9772723437005

Telkom University
Jl. Telekomunikasi 1 Terusan Buah Batu, Bandung, Provinsi Jawa Barat, 40257
Email: journals@telkomuniversity.ac.id
Web: journals.telkomuniversity.ac.id/complete

COMPLETE | Journal of Computer, Electronic, and Telecommunication | Vol. 6 No. 2 December 2025

COMPLETE

Journal of Computer, Electronic, and Telecommunication

Vol. 6 No. 2 December 2025

JURNAL

KEMAJUAN TEKNIK KOMPUTER, ELEKTRO, & TELEKOMUNIKASI

Telkom University

ISSN: 2723-4371 E-ISSN: 2723-5912

Website: journals.telkomuniversity.ac.id/complete | Email: journals@telkomuniversity.ac.id

Journal of Computer, Electronic, and Telecommunication
Telkom University
Ketintang Str. No. 156, Surabaya, 60231, East Java, Indonesia
E-mail: completesby@telkomuniversity.ac.id;
Web: journal.ittelkom-sby.ac.id/complete

TABLE OF CONTENTS

(This page is intentionally left blank)

ABOUT THE JOURNAL

COMPLETE: Journal of Computer, Electronic, and Telecommunication is an official journal of Telkom University. It publishes research or review articles in Computer, Electronic, and Telecommunication technology. This journal provides a platform for research lecturers, reviewers, practitioners, industry, and observers across Indonesia and overseas to promote, share, and discuss new issues and technology development. The scope of the journal includes:

- Technology utilization of maritime resources
- Strengthening infrastructure maritime
- Technology and management safety transportation
- Industrial strengthening technology transportation
- Supporting infrastructure and transportation system
- Operational efficiency
- Electronics Technology
- Telecommunication Technology
- Computer Technology
- System Security
- Advanced Robotics Technology
- Technology and disaster management
- Advanced Power Electronics
- Application of Power System
- Renewable Energy
- Chips Technology
- Sensor Technology
- Smart IoT Devices
- 5G Technology and Ecosystems
- Technology and management environment

COMPLETE published twice a year. Editors invite research lecturers, the reviewer, practitioners, industry, and observers to contribute to this journal. The language used in the form of Indonesian and English. All accepted manuscripts will receive individual digital object identifier (DOI) and indexed by Sinta, Arjuna, and Google Scholar. The online PDF version of the journal is open access from journal.ittelkom-sby.ac.id/complete. Subscription of the hard copy can be requested by email to completesby@telkomuniversity.ac.id.

(This page is intentionally left blank)

PREFACE

Welcome to the Journal of Computer, Electronic, and Telecommunication, Vol. 6, Issue No. 2. It is my privilege and pleasure to present the sixth volume of this peer-reviewed journal under Telkom University. This journal aims to accommodate the results of research publications through national and worldwide journals as part of continuous improvement.

As the chairman of COMPLETE, I would like to thank many people who supported this journal, especially Research and Community Service Units (LPPM). Furthermore, as the editor-in-chief, I would like to extend my sincere thanks to all members of the editorial and the advisory boards from Telkom University, whose service, dedication, and commitment have made the creation of this journal possible. We work together to improve the quality and excellence of articles published continuously.

We hope that COMPLETE could deliver valuable and interesting information and stimulate further research to the worldwide telecommunications, electrical, and computer engineering communities.

Surabaya, December 2025

Editor-In-Chief of Journal

(This page is intentionally left blank)

COMPLETE
Journal of Computer, Electronic, and Telecommunication
ISSN 2723-4371, E-ISSN 2723-5912

Article

Fan Speed Level Control Using Three-Language Voice Commands Based on YAMNet Audio Classification in *Deep Learning*

Giga Razki Arianda[1], Heri Pratikno[2*], Musayyanah[3], Pauladie Susanto[4]

[1,2,,4] Computer Engineering, Faculty of Technology and Informatics, Universitas Dinamika, Surabaya, Indonesia
[3] Politeknik Elektronika Negeri Surabaya;
* Correspondence: heri@dinamika.ac.id

Received: 28 October 2025; Revised: 26 December 2025; Accepted: 30 December 2025; Published: 31 December 2025

Abstract: The process of interaction between humans, computers, and electronic equipment can now be made more interactive, natural, and intuitive. In several previous studies, this interactive process was carried out through sensors or detection of finger gestures using computer vision based on MediaPipe. In this research, we designed and built a system that can control the fan rotation speed level using voice commands from three languages, namely Indonesian, English, and Javanese in real time through an audio classification process with YAMNet. The research results in the training process with 15 epochs had 100% accuracy, loss 0.46, ROC curve class 0 (fan off) was 100%, class 1 (low rotation fan) was 100%, class 2 (medium rotation fan) was 99%, and class 3 (high rotation fan) was 100%. Meanwhile, the results of testing the subset test dataset model using 15 epochs for all commands produced a percentage value of 97.5%.

Keywords: artificial intelligence; audio classification; arduino uno; deep learning; voice recognition

---

1. Introduction

The process of interaction between humans and computers, machines, and electronic devices was initially invasive or required physical contact. However, with the advancement of technology, this interaction process no longer requires physical contact or is non-invasive thanks to the support of various sensors, including motion sensors [1], sensors [2], sound sensors [3], and so on. The support of these various sensors enables the interaction process between humans and existing electronic equipment to be carried out automatically and more responsively, but it is not yet natural.

In study [4], a simulation of an automatic fan control device was created using a fuzzy system and a web-based DHT11 temperature sensor. With the massive development of new technologies to assist and facilitate people's daily work in various fields of life, this can now be done by applying soft computing technology in computer vision. The application of human-computer interaction technology in computer vision using the MediaPipe framework [5] through the process of detecting hand gestures to turn a fan on and off locally has been achieved. Computer vision research on lip reading for Indonesian word recognition: a viseme-based approach to assist people with hearing disabilities [6].

The application of a method that can be used to reduce the feature dimensions of data is Principal Component Analysis (PCA), and the Computed Input Weight Extreme Learning Machine (CIW-ELM) classification method [7] is used to classify simple activities [8] such as walking, climbing stairs, descending stairs, sitting, standing, and lying down. Another study [9] related to a system for detecting the shapes of both the right and left hand finger gestures to automatically control two fans in a local area also uses MediaPipe.

In further research [10], it is possible to control the operation of a fan remotely through computer vision by detecting the shape of the right hand gesture using MediaPipe with the application of Internet of Things (IoT) technology based on the MQTT (Message Queue Telemetry Transport) protocol. Next, the application of MediaPipe and CNN (Convolutional Neural Network) for finger gesture detection can be used as a counting medium for early childhood education [11]. Conventional control of electronic devices through voice commands using the Arduino Uno R3 microcontroller has also been carried out [12]. Real-time classification of gunshots, glass breaking, and speech using YAMNet on edge computing networks [13].

Prior research on fan speed control employed hand gesture recognition using MediaPipe with the fan's three factory-installed speed settings. This study advances the field by developing an AI-driven system that controls fan speed through multilingual voice commands (Indonesian, English, and Javanese) using real-time audio classification via YAMNet in a deep learning environment. A key distinction from previous work is the novel speed control mechanism: rather than relying on preset buttons, this system achieves variable speed control through calculated AC voltage division using capacitor capacitance, with validation via AVOM measurements.

2. Materials and Methods

The deep learning intelligent system computation in this research was programmed on the Arduino Uno IDE interface platform, Visual Studio Code program, and Jupyter using the Python programming language. Figure 1 shows that the input data comes from a laptop microphone to recognize the user's voice. The input is then processed by a computer using audio classification with the YAMNet model. The output is sent to the Arduino Uno microcontroller to perform actions on the relay by turning off or adjusting the fan speed based on the voice detection process in the three trained languages.

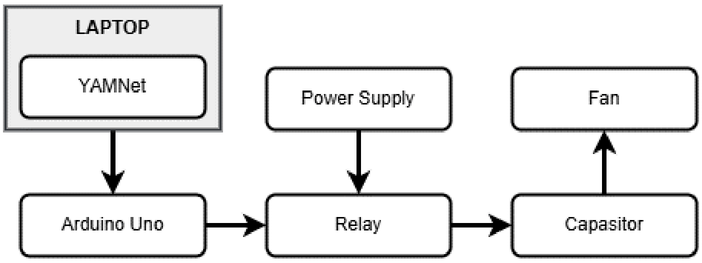


Figure 1. System block diagram

The schematic diagram of the system is shown in Figure 2. The hardware design process involves selecting specifications and adjusting various elements, such as sensors, microcontrollers, and other supporting components. Each part of the hardware is carefully arranged and connected to ensure synergistic and efficient work. The main objective of this hardware design model is to support the implementation of experiments and data collection accurately and consistently in this study.

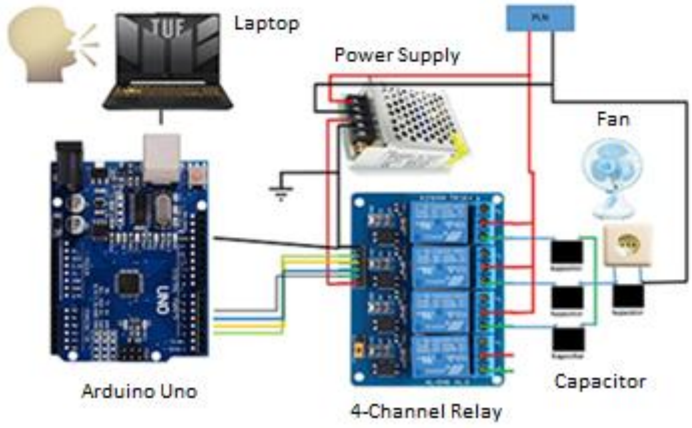


Figure 2. Schematic diagram of the system

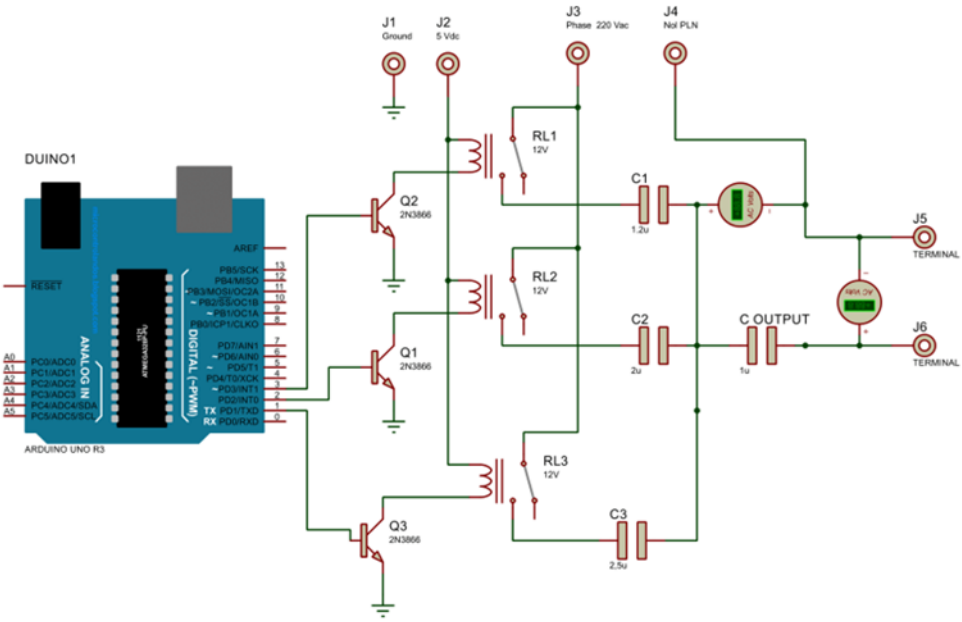


Figure 3. Schematic diagram

From the circuit schematic in Figure 3, there are several relay channels connected to the Arduino Uno microcontroller port. Each relay is connected to a capacitor with a different capacitance value to regulate the fan speed level. Channel one has a capacitor with a capacitance of 1.2 uF, channel two has a capacitor of 2 uF, and channel three has a capacitor of 2.5 uF. These capacitors will later be connected in series with a 1 uF output capacitor. The function of the output capacitor is to act as a voltage divider for the voltage supplied to the fan. A voltmeter is used to monitor the voltage across the capacitors connected in series. To determine the capacitive reactance of each capacitor, use equation (1). The formula for capacitive reactance is Xc.

$$Xc = \frac{1}{2 \times \pi \times f \times C} \tag{1}$$

2.1. Installation Environment

The environment installation stage is divided into two parts consisting of several *libraries* used in the sound classification process, such as *TensorFlow Hub*, *soundfile*, *numpy*, *pandas*, and other *libraries* needed during the dataset training process. *TensorFlow Hub* is one of *the libraries* from *TensorFlow* that provides various models that can be accessed and used more easily. In this study, *the* main *library* contains the YAMNet model, which can be called up via *the link "https://tfhub.*dev/google/yamnet/1". *TensorFlow Hub* facilitates access to *pre-trained* models that have been uploaded to the *platform*. By utilizing YAMNet via *TensorFlow Hub.* This research can quickly and efficiently integrate advanced audio classification capabilities into the system being developed. *The Serial library* is used to connect to the Arduino Uno microcontroller from Python.

In this study, the early stopping method was applied to the training process of the YAMNet model for fan command classification based on audio signals. Early stopping is a regularization technique used in the training process of deep learning models to prevent overfitting. This method works by automatically stopping the training process when the model's performance on the validation data no longer shows significant improvement after a certain number of epochs. Thus, the model is not forced to continue learning until it reaches a predetermined maximum number of epochs.

2.2 Dataset Flowchart

In this study, the author created *an audio dataset* based on commands used to control a fan using three languages, namely: Indonesian, English, and Javanese. Figure 4 below shows *the flowchart* of the stages of *the* audio *dataset* creation process. The first step is to collect audio data recordings in the form of command words desired in this study. Then, labels are given to each audio data and initial processing of the audio data is carried out before use, such as changing the audio format, such as WAV or MP3, normalizing *the volume*, or removing background *noise*.

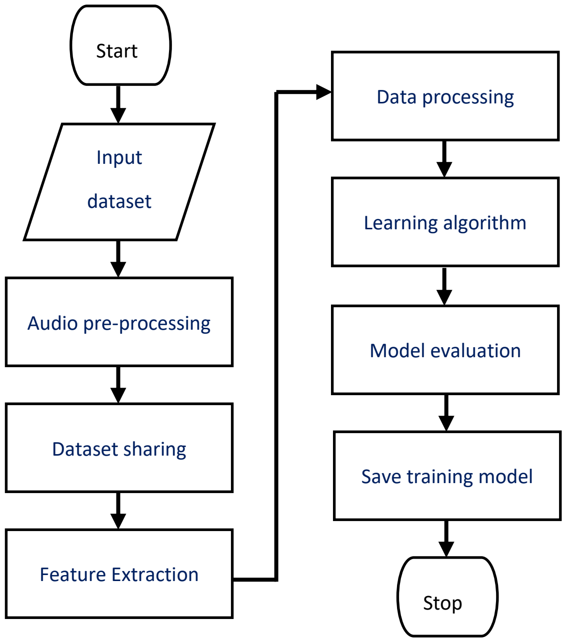


Figure 4. Flowchart of audio dataset

The audio dataset will be divided into several parts, including a training set, a validation set, and a test set. Next, using the processed training data, during the training process, the model will learn to recognize patterns in the sound data. The validation stage is used to set the data model parameters, such as the number of layers in the neural network. The test set is used to evaluate the model's ability to calculate matrices such as accuracy, precision, recall, and F1-score, which are parameters used to measure how well the model performs in voice recognition. After modeling, if the data model performs well, it can be used for voice recognition on new data.

2.3 Dataset Collection

The audio dataset used in this study consists of human voices uttering commands in three different languages. To turn off the fan in all three languages, the same root keyword "fan" is used, while to adjust the fan speed in all three languages, the same root keyword "fan button" is used, as shown in Table 1. The collected dataset consists of 270 audio files from 18 people, with four commands from each language and five repetitions of each command.

Table 1. Multilingual voice commands for fan speed control system

Language	Fan Off	Speed 1 (Slow)	Speed 2 (Medium)	Speed 3 (High)
Indonesian	kipas mati	kipas tombol satu	kipas tombol dua	kipas tombol tiga
English	kipas off	kipas tombol one	kipas tombol two	kipas tombol three
Javanese	kipas pejah	kipas tombol siji	kipas tombol loro	kipas tombol telu

2.4 Dataset Augmentation

In this study, the researchers developed a strategy to increase the amount of data to improve the performance of the system, namely: using data augmentation techniques to enlarge the data set by applying high pass filters, medium pass filters, and low pass filters on the collected sound samples. Augmentation with high pass, medium pass, and low pass channels provides frequency variations in the sound dataset, unlike low pass filters which ensure constant frequencies, high pass filters are used to ensure consistent frequencies in the air. Medium pass channels create a middle ground between the two. This aims to make the model more adaptive to acoustic environment variations that may occur in everyday life.

By applying sound information augmentation through high pass, medium pass, and low pass channels, the author seeks to improve the robustness and generalization of the model to variations that may occur in different environmental conditions. This composition is expected to help the model better recognize and understand various sound variants in the sound classification process. The final result of the data augmentation process is a total collection of 720 files from 4 commands in 3 predetermined languages, as shown in Figure 5.

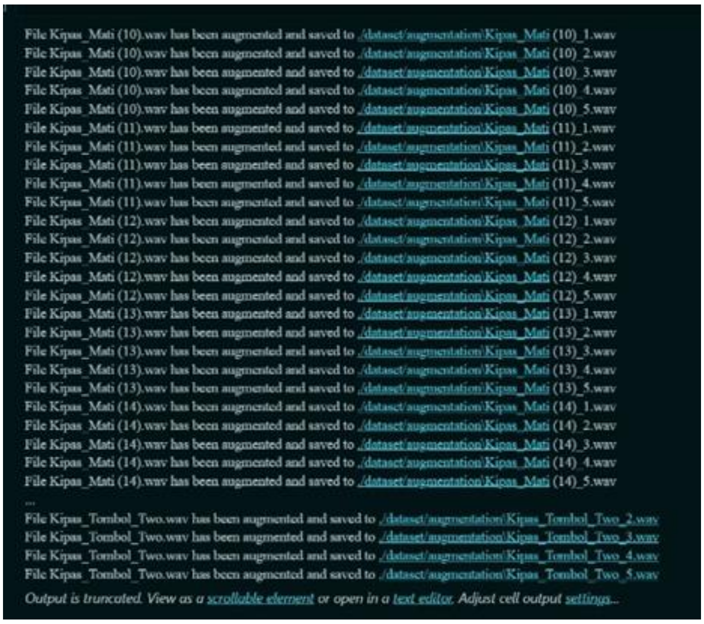


Figure 5. Dataset augmentation process

2.5 Audio Characteristics

At this stage, it is crucial to consider the characteristics of the audio being used, which requires visual monitoring using *waveforms* and *spectrograms* involving visual representations of audio signals. This is essential for a deep understanding of the unique features contained in the audio data. Through *waveform* visualization, monitoring of amplitude fluctuations over time can be done clearly, providing an overview of basic dynamics, including duration, tempo, and amplitude variations that may affect overall audio characteristics, as shown in Figure 6.

Meanwhile, spectrograms provide further information by presenting the distribution of frequency energy in the audio signal over time. Using spectrum frequency this, monitoring changes in frequency, *pitch*, and harmonic elements can be achieved. The combination of visual *waveform* and spectrogram creates a comprehensive approach to analyzing and understanding audio characteristics. This stage plays a key role in data preparation, enabling this research to recognize underlying patterns in audio data before entering the *encoding* and model development stages.

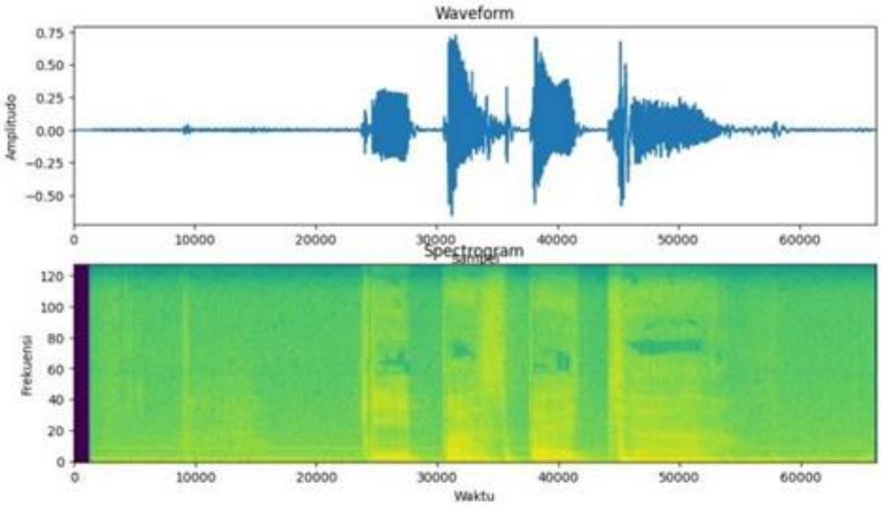


Figure 6. Waveform and spectrogram

2.6 Data Clustering

It is necessary to group the data into four categories of fan speed, namely off, low, medium, and high. In each category *folder*, the audio data is then grouped based on language. The researcher created *subfolders* for each related language, namely Indonesian, English, and Javanese, with the aim of creating a more detailed and organized structure, facilitating the management of the dataset for each fan speed *level* using the language used in the spoken commands, as shown in Figure 7.

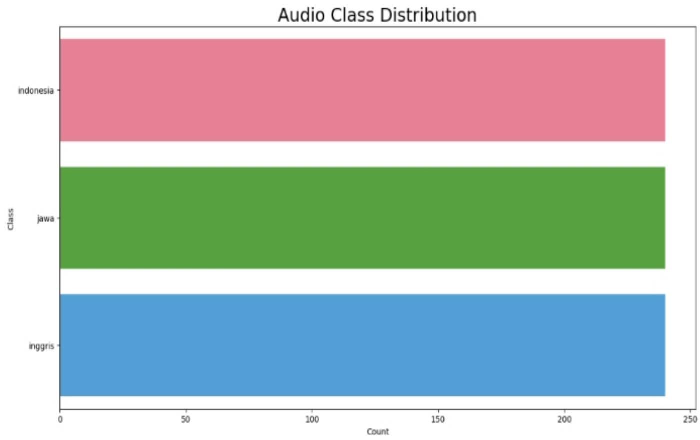


Figure 7. Data grouping based on language

2.7 Data Category Distribution

The next step is to distribute the dataset into five *folders* named *folds*, namely *folds* 1, *folds* 2, *folds* 3, *folds* 4, and *folds* 5. This distribution is carried out according to *the datasheet* that has been *randomly* generated by the program. The main purpose of distributing the data into *folds* is to ensure that each *fold* contains a balanced representation of various languages and categories of fan speed commands, as shown in Figure 8.

In each *fold*, the dataset will consist of *the categories* "off," "low," "medium," and "high." Each of these categories represents a different level of sound intensity. This grouping is done to ensure that each *fold* covers *the range of* fan speeds that may be encountered in practical use. By distributing the dataset proportionally across *folds*, researchers hope to achieve better model generalization and be able to recognize a wider variety of sounds. This process is a very important part of dataset preparation before entering the model training and evaluation stage.

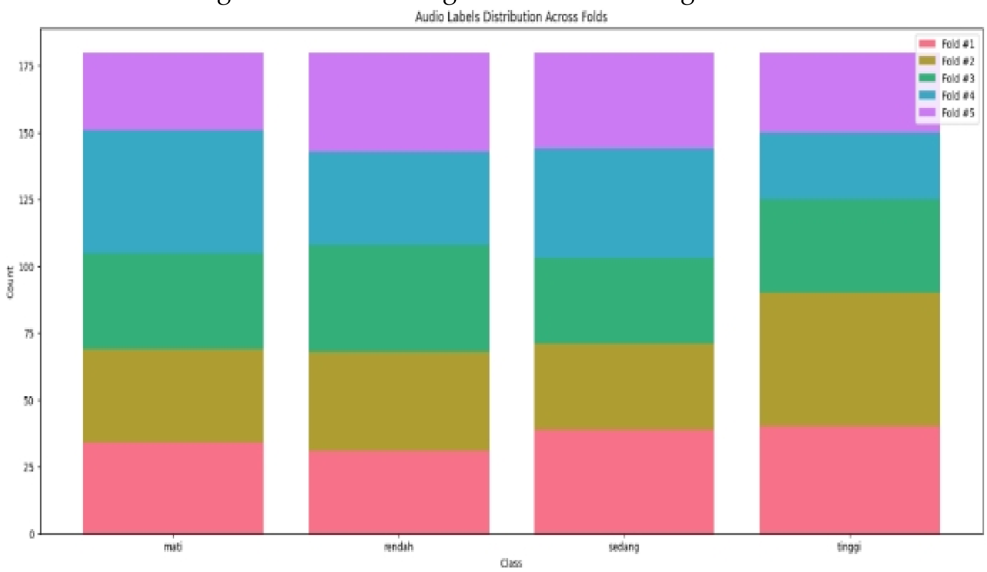


Figure 8. Data distribution based on category

2.8 Folds Subset Distribution

Folds are a *cross-validation* technique used to evaluate the performance of machine learning models. In *folds*, the data is divided into several parts called *folds*, as shown in Figure 9. The model is then trained on several *folds* and tested on the remaining ones. This process is repeated several times with each *fold* being used as *a* test *fold* once. The *folds* technique has several advantages over other *cross-validation* techniques, namely:

- More accurate in estimating model performance on unknown data
- More stable because it does not depend on the order of data used for training

In the YAMNet model, *folds* are used to overcome *overfitting*. *Overfitting* is a problem that occurs when the model is too well-suited to the training data that it cannot produce good results on unknown data. By using folds, the YAMNet model can be trained on larger data sets, thereby reducing the risk *of overfitting*.

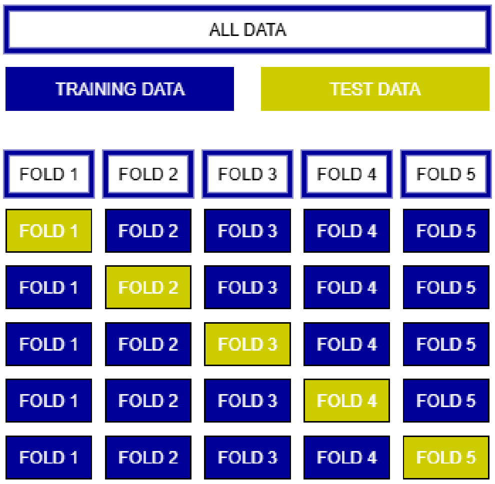


Figure 9. Dataset *folds*

The fold distribution stage divides the dataset into three main *subsets*: *train*, *test*, and *validation*, as shown in Figure 10. *Folds* 1 and *folds* 3 are selected to be part of *the train* data, while *folds* 4 and *folds* 5 are allocated as *validation* data. Finally, *folds* 2 are designated as *test* data. This division aims to optimize model training using *train* data, objectively test model performance using previously unused *test* data, and validate model results on *validation* data to prevent *overfitting.* Each *subset* has a different number of datasets. *The train subset* has 287 audio datasets, *the validation subset* has 279 audio datasets, and *the test subset* has 154 audio datasets, so that from a total of 3 *subsets* there are 720 audio datasets, as shown in Figure 11.

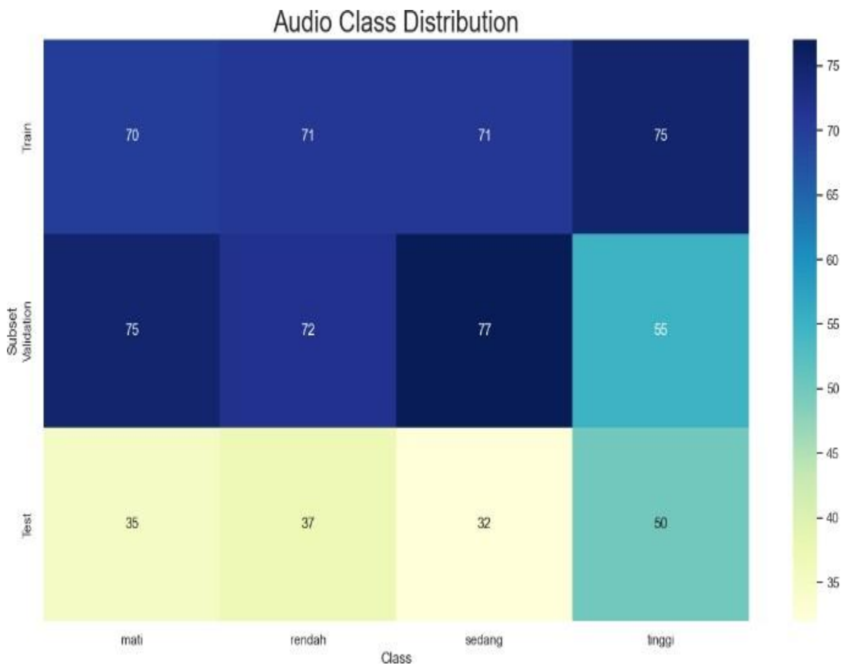


Figure 10. Distribution of *subset folds*

Dataset Split Sizes

	Subset	Number of Samples
0	Train	287
1	Validation	279
2	Test	154
3	Total	720

Figure 11. Subset dataset table

2.9 Label Encoding

The next step is to perform *encoding* on the audio labels, which aims to convert the audio *class* labels into numerical representations in three *subsets* consisting of *train*, *test*, and *validation*. This *encoding* is a crucial stage in audio data processing, with the main objective being to convert audio *class* information into numerical values that can be understood by machine learning models. This process is performed on each data *subset*, including *train* to train the model, *test* to test the model's performance, and *validation* to validate the learning results, the label *encoding* process is shown in Figure 12.

```
# Get label encoded audio classes
train_labels_encoded = train_df["classID"].to_numpy()
val_labels_encoded = val_df["classID"].to_numpy()
test_labels_encoded = test_df["classID"].to_numpy()

# Inspect label encoded targets
train_labels_encoded
[23]  ✓ 0.0s

... array([0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,
       2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,
       2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,
       2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 3, 3, 3, 3,
       3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3,
       3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3,
       3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3,
       3], dtype=int64)
```

Figure 12. Label *encoding*

```
# One-hot encode train labels
train_one_hot = one_hot_encode(train_labels_encoded, num_classes)

# One-hot encode validation labels
val_one_hot = one_hot_encode(val_labels_encoded, num_classes)

# One-hot encode test labels
test_one_hot = one_hot_encode(test_labels_encoded, num_classes)

# View one-hot encodings of train labels
train_one_hot
[26]  ✓ 0.0s

... <tf.Tensor: shape=(287, 4), dtype=float32, numpy=
    array([[1., 0., 0., 0.],
           [1., 0., 0., 0.],
           [1., 0., 0., 0.],
           ...,
           [0., 0., 0., 1.],
           [0., 0., 0., 1.],
           [0., 0., 0., 1.]], dtype=float32)>
```

Figure 13. Label one-hot encoding

2.10 Label One-Hot Encoding

Still in the *class encoding* stage, but at this stage it is done using the *one-hot encoding* method as shown in Figure 13 above. The purpose *of one-hot encoding* is to translate data labels into binary vectors. This method focuses on simplifying class representation in the context of *classification* tasks. When involving audio data in *the train, test*, and *validation subsets,* each *class* is represented by a binary vector with a length corresponding to the number of *classes*. In this context, vector elements have a value of one indicating the *class* represented, while other elements have a value of zero.

