJNMT. Some of the topics raised in this journal are: Analysis and Design of QR Code Based Information System on Plant Identification, Implementation of Backpropagation Method with MLPClassifier to Face Mask Detection Model, Implementation IT Governance Using COBIT 5 Framework at PT. XYZ (Persero), Indonesia Analysis Sentiment on Non Fungible Token (NFT), Predicting the Case of COVID-19 in Indonesia using Neural Prophet Model, Analysis Sentiment in Bukalapak Comments with K-Means Clustering Method.

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

Finally, we would like to thank all contributors to this December 2022 Edition of IJNMT. We hope that scientific articles from research in this journal can be useful and contribute to the development of research and science in Indonesia.

December 2022,

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

TABLE OF CONTENT

Analysis and Design of QR Code Based Information System on Plant Identification

Case Study : Environmental Service Dormitory RW 04 Tegal Alur Sub-District, Kalideres District, West Jakarta

Ahmad Syaikhu[1], Rizky M. L. Soeryaprawira[2], Yoga A. Daswara[3], Cornelius M. Sarungu[4]

Information Systems Department, BINUS Online Learning, Bina Nusantara University, West Jakarta, Indonesia
[1]ahmad.syaikhu@binus.ac.id
[2]rizky.soeryaprawira@binus.ac.id
[3]yoga.daswara@binus.ac.id
[4]cornelius.sarungu@binus.ac.id

Abstract— **At the location of the environmental service dormitory in RW 04 Tegal Alur sub-district, Kalideres district, West Jakarta, the community cultivate various types of plants and there is a requirement to introduce or educate the public about types of plants and their benefits. The purpose of this research project is to analyze the needs, as well as to design the QR Code-based information system, especially in the process of educating the types of plants and their benefits in the area. The design process uses the object-oriented analysis and design (OOAD) method which explores UML modelling (use case, activity diagram, sequence diagram, class diagram). The analysis will be carried out on to be implemented business processes, then the results of the analysis are poured into the design of a QR code-based information system in the form of web and mobile app that can assist educational activities. Web application will be used by the administrator to manage data to be displayed in mobile app. Mobile app will be used by the user to explore plant species and their benefits based on QR code that installed near the plants.**

Index Terms—information system; qr code; object oriented analysis and design;

I. INTRODUCTION

QR code is already used in many aspects of our live. Ranging from digital payment, transportation ticketing, logistics, products and documents identification etc. More or less, it has become part of our daily activities.

In RW 04 of Tegal Alur sub-district, Kalideres district, West Jakarta, the place of a local environmental service dormitory, the local community cultivate various medicinal and other types of plants [1]. Because of the variety of the plants are quite a lot and they want to record and store the information with expectation that every community member or even other people from outside of the community can access and gain knowledge from it. This need spawned up an idea to create a mobile-based application that can reveal the name of the plants by scanning the QR code on the plant with user smartphone, and also can provide additional useful information. The mobile app is expected to help people on knowing the name and the benefits of the plants and also create awareness of the importance of the plants to their environment.

In this project we are utilizing Quick Response code (QR code) technology as a tool to identify plants in the environment, it is hoped that detailed information about plants planted by local residents can be converted into a two-dimensional code that is printed into a more concise and easy media.

According to John W. Satzinger, Robert B. Jackson, Stephen D. Burd., system design is a set of activities that describe in detail how the system will run. It aims to produce software products according to user needs [2]. Producing software includes the activity of analysis and design.

QR code is invented in 1994 and used by a Japanese automotive company Denso Wave. It is come under category of 2D barcode. There are some variants of 2D barcodes, and it got popularized since early 2000. A decade later QR code emerge as the dominant one on the field, especially when it started to be used in digital payment. QR code have some significant advantages such as can contain large amount of information, consume less space, highly resistance to damage and

distortion, readable from any directions, scalable on size, have standardized process [3].

The exploitation of QR code to identify simplify the process of identification of things is already conducted since the emergence of the QR code itself. It already explored to sign and validate certificate of education back in 2011 [4]. In library service, QR code has been explored to store the information of a book complete with its code and location. It speeded up the process of searching or placement in its related rack in the library [5]. Even further, the QR code can also be used for digital library, where the users can scan the QR code of a book of their interest after searching via internal search engine, and then brought them to the download page from their mobile phone [6]. Another use of QR code in this field is seminar registration, student presence identification and recording and education tools registration and information distribution. On seminar registration its features consist of QR code generation, placement and presence recording of the seminar participants [7]. QR code can also be put on education tools of preliminary to high school, either for asset management or information distribution purpose [8]. Presence system is an integral part of the education system, the use of QR code level it up. It can be implemented on either onsite or online classes [9].

In business field, the QR code also has been used to support the daily business process. Employee presence system based on QR code already been used in business to replace the paper based traditional presence system [10]. Asset management system could also use the QR code on registering the assets and track their movement [11]. Another example is on mail administration process, where the QR code is used to replace signature. It increased the security aspect of the digital formal mail delivery between institution or inside the organization [12]. The application in digital payment maybe is the most popular. Even in a small-scale environment like a university, the students that are surveyed mostly ready and eager to use the digital payment using QR code and mobile app [13].

In this project, we try to implement the QR code to help people doing the identification of plants that grow around them, and extract any information about the particular plant, such as its benefits and how to maintain it. Previous research has been conducted in Jompie botanical garden to implement the use of QR code for plant identification and information distribution [14]. The differences with our project are in the exploitation of the cloud platform and printing process of the QR code.

The reasons we choose mobile app to be used in our project are the access to the native functionality such as camera to scan the QR code, and the offline data access support where we could design a local data cache mechanism that allow the user keep using the functionality of the app while offline, and it is relatively faster than web app [17].

II. METHOD

In this project we used simple methodology as presented in Fig. 1. The first step is conducting data gathering. After most required data are captured and recorded, we use it on analysis of the existing system, and it will lead us to the problem identification stage. After the problems are defined, we then jump into the solution design activity that will resulted in new system design.

The methodology used in this project (Fig. 1) is actually is a section of the waterfall method of software development lifecycle. It is not fully adopted because of the scope is only to design phase. The scope of this project is to analysis and design, so we only cover the analysis process, result and design elements based on the object-oriented analysis and design method.
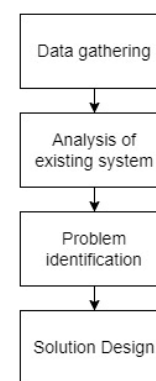
Fig. 1. Methodology used in this project.

The data gathering, analysis of existing system and problem identification are part of the analysis phase, while the design phase is in the solution design.

On the data gathering we conduct the interview and observation of recent process. The results then used to create the process map of existing system, while the copy of documents and any other artifact will be used in designing the database and forms. After the process map is built, analysis is done based on it. Process review together with the environment service team is done to map all problems that already spotted, identify problems that probably still hidden and to get deeper understanding about the whole process. Once all problems are identified, risks related to each problem are determined. Next step is to map the solutions or features for each identified problem. Solution design then conducted to design each feature using object-oriented design method supported with UML diagrams.

III. RESULT AND DISCUSSION

A. Data Gathering

In carrying out the analysis and design of a QR code-based system, several methods are used to collect data, namely as follows:

1. The method of analysis consists of:

    a.   Conduct observations and field studies directly at the dormitory of the environmental service office in the form of interviews with the local people about what they need from an information system to help them find out the names and benefits of each plant that they cultivate and grow around their area.

    b.   Analysis of observation findings and field studies: Analyze the results of observations and field studies data that have been obtained to find weaknesses and problems in the current system

    c.   Identification of information requirements: Identifying the system requirements needed for system design for information exposure of cultivated plants.

    d.   Identification of system requirements: Identify the system requirements required for the design of the expected system.

2. Literature study is also conducted, and it consists of reading and studying literature, books, essays and journal papers that highly related to existing problems.

B. Analysis of Existing System

Current system of plants information management is done by the environmental service team and it was done manually. We could observe from flowchart in Fig. 2 that depicts how current system works.
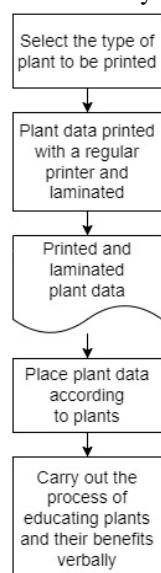
Fig. 2. Process flow of existing system.

    e.   The team identified as many plants as they can around their environment and record them manually in a record book. Then they select which plant will be marked with name and other information. The name and information to

be displayed are printed in a plain paper and laminated, coated with transparent plastic surface (Fig. 3). The lamination is meant to make the stuf weatherproof, but in fact it cannot last long, the endurance just get shorter on rainy season

Fig. 3. Photos of the paper-based information on the plant.

The laminated paper then placed close to the related plant. After most of the plants are marked, the management of environmental service team then launch an education program to the local community. They educate the community about the benefits and the use of each plants. The expectation of this program is that the awareness about useful plants in the community will increase gradually and they can easily utilize those plants when needed. The education still be done verbally with one trainer guide a group of local people and doing a short trip touring around the area. Along the trip the trainer explained the information of each plants to the group.

C. Problem Identification

From the current system flow and the interview results we could identify some problems in it. We used Ishikawa "fishbone" diagram to map all the problems and its categories (Fig.4).

Fig. 4. The Fishbone diagram of the problems.

The identified problems are presented in table below (Table I) along with the risks.

TABLE I. Identified Problems.

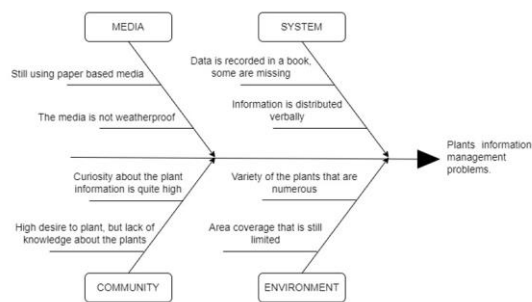| No.	Problem	Risk
1	Plants data management that still being done manually. Plants data still recorded in a book (paper-based media).	Activities done on the data management are not well recorded. Plants data could be gone or erased if something happen with the book.
2	Plants data is printed in a plain paper, and laminated. It's not a weatherproof.	Plants data is susceptible to erased by the rough weather.
3	The community education process is done verbally.	Repeating the same work every time a new group come.
4	There are still many people around the area that do not know about the plants, its benefits and how to utilize it.	The plants will be left ignored and unused, regardless of the potentials and high benefits.
5	High desire to plant but lack of knowledge about the plants.	The people probably will plant any of the vegetation they like without ever knowing the purpose and benefit.
6	The information is distributed verbally.	The information become difficult to remember and accessed.
7	Variety of the plants that are numerous.	Without a storage media and information system, it would be difficult to manage the data.
8	Area coverage that is still limited.	More man power or work hours is needed to enlarge the coverage area if still using the manual method.

D. Solution Design

Based on the problems faced by the environmental service dormitory team, we define a solution equipped with features related with the problem list on Table 1. After discussing with the environment service team and some local informal leaders, we agree on deciding to build an information system to help the team eliminate those problems. The agreed solution to be implemented is a mobile app with web administration and services module as the backend. The features to be put in those app are listed below (Table II).

TABLE II. Solution list.

No.	Problems	Solution / Features
1	Plants data management that still being done manually. Plants data still recorded in a book (paper-based media).	Web application for administration purpose.
2	Plants data is printed in a plain paper, and laminated. It's not a weatherproof.	Replacing the plain paper with synthetic yupo, and use the Zebra GT800 printer with full resin ribbon to create a weatherproof label for the QR code.
3	The community education process is done verbally.	Mobile app to access the information. Any people can install the app from playstore, register and use it immediately.
4	There are still many people around the area that do not know about the plants, its benefits and how to utilize it.	Same as number 3. Enrich the data of plants in the area. There should be a feature to show all the necessary information about the plant, including its name, type, category, latin name, description, benefit, how to plant, grow and maintain, etc.
5	High desire to plant but lack of knowledge about the plants.	Same as number 4.
6	The information is distributed verbally.	Same as number 4.
7	Variety of the plants that are numerous.	Database connected to the web application to store all data. QR generation for each plant.
8	Area coverage that is still limited.	Same as number 4. Still need to go around the area to put the new labels.

The flowchart of system requirements in the environmental service dormitory area can be seen through the flowchart below (Fig.5 and 6).

We design the new process with two separate roles, the administrator and the user. The administrator is person who manage the configuration, information and also the QR code from generating, printing to onsite installation. The process flow is shown in Fig. 5. The other role is the user, person who register, login into the app, scan the QR code and get the information of the plant. The process flow is shown in Fig. 6.

From system architecture perspective, to accommodate those two separate roles we have to connect the mobile app, that function as the frontend, to the web application that serve as the backend. The architecture diagram is shown in Fig. 7.
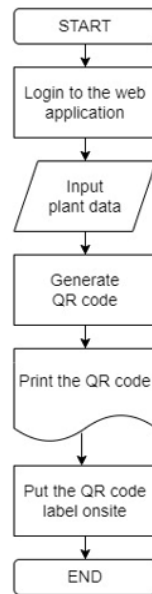
Fig. 5. Process flow of the administrator role

Fig. 7. The architecture diagram of the system.

In our proposed architecture, application server, database server and web service server are hosted in the google cloud platform. After all of those servers are activated, all the services listed in web service registry can be accessed by the mobile app and web client. Known previous research is using proprietary server architecture on its implementation [14]. Apache server is used to host the PHP based web application. Apache is well known for its easy configuration and secure socket layer (SSL) support [16]. The protocol used is secure http (https) for security reason, because this will be used as public service application.

After defining the process flows and the system architecture, we continue into the other design elements of this QR code-based plant information system consists of use case diagrams, detailed use case scenario or description, activity diagrams, sequence diagrams, class diagrams, web user interfaces, mobile user interfaces and system specifications.

The next step for designing a QR code-based system is making use case diagrams for web application and mobile app as follows:

Fig. 6. Process flow of the user role.

Fig. 8. Web application use case diagram.

The administrator can login to the web application and taken into the homepage. On the main menu, the administrator can choose to manage user data, manage plant data and manage articles. Each of the data management use case includes view, add new, edit and delete actions (Fig. 8).

Fig. 9. Mobile app use case diagram.

The user can login into the mobile app and taken into the main menu. The user then can scan QR code, view articles and manage their profile (Fig. 9).

The next step is to create a detailed use case scenario of each use case diagram above. The following are nine examples of the use case scenario, seven are use cases of the web application and two use cases are from the mobile app design (Table III-XI):

TABLE III. Use case scenario of the main menu of the web application.
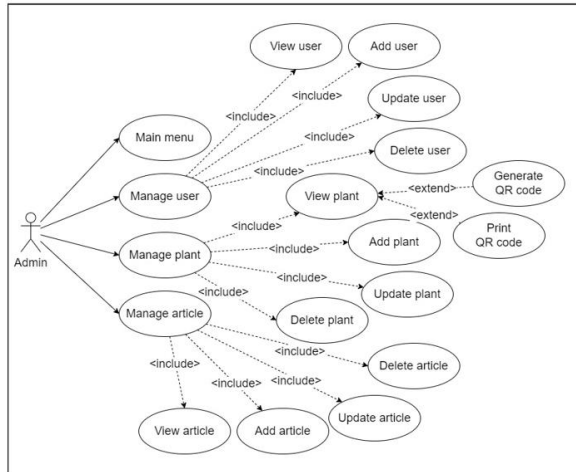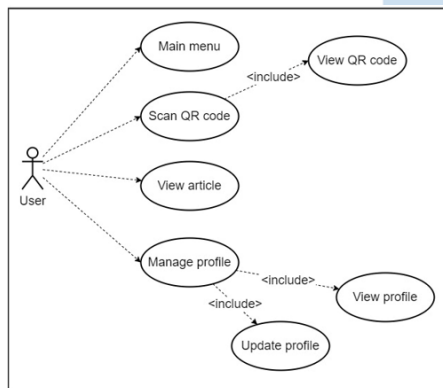
Use Case Name:	Main menu
Scenario:	Actor want to visit the homepage of the web.
Triggering Event:	Actor is Login.
Brief Description:	Actor need to visit the homepage of the web.
Actors:	Administrator
Related Use Case:	-

Stakeholders:	Administrator	
Preconditions:	Actor already register and login.	
Post Conditions:	After actor registered and login, the system will take the actor to the web homepage.	
Flow of activities	**Actor**	**System**
	1. Actor choose Register option. 2. Actor input the data. 3. Click Register button.	1.1 Display the registration form. 2.1 – 3.1 Validate the registration form.
	4. Actor choose Login option. 5. Actor input username and password. 6. Actor click Login.	4.1 Display the login form. 5.1 - 6.1 Validate the input data. 6.2 If input data valid, display the web homepage. 6.3 If input data invalid, show error message and back to the Login page.
Exception Conditions:	2.1 Invalid email/username. 2.2 Invalid password.	

TABLE IV. Use case scenario of manage plant data, part of the web application.

Use Case Name:	Manage plant data	
Scenario:	Actor need to manage the plants data.	
Triggering Event:	Actor choose the Manage Plants Data menu option.	
Brief Description:	Actor need to see the plants list and manage the plants data.	
Actors:	Administrator	
Related Use Case:	Main menu Show plant data Add plant data Edit plant data Delete plant data	
Stakeholders:	Administrator	
Preconditions:	Actor already login and choose Manage Plants Data menu option.	
Post Conditions:	System will display the Manage Plant Data page and the list of plants.	
Flow of activities	**Actor**	**System**
	1. Actor click the Manage Plant Data option.	1.1 Display plants list.
Exception Conditions:	1.1 Database down	

TABLE V. Use case scenario of the user data management of the web application.

Use Case Name:	Manage user data
Scenario:	Actor need to manage user data that already register from mobile app.
Triggering Event:	Actor choose the Manage User Data option.
Brief Description:	Actor need to see the list of users and manage the user data.
Actors:	Administrator
Related Use Case:	Main menu Show user data Add user data Edit user data

	Delete user data	
Stakeholders:	Administrator	
Preconditions:	Actor already login and choose Manager User Data options.	
Post Conditions:	After click Manage User Data, system will display the registered users list.	
Flow of activities	**Actor**	**System**
	1. Actor click Manage User Data menu option.	1.1. Display all the registered users list.
Exception Conditions:	1.1 Database down	

TABLE VI. Use case scenario of article management, part of the web application**.**

Use Case Name:	Manage article	
Scenario:	Actor need to view articles.	
Triggering Event:	Actor choose Manage Article option.	
Brief Description:	Actor need to view and manage articles in the Manage Article menu option.	
Actors:	Administrator	
Related Use Case:	Main menu Show article Add article Edit article Delete article	
Stakeholders:	Administrator	
Preconditions:	Actor already login and choose Manage Article menu option.	
Post Conditions:	System will display the Manage Article page and the list of articles.	
Flow of activities	**Actor**	**System**
	1. Actor click Manage Article menu option.	1.1. Display list of articles.
Exception Conditions:	1.1 Database down	

TABLE VII. Use case scenario of data presentation, part of the web application.

Use Case Name:	Show plant data	
Scenario:	Actor need to view plant data.	
Triggering Event:	Actor choose one of the plant data from the list of plants.	
Brief Description:	Actor need to view a plant data in detail.	
Actors:	Administrator	
Related Use Case:	Manage plant data Generate QR code Print QR code	
Stakeholders:	Administrator	
Preconditions:	Actor already login and click one of the plants in the list of plants.	
Post Conditions:	System shows the detail data of the selected plant.	
	Actor	**System**
Flow of activities	1. Actor select one plant from the list.	1. Display detail data of the selected plant.
Exception Conditions:	1. Database down	

TABLE VIII. Use case scenario of generate QR code, part of the web application.

Use Case Name:	Generate QR code	
Scenario:	Actor need to generate QR code of a plant.	
Triggering Event:	Actor choose a plant and click Generate QR code option.	
Brief Description:	Actor generate QR code of a plant.	
Actors:	Administrator	
Related Use Case:	Show plant data	
Stakeholders:	Administrator	
Preconditions:	Actor already login. Actor already clicked the Manage Plant Data menu option. Actor click generate QR code.	
Post Conditions:	QR code generated and stored in the database.	
Flow of activities	**Actor**	**System**
	1. Actor click Generate QR code in the table row.	1.1 Generate QR code of the selected plant. 1.2 If the QR can be generated, system shows a success notification along with the picture of the QR. QR data is saved in the database. 1.3 If the QR failed to generate, system shows an error notification. Will go back to the Manage Plant Data page when the user clicked the OK button.
Exception Conditions:	1.1 Database down	

TABLE IX. Use case scenario of print QR code, part of the web application**.**

Use Case Name:	Print QR code
Scenario:	Actor wants to print the QR code of the selected plant.
Triggering Event:	Actor click the Print QR on the selected plant data.
Brief Description:	Actor prints the QR code of a selected plant.
Actors:	Administrator
Related Use Case:	Show plant data
Stakeholders:	Administrator
Preconditions:	Actor already login and click one of the plants in the list of plants.

		error message and asking the user to input again the proper username/ password.
Exception Conditions:	1.1 Smartphone is not connected to the internet. 3.1 Duplicated data. 4.1 Smartphone is not connected to the internet. 6.1 Invalid username/password.	

Post Conditions:	QR code of the selected plant is printed.	
Flow of activities	**Actor**	**System**
	1. Actor click the QR code of a plant in table row. 2. Actor click the Print QR code option.	1. Display the QR code in a pop-up window. 2. Print QR Code in selected media.
Exception Conditions:	1.1 Database down 2.1 Disconnected from local network 3.1 Printer cannot be accessed	

The following tables are use case scenario examples from the mobile app module (Table X-XI):

TABLE X. Use case description of main menu in mobile app.

Use Case Name:	Main menu	
Scenario:	The Actor want to visit the main menu in mobile app.	
Triggering Event:	The actor is doing login.	
Brief Description:	The actor is visiting the main menu in mobile app.	
Actors:	User	
Related Use Case:	Show plant data Show user data	
Stakeholders:	User	
Preconditions:	The actor is already registered as user and doing login.	
Post Conditions:	The homepage is displayed and the actor can access the main menu options.	
Flow of activities	**Actor**	**System**
	1. Actor choose Register. 2. Actor input data in Full Name, Email, Password and Confirm Password fields. 3. Actor click "Register" button.	1.1 Display the registration form. 2.1 – 3.1 System will store the data input by the actor.
	4. Actor choose Login. 5. Actor fill the username and password. 6. Actor click the login button.	4.1 System displaying the login form. 5.1 – 6.1 Validate the input data. 6.2 If input data valid then system will display the main menu. 6.3 If the input data is invalid then the system display an

TABLE XI. Use case description of scan QR code, part of the mobile app.

Use Case Name:	Scan QR code	
Scenario:	The actor wants to scan the QR code from the mobile app.	
Triggering Event:	The actor click scan QR code.	
Brief Description:	The actor can access the mobile phone camera to scan the QR code.	
Actors:	User	
Related Use Case:	Login, Show plant data	
Stakeholders:	User	
Preconditions:	The actor is already login and choose the scan QR code.	
Post Conditions:	QR code scanned and the plants data is shown.	
Flow of activities	**Actor**	**System**
	1. Actor click scan QR Code 2. Actor direct the camera to the plants QR code.	1.1 System will display the mobile phone camera. 2.1 QR code data validation. 2.2 If the QR code data is valid then system will display the plant data. 2.3 If the QR code data is invalid then the system will pop up an error message and going back to the mobile camera display to recapture the QR code.
Exception Conditions:	1.1 The camera is broken or error. 2.1 Invalid QR code data.	

Next step we build the activity diagram for each use case that describe the process flow of each use case. The following are examples of activity diagram from the system (Fig. 10-12):
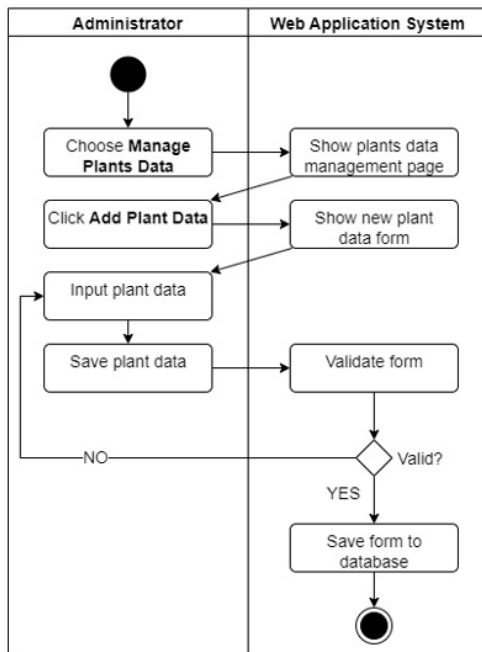
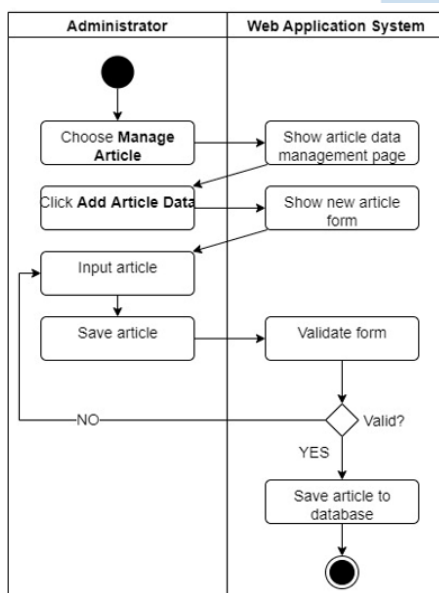Fig. 10. Activity diagram of Add Plant Data.

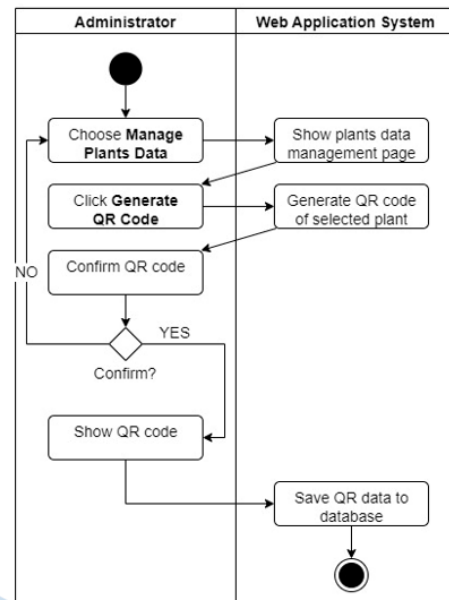Fig. 11. Activity diagram of Add Article Data.

Fig. 12. Activity diagram of Generate QR Code.