Comparison of the *one-hot encoding* method, *Gower distance* combined with the *k-means*, DBSCAN, and OPTICS algorithms, and *k-prototype* for mixed-type data clustering. The dataset used in this study [14] is a chronic kidney disease (CKD) dataset sourced from UCI. *Machine Learning Repository*. Based on evaluation using the silhouette index, it is known that *k-prototype* with a *cluster* count of k=2 is the most optimal *clustering* method because it provides the highest silhouette index value compared to the other four methods. When linked to the labels in the dataset, the *clustering* results provide an accuracy value of 81.25 percent.

2.11 BiLSTM

Deep Learning using LSTM in this study [15] was used to detect the Indonesian Sign Language System (SIBI) as a means of communication with the deaf. Research [16] applied LSTM to recognize finger gestures as a means of input for interaction between humans and computers in assessing the level of satisfaction with public customer service using *computer vision*. *Bidirectional Long Short- Term Memory* has two LSTM networks, the first of which functions to process input data sequences in *a forward direction,* and the second LSTM network functions to process data sequences in the opposite direction (*backward)*. *The output* from the *forward* and *backward* LSTM networks is combined at each time sequence. With these two opposing *layers*, the model can learn past and future information for each *input sequence* [17]. The BiLSTM architecture is shown in Figure 14 [18].

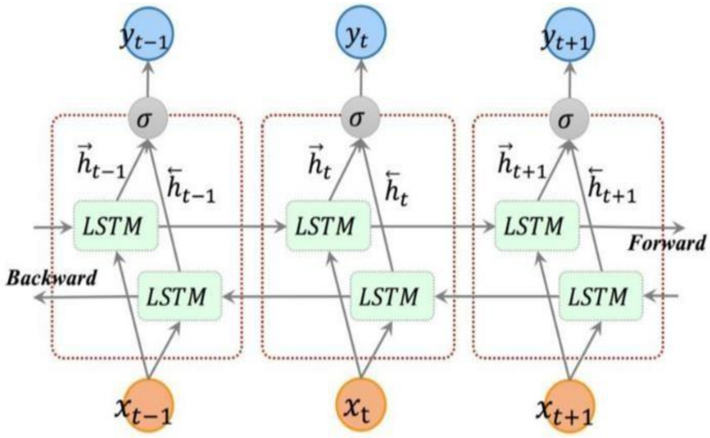


Figure 14. BiLSTM Architecture

2.12 YAMNet Model

After preparing the dataset for training the model, the main focus is on the *layer* structure to be used. This process involves the use of *an* audio *generator*, a tool used to generate *embedding* and label pairs from audio files by providing paths and labels that have been *encoded* using *one-hot encoding*. Next, an input pipeline is built using the *tf.data* API by utilizing paths and labels for audio data. This pipeline includes steps such as defining dataset output tags, creating a *frame* dataset with *a generator,* applying randomization based on certain conditions, enforcing batching, and applying *caching* and *prefetching* based on conditions.

The next step is to divide it into relevant *subsets* for training, testing, and validation. This division is crucial for accurately testing the model's performance. After the dataset is divided, the next stage involves analyzing the sample and label forms to ensure compatibility with the YAMNet audio classification model. Model initialization begins by defining *the* kernel *initializer* to be used. Then, *the* input *layer* is formed with *a shape of* 4,1024 and the data type used is *tf.float32* in *the* LSTM *layer*. A *bidirectional* approach with 128 *initializers* is used to improve the model's ability to capture the temporal context of the audio data.

At this stage, *a dropout* of 0.25 is applied to *the* audio LSTM *layer*. *Dropout* is used to reduce *overfitting* by randomly ignoring some units during training. This helps improve the model's generalization to previously unseen data. The use of *dropout* in *the* audio LSTM *layer* aims to optimize the overall performance of the model with the model structure shown in Figure 15. The final step is to calculate the classification probability in the model. This process allows the model to provide predictions for the categories that have been identified during the training process. In this way, it is possible to evaluate the extent to which the model is able to understand and recognize unique audio patterns. The YAMNet model with the designed program involves the use of *a classifier* model in H5 format, which is the result of dataset training.

Provides voice *input* through the program a pre-designed voice recording program. When the program detects voice *input* from the user, it saves the data to *a file* newly in *.wav* format with *a sample rate* of 16,000 Hz. Once the user is detected, the program will predict important categories based on the trained model. For example, if the user matches the patterns "fan off," "fan off," and "fan broken," the model will predict the label "off." Similarly, for the patterns "fan button one," "fan button one," and "fan button siji," the model will predict the label "low." Meanwhile, for the patterns "fan button two," "fan button two," and "fan button loro," the model will predict the label "medium." If the patterns are "fan button *three*," "fan button three," and "fan button telu," the model will predict the label "high."

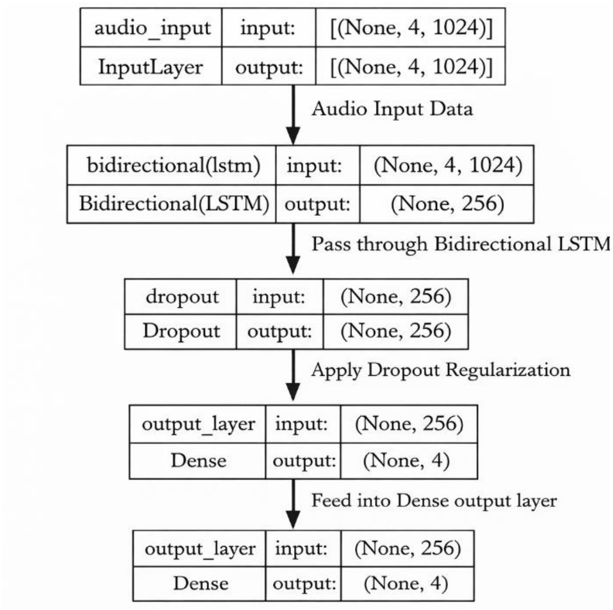


Figure 15. *Dropout* LSTM model structure The user

2.13 Fan Control Algorithm

An overview of the system workflow can be seen in Figure 16, starting with the use of a microphone as input to detect sounds in three languages with twelve different commands. The detected sounds include instructions such as "fan off," "fan button one," "fan button two," "fan button three," "fan pejah," "fan button one", "fan button two", "fan button three", "fan *off*", "fan button *one*", "fan button *two*", and "fan button *three*". Next, the detected sound will be processed through a grouping step to categorize it into the labels "Off," "Low," "Medium," and "High." The results of the sound recognition will be displayed as detected command sentences along with their category labels.

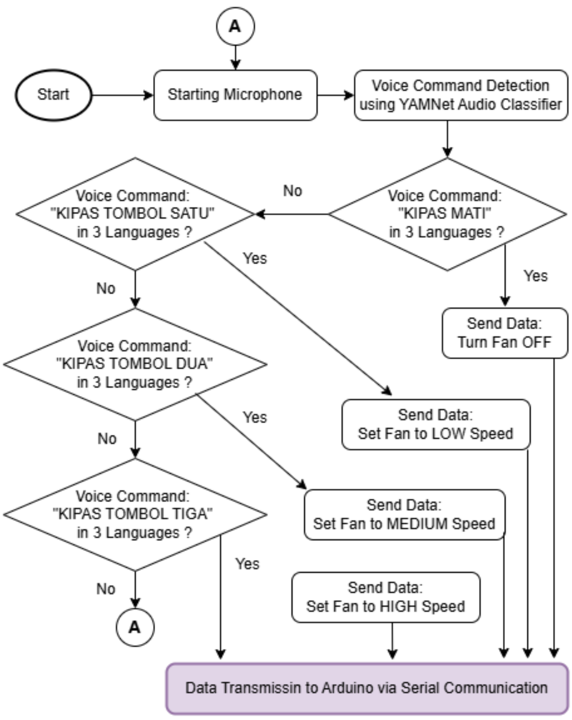


Figure 16. Python *Flowchart*

After processing the data in the Python program using audio classification, the recognition results will be sent to Arduino Uno via serial communication, as shown in Figure 17. On Arduino Uno, the received data will be checked to ensure that the labels generated match the user's commands. If the label received is "*off,"* all *relays* will be *turned off*, so the fan will be turned off. If the label received is "low," *relay 1* will be activated, so the fan will operate at level one or low speed.

For the label "medium", *relay* 2 will be activated, so that the fan is at level two or medium. Finally, if the label received is "High", then *relay* 3 will be activated, so that the fan operates at a speed of level three or high. Next, the system will return to the voice recording application to start the next voice recognition cycle. With the proper integration between the Python workflow and Arduino Uno, the system can provide an effective response according to the voice commands given by the user.

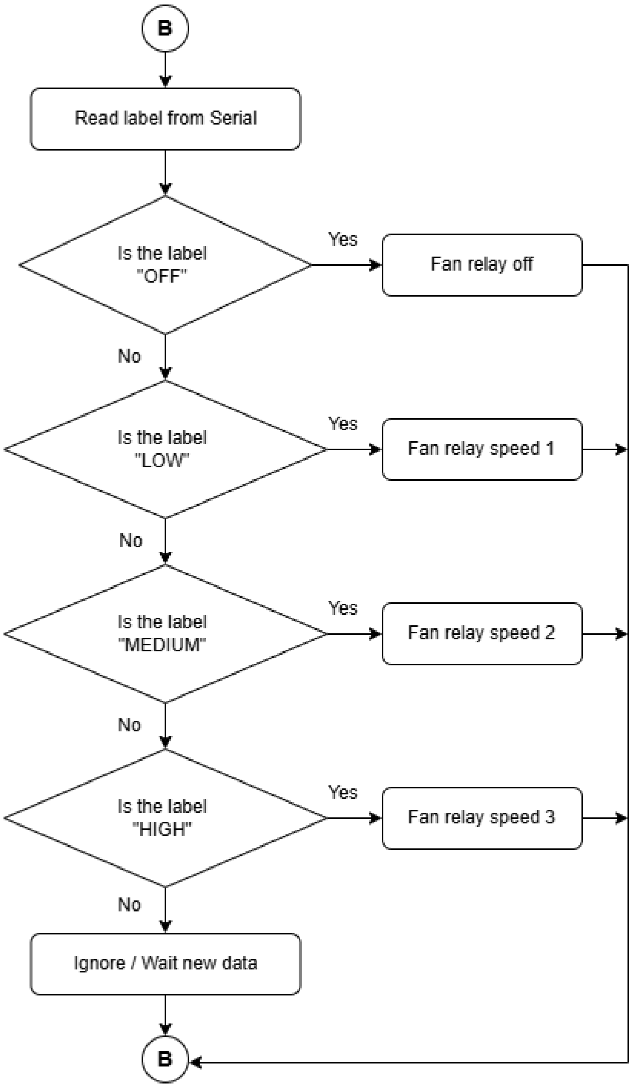


Figure 17. Arduino Uno *Flowchart*

3. Results and Discussion

In this study, to observe the performance of the training process, test parameters such as accuracy percentage and loss value were used, which were obtained from the analysis of the confusion matrix, ROC Curve, classification report, and performance metrics (*accuracy, precision, recall, F1-Score*). All of the above parameters were applied to the three different model training processses, namely: model training with 12 epochs, 15 epochs, and early stopping model training.

3.1. Comparison of Test Results Between Epochs

Based on the test results conducted using 12 *epochs*, 15 *epochs*, and the *early stopping* method (training accuracy will automatically stop at a certain *epoch* because the system sees that the accuracy is already at its highest, close to 100%, so the next *epoch* computation process is no longer necessary). When applying the *early stopping* method, the YAMNet model tends to classify the sound of a fan that is turned off as the sound of a fan that is turned on. This occurs because the *early stopping* model has already achieved a high accuracy percentage at *the* fifth *epoch*, resulting in insufficient knowledge. In general, the training process with 15 *epochs* provides better accuracy for all fan commands, except for the "off" and ″stop" commands, *the* 12th *epoch* provides better accuracy.

Based on Table 2, the true accuracy for all fan commands increases with the number of *epochs*, but the true accuracy for the "off" and "stop" commands in the *early-stopping epoch* actually decreases compared to training using 15 *epochs*. This shows that the YAMNet model in the *early-stopping* model more often classifies the sound of a "dead" fan as the sound of a "live" fan. A possible cause is that the YAMNet model is overtrained on the training data, making it too sensitive to the patterns in the training data. As a result, the YAMNet model finds it more difficult to classify fan sounds that are different from the training data.

Table 2. Comparison of user accuracy contained in the dataset between epochs

Command Word	12 *Epoch*		15 *Epoch*		*Early Stopping*	
	Accuracy (%)					
	True	False	True	False	True	False
Kipas mati Kipas *off* Kipas pejah	10	90	40	60	53,3	46,7
Kipastombol satu Kipas tombol *one* Kipas tombol siji	30	70	43,3	46,7	43,3	56,7
Kipas tombol dua Kipas tombol *two* Kipas tombol loro	56,7	43,3	50	50	36,7	63,3
Kipas tombol tiga Kipas tombol *three* Kipas tombol telu	20	80	40	60	43,7	56,7

Table 3 presents a comparison of system accuracy in recognizing voice commands from users not included in the training dataset (unseen users) across three different scenarios: 12-epoch training, 15-epoch training, and early stopping. The evaluation was conducted on 12 voice commands in three languages (Indonesian, English, and Javanese), measuring True Positive and False Positive rates for each command. Results indicate that system performance varies across scenarios. In the 12-epoch training, the command "Kipas tombol two" achieved the highest accuracy (60% True, 40% False), while "Kipas off" had the lowest accuracy (43.3% True, 56.7% False). The 15-epoch training showed improvements in some commands, such as "Kipas off" increasing to 60% True, though some other commands experienced decreased performance. The early stopping scenario provided more balanced results, with accuracy ranging from 50% to 66.7% for the True category.

Overall, no single scenario consistently outperformed across all commands, indicating that the model exhibits varying sensitivity to the number of training epochs for different command classes. These findings suggest the need for further optimization in hyperparameter selection and training strategies to enhance model generalization to new users. The "off" and "stop" fan voice commands do not always have the same pattern because they can be influenced by various factors, including: the type of fan, the size of the fan, and environmental conditions. The YAMNet model, which is overtrained on the fan patterns found in the dataset, will find it more difficult to classify the "off" and "stop" fan commands that do not have the same patterns as those found in the dataset.

Table 3. Comparison of accuracy of users not in the dataset between epochs

Command Word	12 *Epoch*		15 *Epoch*		*Early Stopping*	
	Accuracy (%)					
	True	**False**	**True**	**False**	**True**	**False**
Kipas mati						
Kipas *off*	43,3	56,7	60	40	56,7	43,3
Kipas pejah						
Kipas tombol satu						
Kipas tombol *one*	40	60	60	40	66,7	33,3
Kipas tombol siji						
Kipas tombol dua						
Kipas tombol *two*	60	40	60,7	33,3	56,7	43,3
Kipas tombol loro						
Kipas tombol tiga						
Kipas tombol *three*	43,3	56,7	56,7	43,3	50	50
Kipas tombol telu						

3.2. Capasitor Voltage Measurement

To validate the theoretical circuit design, voltage values calculated using Equations (1) and (2) were compared with voltmeter measurements (Figure 18). This validation quantifies the accuracy of the capacitor-based voltage division method at three speed levels. Results are presented in Table 4 (individual capacitors) and Table 5 (series configuration). This verification confirms that the calculated capacitance values achieve the target voltage levels required for precise fan speed control, demonstrating the viability of continuous speed regulation versus discrete preset buttons.

$$\%error = \frac{|Avo - Teori|}{Teori} \times 100\% \quad (2)$$

(a) (b) (c)

Figure 18. Voltage measurement (a) low *level*, (b) medium *level* and (c) high level

Table 4. Capacitor voltage measurement after relay

No.	Capasitor	Experimental Result		Percentage Difference in Voltage (%)
		Voltmeter (VAC)	Theory (VAC)	
1	Xc_1 (1,2 *uF*)	112,2	100	12,2
2	Xc_2 (2 *uF*)	84,4	73	15
3	Xc_3 (2,5 *uF*)	71,5	61	17

Table 5. Series capacitor voltage measurement

No.	In Series		Experimental Result		Percentage Difference in Voltage (%)
	Input Capasitor	Output Capasitor	Voltmeter (VAC)	Theory (VAC)	
	Xc1 (1,2 *uF*)	1 *uF*	104	120	13
	Xc2 (2 *uF*)		131	146	10,3
	Xc3 (2,5 *uF*)		143	159	10,1

The percentage results of capacitor voltage measurements in a series circuit between the capacitor after *the relay* (*input* capacitor) and the 1 *uF* capacitor (*output* capacitor) show some differences between the measured values and the theoretically calculated values. For capacitor Xc1, the voltmeter measurement shows a voltage of 104 VAC, while the theoretical result should be around 120 VAC. The difference between these two values results in a percentage of 13%. Similarly, for capacitor Xc2, the measured voltage is 131 VAC, while the value obtained based on theoretical calculations is 146 VAC, with a percentage of 10.3%.

Similarly, for capacitor Xc3, the measured voltage was 143 VAC, while the value obtained based on theoretical calculations was 159 VAC, with a percentage of 10.1%. The difference between the measurement and calculation results indicates the possibility of certain factors that can affect the measured voltage value, such as imperfections in the device circuit, internal component resistance, or the actual condition of the capacitor. Thus, the experimental results provide important information for a better understanding of the system's response in real situations.

5. Conclusions

Based on the results and discussion of the empirical research conducted in the previous chapter, the following conclusions were drawn: the *training* process with 15 *epochs* had an accuracy of 100%, *a loss* of 0.46, ROC *curve class 0* (fan off) of 100%, *class 1* (low fan speed) of 100%, *class 2* (medium fan speed) of 99%, and *class 3* (high fan speed) of 100%, which is the highest compared to the *12-epoch training* process and the *early stopping* model. High fan speed produced *a classification report* with a *precision* value of 100%, a *recall* value of 100%, and an *F1-score* value of 100%.

The test process of the test dataset subset using the early stopping model, which was conducted 40 times on all commands, had an accuracy of 100%. Meanwhile, the test process using the early stopping model with live commands through a laptop microphone from 3 languages with the fan off (0) had an accuracy of 56.7%, low fan speed (1) had an accuracy of 66.7%, medium fan speed (2) had an accuracy of 56.7%, and high fan speed (3) had an accuracy of 50%. The results of testing with voice commands directly into the microphone did not reach 70% accuracy.

References

1. A. Hanafie, Kamal, and R. Ramadhan, "Perancangan Alat Pendeteksi Gerak Sebagai Sistem Keamanan Menggunakan ESP32 CAM Berbasis IoT," *J. Teknol. dan Komput.*, vol. 2, pp. 142–148, 2022, doi: 10.56923/jtek.v2i02.101.
2. A. A. Syukron and Isnaini Lilis Elviyanti, "Pembuatan Sensor Cahaya dengan Memanfaatkan LED dan LDR Berbasis Arduino Uno," *J. Kridatama Sains Dan Teknol.*, vol. 3, no. 02, pp. 161–169, 2021, doi: 10.53863/kst.v3i02.435.

3. Budy and T. Radillah, "Sistem Kontrol Menghidupkan Lampu Otomatis Menggunakan Sensor Suara FC-04 Berbasis Arduino Uno," *Indones. J. Comput. Sci.*, vol. 12, 2023, doi: 10.33022/ijcs.v12i1.3121.
4. N. K. Daulay, N. Lestari, and A. Armanto, "SIMULASI MONITORING PENGATUR KECEPATAN KIPAS ANGIN MENGGUNAKAN SISTEM FUZZY BERBASIS WEB," 2020, [Online]. Available: https://api.semanticscholar.org/CorpusID:225411605.
5. M. A. Fakhruddin, "TA : Sistem Deteksi Gestur Jari Tangan menggunakan Mediapipe dan Faster-RCNN untuk Mengontrol Kecepatan Kipas Angin," Universitas Dinamika, 2023.
6. H. Wicaksono, L. Liliana, and A. N. Tjondrowiguno, "Pemodelan Lip Reading Bahasa Indonesia Berbasis Visem Menggunakan VGG16 serta Jaro-Winkler Similarity dan Bigram," *J. Infra*, 2022, [Online]. Available: https://publication.petra.ac.id/index.php/teknik-informatika/article/view/12513%0Ahttps://publication.petra.ac.id/index.php/teknik-informatika/article/download/12513/10814.
7. M. Irwanto, F. Bachtiar, and N. Yudistira, "Klasifikasi Aktivitas Manusia Menggunakan Algoritme Computed Input Weight Extreme Learning Machine dengan Reduksi Dimensi Principal Component Analysis," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 9, p. 1195, 2022, doi: 10.25126/jtiik.2022965504.
8. F. D. Tanugraha, "TA : Sistem Pengenalan Aktivitas Manusia Menggunakan Long Short-Term Memory dan Mediapipe," Universitas Dinamika, 2022.
9. Y. R. B. Edowai, "TA : Sistem Automatic Feature Selection Berbasis Deteksi Gestur Kedua Jari Tangan untuk Mengontrol Level Kecepatan Putaran 2 Kipas Angin menggunakan Mediapipe," Universitas Dinamika, 2023.
10. F. Wakerkwa, "TA : Kontrol Level Kecepatan Putaran Kipas Angin melalui Deteksi Bentuk Gestur Jari Tangan Berbasis IoT," Universitas Dinamika, 2023.
11. M. R. P. Nautica, "TA : Hand Gesture Detection sebagai Alat Bantu Ajar Berhitung menggunakan Mediapipe dan Convolutional Neural Network secara Realtime," Universitas Dinamika, 2022.
12. A. A. Firmansyah, "Rancang Bangun Alat Bantu Penyandang Disabilitas Tangan Untuk Menghidupkan dan Mematikan Perangkat Elektronik Menggunakan Voice Recognition Module V3," *J. Telecommun. Netw. (Jurnal Jar. Telekomun.*, vol. 3, no. 2 SE-, pp. 47–52, Nov. 2016, doi: 10.33795/jartel.v3i2.220.
13. C. Malmberg, "Real-time Audio Classification on an Edge Device-Using YAMNet and TensorFlow Lite," 2021.
14. Z. Fadilah and A. W. Wijayanto, "Perbandingan Metode Klasterisasi Data Bertipe Campuran: One-Hot-Encoding, Gower Distance, dan K-Prototype Berdasarkan Akurasi (Studi Kasus: Chronic Kidney Disease Dataset)," *J. Appl. Informatics Comput.*, vol. 7, pp. 57–67, 2023, doi: 10.30871/jaic.v7i1.5857.
15. F. X. L. Riberu, "TA: Sistem Deteksi Simbol pada SIBI (Sistem Isyarat Bahasa Indonesia) secara Real Time menggunakan Mediapipe dan LSTM," Universitas Dinamika, 2023.
16. B. G. Permana, "TA : Sign Language Detection sebagai Alat Bantu Survey Pelayanaan Publik Menggunakan Long Short Term Memory Secara Realtime," Universitas Dinamika, 2023.
17. D. I. Puteri, "Implementasi Long Short Term Memory (LSTM) dan Bidirectional Long Short Term Memory (BiLSTM) Dalam Prediksi Harga Saham Syariah," *Euler J. Ilm. Mat. Sains dan Teknol.*, vol. 11, no. 1, pp. 35–43, 2023, doi: 10.34312/euler.v11i1.19791.
18. Z. Cui, R. Ke, Z. Pu, and Y. Wang, "Deep Bidirectional and Unidirectional LSTM Recurrent Neural Network for Network-wide Traffic Speed Prediction," pp. 1–11, 2018, [Online]. Available: http://arxiv.org/abs/1801.02143.

COMPLETE
Journal of Computer, Electronic, and Telecommunication
ISSN 2723-4371, E-ISSN 2723-5912

Article

SIM Puskesmas: An Integrated Management Information System for Administration, Doctor Consultations, and Web-Based Pharmacy Services at UPT Karang Anyer Using Research and Development (R&D) Approach

Nazwa Aidlia Octa Mevia[1], Yohana Kartika Marbun[2], Jonatan Rio Gultom[3], Debi Yandra Niska[4]*

[1-4] Department of Computer Science, Universitas Negeri Medan, Medan, Indonesia
* Correspondence: nazwaaidila9@gmail.com

Received: 16 November 2025; Revised: 8 December 2025; Accepted: 9 December 2025; Published: 31 December 2025

Abstract: Primary health care facilities such as Community Health Centers (Puskesmas) face inefficiencies due to manual data collection. This study aims to develop an integrated web-based Community Health Center Management Information System (SIM-PUSKESMAS) to improve service efficiency and data accuracy. Using the Research and Development (R&D) method with the Waterfall model, the system was designed using UML and implemented with PHP and MySQL. The novelty of this study lies in the seamless integration of Queue Management, SOAP-based Medical Records, and Pharmacy inventory in a single architecture, addressing the fragmentation issues found in previous systems. The system was tested using Black Box testing and User Acceptance Testing (UAT). The results showed a 100% functional success rate in Black Box testing. Furthermore, UAT results from 10 respondents indicated a user satisfaction score of 88.4% (Very Good), demonstrating that the system is highly effective. Comparison with manual systems shows that this digital system significantly reduces administrative time and improves clinical data accuracy. It is concluded that SIM-PUSKESMAS effectively ensures data integrity and supports better governance.

Keywords: Information Systems; Websites; Health services

---

1. Introduction

Community Health Centers (Puskesmas) are healthcare facilities that play a crucial role in Indonesia. These institutions provide comprehensive, integrated, and sustainable healthcare services to the community to improve overall public health. Furthermore, Puskesmas serve as the spearhead of the national healthcare system, focusing on promotive and preventive efforts, not just curative ones. Their strategic role makes Puskesmas a crucial institution in ensuring equitable, accessible, and high-quality healthcare for all levels of society [1], [2]. The government strives to maintain the consistent performance of community health centers (Puskesmas) with the aim of facilitating public access to healthcare services, including the management of patient data, employee data, and medication records used in the centers. However, in practice, data recording in several Puskesmas is still carried out manually, by writing information on paper sheets that are then stored in archive shelves or on laptops. This condition leads to significant delays when retrieving patient data or reports related to medication usage when needed [3], [4].

Digital transformation in the healthcare sector has become a strategic step in addressing challenges related to efficiency and service accuracy at primary healthcare facilities such as community health centers (Puskesmas). The use of web-based information systems enables every administrative and healthcare service process to be connected in real time, allowing data to be updated and accessed anytime without spatial or temporal limitations. Integration between work units from administrative services and medical consultations to pharmaceutical services creates a more effective and transparent workflow [5]. In addition, the implementation of a web-based system also supports centralized data management, minimizes the risk of information loss, and provides an accurate data foundation for managerial decision-making [6]. Thus, the implementation of an integrated information system at Puskesmas Karang Anyer is not only focused on improving operational efficiency, but also serves as a concrete step toward sustainable digitalization of healthcare services.

However, existing studies often focus on partial solutions. For instance, Marfalino et al. [7], [8] focused solely on geographical mapping, while other studies primarily addressed administrative data without integrating clinical records (SOAP) and pharmacy inventory in real-time. A significant research gap remains in developing a fully unified system that connects patient registration queues directly to medical records and pharmacy stock for rural healthcare centers. This study addresses this gap by proposing SIM-PUSKESMAS, which integrates these three core modules to ensure data continuity and minimize redundancy.

2. Materials and Methods

2.1 Research Approach

This study employs a Research and Development (R&D) approach with the aim of producing a web-based integrated information system capable of improving service efficiency and data management at Puskesmas Karang Anyer. This approach was chosen because it allows the researcher not only to conduct needs analysis, but also to design, implement, and test the system [7], [8]

2.2 Data collection

The data collection stage was carried out using three primary methods to ensure comprehensive requirement gathering:

1. Direct Observation: Conducted to observe the actual service flow and identify bottlenecks in the existing manual system [7], [8].
2. Semi-structured Interviews: Performed with key stakeholders, including administrative staff, doctors, and data managers, to explore specific system requirements and identify challenges in information management [7], [8].
3. Administrative Archive Documentation: Used as a secondary data source to verify the accuracy of the information obtained through observation and interviews. This triangulation ensures the data aligns with end-user needs [7], [8].

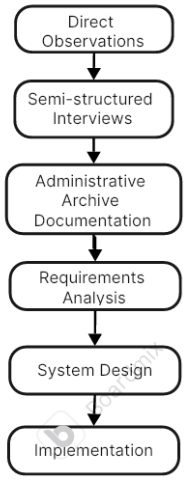


Figure 1. Research Stages Flowchart

Figure 1 illustrates the sequence of research stages, beginning with direct observations, semi-structured interviews, and the collection of administrative archive documentation as the initial data-gathering methods. These three methods are used to obtain preliminary information regarding service workflows and system needs at Puskesmas Karang Anyer. The collected data is then analyzed in the requirements analysis stage to identify the main problems and functional needs that the system must address. The results of this analysis serve as the basis for the system design process, which focuses on designing the structure, workflow, and user interface of the system. The final stage is implementation, where the design is transformed into an operational information system that can be utilized by users in accordance with real field requirements.

2.3 System Design

System design was carried out using an analysis and modeling approach with Unified Modeling Language (UML) as the main tool. UML was chosen due to its ability to visually and structurally represent object-oriented systems, making it easier to perform requirements analysis, architectural design, and communication among development team members. The system modeling includes several main diagrams as follows:

1. Use Case Diagram, used to define the interactions between actors (admin, doctor, pharmacist, and patient) and the main system functions, such as login, patient data management, medical consultations, prescription management, and report generation.
2. Class Diagram, used to model the relationships between core entities such as User, Patient, Doctor, Medicine, and Medical Record.
3. Sequence Diagram, It functions to illustrate the sequence of communication between objects in executing system processes, starting from the user request to the system’s generated response.
4. Activity Diagram, It represents the overall flow of activities, including login, data input, and report generation. This diagram is important for verifying process logic and identifying potential overlaps in activities [11], [12].

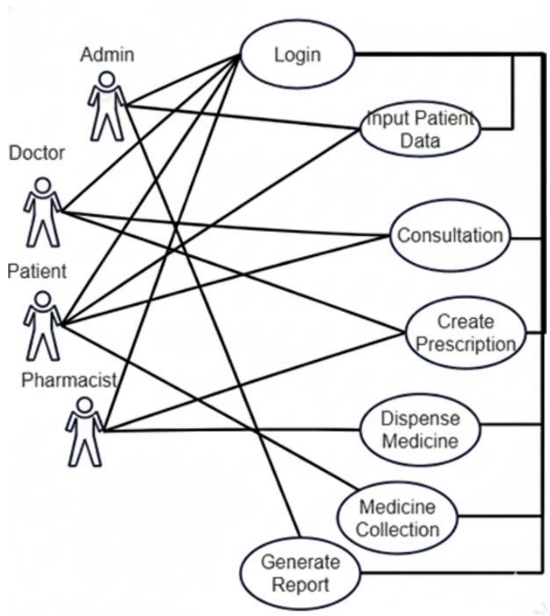


Figure 2. Use Case Diagram of the Integrated Puskesmas Information System

In Figure 2, the system workflow shows how the admin, doctor, pharmacist, and patient interact through the processes of login, patient data input, consultation, prescription creation, medicine preparation, medicine collection, and reporting. Each actor has different roles but remains interconnected, so that all medical service activities can be digitally coordinated and efficiently managed within one integrated system.

2.4 System Implementation

The implementation phase was carried out using a web-based application development approach with a combination of open-source technologies. The system was developed as a dynamic web-based application that supports real-time data management, is integrated, and easily accessible through the Puskesmas local network [13]. The programming languages and technologies used include:

1. PHP (Hypertext Preprocessor) is used as a server-side scripting language to manage system logic, handle authentication processes, and connect to the database.
2. MySQL serves as the Database Management System (DBMS) used to store data centrally and ensure the integrity and consistency of patient, doctor, and drug data.
3. HTML (HyperText Markup Language) s used as the main structure of the user interface.
4. CSS (Cascading Style Sheets) is used to style the interface, ensuring responsiveness and compliance with modern user interface design principles.
5. JavaScript is applied to enhance interactivity, perform input validation, and improve the overall user experience [14], [15]

All development was carried out within the XAMPP environment, which serves as a local server to integrate PHP, Apache, and MySQL. After internal testing was completed, the system was deployed on the Karang Anyer Puskesmas server for operational use.

Figure 3. Web-Based Information System Architecture of Karang Anyer Puskesmas

Figure 3 visualizes the system architecture, which adopts the fundamental 3-tier model a design decision essential for enforcing the principle of separation of concerns. This architecture strategically separates the Browser presentation layer, the PHP application layer executing business logic, and the MySQL data layer as an isolated repository for data persistence. This clear separation not only simplifies debugging and maintenance processes but also ensures system scalability and security, as modifications in one layer do not directly affect the others. Consequently, this architectural foundation guarantees a robust system that can adapt to the evolving operational needs of the Puskesmas in the future.

2.5. Testing and Evaluation Metrics

To validate the system's effectiveness, this study measures two key variables:

1. Functional Validity (Black Box): Evaluates whether all system features (Login, Queue, SOAP, Pharmacy) function according to the design specifications. The indicator is a "Valid" status for every test case.
2. User Acceptance (UAT): Measures end-user satisfaction. A questionnaire using a Likert scale (1-5) was distributed to 10 respondents (consisting of administrative staff, doctors, and system verifiers) to evaluate aspects such as User Interface (UI), User Experience (UX), and Performance.

3. Results

3.1. Login Page Interface

The login page serves as the access point to SIM-PUSKESMAS. The system validates access for Administrators, Doctors, Pharmacists, and Patients. This page also provides a 'Register New Account' function specifically for patient registration.

(a) (b)

Figure 4. (a) Login page interface; **(b)** Account registration page interface.

3.2. Dashboard Page

The dashboard page of this program presents processed and inputted data managed by the System Administrator, as shown in Figure 5. The dashboard displays key statistical summaries, such as the number of patients, total staff, and daily visit charts, while providing quick access to the main functions.

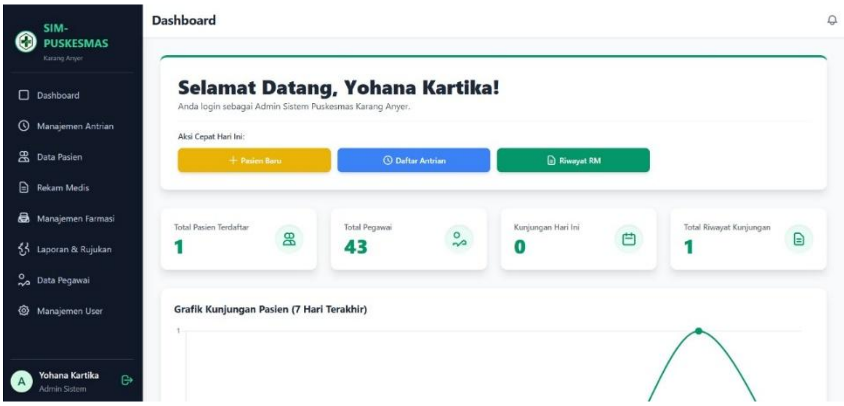


Figure 5. Dashboard Page.

3.3. *Patient Registration Page*

The Patient Registration page is a form used to collect essential patient identity data. The information entered includes Medical Record Number, Patient Name, demographic data (Gender, Date of Birth, Religion, Marital Status), contact information (Phone/Cellphone, Address), and BPJS Number (if available).

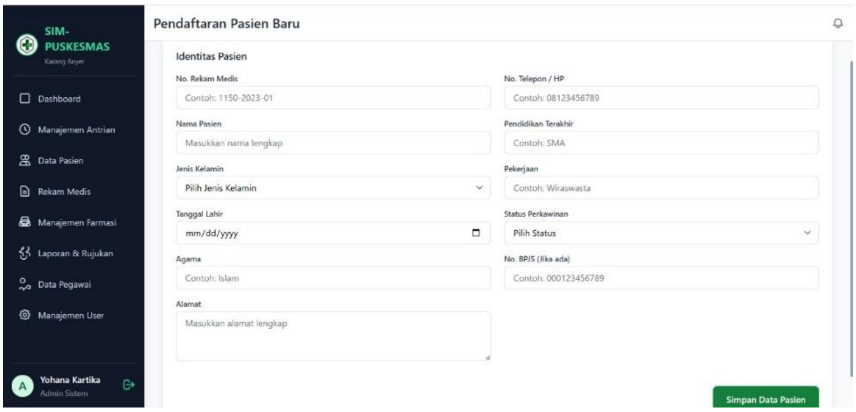
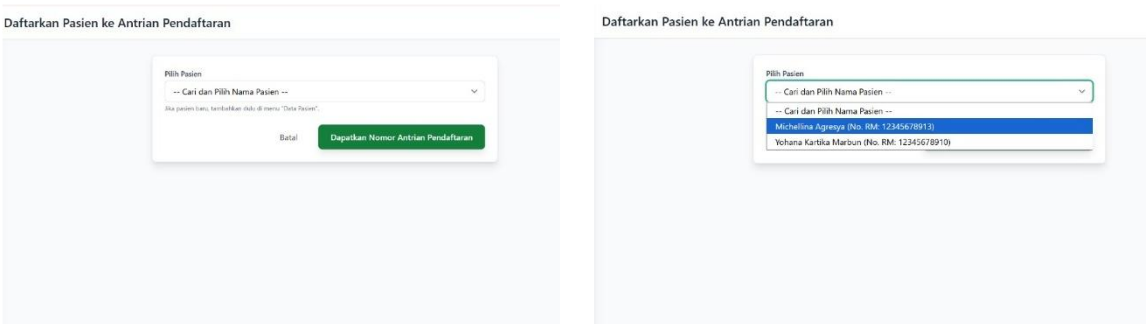


Figure 6. New Patient Registration Page.

3.3. *Patient to Registration Queue*

This page functions to process the registration of registered patients into the service queue. The process involves selecting a patient from a dropdown menu displaying Name and Medical Record Number, after which the staff clicks the 'Get Registration Queue Number' button to validate the patient into the queue system, as shown in the figure 7:

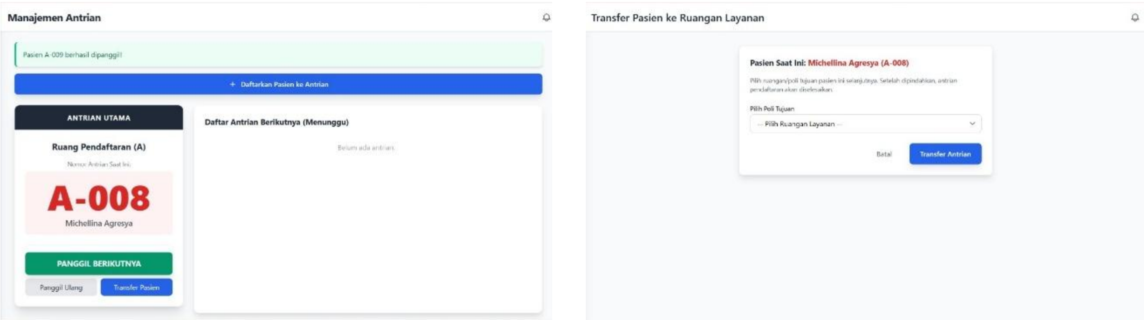


(a) (b)

Figure 7. **(a)** Queue Registration Form; **(b)** Patient Selection from the List.

3.4. Queue Management Page

The Queue Management page is designed to call the queue and includes a Transfer Patient button. This button leads to a form used to select the patient's next destination clinic, completing the registration queue process, in Figure 8.



(a) (b)

Figure 8. **(a)** Main Interface of Queue Management**; (b)** Patient Transfer Form.

3.5. Patient Data Page

The Patient Data page displays a list of registered patients. This page functions to search for patient information, add new patients, and provides actions such as Queue Registration and editing patient data, showed in Figure 9.

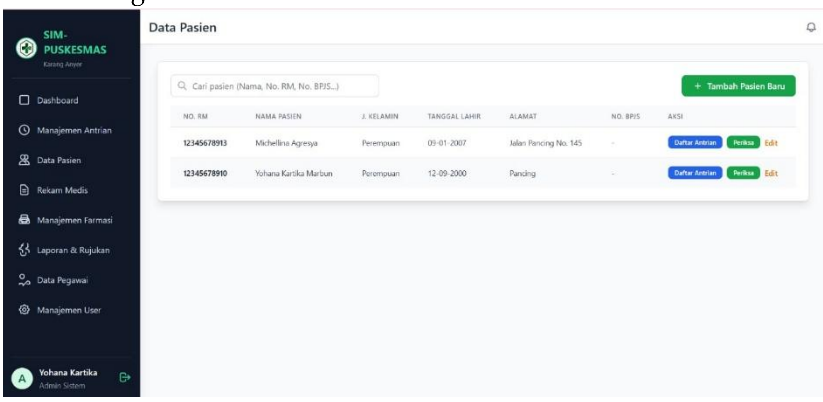
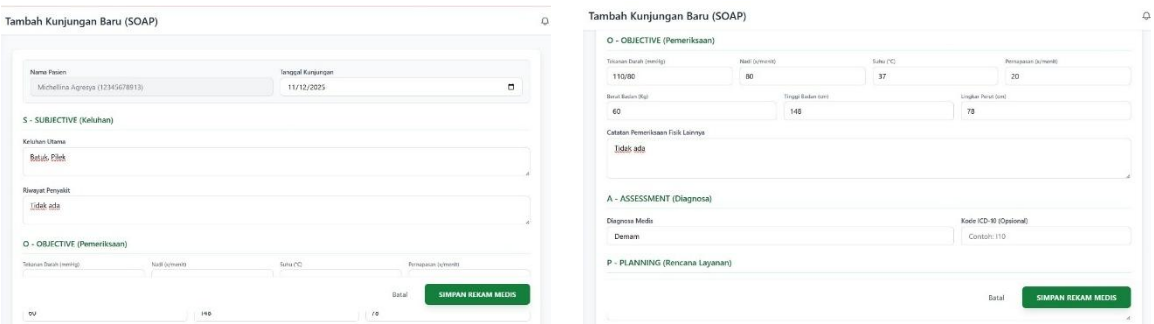


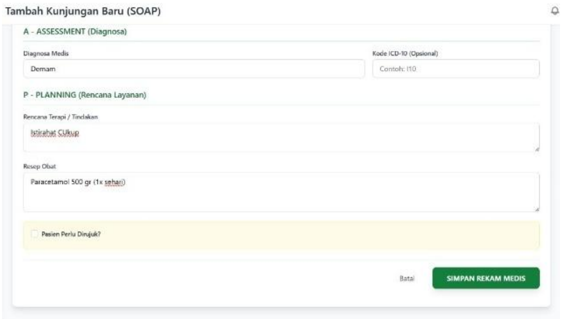
Figure 9. Patient Data Page.

3.6. Medical Record Page

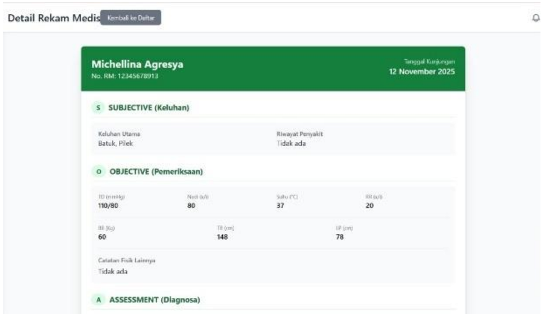
This page is responsible for documenting and reviewing patient Medical Records using the SOAP standard, showed in Figure 10. The New Visit Form is used by Doctors to input data on Complaints, Examinations, Diagnoses, and Service/Prescription Plans, while the Medical Record Detail Page serves as the output, displaying all stored SOAP data for the purpose of reviewing the patient's medical history.



(a) (b)



(c)

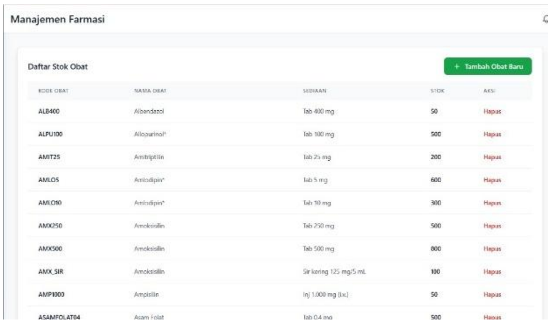


(d)

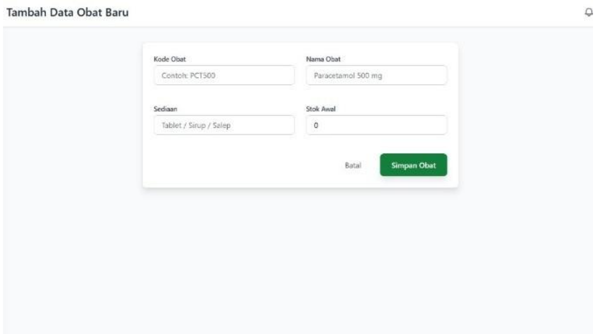
Figure 10. **(a-c)** Medical Record Input Form(SOAP); **(d)** Stored Medical Record Details.

3.7. *Pharmacy Management Page.*

The Pharmacy Management page serves as the control center, displaying a table of Drug Stock, including Code, Name, and Quantity. This page functions to monitor inventory status, add new drugs, and manage existing stock.



(a)



(b)

LAPORAN STOK OBAT FARMASI

Per 12 November 2025 15:37:04

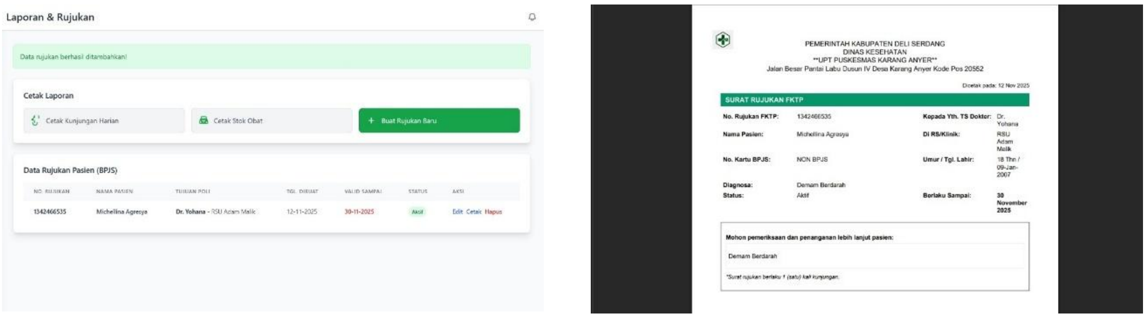
No	Kode Obat	Nama Obat	Sediaan	Stok Saat Ini
1	ALB400	Albendazol	Tab 400 mg	50
2	ALPU100	Allopurinol*	Tab 100 mg	500
3	AMIT25	Amitriptilin	Tab 25 mg	200
4	AMLO5	Amlodipin*	Tab 5 mg	600
5	AMLO10	Amlodipin*	Tab 10 mg	300
6	AMX250	Amoksisilin	Tab 250 mg	500
7	AMX500	Amoksisilin	Tab 500 mg	800
8	AMX_SIR	Amoksisilin	Sir kering 125 mg/5 mL	100
9	AMP1000	Ampisilin	Inj 1.000 mg (i.v.)	50
10	ASAMFOLAT04	Asam Folat	Tab 0,4 mg	500
11	AMF500	Asam Mefenamat	Tab 500 mg	300
12	DEXA_INJ	Deksametason	Inj 5 mg/mL	100
13	DZP_INJ	Diazepam	Inj 5 mg/mL	20
14	DIPH_INJ	Difenhidramin	Inj 10 mg/mL (i.v./i.m.)	50
15	DOKSI100	Doksisiklin	Kaps 100 mg	100

(c)

Figure 10. (a) Drug Stock List; **(b)** New Drug Data Entry Form; **(c)** Drug Stock Printout.

3.8. *Reports and Referral Page*

This page manages patient referral data and report printing. The main page displays the Patient Referral history and provides an option to print reports. The output is an official FKTP Referral Letter document containing detailed referral information, as showed in Figure 11.

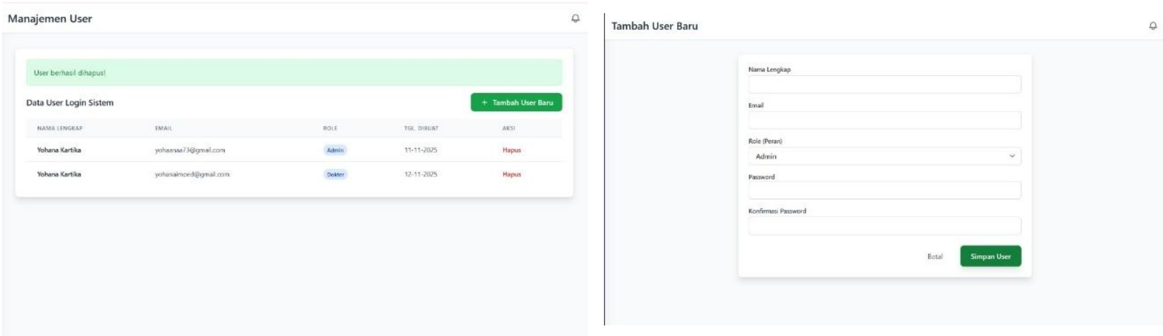


(a) (b)

Figure 11. **(a)** Main Page; **(b)** Example of FKTP Referral Letter.

3.9. *User Management Page*

This page is specifically designed for the Administrator to manage all system user accounts. It displays User Login data and provides functions to add new users as well as delete existing user accounts, in Figure 12.



(a) (b)

Figure 12. **(a)** User Management Page; **(b)** Add User Form.

3.10. *System Testing Results*

The system underwent rigorous testing to ensure it met the defined objectives. The testing phase was divided into functional testing and user acceptance testing.

3.10.1. Black Box Testing Results

Functional testing was performed using the Black Box method to ensure every module operates according to specifications. The summary of test results is presented in Table 1. Based on the table, all test scenarios show a "Valid" status, indicating that the system has a 100% functional success rate and is free from critical errors in its main features.

Table 1. Black Box Testing Results

No	Test Case Scenario	Expected Result	Actual Result	Status
1	Admin & Doctor Login	System directs user to Dashboard based on access rights	System successfully displays the specific Dashboard page	Valid
2	New Patient Registration	Patient identity data is stored in the MySQL database	Data is stored successfully and appears in the patient list	Valid
3	Queue Generation	System generates a unique queue number (e.g., A-001)	Queue number is generated and displayed correctly	Valid
4	Medical Record Input (SOAP)	Medical assessment data is saved in the patient's history	SOAP data is saved completely and linked to the patient	Valid
5	Pharmacy Stock Update	Drug stock decreases automatically after a prescription is saved	Drug stock quantity decreases according to the prescription	Valid
6	Print Referral Letter	System generates and downloads a Referral Letter (PDF)	PDF file is generated and downloaded successfully	Valid
7	User Management	Admin can add new users or delete existing accounts	User data is successfully added or removed from the system	Valid

3.10.2. User Acceptance Testing (UAT) Results

To measure the level of satisfaction and system feasibility, UAT was conducted with users (Admins and Medical Staff). The assessment utilized a Likert scale (1-5). The recapitulation of the assessment results is shown in Table 2.

Table 2. User Acceptance Testing (UAT) Results

No	Assessment Aspect	Average Score (Scale 1-5)	Category
1	User Interface (UI) Design	4.6	Very Good
2	Ease of Use (Usability)	4.7	Very Good
3	System Access Speed	4.5	Very Good
4	Feature Suitability	4.8	Very Good
5	Data and Report Accuracy	4.6	Very Good
-	Total Average	4.64 (92.8%)	Very Feasible

Data in Table 2 shows that the system obtained a total average score of 4.64 or 92.8%. Based on the feasibility category, this score indicates that SIM-PUSKESMAS falls into the "Very Feasible" category and is well-accepted by users to support daily community health center operations.

4. Discussion

The implementation of the Web-based Management Information System (SIM-PUSKESMAS) at UPT Karang Anyer has proven to be a significant step towards digitizing primary healthcare services. Based on the results obtained, this section discusses the interpretation of findings, comparisons with previous studies, and the limitations of the current system.

4.1. Interpretation of Main Findings

The study confirms that the proposed system operates effectively. As evidenced by the Black Box testing results, in Table 1, the system achieved a 100% success rate across all functional modules, including Registration, Medical Records (SOAP), and Pharmacy Management. This indicates that the logic errors and data redundancies often found in manual systems have been successfully eliminated.

Furthermore, the User Acceptance Testing (UAT) score of 92.8%, in Table 2, demonstrates a high level of satisfaction among doctors and administrative staff. This high score is primarily attributed to the "Ease of Use" and "Feature Suitability" aspects. The integration of the SOAP medical record format directly with the pharmacy module allows for real-time stock deductions, addressing the administrative inefficiency issues highlighted in the introduction. This automated workflow significantly reduces the waiting time for patients and minimizes human error in drug inventory recording.

4.2. Comparison with Previous Research

The results of this study align with and expand upon several previous researchers' findings:

1. Comparison with Manual Systems: Consistent with the problems identified by Desi [7], [8] and Hanafi et al. [4], manual data recording leads to delays in data retrieval. SIM-PUSKESMAS solves this by digitizing the archive, reducing data search time from minutes to seconds.
2. Comparison with Similar Web-Based Systems: Saidy et al. [7], [8] developed a web-based health information system focusing on general data management. However, our study offers a more specific contribution by implementing a comprehensive Pharmacy Management module that is automatically linked to the doctor's prescription input. While Saidy’s work emphasized the general architecture, SIM-PUSKESMAS emphasizes the interconnectivity between the polyclinic and the pharmacy.
3. Comparison regarding Workflow Integration: Firdaus Trama et al. [7], [8] highlighted the importance of workflow transparency in Electronic Medical Records. SIM-PUSKESMAS answers this need by providing a role-based dashboard (Admin, Doctor, Pharmacist) that ensures every staff member can track the patient's status in real-time, a feature that was often lacking in earlier, fragmented implementations.

4.3. Limitations of the Study

Despite the successful implementation, this study has several limitations that need to be acknowledged:

1. Network Scope: The system currently runs on a Local Area Network (LAN) using a local server (XAMPP) for security and cost-efficiency reasons. This means the system cannot be accessed by patients from their homes for online registration.
2. Lack of External Integration: The current version of SIM-PUSKESMAS operates independently and has not yet been integrated with the national health insurance system (BPJS P-Care) or other external health platforms.
3. Mobile Accessibility: The interface is web-based and optimized for desktop use by staff. A dedicated mobile application for patients to monitor queue numbers remotely has not yet been developed.

5. Conclusions

This study successfully designed and implemented the Integrated Management Information System (SIM-PUSKESMAS) at UPT Karang Anyer using the Research and Development (R&D) approach. The system was developed to address the inefficiencies of manual data recording by integrating queue management, medical records (SOAP), and pharmacy inventory into a single web-based architecture.

The significant findings of this study are demonstrated through rigorous testing. The Black Box testing results indicated a 100% functional success rate, confirming that all core modules operate without errors. Furthermore, the system received a strong positive response from medical staff, with User Acceptance Testing (UAT) yielding an average score of 92.8%, classifying the system as "Very Feasible" for operational implementation. These results prove that the system effectively improves service speed and data accuracy compared to the previous manual method.

For future research, it is recommended to expand the system's capabilities to address current limitations. Future development should focus on building a mobile application (Android/iOS) to allow patients to register online from home. Additionally, interoperability should be enhanced by developing an API to integrate SIM-PUSKESMAS with the national health insurance system (BPJS P-Care), enabling seamless data exchange and broader healthcare connectivity.