After the activity diagrams are finished and reviewed, we then build the sequence diagram of each use case. The following are the sequence diagram examples of the system (Fig. 13-16):
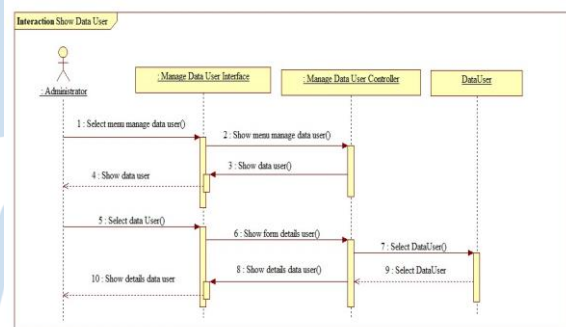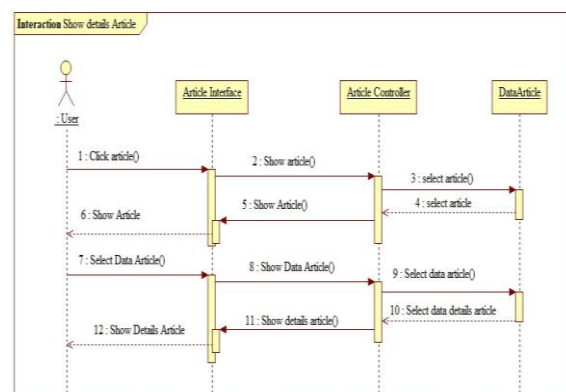

Fig. 13. Sequence diagram of main menu.

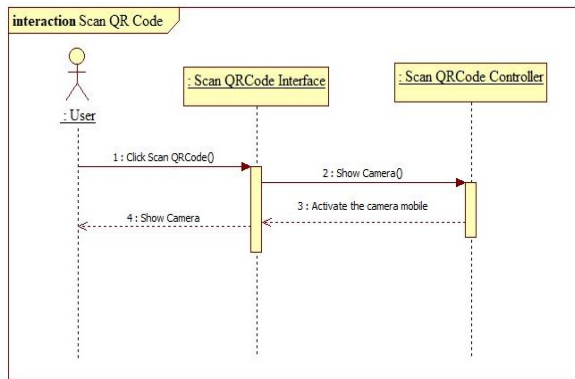Fig. 14. Sequence diagram of view article detail.
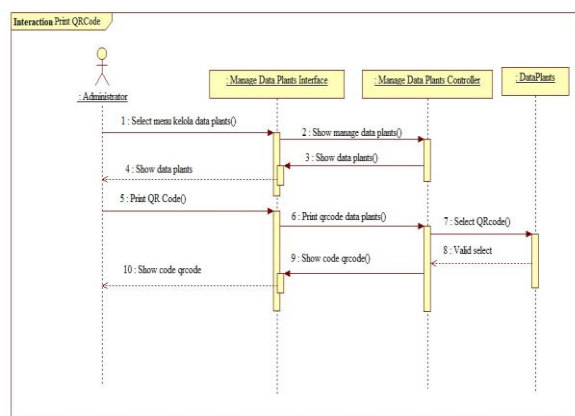
Fig. 15. Sequence diagram of scan QR code.

Fig. 16. Sequence diagram of Print QR code.

After finishing and reviewing the sequence diagrams, we continue to the next step which is building the class diagram of the system. The following is the class diagram design (Fig. 17):
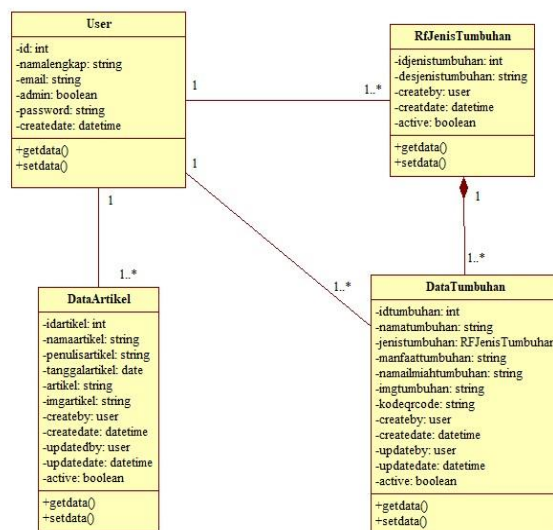

Fig. 17. Domain class diagram of the system.

Next step we define the data dictionary that consists of data items that will be saved into the database. We provide tables below for the data dictionary (Table XII-XV)

TABLE XII. User data.

Field	Type	Null	Default
id	int(6)	No	*Primary Key*
namalengkap	varchar(50)	Yes	
email	varchar(50)	Yes	
admin	bit(1)	Yes	
password	varchar(50)	Yes	
createdate	datetime(8)	Yes	

TABLE XIII. Plant data.

Field	Type	Null	Default
idtumbuhan	int(4)	No	*Primary Key*
namatumbuhan	varchar(100)	Yes	
jenistumbuhan	int(4)	Yes	
manfaattumbuhan	varchar(500)	Yes	
namailmiahtumbuhan	varchar(100)	Yes	
imgtumbuhan	varchar(200)	Yes	
kodeqrcode	varchar(50)	Yes	
createby	int(4)	Yes	
createdate	datetime(8)	Yes	
updateby	int(4)	Yes	
updatedate	datetime(8)	Yes	
active	bit(1)	Yes	

TABLE XIV. Plant category data.

Field	Type	Null	Default
idjenistumbuhan	int(4)	No	*Primary Key*
desjenistumbuhan	varchar(50)	Yes	
createby	int(4)	Yes	
creatdate	datetime(8)	Yes	
active	bit(1)	Yes	

TABLE XV. Article data.

Field	Type	Null	Default
idartikel	int(4)	No	*Primary Key*
namaartikel	varchar(100)	Yes	
penulisartikel	varchar(50)	Yes	
tanggalartikel	date(3)	Yes	
artikel	varchar(500)	Yes	
imgartikel	varchar(200)	Yes	
createby	int(4)	Yes	
createdate	datetime(8)	Yes	
updateby	int(4)	Yes	
updatedate	datetime(8)	Yes	
active	bit(1)	Yes	

After done with the structural design of the system which represented by the class diagram and data dictionary, we continue the design process to the mock up or user interface design.

In web application homepage (Fig. 18), there are three menu options that can be selected, they are : Manage plant (kelola tumbuhan), Manage users (kelola user) and manage article (kelola artikel).
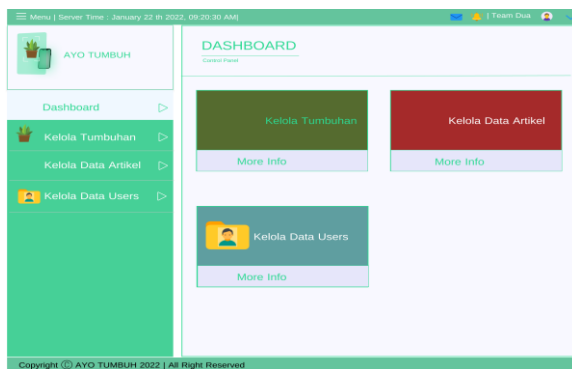
Fig. 18. Web application homepage.

In plant management page (Fig. 19), there is a table containing list of plants that have been stored in the database. We could see most important information about the plant (plant name, category, benefit of the plant, latin name, image, QR code, audit logs and action icons).
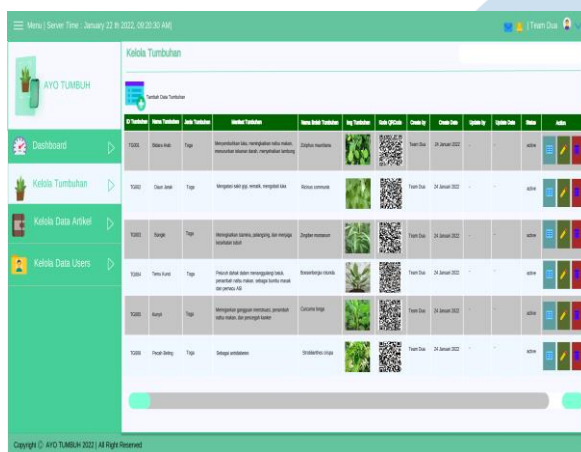

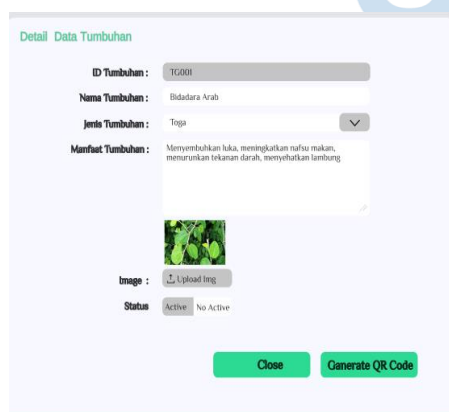Fig. 19. Plant management page.

Fig. 20. Plant data detail page.

When we select one of the rows in the table and click the plant picture or edit icon, the system will display a pop-up form containing the complete information about the plant (Fig. 20). The administrator could

upload the plant image, set plant data status (active/inactive) and generate the QR code. After the administrator generate the QR code, the system will display the result along with close, print and save buttons. Printing is allowed in this stage (Fig. 21). The size of the QR code should be big enough to get the public attention and trigger the desire to scan it [15].

Fig. 21. Print QR code UI.

In manage articles menu, the administrator can add new article, view, edit and delete existing articles (Fig. 22). The display is the same with user management and plant management menu, there is a list of items and the administrator can select one item to do related actions on it.
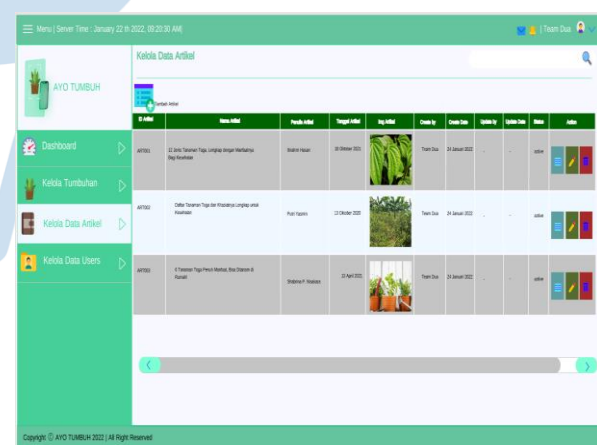

Fig. 22. Article management page.

When the administrator wants to add new article, they can click the add article icon in the top left corner above the table. System will display an article form to be filled, then can be saved or cancelled (Fig. 23).
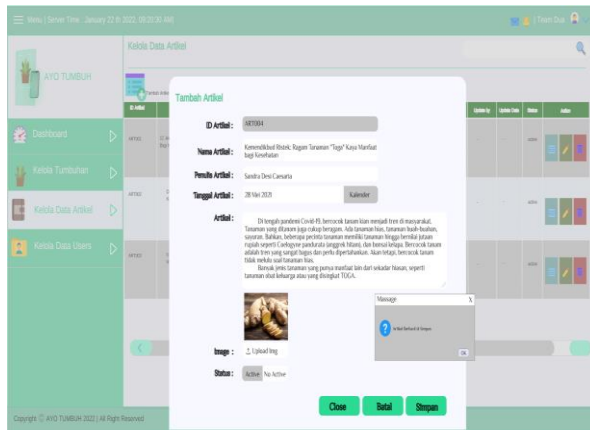
Fig. 23. Add new article page.

On the user side, after the app is installed in the mobile phone, the user opened the app and brought to the welcome page that shows some information (Fig. 24.a-c). The user then can register (Fig. 25.a), login (Fig. 25.b) and also click forgot password if they forget their password (Fig. 25.c). After successful login, they will be brought to the homepage (Fig. 26.a). In the area of service, around the environmental service dormitory, the user can scan any QR code that already been installed on any plants they found along the way and related plant information will be displayed immediately (Fig. 26.b-c).
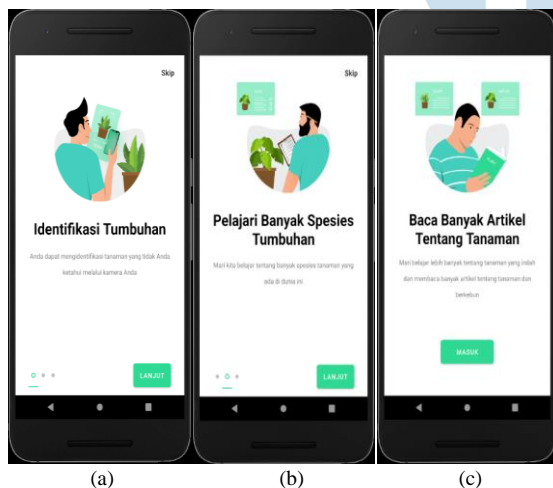

(a)   (b)   (c)

Fig. 24. Front page of the mobile app.

(a)   (b)   (c)
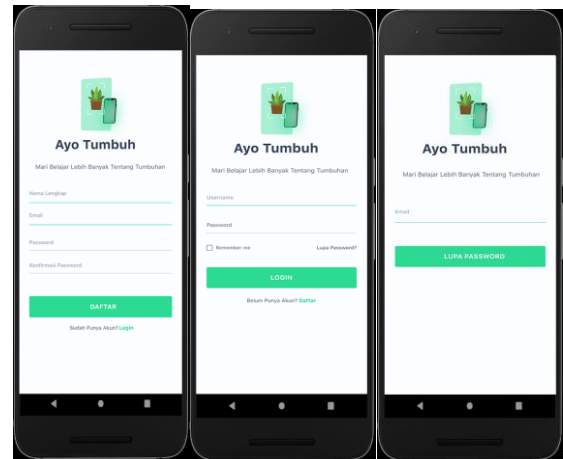
Fig. 25. The register, login and forgot password pages of the mobile app.

(a)   (b)   (c)
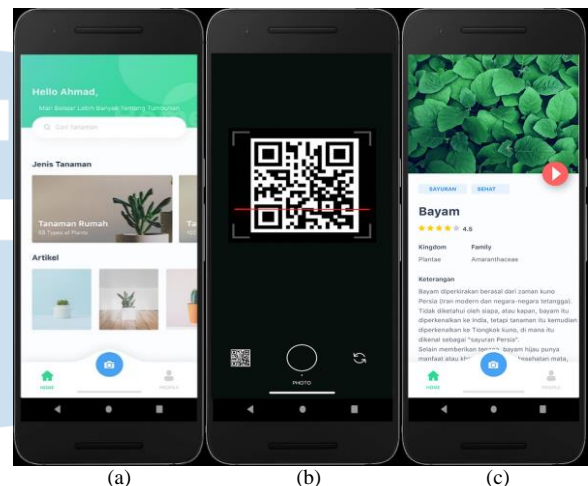
Fig. 26. The mobile app UIs (Homepage, Scan QR and Display Information).

Given below are the specifications for hardware and software to operate the system:

TABLE XVI. Web system specification.

	Hardware	Software
Cloud Server	1. Processor: Intel Xeon E3-1270 v6 4 Cores, 3.80 GHz	1. Windows Server
	2. Memory: 16 GB RAM	2. MYSQL
	3. Storage: 1 x 1 TB HDD	3. NET Framework 4.5
	4. bandwidth: 20 TB bandwidth	

TABLE XVII. Mobile app specification.

	Hardware	Software
HP Android	1. Processor: Dual-core 1 GHz Cortex-A9	1. Jelly Bean (4.2.2) – Red *Velvet* (*Android 11*) 2. iOS 10 – iOS 15
	3. Memory: Internal 4 GB, 512 MB RAM	2. Android Studio dan iOS
	4. Display: Capacitive touchscreen, 480x800 pixels, 4.0 Inches	

	5. Camera 5 MP, 2592 x 1944 pixels	

IV. CONCLUSION

The results obtained from the analysis and design during the implementation of this project are as follows:

1. With the design of a QR code-based information system on plants, can help the dormitory of the environmental service RW04 sub-district Tegal Alur, district Kalideres, West Jakarta to increase and widened their services to the pulic of the area.

2. The implementation of QR code scanning on plant identification and information extraction increase the efficiency of information distribution to the public.

3. The use of mobile app expanded the coverage area of services, but it should be followed by plant identification actions that include the placement of the QR code on the plants.

4. The change on the QR code printing media increase durability and the resistance against the weather.

5. Data on plant species obtained during the research at the dormitory of the environmental service office RW04 sub-district Tegal Alur, district Kalideres, West Jakarta, only a few RTs namely RT02, RT04 and RT05 out of a total of 8 RTs that can provide information related to plant species, so the data obtained can still be optimized and completed in the future for educational facilities about the plants they grew.

6. This model still has a risk of loosing the label because of stealing attempt.

7. Further research project could be initiated by exploring the use of artificial intelligence (AI) to detect the plant, or at least detecting the leave pattern. It will remove the need of using QR or barcode.

ACKNOWLEDGMENT

REFERENCES

[1] Kesatu.co. "RW 04 asrama DLH Tegal Alur, kampung asri di tengah kota Jakarta," (Online). Accessed: 12-17-2021. Available: https://kesatu.co/rw-04-asrama-dlh-tegal-alur-kampung-asri-di-tengah-kota-jakarta/.

[2] John W. Satzinger, Robert B. Jackson, Stephen D. Burd. "System analysis and design in a changing world," (7th edition). Boston: Cengage Learning. 2016.

[3] M. Querini, A. Grillo, A. Lentini, G.F. Italiano. "2D color barcodes for mobile phones," International journal of computer science and applications, 8(1), 136-155. 2011.

[4] Anita Rahmawati, Arif Rahman. "Sistem pengamanan keaslian ijasah menggunakan qr-code dan algoritma base64," Jurnal sistem informasi Indonesia, 1(2), 105-112. 2011.

[5] Rastri Prathivi. "Analisa sistem QR code untuk identifikasi buku perpustakaan," Jurnal Pengembangan Rekayasa dan Teknologi, 14(2), 37-40. December, 2018.

[6] Muhammad Fabio Armadani, Dedi Mulyadi. "Digitalisasi manajemen sistem dokumen pengelolaan perpustakaan menggunakan qr code dan digital signature," Jurnal buana ilmu, 6(1), 13-19. 2021.

[7] Zen Munawar. "Aplikasi registrasi seminar berbasis web menggunakan qr code pada universitas xyz," Jurnal teknologi informasi dan komunikasi, 6(2), 128-150. 2019.

[8] Barra Purnama Pradja, Abdul Baist, Aan Subhan Pamungkas. "Pengembangan alat peraga matematika terintegrasi QR code sebagai sumber belajar bagi mahasiswa calon guru," Jurnal Penelitian Pembelajaran Matematika, 14(1), 1-11. 2021.

[9] Darul Mukminin, Ameliana, Adiiba Dwikanthi, Lukas Umbu Zogara. "Aplikasi kehadiran siswa berbasis android menggunakan qr code pada bimbel excellent institute," Jurnal teknoinfo, 16(1), 66-71. 2022.

[10] Renaldi P. Tumanggor, Elvira Asril, Guntoro. "Sistem informasi absensi karyawan menggunakan qr code (quick response) di pt. redes jaya perssada," Jurnal JITaCS, 1(1), 20-29. 2021.

[11] Muhammad Luthfi Syam, Erdisna. "Sistem informasi stok barang menggunakan qr-code berbasis android," Jurnal informatika ekonomi bisnis, 4(1), 17-22. 2022.

[12] Wahyudi Agustiono, Intan Rofika Putri, Devie Rosa Anamisa. "Desain layanan e-surat untuk desa waru barat, pamekasan, madura menggunakan qr code," Jurnal teknologi informasi dan ilmu komputer (JTIIK), 8(6), 1127-1136. 2021.

[13] Santa Frita Saragih, Elmor Benedict Wagiu. "Analisa perencanaan pembayaran menggunakan sistem qrcode di industrial universitas advent indonesia," Jurnal TeIKa, 9(1), 15-29. April, 2019.

[14] Muh. Ismail, Masnur, Al Ghazali Syam. "Aplikasi qr code sebagai sarana penyampaian informasi pohon dikebun raya jompie," Jurnal sintaks logika -1(1), 33-41. January, 2021.

[15] Dedek Arianti, Mahyuzar. "Efektivitas penggunaan aplikasi QR code (studi sosialisasi oleh harian serambi indonesia dan pemahaman masyarakat kecamatan kuta alam banda aceh)," Jurnal Ilmiah Mahasiswa FISIP Unsyiah, 2(2), 79-94. May, 2017.

[16] Irmayani Syafitri. "Pengertian apache beserta fungsi, kelebihan dan kekurangan apache yang perlu anda ketahui," (Online). Accessed: 19 March 2022 from https://www.nesabamedia.com/pengertian-apache/

[17] Naveen Kumar, "Mobile apps vs web apps - what's the difference?," (Online). Accessed: 25 July 2022 from https://www.joomdev.com/mobile-apps-vs-web-apps/.

Implementation of Backpropagation Method with MLPClassifier to Face Mask Detection Model

Wendy Hendra Wijaya[1] , Raymond Sunardi Oetama[2], Fransiscus Ati Halim[3]

Information System, Faculty of Information and Technology, Universitas Multimedia Nusantara, Tangerang, Indonesia

[1]wendy.wijaya4@student.umn.ac.id, [2]raymond@umn.ac.id, [3]fransiscus.ati@lecturer.umn.ac.id

Abstract— **The COVID-19 virus is highly contagious and can be spread through respiratory droplets when an infected person coughs or sneezes. To prevent the spread of the virus, the Minister of Health of the Republic of Indonesia has issued guidelines, including the use of personal protective equipment in the form of masks that cover the nose and mouth. This study aimed to develop a backpropagation neural network method from MLPClassifier and use the CRISP-DM framework to detect masks on human faces. The study used the RMFD (Real-World Masked Face Dataset) to train the model, which contains images of human faces with and without masks. The results of the study showed that the backpropagation neural network method had an accuracy of 94.4% in detecting masks on human faces. This accuracy outperformed other algorithms, such as the DNN algorithm from the paper with the title "SSDMNV2: A real time DNN-based face mask detection system using single shot multibox detector and MobileNetV2". This research is expected to contribute to the efforts to prevent the spread of COVID-19 by detecting the use of masks.**

Index Terms—**Backpropagation; COVID-19; CRISP-DM; Face Mask; Neural Network; MLPClassifier;**

I. INTRODUCTION

The COVID-19 pandemic has had a significant impact on global health and economies [1]. As of August 2021, over 218 million people have been infected and over 4.5 million have died. The virus interferes with the respiratory system and can lead to severe illness and death [2]. It has also caused widespread economic disruption due to measures such as lockdowns and travel restrictions. The Minister of Health of the Republic of Indonesia has issued guidelines to prevent the spread of COVID-19, including the use of masks [3]. However, data from covid19.go.id shows that the number of cases in Indonesia is still increasing, indicating low compliance with these guidelines in the community [4]. A survey also found low use of masks in the community. There is a need for increased awareness and compliance with guidelines to prevent the spread of COVID-19 in Indonesia [5].

One way to increase the level of compliance with guidelines to prevent the spread of COVID-19 is to improve supervision of the use of masks. Previous research has shown that machine learning models can be used for this purpose with different algorithms and methods. For example, the research "SSDMNV2: A real time DNN-based face mask detection system using single shot multibox detector and MobileNetV2" by Nagrath Preeti and colleagues has an accuracy of 93%. Data augmentation was performed in this research, without using HOG extraction and using the DNN (Deep Neural Network) algorithm [6]. The second research "Face Mask Detection by using Optimistic Convolutional Neural Network" by Suresh and colleagues has an accuracy of 98%. the Optimistic CNN (Convolutional Neural Network) algorithm was used without data augmentation and without using HOG extraction in this research [7].

The idea to use the backpropagation method in developing a machine learning model that can detect the use of masks on human faces is from a previous study named "Detection of Fire With Image Processing Using Backpropagation Method", which had an accuracy of 95% in detecting fire images. The dataset of that research was composed of images of fires that

were processed using HOG extraction and without data augmentation [8].

The model for this research will be based on the CRISP-DM framework and will use the MLPClassifier from Scikit learn library to use backpropagation neural network algorithm. The dataset used will be the RMFD (Real-World Masked Face Dataset) from GitHub, and image augmentation and HOG extraction will be applied to the data. The goal is to create a model with high accuracy in detecting the use of masks on human faces. This approach can help automate the supervision of the use of masks as personal protective equipment to prevent the spread of COVID-19.

II. LITERATURE REVIEW

A. Face Mask Detection

This study focuses on detecting the use of masks on human faces to help control the spread of COVID-19. COVID-19 itself is a virus that spreads widely and became a global pandemic starting in the city of Wuhan, China in December 2019 [1]. COVID-19 can be spread through droplets made from coughs or sneezes, which are inhaled by other humans. This virus can live on droplets for up to 8 hours, which makes it very easy for the virus to spread. Therefore, the use of masks is recommended in preventing the spread of COVID-19 [9]. This research uses RMFD data as the research object. The dataset contains image data with 2 classifications, namely faces that use masks and faces that do not use masks [10].

B. Histogram of Gradient (HOG)

A histogram of gradient is a technique used in image processing to detect objects by separating them from the background or noise. This separation helps improve the accuracy of the detection by focusing on the object of interest. The method uses the features and characteristics of the image, such as lines, contours, and shape, to measure the object. The measurement is divided into several areas, each of which is divided into pixels. Each pixel contains a value that is represented in the form of a histogram. This information is used to identify the object in the image and distinguish it from the background or other irrelevant objects [11].

$$f(x,y) = \begin{cases} fx(x,y) = f(x+1,y) - f(x-1,y) \\ fy(x,y) = f(x+1,y) - f(x-1,y) \end{cases}$$
(1)

$$Arg(x,y) = \sqrt{fx(x,y)^2 + fy(x,y)^2}$$
(2)

$$(x,y) = arctan\frac{fx(x,y)}{fy(x,y)}$$
(3)

The value of the brightness level in the image is based on equation 1, while equations 2 and 3 are calculations that are used for gradient-oriented gradient values [12].
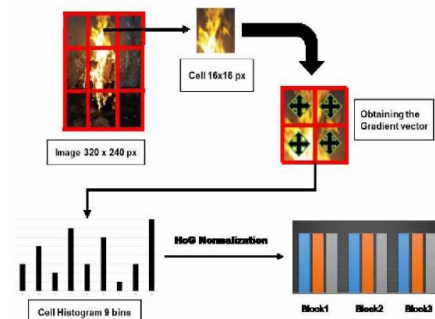

Figure 1. Example of HOG extraction [8]

C. Image Augmentation

Image Augmentation is a procedure used to add to existing image data by transforming the initial image data used. The result of image augmentation itself is new data that comes from the initial image data. This serves to increase the variety of image data used by not merely duplicating existing image data. The limitation of the data used is one of the problems that can be overcome using image augmentation so that more data can be studied [13]. Things that are done, for example, are rotating images based on degrees (90 degrees, 180 degrees, 270 degrees) and mirroring or flipping image data.