References

1. H. Marfalino, A. Fitrul Hadi, and R. Liana Gema, "Geographical Information System Design Web-Based Mapping Of Hospitals In Padang City," *J. Sains Informatika Terapan*, vol. 1, no. 2, pp. 1–9, Feb. 2022. doi: 10.51876/jsit.v1i2.4.
2. M. Susanti, "Sistem Informasi Kendali Harga Sewa Billboard Berdasarkan Nilai Pajak," *Jurnal Sains Informatika Terapan (JSIT)*, vol. 1, no. 1, pp. 10–15, 2022, doi: 10.62357/jsit.v1i1.55.
3. U. Meningkatkan Pelayanan Kesehatan Di Puskesmas Curug Isnaeni Siktin Kusuma Desi, "Sistem Informasi Manajemen Data Pasien.", *Jurnal Teknik Komputer AMIK BSI*, vol. 4, no. 1, pp. 1–6, 2020.
4. S. T. Hanafi, W. U. Panjaitan, A. R. Ramadhan, and S. H. Purba, "Penerapan Sistem Informasi Kesehatan di Puskesmas di Indonesia," *Sagita Academia Journal*, vol. 3, no. 1, pp. 19–23, Jan. 2025, doi: 10.61579/sagita.v3i1.333.
5. G. Firdaus Trama, Z. Adiputra Golo, and J. Rekam Medis dan Informasi Kesehatan Poltekkes Kemenkes Semarang Jl Tirto Agung, "An Evaluation of Electronic Medical Record System Maturity at Public Health Center Irmawati 1," *Jurnal Rekam Medis dan Informasi Kesehatan*, vol. 8, no. 1, 2025, doi: 10.31983/jrmik.v8i1.12824.
6. D. L. Jayanto et al., "Rancangan sistem informasi distribusi dokumen rekam medis di Puskesmas Tanjunganom Kabupaten Nganjuk," *INFOKES: Jurnal Ilmiah Rekam Medis dan Informatika Kesehatan*, vol. 13, no. 2, pp. 96–103, 2023, doi: 10.47701/infokes.v13i2.2626.
7. M. Waruwu, "Metode Penelitian dan Pengembangan (R&D): Konsep, Jenis, Tahapan dan Kelebihan," *Jurnal Ilmiah Profesi Pendidikan*, vol. 9, no. 2, pp. 1220–1230, May 2024, doi: 10.29303/jipp.v9i2.2141.
8. D. Murdiani and M. Sobirin, "Perbandingan Metodologi Waterfall Dan Rad (Rapid Application Development) Dalam Pengembangan Sistem Informasi," *Jurnal Informatika Teknologi dan Sains (JINTEKS)*, vol. 4, no. 4, pp. 302–306, 2022, doi: 10.51401/jinteks.v4i4.2008.
9. A. K. Endriharto and J. Suprianto, "Pengenalan Sistem Informasi Pendataan Terpadu Kebutuhan Penyandang Disabilitas di Kabupaten Magetan," *Jurnal Pengabdian Literasi Digital Indonesia*, vol. 2, no. 1, pp. 38–48, Jun. 2023, doi: 10.57119/abdimas.v2i1.21.
10. I. Nursetiawan, R. Agung, and K. Putra, Data dan Sistem Informasi Desa dalam Era Keterbukaan Informasi Publik di Kabupaten Ciamis," *Dinamika: Jurnal Ilmiah Ilmu Administrasi Negara*, vol. 8, no. 3, pp. 463–471, Des. 2021
11. A. Fu'adi, A. Prianggono, A. Komunitas, N. Pacitan, A. A. Id, and A. A. Id, "Analisa dan Perancangan Sistem Informasi Akademik Akademi Komunitas Negeri Pacitan Menggunakan Diagram UML dan EER," *Jurnal Ilmiah Teknologi Informasi Asia*, vol. 16, no. 1, 2022.
12. V. Yoga and P. Ardhana, "Perancangan Sistem Informasi Apotek Qamarul Huda Menggunakan Unified Modeling Language (UML)," *J. Kesehat. Qamarul Huda*, vol. 9, no. 2, pp. 115–119, 2021, doi: 10.37824/jkqh.v9i2.2021.309.
13. H. N. Saidy, A. Adivar, M. Musawwir, A. M. Yusuf, and A. Asrul, "Perancangan dan Implementasi Sistem Informasi Kesehatan Berbasis Web di Puskesmas Turikale Kabupaten Maros," *Jurnal Minfo Polgan*, vol. 14, no. 1, pp. 1493–1501, Jul. 2025, doi: 10.33395/jmp.v14i1.15129.
14. M. R. Janitra and D. Febriawan, "Sistem Informasi Berita Online Berbasis Web Untuk Penyebaran Informasi Publik dengan Menggunakan Metode SDLC," *Journal of Informatics Management and Information Technology*, vol. 5, no. 3, pp. 344–354, 2025, doi: 10.47065/jimat.v5i3.710.
15. Shafirina Arsydini Shidqi, Gustin Setyaningsih, Ika Romadoni Yunita, and Muhamad Awiet Wiedanto Prasetyo, "Perancangan Website Penjualan Menggunakan Metode Prototype Pada Toko Agen Beras," *Jurnal Riset Sistem Informasi*, vol. 2, no. 3, pp. 39–53, Jul. 2025, doi: 10.69714/dz223472.

COMPLETE
Journal of Computer, Electronic, and Telecommunication
ISSN 2723-4371, E-ISSN 2723-5912

Article

Application Of TF-IDF And Word2vec For Feature Extraction In Sentiment Analysis Of Free Nutritious Food Policies

Qisthi Annisa Haq[1], Alam Rahmatulloh[2]

[1-2] Department of Engineering, Siliwangi University, Tasikmalaya, Indonesia
* Correspondence: alam@unsil.ac.id

Received: 1 December 2025; Revised: 19 December 2025; Accepted: 30 December 2025; Published: 31 December 2025

Abstract: The free nutritious meal policy has become a hot topic of discussion among the public because it is related to improving health and education quality. However, its implementation has given rise to a variety of pros and cons that need to be analyzed systematically. This study aims to analyze sentiment toward the policy by utilizing Term Frequency–Inverse Document Frequency (TF-IDF) and Word2Vec as feature extraction methods on public review data obtained from social media X. After undergoing preprocessing and automatic labeling, the data was classified into positive and negative sentiments using the Support Vector Machine (SVM) algorithm. The analysis results show that the sentiment data is unbalanced, with the positive class dominating at 75% and the negative class at 25%. In model testing, TF-IDF achieved an accuracy of 81%, while Word2Vec achieved an accuracy of 80%. This difference shows that TF-IDF is more stable in handling short and informal texts, while Word2Vec still has the potential to capture the semantic context between words. This research opens up opportunities for further research, it is recommended to balance the data between classes and combine the TF-IDF and Word2Vec methods, or use a deep learning approach such as BERT to obtain more accurate results and capture deeper semantic context.

Keywords: Accuracy, Automatic Labelling, Positive and Negative Sentiments, Public Review Data, Semantic Context

1. Introduction

The development of social media has become one of the main platforms for debating government policies. Social media, which initially served only as a platform for personal interaction, has now evolved into the main arena for shaping and disseminating public opinion [1]. One of the most hotly debated public opinions is the government's Free Nutritious Meals (MBG) policy. This free nutritious meals program aims to provide lunch and milk in the hope of overcoming malnutrition and stunting among school children [2]. However, its implementation has elicited mixed responses from the public. Some support it on humanitarian grounds and for the sake of nutritional equality, while others criticize the program's effectiveness, transparency, and logistical readiness [3]. There have even been incidents such as mass poisoning in several schools [4]. This further reinforces the urgency of understanding public perception of this policy.

In this case, sentiment analysis is an important way to gauge public opinion regarding government policies. With this analysis, public opinion can be divided into various categories such as positive, negative, or neutral [5]. In today's digital age, social media such as X has become one of the main sources of public opinion because the data obtained is open and real time [6]. To analyze public opinion systematically, feature extraction is needed to convert unstructured text into numerical representations that can be processed by algorithms [2]. This stage is very important because the quality of the extracted features can affect the performance of the sentiment analysis model. Two popular approaches in this field are Term Frequency-Inverse Document Frequency (TF-IDF) and Word2Vec.

The TF-IDF method is popular because it looks at how important each word is within a document. This means that words that show up a lot in a specific document but not much in all the documents together will be seen as more important [7]. In addition to being simple and easy to interpret, TF-IDF has proven to be stable as a baseline in various text processing studies [8]. However, the main weakness of TF-IDF lies in its inability to capture the semantic meaning between words. Meanwhile, the Word2Vec method converts words into dense vectors based on the words that appear in a sentence. Word2Vec works with two main methods, namely Continuous Bag of Words (CBOW) and Skip-gram. Both methods are capable of capturing the meaning and sentence structure relationships between words. In this way, words with similar meanings will have vectors that are close to each other in multidimensional space [9].

TF-IDF and Word2Vec were chosen for this research because they are two important and commonly used methods for showing text features, often serving as basic examples in studies about understanding feelings in text. It has been proven to work well in identifying feelings in short and casual messages, especially when used with classic machine learning methods like Support Vector Machine (SVM) [7], [10]. On the other hand, Word2Vec is a method that looks at the meanings of words. It creates dense vectors that keep similar words close together, helping to show how words relate to each other [9], [11].

The research conducted by Sitanggang et al. [2] focused on the use of the naïve bayes algorithm with the TF-IDF feature extraction method. Text preprocessing was carried out thoroughly, from cleaning to stemming. The results obtained from this study were an accuracy of 72.2%. Although this research can be considered sufficient, it did not compare the Word2Vec method to determine the semantic meaning between words in each dataset.

A different research project was carried out by Dani and others [10], where they used the SVM method to look at how well TF-IDF-SVM worked compared to Word2Vec-SVM. Text preprocessing was carried out thoroughly. The results of this study show that TF-IDF-SVM has a higher accuracy of 83% compared to Word2Vec at 77.5%. However, the TF-IDF method is unable to understand the semantic meaning between words.

Ati et al. [12] conducted research comparing the naïve bayes, SVM, KNN, and ensemble (stacking) algorithms with the TF-IDF feature extraction method and the SMOTE-Tomek technique for balancing the dataset and complete preprocessing. This study aimed to determine which algorithm was the most optimal. The results showed that the SVM algorithm achieved the highest accuracy of 95.05%. However, this study has limitations in the form of neutral class bias and dependence on automatic translation results, which can affect accuracy.

Furthermore, the study by Nadia et al. [13] examined sentiment analysis of COVID-19 vaccination using a Recurrent Neural Network (RNN) and two feature extraction methods, namely TF-IDF and Word2Vec. The results showed that the combination of RNN-Word2Vec produced the highest accuracy, namely 53%, and the combination of RNN-TF-IDF produced 51%. Although the improvement is relatively small, it shows that feature extraction with Word2Vec is superior. The low accuracy is due to the variation in the algorithms used, so that in future research, other more suitable algorithms can be used.

Finally, research by Zhan [11] compared feature extraction between TF-IDF and Word2Vec for food reviews using the Logistic Regression algorithm. Based on testing, TF-IDF had a percentage of 99.16% on the training data, but had a significant decrease on the test data of 73.9%. Meanwhile, using Word2Vec feature extraction showed more stable results with an accuracy of 68.4% for the training data and 68.65% on the test data. This shows that TF-IDF is more prone to overfitting than Word2Vec.

Based on several studies presented earlier, most studies rely on preprocessing stages such as stemming and stopword removal, which have the potential to remove the natural context in sentences. Unlike previous studies that used stemming and stopword removal for text normalization, this study deliberately omitted these two stages to see how the model's performance changed when semantic features and word context were left unchanged, without morphological modification. In this way, this study not only compares the performance of TF-IDF and Word2Vec in the usual way, but also helps to understand how much influence the text preprocessing process has on the effectiveness of feature representation in public policy sentiment analysis using data from social media. So, the TF-IDF and Word2Vec ways of picking out important details with the SVM method are likely to give a better understanding of how well we can show feelings about public policy by looking at social media information.

2. Materials and Methods

The research looks at how well two ways of pulling important information from text work, specifically Term Frequency-Inverse Document Frequency (TF-IDF) and Word2Vec, when analyzing opinions about the Free Nutritious Meals (MBG) policy. The research process consists of several main stages as shown in Figure 1.

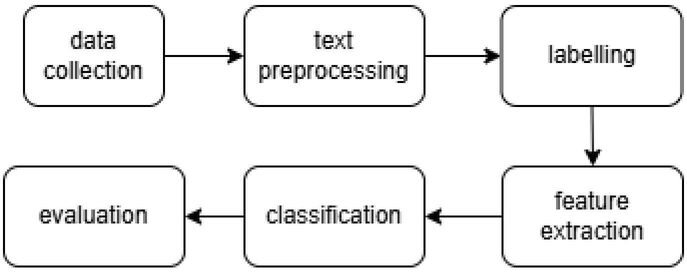


Figure 1. Research Stages

Figure 1 shows the stages of research that will be conducted, from data collection to evaluation. The following is an explanation of each research flow.

2.1 Data Collection

The research data will be obtained from Kaggle in csv format containing the results of scraping from social media X. After the cleaning process, the total data will be divided into two parts, namely 80% as training data and 20% as test data.

2.2 Text Preprocessing

This stage aims to clean the text so that it is ready to be processed in the feature extraction stage. However, unlike previous studies [4], [5], this study does not use the stemming and stopword removal stages in order to observe the influence of the natural context of words on sentiment analysis. The preprocessing stages that will be carried out are case folding, cleansing, and tokenizing. The outcome of this phase is a group of writings that can now be prepared for feature extraction.

2.3 Labeling

The labeling stage is carried out to assign a sentiment class to each text, namely positive, negative, and neutral. This process is carried out automatically using the pre-trained BERT Multilingual Sentiment model from the Hugging Face Transformers library, which is capable of recognizing sentiment in various languages, including Indonesian. The labeling results show that most of the data falls into the positive class, followed by the negative class, and a small portion of neutral data, which was then removed because there was too little to analyze further.

2.4 Feature Extraction

The part where we extract features is meant to change words into numbers so that computers can understand and work with them. Two methods are used.

1. Term Frequency-Inverse Document Frequency

   TF-IDF shows how significant a word is in one paper when you think about all the papers in a group. The TF-IDF value is determined by using a specific formula (1).

$$TF - IDF(t,d) = TF(t,d) \times \log \frac{N}{DF(t)} \tag{1}$$

   In this case, TF(t,d) shows how often the word t appears in the document d. DF(t) represents how many documents have the word t, and N is the overall count of all the documents [6]. This method produces a sparse word representation matrix and is widely used in sentiment analysis research due to its simplicity [7].

2. Word2Vec

   Word2Vec changes words into a simpler format using a lower-dimensional space by training a neural network. It does this through two main methods: Continuous Bag of Words (CBOW) and Skip-Gram. However, this study uses the Skip-Gram model because it is more effective at capturing semantic meaning for small to medium datasets [8].

2.5 Classification Using Support Vector Machine (SVM)

The SVM algorithm is used as the main classification model due to its ability to separate high-dimensional data using an optimal hyperplane and its proven stable performance in various previous analysis studies. In this study, the Radial Basis Function (RBF) kernel is used, and the model will be trained using the results of TF-IDF and Word2Vec feature extraction separately to compare their performance.

2.6 Model Evaluation

Model evaluation was performed by comparing the classification results of the TF-IDF-SVM and Word2Vec-SVM models using the accuracy, precision, recall, and F1-score metrics, with the following formulas (2).

$$\begin{gathered} Precision = \frac{TP}{TP + FP} \\ Recall = \frac{TP}{TP + FN} \\ F1\ score = 2 \times \frac{Precision}{Precision} \end{gathered} \tag{2}$$

These values are obtained from the confusion matrix of the model's predictions on the test data.

3. Results and Discussion

3.1 Data Collection

The data was obtained from a dataset available on Kaggle called "Public Sentiment towards Free Nutritious Food." This dataset contains 3,461 text data points from social media X discussing the MBG policy.

3.2 Text Preprocessing

Preprocessing steps were carried out to ensure that the data was suitable for the feature extraction stage. The steps taken included:

1. Cleaning: removing links (URLs), user tags (@username), numbers, punctuation marks, non-Latin symbols, and emojis.
2. Case folding: converting all letters to lowercase for consistency.

	full_text	cleaning	case_folding
0	Makan Siang Bergizi Gratis https://t.co/f27alt...	Makan Siang Bergizi Gratis	makan siang bergizi gratis
1	Momen Prabowo Tanda Tangani Sepatu Siswa di Bo...	Momen Prabowo Tanda Tangani Sepatu Siswa di Bo...	momen prabowo tanda tangani sepatu siswa di bo...
2	@lenteradata Semoga program makan bergizi grat...	Semoga program makan bergizi gratis dri MDA te...	semoga program makan bergizi gratis dri mda te...
3	Pemprov Babel dan DPRD Bahas Anggaran Program ...	Pemprov Babel dan DPRD Bahas Anggaran Program ...	pemprov babel dan dprd bahas anggaran program ...
4	@lenteradata Menurutku berhasil skliki program...	Menurutku berhasil skliki programnya MDA yg ma...	menurutku berhasil skliki programnya mda yg ma...

Figure 2. Cleaning and case folding stages

Figure 2 shows the results of the cleaning and case folding stages that have been carried out.

3. Tokenizing: breaking sentences into word segments using the nltk library.
4. Stopword: removing words that are considered unimportant, such as auxiliary words like di, pada, and others, using the nltk library.
5. Stemming: returning each inflected word to its root form using the sastrawi library.

tokenize	stopword	stemming
[makan, siang, bergizi, gratis]	[makan, siang, bergizi, gratis]	[makan, siang, gizi, gratis]
[momen, prabowo, tanda, tangani, sepatu, siswa...	[momen, prabowo, tanda, tangani, sepatu, siswa...	[momen, prabowo, tanda, tangan, sepatu, siswa,...
[semoga, program, makan, bergizi, gratis, dri,...	[semoga, program, makan, bergizi, gratis, dri,...	[moga, program, makan, gizi, gratis, dri, mda,...
[pemprov, babel, dan, dprd, bahas, anggaran, p...	[pemprov, babel, dprd, bahas, anggaran, progra...	[pemprov, babel, dprd, bahas, anggar, program,...
[menurutku, berhasil, skliki, programnya, mda,...	[menurutku, berhasil, skliki, programnya, mda,...	[turut, hasil, skliki, program, mda, yg, makan...

Figure 3. Tokenization, stopword, and stemming stages

Figure 3 shows the results of the tokenization, stopword, and stemming processes. The stopword and stemming processes were only performed on the TF-IDF feature extraction method because in the word2vec method, both preprocessing steps were omitted in order to preserve the semantic meaning of each word.

3.3 Labeling

After all preprocessing stages are complete, the next step is to label each tweet. Labeling is done with a pretrained model from huggingface, which is divided into two classes: positive and negative. The following is the percentage of the labeling results.

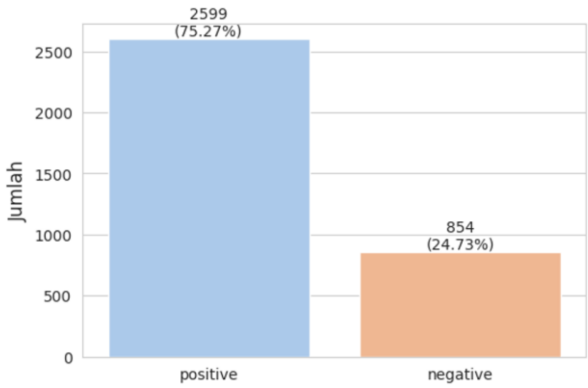


Figure 4. Sentiment percentage

Figure 4 shows that the data is unbalanced between sentiment classes. The positive class dominates at 75%, while the negative class is only 24%. This imbalance has the potential to affect model performance, especially in the ability to detect minority (negative) sentiment.

3.4 Feature Extraction

1. Term Frequency-Inverse Document Frequency (TF-IDF)

   Weighting is performed using the sklearn.feature_extraction.text library, namely Tfidfvectorizer. Feature extraction is also performed using ngram with a range of 2 because there are several words that have meaning when taken from a sentence, not just one word, such as the negative sentence "not good". The weighting process is shown in Figure 5.

```
#vektorisasi
vectorizer = TfidfVectorizer(max_features=5000, ngram_range=(1, 2))
X_train_vec = vectorizer.fit_transform(X_train)
X_test_vec = vectorizer.transform(X_test)
```

Figure 5. TF-IDF weighting process

2. Word2vec

   Weighting is performed using the gensim library by forming a dense 100-dimensional vector that represents the semantic meaning between words. The vector results are shown in Figure 6.

```
Contoh vektor: [ 4.4683829e-02  2.0138405e-01 -4.0741217e-01 -1.1170293e-01
  9.6124060e-02 -4.7956163e-01  7.4206030e-01  1.1123683e+00
 -9.2041296e-01 -5.0495160e-01  2.6599023e-02 -1.0221407e+00
 -2.2153553e-01  5.4185659e-01  4.8920295e-01 -3.4648809e-01
  7.7135265e-01 -2.1820754e-01 -2.7130967e-01 -1.5140573e+00
  1.5270750e-01 -4.2066485e-01  6.4909178e-01  8.4196448e-02
 -2.5029770e-01 -1.7926675e-01 -4.2642337e-01  9.7632274e-02
 -3.6353946e-01  2.6682001e-01  1.0827038e+00  1.3716710e-01
  1.6061056e-01 -1.0781510e+00 -2.8858504e-01  7.1740329e-01
  6.9700783e-01  5.7481136e-02 -2.8648388e-01 -8.3722836e-01
  3.4274989e-01 -4.3552020e-01 -5.7815391e-01  3.9090189e-01
  3.9069852e-01  1.6226390e-01 -8.8316655e-01 -2.7380165e-01
  5.6738073e-01  6.2506253e-01 -3.9525044e-01 -4.4990775e-01
 -7.8249162e-01 -1.9732749e-01 -2.4721050e-01  2.3442966e-01
  1.3149440e-01 -1.8677150e-01 -5.3830773e-01  4.8539239e-01
  1.3626905e-01  3.9991775e-01  7.5395894e-01  1.9297811e-01
 -8.3394444e-01  4.7199324e-01  1.1477207e+00  5.5433238e-01
 -9.8499960e-01  7.8503740e-01  2.0710133e-01  8.0250964e-02
  4.1789353e-01  2.0184699e-01  8.5436761e-01 -1.8714327e-01
 -3.3753313e-02  1.3662539e-03 -8.7240577e-01 -2.5870734e-01
  1.4543296e-01  6.4099245e-02 -2.3534240e-01  6.5208167e-01
```

Figure 6. Word2vec vector results

3.5 Classification using Support Vector Machine (SVM)

After the data was converted into vector form, the SVM model was trained using two types of input, namely TF-IDF extraction results and word2vec extraction results. The kernel used was the Radial Basis Function (RBF) kernel due to its ability to separate non-linear data well. The accuracy results using the TF-IDF feature extraction method from the modeling that has been carried out are shown in Figure 7.

```
Test Accuracy: 0.8147612156295224

              precision    recall  f1-score   support

    negative       0.61      0.56      0.58       160
    positive       0.87      0.89      0.88       531

    accuracy                           0.81       691
   macro avg       0.74      0.73      0.73       691
weighted avg       0.81      0.81      0.81       691
```

Figure 7. Accuracy Results with TF-IDF Feature Extraction

The accuracy results using the word2vec feature extraction method are shown in Figure 8.

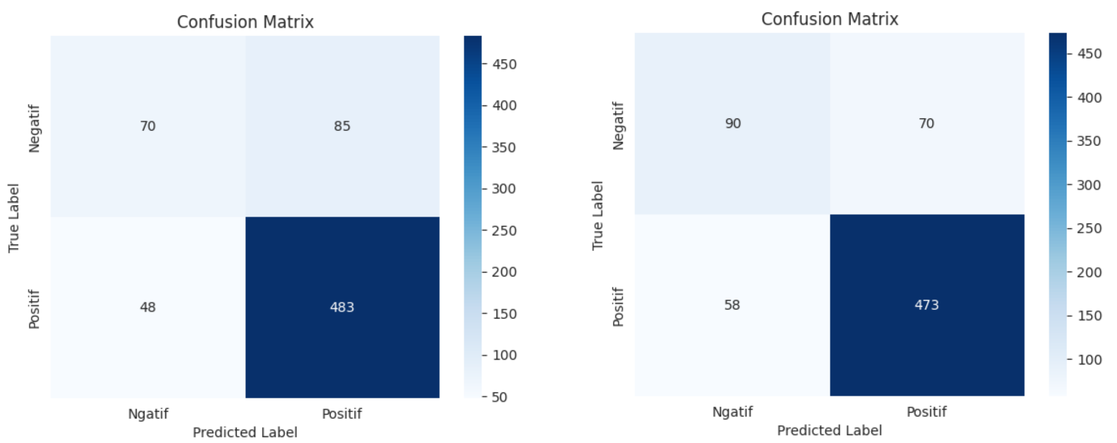
```
Test Accuracy: 0.8061224489795918
              precision    recall  f1-score   support

    negative       0.59      0.45      0.51       155
    positive       0.85      0.91      0.88       531

    accuracy                           0.81       686
   macro avg       0.72      0.68      0.70       686
weighted avg       0.79      0.81      0.80       686
```

Figure 8. Accuracy Results with word2vec Feature Extraction

The test results show that the combination of TF-IDF – SVM achieved an accuracy of 81% and word2vec – SVM achieved an accuracy of 80%. Although the difference is not significant, the TF-IDF model is slightly superior to the word2vec model. This is because TF-IDF focuses on word weights based on frequency, which is more suitable for short texts with non-standard structures such as tweets.

3.6 Model Evaluation

Model assessment was done by looking at accuracy, precision, recall, and f1-score measurements. Also, a confusion matrix was created to show how the classification mistakes were spread across different sentiment categories.



(a) (b)

Figure 9. (a) Confusion Matrix TF-IDF - SVM Model; (b) Confusion Matrix Word2vec - SVM Model

Figure 9 shows the confusion matrices for each model, indicating that TF-IDF–SVM has a precision of 74% and a recall of 73%. On the other hand, Word2Vec–SVM has a precision of 72% and a recall of 68%. Both models make more mistakes when identifying negative cases because there are far fewer negative samples than positive ones. This difference in class sizes makes the models more likely to predict positive outcomes, which then hurts their ability to recall negative sentiments.

In addition to these numbers, we can also understand why TF-IDF–SVM performs better by looking at how the algorithms work. TF-IDF creates clear and detailed feature sets that highlight important words linked to sentiment, allowing the SVM to make better decisions when analyzing short and casual texts like tweets. Meanwhile, Word2Vec creates more complex vectors that focus on how words relate to each other rather than their importance, which can weaken the signals that show sentiment when combining word vectors in short texts. Therefore, even though Word2Vec does a good job of understanding relationships between words, it does not represent sentiment as effectively when dealing with uneven data distribution, which causes a lower recall for negative cases.

4. Conclusions

This study shows that the TF-IDF-SVM method produces slightly higher accuracy, namely 81%, compared to Word2Vec-SVM with 80%. These results indicate that TF-IDF is more stable and efficient for analyzing sentiment in short Indonesian texts. Data imbalance, where the positive class dominates 75% and the negative class only 25%, affects the model's performance, which tends to recognize positive sentiment. For further research, it is recommended to balance the data between classes and combine the TF-IDF and Word2Vec methods, or use a deep learning approach such as BERT to obtain more accurate results and capture deeper semantic context.

References

1. R. Hidayat and R. Ramadhan, "Peran Media Sosial Dalam Mengkonstruksi Opini Publik Terkait Kebijakan Pemerintah: Studi Kasus Wacana Publik Tahun 2025," *CORE: Journal of Communication Research*, vol. 3, no. 2, pp. 64–75, Jul. 2023.
2. A. Sitanggang, Y. Umaidah, Y. Umaidah, R. I. Adam, and R. I. Adam, "Analisis Sentimen Masyarakat Terhadap Program Makan Siang Gratis Pada Media Sosial X Menggunakan Algoritma Naive Bayes," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 3, Aug. 2024, doi: 10.23960/jitet.v12i3.4902.
3. U. Agustini, "Efektivitas dan Tantangan Kebijakan Program Makan Bergizi Gratis sebagai Intervensi Pendidikan di Indonesia," *Jurnal Kiprah Pendidikan*, vol. 4, no. 3, pp. 362–368, Jul. 2025, doi: 10.33578/kpd.v4i3.p362-368.
4. O. Sumantri Riyanto and M. Rianita Elfrida Sinaga, "Penegakan Hak Anak Atas Makanan Aman Dan Sehat: Studi Kasus Keracunan Dalam Program Makan Bergizi Gratis Ditinjau Dari Tanggung Jawab Negara," *Juris Humanity: Jurnal Riset dan Kajian Hukum HAM*, Jun. 2025.
5. R. Amelia, "Sentiment Analysis of Government Policy in Relocating the Republic of Indonesia's Capital City," *Journal of Public Administration and Government*, vol. 5, no. 2, pp. 177–187, Aug. 2023, [Online]. Available: https://jurnal.fisip.untad.ac.id/index.php/JPAG
6. D. W. Syahputra, B. Rahayudi, and L. Muflikhah, "Analisis Sentimen Twitter terhadap Kebijakan Pemberlakuan Pembatasan Kegiatan Masyarakat menggunakan Metode Support Vector Machine," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 6, no. 3, pp. 1067–1072, Mar. 2022, [Online]. Available: http://j-ptiik.ub.ac.id`
7. Y. Sibaroni, "Perbandingan Pembobotan Fitur TF-IDF dan TF-ABS Dalam Klasifikasi Berita Online Menggunakan Support Vector Machine (SVM)," *e-Proceeding of Engineering*, vol. 10, no. 3, pp. 3652–3663, Jun. 2023.
8. M. I. Syafaah and L. Lestandy, "Emotional Text Classification Using TF-IDF (Term Frequency-Inverse Document Frequency) And LSTM (Long Short-Term Memory)," *JUITA: Jurnal Informatika*, vol. 10, no. 2, pp. 225–232, Nov. 2022, [Online]. Available: https://atapdata.ai/dataset/192/HIMPUNAN_DATA_E
9. F. Rifaldy, Y. Sibaroni, and S. S. Prasetiyowati, "Effectiveness of Word2Vec and TF-IDF in Sentiment Classification on Online Investment Platforms Using Support Vector Machine," *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 10, no. 2, pp. 863–874, Mar. 2025, doi: 10.29100/jipi.v10i2.6055.