Figure 2. Example of augmentation on tumor images [14]

D. Backpropagation

The purpose of this method is to minimize the mean-squared error that occurs between the network output and the actual output for input to the network or the network itself [15]. This algorithm is very popular in learning artificial neural networks. This learning can also be referred to as error-correction learning, which has a forward (forward) and backward (backward) path [16], [17]. The workings of this method can be seen as shown in Figure 3 which has 2 paths, namely forward and backward.
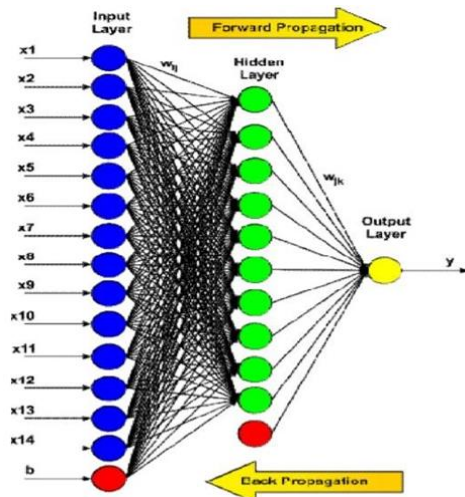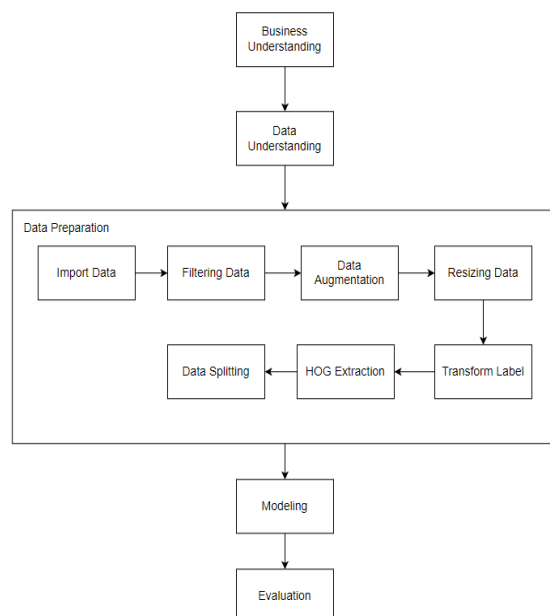
Figure 3. Backpropagation neural network [8]

Figure 4 Research flow from CRISP-DM framework

Figure 4 backpropagation neural network explanation [8]:

- X1 – X14 : Data input (input layer)

- b (red): Bias

- Wij : Weight to hidden node

- Wjk : Weight to output node

- Z1 – Z10 (Green): Hidden Layer

- Y : Output (output)

The first flow which is forward propagation is to update the weights for each hidden layer that is built. The hidden layer itself is given weight from the data that has been prepared, up to the output stage. In the second flow which is backward propagation (backpropagation) begins. backpropagation is the flow of returning the training pattern received by the artificial neural network to activate output propagation. From the output results, the artificial neural network uses the output as a target for the training pattern of the hidden layer.

III. METHODOLOGY

A. Object of Research

This research focuses on detecting the use of masks on human faces to help control of the spread COVID-19. COVID-19 itself is a virus that spreads widely and became a global pandemic starting in the city of Wuhan, China in December 2019 [1]. COVID-19 can be spread through droplets made from coughs or sneezes, which are inhaled by other humans. This virus can live on droplets for up to 8 hours, which makes it easy for the virus to spread. Therefore, the use of masks is recommended in preventing the spread of COVID-19 [9]. This research uses RMFD data as the research object. The dataset contains image data with 2 classifications, namely faces that use masks and faces that do not use masks [10].

B. Research Method

Figure 5 explains CRISP-DM or Cross Industry Standard Process for Data Mining used in this research. CRISP-DM is a standardization process for the life cycle of a data mining project [18].
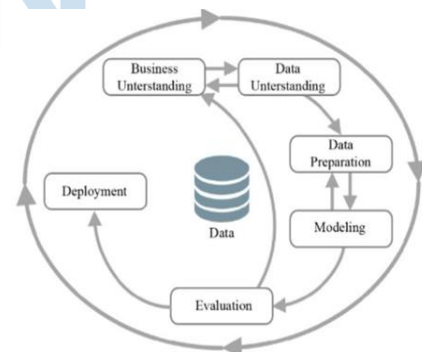

Figure 5. CRISP-DM [19]

The framework used in this research can be seen as the flow in Figure 4, which is a framework based on the CRISP-DM framework. This research only does the life cycle of CRISP-DM until the Evaluation

1) *Business Understanding*

Business Understanding is useful for identifying and determining the targets and objectives to be researched [19]. In this research, the main objective is to detect the use of masks on the human face, which has 2 classifications. The two classifications are human faces that use masks and human faces without a mask. The classification used the backpropagation method.

2) Data Understanding

In this second stage, there is a process of data collection, data understanding, and data analysis [19]. This study uses data called RMFD dataset [10]. The data comes from GitHub (https://github.com/X-zhangyang/Real-World-Masked-Face-Dataset) which contains photos of human faces using masks such as surgical masks, KN-95 masks, and others. And the data contains the image of human faces without masks. Figure 6 and Figure 7 are the example of the RMFD data.

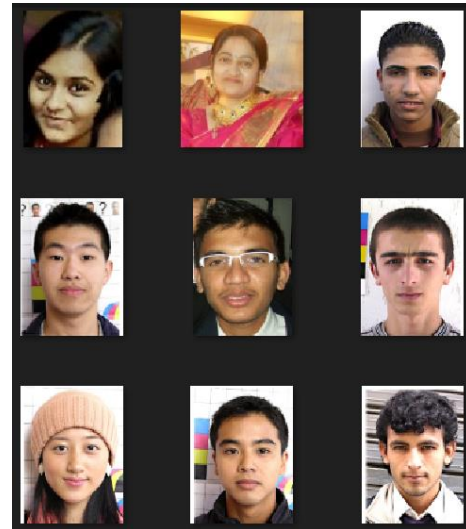Figure 6. RMFD human faces using a mask

Figure 7. RMFD human faces without a mask

3) Data Preparation

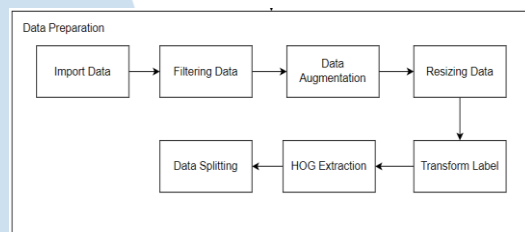The flow of data preparation in this research can be seen in Figure 8.

Figure 8. Data Preparation

The first process is importing the data to Google Colab. Filtering the image data that has been obtained is the second process. The dataset will be filtered based on their data type and features. The data type used in this research is "jpg" data type, and the second filter is a feature of the image itself. The only features used are images that have only "RGB" feature. When the data type is not "jpg", then the data will be changed to "jpg", also the data feature is not RGB, then the data will be deleted. The result of filtering process can be seen in Table I.

TABLE I. The result of filtering process

Data Classifi-cation	Data Type			Data Feature		
	jpg	*png*	*jpeg*	*RGB*	*RG BA*	*P*
With_ mask	1613	224	88	1798	117	15
Without_ mask	1930	0	0	1930	0	0

After filtering the image, the next process is data augmentation. This stage is useful for multiplying existing data up to 10000 image data (each label/ classification has 5000 image data). Augmentation is done by flipping the image from bottom to top, from left to right, and rotating the image by 90 degrees, 180 degrees, and 270 degrees at random. The augmented data are shown in Figure 9.

Figure 9. Augmented image

After augmenting the image, the image data will also be resized to a size of 320x240 pixels and converted the data into an array. This process is based on previous research that became the reference in this study, named "Detection of Fire With Image Processing Using Backpropagation Method" [8]. Label/ Classification from the dataset also transformed into binary class from multi-class, using LabelBinarizer library.

HOG extraction will be carried out on the image data that has been successfully converted into an array. HOG extraction is a descriptor of the elements in the image or photo by separating the elements that are important in the data (which is the main goal) and removing elements that are not important in the image under study [20]. Data from the previous process will be divided into 3 according to its function, data for the training model, test model, and validation model with a portion of 10% for the validation model, 10% for the test model, and 80% for the training model.

4) Modeling

The modeling process will use Google Colab tools that use Python language. Initially, the existing data, named photo data of humans using masks and photo data of humans not wearing masks will be imported into Google Colab. The data will then be labeled into 2 classifications according to the classification of each photo, named photos of masks and photos without masks. The data will also be split into 3, train data which is useful for training the model, validation data is for validating the model, and test data which is useful for evaluating the model. The results of the model itself are in the form of accuracy.

Modeling with the backpropagation neural network method will use the MLPClassifier or Multi-layer

Perceptron Classifier derived from the Scikit Learn library. MLPClassifier itself is a supervised learning that has one hidden layer or even more with stacked links or neurons originating from the received input. The model created using the MLPClassifier will be optimized again by using the log-loss function obtained in the first model creation [21].

5) Evaluation

In this fifth stage, the results of the model will be assessed and evaluated. This is useful as a benchmark, whether the model has been made in accordance with the initial objectives of this research with the accuracy as the output. The evaluation of this research itself is the level of mask detection accuracy on the human face. The purpose of this research is whether this model can detect faces wearing a mask, and those who don't wear a mask.

6) Deployment

This research did not reach the deployment stage. This is because this study only focuses on comparing the level of accuracy of previous studies.
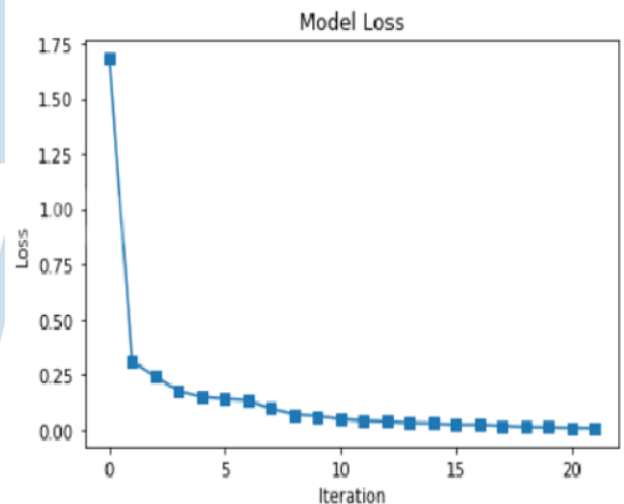
IV.   RESULT AND DISCUSSION
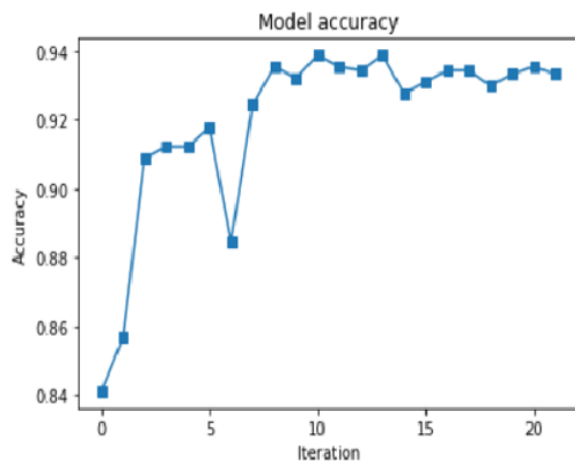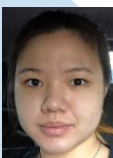

Figure 10. Loss from the model
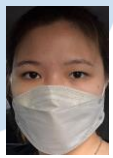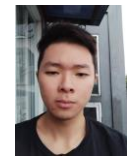
Figure 11. Accuracy from the model

The results of the model can be seen in Figure 10 and Figure 11. There were 22 iterations with a loss of 0.008 and a validation score of 0.933. As for the accuracy score obtained in the testing data, it is 0.944 with a confusion matrix as shown in Figure 13. The model that has been built produces a confusion matrix like Figure 12, with 465 true positives, 479 true negatives, 35 false positives, and 21 false negatives.

Figure 12. Confusion matrix from the model

Photo	Actual	Prediction	Photo	Actual	Prediction
	With_mask	With_mask		Without_mask	Without_mask
	Without_mask	Without_mask		With_mask	With_mask
	Without_mask	Without_mask		Without_mask	Without_mask
	With_mask	With_mask		With_mask	Without_mask
	With_mask	With_mask		Without_mask	Without_mask
	Without_mask	Without_mask		With_mask	With_mask

	Without_mask	Without_mask		Without_mask	Without_mask
	With_mask	With_mask		With_mask	With_mask
	With_mask	With_mask		Without_mask	Without_mask
	Without_mask	Without_mask		With_mask	Without_mask

Test to the model also has been done with photos owned by the researcher. There are 20 photos taken by camera handphone as shown in Figure 13. The 20 photos contain photos of human faces that wear a mask and do not wear a mask, in pairs and the results are shown in Table II

Figure 13. Sample of photos

Testing the data that has been tested on the 20 image data managed to guess the image correctly as many as 18 data. There is a predicting error in the 13th and 20th data. The image data used has taken the same pre-

processing stage as the modeling stage (such as resizing data, converting data into arrays, and performing HOG extraction). From the predictions that have been made, it can be concluded that the predictions for the data from the model testing results get an accuracy of 90%.

V. CONCLUSION

The application of the backpropagation neural network method using the CRISP-DM framework, as well as modeling is carried out through several pre-processing stages such as filtering data based on the data type and features of the image data itself, multiplying the data using an augmentor library in augmenting research image data, and processing data using HOG extraction before building a neural network model with the backpropagation method has been done with an accuracy of 94.4%.

Based on the results, the backpropagation method can outperform the DNN algorithm by 1.4% in detecting masks on human faces. Even though the Optimistic CNN algorithm is still superior to the backpropagation method, with a difference of 3.6%. From the results that have been achieved, it is can be concluded that this research has succeeded in achieving its goal, which is detecting face masks on the human face by obtaining an accuracy of 94.4% using the backpropagation method. This research also succeeded in manually validating to detect face mask on the human face and this model was able to predict 18 out of

20 photos provided by the researcher, which means getting 90% accuracy from the test.

ACKNOWLEDGMENT

REFERENCES

[1] R. Sitharthan, D. Shanmuga Sundar, M. Madurakavi, Karthikeyan, I. Jacob Raglend, J. Belwin Edward, R. Raja Singh, R. Kumar., "Assessing nitrogen dioxide (NO2) impact on health pre- and post-COVID-19 pandemic using IoT in India," *Int. J. Pervasive Comput. Commun.*, 2020, doi: 10.1108/IJPCC-08-2020-0115/FULL/HTML.

[2] A. Rendeiro, C. Vorkas, J. Krumsiek, Singh, K. Harjot, C. Shashi, V. Luca, M. T. Cacciapuoti, I. Giorgio, O. Elemento, M. Salvatore., "Metabolic and immune markers for precise monitoring of COVID-19 severity and treatment," *medRxiv*, Sep. 2021, doi: 10.1101/2021.09.05.21263141.

[3] "Keputusan Menteri Kesehatan Republik Indonesia, Nomor HK.01.07/MENKES/382/2020," 2020. http://hukor.kemkes.go.id/uploads/produk_hukum/KMK_No__HK_01_07-MENKES-382-2020_ttg_Protokol_Kesehatan_Bagi_Masyarakat_di_Tempat_dan_Fasilitas_Umum_Dalam_Rangka_Pencegahan_COVID-19.pdf (accessed Oct. 8, 2021).

[4] Satuan Tugas Penanganan Covid-19, "Analisis Data COVID-19 Indonesia (Update Per 14 Maret 2021) - Berita Terkini | Covid19.go.id," 2021. https://covid19.go.id/p/berita/analisis-data-covid-19-indonesia-update-13-juni-2021 (accessed Oct. 06, 2021).

[5] R. Emkaen, "Badan pusat statistik kabupaten enrekang," 2020. https://bandungkab.bps.go.id/publication/2021/09/22/a45e7ec69a9b9e48b91e3945/perilaku-masyarakat-kabupaten-bandung-pada-masa-ppkm-darurat.html (accessed Oct. 06, 2021).

[6] P. Nagrath, R. Jain, A. Madan, R. Arora, P. Kataria, and J. Hemanth, "SSDMNV2: A real time DNN-based face mask detection system using single shot multibox detector and MobileNetV2," *Sustain. Cities Soc.*, vol. 66, p. 102692, Mar. 2021, doi: 10.1016/J.SCS.2020.102692.

[7] K. Suresh, M. B. Palangappa, and S. Bhuvan, "Face Mask Detection by using Optimistic Convolutional Neural Network," *Proc. 6th Int. Conf. Inven. Comput. Technol. ICICT 2021*, pp. 1084–1089, Jan. 2021, doi: 10.1109/ICICT50816.2021.9358653.

[8] M. Iqbal, B. Irawan, and C. Setianingsih, "Detection of Fire with Image Processing using Backpropagation Method," *2019 Int. Conf. Adv. Mechatronics, Intell. Manuf. Ind. Autom. ICAMIMIA 2019 - Proceeding*, pp. 344–349, 2019, doi: 10.1109/ICAMIMIA47173.2019.9223392.

[9] R. G. Babu, A. Balaji, K. K. Kumar, S. Maurya, and S. K. Saravana, "Smartphone-based electrochemical sensor for assessing COVID-19 infected patients," *Int. J. Pervasive Comput. Commun.*, 2020, doi: 10.1108/IJPCC-10-2020-0169/FULL/HTML.

[10] X. Zhang, "GitHub - X-zhangyang/Real-World-Masked-Face-Dataset: Real-World Masked Face Dataset, 口罩人脸数据集." https://github.com/X-zhangyang/Real-World-Masked-Face-Dataset (accessed Jun. 02, 2022).

[11] X. Zhang, H. Gao, C. Xue, J. Zhao, and Y. Liu, "Real-time vehicle detection and tracking using improved histogram of gradient features and Kalman filters:," *https://doi.org/10.1177/1729881417749949*, vol. 15, no. 1, Jan. 2018, doi: 10.1177/1729881417749949.

[12] A. Chitlangia and G. Malathi, "Handwriting Analysis based on Histogram of Oriented Gradient for Predicting Personality traits using SVM," *Procedia Comput. Sci.*, vol. 165, pp. 384–390, Jan. 2019, doi: 10.1016/J.PROCS.2020.01.034.

[13] M. D. Bloice, P. M. Roth, and A. Holzinger, "Biomedical image augmentation using Augmentor," *Bioinformatics*, vol. 35, no. 21, pp. 4522–4524, Nov. 2019, doi: 10.1093/BIOINFORMATICS/BTZ259.

[14] C. Han, L. Rundo, R. Araki, N. Yudai, F. Yujiro, M. Giancarlo, N. Hideki, H. Hideaki., "Combining noise-to-image and image-to-image GANs: Brain MR image augmentation for tumor detection," *IEEE Access*, vol. 7, pp. 156966–156977, 2019, doi: 10.1109/ACCESS.2019.2947606.

[15] R. Rosalina and H. Jayanto, "Maximal Overlap Discrete Wavelet Transform, Graph Theory And Backpropagation Neural Network In Stock Market Forecasting," *IJNMT (International J. New Media Technol.*, vol. 5, no. 1, pp. 41–46, Jun. 2018, doi: 10.31937/IJNMT.V5I1.679.

[16] N. Hadianto, H. B. Novitasari, and A. Rahmawati, "Klasifikasi Peminjaman Nasabah Bank menggunakan Metode Neural Network," *Pilar Nusa Mandiri J. Comput. Inf. Syst.*, vol. 15, no. 2, pp. 163–170, Sep. 2019, doi: 10.33480/PILAR.V15I2.658.

[17] S. Amirsadri, Seyed, J. Mousavirad, and K. E. Komleh, "A Levy flight-based grey wolf optimizer combined with back-propagation algorithm for neural network training," *Neural Comput. Appl.*, doi: 10.1007/s00521-017-2952-5.

[18] C. P. Kuncoro, "Analysis Of UMN Student Graduation Timeliness Using Supervised Learning Method," *IJNMT (International J. New Media Technol.*, vol. 8, no. 2, pp. 89–95, 2021, doi: 10.31937/IJNMT.V8I2.2366.

[19] A. Shi-Nash and D. R. Hardoon, "Data analytics and predictive analytics in the era of big data," *Internet Things Data Anal. Handb.*, pp. 329–345, Feb. 2017, doi: 10.1002/9781119173601.CH19.

[20] D. Prasanna and M. Prabhakar, "An effiecient human tracking system using Haar-like and hog feature extraction," *Clust. Comput. 2018 222*, vol. 22, no. 2, pp. 2993–3000, Jan. 2018, doi: 10.1007/S10586-018-1747-5.

[21] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, "Scikit-learn: Machine Learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011

Implementation IT Governance Using COBIT 5 Framework at PT. XYZ (Persero)

Drago Tinus[1], Johan Setiawan[2]

[1,2] Information System, Faculty of Technology and Information,
Universitas Multimedia Nusantara, Tangerang, Indonesia
[1]drago.tinus@student.umn.ac.id, [2]johan@umn.ac.id

Abstract— **PT. XYZ (Persero) is one of the state-owned companies that move in insurance & guarantee, capital market & investment. PT. XYZ (Persero) conducts an assessment of information technology governance with the aim of meeting the implementation of GCG (Good Corporate Governance) which are listed through KBUMN (Ministry of State-Owned Enterprises) regulations PER-02/MBU/2013 & PER-03/MBU/02 /2018. The regulation PER-02/MBU/2013 has been stipulated on page 6 section 3.2. The target is that "The maturity level target in the next 5 years is a minimum maturity level of 3 in accordance with the specified maturity level". PT. XYZ will conduct a COBIT 5 assessment again based on the results of the assessment that has been carried out by the previous assessor. COBIT 5 assessment will be carried out with a target of level 4 so that it is expected to get a maturity value above 3 and can meet the request of KBUMN (Ministry of State-Owned Enterprises). The assessment is done by identifying company objectives, Enterprise Goals, mapping IT Related Goals, mapping and determining process domains, analyzing the previous assessor's COBIT 5 capability level measurement, measuring the process capability level of EDM01, EDM02 and EDM04, results and gap analysis as well as improvement recommendations. Based on the assessment that has been carried out on the EDM01, EDM02 & EDM04 process, the final result is obtained with a final value of 97.08% so that it gets a Capability Level of 3,88 and it has a gap of 0,12 from the expected target.**

Keywords— Capability Level, COBIT 5, EDM, IT Governance, PT. XYZ (Persero).

I. INTRODUCTION

In the era of globalization, information technology has developed rapidly. The use and utilization of information technology cannot be ignored in today's human life. In general, the use of information technology is not only used individually, but is also used for organizational or company purposes in increasing efficiency, effectiveness and assisting existing work processes. However, the utilization and use of information technology itself must have a large amount of investment accompanied by a high risk. Risk itself is a negative impact that can hinder the achievement of desired goal. Therefore, the implementation of the use and utilization of information technology needs to be controlled or supervised thoroughly to achieve the desired thing. Therefore, proper and good management of information technology is needed in order to feel the use and utilization of information technology that is efficient, effective and can help the work process for the company or organization [1].

Management of information technology is an important part in a company or organization, where the management of information technology is one part of the strategic resources needed in the development process. Management of information technology that is controlled and carried out properly and correctly is a key to success as a means to manage information technology work carried out with a systems approach.

PT. XYZ (Persero) is one of BUMN (State Owned Enterprises) which is engaged in insurance & guarantee as well as capital market & investment. PT. XYZ has the task of processing, collecting and presenting data services from each subsidiary as well as providing information technology systems and infrastructure. Based on the importance of the task of PT. XYZ until now there is no parameter or measurement scale that measures the maturity level of the use and utilization of information technology owned.

The role of IT Governance is very much needed to ensure the use and utilization of information technology that is owned can assist in meeting all

company needs. This also needs to be carried out with the aim of using and utilizing information technology at PT. XYZ can generate business value that can generate profits or facilitate business processes for the company. In addition, it is necessary to ensure that the use and utilization of information technology resources used can run in the long term or sustainale. Even so, cooperation is needed between the information technology division and other divisions such as human resources, risk management and legal in monitoring the use and utilization of information technology activities at PT. XYZ.

The development use, and utilization of information technology by institutions or government agencies has been going on for quite a long time, with a increasing level of seriousness [2]. To ensure the use and utilization of information technology supports the fulfillment of company needs, it is necessary to pay attention to the level of efficiency in allocating and using resources, managing risks and good governance related to information technology [2]. In addition, in order to obtain maximum results in the use and utilization of information technology, companies need to carry out information technology governance.

In implementing or designing information technology governance, companies need to look at all conditions and factors for the use and utilization of current and future information technology. In addition, it is also necessary to adjust the information technology strategy with the business strategy that determines the level of information technology needs to support current and future business needs. Therefore, to ensure the use and utilization of information technology at PT. XYZ which is already running and can generate business value and ensure that risk management, legal and human resources are capable of fulfilling all company needs, it is necessary to carry out an information technology governance assessment process.

Indirectly, the IT governance process is a structured activity to assess or evaluate all information technology components and activities that are planned, owned and executed. It also helps in ensuring that the use and utilization of IT had met and can support the needs of the company.

As a result, an audit of information technology governance is necessary. This is intended by conducting an assessment to determine the maturity level of the use and utilization of information technology that is owned and currently running. In this study, PT. XYZ will implement IT Governance using the COBIT 5 framework. The COBIT framework itself has the function to measure the maturity level of IT utilization in an organization [3]. COBIT is a framework that is used to determine the level / value of

the maturity of IT process control. The COBIT 5 framework divides information technology activities into 2 areas, namely governance (Evaluate, Direct, Monitor) and management (Plan, Build, Run, Monitor). In addition, the existing information technology processes are divided into 5 process domains, including EDM (Evaluate, Direct, Monitor), APO (Align, Plan, Organise), BAI (Build, Acquire, Implement), DSS (Deliver, Service, Support) and MEA (Monitor, Evaluate, Assess). The function of these five domains aims to determine the alignment of values between different stakeholders, business goals and the value of information technology that is owned and will be used [4].

The assessment of IT governance at PT.XYZ is carried out with the hope of meeting and implementing good IT governance for state-owned companies which is stated in the KBUMN regulation PER-02/MBU/2013 concerning "Guidelines for the Preparation of Information Technology Management for State-Owned Enterprises" & PER-03/MBU/02/2018 concerning "Amendment to the Regulation of the Minister of State-Owned Enterprises Number PER-02/MBU/2013 concerning Guidelines for the Preparation of Information Technology Management for State-Owned Enterprises". The regulation PER-02/MBU/2013 has been stipulated on page 6 section 3.2. The target is that "The maturity level target of SOE IT Governance in the next 5 years is a minimum maturity level of 3 in accordance with the specified maturity level.