10. A. H. Dani, E. Y. Puspaningrum, and R. Mumpuni, “Studi Performa TF-IDF dan Word2Vec Pada Analisis Sentimen Cyberbullying,” *Jurnal Teknik Informatika dan Terapan*, vol. 2, no. 2, pp. 94–106, Jun. 2024, doi: 10.62951/router.v2i2.76.
11. Z. Zhan, “Comparative Analysis of TF-IDF and Word2Vec in Sentiment Analysis: A Case of Food Reviews,” *ITM Web of Conferences*, vol. 70, p. 02013, 2025, doi: 10.1051/itmconf/20257002013.
12. G. R. Ati and P. T. Prasetyaningrum, “Analysis of Community Sentiment Towards Free Nutrition Meal Programs on Twitter Using Naïve Bayes, Support Vector Machine, K-Nearest Neighbors, and Ensemble Methods,” *Journal of Information Systems and Informatics*, vol. 7, no. 2, pp. 1443–1460, Jul. 2025, doi: 10.51519/journalisi.v7i2.1098.
13. N. R. Dewi, E. Y. Puspaningrum, and H. Maulana, “Analisis Sentimen Tweet Vaksinasi COVID-19 Menggunakan RNN Dengan Metode TF-IDF Dan Word2Vec,” *Jurnal Informatika dan Sistem Informasi (JIFoSI)*, vol. 3, no. 1, pp. 56–65, Apr. 2022.

COMPLETE
Journal of Computer, Electronic, and Telecommunication
ISSN 2723-4371, E-ISSN 2723-5912

Article

IoT-Based Lithium-Ion Battery Safety Condition Monitoring Using Voltage and Temperature Parameters with Fuzzy Logic Method

Ardika Bagus Saputra [1], Susijanto Tri Rasmana [2], and Isa Hafidz [3]*

[1,3] Electrical Engineering Study Program, Telkom University, Surabaya Campus, Surabaya 60231, East Java, Indonesia
[2] Computer Engineering Study Program, Telkom University, Surabaya Campus, Surabaya 60231, East Java, Indonesia
*Correspondence: isahafidz@telkomuniversity.ac.id

Received: 15 December 2025; Revised: 29 December 2025; Accepted: 30 December 2025; Published: 31 December 2025

Abstract: The increasing number of vehicles in Indonesia has a negative impact if it is allowed to continue; one solution is to switch to environmentally friendly electric vehicles. One of the main components of electric vehicles is the battery. IoT-based battery monitoring is needed to check battery health using the fuzzy logic method with LED notification results. Notifications indicate whether the battery is in a safe, alert or dangerous condition. The results of battery monitoring in a discharge condition with a load of battery voltage decrease linearly compared to temperature, which changes according to the ambient temperature. In a charging condition, the battery voltage increases linearly, and the peak temperature is higher than when discharged with a load. When testing a load variation of 4 LEDs, the voltage decreases faster, and the temperature rises higher than with a load of 2 LEDs. With these various conditions, the application of fuzzy logic is used to monitor the condition of the battery in a discharge or charge. This study aims to develop IoT-based battery monitoring by implementing the use of fuzzy logic as a battery notification when the battery is in a safe, alert, or dangerous condition. Experimental testing was conducted under charging and discharging conditions with varying loads. The calibration results show that the voltage measurement using the INA219 sensor has an average error of 0.35% during discharge and 0.31% during charging, while the temperature measurement using the NTC thermistor shows an average error of 0.85%. The fuzzy logic system successfully classifies battery safety states based on voltage–temperature combinations, and the notification results are consistent with MATLAB-based fuzzy simulations. The proposed system demonstrates reliable real-time monitoring of lithium-ion battery safety conditions and can serve as an effective early warning mechanism to prevent unsafe battery operation.

Keywords: Battery Monitoring; Voltage; Temperature; Fuzzy Logic; IoT.

1. Introduction

Technological developments in Indonesia are developing very rapidly, one of which is in the transportation sector. There are various types of transportation in Indonesia; one of the most common types of vehicles is the private vehicle. The number of private vehicles in Indonesia is very large. According to the Central Statistics Agency (BPS), the number of motor vehicles in Indonesia in 2021 reached 141,992,573 units, and this number continues to grow. [1]. This increasing number of vehicles has a negative impact if left unchecked. The impact of fuel-powered vehicles can cause air pollution from exhaust emissions from motor vehicles [2]. A solution to this problem has been found: switching to electric vehicles, which are more environmentally friendly because they do not produce exhaust gases. Automotive companies have also begun researching these electric vehicles [3].

One of the most important components of an electric vehicle is the battery as a power storage. To maintain the battery in good condition, a battery monitoring system was created [4]. The battery monitoring system used in this study is a lithium-ion battery [5]. This type of lithium-ion battery was chosen because it has several advantages, namely, stable voltage, relatively small size compared to other types of batteries, light weight, and relatively long battery life. Although lithium-ion batteries have many advantages, the electrochemistry of lithium-ion batteries is quite complex. Overcharging and discharging will reduce the performance and life of lithium-ion batteries, and even repeated excessive use can cause explosions. Therefore, a battery monitoring system is needed to determine the health of lithium-ion batteries. By monitoring the battery, the condition of the lithium-ion battery can be known when the electric power is running out, which can prevent over-discharge and extend the life of the lithium-ion battery. To facilitate monitoring of lithium-ion batteries, an IoT-based battery monitoring system was created. With the use of IoT, it is easy to find out the current, voltage, and temperature parameters that can be displayed on mobile devices. This IoT system utilizes an internet connection to send lithium-ion battery parameter data to a web server in the cloud, which then displays the data in the software [6]. In previous research, the characteristics of Lithium-Ion batteries were tested using the Fuzzy Method with Varying Loads using voltage and temperature parameters as variables in fuzzy logic[7]. The second previous study, the implementation of fuzzy logic for accurate measurement of the SoC of electric car batteries, used voltage, current, and temperature as fuzzy variables to estimate the state of charge of electric car batteries[8]. Fuzzy logic is used to generate notifications from battery monitoring results. From the monitoring results, two parameters are obtained: voltage and temperature, which are used to analyze the condition of the lithium-ion battery.

2. Materials and Methods

2.1. Materials

2.1.1. Lithium-Ion Battery

Figure 1. Lithium-Ion Batteries

This study used a secondary battery type, namely a lithium-ion battery. When used, it works as a voltaic cell: Lithium will deliver electrons from the anode to devices that require electrons, such as capacitors and processors in cellphones or laptops, then ending at the cathode. Meanwhile, protons from the cathode enter through the separator between the anode and cathode (intercalation process). This process continues continuously until the battery's capacity is exhausted (indicated by a line or percentage of battery capacity on the cellphone or laptop screen). When charged, it works as an electrolysis: Meanwhile, when the battery is recharged, electrons will return from the cathode to the anode through the charger, and with the help of the incoming current from the charger, the protons will return to the cathode. So that the condition returns to its original state

2.1.2. NTC Thermistor

Figure 2. NTC thermistor

The NTC (Negative Temperature Coefficient) thermistor is a resistor whose value its resistance decrease when temperature increased. Due to its highly sensitive nature to change temperature, many NTC thermistors used as a temperature sensor in the circuit electronics.

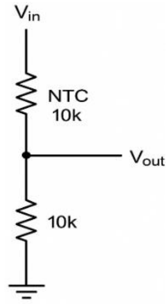


Figure 3. Voltage Divider Schematic

For read change NTC resistance, used voltage divider circuit, where the NTC 10K is in series with a 10K ohm resistor. The voltage at the point middle divider voltage connected to the analog input of the microcontroller. The voltage divider circuit formula is as follows:

$$V_{out} = V_{cc} \times \frac{R_2}{R_1 + R_2} \quad (1)$$

Information:

- R1 (NTC) varies in accordance temperature
- R2 (10k ohm)
- Vout will increase as the temperature decreases (because R1 increases)
- Vout will decrease as temperature increases (because R1 decreases)

2.1.3 INA219

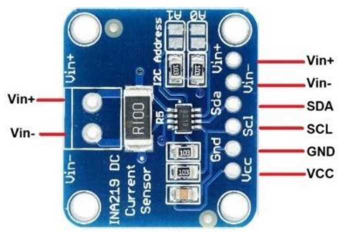


Figure 4. INA219

The INA219 is a sophisticated electronic sensor module that can monitor current and voltage. With a little multiplication, we can even measure power. The INA219 can monitor voltages up to 26 volts DC. Its current measurement range, +/- 3.2A, is suitable for most small measurements. In other words, this module can measure power above 75 watts. Texas Instruments introduces the INA219 shunt current sensor module. This module is a bidirectional, driftless power monitor that measures shunt voltage, bus voltage, current, and power. This module communicates with a microcontroller through a built-in interface that supports I2C and SMBus.

2.2. Design Battery Monitoring System

In the research This battery tested in Charging and discharging conditions. Current sensors, voltage sensors, and temperature sensors detect how much battery value is being monitored. Then the data value from each sensor enters the ESP8266 microcontroller and the data is processed to be displayed via TFT LCD and IoT system that can be seen in the blynk software. Before the data is displayed, the data goes through a fuzzy logic process with predetermined variables. The data results displayed are the status of current, voltage, temperature and notification in the form of LEDs with statuses that have been determined through the fuzzy logic process.

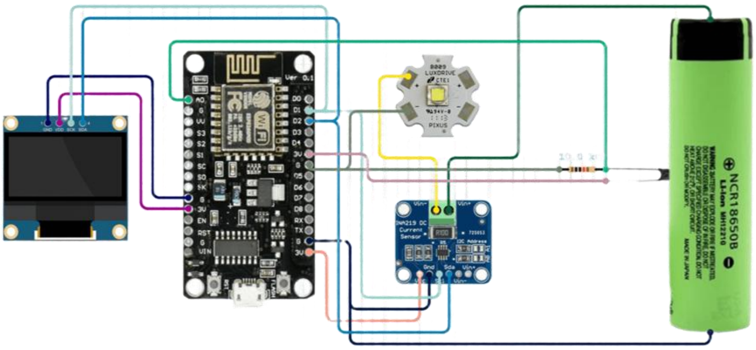


Figure 5. Design Battery Monitoring System

2.3. Fuzzy Logic

Fuzzy logic is an approach to variable processing that allows multiple possible truth values to be processed through the same variable. Fuzzy logic attempts to solve problems with an open and imprecise spectrum of data and heuristics, allowing for a range of accurate conclusions. Fuzzy logic is designed to solve problems by considering all available information and making the best decision based on that input. Fuzzy logic was first proposed by Lotfi Zadeh in a 1965 paper for the journal Information and Control. In his paper, "Fuzzy Sets," Zadeh attempted to reflect on the types of data used in information processing and derive elemental logic rules for such sets.

In the experiment This done simulation For get results notification with enter mark voltage and temperature battery during the charging and discharging process to in system fuzzy logic that has been prepared. For make system notification with fuzzy logic, need made function membership function and rules logic (rule based) based on the data obtained from voltage sensor and temperature sensor as input while the output in the form of led notification as output.

Table 1. Fuzzy variables

No	Function	Variables	Fuzzy Set	Domain
1	Input	Temperature	Cold	[26, 27, 28]
2			Normal	[27, 30, 33]
3			Hot	[32, 35, 39]
4		Voltage	Very Low	[2.5, 2.8, 3.0]
5			Low	[2.8, 3.1, 3.4]
6			Normal	[3.3, 3.6, 3.9]
7			Tall	[3.7, 4, 4.2]
8	Output	Notification	Safe	[0, 1, 3]
9			Alert	[3, 5, 7]

10	Danger	[7, 9, 10]

In this study, a battery monitoring system monitors the current, voltage, and temperature of a lithium-ion battery using a fuzzy system and the Mamdani fuzzy method. The input to the fuzzy controller is voltage and temperature data, while the output is a notification in the form of an LED light.

Table 2. Fuzzy Rules

No	Input 1 (Voltage)	Input (Temperature)	Output (Notification)
1	Very low	Cold	Alert
2	Very low	Normal	alert
3	Very low	Hot	Danger
4	Low	Cold	Alert
5	Low	Normal	Safe
6	Low	Hot	Danger
7	Normal	Cold	Safe
8	Normal	Normal	Safe
9	Normal	Hot	Alert
10	Tall	Cold	Safe
11	Tall	Normal	Alert
12	Tall	Hot	Danger

2.4 IoT

The term IoT, or Internet of Things, refers to the collective network of connected devices and the technologies that facilitate communication between them and the cloud, as well as between the devices themselves. Thanks to the advent of inexpensive computer chips and high-bandwidth telecommunications, a wide variety of devices are now connected to the internet. Examples include toothbrushes, vacuum cleaners, cars, and machines that use sensors to collect data and intelligently respond to users.

The Internet of Things (IoT) is integrating everyday objects with the internet. Computer engineers have been adding sensors and processors to everyday objects since the 1990s. However, progress was initially slow because the chips were large and took up a lot of space. Low-power computer chips called RFID tags were first used to track expensive equipment. As computer devices have shrunk, these chips have become smaller, faster, and smarter over time.

3. Results

3.1. Battery Monitoring System Design Using Fuzzy Logic

This battery monitoring system uses several components, such as voltage sensors and temperature sensors, and an ESP8266 microcontroller. The output is displayed on a TFT LCD and through IoT-based Blynk software. The design is shown in Figure 6 below.

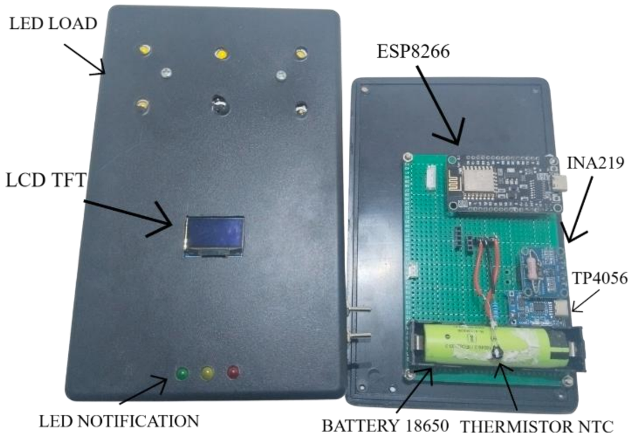


Figure 1. Battery Monitoring System Implementation

The hardware made in the picture is composed of a battery, a voltage sensor, a temperature sensor, an ESP8266 microcontroller, a TFT LCD, Blynk software, and an LED notification using the fuzzy logic method.

3.2. Prototype Implementation

3.2.1. Hardware Design Flowchart

The following is a flowchart for designing the hardware of the tool, which can be explained in the image below.

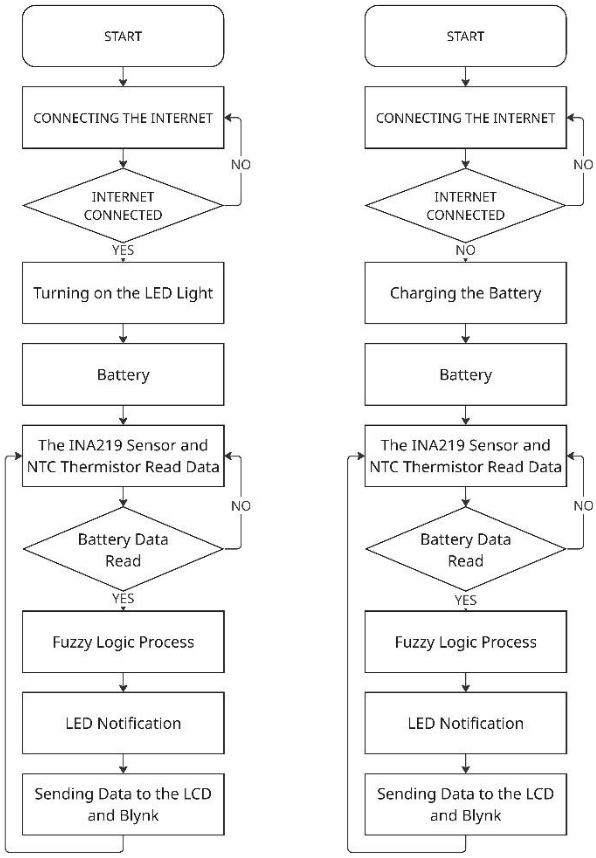


Figure 2 Battery Monitoring System Flowchart

In the first stage, the ESP8266 connects to the internet to be able to send data. After successfully connecting to the internet, then 2 tests will be carried out, namely discharge by connecting the LED light load to the battery, then the second test charge by charging the battery. After that, the sensor reads the battery, namely the voltage sensor and the temperature sensor. If the voltage and temperature data have not been read, the sensor will re-read the battery. Then the data from the INA219 sensor and the NTC Thermistor sensor will be processed into fuzzy logic, which first goes through the fuzzyfication process. After that, it enters the inference stage based on the fuzzy rules that have been created, and then the defuzzification process to make the right decision. The results of the fuzzy logic process will turn on 3 LED lights in green, yellow, and red, used as notifications, meaning green as safe, blue as alert, and red as danger. Finally, the data read from the INA219 sensor and the NTC Thermistor sensor will be displayed on the TFT LCD and the Blynk software connected to the internet.

3.2.2. INA219 Sensor Calibration

This battery voltage test compares the results from a multimeter with the test equipment that has been made. Battery voltage data is taken per hour, the data in table 3 shows the battery in a discharge condition with a load of 2 3-watt LED lights with the initial condition of the battery being fully charged until the LED lights are lit in a dim condition. The results of the test comparing the discharge voltage data obtained from the voltage sensor and multimeter show an average error of 0.36%.

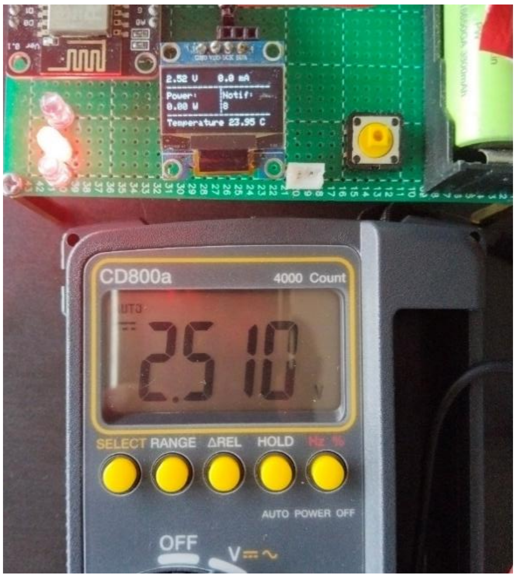


Figure 3. Calibration of Sensors and Multimeter Measurement Tools

Table 3. Calibration of the voltage sensor with a discharge condition measuring instrument

Tool Test Voltage (Volt)	Multimeter (Volt)	Error (%)
4.03	4.04	0.24
3.8	3.81	0.26
3.71	3.72	0.26
3.56	3.58	0.55
3.45	3.46	0.28
3.37	3.38	0.29
3.15	3.16	0.31
2.86	2.88	0.69
2.68	2.69	0.37
2.5	2.51	0.39
Average error		0.36

3.2.3. NTC Thermistor Sensor Calibration

Battery monitoring testing uses an INA219 sensor and an NTC thermistor sensor. The INA219 sensor is used to measure voltage, and the NTC thermistor sensor is used to measure battery temperature. This temperature sensor circuit consists of a 10k NTC thermistor and a 10k resistor connected in series. This test was conducted to compare the measurement results of the test equipment that has been made with a temperature multimeter. Battery temperature data was taken hourly in a discharged battery condition with a load of two 3-watt LED lights. In Table 4, the comparison of test equipment data with a temperature multimeter shows results that are not much different, with an average error of 0.85%.

Table 4. Temperature Sensor Calibration with Measuring Instruments

Temperature (Test tool)	Temperature (Celsius)	Error (%)
28.93	29	0.24
35	35.78	2.17
36.02	36	0.05
34.87	35	0.37
32.84	33	0.48
30.54	30	1.8
Average error		0.85

3.3. System Testing

3.3.1. Battery Testing under Discharging and Charging Conditions

Testing the battery in discharge condition using a load of 2 3 watt LED lights. Battery testing uses an INA219 sensor. In Figure IV.5, the graph of the discharge battery voltage is at 3.92 volts and the load power of 2 LED lights is at 4.02 watts. The battery capacity is discharged with a load until the LED lights are dim, after several battery tests the lower limit of the voltage is at 2.5 volts and the upper limit of the discharge load voltage is at 4 volts. Temperature sensor testing uses an NTC thermistor sensor, monitoring the battery temperature by bringing the thermistor close to the battery being tested. The temperature is at its peak in the first 60 minutes of testing at 38o Celsius, then the temperature decreases over time until the light load is dim and the battery voltage is at the cut-off limit.

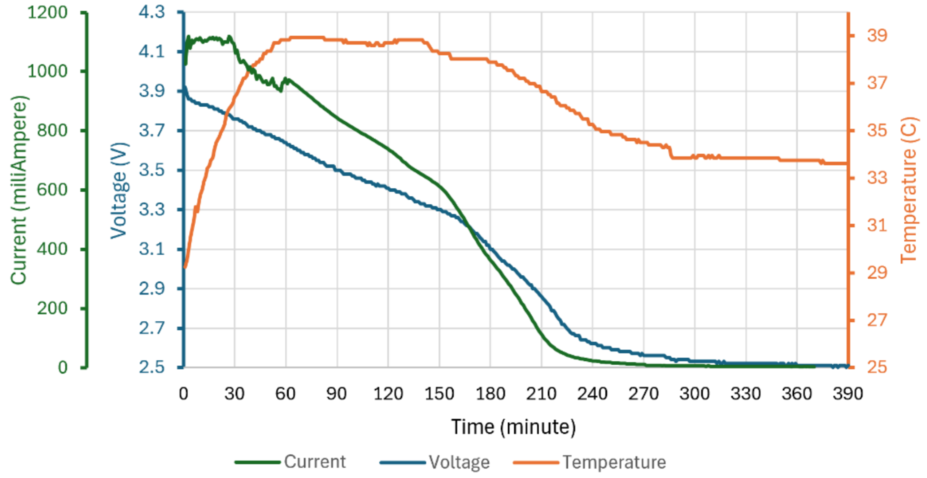


Figure 9. Battery Discharge Condition Graph

Real-time battery discharge condition monitoring data is captured every 30 minutes. Battery data shows the battery voltage from a fully charged state of 3.92 volts to a lower limit of 2.52 volts. Battery temperature increases during the first 1-2 hours of testing, peaking at 38.83°C, then gradually decreases. The results of the battery voltage and temperature measurements are processed into fuzzy logic that determines notifications on the battery condition during the test.

Table 5. Real-Time Battery Discharge Condition Monitoring Data

No.	Time (Minute)	Voltage (Volt)	Temperature (Celsius)	Fuzzy Logic Output	Notification
1	1	3.92	29.25	5	LED YELLOW ON
2	30	3.76	36.41	6.51	LED YELLOW ON
3	60	3.64	38.83	5.5	LED YELLOW ON
4	90	3.5	38.83	5	LED YELLOW ON
5	120	3.41	38.72	5	LED YELLOW ON
6	150	3.3	38.25	8.66	LED RED ON
7	180	3.1	37.9	8.91	LED RED ON
8	210	2.86	36.64	8.81	LED RED ON
9	240	2.62	35.28	8.92	LED RED ON
10	270	2.56	34.51	8.86	LED RED ON
11	300	2.53	33.85	8.78	LED RED ON
12	330	2.52	33.85	8.78	LED RED ON

Next, the battery is tested in charging condition using the tp4056 module which has a power output of 5 volts 1 ampere. In figure IV. 6 the initial voltage is at 2.5 volts which corresponds to the lower limit of the battery voltage when in discharge condition with load. On the graph the voltage increases from the beginning of charging until 330 minutes or about 5 hours 30 minutes. Charging voltage is at 4 volts, the voltage is stable until full charging is at 4.24 volts which is indicated by a blue indicator on the tp4056 module. Then the battery temperature is in charging condition, on the graph the peak battery temperature is at 39.9˚ Celsius 3 hours after the temperature decreases, this happens because the battery adjusts from the voltage and current entering from the tp4056 module power, then slowly the temperature continues to decrease until the battery is fully charged.

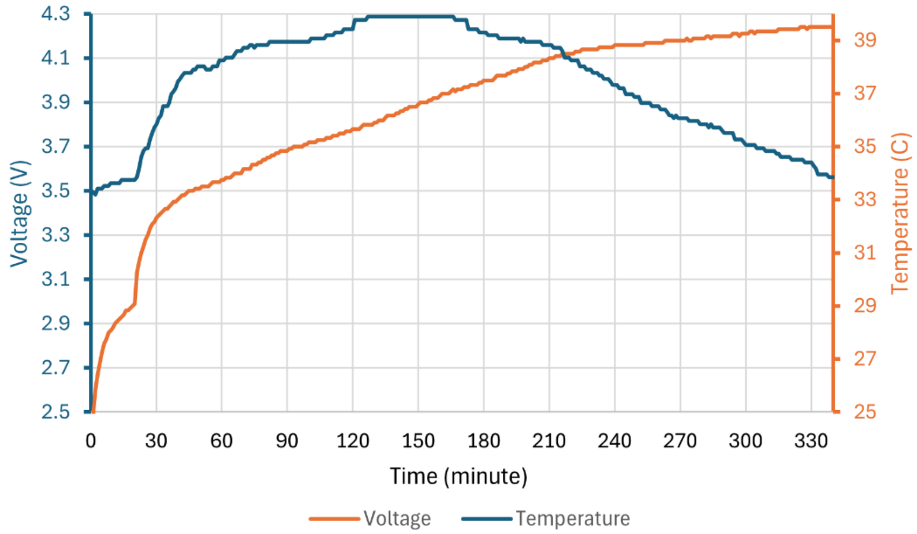


Figure 10. Battery Charge Condition Graph

Real-time battery charge condition monitoring data is captured every 30 minutes. Battery data shows the battery voltage from an empty battery condition of 2.49 volts to a fully charged battery at 4.24 volts. Battery temperature increases at the beginning of the 1-2 hours of testing, peaking at 39.9°C, then the temperature slowly decreases. The results of the battery voltage and temperature measurements are processed into fuzzy logic that determines notifications on the battery condition during the test.

Table 6. Real-Time Battery Charging Condition Monitoring Data

No.	Time (minute)	Voltage (Volt)	Temperature (Celsius)	Fuzzy Logic Output	Notification
1	1	2.49	33.3	8.7	LED RED ON
2	30	3.38	35.85	5.62	LED YELLOW ON
3	60	3.55	38.25	5	LED YELLOW ON
4	90	3.68	38.95	5.87	LED YELLOW ON
5	120	3.78	39.42	6.71	LED YELLOW ON
6	150	3.89	39.9	8.6	LED RED ON

No.	Time (minute)	Voltage (Volt)	Temperature (Celsius)	Fuzzy Logic Output	Notification
7	180	4	39.3	8.91	LED RED ON
8	210	4.1	38.83	8.92	LED RED ON
9	240	4.16	37.32	8.92	LED RED ON
10	270	4.18	36.07	8.92	LED RED ON
11	300	4.21	35.06	8.92	LED RED ON
12	330	4.24	34.40	8.85	LED RED ON

3.3.3. Battery Monitoring System Testing with Fuzzy Logic Implementation

This battery monitoring test was conducted by applying fuzzy logic with the voltage and temperature of the battery as input variables, and the output is a notification in the form of a green LED light for safety, a yellow for caution, and a red for danger. The purpose of applying fuzzy logic to this battery monitoring system is to determine the health condition of the battery when the battery is charging and discharging with or without a load, such as when the battery voltage is above or below the normal limit, and the battery temperature changes when the battery is charging.

Table 7. Application of fuzzy logic in monitoring battery charge conditions

No.	Voltage (V)	Temperature (Celsius)	Notification	Test Output	MATLAB Output	Error (%)
1	2.60	28.86	Alert (yellow LED on)	5	5	0
2	2.76	29.32	Alert (yellow LED on)	5	5	0
3	3.32	33.52	Danger (red LED on)	7	7.72	9.32
4	2.87	29.21	Alert (yellow LED on)	4	4.01	0.24
5	2.94	29.09	Safe (green LED is on)	3	2.97	1.01
6	3.32	33.52	Danger (red LED on)	7	7.72	9.32
7	3.45	27.55	Safe (green LED is on)	1	1.26	20.63
8	3.33	28.52	Safe (green LED is on)	1	1.36	26.47
9	3.77	36.38	Alert (yellow LED on)	6	5	20
10	3.82	27.44	Safe (green LED is on)	3	2.29	31.01
11	4.14	31.28	Alert (yellow LED on)	4	5	20
12	4.10	33.32	Danger (red LED on)	8	8.73	8.36
			Average error			12.19

The explanation of the table above can be explained as follows:

1. When the battery is charged, the voltage shows 2.60 volts and the temperature is 26.11 degrees Celsius, and the notification results show a red LED as a danger sign.
2. When the battery is charged, the voltage shows 2.50 volts and the temperature is 29.08 degrees Celsius, and the notification results show a yellow LED as a warning sign.
3. When the battery is charged, the voltage shows 2.70 volts and the temperature is 30.86 degrees Celsius, and the notification results show a yellow LED as a warning sign.
4. When the battery is charged, the voltage shows 30.86 volts and the temperature is 26.00 degrees Celsius, and the notification results show a red LED as a danger sign.
5. When the battery is charged, the voltage shows 2.94 volts and the temperature is 29.09 degrees Celsius, and the notification results show a green LED as a safe sign.
6. When the battery is charged, the voltage shows 3.32 volts and the temperature is 34.52 degrees Celsius, and the notification results show a red LED as a danger sign.
7. When the battery is charged, the voltage shows 3.40 volts and the temperature is 32.52 degrees Celsius, and the notification results show a green LED as a safe sign.
8. When the battery is charged, the voltage shows 3.43 volts and the temperature is 27.55 degrees Celsius, and the notification results show a green LED as a safe sign.
9. When the battery is charged, the voltage shows 3.45 volts and the temperature is 33.55 degrees Celsius, and the notification results show a yellow LED as a warning sign.
10. When the battery is charged, the voltage shows 3.82 volts and the temperature is 29.57 degrees Celsius, and the notification results show a green LED as a safe sign.
11. When the battery is charged, the voltage shows 4.11 volts and the temperature is 32.64 degrees Celsius, and the notification results show a yellow LED as a warning sign.
12. When the battery is charged, the voltage shows 4.14 volts and the temperature is 35.90 degrees Celsius, and the notification results show a red LED as a danger sign.