PT. XYZ has previously conducted a COBIT 5 assessment of twenty-one domain processes which were carried out together with consultants which ended on October 5, 2021 with a score of 3.43 as an assessor. Furthermore, PT. XYZ will conduct a COBIT 5 assessment again based on the results of the assessment that has been carried out by the previous assessor. COBIT 5 assessment will be carried out with a target of level 4 so that it is expected to get a maturity value above 3 and can meet the request of KBUMN (Ministry of State-Owned Enterprises) where the minimum maturity level obtained is level 3.

II. LITERATURE REVIEW

A. IT Governance

IT governance is technology-related proces that are carried out to regulate and manage the development and utilization of IT in an organization. In carrying out the use and utilization of IT in an organization, its performance needs to be assessed so that the IT management mechanism can run in accordance with the business objectives of the agency or government organization [5].

Information technology governance also needs to be implemented to direct IT activities to fulfill the mission and vision of government agencies or organizations. IT management assessment can be done with IT governance assessment. IT governance is a theory formed as a result of the development and use of IT in companies or organizations that function to facilitate the achievement of goals [6]. The application of IT governance is needed in the delivery of good and quality services if they can be managed in the right and structured way so as to increase user satisfaction. [6]

System audit or information technology governance is a means to test or examine information systems to determine whether an organization's information system operates in accordance with its mission, and objectives or carry out performance tests, uncovering potential risks and impacts that may occur within an organization [7]. An audit of information technology systems or governance is a way to collect and evaluate evidence to prove that the existing system can maintain data integrity effectively and efficiently.

B. Framework COBIT5

COBIT (Control Objective for Information and related Technology) is a framework that leads to service functions, management, audit functions, IT controls and aims to establish the availability, integrity and confidentiality of data [8]. Based on this, it can be concluded that COBIT 5 is a framework used by companies or organizations to carry out information technology control processes to ensure data availability, integrity and confidentiality as well as the execution of IT utilization and use.

There are five principles from the COBIT 5 framework, which are [9]:
1. Meeting Stakeholder Needs
   Aims to provide and fulfill the needs or all processes desired by stakeholders into structured, detailed, clear and practical goals with the utilization and use of IT.
2. Covering the Enterprise End-to-end
   Manage the governance and management of information technology as a whole.
3. Applying a Single Integrated Framework
   Designed as a framework that has standards covering various IT governance frameworks.
4. Enabling a Holistic Approach
   Has several enablers that can help implement IT governance and management that is easy to understand and understand to achieve company or organizational goals.
5. Separating Governance From Management
   Separate governance from management, because both have different activities, functions and organizational structures.

C. COBIT5 Process Reference Model

COBIT 5 divides IT governance and management processes into 2 main processes, which are [8]:
1. Governance has five processes that are determined through the EDM process practices.
2. Management, including Plan, Build, Run and Monitor. This domain is the level of the domain and process structure contained in COBIT 4.1, namely APO, BAI, DSS and MEA.

After that, there are three stages in determining the domain process that will be assessed using the COBIT 5 framework, which are [11]:
1. Adjusting or aligning COBIT 5 strategic objectives with the organization's strategic objectives.
2. Comply with applicable regulations, both from internal regulations and external regulations.
3. Melakukan pendekatan risiko dalam menentukan prioritas *domain process*.

D. COBIT5 Implementation Lifecycle

There are seven phase of implementing information technology governance using the COBIT 5 Framework [12], which are:
1. Initiate Programme – What are the drivers?
   Analyze the cause of the problem and decide if there is a need for implementation or improvement initiatives.
2. Define Problems and Opportunities
   Focusing on defining the scope of implementation or initiatives in improving the organization's or company's goals to IT-related goals into IT processes and considering the risks that may occur.
3. Define Road Map
   Set targets for improvement or development to be implemented as well as gap analysis in order to provide recommendations for improvement of the gaps that occur.
4. Plan Programme
   Design the best practical solution based on the business case/project being executed. A well-developed project can assist in project monitoring and identification.
5. Execute Plan
   Providing solutions based on the design of solutions that have been formed that can be carried out in daily activities and ensuring that business goals are aligned, can be maintained, achieved and can be measured. To achieve success, commitment from top management, stakeholders and influential business owners is required.

6. Realise Benefits

   Focusing on sustainable operations by new enablers or to be improved and monitoring the achievement of the expected benefits.

7. Review Effectiveness

   Analyze IT governance and management requirements with the aim of managing continuous improvement.

E. COBIT5 Process Capability Attribute

Based on the identified IT processes, COBIT 5 can perform measurements with six levels of maturity (Capability Level), which are [13]:

1. Level 0 (Incomplete Process)

   Not implementing the IT process as intended and not meeting the objectives of the IT process.

2. Level 1 (Performed Process)

   Have implemented IT processes but have not achieved the goals of IT processes.

3. Level 2 (Managed Process)

   Have implemented IT processes and achieved the goals to be achieved by means of good management.

4. Level 3 (Established Process)

   Already have and implement IT processes that have been standardized into the overall scope of the organization. As well as having standardized processes that apply as a whole.

5. Level 4 (Predictable Process)

   Has implemented IT processes into a clear boundary. This limit is obtained based on previous assessments when implementing standardized IT processes.

6. Level 5 (Optimizing Process)

   Have tried new things and made continuous improvements to achieve maximum capabilities.

There is a scale for assessing the level of maturity (Capability Level) that a process has achieved its objectives, as shown in Table 1 below [11]:

Table 1. Maturity Rating Scale COBIT 5

Code	Description	Scale
F	Fully Achieved	More than 85% up to 100%
L	Largely Achieved	More than 50% up to 85%
P	Partially Achieved	More than 15% up to 50%
N	Not Achieved	0% to 15%

F. COBIT 5 RACI *Chart*

The RACI diagram or RACI Chart is part of the Assignment of Responsibility Matrix (RAM), which is a mapping between resources and activities in each procedure [9]. RACI itself stands for [10]:

- Responsibility

  Determine who will complete the task. It refers to an important operational part in fulfilling activities and creating the desired results.

- Accountable

  Who determines success or success in this task. Which is given overall responsibility to complete this task according to the level of accountability.

- Consulted

  Who provides feedback or input which is a key role for all inputs received. In addition to providing input, it is also necessary to obtain information from external partners or other outside parties for comparison or consideration.

- Informed

  Who receives the information. Is the party who receives and is notified of the results of the achievements or the results of the tasks carried out.

III. METHODOLOGY

A. Identify Company Goals

Seek and find out information about the company's goals and objectives in the use and utilization of information technology. The information obtained is related to the goals and objectives of the use and utilization of information technology at PT. XYZ will be used in determining Enterprise Goals, IT Related Goals and determining the process domain.

B. Enterprise Goals Mapping

Figure 1. Enterprise Goals Mapping

Mapping the goals and objectives of the use and utilization of information technology at PT. XYZ with Enterprise Goals framework COBIT 5. Mapping Enterprise Goals based on the goals and objectives of the use and utilization of information technology at PT.

XYZ. Figure 1. are the 17 Enterprise Goals that have been determined by the COBIT 5 framework.

C. IT Related Goals Mapping

The previously determined Enterprise Goals are used as material for mapping IT Related Goals.

Figure 2. IT Related Goals Mapping

Mapping of IT Related Goals is carried out in accordance with the previously determined COBIT 5 Enterprise Goals.

D. Determine the process to be assessed

Determination of the domain process that will be used in accordance with the information technology activities carried out by the company. The process domain mapping is carried out in accordance with the predetermined COBIT 5 IT Related Goals.

Figure 3. COBIT 5 Process Mapping

Figure 4. COBIT 5 Process Mapping

Mapping process domains according to the previously determined COBIT 5 IT Related Goals.

E. Filling out Questionnaires and doing Interview

Conduct interviews or seek information related to the established COBIT 5 process domain. Domain process that has been determined will be used to evaluate information technology activities carried out by the company.

Interviews were conducted with the company regarding the information technology activities carried out. Interviews were conducted regarding the activities carried out and based on the COBIT 5 Process

Reference Model. The end result is information related to the company's information technology activities.

F. Assessment

Carry out COBIT 5 assessment based on the results of interviews related to information technology activities that have been carried out information related to the company's information technology activities becomes the basis for carrying out the assessment. This is done by analyzing previously obtained information related to information technology activities with the activities listed in COBIT 5 and checking the completeness of the required documents. The end result is to provide an assessment score for each activity carried out by the company.

G. Leveling/Ranking

Determine the level to determine the maturity level of the information technology activities carried out. The assessment score for each information technology activity carried out by the company becomes the basis for determining the level of the company in implementing information technology activities contained in COBIT 5.
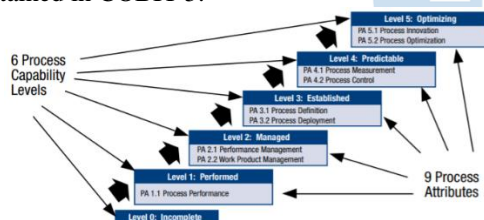
Figure 5. Process Rating on COBIT5

H. Giving Improvement Recommendation

Provide recommendations for improvement based on all findings found in information technology activities carried out by the company as a solution for future improvements.

All information related to activities, scoring of each process domain and the final level of maturity level of information technology activities carried out by the company become a reference in providing recommendations for improvement.

IV. RESULT AND DISCUSSION

A. Identify Company Goals

Below are the results of the identification of company goals obtained from Mr. Wahyu Eko as the Coordinator of the IT Planning and Governance Department in the IT division of PT. XYZ:

1. Manage and regulate the strategy and activities of information technology in the company.
2. Ensuring the function of information technology in meeting the daily operational activities of the company.
3. Manage and manage risk management related to information technology in the company.
4. Prepare and manage the Company's Work Plan and Budget.
5. Establish and review SOPs related to information technology in the company.
6. Fulfill KPI (Key Performance Indicator) as a fulfillment of IT audit and follow up on audit results.

B. Mapping Enterprise Goals

After the identification of the company's goals is obtained, then proceed with selecting the COBIT 5 Enterprise Goals. The results of the Enterprise Goals mapping carried out in Figures 6 are as follows [10]:

Enterprise Goals COBIT 5

BSC Dimension		Enterprise Goals	Peringkat
Financial	1	Nilai pemegang saham dari investasi bisnis	3
	2	Portofolio layanan dan prooduk yang kompetitif	8
	3	Mengendalikan risiko bisnis	16
	4	Kepatuhan akan regulasi hukum maupun peraturan eksternal	2
	5	Transparansi keuangan	5
Customer	6	Budaya layanan yang berorientasi terhadap pelanggan	6
	7	Ketersediaan dan keberlanjutan layanan bisnis	1
	8	Cepat tanggap terhadap lingkungan bisnis yang mengalami perubahan	17
	9	Penggunaan informasi sebagai penentuan pengambilan keputusan strategis	14
	10	Memaksimalkan biaya pemberian layanan	10
Internal	11	Memaksimalkan proses bisnis secara fungsional	4
	12	Memaksimalkan biaya proses bisnis	12
	13	Mengendalikan proses bisnis yang mengalami perubahan	13
	14	Operasional dan produktivitas staf	9
	15	Kepatuhan dan produktivitas staf	11
Learning and Growth	16	SDM yang memiliki motivasi dan terampil	15
	17	Budaya inovasi bisnis dan produk	7

Figure 6. Enterprise Goals Mapping

C. IT Related Goals Mapping

The results of the previous Enterprise Goals mapping will be used as the basis for determining IT Related Goals. IT Related Goals will be determined based on the primary key (P) in Figure 7 as follows [10]:

Figure 7. Mapping IT Related Goals

The following is the result of mapping IT Related Goals in Figure 8, 9 and 10 below based on the Enterprises Goals mapping that has been done previously:

BSC Dimension	Enterprise Goals	Peringkat	IT Related Goals
Customer	7 Ketersediaan dan keberlanjutan layanan bisnis	1	4 Mengendalikan risiko bisnis terkait TI
			10 Keamanan informasi, aplikasi dan infrastruktur
			14 Ketersediaan informasi yang terjamin serta dapat membantu dalam pengambilan keputusan
Customer	4 Kepatuhan akan regulasi hukum maupun peraturan eksternal	2	2 Kepatuhan dan dukungan TI terhadap kepatuhan bisnis terhadap peraturan eksternal serta hukum
			10 Keamanan informasi, aplikasi dan infrastruktur
Financial	1 Nilai pemegang saham dari investasi bisnis	3	1 Sinkronisasi TI dengan strategi bisnis
			3 Komitmen manajemen eksekutif dalam membentuk keputusan terkait TI
			5 Perwujudan manfaat portofolio layanan dan investasi yang mendukung TI
			7 Menyediakan layanan TI sesuai kebutuhan bisnis
			11 Memaksimalkan sumber daya, aset dan kapabilitas TI
			13 Penyampaian program yang sesuai anggaran, tepat waktu, memberikan manfaat, memenuhi standar kualitas dan persyaratan

Figure 8. Mapping IT Related Goals

Internal	11 Memaksimalkan proses bisnis secara fungsional	4	1 Sinkronisasi TI dengan strategi bisnis
			7 Pengiriman layanan TI sesuai dengan kebutuhan bisnis
			8 Penerapan informasi dan teknologi serta solusi aplikasi yang layak
			9 Kecekatan TI
			12 Pemberdayaan serta mendukung proses bisnis dengan memadukan teknologi maupun aplikasi ke dalam proses bisnis

Figure 9. Mapping IT Related Goals

Financial	5 Transparansi keuangan	5	6 Transparansi biaya, risiko serta manfaat TI

Figure 10. Mapping IT Related Goals

D. Mapping and Determination of Domain Process

Based on the results of the mapping of IT Related Goals that have been carried out, the results of the mapping will be used as the basis for determining the Domain Process. The Domain Process will be

determined based on the primary key (P) in Figure 11 and 12 as follows [10]:

Figure 11. Mapping IT Related Goals to COBIT 5 Process

Figure 12. Mapping IT Related Goals to COBIT 5 Process

Based on Figure 11 and figure 12 above, it can be determined which processes can be assessed. The following is the result of the process mapping carried out in Figure 13 and 14 below this:

No	IT Related Goals	EDM	APO	BAI	DSS	MEA
1.	IRG 1 Sinkronisasi TI dengan strategi bisnis	01 02	01 02 03 05 07 08	01 02	-	-
2.	IRG 2 Kepatuhan dan dukungan TI terhadap kepatuhan bisnis terhadap peraturan eksternal serta hukum	-	01 12 13	10	05	02 03
3.	IRG 3 Komitmen manajemen eksekutif dalam membentuk keputusan terkait TI	01 05	-	-	-	-
4.	IRG 4 Mengendalikan risiko bisnis terkait TI	03	10 12 13	01 06	01 02 03 04 05 06	01 02 03
5.	IRG 5 Perwujudan manfaat portofolio layanan dan investasi yang mendukung TI	02	04 05 06 11	01	-	-
6.	IRG 6 Transparansi biaya, risiko serta manfaat TI	02 03 05	06 12 13	09	-	-
7.	IRG 7 Menyediakan layanan TI sesuai kebutuhan bisnis	01 02 05	02 08 09 10 11	02 03 04 06	01 02 03 04 06	01
8.	IRG 8 Penerapan informasi dan teknologi serta solusi aplikasi yang layak	-	04	05 07	-	-

Figure 13. Mapping IT Related Goals to COBIT 5 Process

No	IT Related Goals	EDM	APO	BAI	DSS	MEA
9.	IRG 9 Kecekatan TI	04	01 03 04 10	08	-	-
10.	IRG 10 Keamanan informasi, aplikasi dan infrastruktur	03	12 13	06	05	-
11.	IRG 11 Memaksimalkan sumber daya, aset dan kapabilitas TI	04	01 03 04 07	04 09 10	01 03	01
12.	IRG 12 Pemberdayaan serta mendukung proses bisnis dengan memadukan teknologi maupun aplikasi ke dalam proses bisnis	-	08	02 07		
13.	IRG 13 Penyampaian program yang sesuai anggaran, tepat waktu, memberikan manfaat, memenuhi standar kualitas dan persyaratan	-	05 07 11 12	01 05	-	-
14.	IRG 14 Ketersediaan informasi yang terjamin serta dapat membantu dalam pengambilan keputusan	-	09 13	04 10	03 04	-

Figure 14. Mapping IT Related Goals to COBIT 5 Process

After mapping the COBIT 5 process which was then handed over to Mr. Wahyu Eko as the IT Governance Coordinator in the IT division of PT. XYZ, Mr. Wahyu Eko submitted a suggestion to conduct an IT Maturity Level assessment using the COBIT 5 framework in the governance process (Governance). Taking this into account, then the students and the company through Mr. Wahyu Eko as the IT Governance Coordinator of the IT division of PT. XYZ made an agreement to assess 3 Evaluate, Direct and Monitor (EDM) domain processes to be assessed, namely EDM01, EDM02 and EDM04 with the assessment target reaching level 4 (Predictable Process) where previously the assessment by third parties only reached level 3 (Established). Process).

E. COBIT 5 Process Capability Measurement Results by Third Parties

The final result of measuring the level of COBIT 5 process capability that has been carried out by a third party previously with the target of achieving level 3. The following is a summary of the IT Capability Level PT. XYZ (Persero) by third parties, as shown in Figure 15:

No	Nama Proses	Level Penilaian
Evaluate, Direct and Monitor		
1	EDM01: Ensure Governance Framework Setting and Maintenance	3
2	EDM02: Ensure Benefits Delivery	3
3	EDM03: Ensure Risk Optimisation	4
4	EDM04: Ensure Resource Optimisation	3
Keterangan		
Hijau		Nilai *Assessor* Lebih Ting
Biru		Nilai *Assessor* Lebih Rendah
Putih		Nilai *Assessor* sama denga self-assessment

Figure 15. Summary Value of IT Capability Level

The summary results produce findings that will be part of the fulfillment of the next level 4 target that will be implemented, as shown in Figure 16:

Tingkat (Level)	Atribut	Kriteria	Target Pemenuhan
4 (*Predictable Process*)	PA 4.1	a) Persyaratan pemrosesan informasi untuk mendukung tujuan bisnis yang ditetapkan, terkait dan telah ditetapkan.	Dibutuhkan pemetaan antara *process goals* dengan *business goals*.
		b) Tujuan pengukuran proses diturunkan dari kebutuhan informasi proses.	Dibutuhkan *cascading* penurunan matrik KPI dari *process goals*.
		c) Tujuan kinerja proses kuantitatif yang mendukung tujuan bisnis yang relevan telah ditentukan.	Dibutuhkan Obyektif dan matrik KPI yang bersifat kuantitatif (bukan lagi kualitatif).
		d) Ukuran dan frekuensi pengukuran telah teridentifikasi dan ditentukan sesuai dengan tujuan pengukuran proses kuantitatif dan tujuan kinerja.	Dibutuhkan pengukuran kinerja proses serta frekuensinya yang sudah ditentukan dan tercantum dalam SOP.
		e) Hasil pengukuran dikumpulkan, dianalisis dan dilaporkan untuk memantau seberapa baik tujuan kinerja proses kuantitatif terpenuhi.	Dibutuhkan hasil pengukuran kinerja yang telah dilaporkan dan dilakukan *review*.
		f) Hasil pengukuran digunakan untuk mengkarakterisasi kinerja proses.	Dibutuhkan hasil pengukuran untuk mendukung pencapaian obyektif proses yang tercantum dalam SOP.
	PA 4.2	a) Teknik analisis serta kontrol diidentifikasi dan diaplikasikan jika dapat dijalankan.	Dibutuhkan metode, formula serta sumber data yang diperlukan untuk pengukuran kinerja.
		b) Batas kontrol variasi ditetapkan untuk kinerja proses normal.	Dibutuhkan suatu target atau batasan target sebagai kriteria keberhasilan proses.
		c) Data pengukuran dianalisis untuk penyebab khusus variasi.	Dibutuhkan catatan analisis atau tindak lanjut atas ketidaktercapaian kinerja proses.
		d) Tindakan korektif diambil untuk mengatasi penyebab spesifik atau khusus dari perubahan maupun variasi.	Dibutuhkan daftar *corrective action* untuk variasi kinerja yang tidak sesuai dengan target.
		e) Batas kontrol didefinisikan ulang sesuai kebutuhan setelah tindakan korektif.	Dibutuhkan *review* serta *rebaselining* ukuran (target) kinerja jika terdapat variasi.

Figure 16. Level 4 Capability Level Fulfillment Target

F. Assessment

The EDM01 (Ensure Governance Framework Setting and Maintenance) process aims to provide a consistent, aligned and integrated approach through a corporate governance approach. The capability assessment results in EDM01 process using self assessment template framework COBIT 5 [20] listed in Figure 17 below.

Nama Proses	Tingkat							
	0	1	2		3		4	
EDM01		PA 1.1	PA 2.1	PA 2.2	PA 3.1	PA 3.2	PA 4.1	PA 4.2
Penilaian Berdasarkan Kriteria	F	F	F	F	F	F	F	F
Penilaian Berdasarkan Persentase	100%	100%	100%	100%	100%	100%	86.66%	90%
Pencapaian Tingkat Kapabilitas								4

Figure 17. Process Capability Assessment on EDM01

The score reached by EDM01 process, is 97.08% and reaches capability level at 3.88, as shown in Figure 18 below.

Process Name	Level	Nilai per Level	Nilai Akhir	Capability Level
EDM 01	1	100%	97.08%	3,88
	2	100%		
	3	100%		
	4	88.33%		

Figure 18. EDM01 Final Result

The EDM02 (Ensure Benefits Delivery) process has the aim of maximizing the contribution related to business value from business processes, assets and information technology services generated by investments made at an affordable and reasonable cost. The capability assessment results in EDM02 process using self assessment template framework COBIT 5 [20] listed in Figure 19 below.

Nama Proses	Tingkat							
	0	1	2		3		4	
EDM02		PA 1.1	PA 2.1	PA 2.2	PA 3.1	PA 3.2	PA 4.1	PA 4.2
Penilaian Berdasarkan Kriteria	F	F	F	F	F	F	F	F
Penilaian Berdasarkan Persentase	100%	100%	100%	100%	100%	100%	86.66%	90%
Pencapaian Tingkat Kapabilitas								4

Figure 19. Process Capability Assessment on EDM02

The score reached by EDM02 process, is 97.08% and reaches capability level at 3.88, as shown in Figure 20 below.

Process Name	Level	Nilai per Level	Nilai Akhir	Capability Level
EDM 02	1	100%	97.08%	3,88
	2	100%		
	3	100%		
	4	88.33%		

Figure 20. EDM02 Final Result

The EDM04 (Ensure Resource Optimisation) has the objective of ensuring sufficient IT-related capabilities (human resources, technology and processes) to effectively support business objectives with optimal costs and opportunities to increase

benefits and realization of future changes. The capability assessment results in EDM04 process using self assessment template framework COBIT 5 [20] listed in Figure 21 below.

Nama Proses	Tingkat							
	0	1	2		3		4	
EDM04		PA 1.1	PA 2.1	PA 2.2	PA 3.1	PA 3.2	PA 4.1	PA 4.2
Penilaian Berdasarkan Kriteria	F	F	F	F	F	F	F	F
Penilaian Berdasarkan Persentase	100%	100%	100%	100%	100%	100%	86.66%	90%
Pencapaian Tingkat Kapabilitas								4

Figure 21. Process Capability Assessment on EDM04

The score reached by EDM04 process is 97.08% and reaches capability level at 3.88, as shown in Figure 22 below.

Process Name	Level	Nilai per Level	Nilai Akhir	Capability Level
EDM 04	1	100%	97.08%	3,88
	2	100%		
	3	100%		
	4	88.33%		

Figure 22. EDM04 Final Result

G. Gap Analysis

After obtaining the Capability Level of each assessment process, a gap analysis will be carried out to identify the results obtained have reached the Level 4 (Predictable) target that has been determined together with the IT Governance Coordinator of the IT division of PT. XYZ, as shown in Figure 23 & Figure 24:

No	Process Name	Level		Gap Analysis
		Target	Assessed	
1	EDM01	4	3,88	0,12
2	EDM02	4	3,88	0,12
3	EDM04	4	3,88	0,12

Figure 23. Gap Analysis
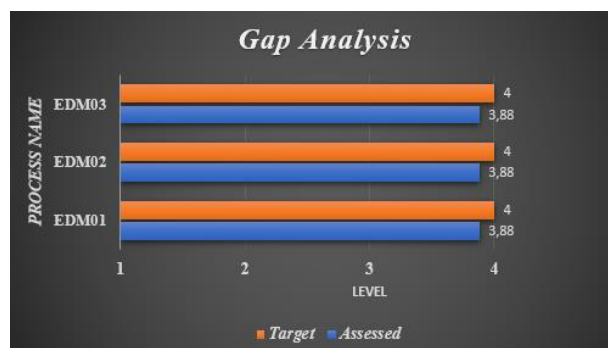
Figure 24. Bar Chart Gap Analysis

H. Finding

The findings obtained, as shown in Figure 25:

No	Tingkat	Atribut	Kriteria	Temuan
1.	4	PA 4.1	d	Belum adanya frekuensi pengukuran kinerja dalam dokumen SK Dir XYZ 047-2019 Pedoman HC – Combined.
2.	4	PA 4.2	e	Belum adanya mekanisme rebaselining dalam Pengisian Evaluasi RKAP 2021 7M21 Divisi XYZ ITrev20210901.

Figure 25. Gap Analysis Finding

I. Improvement Recommendation for EDM01, EDM02 and EDM04

Based on the measurement of IT governance using the COBIT 5 framework conducted at PT. XYZ, then the following conclusions can be drawn:

1. An update is needed regarding the addition of the frequency of performance measurement in the SK Dir XYZ 047-2019 Guidelines for HC – Combined to meet PA 4.1 criteria d).

2. It is necessary to update regarding the rebaselining mechanism (main alignment of the performance measurement basis to increase the correlation between work plans, budgets, scope and schedule) to variations that have not been clearly analyzed in the Completion of Evaluation of RKAP 2021 7M21 Division XYZ_ITrev20210901 to meet PA 4.2 criterion e).

V. CONCLUSION

The conclusions obtained from this research are:

1. Based on the identification of company goals, identification and mapping of enterprise goals, mapping of IT related goals, as well as mapping and determination of the COBIT 5 process domain. Selected and agreed jointly with the IT Governance Coordinator of the IT division of PT. XYZ to assess or conduct an assessment on the main processes of governance (governance), including EDM01, EDM02 and EDM04 where the three processes are given a target to reach level 4 (Predictable).

2. Based on the assessment that has been carried out, the EDM01, EDM02 & EDM04 process reached the final value with a large percentage level or in other words reached Fully Achieved (F) with a capability level that almost reached the expected target and had a very small gap value.

3. To meet the gap / gap values that exist in EDM01, EDM02 and EDM04:

   - An update is needed regarding the addition of the frequency of performance measurement in the SK Dir XYZ 047-2019 Guidelines for HC – Combined to meet PA 4.1 criteria d).

- It is necessary to update regarding the rebaselining mechanism (main alignment of the performance measurement basis to increase the correlation between work plans, budgets, scope and schedule) to variations that have not been clearly analyzed in the Completion of Evaluation of RKAP 2021 7M21 Division XYZ_ITrev20210901 to meet PA 4.2 criterion e).

REFERENCES

[1] T. Oktarina, "TATA KELOLA TEKNOLOGI INFORMASI DENGAN COBIT 5," *Jurnal Informanika,* vol. 3, pp. 30-38, 2017.

[2] K. Rizki and N. Bahtiar, "Analisis Tata Kelola Teknologi Informasi (IT Governance) Menggunakan COBIT 5 (Studi Kasus di UPT Puskom Universitas Diponegoro)," *Jurnal Masyarakat Informatika,* vol. 11, pp. 49-58, 2017.

[3] R. R. Suryono, D. Darwis and S. I. Gunawan, "AUDIT TATA KELOLA TEKNOLOGI INFORMASI MENGGUNAKAN FRAMEWORK COBIT 5 (STUDI KASUS: BALAI BESAR PERIKANAN BUDIDAYA LAUT LAMPUNG)," *Jurnal TEKNOINFO,* vol. 12, pp. 16-22, 2018.

[4] D. A. O. Turang and M. C. Turang, "ANALISIS AUDIT TATA KELOLA KEAMANAN TEKNOLOGI INFORMASI MENGGUNAKAN FRAMEWORK COBIT 5 PADA INSTANSI X," *Kumpulan jurnaL Ilmu Komputer (KLIK),* vol. 7, pp. 130-144, 2020.

[5] H. M. Rumere, A. R. Tanaamah and M. N. N. Sitokdana, "ANALISIS KINERJA TATA KELOLA TEKNOLOGI INFORMASI PADA DINAS PERPUSTAKAAN DAN KEARSIPAN DAERAH KOTA SALATIGA MENGGUNAKAN FRAMEWORK COBIT 5.0," *Sebatik,* vol. 24, pp. 14-21, 2020.

[6] A. P. Agustian, "Penerapan Tata Kelola Teknologi Informasi Menggunakan Framework COBIT5 pada Urusan Bisnis dan Strategi TI di PT. XYZ," *Doctoral dissertation, Universitas Multimedia Nusantara,* 2021.

[7] A. P. Utomo, I. Nugroho and Saefurrohman, "Analisa dan Perancangan Audit Sistem Informasi Akademik dengan Framework COBIT 5," *Jurnal IKRA-ITH Informatika,* vol. 4, pp. 6-12, 2020.

[8] D. M. Efendi, S. Mintoro and IinSeptiana, "AUDIT SISTEM INFORMASI PELAYANAN PERPUSTAKAAN MENGGUNAKAN FRAMEWORK COBIT 5.0," *Jurnal Informasi Dan Komputer,* vol. 7, pp. 31-36, 2019.

[9] A. Wiraniagara and A. F. Wijaya, "ANALISIS TATA KELOLA TEKNOLOGI INFORMASI MENGGUNAKAN FRAMEWORK COBIT 5 DOMAIN DELIVER SUPPORT AND SERVICE (STUDI KASUS: YAYASAN EKA TJIPTA, JAKARTA)," *Sebatik,* vol. 23, pp. 663-671, 2019.

[10] ISACA, COBIT 5: Enabling Processes, ISACA, 2012.

[11] V. Felayati, R. Mulyana and R. W. Witjaksono, "Analisis dan Perancangan Tata Keloladan Pengelolaan Teknologi Informasi Berbasis Kerangka COBIT 5 DomainAlign-Plan-Organise (APO)," *JISA (Jurnal Informatika dan Sains),* vol. 1, pp. 14-24, 2018.

[12] B. Nadhiroh, O. Purwaningrum and S. Mukaromah, "STUDI LITERATUR : FRAMEWORK COBIT 5 DALAM TATA KELOLA TEKNOLOGI INFORMASI," *SCAN-Jurnal Teknologi Informasi dan Komunikasi,* vol. 16, pp. 7-14, 2021.

[13] E. Ekowansyah, Y. H. Chrisnanto and N. S. Puspita, "Audit Sistem Informasi Akademik Menggunakan COBIT 5 di Universitas Jenderal Achmad Yani," *Seminar Nasional Komputer dan Informatika (SENASKI),* pp. 201-206, 2017.

[14] W. Reynard, "COBIT 5: Tingkat Kapabilitas pada PT Supra Boga Lestari," *ULTIMA InfoSys,* vol. IX, pp. 18-23, 2018.

[15] M. H. Hassor and M. N.N.Sitokdana, "ANALISIS TATA KELOLA TEKNOLOGI INFORMASI PADA TVRI PAPUA MENGGUNAKAN COBIT5.0 DOMAIN MEA," *Sebatik,* vol. 25, pp. 373-381, 2021.

[16] R. M. Ali and D. A. R, "EVALUASI TATA KELOLA TEKNOLOGI INFORMASI PADA SISTEM INFORMASI AKADEMIK FAKULTAS TEKNIK UNIVERSITAS KHAIRUN TERNATE MENGGUNAKAN FRAMEWORK COBIT 5," *JUTEKIN,* vol. 6, pp. 91-100, 2019.

[17] C. M. Sufyana and E. Suharto, "ANALISIS PENGUKURAN TINGKAT KEMATANGAN SISTEM INFORMASI AKADEMIK MENGGUNAKAN COBIT 5.0 DI POLITEKNIK X," *Jurnal E-KOMTEK(Elektro-Komputer-Teknik),* vol. 2, pp. 101-116, 2018.

[18] L. D. Oktaviana, P. Pribadi and M. Sabrinawati, "Evaluasi IT Governance Menggunakan

Framework COBIT 5 (Studi Kasus : PT. XYZ)," *Jurnal Pro Bisnis,* vol. 12, pp. 56-68, 2019.

[19] A. N. Thamrin, Kusrini and Rismayan, "Framework Cobit 5 untuk Audit Tata Kelola Teknologi Informasi (Studi Kasus: Diskominfo Kota Palopo)," *Jurnal Pekommas,* vol. 6, pp. 9-15, 2021.

[20] ISACA, COBIT® 5: Self-assessment Guide Using COBIT® 5, ISACA, 2013.
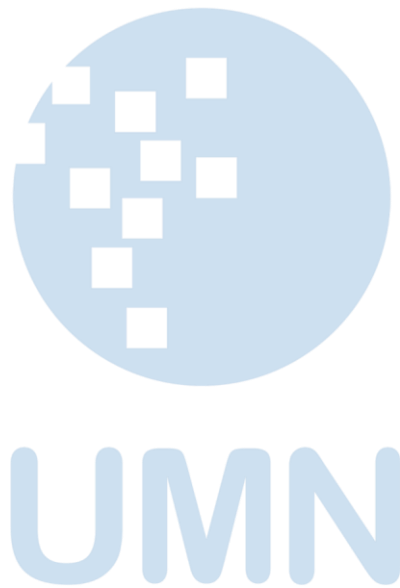
Indonesia Analysis Sentiment on Non Fungible Token (NFT)

Adhitia Erfina[1], Dinda Tasya Mahardika[2]

[1] Information Systems Study Program, Universitas Nusa Putra, Sukabumi, Indonesia

adhitia.erfina@nusaputra.ac.id

Abstract— NFT or Non-Fungible Token is a unique token attached to a digital asset that is connected to the blockchain system. Various assets, such as digital art, music cover, memes, sold as NFT. NFT has been widely discussed on various social media, Youtube is one of them. NFT has become a new trend for the Indonesian people, based on the fact that someone who sells a selfie photos at the Open Sea platform is going viral because it is an unsual event about why do they produce it. But people actually accept the trend as a mistake, they intentionally upload their identity on the platform. This happened because there was little information related to NFT and the public did not really understand that NFT could be a bridge for criminals. But in this case, many people as artists have been greatly helped in the marketing of their art. And even when the stock market is down, NFT remains one of the digital assets that attracts the attention of the world community. Therefore, this study was made to analyze the public's response with sentiment analysis. Data obtained from Youtube content comments and then classified into positive, negative, and neutral classes with Term Frequency-Inverse Document Frequency (TF-IDF) for the process of word weighting and classification using the Naïve Bayes Classifier algorithm. The test is carried out by calculating accuracy, precision, recall and F1-score, using a variety of data training and data testing. And the overall accuracy results are 64%, for positive prediction class precision is 63%, neutral class precision is 83%, while for negative prediction is 0% and recall obtained from positive is 99%, neutral recall is 0.7% while negative is 0%. These results are the data obtained on Youtube comment.

Keywords: Non Fungible Token, Naïve Bayes Classifier, Open Sea, Youtube;

I. INTRODUCTION

The transaction of Non Fungible Token (NFT) currently being widely in return, caused by the virality of a person named Ghozali who can sell his selfies for millions of rupiah on the Open Sea platform. Due to this virality, many Indonesian people are starting to be curious and want to know what NFTs really are and how to get money through NFT.[1] NFT is a digital content that is connected to the blockchain system. what is meant by digital content is photos, art, video, audio and others. NFT converts digital content into a digital asset that is assigned a special token number and verified through blockchain. NFT also has its own uniqueness, an NFT will not be the same as other NFTs, both in terms of its work or value.[2]

The difference between Fungible and Non-Fungible Assets are: Fungibility or fungible that means the equivalent is exactly the same as an asset or whatever can be exchanged for the same value. For example the currency value. if someone have 10,000 rupiah then it can be bought for an item that worth 10,000 rupiah. While Non-Fungible is an asset that has different values. For example, someone's autographed t-shirt from a favorite artist's concert, a polaroid album of someone's cherished memories, or a seat ticket someone bought to attend a sporting event. None of those items are interchangeable in the same way as there is an aspect of uniqueness in those items.[3]

At this time social media has become a very popular communication tool among internet users in Indonesia. Social media is a place to express and argue about various topics. YouTube is one of those things.[4]

Youtube is a platform that allows users to upload, watch, and share videos. Youtube also allows users to express their opinion on a video that can be commented, especially based on the comment string provided by Youtube. It can be used to analyze whether public opinion is positive or negative [5]. This process can be manually determined whether an opinion is positive or negative. But with the large number of public opinions, it takes more time and effort to classify the opinion. Therefore, the application of machine learning

techniques has been proposed to classify the various data of opinion. This research is expected to classify YouTube user comments on an NFT video by performing the Text Mining function to process the data classifying documents, as well as the application of the Naive Bayes Classifier (NBC) algorithm using probability calculations.

II. LITERATURE REVIEW

A. Non Fungible Token (NFT)

NFT (Token Non-Fungible) is a digital asset owned by an individual or group in which the Ethereum blockchain is used to record everything that occurs in it. NFT is one of the valuable or unique items with an unchangeable exchange rate. Artwork, game assets, photos, videos, music, etc are NFT products that can be sold. People can also convert document assets into NFTs. In terms of selling price, NFT is determined by subjective factors such as quality, creativity, and the artist's reputation. [6]

B. Python

Programming in the Python programming language is one of the high-level stages because Python is an interpretive and multipurpose language. Unlike other programming languages, Python is extremely simple to understand and learn. Python language prioritizes user understanding of code and syntax to make it easier to understand. As a result, Python is very simple to learn for both beginners and those who have mastered other programming languages. Guidovan Rossum, the creator of this language, first introduced it in 1991.[7]

C. Sentiment Analysis

Sentiment analysis refers to techniques for assessing and identifying positive and negative emotions and opinions [8]. Opinion mining is another term for sentiment analysis. This section is also known as "exploring the emotions behind user's words." Users are currently express themselves through online platforms such as social media, e-commerce, and websites. As a result, sentiment analysis on a social media platform is an option. One benefit of sentiment analysis is that it can save a significant amount of time and effort. Automatic sentiment analysis is also possible. A wide range of tools are outfitted with specialized algorithms for analyzing large amounts of data.[9]

D. Youtube Scraping

In Indonesia, 88% of the total population uses the YouTube social media platform. Social media features like providing information in the form of short videos and captions are supplemented by a variety of creations and innovations that can pique user interest and display video comments. Written expressions of user attitudes and emotions. Many researchers use social media to collect data in the form of user comments on specific topics and to investigate the opinions of social media users. To obtain data for these comments, a method known as "Scrapping" is required. Python, a programming language, can be used to perform this scraping method. To perform scraping methods, this programming language makes use of the Selenium and Beautiful Soup libraries and scanning data in comment format based on ID/name in HTML syntax.[10]

E. Preproccesing Data

This phase includes data selection and comment data cleaning. The following are the stages of the preprocessing process :

- Cleaning: Unwanted characters such as punctuation, periods, commas, question marks, exclamation points, HTML and URLs, hashtags and mentions, emojis, and irrelevant characters were removed.
- Tokenization: Separation of sentences into word by word, so that the analysis process can be easier.
- Case Folding: Equalizing words to lowercase in each prepared sentence or opinion in the data.
- Stopwords Removal: Deleting or eliminating words that are not important or clear. As a result, because fewer processes are passed, this classification stage can be considered more efficient.
- Stemming: Changing affixes into initial or basic words.

F. Lexicon Based

Lexicon Based in this study serves to classify an opinion [11]. One of the benefits of applying the Lexicon method is the documents are converted into word dictionaries in sentences, which are then compared spontaneously with opinion dictionaries created using the lexicon method. If the sentence in the prepared data is an opinion, it will be assigned the same value.

G. TF IDF

TF-IDF (Term Frequency-Inverse Document Frequency) weighting is a step-by-step method of converting text data that transforms data from textual data into numeric data for weighting of each word or feature. TF-IDF is a statistical measurement whose function is used as an assessment of how important words are in a document. TF counts how often a word appears in each sentence in the given data and shows the importance of that word in each document. DF counts documents containing words showing how often they occur. However, IDF is the opposite of DF calculation. The calculation of word weighting using TF-IDF is the result of multiplying TF with IDF. The

weight or value of a word will be high if it always appears in the document.[12]

H. Naïve Bayes Classifier

Naive Bayes classifier is one way of classification with Bayes' theorem. This classification uses a probabilistic and statistical method originally proposed by the British scientist Thomas Bayes. This method can also be called the Bayes theorem because it is a method of predicting future possibilities based on past experience. The basic function of Naïve Bayes Classifier is to assume a very strong independence from all conditions and events. Olson and Delen explain in their book that Naïve Bayes calculates the probabilities of each class of decisions. However, only if the decision class is true. This algorithm is based on the assumption that object attributes are independent. The probabilities included in the final estimate are calculated as the sum of the frequencies from the "master" decision table. The Naïve Bayes Classifier is more accurate than other classifier models. Journal of Naïve Bayes by Xhemali, Hinde, and Stone. "Decision Trees and Neural Networks in Training Web Page Classification" states that "the Naïve Bayes Classifier is more accurate than other classifier models".[13]

I. Confusion matrix

The classification procedure must, of course, provide an explanation as to whether or not a work process is efficient in the classification of this data. There are several methods of calculating the performance evaluation of the classification algorithm that can be used, one of which is the confusion matrix method. The results of the calculations contained in the application of the confusion matrix is the comparison of the classification results carried out by the prediction classification process with the actual classification results.

The confusion matrix which is the final result of this research calculation applies performance metrics which include Precision, Recall, F1-Score and Accuracy [14][15].

III. METHODOLOGY

The research method has guidelines in the form of research procedures or steps to align the expected results with the original purpose. In general, the flow of this research starts from determining the research topic to be raised and conducting a literature study on related research. The topic of Sentiment Analysis of NFT for the Indonesian population was raised because this issue was considered still hot and worthy of discussion given the many irregularities committed by the community. After the data source for the need for analysis is

obtained, the next step is to preprocess the data to clean up incomplete data. After that the results obtained will be tested for correctness using the measurement confusion matrix. Figure 1 shows the flow of this research.
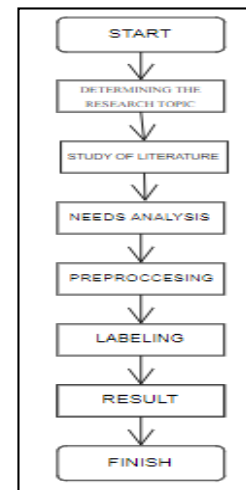
Figure 1. Research Flow

It can be seen that the design flow in the research uses a research methodology that has an appropriate and structured process design.

A. Data Collection

Data collection process uses the Scraping method where the data set comes from Youtube social media comments regarding "Non Fungible Tokens". Figure 2 shown description of the scraping process.
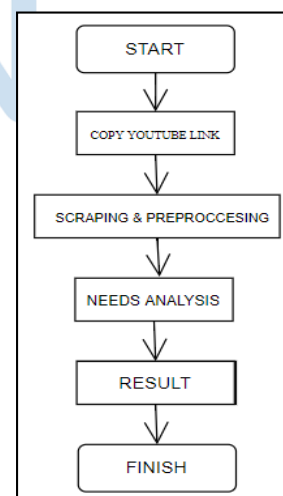
Figure 2. Data Collection Process

Data was scraped from Youtube website based on user comments using the selenium web driver library dictionary with the Non-Fungible Token (NFT)

keyword applied in Python and saved in.csv/.xls format. The obtained data will continue the preprocessing phase.

B. Preprocceing

Figure 3 describes the pre-processing phase, which is a continuation of the data scraping phase. Data are cleansed, tokenized, case-folded, and stopwords are removed such that a new dataset containing only clean data is produced.
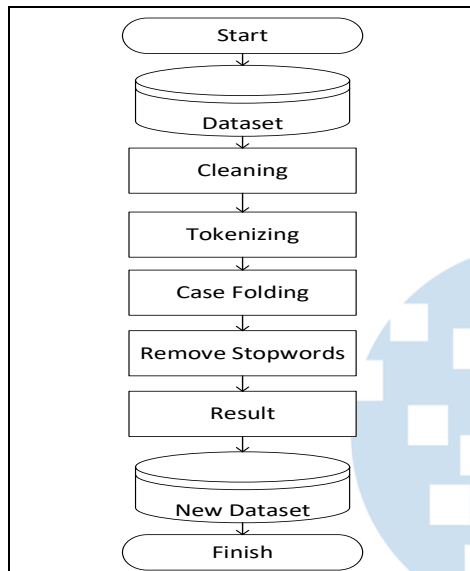
Figure 3. Scarping Process and Data Preprocessing

A new clean dataset is then obtained. This data will be used as the main data for sentiment analysis process. The data is obtained from YouTube video comments that contain NFT content and related variables. The data is then saved in .csv/.xlsx format so that it can be identified by Python.

C. Translate

Furthermore, the process of words translating on the clean data from Indonesian into English. This is done because in the python programming language there is only a vader sentiment library to bring up sentiment variables. This also called sentiment labeling. Example of Translated Youtube comment data results shown by table 1:

TABLE I.   RESULTS TRANSLATE

Comment
(before and after translating)
Saya melihat Tekhnologi NFT ini untuk melindungi para pekerja seni yg sering mengalami

pembajakan. Sehingga bisa menjaga originalitas sebuah karya di dunia digital.

I see this NFT technology to protect art workers who often experience piracy. So that it can maintain the originality of a work in the digital world.

D. Data labeling

The data is converted into English by carrying out the previous process, then the steps in labeling this data Youtube comment data must generate a polarity score from the results of this labeling calculation using a lexical dictionary. After calculating the polarity value, the sentiment variable analysis will appear positive, negative or neutral. The calculation raises the lexical sentiment.

$$Sentence\ sentiment \begin{cases} positif\ if\ Spositif\ >\ Snegatif \\ netral\ if\ Spositif\ =\ Snegatif \\ negatif\ if\ Spositif\ <\ Snegatif \end{cases} \quad (1)$$

E. Naïve Bayes Classifier

Naïve Bayes Classification is also included in an algorithm that is easy to use and simple and can predict an event based on the results of a good classification [16]. Naïve Bayes classification is also included in an algorithm that is easy to use and simple and can predict an event based on the results of a good classification.

F. Visualization

The next process is visualizing the data using the Matplotlib and Wordcloud libraries. The wordcloud visualization shows the words that often appear in the sentiment data. The output is a visualization in the form of a histogram that shows the results of the percentage accuracy of the polarity generated by each sentiment generated [17][18].

G. Confusion Matrix

In a data classification process, the use of methods must have an overview of the performance of each method. The confusion matrix method was chosen in this study. Because the Confusion Matrix method can calculate the accuracy and then compare it with the results of the classification of the actual procedure carried out with the classification results from the method. In this study, manual labeling is carried out so that the actual data classification can be seen from the polarity value. Visual or reference calculations from the confusion matrix method are Precision, Recall, F1-Score, and Accuracy.

- Precision model calculation results to determine the accuracy of estimates per value from the data compared to the model causation results. The formula for precision:

$$precision = \frac{TP}{TP + FP} \qquad (2)$$

- Recall is the results of model calculations related to the visualization of the suitability of refinding classification information. The recall formula is:

$$recall = \frac{TP}{TP + FN} \qquad (3)$$

- Accuracy is the result of data calculation that show how much accuracy is generated from this classification model. The formula is:

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \qquad (4)$$

IV. RESULTS AND DISCUSSION

The explanation in this section is the results and discussion, starting with the Scraping, preprocessing, sentiment labeling, visual data, word weighting and the classification process using the Naive Bayes Classification algorithm, using the Python language.

A. Data Collection

Youtube Comments data collection with Selenium Web Driver is manually using the inspect element by taking the comments class on the youtube web browser. It can be seen in figure 4 :
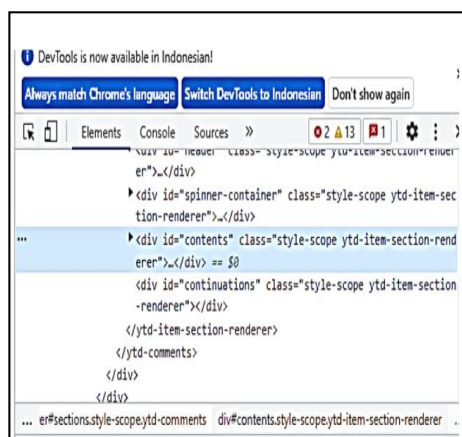
Figure 4. The process of taking youtube classes

Next, go through the youtube class properties to retrieve the data from the inspect element and paste it to the python coding section. The process of taking this youtube comment data must go through the stage of taking the comment class on the youtube website by using inspect element features on the web browser. This will make the comment data class (dataset) that just scraped is more relevant and just right.

The data retrieval uses several youtube links which produce 1188 data with the title of video content containing the word "Non Fungible Token NFT". After the scraping stage is executed, the data does not appear immediately because it has to go through the preprocessing stage. A visual representation of the scraping process, and preprocessing data retrieval with the selenium web driver as shown by figure 5:
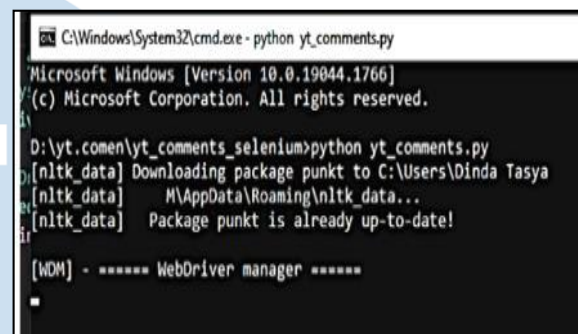
Figure 5. Run Scraping and Preprocessing

Figure 6. Scraping data in the web browser

The process in Figure 6 shows the process of taking comment data for further entry into the preprocessing stage. This data is unclean data that needs to be cleaned before entering the sentiment analysis stage. Data needs to be cleaned to get optimal results.

The data generated from scraping is text captured exactly as it appears in a YouTube comment. Hence, there is still noise or unnecessary information in the data. In this study, the generated data required preprocessing in order to be more organized.

Non-standard terms, such as slang, links, emoji, numerals, and punctuation marks, were the most commonly removed characters following the preprocessing phase. After removing these characters, the data are compiled into a new dataset with additional supporting attributes.

id	text_before	text_after	username	created_at
1	Pak Indrawan Nugraha mahaguru yang selalu kan	indrawan nugraha mahaguru tonton konten maju edukasi indonesia indra	unknown	2022-06-12 00:00:00
2	Nice discussion about NFT trending. Saya usul ba	nice discussion about nft trending usul bang deddy invite salah pakar cryp	Adif Laksa	2022-06-12 00:00:00
3	Saya melihat Tekhnologi NFT ini untuk melindun	tekhnologi nft lindung kerja seni alami baja jaga originalitas karya dunia d	Ahsin Arif	2022-06-12 00:00:00
4	Intinya nft ada untuk mendigitalisasi hak kekayaa	inti nft digitalisasi hak kaya intelektual nft digital seni fisik hak karya mud	Panggil sa	2022-06-12 00:00:00
5	mas dedi sebenernya tau jawaban dr pertanyaan	dedi sebenernya tanya krna phd bidang psycology problem topik bicara d	Ibnus Shir	2022-06-12 00:00:00
6	Om Ded, mohon izin saya kasih masukan. Sebelu	om ded mohon izin kasih masuk video doktor indrawan metaverse keren	Nico	2022-06-12 00:00:00
7	Kalo meneurutku om Ded sama Pak Indra memar	meneurutku om ded indra awam nft karya seni digital unik digital contrac	Ninoy Gra	2022-06-12 00:00:00
8	Konsep saat ini banyak NFT yang dapat digunakar	konsep nft konsep game play to earn metaverse nft ubah character gerak	Hana Veri	2022-06-12 00:00:00
9	Ada satu quotes yang saya ingat bener di dunia m	quotes bener dunia marketing barang beli logika nilai rendah barang beli	Ihfal Rama	2022-06-12 00:00:00
10	Master Dedi & Maha Guru Indra duet yang top ba	master dedi maha guru indra duet top banget buka paradigma konten kua	Yutra 888	2022-06-12 00:00:00
11	Banyak mungkin yang belum tahu kalau terkadan	terkadang nft utility balik harga project kasih akses holdernya community	Pradana D	2022-06-12 00:00:00
12	Pak Indrawan Nugroho, selalu dari sudut pandan	indrawan nugroho sudut pandang cerdas bahas detail asai tarik angkat	Nuri Hend	2022-06-12 00:00:00
13	Pak Indrawan dan bro Deddy terima kasih buat sh	indrawan deddy terima kasih sharingnya pikir sederhana bantu inovasi te	novianta	2022-06-12 00:00:00
14	Baru sedikit paham soal NFT, mendengarkan disk	paham nft dengar diskusi coba paham sudut pandang gambar gerak harga	Ilyazagus	2022-06-12 00:00:00
15	Semua sudah berubah mengikuti kemajuan jama	ubah ikut maju jaman bicara digital sisi kode pemrograman kreatifitas sisi	Agung Eka	2022-06-12 00:00:00
16	Setiap orang percaya apa yang mau dia percaya, r	orang percaya percaya nilai barang gantung perspektif nikmat alami harga	Anti Idiot	2022-06-12 00:00:00
17	Terimakasih sudah memberikan gambaran dan m	terimakasih gambar motivasi om dedy indra sehat serta keluarga amin rat	R waluya	2022-06-12 00:00:00
18	Saya rasa value dari NFT yang dibahas sekarang a	value nft bahas podcast bahas forum milik nft bangga kenal gila	Abdul Har	2022-06-12 00:00:00
19	I think the key is a Value . Like Pak Indrawan said	think the key is value like indrawan said at his content on rise above the c	Caesarzkn	2022-06-12 00:00:00
20	8:12 thank you mas indra karena menghemat 20 r	thank you indra hemat menit it takes humility to really admit own limitati	adrian bay	2022-06-12 00:00:00
21	penjelasan nilai NFT nya mungkin begini, untuk s	jelas nilai nft karya nft susah logic compare karya fisik monalisa karya pica	Rochale	2022-06-12 00:00:00
22	Best conversation om ded:"Nasib kita dibatasi ol	best conversation om ded nasib batas pikir pikir plafon tinggi capai hidup	Irny Karen	2022-06-12 00:00:00
23	Sukses selalu untuk #CLOSETHEDOORuntuk Om D	sukses om dedi indra moga senantiasa sehat edukasi pribadi bangsa indor	Bang AW	2022-06-12 00:00:00
24	Mengenal NFT atau Karya seni digital. Memang s	kenal nft karya seni digital susah pikir nalar coba ambil sisi positif bantu k	FATTL Bon	2022-06-12 00:00:00

Sheet1 / Sheet2 / Sheet3

Figure 7. Scraping and preprocessing results

In Figure 7 shows the data results that have gone through the scraping and preproccsing stages. So that a data that collected will be made a data table in order to facilitate the next stage of the process. The data frame contains five attributes, that is:

- *Id*: the id attribute contains the serial number of the data starting from 1-finish so that the data is easier to know how many there are

- *Text_before*: This attribute contains text or YouTube comments from scraping before preprocessing

- *Text_after*: this attribute contains text or youtube comments after preprocessing

- *Username*: username is the name created by the user on the comment maker's account on the youtube.