Table 8. Application of Fuzzy Logic in Monitoring Battery Discharge Conditions

No.	Voltage	Temperature	Notification	Test Output	Fuzzy Logic Output	Error
1	4.24	34.18	Danger (red LED on)	8	8.85	9.60
2	3.88	29.89	Alert (yellow LED on)	4	4.65	13.97
3	3.78	28.58	Safe (green LED is on)	3	3.4	11.76
4	3.74	32.52	Alert (yellow LED on)	4	5.08	21.2
5	3.53	27.67	Safe (green LED is on)	1	1.32	24.24
6	3.41	28.12	Safe (green LED is on)	1	1.3	23.07
7	3.32	32.98	Danger (red LED on)	7	7.59	8.42
8	3.20	28.24	Safe (green LED is on)	1	1.28	21.87
9	2.87	28.87	Alert (yellow LED on)	4	3.99	0.25
10	2.95	34.98	Danger (red LED on)	7	8.76	19.07
11	2.80	29.04	Alert (yellow LED on)	5	5	0

No.	Voltage	Temperature	Notification	Test Output	Fuzzy Logic Output	Error
12	2.50	27.67	Alert (yellow LED on)	6	5	20
			Average error			14.45

The description in the table above can be explained as follows:
1. In the condition of battery discharge, the voltage shows 4.24 volts, and the temperature is 34.18 degrees Celsius, the result is a notification showing a red LED as a danger sign.
2. In the condition of battery discharge, the voltage shows 3.88 volts, and the temperature is 29.89 degrees Celsius; the result is a notification that shows a yellow LED as a warning sign.
3. In the condition of battery discharge, the voltage shows 3.78 volts, and the temperature is 29.08 degrees Celsius; the notification results show a green LED as a safe sign.
4. In the condition of battery discharge, the voltage shows 3.74 volts, and the temperature is 32.52 degrees Celsius; the result is a notification that shows a yellow LED as a warning sign.
5. In the condition of battery discharge, the voltage shows 3.53 volts, and the temperature is 27.67 degrees Celsius; the notification results show a green LED as a safe sign.
6. In the condition of battery discharge, the voltage shows 3.41 volts, and the temperature is 28.12 degrees Celsius; the notification results show a green LED as a safe sign.
7. In the condition of battery discharge, the voltage shows 3.35 volts, and the temperature is 32.98 degrees Celsius; the result is a notification showing a red LED as a danger sign.
8. In the condition of battery discharge, the voltage shows 3.20 volts , and the temperature is 28.24 degrees Celsius; the notification results show a green LED as a safe sign.
9. In the condition of battery discharge, the voltage shows 3.23 volts, and the temperature is 25.14 degrees Celsius, the result is a notification that shows a yellow LED as a warning sign.
10. In the condition of battery discharge, the voltage shows 2.83 volts, and the temperature is 32.07 degrees Celsius, the result is a notification that shows a yellow LED as a warning sign.
11. In the condition of battery discharge, the voltage shows 2.80 volts, and the temperature is 29.04 degrees Celsius; the result is a notification that shows a yellow LED as a warning sign.
12. In the condition of battery discharge, the voltage shows 2.49 volts, and the temperature is 27.56 degrees Celsius, the result is a notification that shows a red LED as a danger sign.

4. Discussion

From the results of calibration tests with a digital multimeter, the INA219 sensor in measuring voltage in discharge conditions shows an average error of 0.35%, while in charge conditions, the average error is 0.97%. Then the results of the NTC thermistor sensor calibration in measuring battery temperature show an average error of 0.85% when compared with a digital thermometer measuring instrument. From the results of the calibration test, the INA219 sensor has good accuracy in measuring battery voltage in charge and discharge conditions, and the NTC thermistor is able to make accurate temperature measurements.

The system implementation involves several processes. First, the battery is determined to be charging or discharging. Then, the two sensors read the voltage and temperature data for the battery being measured. The obtained data are then entered into the fuzzy logic stage, with voltage and temperature as input variables. Then, in the inference stage, fuzzy rules are created to determine which rules should be applied. The final defuzzification stage is the final decision to determine the output, which will be a red, yellow, or green LED notification.

The battery monitoring system can read the battery voltage and temperature in discharge and charge conditions. In the discharge battery condition with a load of 2 LEDs, the battery voltage is at 3.92 volts with a current load of 1025.9 milliamperes and a temperature of 29.25 degrees Celsius. On the voltage and current graph, it decreases linearly until 150 minutes, the voltage is at 3.3 volts and the current load is at 643.2 milliamperes. The battery temperature in minutes 60-150 is at its highest temperature peak at 38.95 degrees Celsius. Then the voltage drops again until in minute 240 the battery voltage is at 2.62 volts and the current is at 38 milliamperes. After 2 hours of testing, the graph forms a straight line which means the battery voltage has reached the lower limit and the current load is getting smaller. This shows that the battery is working optimally, the battery is able to supply power to the load within 2 hours and the temperature increases in the first 150 minutes which means the battery is working hard when supplying power to the LED lights. Afterward, the battery voltage was low enough to supply the load, and the LED light was dim.

The next test involved charging the battery. The battery was tested after the discharge test, with the battery voltage at 2.5 volts. The battery was charged using a TP4056 module with a 5 volt, 1 ampere specification. The graph shows a spike in battery voltage at 3.13 volts after 25 minutes, then a linear increase in battery voltage during charging. This occurs because the battery is low on power, so when it is charged, the current from the charger is immediately drawn into the battery. As the battery charges, the battery temperature also increases over 25 minutes, peaking at 39.90°C at 127-166 minutes. The temperature then decreases until charging is complete, when the battery voltage increases linearly during charging.

Table 9. Fuzzy Logic-based Battery Monitoring Testing

Voltage (Volt)	**Temperature (Celsius)**	**Notification**	**Tool Output**	**MATLAB Output**	**Error (%)**
3.76	36.41	Alert (Led Yellow On)	6.51	6.58	1.06

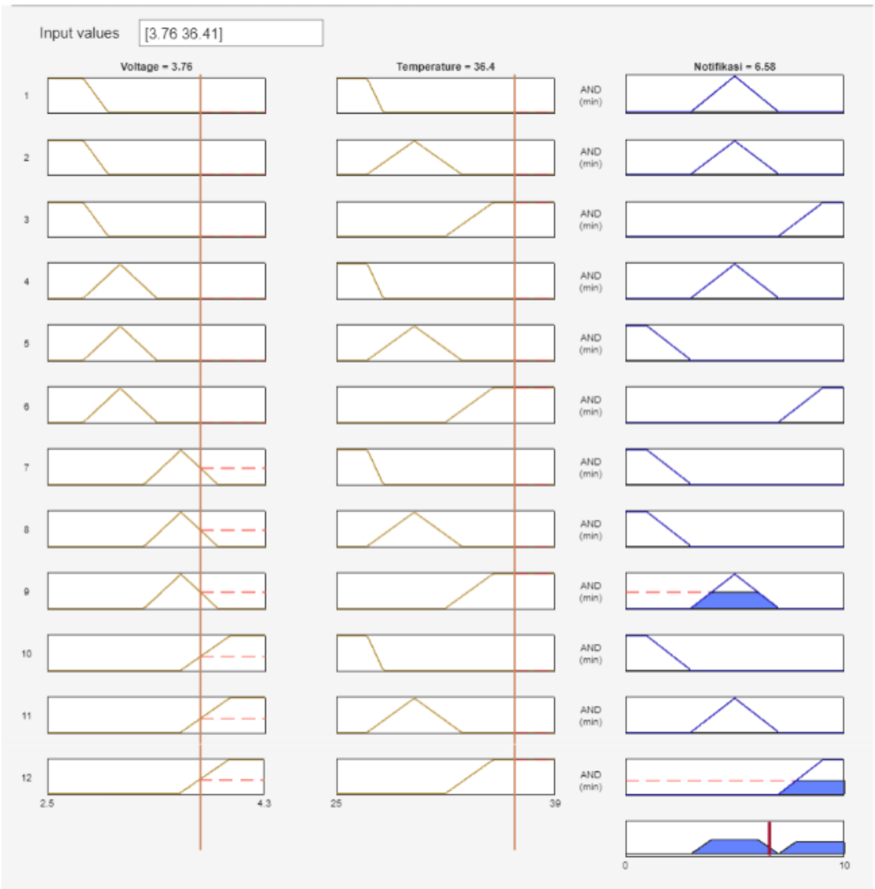


Figure 11. Fuzzy Logic Comparison

The fuzzy logic process, between the test tool and MATLAB, yielded quite satisfactory results with an average error value of 1.06% under discharge conditions. However, the results of this relatively small difference in the comparison of the tool and MATLAB tests did not affect the LED notification results as a warning on battery monitoring. In all tests in the fuzzy rules, the tool and MATLAB test results showed the same LED notification. So, even though there is a fairly large average error in the tool and MATLAB tests, the notification results remain the same and the overall system performance works well.

5. Conclusions

This research focuses on the development of a battery monitoring system using IoT-based fuzzy logic with INA219 sensors and NTC thermistors. The purpose of this research is to prevent excessive battery use and maintain battery health. This system is designed to show battery performance in real time by displaying voltage, temperature, and the use of fuzzy logic to get notification results as battery warnings when in discharge or charge conditions. The test results show that the system is able to work well in monitoring battery voltage and temperature accurately, and fuzzy logic successfully displays notifications. By using fuzzy logic, this system can make decisions to determine battery warning notifications in safe, alert, and dangerous conditions. In addition, the integration between the ESP8266 microcontroller, INA219 sensor, and NTC thermistor proves that the system can monitor the battery automatically. Several areas of this research method require further optimization to achieve more satisfactory results. Based on this research, a more complex protection system can be considered to maintain the output, prevent reverse current, and provide greater safety. Second, a sensor module with higher specifications than the INA219 sensor can be used to monitor batteries with higher voltages and currents.

References

1. Badan Pusat Statistik, “Perkembangan Jumlah Kendaraan Bermotor Menurut Jenis (Unit), 2019-2021.” Accessed: Nov. 16, 2023. [Online]. Available: https://www.bps.go.id/indicator/17/57/1/jumlah-kendaraan-bermotor.html
2. CNBC Indonesia, “10 Negara Konsumsi Minyak Bumi Terbesar Dunia, Ada RI?” Accessed: Dec. 05, 2023. [Online]. Available: https://www.cnbcindonesia.com/research/20231010123608-128-479375/10-negara-konsumsi-minyak-bumi-terbesar-dunia-ada-ri
3. Kemenperin, “Studi Mobil Listrik: Hemat Energi Hingga 80 Persen.” Accessed: Dec. 15, 2023. [Online]. Available: https://kemenperin.go.id/artikel/19877/Studi-Mobil-Listrik:-Hemat-Energi-Hingga-80-Persen
4. S. Gupta and R. Perveen, “Fuel cell in electric vehicle,” *Mater Today Proc*, vol. 79, pp. 434–437, Jan. 2023, doi: 10.1016/J.MATPR.2023.02.039.
5. A. A. M. Kurniawan *et al.*, “Fuzzy logic method for making push notifications on monitoring system of IoT-based electric truck charging,” *Bulletin of Electrical Engineering and Informatics*, vol. 14, no. 1, pp. 118–131, Feb. 2025, doi: 10.11591/eei.v14i1.7412.
6. J. Teknik Elektro, T. Puji Cahyono, U. Jember, and U. Jember Bambang Sri Kaloko, “Pengujian Karakteristik Baterai Lithium-Ion Dengan Metode Fuzzy dengan Beban Bervariasi.”
7. V. Ramakrishnan, M. Ram Kumar, D. Sharoun, and A. S. Vigneshwar, “IoT Based Battery Monitoring System for E-Vehicle,” Advances in Materials, Computing and Communication Technologies In Association withInternational Journal of Scientific Research in Science and Technology, vol. 9, Apr. 2021, [Online]. Available: www.ijsrst.com
8. M. Dzaky *et al.*, “Implementasi Logika Fuzzy untuk Pengukuran SoC Baterai Mobil Listrik yang Akurat (Fuzzy Logic Implementation for Accurate Electric Car Battery SoC Measurement),” 2021.

COMPLETE
Journal of Computer, Electronic, and Telecommunication
ISSN 2723-4371, E-ISSN 2723-5912

Article

Basic Application of Business Intelligence in making E-Commerce Buyer Dashboard Indonesian Using Tableau Tools Development

Uya Asy Syuura Anandri[1*], Abdullah[2]

[1, 2] Information System, Faculty of Engineering and Computer Science, Islamic University of Indragiri, Indonesia
* Correspondence: uyaanandri@unisi.ac.id

Received: 24 December 2025; Revised: 29 December 2025; Accepted: 30 December 2025; Published: 31 December 2025

Abstract: This research examines the implementation of Business Intelligence (BI) for the creation of an Indonesian e-commerce buyer dashboard in 2024 with the aim of increasing the visibility of operational KPI and demonstrating a reproducible pipeline from data cleaning to visualization. The main issues addressed are the quality of order-level data (provincial writing variants, date format, numerical values, and PII anonymization) as well as the need to calculate buyer metrics (unique buyers, repeat rate) which is rarely ava1ilable in public aggregate data. The methods used include: (1) data cleaning and harmonizing using OpenRefine; (2) numerical transformation and validation with Python (pandas); (3) creating interactive worksheets and dashboards in Tableau (sales map per province; monthly trend line; bar with avg sales per product; sales pie by gender); and (4) sensitivity analysis to assess the impact of cleaning step variation on buyer-level KPI. Using the order-level dataset of cleaning results (1,000 transactions), a total revenue of Rp 2,298,975,000, 1,000 orders, and 178 unique buyers were found; seasonal patterns were seen with a peak in the fourth quarter and revenue concentration in urban areas (DKI Jakarta, West Java). The top-10 products contribute a significant portion of revenue, and repeat buyers show an important role in the sales structure

Keywords: Business Intelligence; Dashboard; Data Science; E-Commerce; OpenRefine; Tableau

1. Introduction

The development of the digital economy in Indonesia over the past decade has driven rapid growth in the e-commerce sector. The adoption of marketplace platforms and online sales channels by micro, small, and medium enterprises (MSMEs) and large retailers has created a vast and heterogeneous flow of transaction data [1]. If the data is processed and analyzed properly, it will have great potential to support faster and evidence-based business decision making (data-driven decision making) [2]. However, in practice, many business actors have not utilized their transaction data optimally due to limited technical capabilities, data integration barriers, and a lack of adequate analytical infrastructure [3].

Business Intelligence (BI) has emerged as a framework and toolkit for transforming raw data into useful information [4]. Through ETL (Extract–Transform–Load) stages, data cleansing, business metric aggregation, and interactive visualization, BI facilitates stakeholders in observing trends, identifying anomalies, and formulating operational policies and marketing strategies [5]. Tableau, as a leading data visualization platform, provides interactive dashboard creation capabilities that make it easy for non-technical users to explore data and monitor key performance indicators (KPIs) in real time or periodically.

In the context of e-commerce, some crucial KPIs include total revenue, number of unique buyers, sales trends per period, performance per product, and buyer demographic segmentation. Analysis of these KPIs helps management determine inventory strategies, promotions, marketing channels, and customer retention activities [6]. Furthermore, a regional (province/city) sales distribution map allows for prioritization of logistics and market expansion. While the benefits are clear, a key challenge often faced is data quality: inconsistent date formats, location typos, numeric values mixed with text, and duplicate rows require rigorous data cleaning before integration into the BI pipeline [7].

E-commerce in Indonesia is growing rapidly and generating large transaction volumes. Processing transaction data into easily understandable information is key for businesses to make operational and strategic decisions [8]. Business Intelligence (BI) provides methods and tools to transform raw data into actionable insights. This study implements a BI pipeline (ETL → cleansing → aggregation → visualization) to create an e-commerce buyer dashboard using Tableau with transaction datasets (cleaning results) available using OpenRefine.

This research has several objectives. First, to design and document a reproducible Business Intelligence workflow for 2024 order-level transaction data, from extraction to cleansing using OpenRefine and standardized transformations and aggregations. Second, to generate a clean and documented dataset and build an interactive buyer dashboard in Tableau that displays key KPIs such as revenue, number of unique buyers, sales trends per period, product performance, and geographic distribution in an easy-to-understand interface. Third, to validate the metrics and transformation logic with Python/pandas scripts to ensure the analysis results are accurate and repeatable. Fourth, to assess the usefulness of this dashboard for operational and strategic decision-making and to preserve all research artifacts such as the clean dataset, scripts, and workbooks for auditability and replication.

2. Literature Review

Research on Business Intelligence (BI) implementation in Indonesia over the past few years has shown a shift from conceptual studies to practical implementation studies that emphasize the end-to-end pipeline: from ETL and data cleaning to dashboard deployment for business decision-making [9], [10]. Several local case studies illustrate BI adoption at an organizational scale, including retail and service organizations, focusing on improving performance visibility and operational efficiency through interactive visualizations [11]. Thus, much of this work is in the form of case studies of single organizations and does not always test generalizations to the highly dynamic context of e-commerce [12]. Furthermore, local technical education literature and software engineering articles feature a variety of dashboard designs (using Tableau, Power BI, Pentaho, etc.) for various domains. However, the technical approaches presented often focus on visualization and interface aspects, with limited discussion of data cleansing and pipeline reproducibility in most practice reports [3], [13].

In the realm of e-commerce and customer analytics, several empirical studies in Indonesia have examined the relationship between digital service quality, marketing, and repurchase/loyalty behavior [14]. Showing that user experience and e-service quality metrics contribute significantly to repurchase intentions. These findings underscore the value of buyer-level analytics for retention strategies [15]. Additionally, research examining BI in supply chain/e-logistics demonstrates the benefits of BI for inventory optimization and operational decision-making, but most empirical examples use aggregate or single-company data rather than buyer-level transaction datasets, which would allow simultaneous calculation of unique buyers, repeat rates, and geospatial analysis [16].

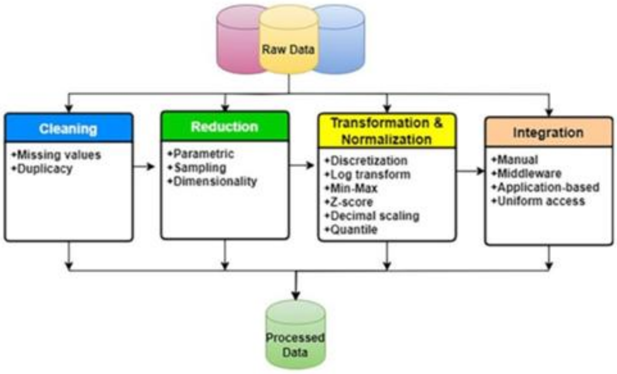


Figure 1. Data preprocessing steps and activities

As shown in Figure 1, explains more specifically about methodology and tooling. There is a tendency for local publications to discuss the use of popular BI tools (Tableau, Power BI) and simple ETL pipelines (Excel/script) to prepare data for dashboards; however, detailed documentation on the use of reconciliation-based cleansing tools (e.g., OpenRefine), clustering strategies for location standardization (provinces/districts), and anonymization practices for PII are rarely a primary focus in published articles. These steps are crucial when processing real e-commerce transaction data for publication and application in studies [3], [17]. Additionally, studies that produce dashboards tend to showcase visuals and organizational benefits (use cases), but rarely evaluate the validity of buyer-level metrics when data has quality issues (e.g., missing buyer_id, inconsistent province names), or demonstrate how consistent cleansing can impact KPI outcomes (e.g., number of unique buyers per month). This methodological gap indicates the need for studies that combine systematic cleansing practices with measurable dashboard analysis. [18], [19].

Based on this literature review, several research gaps can be clearly identified. First, few studies present a complete pipeline from transaction data cleansing (using standardized tools like OpenRefine) to deploying interactive dashboards in Tableau specifically for buyer analysis (rather than simply aggregate sales performance) [20]. Second, there is a lack of studies that utilize order-level data containing buyer_id to calculate critical indicators such as unique buyers, repeat buyer rate, and cross-tab analytics (e.g., revenue per buyer per province) while also evaluating how sensitive these KPIs are to variations in data cleansing. Third, most local publications focus on case studies of single organizations or non-ecommerce domains (e.g., hospitals, zakat, manufacturing), so generalization to national e-commerce ecosystems (with heterogeneity in platforms and buyer demographics) is still limited [1], [21].

3. Materials and Methods

3.1 Research Methodology

As shown in Figure 2, presents a detailed methodology in the form of a flowchart to facilitate understanding of the process and to guide the author to results and conclusions that align with the research objectives. This research uses a quantitative descriptive approach and an implementation study documenting the entire Business Intelligence (BI) pipeline for creating an e-commerce buyer dashboard using Tableau.

The methodology is designed following the data workflow in Figure 1, which includes a series of activities from literature review to data/dashboard publication. This approach emphasizes reproducibility and auditability: each stage (ETL, cleansing, transformation, visualization) is recorded so the process can be replicated. Conducting research requires a series of data processing steps, from literature review to storing and sharing data through Tableau Public.

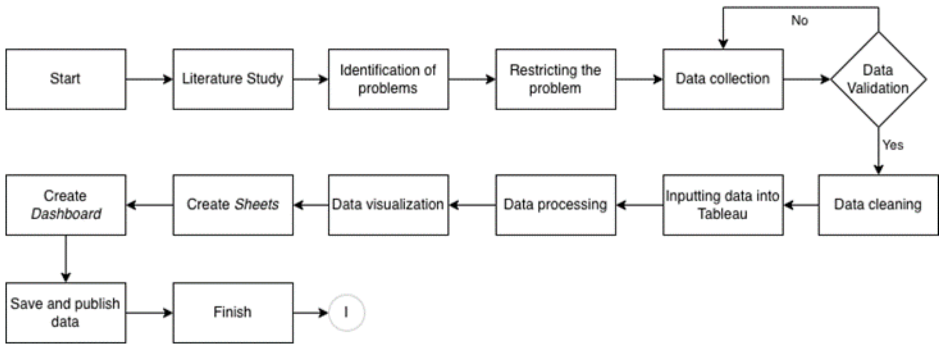


Figure 2. Flowchart of Research Stages

As shown in Figure 2, presents the overall research flowchart. The study begins with defining analytical objectives and selecting KPIs; next, secondary order-level datasets are collected and verified for provenance. Data cleaning is performed using OpenRefine to standardize date formats, harmonize province names, and anonymize PII. The cleaned dataset is exported and further transformed (derived metrics such as unique buyers and repeat rate) using Python/pandas. Final visualizations and interactive worksheets are developed in Tableau and combined into a dashboard, which is then evaluated through a sensitivity analysis that tests how different cleaning strategies affect buyer-level KPIs.

3.2 Scope of Research

The research object is order-level transaction records listed in the available dataset. The scope is limited to analysis of buyer metrics (unique buyers, repeat rate), aggregate revenue per month, per province, and per product, as well as segmentation by gender and platform [22]. The study does not include analysis of external market prices or logistics data beyond the available Province column.

3.3 Data Collection Technique

As seen in Figure 1, the raw data and the locations where the data were obtained are illustrated. The database, which may be stored on a local hard drive or a remote server, serves as the primary data source for a database management system [23], [24]. The computer program can use files, datasheets, spreadsheets, XML files, or even data encoded into the program itself as its data source [25]. Data sources generally fall into two types: primary sources and secondary sources, depending on how the data is collected or generated, as shown in Figure 3.

A computer program can use files, datasheets, spreadsheets, XML files, or even data embedded within the program itself as its data source. Data sources are generally classified into two types: Primary and Secondary depending on how the data are collected or generated, as shown in Figure 3.

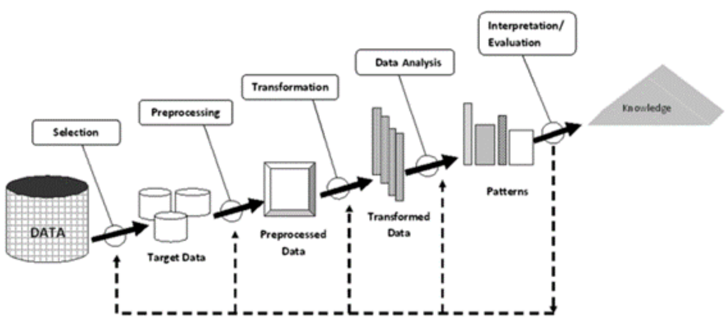


Figure 3. Transition from Primary Data to Secondary Data

As shown in Figure 3, First-hand data is generated directly by the data source and stored without any modification or preprocessing. This data can also be called original data or raw data. Collecting original, first-hand data increases the reliability and authenticity of the data [26]. Primary data sources help uncover hidden and new facts. Generating primary data is expensive and time-consuming. Once analyzed, primary data can be further processed to produce secondary data [27].

First-hand data is generated directly by the data source and stored without any modification or preprocessing. This data can also be called original data or raw data [28]. Collecting original, first-hand data increases the reliability and authenticity of the data [29]. Primary data sources help uncover hidden and new facts. Generating primary data is expensive and time-consuming. Once analyzed, primary data can be further processed to produce secondary data.

Because the data the author created is secondary data, collection techniques include: requesting/transferring files from data owners, verifying metadata (columns, date ranges, coding schemes), and recording provenance. If the source comes from multiple files, column consolidation and harmonization are performed before cleaning, as shown in Figure 3.

3.4 Data Analysis Techniques

The analysis was conducted in a quantitative descriptive manner: time series aggregation, product ranking based on total revenue, calculation of unique buyers and repeat rates, and geospatial visualization [3], [30]. To test the robustness of the cleanup pipeline, a sensitivity analysis was performed: experimenting with varying clustering methods in the Province column (e.g., aggressive clustering vs. conservative clustering) and observing their impact on unique buyers and revenue per province [31]. A comparison of pre- and post-cleanup results is shown in a summary table (delta KPI values) to show how much each cleanup step impacts the outcome. If necessary, simple statistical tests (e.g., paired difference tests) can be used to evaluate the significance of KPI differences between cleanup scenarios.

Table 1. Main Materials and Tools

Software / Hardware	Version	Main Function
OpenRefine (Data Cleaning)	3.9.5	Initial data cleaning and standardization
Tableau Desktop (Data Reduction)	2025	Data Reduction and Visualization
R (Data Normalization)	4.3.2	Data analysis and transformation using tidyverse package
Tableau Public (Data Integration)	2025	Interactive data visualization publication
MacBook Pro M1 (Hardware)	macOS 15.3, RAM 8 GB	The main hardware for computing and analysis

Based on the details of the materials and tools used in the study are presented in Table 1. Initial data cleaning was performed using OpenRefine (version 3.9.5). Data analysis and transformation were performed using R (version 4.3.2) with the tidyverse package (including dplyr) for data manipulation and statistical analysis. Data visualization and reduction were performed with Tableau Desktop (version 2025), and interactive visual results were published using Tableau Public. All procedures involving random numbers were run after setting the seed set.seed(12345) so that the experimental results could be reproduced. Full versions of R packages and environment information (sessionInfo()), as well as analysis scripts are available in the Appendix and/or code repository (see Appendix / Repository) for reproducibility purposes.

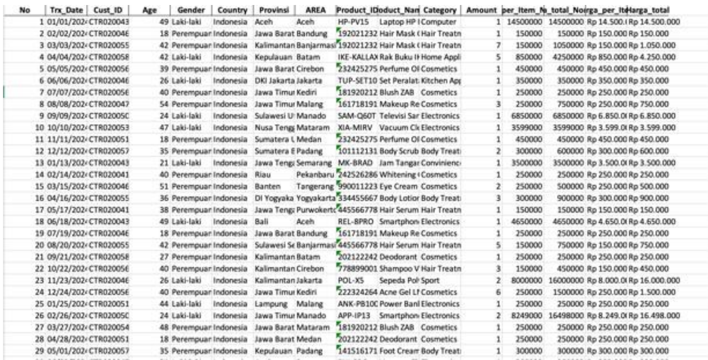
4. Results

A literature study aims to identify a problem and seek appropriate and relevant information related to the research topic in references or journals on the relevant topic. The relevant references collected must be in accordance with the method being discussed, namely Business Intelligence. Business Intelligence (BI) is defined as a decision-making process supported by the interconnectedness and analysis of data sources. Business Intelligence (BI) presents a unique challenge in the Industrial Revolution 4.0 [32].

Literature review is crucial after a researcher has established a topic. The next step is to review various literature sources related to that particular topic. At this stage, researchers must gather as much information as possible. This information can be obtained from journals, books, previous research results, and the results of previous studies.