- *Created_at* : this attribute contains the time of scarping data

B. Labeling

This labeling stage uses a dictionary to reference the core or main language. Determining the origin of this study, to classify opinions and polarity scores using the Vader Sentiment dictionary. The way to calculate sentiment is by calculating the formula for polarity score < 0 for negative classification, polarity score = 0 for neutral classification, and polarity score > 0 for positive classification. Table II shows the example of the result in sentiment classification using the lexicon based method.

TABLE II. VADER SENTIMENT LABELING

Comment (before and after translating)	Score Polarity	Sentiment
saya lihat tekhnologi nft ini untuk lindungi para pekerja seni yg sering mengalami bajakan hingga bisa jaga originalitas buah karya di dunia digital.	0,8176	Positve
nft technology protect steel natural artworks keep the originality of digital world works	0,8176	Positve

The results of the data after being given a level with the library vader sentiment. The results of the data table appear containing youtube comments, Score polarity, and Sentiment. After all the data are labeled, then we will issue the positive and negative label data results with a bar plot as shown in the Figure 8:

Figure 8. Sentiment classification percentage

From the total dataset that has been labeled, it can be seen in Figure 8, from 100% of the data that has been labeled the results are 64.5% positive, 25.3% Neutral, and 10.2 Negative

C. Word Cloud Visualization

The results of the data above are then searched for what words that often appear using the word cloud. Figure 9 showns uses the python programming language and the matplotlib library.

Figure 9. Word Cloud Results

D. TF IDF

After the data has gone through the preprocessing stage, then there are 2 stages of this process, namely TF (Term Frequency) and IDF (Inverse Document Frequency). The number of words that exist in a document, and the more words that appear in each document, the higher the TF value. While the IDF is inversely proportional where if the words that appear not often or few in the document then the IDF value will be more than the words that often appear.[17] Figure 10 is an example of a TF-IDF script:


```
1   # proses TF-IDF dan
2   from sklearn.feature_extraction.text import CountVectorizer
3   from sklearn.feature_extraction.text import TfidfVectorizer
4
5   tf = TfidfVectorizer()
6   text_tf = tf.fit_transform(data['comment'].astype('U'))
7   text_tf
```

```
<1186x3324 sparse matrix of type '<class 'numpy.float64'>'
    with 15060 stored elements in Compressed Sparse Row format>
```

Figure 10. TF IDF process script

E. Data Spliting

In this data split stage, it means that the distribution of varied data according to the needs in the classification results later because this data split is very influential on the classification results in this study. The data is divided into 2 parts in the form of test data as well as training. sample data spliting script with 20% test data shown by figure 11.

```
# splitting data
from sklearn.model_selection import train_test_split

x_train, x_test, y_train, y_test = train_test_split
(text_tf, data['Value'], test_size=0.2, random_state=0)
```

Figure 11. Script for data split process

F. Naïve Bayes Classifier

After going through al the pre-processing and data spliting stages, the next stage of classification will be carried out using the Naive Bayes algorithm. Figure 12 shown the stage of classification using Naive Bayes Classifier.

```
# naïve bayes
from sklearn.naive_bayes import MultinomialNB
from sklearn.metrics import accuracy_score,
precision_score, recall_score, f1_score
from sklearn.metrics import classification_report
from sklearn.metrics import confusion_matrix

clf = MultinomialNB().fit(x_train, y_train)
predicted = clf.predict(x_test)
```

Figure 12. Naive Bayes Classifier Script

G. Classification results

At this stage the YouTube comment data has been calculated for its classification by the Naïve Bayes Classification algorithm using python programming. The results of the classification report is :

	precision	recall	f1-score	support
Negatif	0.00	0.00	0.00	18
Netral	0.83	0.07	0.13	72
Positif	0.63	0.99	0.77	148
accuracy			0.64	238
macro avg	0.49	0.35	0.30	238
weighted avg	0.65	0.64	0.52	238

Figure 13. Result of classification report

Figure 13 shows the results of sentiment analysis testing with Nave Bayes on YouTube social media concerning Non-Fungible Tokens with an accuracy value of 64%. Precision of positive prediction class 63%, precision of neutral prediction class 83%, and negative predictions class of 0%. Positive recall was acquired by 99%, neutral recall by 0.7%, and negative recall by 0%. These are the results derived from YouTube comments. The confusion matrix produced the following results:
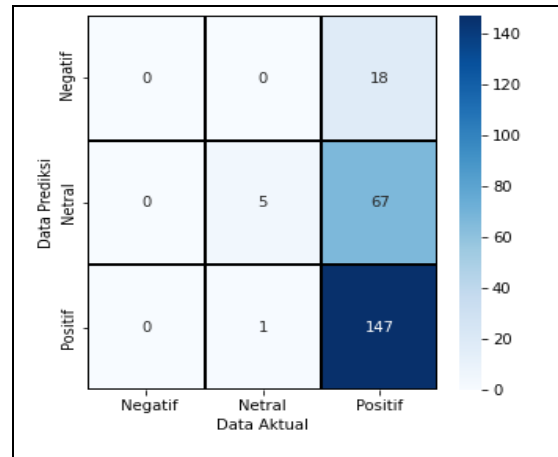
Figure 14. Confusion Matrix Results

V. CONCLUSION

Due to the success of a young man named Ghazali selling his selfie photo for billions of rupiah, the results indicate that a large number of Indonesians responded positively to the viral NFT event in Indonesia. This study analyzes public sentiment about Non-Fungible Tokens from YouTube comment using the Naive Bayes Classifier algorithm which achieves an accuracy of 64%. Precision of positive prediction class 63%, precision of neutral prediction class 83%, and negative predictions class of 0%. Positive recall was acquired by 99%, neutral recall by 0.7%, and negative recall by 0%. The results of sentiment analysis data with the Naïve Bayes algorithm are also assisted by TF-IDF in processing word weighting in text analysis

REFERENCES

[1] V. Maisyarah, "Viral Nih Bos, Ghozali Everyday Bisa Kaya Raya Dari Hasil Jual Foto Selfie Dalam Bentuk NFT." 2022. [Online]. Available: https://jurnalsumsel.pikiran-rakyat.com/lifestyle/pr-743466683/viral-nih-bos-ghozali-everyday-bisa-kaya-raya-dari-hasil-jual-foto-selfie-dalam-bentuk-nft

[2] "Mengenal Lebih Dekat Apa Itu NFT," *Kompas.com*. 2022. [Online]. Available: https://money.kompas.com/read/2022/01/13/083234526/mengenal-lebih-dekat-apa-itu-nft

[3] S. Livikacansera, "Ceruk Menjanjikan Bisnis Konten," *Republika.com*. 2022. [Online]. Available: https://www.republika.id/posts/16529/ceruk-menjanjikan-bisnis-konten

[4] A. Sentimen, S. Pekerja, and P. Tolak, "PADA MEDIA SOSIAL YOUTUBE MENGGUNAKAN ALGORITMA," no. 02, pp. 99–105.

[5] E. David, M. Sondakh, and S. Harilama, "Pengaruh Konten Vlog dalam Youtube terhadap Pembentukan Sikap Mahasiswa Ilmu Komunikasi," *Acta Diurna*, vol. 6, no. 1, 2017, [Online]. Available: https://ejournal.unsrat.ac.id/index.php/index/index

[6] D. P. Sari, "PEMANFAATAN NFT SEBAGAI PELUANG BISNIS PADA ERA METAVERSE," vol. 7, pp. 237–245, 2022.

[7] S. Falentino and E. Adhitia, *BAHASA ULAR UNTUK PEMROGRAMAN PYTHON*. 2021.

[8] S. Y. Nursyi'ah, A. Erfina, and C. Warman, "Analisis Sentimen Pembelajaran Daring Pada Masa Pandemi Covid-19 Di Twitter Menggunakan Algoritma Naïve Bayes," *J. Media Inform. Budidarma*, pp. 117–123, 2021.

[9] G. N. Arviana, "Sentiment Analysis, Teknik untuk Pahami Maksud di Balik Opini Pelanggan." 2021.

[10] P. Anjani, "Tutorial Scraping Komentar pada Media Sosial Youtube." 2020. [Online]. Available: https://predatech.org/tutorial-scraping-komentar-pada-media-sosial-youtube/

[11] A. H. Mulyadi and S. Lestari, "Analisis Sentimen Terhadap Sekolah Saat Covid-19 Pada Twitter Menggunakan Metode Lexicon Based," vol. 03, no. 01, pp. 17–23, 2022.

[12] J. A. Septian, T. M. Fahrudin, and A. Nugroho, "Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF - IDF dan K - Nearest Neighbor," pp. 43–49.

[13] C. Journal, M. S. Mustafa, M. R. Ramadhan, A. P. Thenata, P. Studi, and T. Informatika, "Implementasi Data Mining untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier," vol. 4, no. 2, 2017.

[14] G. N. Arviana, "Sentiment Analysis, Teknik untuk Pahami Maksud di Balik Opini Pelanggan." 2021.

[15] P. Anjani, "Tutorial Scraping Komentar pada Media Sosial Youtube." 2020. [Online]. Available: https://predatech.org/tutorial-scraping-komentar-pada-media-sosial-youtube/

[16] A. H. Mulyadi and S. Lestari, "Analisis Sentimen Terhadap Sekolah Saat Covid-19 Pada Twitter Menggunakan Metode Lexicon Based," vol. 03, no. 01, pp. 17–23, 2022.

[17] J. A. Septian, T. M. Fahrudin, and A. Nugroho, "Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF - IDF dan K - Nearest Neighbor," pp. 43–49.

[18] P. P. E. Indarbensyah and N. Rochmawati, "Penerapan N-Gram menggunakan Algoritma Random Forest dan Naïve Bayes Classifier pada Analisis Sentimen Kebijakan PPKM 2021," J. Informatics Comput. Sci., vol. 2, no. 04, pp. 235–244, 2021, doi: 10.26740/jinacs.v2n04.p235-244

Predicting the Case of COVID-19 in Indonesia using Neural Prophet Model

Efraim Yahya Wijaya[1], Alethea Suryadibrata[2]

[1] Department of Informatics, Universitas Multimedia Nusantara, Tangerang, Indonesia
efraim.wijaya@student.umn.ac.id
[2] Department of Informatics, Universitas Multimedia Nusantara, Tangerang, Indonesia
alethea@umn.ac.id

Abstract— **The spread of Coronavirus disease (COVID-19) is constantly changing in Indonesia. It is important to know the trends to help hospitals handle the crisis during outbreaks. Through this research, the prediction method is used to find increasing and decreasing of COVID-19 cases using the Neural Prophet model. Then the model is compared with the Facebook Prophet as comparison model. In this study dataset is used from (covid19.go.id) which was taken on 23 June 2022 with scraping technique. The results of this study indicate that the Neural Prophet model has better value in RMSE, and MAE compared to the Facebook Prophet.**

Index Terms—COVID-19; Facebook Prophet; MAE; Neural Prophet; RMSE;

I. INTRODUCTION

Since the initial entry of coronavirus disease 2019 (COVID-19) in Indonesia, the spread of virus is continuing to increase. The virus can quickly spread in the air through respiratory of an infected person [1]. This disease can cause fever, shortness of breath, even death. To suppress the spread of disease, the government makes a policy such physical distancing and wearing a mask. However, the emerge of new variant mutation such B.1.1.7 (Alpha) and B.1.617.2 (Delta) on 03 May 2021, causing the case to increase up to 56,757 active case on 15 July 2021 (covid19.go.id). This has significant impact on hospitals that provide medical treatment to patient that exposed to the COVID-19. Some patients with severe symptoms require advanced health resources, including respiratory support and intensive care.

Until this day the spread of virus can be controlled by policies dan regularization in every province such as limiting mobility with PPKM, self-isolation for people exposed with the COVID-19 and giving vaccine to the public. These factors create uncertainty to the spread of the COVID-19 virus. Based on COVID-19 information from the government website (covid19.go.id), shows the case of spread increased significantly at the time of long holidays. In addition, the mutation of variant B.1.617.2 (Delta) appeared in May 2021, and a new variant B.1.1.529 (Omicron) appeared in December 2021. Therefore, the forecasting method is needed to predict the number of infected cases in Indonesia.

Auto Regressive Integrated Moving Average (ARIMA) is one of the classical models and commonly used in various forecasting models [2]. Unfortunately, the ARIMA accuracy decreases when performing long range forecast. Moreover, data is not always linear. Therefore, many time series practitioner combining classical model with other model to make high quality forecast [3] [4]. However, the use of the model requires a high domain knowledge and expertise.

Recurrent Neural Network (RNN), Long Term Short Memory (LTSM), and Gated Recurrent Unit (GRU) is one of the deep learning models that can be used for prediction. This model is more robust compared with classical time series. The advantage of this model is that can remember important feature of sequential time series input. This model can produce good forecast in various forecasting models [5] [6]. Even so, this model is overly complex and could reducing the interpretability. Moreover, this model prone to underfit due small amount of data.

Facebook Prophet is one of the most popular models to predict time series data. This model can decompose time series to three main components: trend, seasonality, and holidays [7].This model use analysis-in-the-loop approach which can automatically to

modeling time series. Moreover, this model could provide insight and easy to interpret.

Neural Prophet is successor of the Facebook Prophet model [8]. This model has similar characteristic feature as Facebook Prophet but more extensibility like auto-regression (AR), lagged regressor, and feature regressor. This model capable to increase model accuracy with used of neural network. In addition, this model uses Pytorch so that this model can be developed further by time series practitioner.

There have been several studies to predict cases of COVID-19. Wahyudi and Palupi predict the peak of COVID-19 pandemic using Susceptible, Infected, and Removed (SIR) model [9]. Research by Khurana et al. [10] said that the prediction of Neural Prophet is better than other machine learning models. The research by Wildhanrahman et al. [11], said model Facebook Prophet cannot predict the peak of case because the case of COVID-19 is still increasing that time. Then the research by Harahap et al. [12], said Facebook Prophet prediction still need more dataset. Therefore, if dataset updated in range 2 years, the result could be more accurate.

In this research will be using Neural Prophet for predicting the case of COVID-19 in Indonesia by days. The case is from every positive confirmed case for all province in Indonesia. This model was preferred because it easy to interpret and can increase the accuracy of prediction with extensibility of neural network. This model will be compared with Facebook Prophet as a comparison model. For the metrics, it will use scale dependent error such as Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). The dataset which was taken from government website (covid19.go.id). This is public data which was updated every day from Kementerian Kesehatan (Kemenkes) to inform the surge increase of COVID-19 in Indonesia. The data is taken from 23 June 2022 with scraping technique. The model is expected to be used as anticipatory action against the surge increase of COVID-19 cases in Indonesia.

II. LITERATURE REVIEW

A. *Coronavirus disease (COVID-19)*

Coronavirus disease (COVID-19) are infectious diseases by SARS-CoV-2 viruses [1]. This virus first time appear in Wuhan, China in 2019. This virus can spread quickly through respiration of infected person. According to WHO [1], people who infected by this virus will experience mild respiratory illness and recover without requiring medical attention. Generally, people who has underlying medical condition such cardiovascular disease, diabetes, chronic respiratory

disease, or cancer are more likely to develop serious illness. Unfortunately, every people who infected by this virus will seriously ill even death. This virus first entry on 02 March 2020. Afterwards the virus mutated and become B.1.617.2 (Delta) in May 2021, and B.1.1.529 (Omicron) in December 2021. Until this day, virus spread continue. Therefore, we need forecast model to predict the future.

B. *Time Series Forecasting*

Forecasting is used to get information and used to as a reference to take decision in long term strategic planning. According to Hyndman et al. [13], forecast model is very dependent on availability of data. If the data is not available and not relevant, then the qualitative method preferred. Otherwise, if data is available and relevant, then quantitative method is preferred. Time series is one method to forecast quantitative method. As Hyndman et al. [13], said time series have criteria such numerical information about the past and having similar pattern in the future. Time series have composition such trend, seasonality, and cyclic

1. Trends occur when there is sudden increase or decrease in long term sequence of time series.

2. Seasonality occurs when time series affected by seasonal factor such a week, month, or even year. Seasonality occurs if data appear to be rise or fall and not in fixed frequency.

3. Cyclic normally occur by the fluctuation of economic, and always associated with business cycle. Cyclic happen with unknown factor.

C. *Facebook Prophet*

Facebook Prophet is a model that develop by Facebook Ai Researcher for solve a problem in forecasting. The first problem in time series forecasting is choosing a method for solving time series problems. Second, expertise required in designing a time series model. Therefore, Facebook Ai Researcher developing a framework that resulting a high-quality forecast. The Facebook Prophet uses analyst-in-the-loop which can allow time series practitioner to automate the model. In other word, if there is an error in the model, Facebook Prophet will give feedback to practitioner for model inspection [7]. In Taylor and Letham research, said that Facebook prophet uses decomposable time series or dividing time series to be three main model: trend, seasonality, and holidays. This model is similar with Generalized Additive Models (GAM), which the output from function sum is the result of prediction. Here is the following formulation of Facebook Prophet (1).

$$y(t) = g(t) + s(t) + h(t) + \varepsilon(t) \qquad (1)$$

Where, $g(t)$ is a trend function, which model nonperiodic changes in value of the time series. $s(t)$ is seasonality function, which has periodic change in (days, weekly, and yearly). $h(t)$ is a holidays function, that effect time series in one day or more days. $\varepsilon(t)$ is a error term function, where represent any idiosyncratic changes which was not accommodate by the model.

D. Auto-regressive Neural Network (AR-Net)

Auto-regressive Neural Network (AR-Net) is a model that developed by Triebe et al. [14], this model aims to

In Fig. 1, Triebe et al. [14], said that AR-Net model can mimic the process of traditional AR with neural network. AR-Net is designed so that the parameters of the first layer is the same as AR-coefficient. AR-Net could be extended to many hidden layers for increasing model accuracy. Where AR-Net architecture with $lag\ y(t-1), \dots, y(t-p)$ is input and y_t is target, every weight in line is $w_t, \dots, w_p$, and $H_1, \dots, H_k$.
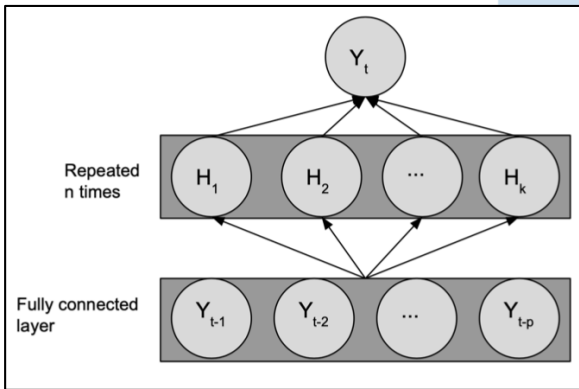

Fig. 1. AR-Net with Hidden Layers [14]

E. Neural Prophet

Neural Prophet is not much different from Facebook Prophet. This model was developed by Triebe et al. [8], to simplify the forecasting models and maintaining the purpose of Facebook Prophet such as interpretation and configuration. The Neural Prophet provides more extensibility features such as automatic differencing with Pytorch as backend. Neural Prophet is made fully modular, so that people could scale by adding more components.

According to Triebe et al. [8], the main concept of Neural Prophet model is modular composability, where every module will contribute to additive component which will be used in forecast. Every module has input and their process models. However, all modules must generate h output, where h is number of steps which will be forecasted in the future simultaneously [8]. Neural Prophet formulation found in Equation 2.

$$\check{y} = T(t) + S(t) + E(t) + F(t) + A(t) + L(t) \quad (2)$$

solve the scalability problem in Classical Autoregression (AR) model. The problem occurs when dealing with long-range dependencies, Classical AR could be slow when fitting data with large scale. According to Triebe et al. [14], model sequence-to-sequence such as Recurrent Neural Network (RNN) could solve the problem. However, Triebe said that RNN is too complex for typical time series problem and reduce interpretability. Therefore, Triebe et al. created the AR-Net model by combining Classical AR and feed-forward neural network.

Where, $T(t)$ is trend function in time t. $S(t)$ is seasonality in time t. $E(t)$ is event function in time t. $F(t)$ is regression function in time t for future-known exogenous variable. $A(t)$ is autoregression function in time t based on previous observation. $L(t)$ is regression function in time t for lagged observation of exogenous variable. All the components can be configurable and combined to make a model. If all component not configurable then only trend and seasonality are used. [8]

F. Root Mean Square Error (RMSE)

Root Mean Square Error (RMSE) is scale dependent error, where the error value is dependent on scale and cannot be compared with different scale. For calculating the RMSE is square root of mean value of predicted and actual.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \check{y}_i)^2}{n}} \quad (3)$$

Where,

n : Amount of data.

$\check{y}_i$: Prediction value.

y: Actual value.

G. Mean Absolute Error (MAE)

Mean Absolute Error (MAE) is scale depended on error, where the error value dependent with scale and cannot be compared with different scale. For calculating the MAE is mean of absolute value from data predicted and actual.

$$MAE = \frac{\sum_{i=1}^{n}|y_i - \check{y}_i|}{n} \quad (4)$$

Where,

n : Amount of data.

$\check{y}_i$: Prediction value.

y: Actual value.

III. METHODOLOGY

A. Data Retrieval

The data was taken from the government website (covid19.go.id) with scraping technique. The observation data are only daily confirmed cases from all provinces in Indonesia. The daily case is including holidays and weekend. In this research, the range of data used is from 02 March 2020 to 23 June 2022 or 843 days since the initial entry of COVID-19 in Indonesia.

B. Model Flowchart

Fig. 2, explain the approach in this research. First is data retrieval from the website government. Then changing the format data. After that, adding holidays into the models. Next, if we do experiment 1 then we split dataset into two, namely train data and test data. Different with experiment 2 we use all the datasets. After that, we fit the model with dataset. Then plotting all model result. Lastly, the models will be compared with RMSE and MAE.
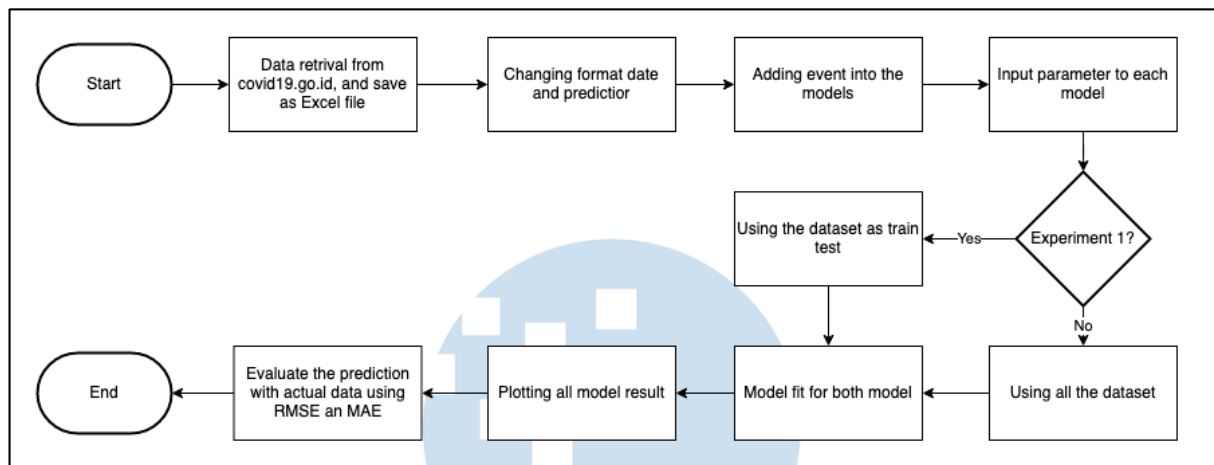
Fig. 2. Model Flowchart to Predict COVID-19 Case in Indonesia.

C. Implementation Method

In this research, we aim to implement Neural Prophet model. This model is easy to interpret like Facebook Prophet and it can increase model accuracy by extensibility of autoregression (AR) and neural network. First, we input the model parameter. Then, split the data. Adding Event. After that, creating a model. Next, plotting the prediction result. Lastly, we compared the model with Facebook Prophet as comparison. RMSE and MAE will be used as a metrics. Table 2 explain the model parameter that will be used.

TABLE I.   MODEL PARAMETER

Models	Parameters			
	Trend	Seasonality	Event	AR
Neural Prophet	Piece-wise linear	Weekly & yearly	Holidays & Event	AR-NN
Facebook Prophet	Piece-wise linear	Weekly & yearly	Holidays & Event	-

1. Experimentation

In this research there are two experiments. The first experiment, we split the data into two data, such as train data and testing data. The second experiment we use all the datasets. The result of the experiment will be resulting prediction value and model component.

2. Data split

Data split is used to determine the accuracy of the models. Table 2 explain the number of distribution dataset. In the first experiment, we split 776 datasets into 80:20, or 80% train data and 20% test data. The train data is 620 days, calculated from 02 March 2020 to 12 November 2021. While the test data is 156 days, calculated from 12 November 2021 to 17 April 2022. In the second experiment, we feed all data to the models with additional 67 days, so the total would be 843 days, calculated from 02 March 2020 to 23 June 2022.

3. Adding Event

In event we add the date of COVID-19 variant of concern first time found in Indonesia. In this case Alpha, Beta, Delta, Omicron, with additional Son of Omicron (BA1) and (BA.2). In holidays we add every holidays in Indonesia.

4. Creating Model

Modeling is done by adding all the model parameters into the model. For Neural Prophet, the model will be trained with mini batch SGD, the model will be measured with L1Loss or MAE. We

use L1Loss because it robust against outliers. For Facebook Prophet, model will fit directly with datasets.

Hyper-parameter will be used to tune the models. By doing so we could increase model accuracy. For Neural Prophet we can adjust learning rate, batch size, epoch, number of hidden layers, and number of nodes. For Facebook Prophet we can adjust changepoint prior scale, seasonality prior scale, and holidays prior scale.

5. Model Prediction

In first experiment, we use all models to predict 156 days ahead or all the data test (in-sample). In second experiment, we use all models to predict all the datasets. Lastly, in every experiment we predict 30 days ahead (out-sample).

D. Evaluation Methods

After experiment, the models can be compared with actual data. This is done because no observed value in future, so that the data prediction will be compared with actual data. We evaluate the prediction and actual value using RMSE and MAE. We used this metrics, because its commonly used in regression models. For comparison, the smallest result is the best model.

IV. RESULT AND DISCUSSION

A. Computational Resources

We use these computational resources to create a model. This specification is related to hardware and software. Here is the specification:

- Hardware: Macbook Air 2017: Intel i5, RAM 8 GB, Storage 512 GB.

- Software: Google Collab with Python 3.7 installed, and Python library (such as: Pandas, Matplotlib, NumPy, and SNS).

B. First Experiment

In first experiment we divide dataset to 80% train data, and 20% test data from 776 days. Then every model we add parameter. After that, we train the model with 620 data to predict 156 days ahead. The next step is plotting the prediction results. Then we evaluate the result with actual data (data set). Lastly, every model will predict 30 days ahead, start from 17 April 2022 to 17 May 2022. Here is the result of each model:

1. Neural Prophet

First, the Neural Prophet model will train the data. Then the model will train with 500 epoch data. Fig. 3, showing the error function with log-L1Loss, the log loss is enabled by default in Neural Prophet. In the graph show that the model is learning in 70 iteration and diverge after. So that, this model did not learn the

pattern of dataset. This happen because if dataset train and dataset test are imbalance then the plot loss will have a gap. If the dataset has sufficient data, then the result will be good.
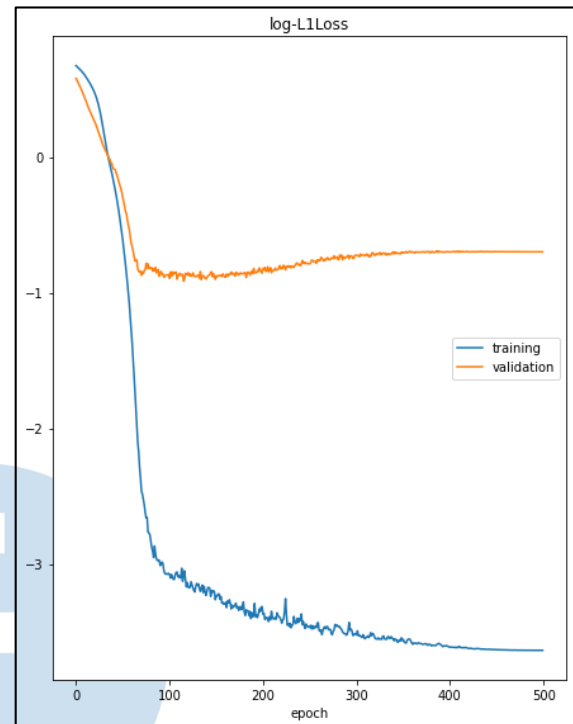
Fig. 3. Loss Function in Neural Prophet

Fig. 4, showing that the prediction of Neural Prophet is closely too accurate. This model cannot directly predict 156 days into the future with train data, because of the limited lag data in AR component. Therefore, this model will predict the test data (in-sample). In the figure show that Neural Prophet model can predict the increase in second wave of COVID-19 in data test. Unfortunately, the result sloping to negative values. This happen because the model does not have restriction to predict.
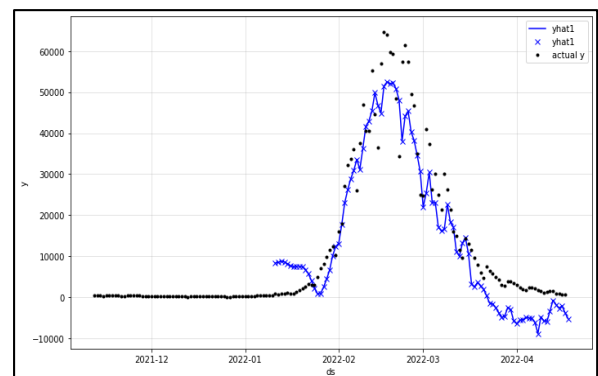
Fig. 4. Prediction of Neural Prophet
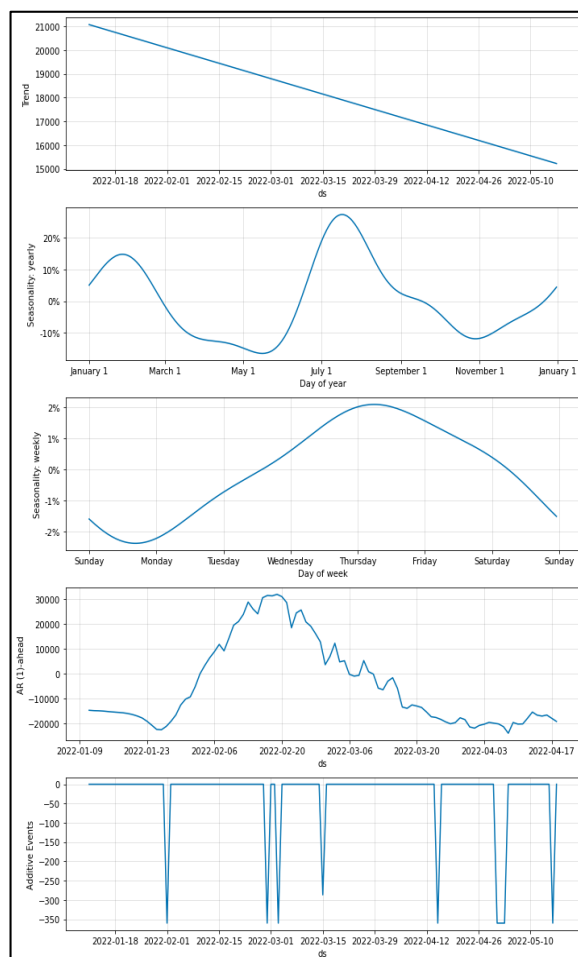
Fig. 5.   Model Decompose in Neural Prophet

Second, the model will decompose the main model to trend, seasonality (weekly, and yearly), AR, and event. In Fig. 5, the trend shows detrending case from January to May 2022. In seasonality yearly, show the surge increase in January and July. In seasonality weekly, show the peak of the case is mostly on Thursday. In AR 1-th step ahead, show that the prediction in AR with range of data test. In additive event, show that the event which affect the trend forecast.

Lastly, in Fig. 6, shows the prediction of 30 days ahead (out-sample) using test data. The prediction shows that the case decreasing to negative value for the next 30 days ahead based on 10 predictions of 30 *nth step forecast* parameters.
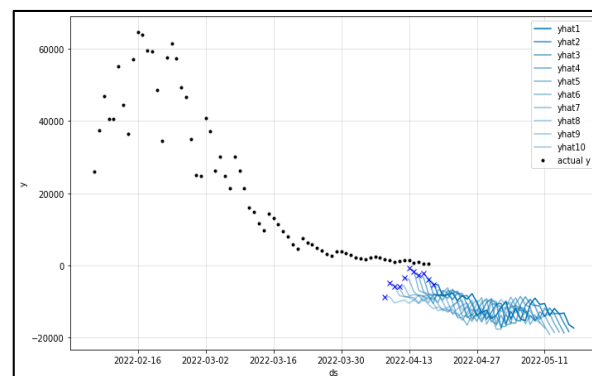
Fig. 6.   Prediction 30 days ahead Neural Prophet

2.   Facebook Prophet

First, the model will fit the data with 620 days of train data. This model does not have training process, so we directly using train data to predict into 156 days into the future. Because of this, the model will add the train data plus 156 days prediction into the trend component. In Fig. 7, showing that the Facebook Prophet is able to fit in train data but failed to predict the test data. This happen because model is overfit and resulting negative prediction.

Fig. 7.   Prediction of Facebook Prophet

Second, the model will decompose the main model into trend, event, seasonality weekly, and seasonality yearly. In Fig. 8, show the trend is declining in the next 156 days after train data. In event, shows the date of holidays that affect the trend. In seasonality weekly, show that the peak of the case is on Friday. In seasonality yearly, show the peak of the case is in July to September.

Fig. 8.   Model Decompose of Facebook Prophet.

Lastly, in Fig. 9, shows that the next 30 days ahead using data test (out-sample). The prediction shows negative result in the next 30 days ahead.
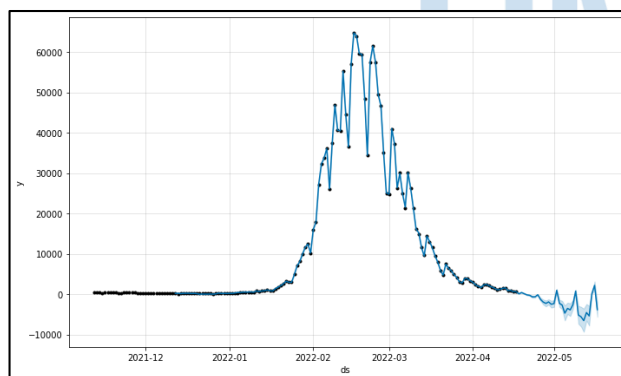
Fig. 9.   Prediction 30 days ahead Facebook Prophet

C.   Second Experiment

In second experiment we will use all the dataset to fit both models. Then every model we add parameter and hyper-parameter. Next, plotting the prediction results. After that, we try to predict 30 days ahead (out-sample). Then, we evaluate the result of the prediction with all the data. Here is the result of each model:

1.   Neural Prophet

First, the model will train all the dataset from 02 March 2020 to 23 June 2022 or 843 days after the first case of COVID-19 in Indonesia. Then the model will be fine-tuned with hyper-parameters. In Fig. 10, shows that the Neural Prophet model can predict all datasets accurately. This model can accurately predict the first, second, and third wave of COVID-19 in Indonesia.
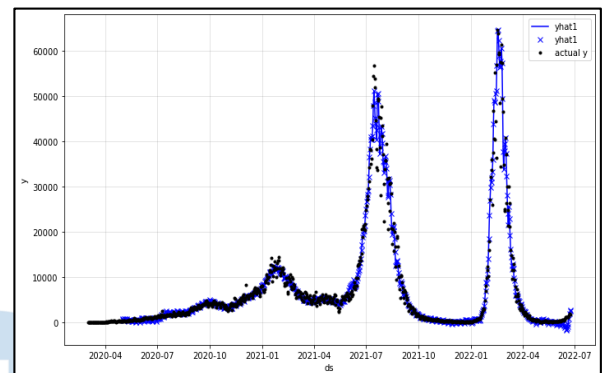
Fig. 10. Prediction of Neural Prophet

Second, the model will decompose the main model into trend, seasonality yearly, seasonality weekly, AR, and Event. In Fig. 11, shows the almost similar component in experiment 1 (Fig. 5), but with all the datasets. Thus, this model can capture more insight.
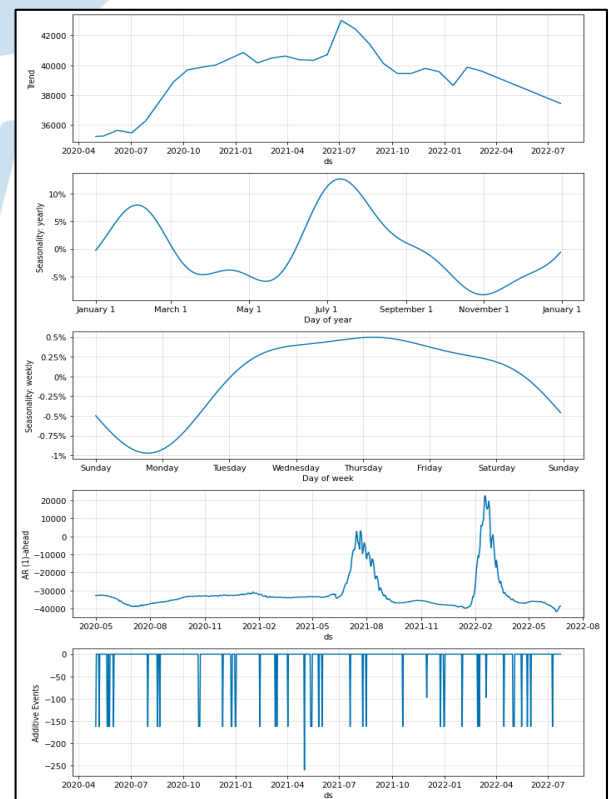
Fig. 11. Model Decompose of Neural Prophet

Lastly, in Fig. 12, shows that the next 30 days ahead (out-sample) prediction. The prediction shows surge increase in the next 30 days ahead. This prediction based on AR nth step ahead which n is 30 steps from forecast origin. In this prediction only 10 step origin forecasts shown. The yhat1 is the recent origin and yhat10 is previous forecast.
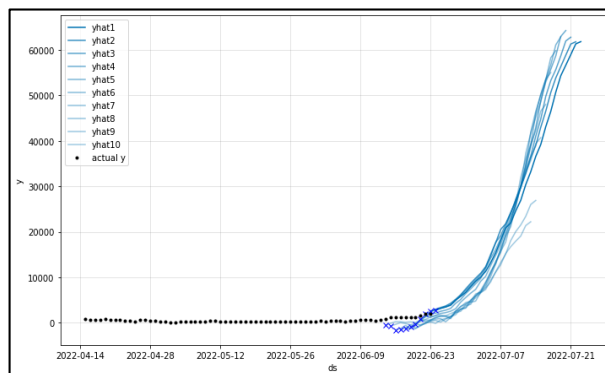
Fig. 12. Prediction 30 Days Ahead Neural Prophet

2. Facebook Prophet

First, the model will be fitted to all datasets from 02 March 2020 to 23 June 2022 or 843 days. Then the model will be fine-tuned with hyper-parameters. In Fig. 13, shows the Facebook Prophet can accurately predict all the dataset.
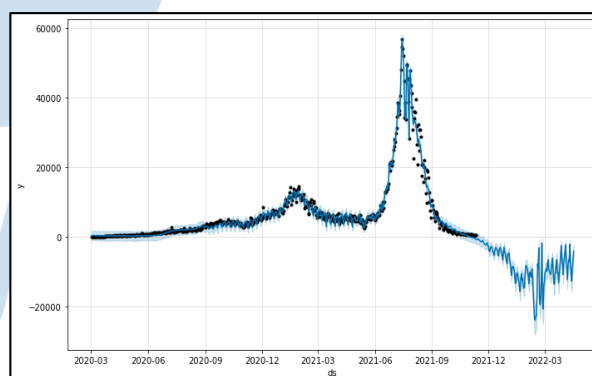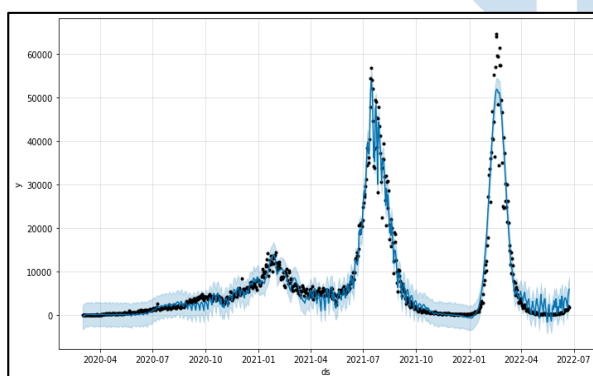

Fig. 13. Prediction of Facebook Prophet

Second, the model will decompose the main model into trend, holidays, seasonality weekly, and seasonality yearly. In Fig. 14, shows the smoother component after adding more data and fine-tuned hyper parameter. This shows that Facebook Prophet could compete with Neural Prophet with more data and proper fine-tune. Unfortunately, Facebook Prophet fail to capture yearly seasonality data.
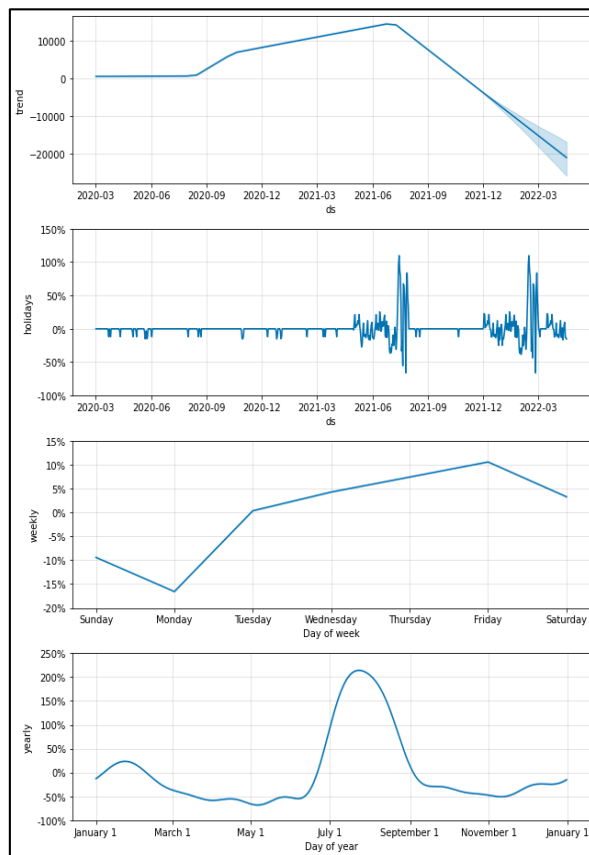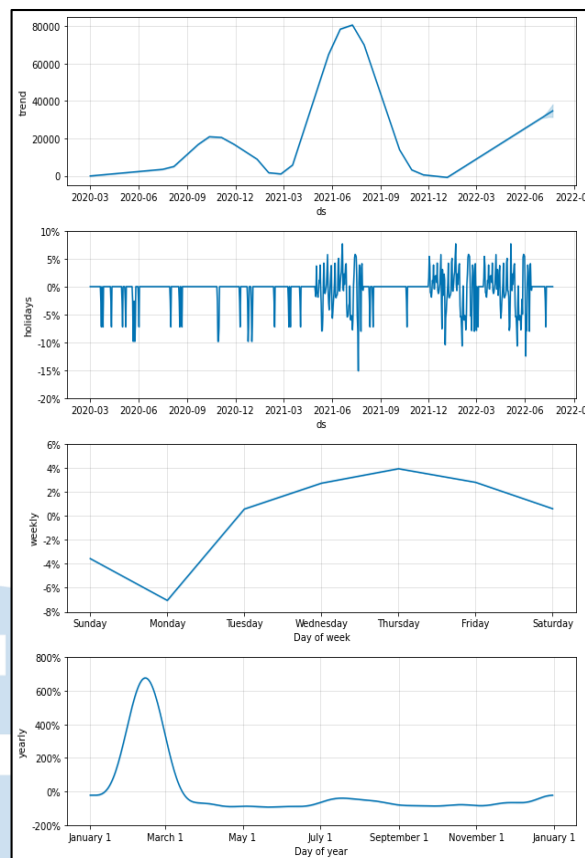

Fig. 14. Model Decompose of Neural Prophet

Lastly, in Fig. 11, shows that the next 30 days ahead (out-sample). The prediction shows that COVID-19 case will be increased in the next 30 days.
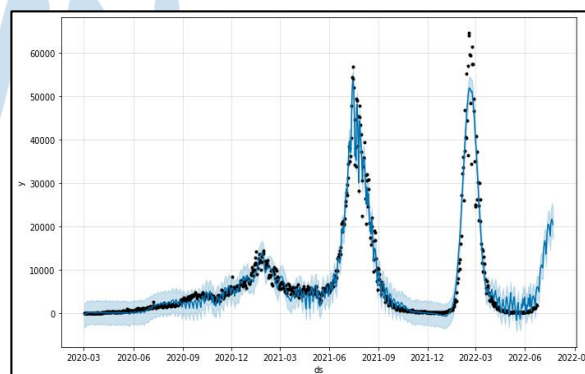
Fig. 15. Prediction 30 Days Ahead Facebook Prophet

D. *Model Evaluation*

The result of the model will be split into two parts, namely the first experiment and the second experiment. For first experiment, the model will be compared with data test. For the second experiment, the model will be compared with all datasets. The testing data is done by calculating the actual value with the predicted value,

then calculating the error with MAE and RMSE. The result of the experiment is seen in Table 2. From the prediction results, the Neural Prophet model has the smallest MAE and RMSE compared with Facebook Prophet. But with second experiment by adding more data, the Facebook Prophet can compete with Neural Prophet.

TABLE II.   PREDICTION RESULTS

Metrics	First Experiment		Second Experiment	
	Facebook Prophet	*Neural Prophet*	*Facebook Prophet*	*Neural Prophet*
RMSE	28,798.97	5,728.69	2,122.16	2,040.5
MAE	19,702.26	4,007.67	1,239.67	941.5

V.   CONCLUSION

Implementation of the Neural Prophet model to predict COVID-19 cases in Indonesia has been completed. The experiment show that the evaluation of Neural Prophet has better RMSE, and MAE compared to Facebook Prophet. The first experiment, the Neural Prophet has the result values of RMSE 5,728.69 and MAE 4,007.66. Then for the Facebook Prophet has the following result values of RMSE 28,798.97 for RMSE and 19,702.26 for MAE. The second experiment, the Neural Prophet model has the result of RMSE 2,040.5 and MAE 941.73. While the Facebook Prophet model has the result of RMSE 2,122.16 and MAE 1,239.67. In the future, the forecast method could be implemented with UI so that people can easily see the prediction. Also in future, the model could automatically add the data or holidays. In addition, if the number of datasets sufficient, the deep learning models could be used in future research.

REFERENCES

[1] WHO, "Coronavirus disease (COVID-19): how is it transmitted?," www.who.int, 13 December 2020. [Online]. Available: https://www.who.int/news-room/questions-and-answers/item/coronavirus-disease-covid-19-how-is-it-transmitted. [Accessed 7 February 2022].

[2] Y. Ardila, "Hybrid ARIMA and multilayer perceptron for short term forecasting of electricity consumption in East Java," Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia, 2016.

[3] M. Khashei and Z. Hajirahimi, "A comparative study of series ARIMA/MLP hybrid models for stock price forecasting," *Communications in Statistics - Simulation and Computation,* vol. 48, no. 9, pp. 2625-2640, 2018.

[4] C. B. A. Satrioa, W. Darmawan, B. T. U. Nadia and N. Hanafia, "Time series analysis and forecasting of coronavirus disease in Indonesia using ARIMA model and PROPHET," *Procedia Computer Science,* vol. 179, p. 524–532, 2021.

[5] K. ArunKumar, D. V. Kalaga, C. M. S. Kumar, M. Kawaji and T. M. Brenza, "Forecasting of COVID-19 using deep layer recurrent neural networks (RNNs) with gated recurrent units (GRUs) and long short-term memory (LSTM) cells," *Chaos, Solitons & Fractals,* vol. 146, p. 110861, 2021.

[6] S. Hansun and A. Suryadibrata, "Gold price prediction in COVID-19 era," *International Journal of Computational Intelligence in Control,* vol. 13, no. 2, pp. 1-4, 2021.

[7] S. J. Taylor and B. Letham, "Forecasting at Scale," *The American Statistician,* vol. 72, no. 1, pp. 37-45, 2017.

[8] O. Triebe, H. Hewamalage, P. Pilyugina, N. Laptev, C. Bergmeir and R. Rajagopal, "NeuralProphet: explainable forecasting at scale," arXiv:2111.15397, 2021.

[9] B. A. Wahyudi, and I. Palupi, "Prediksi puncak pandemi Covid-19 di Indonesia dengan model SIR," *Jurnal Teknologi dan Sistem Komputer,* vol. 9, no. 1, pp. 49-55, Jan. 2021.

[10] S. Khurana, G. Sharma, N. Miglani, A. Singh, A. Alharbi, W. Alosaimi, H. Alyami and N. Goyal, "An intelligent fine-tuned forecasting technique for COVID-19 prediction using neuralprophet model," *Computers, Materials & Continua,* vol. 71, no. 1, pp. 629-649, 2022.

[11] F. I. Wildhanrahman, C. L. Abdulbaaqiy and M. F. Dirhami, "Prediction of the peak of the COVID-19 outbreak in Indonesia using a logistic model and forecasting with FB prophet," *Unri Conference Series: Community Engagement,* vol. 2, pp. 352-358, 2020.

[12] M. Harahap, A. Z. Andika, A. M. Husein and A. Dharma, "Analisis tren dan perkiraan pandemi COVID-19 di Indonesia menggunakan peramalan metode prophet :sebelum dan sesudah aturan new normal," *Jurnal Teknologi Informasi Dan Ilmu Komputer,* vol. 9, no. 1, pp. 51-60, 2019.

[13] R. J. Hyndman and G. Athanasopoulos, Forecasting: principles and practice, Melbourne, Australia: OTexts, 2021.

[14] O. Triebe, N. Laptev and R. Rajagopal, "AR-Net: a simple auto-regressive neural network for time series," arXiv:1911.12436, 2019.

[15] S. P. COVID-19, "COVID-19 Response Acceleration Task Force," Response Acceleration Task Force, [Online]. Available: https://covid19.go.id. [Accessed 20 02 2022].

Analysis Sentiment in Bukalapak Comments with K-Means Clustering Method

Rena Nainggolan[1], Fenina Adline Twince Tobing[2], Eva J.G. Harianja[3]
[1,3] Komputer Akuntansi, Universitas Methodist Indonesia, Medan, Indonesia
filipiopat@gmail.com
[2] Prodi Informatika, Fakultas Teknik dan Informatika,
Universitas Multimedia Nusantara, Tangerang, Indonesia
fenina.tobing@umn.ac.id

Abstract— Technological development is very fast this era of globalization, to facilitate the work of many aspect that can be utilized, as well as for the flow of information. By applying computer technology in various fields, such as education, entertainment, health, tourism and culinary. Clustering is one of the Data Mining techniques. Clustering works by combining many data or objects into one cluster, with the aim that each data in one cluster will have data that is as similar as possible and different from data or objects in other groups. K-Means Clustering can to perform computations that are relatively fast and efficient in combining large amounts of data. In this research, there are 1407 comments which will be training data and testing data.