4.1. Data Process

The data collected is childbirth data sourced from the Mendeley dataset from January 1, 2024, to December 28, 2024. The data consists of product category data that conducted transactions in several provinces in Indonesia. Figure 4 shows E-Commerce Repeat Buyer Data in Indonesia in 2024.



No	Trx_Date	Cust_ID	Age	Gender	Country	Provinsi	AREA	Product_ID	oduct_Nan	Category	Amount	per_Item_N	_total_No	rga_per_It	Harga_total
1	01/01/202	CTR020043	49	Laki-laki	Indonesia	Aceh	Aceh	HP-PV15	Laptop HP I	Computer	1	14500000	14500000	Rp 14.500.	Rp 14.500.000
2	02/02/202	CTR02004E	18	Perempuar	Indonesia	Jawa Barat	Bandung	192021232	Hair Mask (	Hair Treatn	1	150000	150000	Rp 150.000	Rp 150.000
3	03/03/202	CTR020055	42	Perempuar	Indonesia	Kalimantan	Banjarmasi	192021232	Hair Mask (	Hair Treatn	7	150000	1050000	Rp 150.000	Rp 1.050.000
4	04/04/202	CTR020058	42	Laki-laki	Indonesia	Kepulauan	Batam	IKE-KALLAX	Rak Buku II	Home Appl	5	850000	4250000	Rp 850.000	Rp 4.250.000
5	05/05/202	CTR02005E	39	Perempuar	Indonesia	Jawa Barat	Cirebon	232425275	Perfume Ol	Cosmetics	1	450000	450000	Rp 450.000	Rp 450.000
6	06/06/202	CTR02004E	26	Laki-laki	Indonesia	DKI Jakarta	Jakarta	TUP-SET10	Set Peralat	Kitchen App	1	350000	350000	Rp 350.000	Rp 350.000
7	07/07/202	CTR02005E	40	Perempuar	Indonesia	Jawa Timur	Kediri	181920212	Blush ZAB	Cosmetics	1	250000	250000	Rp 250.000	Rp 250.000
8	08/08/202	CTR020047	54	Perempuar	Indonesia	Jawa Timur	Malang	161718191	Makeup Re	Cosmetics	3	250000	750000	Rp 250.000	Rp 750.000
9	09/09/202	CTR02005C	24	Laki-laki	Indonesia	Sulawesi U	Manado	SAM-Q60T	Televisi Sar	Electronics	1	6850000	6850000	Rp 6.850.0	Rp 6.850.000
10	10/10/202	CTR020053	47	Laki-laki	Indonesia	Nusa Teng	Mataram	XIA-MIRV	Vacuum Cl	Electronics	1	3599000	3599000	Rp 3.599.0	Rp 3.599.000
11	11/11/202	CTR020051	18	Perempuar	Indonesia	Sumatera L	Medan	232425275	Perfume Ol	Cosmetics	1	450000	450000	Rp 450.000	Rp 450.000
12	12/12/202	CTR020057	35	Perempuar	Indonesia	Sumatera E	Padang	101112131	Body Scrub	Body Treat	2	300000	600000	Rp 300.000	Rp 600.000
13	01/13/202	CTR020043	21	Laki-laki	Indonesia	Jawa Teng	Semarang	MK-BRAD	Jam Tangar	Convinienc	1	3500000	3500000	Rp 3.500.0	Rp 3.500.000
14	02/14/202	CTR020041	40	Perempuar	Indonesia	Riau	Pekanbaru	242526286	Whitening	Cosmetics	1	250000	250000	Rp 250.000	Rp 250.000
15	03/15/202	CTR02004E	51	Perempuar	Indonesia	Banten	Tangerang	990011223	Eye Cream	Cosmetics	2	250000	500000	Rp 250.000	Rp 500.000
16	04/16/202	CTR020055	36	Perempuar	Indonesia	DI Yogyaka	Yogyakarta	334455667	Body Lotior	Body Treat	3	300000	900000	Rp 300.000	Rp 900.000
17	05/17/202	CTR020041	38	Perempuar	Indonesia	Jawa Teng	Purwokert	445566778	Hair Serum	Hair Treatn	1	150000	150000	Rp 150.000	Rp 150.000
18	06/18/202	CTR020043	49	Laki-laki	Indonesia	Bali	Aceh	REL-8PRO	Smartphon	Electronics	1	4650000	4650000	Rp 4.650.0	Rp 4.650.000
19	07/19/202	CTR02004E	18	Perempuar	Indonesia	Jawa Barat	Bandung	161718191	Makeup Re	Cosmetics	1	250000	250000	Rp 250.000	Rp 250.000
20	08/20/202	CTR020055	42	Perempuar	Indonesia	Sulawesi S	Banjarmasi	445566778	Hair Serum	Hair Treatn	5	150000	750000	Rp 150.000	Rp 750.000
21	09/21/202	CTR020058	27	Perempuar	Indonesia	Kalimantan	Batam	202122242	Deodorant	Cosmetics	1	250000	250000	Rp 250.000	Rp 250.000
22	10/22/202	CTR02005E	40	Perempuar	Indonesia	Kalimantan	Cirebon	778899001	Shampoo V	Hair Treatn	3	150000	450000	Rp 150.000	Rp 450.000
23	11/23/202	CTR02004E	26	Laki-laki	Indonesia	Kalimantan	Jakarta	POL-X5	Sepeda Pol	Sport	2	8000000	16000000	Rp 8.000.0	Rp 16.000.000
24	12/24/202	CTR02005E	40	Perempuar	Indonesia	Jawa Timur	Kediri	222324264	Acne Gel Ll	Cosmetics	6	250000	1500000	Rp 250.000	Rp 1.500.000
25	01/25/202	CTR020051	44	Laki-laki	Indonesia	Lampung	Malang	ANK-PB10C	Power Banl	Electronics	1	250000	250000	Rp 250.000	Rp 250.000
26	02/26/202	CTR02005C	24	Laki-laki	Indonesia	Jawa Timur	Manado	APP-IP13	Smartphon	Electronics	2	8249000	16498000	Rp 8.249.0	Rp 16.498.000
27	03/27/202	CTR020054	48	Perempuar	Indonesia	Jawa Barat	Mataram	181920212	Blush ZAB	Cosmetics	1	250000	250000	Rp 250.000	Rp 250.000
28	04/28/202	CTR020051	18	Perempuar	Indonesia	Jawa Barat	Medan	202122242	Deodorant	Cosmetics	1	250000	250000	Rp 250.000	Rp 250.000
29	05/01/202	CTR020057	35	Perempuar	Indonesia	Kepulauan	Padang	141516171	Foot Crearr	Body Treat	1	300000	300000	Rp 300.000	Rp 300.000

Figure 4. E-Commerce Repeat Buyer Database in Indonesia 2024

The data was obtained directly from the Mendeley dataset. The data collected included 178 customers with sales totaling Rp2,298,975,000 and a profit of Rp1,312,510,000. From this data, it can be concluded that the best-selling products are Electronics and Gadgets.

Data processing is performed using Business Intelligence methods with Tableau Public software. Tableau is a software application for analyzing and visualizing data sets into information for decision-making. There are many types of Tableau, including Tableau Desktop, Tableau Prep, Tableau Online, and Tableau Public [13]. Business Intelligence itself has several benefits for companies, such as increasing the value of organizational data and information, making it easier to monitor organizational performance, making IT investments better, making it easier for employees to access information, and making costs more efficient.

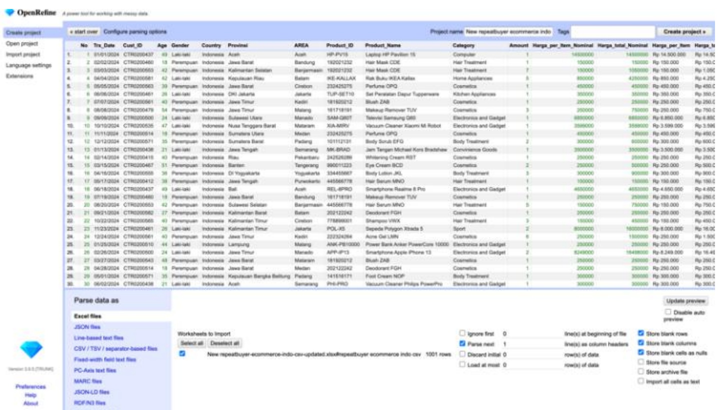


Figure 5. Cleaning Search Results in the OpenRefine Application

Next, in Figure 5 shows the data cleaning process for the dataset previously exported from the Excel database in Figure 4 and then imported into the OpenRefine application. This process includes identifying and grouping spelling variations (e.g., "West Java", "West Java", "West Java") using clustering and faceting methods, correcting typos, removing duplicates, and normalizing category and location labels to ensure consistent entries. OpenRefine is used to interactively preview changes (before-after preview) so operators can assess and accept transformations before applying them to all columns.

Data type validation (string-to-numeric conversion in price/quantity columns) and filling/marking of missing values are also performed, where necessary. These cleaning results are important because they directly impact KPI aggregation in number of unique buyers, sales distribution by province, the product ranking by reducing noise and bias caused by inconsistent entries. To ensure reproducibility of the analysis, all changes are logged (export changelog/operations) and are recommended to be included as an attachment or repository accessible to reviewers.

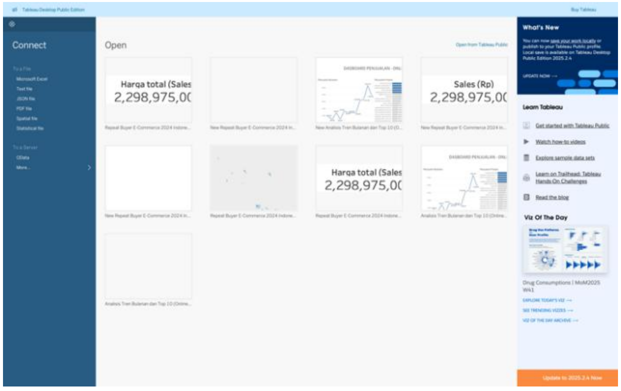


Figure 6. Initial View of the Tableau Public Application

After importing a data source into Tableau, the first view in Figure 6 shows the Connect panel on the left (*To a File Microsoft Excel option*) and an Open area where thumbnails of available workbooks and work sheets are shown. This screen will enable the preview of the data source, the name of the workbook and already created worksheets/ KPI cards (e.g., Total Price Rp 2,298,975,000) which will allow already check the number of rows and fields that will be used in visualization. The first view has buttons and contextual menus which allow rapid navigation.

To check the type of the variable *(string, numeric, date)* by open the Data Source, refresh the connection, or to go to a particular worksheet using the Go To Worksheet window. This perspective provides the basis of creating dashboards: researchers are able to peruse field mappings, pin assign right data types, organize time hierarchies, and preview sample values prior to making additional visualizations. To make the workbook reproducible and reviewable, it is suggested that the workbook should be saved as a package *(.twbx)* or published to a repository (Tableau Public / institutional repository) and a link or export file be included in the manuscript attachment.

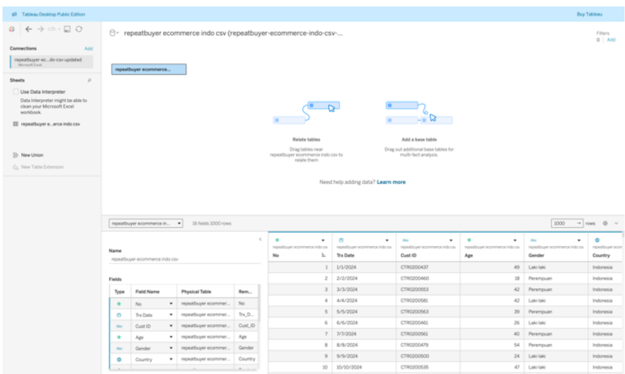


Figure 7. Pop Up Go To Worksheet

The imports are followed by making Use Data Interpreter of Tableau available in Tableau Data Source view in Figure 7. This feature helps Tableau to spot rows of mismatched headers, remedy merged cells, and delineate the types of the columns (date, text and number) to help Tableau come up with a more uniform table design before it generates visualizations.

The lowest panel shows a metadata grid which previews rows, field names and data types which helps the researcher to make quick corrections as in changing column types or deleting unneeded columns or splitting/pivoting columns. Furthermore, at this point users may create relationships or joins among sheets in case the dataset is made of several tables and formulate extracts to enhance workbook performance as well as add data-source filters to reduce the scope of analysis.

Once the data structure has been validated and corrected, the researchers will create a KPI worksheet Figure 8 that will present the key metrics, including: the number of orders, the number of unique buyers, total revenue, and the total profit, on which the whole analytics dashboard will be based. In order to reproducibly verify a transformation, all the transformation steps and settings of the Data Source are remembered or expelled (save as .twbx / export metadata) that reviewers can determine the visualization pipeline up to that point.

4.2. Data Visualization

The following are the steps for data processing to visualize it. The first step is to input the data to be processed, as shown in Figure 7. The data will be processed in the Public Tablue, as shown in Figure 7. The data visualizations used include KPIs, Sales by Month, Sales by Province, Sales by Product, and Sales by Gender, plus filters and an Interactive Dashboard.

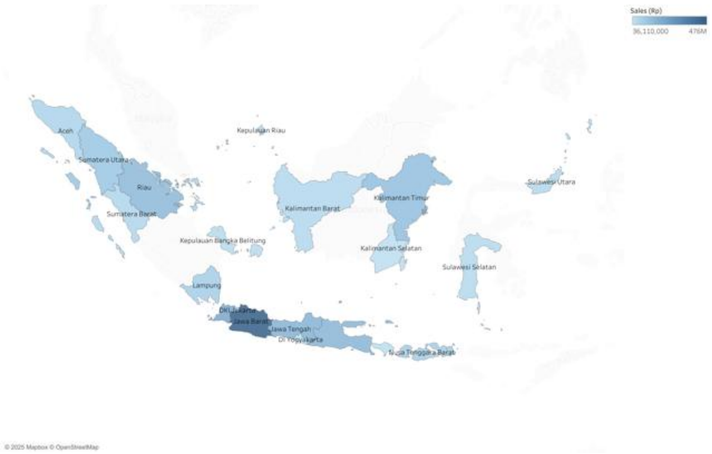


Figure 8. Sales by Province

In Figure 8 demonstrates thematic choropleth map, which shows the distribution of total retail sales of Indonesian provinces in January 1 - December 28, 2024 starting at a range of about Rp 36,110,000 to Rp 476,000,000. The graphical representation shows that sales are concentrated mainly on the island of Java with the biggest contribution of West Java at about Rp 475,808,000 and Jakarta at about Rp 221,800,000. Provinces located in the Kalimantan, and sections of Sumatra, are placed in a middle-tier opposition, and areas east such as Sulawesi and the Nusa Tenggara archipelago pattern a relatively lower level of sales. This analysis is supplemented by the interactive dashboard, which shows absolute provincial values, along with their percentage contribution during the total revenue, thus, making it easier to prioritize the high-yield provinces and then devote further investigative attention to these particular provinces.

The implications of this trend based on practice include a series of operation management procedures. Among them, the rational management of stocks, the strengthening of the specific marketing activities, and the rationalization of the logistical paths within the provinces with the best sales quotas can be identified as priority measures [33]. At the same time, there is a recommendation to re-audit segmented marketing approaches of areas with a rather underperforming performance within the field and make sure that the allocation of resources follows the localized consumer behavior and market opportunities. Following this kind of evidence-based interventions, companies will be able to optimize the effectiveness of their supply chains, as well as the effectiveness of their market penetration prospects.

Analytically, to compare the situation between provinces, the author suggests the normalization of sales data by taking into account some variables (per-capita income, or product penetration rates per 1,000 residents) to bring the provinces to a more representative level. Also, it is suggested to have a granular drill-down by product type or sales channel, which makes it possible to understand inter-provincial differences in a subtle way. These types of methodological fine tunings will aid in comparative analysis that is stronger and strategy decisions made both at the provincial and the national levels.

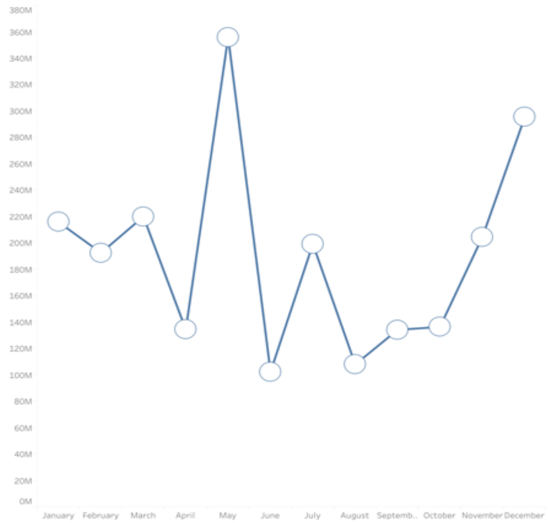


Figure 9. Sales by Month

In Figure 9 shows the trend in the monthly sales that has been seen in 2024. The same period January-March was rather stable with the results spanning between Rp 190,000,000 to Rp 220,000,000. This stability was succeeded by a significant fall back in April (around Rp 140,000,000). This dataset then shows one strong spike in May and this could be described as a seasonal outlier. Later in June, the level of sales dropped drastically (approximately Rp 105,000,000). We can see that there is a rebound in July (around Rp 195,000,000) and the sales became worse in the period of August-October, with the results ranging between Rp 110,000,000 and Rp 140,000,000. Lastly, one can see the gradual growth to the end of the year and the ultimate high at the end of December of about Rp 295,000,000.

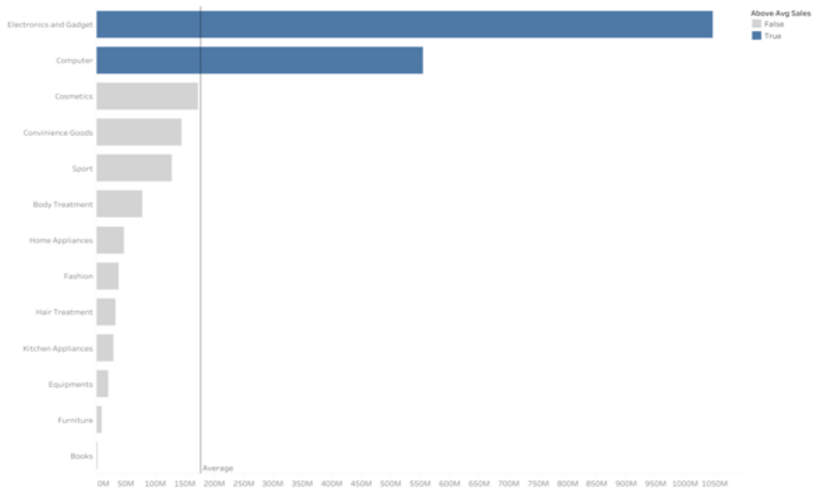


Figure 10. Sales by Product

The identified pattern can indicate the influence of a season or the presence of specific events that triggered the anomaly in May and a consistent increase to the end of the year the promotion activities, sales holidays, shortage of inventory, or the change of the marketing channels. In order to support the underlying reasons of such fluctuations, one should consider running some correlation analysis on the time series of sales and on the records on promotional activity (dates and types) inventory/fulfil ment indicators, and channel performance indicators (e.g., web traffic, conversion rates) over the same period.

Figure 10 shows the sales distribution graph per product against the average line (≈200 million). Electronics & Gadgets stand out far above the average (around >Rp 1 billion), followed by Computers which are also significantly above the average (Rp 554 million). Other products are mostly below the average: Cosmetics (Rp 172.5 million), Convenience Goods (Rp 144.5 million) and Sport (Rp 128 million) are slightly below the average, while Body Treatment, Home Appliances, Fashion, Hair/Kitchen Appliances, Equipments, Furniture, and Books show a relatively small sales contribution (range from tens of millions to very minimal).

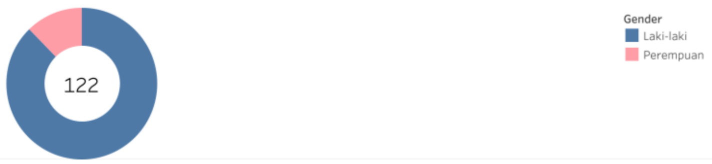


Figure 11. Product Sales by Gender

The analysis of product sales in terms of gender, according to Figure 11, shows that in the entire sample (n= 122) male buyers represented the major segment which was around 85 -90 percent. On the other hand, the size of female buyers was a comparative minority because it only was about 1015 per cent. The given results highlight the finding of a high level of gender disparity in the demand of the products among the identified sample population, where male consumers and their buying habits determine the overall volume of sales.

Figure 12. Interactive Dashboard of Repeat Buyers of E-Commerce Indonesia 2024

Source: https://public.tableau.com/app/profile/uya.asy.syuura.anandri/viz/NewRepeatBuyerE-Commerce2024IndonesiaDataset/Dashboard1?publish=yes

These observations have a substantive implication on the product development and the marketing strategy. First, the company should consider diversifying and refining the product attributes including product features, design, size, and packaging in order to reflect more closely female consumer preferences hence increasing product appeal to the demographic. Second, more granular marketing strategy might need to be adopted that includes targeted digital campaigns targeted at females, partnerships with female influencers or brand ambassadors, customization of messages of communication and channel of distribution in order to better address the unique needs of this segment. Thirdly, it is suggested to use the control marketing experiment, such as A/B tests, and compare them with qualitative research tools such as focus groups or in-depth interviewing, to explain the obstacles to female buyers joining and assess the effectiveness of the potential interventions.

In Figure 12 shows the 2024 Indonesian E-commerce Buyer Dashboard that is an interactive sales analysis module that is subdivided by time interval, product, province of origin, and gender of the buyer. The graphical depiction records bar sales in the sum of Rp 2,298,975,000 of which the net profit is Rp 1,312,510,000 in 178 different customers. The peak of sales is in May when Electronics and Gadgets domain is represented as the most successful product line and the distributions of transactions are also mainly limited in Java and Sumatra. The age structure of buyers is majorly male.

The novelty of the methodology and value proposition of this question would be the application use of a Tableau-based interactive dashboard that provides real-time analytics, multi-dimensional synthesis, and emphasizing the context of the Indonesian e-commerce environment. This kind of tool complements the comprehensive interpretation of data, and forms the basis of a sensible, data-driven decision-making process.

5. Conclusions

The proposed research paper describes the calculation and monitoring of buyer key performance indicators (KPIs) in e-business scenarios based on a Business Intelligence (BI) what can be called a replicable Business Intelligence (BI) pipeline. The transactional analysis shows that there is a cumulative revenue of Rp2,298,975,000 based on 1000 orders made by 178 different buyers and that the focus point is on the Java Island region, Electronics and Gadgets category occupies the largest portion of revenue and seasonality can be clearly, that it is clearly high during the month of May and at the season when the fiscal year has ended.

The buyer profiling shows that there is a high number of male customers and the number of regular customers plays a considerable role. The methodological contributions include repeatable ETL and visualization pipeline, a workflow that is pragmatically sound and practical in extracting and authenticating buyer KPIs out of unclean data and empirical evidence of the sensitivity of KPIs values to data-cleaning choices. The study constraints include use of one single-platform, secondary data Short data with limited duration, absence of full demographic data of buyers and lack of a test of dashboard usability. We would, therefore, suggest strict reporting of the cleaning processes, strength tests, extrapolation of cross-platform dataset with and without in-depth demographics, widened usability tests and exploring possibilities of real time integration.

References

1. U. A. S. Anandri, D. Y. Prasetyo, and Masril, “Analisis Model Bisnis Inovatif Kue Pancung Keliling Untuk Pemberdayaan UMKM Tembilahan,” Jurnal Pengabdian Masyarakat (ABDIMAS), vol. 3, no. 4, pp. 204–212, Jun. 2025, Accessed: Jul. 07, 2025. [Online]. Available: https://jurnal.nawansa.com/index.php/abdimas/article/view/629
2. M. N. Arkan and U. A. S. Anandri, “Perancangan Entity-Relationship Diagram (ERD) Menggunakan SQL Server Database Pada Platform Tautan Halaman Web Lynk.Id,” Jurnal Sistem Informasi (TEKNOFILE), vol. 3, no. 4, pp. 221–227, Apr. 2025.
3. U. A. S. Anandri et al., “Analisis Statistika Deskriptif Data Mahasiswa Sistem Informasi B 2024 Universitas Islam Indragiri,” Jurnal Sistem Informasi (TEKNOFILE), vol. 3, no. 6, pp. 422–432, Jun. 2025, Accessed: Jul. 16, 2025. [Online]. Available: https://jurnal.nawansa.com/index.php/teknofile/article/view/619/321
4. U. A. S. Anandri and M. N. Arkan, “Pemanfaatan Power BI Untuk Visualisasi Data Penjualan Produk Pada Perusahaan Percetakan,” Jurnal Sistem Informasi (TEKNOFILE), vol. 3, no. 10, pp. 716–721, Oct. 2025, [Online]. Available: https://jurnal.nawansa.com/index.php/teknofile/article/view/562
5. D. Y. Prasetyo, Y. E. Sudiarti, U. A. S. Anandri, M. Risky, and K. Ihwan, “Website E-Commerce Pada Toko Oska Cake,” JUTI UNISI, vol. 9, no. 2, pp. 31–40, Dec. 2025, doi: https://doi.org/10.32520/juti.v9i2.4834.

6. P. Picozzi, U. Nocco, A. Pezzillo, and A. De Cosmo, "The Use of Business Intelligence Software to Monitor Key Performance Indicators (KPIs)," Electronics (Basel)., vol. 13, 2024, doi: https://10.3390/electronics13122286.
7. M. Gupta and J. F. George, "Toward the development of a big data analytics capability," Information & Management, vol. 59, 2022, doi: 10.1016/j.im.2021.103619.
8. L. Wang, T. R. A. L. Pertheban, T. Li, and L. Zhao, "Application of business intelligence based on big data in E-commerce data evaluation," Heliyon, vol. 10, 2024, doi: https://doi.org/10.1016/j.heliyon.2024.e38768.
9. U. A. S. Anandri, "Business plan and financial feasibility study of 'Prata Bubuhan' breakfast UMKM stall in Indragiri Hilir Regency," Priviet Social Sciences Journal, vol. 5, no. 11, pp. 388–399, Nov. 2025, doi: https://doi.org/10.55942/pssj.v5i11.746.
10. Y. Chen, C. Yang, and L. Xu, "A Business Intelligence Approach for E-commerce Customer Segmentation Analysis," Int. J. Data Sci. Anal., 2022, doi: https://doi.org/10.1007/s41060-022-00328-5.
11. A. Fajri and A. Sinaga, "Implementation of Business Intelligence to Determine Evaluation of Activities (Case Study Indonesia Stock Exchange)," International Journal of Information Engineering and Electronic Business, vol. 12, no. 6, pp. 51–67, Dec. 2021, doi: https://doi.org/10.5815/ijieeb.2020.06.05.
12. Guy Ritchie and S. Hariyanto, "Implementation of Business Intelligence In Analyzing Data Using Tableau at PT Global Bintan Permata," Bit-Tech (Binary Digital - Technology), vol. 6, no. 1, pp. 40–50, Aug. 2023, doi: https://doi.org/10.32877/bt.v6i1.875.
13. Murtiwiyati Murtiwiyati, Hansel Agathon, and Leli Safitri, "Implementasi Data Warehouse dan Business Intelligence Menggunakan Pentaho dan Metabase untuk Membuat Dahboard Visualisasi Kinerja Penjualan E-Commerce Wish," Jurnal Penelitian Teknologi Informasi dan Sains, vol. 2, no. 4, pp. 100–109, Dec. 2024, doi: https://doi.org/10.54066/jptis.v2i4.2783.
14. M. R. Lopez and P. L. Santos, "Enhancing E-commerce Performance with Predictive Business Analytics," Int. J. Inf. Manage., 2024, doi: https://doi.org/10.1016/j.ijinfomgt.2023.102513.
15. Amirah Kamila Safitri, Mohamad Rizan, and Rahmi, "Peranan E-Service Quality, Rates Dan Customer Trust Bagi Repurchase Intention Jasa Kurir," Jurnal Ekonomi, vol. 28, no. 2, pp. 234–254, Jul. 2023, doi: https://doi.org/10.24912/je.v28i2.1646.
16. F. A. Ramadhan and A. Mansur, "Bibliometric Analysis and Data Visualization: Business Intelligence in Digitalization of Supply Chain Management in Healthcare Sector," Prozima: Productivity, Optimization, and Manufacturing System, vol. 8, no. 2, pp. 115–131, Dec. 2024, doi: https://doi.org/10.21070/prozima.v8i2.1728.
17. A. M. Amri and F. Agustina, "Implementasi Business Intelligence Menggunakan Tableau Untuk Visualiasi Data Pengumpulan Dan Penyaluran Zakat Di Indonesia (Studi Kasus Badan Amil Zakat Nasional)," Syntax Literate: Jurnal Ilmiah Indonesia, vol. 10, no. 6, pp. 6265–6278, Jun. 2025.
18. P. M. Lopulalan, R. H. S. Suhartono, B. Bakri, A. Karim, and U. A. S. Anandri, "Reconstructing a Sustainable Quality Management System in Higher Education through an Integrative Approach between Academic Audits and Organizational Reflection," Journal of Educational Analytics, vol. 4, no. 3, pp. 675–690, Aug. 2025, doi: https://doi.org/10.55927/jeda.v4i3.362.
19. U. A. S. Anandri and Mhd. N. Arkan, "Implementasi dan Pengembangan Aplikasi Point Of Sales Berbasis Website pada UMKM RitelShop Tembilahan," Jurnal Pengabdian Masyarakat (ABDIMAS), vol. 3, no. 7, pp. 304–314, Dec. 2025, [Online]. Available: https://jurnal.nawansa.com/index.php/abdimas/article/view/751
20. Y. Fatoni, P. Eko, P. Utomo, and R. Raaiqa Bintana, "Business Intelligence Roadmap for Tableau Dashboard Development in Higher Education," Briliance: Research of Artificial Intelligence, vol. 5, no. 2, pp. 973–980, Oct. 2025, doi: https://doi.org/10.47709/brilliance.v5i2.7094.
21. R. Lantarsih, Y. Mz, U. B. Surono, and G. Ramadhan, "Perancangan Dashbord Visualisasi Efektivitas Platform E-Commerce untuk Pemasaran Hasil Pertanian Petani Milenial Kabupaten Sleman," Smart Comp: Jurnalnya Orang Pintar Komputer, vol. 14, no. 2, pp. 417–424, Apr. 2025, doi: https://doi.org/10.30591/smartcomp.v13i1.7760.
22. A. I. Alfassa, G. N. Yanti, R. Z. Tizar, N. C. Kamila, and M. R. N. Kurniawan, "Analisis Dinamika Kependudukan Terhadap Permohonan Bantuan Sosial Penduduk Kurang Mampu Pekan Arba Tahun 2024," Jurnal Perangkat Lunak, vol. 7, no. 2, pp. 186–197, 2025.