Keywords-; Clustering; K-Means Clustering; Sentiment;

I. INTRODUCTION

With the convenience and advantages of using E-Commerce, companies carry out online promotions using electronics and internet marketing by designing their own websites and entering information about the facilities and advantages possessed by theis companies on well known websites, intending to make appeal to prospective customers in the hope that prsopective customers will be interested in choosing and deciding to use it. [1]

Especially in the current, it is very demanding for business people to use computer technology so they can complete. In the current era of globalization, which is called e-commerce, transactions in the internet world are defined as e-commerce [2]

For example Bukalapak.com, this site helps potential customers who want to order goods or services because it make it easier for potential customers to make goods purchase transactions

In addition to promoting products owned by each seller, the e-commerce department also handles reviews from E-Commerce users. It can be seen from the use of the internet that it has now become a communication tool in everyday life, both in seeking information uses.

Bukalapak is an Indonesian technology company whose mission is to create a fair economy for all. Through its online and offline platforms, Bukalapak provides opportunities and choices for everyone to achieve a better life.

With the increasing use of the internet encouraging e-commerce [3] managers to further increase the effectiveness of promotions, one of the services is a platform for ordering goods, purchasing, and payment systems such as credit card installments, virtual Accounts, Noncredit card installment, in-store payment, internet banking, payment gateways, direct debit, electronic money, chat, shipping and returning gods if the goods do not match the order. With online purchases, online customer reviews appear which help consumers to be able to interact directly with service providers or tourist attractions, consumers can bargain, provide suggestions, or give impressions or feedback in using the facilities provided, one of which is open stalls.

Over time, this site is used by consumers or buyers, to share experiences while making purchases of goods using the site's facilities. To minimize the negative impact, before purchasing goods. Prospective buyers must find as many information as possible about the items to be purchased. The easiest way is to see ratings from reviews or ratings on products that will be purchased by potential customers.

II. METHOD

A. Sentimen Analisis

Sentiment analysis is the process of extracting, process, and understanding data in the form of text that is not structured automatically to retrieve information and sentiment contained in a sentence of opinion.[4] Sentiment analysis is performed to assess opinions and the tendency of an opinion towards a good topic negative or positive. Sentiment analysis can be applied to opinions in all fields such as economics, political, society, and law. Twitter's social media opens a window for researchers to study emotions, moods hearts, and public opinion through sentiment analysis. Sentiment Analysis is a field of Natural Language Processing (NLP) that builds a system to recognize and extract opinions in text form. [5]

B. Data Mining

There are two terms in data mining namely; such as knowledge discovery and pattern recognition. Each of them has a different meaning and has a definition from each other. [6]According to [7] the purpose of data mining is to obtain knowledge that is still hidden in chunks of data, where as according to [8] it is pattern recognition in chunks. The data to be extracted is called pattern recognition

C. Clustering

The way clustering works is to form a group from a collection of various physical or abstract objects into the same group [9]. The one group consists of a set of data that is as similar as possible between one data and another and must be different from the data or objects in the other groups[10]

Clustering performs the grouping of data without being based on certain that have been defined from the start. This process is very different from the classification process which at the beginning of the process must provide data classes. So clustering is often referred to as unstructured data grouping

D. TF-IDF

At this stage, the Tfidf Vectorizer function from the library is used to carry out the word weighting process [11]. After the TFIDF process has been carried out and the weight value of each word in each tweet sentence has been successfully obtained, the sentiment data classification process will be carried out.[12]

E. K-Means Clustering

The way the K-means method works is by grouping data or objects that have very similar characteristics in one cluster/group and will group data/objects into other clusters that have different characteristics and will eventually produce a cluster or group that has a different level. Very high resemblance [13]. The steps

for clustering with the K-Means method are as follows [14]

1. Determine K or the number of clusters 2.

2. Cluster centers are given an initial value with the random number. But in providing initialization of K cluster centers can be done in various ways, and the most common is randomly

3. In placing objects in which cluster they are placed, then it is calculated based on the distance between the two objects, as well as the distance between the cluster center and the object. To calculate the distance of all data to each cluster center point, you can use the Euclidean distance theory which is formulated as follows:

4. Recalculating the distance of the current cluster membership to the cluster center, the average of all data/objects in a particular cluster is the cluster center, using the mean (average value) or median

5. Do it again for each object with a new cluster, if the new cluster center does not change or change, then the clustering process stops, otherwise it will return to the third stage until the cluster center does not change.

$$D(i,j)= ((X1i−X1j)_↓+(X2i−X2j)_↓+..+(Xki−Xkj)$$

(1)

The following is an explanation of the flow of the research method

1. Identification of problems observing and finding problems that occur in the Bukalapak.com Application seen from the comments of the Bukalapak.com application user.

2. The determination of objectives serves to further clarify what framework is the target of this research. As already written in chapter I, the purpose of this research is to classify the comments of Bukalapak.com using the K-Means Clustering.

3. Library study aims to find out what theories will be used to solve the problems to be studied, as well as get strong reference bases for researchers.

4. Data collection the process of data collection and input in this system begins with web scraping data, the web scraping process is carried out using a data miner, a chrome extension that can be used for web data scraping. The output of this system is in the form of data testing with output values in the

form of positive and negative sentiments which are classified by the system based on training data learning.

5. Preprocessing

Preprocessing is one of the important stages for data in the mining process. The data used in the mining process is not always in ideal conditions for processing. Sometimes in the data, there are various problems that can interfere with the results of the mining process itself such as missing values, redundant data, outliers, or data formats that are not under the system. Therefore, to overcome these problems, the preprocessing stage is needed. Preprocessing is one of the stages of eliminating problems that can interfere with the result of data processing. This stage includes the following stages:

A. Cleaning Process

The cleaning Process that is carried out is changing text to lowercase (case folding, removing characters other than letters, removing user, usernames, or mention (@ 0, removing hashtags (#), and removing URLs or links from each comment.

B. Filtering

Filtering removes unnecessary words from token results. Apart from that, punctuation marks and stopwords were also removed. Stopwords are processed in a sentence if they contain words that often come out and are considered unimportant such as time, liaisons, and so on.

C. Stopword Removal

Stop words are common words that usually appear in large numbers but have no meaning. In indonesia like in, with, to, which, if, will and so on [15]. For that it is necessary to deletion process this word is required a data or list of words that you want to delete [16]

D. Stemming

At the stemming stage, namely changing affixed words into basic words. The stem process is removing the prefix and suffix words.[17]
Example:
Mempermainkan peranan 10 kuda di pementasan seni
Menjadi:
Main peran 1o kuda di pentas seni

After all the data is transformed into numbers, then the data has been obtained and grouped using the K-Means Clustering method. To be able to do to group the data into several clusters, several clusters, several steps need to be carried out, namely:

1. Determine in advance the number of clusters you want. In this study the data there into will be grouped into two clusters.

Determine the starting point of each cluster. In this study, the starting point was randomly generated.

III. RESULT AND DISCUSSION

A. Dataset

To carry out the clustering proposed in this and process ini this study. The authors tested the dataset based on the method proposed in this study, namely by using the K-Means Clustering method. The dataset used is crawling data from Bukalapak.com

Result of data collection the research was conducted using online customer reviews consisting of 1407 review

1. Barang bagus banget *_*
2. Alhamdulillah cocok 😊
3. Harga dan kualita ok
4. Makasih moga cocok @_@
5. Kualitas fungsi produk joss mantap
6. Respon cepat
7. JELEK!!
8. Barang sampe tujuan dengan cepat
9. Rekomendasi lapak ini mantap
10. Semoga berkhasiat dan cepat sehat
11. Lebih cepat dari perkiraan *8-)
12. Memuaskan
13. KECEWA…☹
14. Pasti order lagi
15. BARANG RUSAK %%

Dataset Processing method the dataset that has been obtained from the scrapping process still has a data structure of unsstructured, arbitrary and irregular data. Because of this, it is needed the data preprocessing stage is before the data set is tested with a model. This stage is done to clean data from noise and transform data into structured data, as for the stage in the processing of the data in this study namely:

1. Tokenize

Where in this process word splitting is done in the review sentence. This stage also removes certain character such as punctuation marks as well as filter by text length.

2. Stemming

Stemming in which the process find basic words by eliminating them all the affixes attached to the word.

3. Case Folding

Case Folding where the process utilizes the transform cases feature aims to homogenize the entire text into all lowercase

4. Stopwords Removal

Stopwords Removal where in this process the removal of the word is included in the stopword category. Stopword is words that appear frequently but is considered meaningless.

1. Paket hancur
2. Barang cepat sampai
3. Barang datang lama
4. Sesuai ukuran
5. Sangat kecewa
6. Barang bagus
7. Barang dikemas rapi
8. Penjual ramah
9. Respon cepat
10. Barang rusak

Define abbreviations and acronyms the first

C. String to word vector

The TF-IDF algorithm is applied to convert string data into word vectors, and generate a data matrix with the dimension of 25 attributes x 1407 data, from TF-IDF implementation. There are 25 terms in the data as shown in the following figure

Kece, persis, pas, cepat, profesional, elegan, mahal, murah, trendi, tukar, nyasar, parah, cocok, rapi, tebal, sobek, terimakasih, keras, enak, jelas, komplit, rusak, longgar, sama, salah

D. Converting Data to Numeric

The following table is an example of the final conversion result data.

18 0.004276, 23 0.005957, 27 0.005897, 34 0.025563, 37 0.01063, 43 0.084699, 72 0.504275, 91 0.017341, 94 0.006041, 95 0.005957, 99 0.0107 38,102 0.0188 66, 112 0.957791, 128 0.027989, 131 0.034241, 136 0.021461, 148 1.479877, 155 0.00586, 169 0.121277, 170 0.005802, 176 0.025946, 178 0.005328, 188 0.026011, 193 0.005868, 206 0.005904, 211 0.131904, 1 9 0.12829, 13 0.035806, 18 0.005942, 26 0.01387, 29 0.014772, 30 1.685916

E. Attribute Selection

The data above is still too large and ineffective, so the existing attributes must be filtered. By using the Cfs algorithm, there are 10 attributes

pas, cepat, parah, cocok, jelek, jelas, terimakasih, tebal, komplit, salah

F. K-Means Clustering Coding

Table 1 K-Means Clustering Coding

Data Ke-	Cluster 0	Cluster 1	Min Cluster
1	0.7148	0.8465	C0
2	1.3117	1.32	C0
3	2.1711	0.4824	C0
4	0.1005	0.2815	C0
5	1.1461	0.6353	C1
6	2.1636	2.2118	C0
7	3.1182	3.1818	C0
8	2.772	2.8258	C0
9	1.5869	1.4849	C1
10	2.3618	2.5538	C0
1407	1.2121	1.303	C0

Cluster Instance

Table 2 Cluster instances are shown in the table

	Cluster	Percentage
1	1278	91 %
2	129	9%

G. Graph Comparison

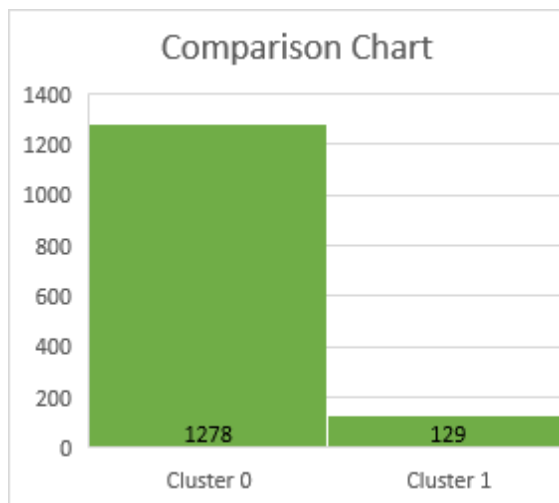Graph comparison of the result of each cluster from 1407 test data with 2 clusters is shown in the picture below

Fig 1. Graph Comparison Chart cluster 0 and Cluster 1

Comparison chart from the table above can be concluded that the result of testing on the1407 review yielded 2 clusters namely:

a. Cluster 0 produces 1278 (91%) reviews that was similar and very high which are grouped into 0 cluster
b. Cluster 1 resulted in 129 (9%) reviews which had very high similarity grouped into 1 cluster group

IV. CONCLUSION

Based on a test conducted on sentiment analysts at the store online with a total of 1407 data reviews, this study it based on the test. Conducted on sentiment analysts at the store online with a total of 1407 data reviews, this study offers a model clustering, and testing produces 2 clusters, and testing produces a tool to assist consumers in deciding to buy a product or service, because of the importance of sentiment analysis for increasing customer confidence. The information helps consumers to get the quality of the product they are looking for from reviews and experiences written by other consumers who have purchased the product from the previous online sellers.

The result obtained is 2 namely:

1. Cluster 0, which consists of 1728 comments or 91 % of customers who gives positive comments,
2. Cluster 1 consists of 129 comments or 9 % of customers who give negative comments.
3. By applying the K-means Clustering method to analyzing sentiment data products, it can help potential customers to determine whether the product is feasible or not to buy because sentiment data is used as a means for consumers to search and obtain information that will later influence decisions to purchase sentiment data also has a function as a tool decision making

REFERENCES

[1] E. Y. Nasution, P. Hariani, L. S. Hasibuan, and W. Pradita, "Perkembangan Transaksi Bisnis E-Commerce terhadap Pertumbuhan Ekonomi di Indonesia," *Jesya*, vol. 3, no. 2, pp. 506–519, 2020, doi: 10.36778/jesya.v3i2.227.

[2] K. Kasmi and A. N. Candra, "Penerapan E-Commerce Berbasis Business To Consumers Untuk Meningkatan Penjualan Produk Makanan Ringan Khas Pringsewu," *J. Aktual*, vol. 15, no. 2, p. 109, 2017, doi: 10.47232/aktual.v15i2.27.

[3] C. F. C. Dewantara, "ANALISIS DAMPAK PENGGUNAAN SITUS Bukalapak. Com, Perilaku, Terhadap Perilaku Pembelian Pada Komunitas Samarinda Photographer," *eJournal Ilmu Komun.*, vol. 3, no. 2, pp. 488–502, 2015.

[4] N. Ransi, L. Surimi, A. Tenriawaru, and L. O. Saidi, "Analisis Sentimen Masyarakat Terhadap Toko Online Aplikasi," pp. 1–8, 2020.

[5] F. V. Sari and A. Wibowo, "Analisis Sentimen Pelanggan Toko Online Jd.Id Menggunakan Metode Naïve Bayes Classifier Berbasis Konversi Ikon Emosi," *J. SIMETRIS*, vol. 10, no. 2, pp. 681–686, 2019.

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

[7] D. Firdaus, "Penggunaan Data mining dalam kegiatan pembelajaran," vol. 6, no. 2, pp. 91–97, 2017.

[8] A. S. Aribowo, "Metode Data Mining Untuk Klasifikasi Kesetiaan," vol. 10, no. 1, pp. 2–4, 2013.

[9] H. Priyatman, F. Sajid, and D. Haldivany, "Klasterisasi Menggunakan Algoritma K-Means Clustering untuk Memprediksi Waktu Kelulusan," vol. 5, no. 1, pp. 62–66, 2019.

[10] T. Hardiani, "ANALISIS CLUSTERING KASUS COVID 19 DI INDONESIA MENGGUNAKAN ALGORITMA K-MEANS Jurnal Nasional Pendidikan Teknik Informatika : JANAPATI | 157," vol. 11, pp. 156–165, 2022.

[11] M. Nurjannah and I. F. Astuti, "PENERAPAN ALGORITMA TERM FREQUENCY-INVERSE DOCUMENT FREQUENCY (TF-IDF) UNTUK TEXT MINING," vol. 8, no. 3, pp. 110–113, 2013.

[12] M. A. Rofiqi, A. C. Fauzan, A. P. Agustin, and A. A. Saputra, "Implementasi Term-Frequency Inverse Document Frequency (TF- IDF) Untuk Mencari Relevansi Dokumen Berdasarkan Query," vol. 1, no. 2, pp. 58–64, 2019.

[13] I. Sumadikarta and E. Abeiza, "PENERAPAN ALGORITMA K-MEANS PADA DATA MINING UNTUK MEMILIH PRODUK DAN PELANGGAN POTENSIAL (Studi Kasus : PT Mega Arvia Utama)," pp. 12–23, 2014.

[14] R. K. Dinata, N. Hasdyna, and N. Azizah, "Analisis K-Means Clustering pada Data Sepeda Motor," vol. 5, no. 1, 2020.

[15] A. F. Hidayatullah, "Pengaruh Stopword Terhadap Performa Klasifikasi Tweet Berbahasa Indonesia," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 1, no. 1, pp. 1–4, 2016, doi: 10.14421/jiska.2016.11-01.

[16] A. Setiawan, E. Kurniawan, and W. Handiwidjojo, "Implementasi Stop Word Removal Untuk Pembangunan

Aplikasi Alkitab Berbasis Windows 8," *J. EKSIS*, vol. 6, no. 2, pp. 1–11, 2013.

[17]   H. R. Pramudita, "Penerapan Algoritma Stemming Nazief & Adriani Dan Similarity Pada Penerimaan Judul Thesis," *J. Ilm. Data Manaj. dan Teknol. Inf.*, vol. 15, no. 4, pp. 15–19, 2014.

[18]   Nainggolan, R., & Tobing, F. A. (2020). ANALISIS CLUSTER DENGAN MENGGUNAKAN K-MEANS UNTUK PENGELOMPOKKAN ONLINE CUSTOMER REVIEWS (OCR) PADA ONLINE MARKETPLACE. *METHODIKA: Jurnal Teknik Informatika Dan Sistem Informasi*, 6(1), 1-5.
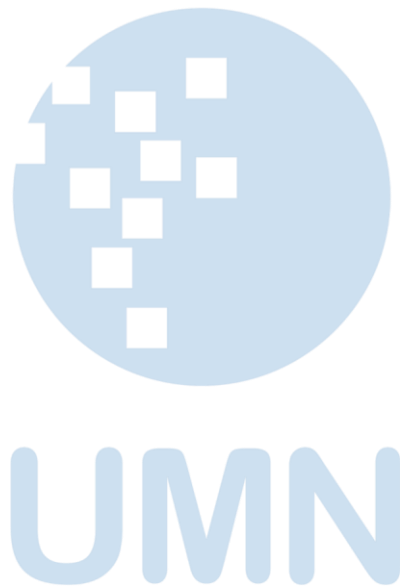
AUTHOR GUIDELINES

1. Manuscript criteria

- The article has never been published or in the submission process on other publications.
- Submitted articles could be original research articles or technical notes.
- The similarity score from plagiarism checker software such as Turnitin is 20% maximum.

2. Manuscript format

- Article been type in Microsoft Word version 2007 or later.
- Article been typed with 1 line spacing on an A4 paper size (21 cm x 29,7 cm), top-left margin are 3 cm and bottom-right margin are 2 cm, and Times New Roman's font type.
- Article should be prepared according to the following author guidelines in this template. Article contain of minimum 3500 words.
- References contain of minimum 15 references (primary references) from reputable journals/conferences

3. Organization of submitted article

The organization of the submitted article consists of Title, Abstract, Index Terms, Introduction, Method, Result and Discussion, Conclusion, Appendix (if any), Acknowledgment (if any), and References.

- Title
  The maximum words count on the title is 12 words (including the subtitle if available)
- Abstract
  Abstract consists of 150-250 words. The abstract should contain logical argumentation of the research taken, problem-solving methodology, research results, and a brief conclusion.
- Index terms
  A list in alphabetical order in between 4 to 6 words or short phrases separated by a semicolon (;), excluding words used in the title and chosen carefully to reflect the precise content of the paper.
- Introduction
  Introduction commonly contains the background, purpose of the research, problem identification, research methodology, and state of the art conducted by the authors which describe implicitly.

- Method
  Include sufficient details for the work to be repeated. Where specific equipment and materials are named, the manufacturer's details (name, city and country) should be given so that readers can trace specifications by contacting the manufacturer. Where commercially available software has been used, details of the supplier should be given in brackets or the reference given in full in the reference list.
- Results and Discussion
  State the results of experimental or modeling work, drawing attention to important details in tables and figures, and discuss them intensively by comparing and/or citing other references.
- Conclusion
  Explicitly describes the research's results been taken. Future works or suggestion could be explained after it
- Appendix and acknowledgment, if available, could be placed after Conclusion.
- All citations in the article should be written on References consecutively based on its' appearance order in the article using Mendeley (recommendation). The typing format will be in the same format as the IEEE journals and transaction format.

4. Reviewing of Manuscripts

Every submitted paper is independently and blindly reviewed by at least two peer-reviewers. The decision for publication, amendment, or rejection is based upon their reports/recommendations. If two or more reviewers consider a manuscript unsuitable for publication in this journal, a statement explaining the basis for the decision will be sent to the authors within six months of the submission date.

5. Revision of Manuscripts

Manuscripts sent back to the authors for revision should be returned to the editor without delay (maximum of two weeks). Revised manuscripts can be sent to the editorial office through the same online system. Revised manuscripts returned later than one month will be considered as new submissions.

6. Editing References

- **Periodicals**
  J.K. Author, "Name of paper," Abbrev. Title

of Periodical, vol. x, no. x, pp. xxx-xxx, Sept. 2013.

- **Book**
  J.K. Author, "Title of chapter in the book," in Title of His Published Book, xth ed. City of Publisher, Country or Nation: Abbrev. Of Publisher, year, ch. x, sec. x, pp xxx-xxx.

- **Report**
  J.K. Author, "Title of report," Abbrev. Name of Co., City of Co., Abbrev. State, Rep. xxx, year.

- **Handbook**
  Name of Manual/ Handbook, x ed., Abbrev. Name of Co., City of Co., Abbrev. State, year, pp. xxx-xxx.

- **Published Conference Proceedings**
  J.K. Author, "Title of paper," in Unabbreviated Name of Conf., City of Conf., Abbrev. State (if given), year, pp. xxx-xxx.

- **Papers Presented at Conferences**
  J.K. Author, "Title of paper," presented at the Unabbrev. Name of Conf., City of Conf., Abbrev. State, year.

- **Patents**
  J.K. Author, "Title of patent," US. Patent xxxxxxx, Abbrev. 01 January 2014.

- **Theses and Dissertations**
  J.K. Author, "Title of thesis," M.Sc. thesis, Abbrev. Dept., Abbrev. Univ., City of Univ., Abbrev. State, year. J.K. Author, "Title of dissertation," Ph.D. dissertation, Abbrev. Dept., Abbrev. Univ., City of Univ., Abbrev. State, year.

- **Unpublished**
  J.K. Author, "Title of paper," unpublished. J.K. Author, "Title of paper," Abbrev. Title of Journal, in press.

- **On-line Sources**
  J.K. Author. (year, month day). Title (edition) [Type of medium]. Available: http://www.(URL) J.K. Author. (year, month). Title. Journal [Type of medium]. volume(issue), pp. if given. Available: http://www.(URL) Note: type of medium could be online media, CD-ROM, USB, etc.

7. **Editorial Adress**

Paper Title

Subtitle (if needed)

Author 1 Name[1], Author 2 Name[2], Author 3 Name[2]

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

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

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

I. INTRODUCTION

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

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

II. EASE OF USE

A. Selecting a Template

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

B. Maintaining the Integrity of the Specifications

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

III. PREPARE YOUR PAPER BEFORE STYLING

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

A. Abbreviations and Acronyms

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

B. Units

- Use either SI (MKS) or CGS as primary units (SI units are encouraged).

- Do not mix complete spellings and abbreviations of units: "Wb/m2" or "webers per square meter," not "webers/m2." Spell units when they appear in text: "...a few henries," not "...a few H."

- Use a zero before decimal points: "0.25," not ".25."

C. Equations

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

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

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

Note that the equation is centered using a center tab stop. Be sure that the symbols in your equation have been defined before or immediately following the equation. Use "(1)," not "Eq. (1)" or "equation (1),"

except at the beginning of a sentence: "Equation (1) is ..."

D. Some Common Mistakes

- The word "data" is plural, not singular.

- The subscript for the permeability of vacuum μ_0, and other common scientific constants, is zero with subscript formatting, not a lowercase letter "o."

- In American English, commas, semi-/colons, periods, question and exclamation marks are located within quotation marks only when a complete thought or name is cited, such as a title or full quotation. When quotation marks are used, instead of a bold or italic typeface, to highlight a word or phrase, punctuation should appear outside of the quotation marks. A parenthetical phrase or statement at the end of a sentence is punctuated outside of the closing parenthesis (like this). (A parenthetical sentence is punctuated within the parentheses.)

- A graph within a graph is an "inset," not an "insert." The word alternatively is preferred to the word "alternately" (unless you really mean something that alternates).

- Do not use the word "essentially" to mean "approximately" or "effectively."

- In your paper title, if the words "that uses" can accurately replace the word using, capitalize the "u"; if not, keep using lower-cased.

- Be aware of the different meanings of the homophones "affect" and "effect," "complement" and "compliment," "discreet" and "discrete," "principal" and "principle."

- Do not confuse "imply" and "infer."

- The prefix "non" is not a word; it should be joined to the word it modifies, usually without a hyphen.

- There is no period after the "et" in the Latin abbreviation "et al."

- The abbreviation "i.e." means "that is," and the abbreviation "e.g." means "for example."

IV. USING THE TEMPLATE

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

IJNMT_firstAuthorName_paperTitle.

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

A. Authors and Affiliations

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

B. Identify the Headings

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

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

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

C. Figures and Tables

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

TABLE I. TABLE STYLES

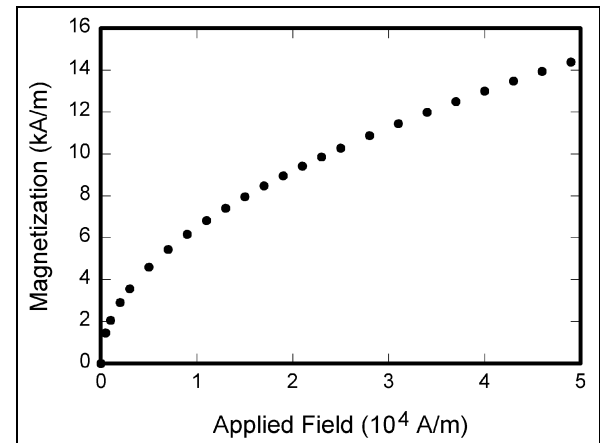| Table Head	Table Column Head		
	Table column subhead	Subhead	Subhead
copy	More table copy		

Fig. 1. Example of a figure caption

V. Conclusion

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

Appendix

Appendixes, if needed, appear before the acknowledgment.

Acknowledgment

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

References

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

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

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

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

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

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

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

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

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

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

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