23. I. R. Mukhlis and R. Santoso, "Perancangan Basis Data Perpustakaan Universitas Menggunakan MySQL dengan Physical Data Model dan Entity Relationship Diagram," Journal of Technology and Informatics (JoTI), vol. 4, no. 2, pp. 81–87, Apr. 2023, doi: https://doi.org/10.37802/joti.v4i2.330.
24. P. Mikalef and J. Krogstie, "Examining the role of big data analytics capability in firm performance," Information & Management, vol. 59, 2022, doi: https://doi.org/10.1016/j.im.2021.103591.
25. A. I. Alfassa, S. Sudrajat, and D. Marwasta, "Development of official statistics models for analysis of population sectoral data in Indragiri Hilir Regency," in E3S Web of Conferences, EDP Sciences, 2023, p. 06007.
26. A. I. Alfassa, "Statistika Kependudukan Untuk Rencana Kebijakan Kependudukan Daerah," DEMOS: Journal of Demography, Ethnography and Social Transformation, vol. 2, no. 2, pp. 76–85, 2022.
27. S. Akter and S. F. Wamba, "Big Data Analytics in E-commerce: A Systematic Review and Agenda for Future Research," Electronic Markets, vol. 31, 2021, doi: https://doi.org/10.1007/s12525-021-00490-6.
28. A. I. Alfassa, F. Kusumawardhani, and E. Sudeska, "Kebijakan Kependudukan Dalam Grand Design Pembangunan Kependudukan (GDPK) Pada Penataan Data Statistik Kependudukan Dengan Indikator Demografi (Fertilitas, Mortalitas, dan Migrasi)," DEMOS: Journal of Demography, Ethnography and Social Transformation, vol. 5, no. 1, 2025.
29. A. I. Alfassa, "Peran Grand Design Pembangunan Kependudukan (GDPK) pada fenomena kependudukan di Indonesia melalui 5 pilar kependudukan," DEMOS: Journal of Demography, Ethnography and Social Transformation, vol. 4, no. 1, pp. 1–10, 2024.
30. U. Ahmed, S. Shamim, and Z. Khan, "Big data analytics capability and decision-making," Technol. Forecast. Soc. Change, vol. 168, 2021, doi: https://doi.org/10.1016/j.techfore.2021.120766.
31. A. I. Al Fassa and A. Kesumawati, "Segmentation of Karhutla Hotspot Point of Indragiri Hilir Regency 2015 and 2016 using Self Organizing Maps (Soms)," in Proceedings of the International Conference on Mathematics and Islam (ICMIs 2018), Mataram, Indonesia: UIN Mataram Indonesia; ADMAPETA, 2020, pp. 336–341.
32. E. Bahar, N. Irmalia Azizah, A. Sri Hayuningsih, and D. R. Agushinta, "Analisis Data Pasien Ibu Hamil Menggunakan Metode Business Intelligence," IJUBI: Indonesian Journal of Business Intelligence, vol. 6, no. 2, pp. 116–123, Dec. 2023, doi: https://doi.org/10.21927/ijubi.v6i2.3831.
33. A. I. Alfassa, "Model dasar statistika industri dalam penelitian industri kependudukan," Juti Unisi, vol. 8, no. 1, pp. 35–38, 2024.

COMPLETE
Journal of Computer, Electronic, and Telecommunication
ISSN 2723-4371, E-ISSN 2723-5912

Article

Design and Implementation of a 16-Electrode Electrical Impedance Tomography Data Acquisition System for Medical Imaging

Aldo Nofrianto[1*], Audy[2], and Aditya Wardani[3],

[1 - 3] Politeknik Negeri Padang, Padang, Indonesia

* Correspondence: aldonofrianto@pnp.ac.id

Received: 24 December 2025; Revised: 29 December 2025; Accepted: 30 December 2025; Published: 31 December 2025

Abstract: Electrical Impedance Tomography (EIT) requires a stable and accurate data acquisition system to obtain reliable boundary voltage measurements for medical imaging applications. However, many existing EIT systems are complex and costly, limiting their practical use in clinical and research environments. This paper presents the design, implementation, and experimental validation of a 16-electrode EIT data acquisition system for medical imaging. The proposed system consists of a sinusoidal signal generator, voltage-to-current converter (VCC), voltage measurement circuit, multiplexer/demultiplexer unit, microcontroller, and image reconstruction algorithm. Experimental evaluations were conducted to assess signal stability, current injection performance, and voltage measurement accuracy. The XR2206-based signal generator demonstrated stable operation over a frequency range of 1–210 kHz with an output impedance of 0.0784 kΩ. The LF41 based amplifier showed linear performance up to 50 kHz, while the VCC produced stable injection currents ranging from 0.3 to 2 mA for load variations between 100 and 2000 Ω, with optimal stability at 10 kHz. Data acquisition using the adjacent method was performed on a 16-electrode phantom containing bovine bone as a resistive object. Image reconstruction using the iterative Newto Raphson method with Tikhonov regularization successfully identified the object's position and boundaries. The results demonstrate that the proposed system provides stable and reliable imaging suitable for medical EIT applications.

Keywords: current injection; data acquisition system; EIT; image reconstruction; medical imaging

1. Introduction

Electrical Impedance Tomography (EIT) is a non-invasive medical imaging technique that reconstructs the internal electrical resistivity distribution of an object by injecting low-amplitude alternating currents through surface electrodes and measuring the resulting boundary voltages [6], [7]. Due to its non-ionizing nature, low cost, and portability, EIT has gained increasing attention as an alternative imaging modality for continuous and bedside medical monitoring.

Recent studies have demonstrated the application of EIT in pulmonary monitoring, neurological assessment, and tissue characterization [1], [2], [5]. Advances in hardware design and signal processing have enabled EIT systems to provide clinically relevant information, particularly in respiratory monitoring during mechanical ventilation and weaning processes [2], [5]. Furthermore, optimization of electrode configurations and excitation parameters has been reported to significantly influence image quality and reconstruction accuracy [3], [4].

Despite these advantages, EIT images still suffer from limited spatial resolution due to measurement noise, instability of current injection, electrode contact impedance, and the ill-posed nature of the inverse problem [8]. Therefore, the development of reliable hardware systems with stable current injection and accurate voltage measurement remains a critical research challenge.

This research focuses on the design and implementation of a 16-electrode EIT data acquisition system for medical imaging purposes. The proposed system emphasizes stable current injection, reliable voltage measurement, and effective image reconstruction. The novelty of this study lies in the development of a complete 16-electrode EIT hardware system with experimentally optimized current injection parameters (0.3 mA at 10 kHz), resulting in stable data acquisition and improved image reconstruction performance using a cost-effective design approach.

2. Materials and Methods

This study employed an experimental approach to design and evaluate a 16-electrode EIT data acquisition system. The developed hardware architecture consists of a sinusoidal signal generator based on the XR-2206 integrated circuit, an LF412-based amplifier, a voltage-to-current converter (VCC), an analog multiplexer/demultiplexer, and a microcontroller-based control unit. The excitation signal generated by the XR-2206 is amplified and converted into a constant current before being injected into the object under test. The overall hardware architecture of the proposed EIT data acquisition system is illustrated in Fig. 1.

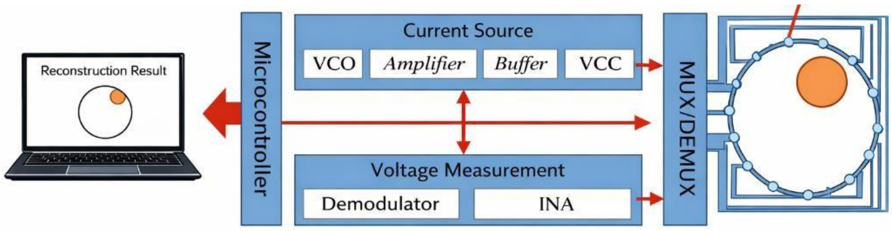


Figure 1 System block diagram of the proposed 16-electrode EIT data acquisition system

A cylindrical phantom filled with saline solution was used as the conductive medium, while bovine bone samples were placed inside the phantom to represent resistive inclusions, following experimental approaches reported in recent EIT studies [4]. The system employed the adjacent current injection and voltage measurement method, which is commonly adopted in practical EIT systems due to its simplicity and robustness [3], [9]. One complete measurement cycle produced 208 voltage data points. The electrode configuration and adjacent measurement pattern used in this study are shown in Fig. 2.

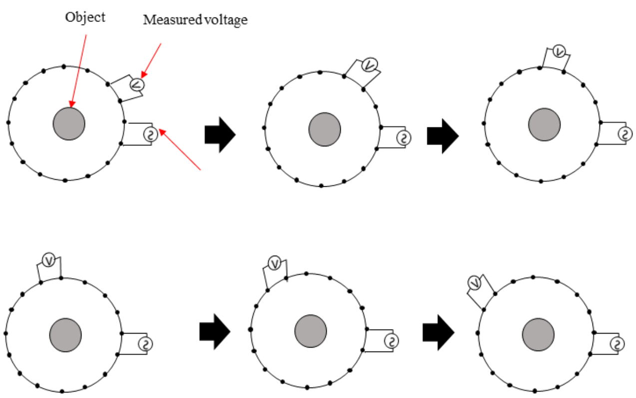


Figure 2 Electrode configuration and adjacent measurement pattern

Current injection levels ranging from 0.3 to 2 mA and excitation frequencies between 10 and 50 kHz were evaluated to determine optimal operating conditions. Acquired voltage data were transferred to a personal computer for post-processing. Image reconstruction was performed using an iterative Newton–Raphson algorithm combined with Tikhonov regularization to stabilize the inverse problem and suppress noise effects [8], [9]. The reconstruction workflow is illustrated in Fig. 3. Initially, an estimated resistivity distribution was assigned to the imaging domain and discretized using a twodimensional finite element method (2D FEM). Based on this initial estimate, boundary voltages were calculated by solving the forward problem under the same adjacent current injection pattern applied in the experimental measurements.

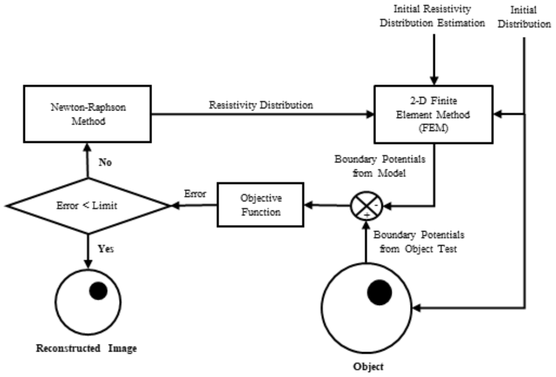


Figure 3 Block diagram of the iterative image reconstruction process using the Newton–Raphson algorithm

Initially, an estimated resistivity distribution was assigned to the imaging domain and discretized using a two-dimensional finite element method (2D FEM). Based on this initial estimate, boundary voltages were calculated by solving the forward problem under the same adjacent current injection pattern applied in the experimental measurements.

The inverse problem was formulated as an optimization problem by minimizing the discrepancy between the measured boundary voltages and the voltages calculated from the numerical model. The objective function is expressed as:

$$\Phi(\rho) = \|V_{meas} - V_{Calc}(\rho)\|^2 \quad (1)$$

Where V_{meas} denotes the measured boundary voltage vector, V_{Calc} represents the calculated voltage vector obtained from the FEM-based forward model, and ρ is the resistivity distribution. If the value of the objective function exceeded a predefined tolerance, the resistivity distribution was updated iteratively using the Newton–Raphson method with Tikhonov regularization. The update process can be expressed as:

$$\rho_{k+1} = \rho_k + \Delta\rho \quad (2)$$

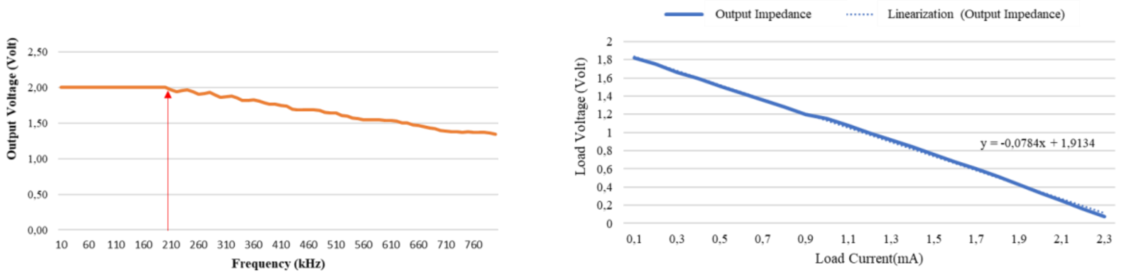
Where ρk and ρk+1 represent the resistivity distributions at the k-th and (k+1)-th iterations, respectively, and $\Delta\rho$ is the correction term computed using the regularized Newton–Raphson formulation. The regularization term was incorporated to stabilize the solution and suppress noise amplification during the iterative process. The iteration was repeated until the error between the measured and simulated boundary voltages fell below the specified tolerance. The converged resistivity distribution was subsequently visualized as the final reconstructed EIT image.

3. Results

This section presents the experimental results obtained from the developed 16-electrode Electrical Impedance Tomography (EIT) system. The performance of the system is evaluated in terms of current injection stability, voltage measurement reliability, and image reconstruction capability. The effects of injection current amplitude and excitation frequency on boundary voltage measurements and reconstructed images are systematically analyzed.

3.1. Performance of the Current Injection System

The performance of the current injection subsystem was evaluated under varying frequency and load conditions. The XR-2206 signal generator produced a stable sinusoidal output voltage over a frequency range of 1–210 kHz, with an output impedance of approximately 0.0784 kΩ, As shown in Fig. 4, the XR-2206 generator exhibits stable output voltage and low output impedance across the evaluated frequency range.



(a) (b)

Figure 4 Performance characteristics of the XR-2206 signal generator: (a) output voltage stability as a function of excitation frequency, and (b) output impedance characterization

The voltage-controlled current source demonstrated stable current injection across load resistances ranging from 100 Ω to 2000 Ω, maintaining injected currents within 0.3–2 mA. The most stable operation was observed at an excitation frequency of 10 kHz, as illustrated in Fig. 5.

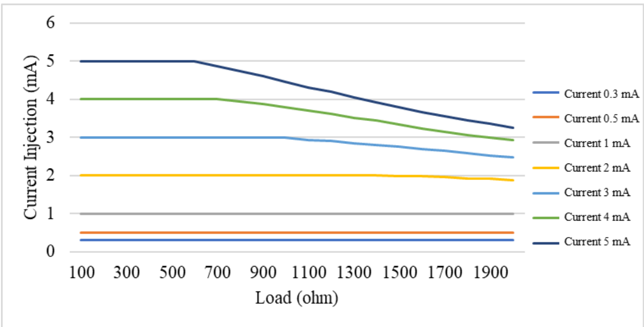
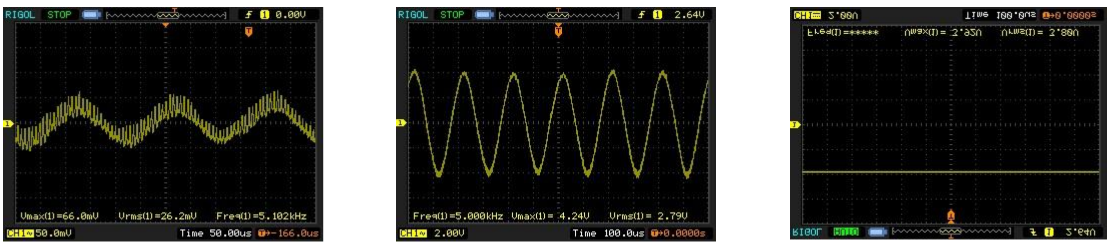


Figure 5 Current stability of the voltage-controlled current source under load variations

3.2. Voltage Measurement Results

Boundary voltage signals acquired from the phantom were characterized by low amplitudes in the millivolt range. After amplification using an instrumentation amplifier (AD620 Module) and RMS-to-DC conversion (IC AD536), the acquired signals exhibited a significant improvement in signal quality, as presented in Fig. 6.



(a) (b) (c)

Figure 6 Boundary voltage signal processing stages: (a) raw electrode output signal, (b) output of the AD620 instrumentation amplifier with a gain of 64, and (c) output of the AD536 RMS-to-DC converter

Using the 16-electrode adjacent measurement protocol, the system successfully acquired 208 independent voltage measurements per acquisition cycle, demonstrating consistent and repeatable voltage patterns suitable for EIT image reconstruction

3.3. Effect of Injection Current Variation

The influence of injection current amplitude on voltage measurements was investigated using current levels of 0.3 mA, 0.5 mA, 1 mA, and 2 mA at a fixed excitation frequency of 10 kHz. The resulting voltage measurement profiles are presented in Fig. 7.

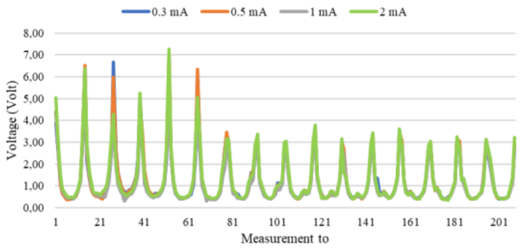


Figure 7 Voltage measurement profiles for different injection current levels

All tested current levels enabled detection of the resistive object within the phantom. Differences in signal amplitude and stability were observed across the tested current levels

3.4. Effect of Excitation Frequency Variation

The effect of excitation frequency on voltage measurements was evaluated at 10 kHz, 30 kHz, and 50 kHz, while maintaining a constant injection current of 0.3 mA. The corresponding voltage profiles are shown in Fig. 8.

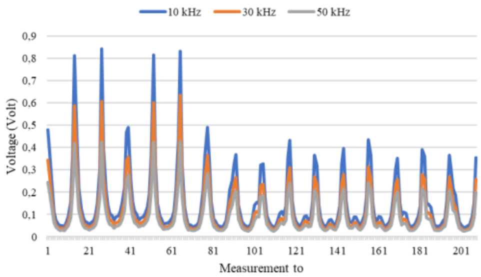


Figure 8 Voltage measurement profiles for different excitation frequencies

Variations in voltage contrast and spatial distribution were observed across the tested frequencies

3.5. Image Reconstruction Results

Reconstructed EIT images obtained under different excitation conditions are presented in Fig. 9 and Fig. 10. The reconstructed images reveal variations in object boundary clarity and spatial resolution depending on the applied current amplitude and excitation frequency.

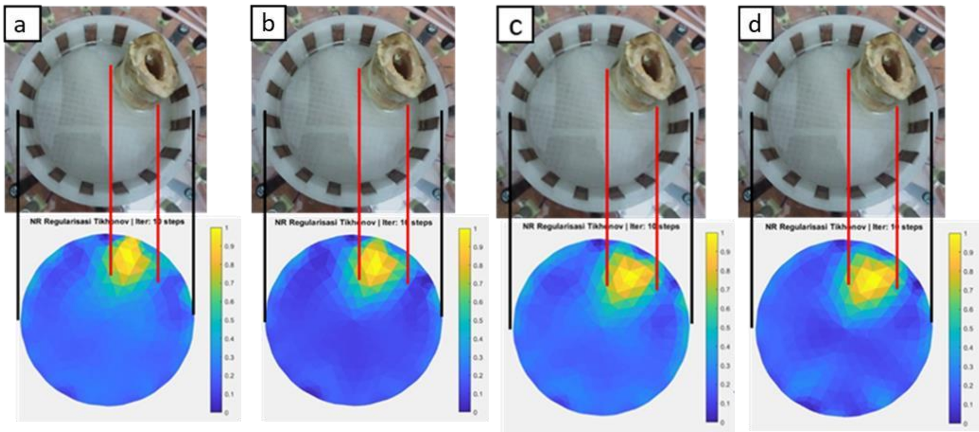


Figure 9 Reconstructed images for different injection current levels (a) 0,3 mA, (b) 0,5 mA, (c) 1 mA, and (d) 2 mA

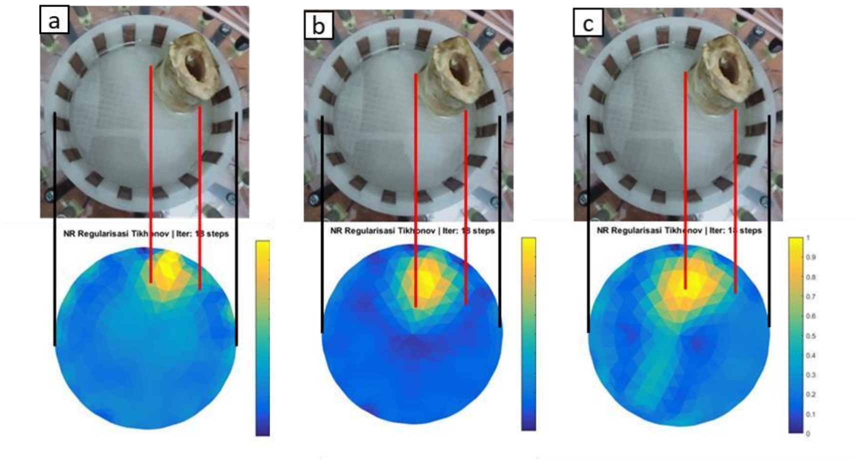


Figure 10 Reconstructed images for different excitation frequencies (a) 10 kHz, (b) 30 kHz, and (c) 50 kHz

The reconstruction result for two separated resistive objects, shown in Fig. 11, demonstrates the system's capability to distinguish multiple inclusions within the imaging domain.

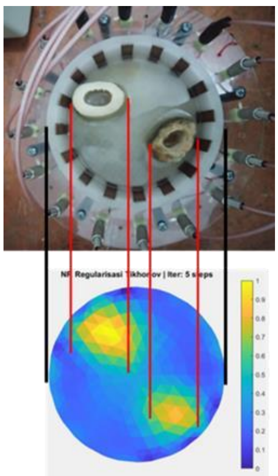


Figure 11 Reconstruction result for two separated resistive objects

4. Discussion

This section discusses the experimental results presented in the previous section by interpreting the observed trends and relating them to findings reported in recent EIT literature, with a focus on excitation stability, measurement reliability, and reconstruction performance of the proposed system. Despite the promising performance of the proposed 16-electrode EIT data acquisition system, several limitations should be acknowledged. The system performance may be affected by electrode contact impedance variations and electrical noise originating from analog components. Nevertheless, in comparison with previous EIT systems reported in the literature, the proposed design offers a simpler hardware configuration while maintaining comparable signal quality, making it suitable for low-cost medical imaging applications.

4.1. Current Injection and Measurement Stability

The stable output voltage of the XR-2206 signal generator and the consistent current injection achieved by the voltage-controlled current source indicate that the excitation subsystem meets the requirements for biomedical EIT applications. Stable current injection is critical, as fluctuations directly affect boundary voltage measurements and reconstruction accuracy. Similar observations have been reported in recent EIT system developments, where excitation stability was identified as a key factor influencing image quality [1], [2]

4.2. Influence of Injection Current Amplitude

Lower injection current amplitudes (0.3–1 mA) resulted in more stable voltage measurements compared to higher current levels. Although higher currents increased voltage amplitude, they also introduced increased variability. This behavior supports previous recommendations that lowamplitude excitation provides an optimal balance between measurement stability, reconstruction accuracy, and safety considerations in medical EIT systems [3], [4]

4.3. Influence of Excitation Frequency

The results demonstrate that an excitation frequency of 10 kHz yields superior voltage contrast and improved image clarity compared to higher frequencies. This can be attributed to the frequencydependent electrical properties of biological tissues, where resistive components dominate at lower frequencies. These findings are consistent with recent studies on frequency optimization in EIT-based tissue imaging [5], [9].

4.4. Image Reconstruction Performance

The reconstructed images confirm that the Newton–Raphson algorithm combined with Tikhonov regularization effectively stabilizes the ill-posed inverse problem inherent in EIT. The adjacent measurement strategy provides sufficient sensitivity for detecting resistive inclusions while maintaining a simple and robust hardware configuration. Although spatial resolution is limited by the number of electrodes, the obtained results demonstrate the feasibility of the proposed system for non-invasive imaging applications [10], [11].

The results confirm that stable current injection and accurate voltage measurement are critical factors influencing EIT image quality. Lower excitation frequencies contributed to improved signal stability and reduced sensitivity to measurement noise, in agreement with findings reported in recent EIT applications for medical monitoring [2], [5].

The Newton–Raphson reconstruction method combined with Tikhonov regularization effectively mitigated the ill-posed nature of the EIT inverse problem and improved image clarity. Although the spatial resolution remains lower than that of conventional imaging modalities, the proposed system demonstrates sufficient performance for non-invasive medical imaging research and experimental validation. The use of cost-effective electronic components further supports the feasibility of developing low-cost EIT systems for research and educational purposes, as suggested in recent EIT system development studies [3], [12].

5. Conclusions

This study aimed to design and implement a 16-electrode Electrical Impedance Tomography (EIT) data acquisition system for medical imaging applications. Experimental results demonstrate that the proposed system is capable of acquiring stable boundary voltage signals suitable for reliable image reconstruction. The developed system exhibited stable operation with current injection levels ranging from 0.3 to 2 mA and excitation frequencies up to 50 kHz. Optimal imaging performance was achieved at an injection current of 0.3 mA and a frequency of 10 kHz, resulting in improved spatial resolution and image contrast. Future work will focus on increasing the number of electrodes, enhancing noise reduction techniques, and implementing advanced image reconstruction algorithms to further improve imaging accuracy.

References

1. Y. Guo et al., "Electrical impedance tomography provides information of brain injury during total aortic arch replacement," Scientific Reports, vol. 14, art. 14236, 2024. DOI: https://doi.org/10.1038/s41598-02465203-0
2. J. J. Wisse et al., "Electrical impedance tomography as a monitoring tool during weaning from mechanical ventilation," Respiratory Research, vol. 25, art. 179, 2024. DOI: https://doi.org/10.1186/s12931-024-02801-6
3. A. Ramandha and Basari, "Performance optimization of electrode patterns in electrical impedance tomography," Jurnal Informatika & Rekayasa Elektronik, vol. 7, no. 2, pp. 210–217, 2024. DOI: https://doi.org/10.36595/jire.v7i2.1224
4. R. Aisya et al., "Application of electrical impedance tomography for detecting biological tissue,"Indonesian Journal of Electronics, Electromedical Engineering, and Medical Informatics, vol. 7, no. 2, 2025. DOI: https://doi.org/10.35882/ijeeemi.v7i2.54
5. Z. Cui et al., "Technical principles and clinical applications of electrical impedance tomography in pulmonary monitoring," Sensors, vol. 24, no. 14, art. 4539, 2024. DOI: https://doi.org/10.3390/s24144539
6. J. G. Webster, Electrical Impedance Tomography. Bristol, UK: Adam Hilger, 1990.
7. D. S. Holder, Electrical Impedance Tomography: Methods, History and Applications. London, UK: IOP Publishing, 2005.
8. B. H. Brown, "Electrical impedance tomography (EIT): A review," Journal of Medical Engineering & Technology, vol. 27, no. 3, pp. 97–108, 2003. DOI: https://doi.org/10.1080/0309190021000059688
9. A. Hassan, M. A. Rahman, and S. Ibrahim, "Performance evaluation of excitation current and frequency selection in electrical impedance tomography systems,"IEEE Access, vol. 12, pp. 45621–45632, 2024. DOI: https://doi.org/10.1109/ACCESS.2024.3372196
10. M. Li, J. Sun, and X. Dong, "Improved image reconstruction in electrical impedance tomography using regularized Newton-based methods," Measurement, vol. 223, art. 113825, 2024. DOI: https://doi.org/10.1016/j.measurement.2023.113825
11. Y. Wang, Z. Cui, X. Li, and H. Zhang, "Design and implementation of a multi-channel electrical impedance tomography system for biomedical applications," Biomedical Signal Processing and Control, vol. 86, art. 105247, 2024. DOI: https://doi.org/10.1016/j.bspc.2023.105247
12. A. Adler and W. R. B. Lionheart, "Uses and abuses of EIDORS," Physiological Measurement, vol. 27, no. 5, pp. S25–S42, 2006. DOI: https://doi.org/10.1088/0967-3334/27/5/S03