



Implementation of Natural Language Processing (NLP) in Developing a Chatbot Application for Classical Islamic Text Learning at Pesantren El-Huda El-Islamy

Yan Sofyan Andhana Saputra^{1*}, Alwi Fachri Ibnu Arroyan²

^{1,2} Department of Information Technology, Faculty of Engineering, Darma Persada University

^{1,2} South Malaka Park Road No. 8, East Jakarta 13450, Indonesia

*yansofyan@gmail.com

Abstract — El-Huda El-Islamy Islamic Boarding School is an Islamic educational institution that aims to mould students into competent and noble religious preachers. This institution focuses on learning and deep understanding of Islamic teachings, including the study of classical books such as the Yellow Book of Fathul Qorib. This research aims to develop and evaluate an Android-based chatbot application that uses the BERT model to help the learning process of the Yellow Book of Fathul Qorib at El-Huda Islamic Boarding School. In this research, the BERT model is integrated into the chatbot application to understand and respond to user questions appropriately. Testing was done by asking 25 questions to the chatbot, which successfully answered 11 questions with a success rate of 44%. Evaluation of the model performance using confusion matrix showed that the chatbot had 90% accuracy, 87% precision, 87% recall, and 86% F1-score. These results show that the chatbot has not been able to provide relevant and accurate responses, and recognise most of the questions asked. This research concludes that this chatbot application is not yet an effective tool to support the learning process at El-Huda Islamic Boarding School.

Keywords – BERT, Chatbot, Kitab Kuning Fathul Qorib, Pondok Pesantren.

Copyright © 2025 TIFDA JOURNAL
All rights reserved.

I. INTRODUCTION

In the era of rapid advancement in science and technology (IPTEK), educational institutions, companies, and other organizations are increasingly required to enhance the quality of their facilities and infrastructure. This is particularly evident in the education sector, where institutions are competing to improve the quality of their systems and facilities in order to attract a greater number of students. This trend also extends to Islamic boarding schools (pesantren), where technological development plays a crucial role in enhancing the quality of education provided [1]. In response to this, pesantren must adapt by upgrading their facilities and infrastructure to meet the evolving demands of the times and the academic needs of their students (santri).

A chatbot is a computer program designed to simulate human conversation, either through text or voice. Its ability to store large volumes of data without loss, to query information sources directly rather than relying solely on traditional search mechanisms, and its

learning capability make it a highly effective tool in customer service. Furthermore, chatbots have proven valuable in academic contexts, offering automated responses to user inquiries without the need for human intervention.

Pondok Pesantren El-Huda is one of the many Islamic boarding schools located in Bekasi Regency, specifically in the village of Satria Jaya, North Tambun District, and is led by Drs. KH. Nur Hasyim Ilyas, M.Pd.I. The pesantren includes the study of various classical Islamic texts, commonly referred to as kitab kuning. Although the study of kitab kuning is fundamental to preserving Islamic scholarly traditions, it is often accompanied by several challenges. These include difficulties in understanding complex textual content and limited access to relevant classical literature.

The development of a chatbot application presents an innovative solution for facilitating kitab kuning learning at Pondok Pesantren El-Huda El-Islamy. This application offers students an opportunity to engage in

self-directed, interactive learning tailored to individual needs, unbounded by temporal or spatial constraints [2]. It is anticipated that the implementation of such a chatbot will significantly improve the effectiveness and accessibility of kitab kuning education.

Various methodologies can be employed in the design and development of chatbots, one of which is Natural Language Processing (NLP). NLP is a subfield of artificial intelligence concerned with the interaction between computers and human (natural) language, enabling machines to understand, interpret, and generate human language [3]. One notable NLP model is Bidirectional Encoder Representations from Transformers (BERT), developed by Google. BERT is a transformer-based model that utilizes attention mechanisms in place of traditional recurrent networks and is capable of capturing contextual relationships between distant words within a sentence [4]. Each word is represented by BERT in a contextually rich manner, enhancing performance in complex, sequential NLP tasks.

In light of the aforementioned context, this study aims to develop a chatbot to support kitab kuning learning at Pondok Pesantren El-Huda El-Islamy.

II. METHODOLOGY

CRISP-DM (Cross-Industry Standard Process for Data Mining) is a well-established methodology used for the development of data mining projects. Although originally designed specifically for data mining initiatives, this methodology can also be effectively applied to the development of chatbot systems or other types of applications. The CRISP-DM framework consists of several key phases: Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation, and Deployment [5].

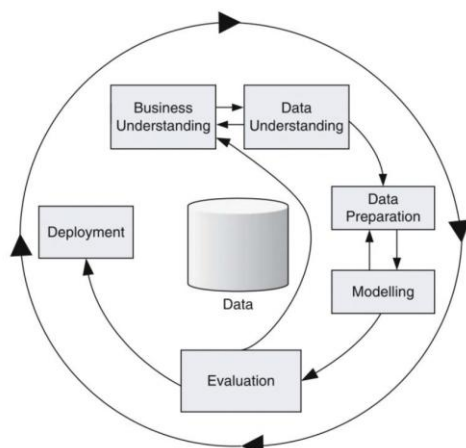


Figure 1. Cross Industry Standard Process for Data Mining [6]

III. RESULTS AND DISCUSSION

A. Business Understanding

In this phase, the business objectives and requirements are clearly defined [7]. For this project, the primary objective is to enhance the learning process of kitab kuning through the use of chatbot technology. A comprehensive understanding is required regarding how kitab kuning is traditionally studied, as well as the challenges faced by students (santri) in the process. Specifically, the kitab kuning utilized in the development of this application is Fathul Qorib authored by Muhammad bin Qasim Al-Ghazi.

B. Data Understanding

This stage involves the collection of relevant data for the project, such as textual content from the kitab kuning. This stage includes data exploration, and the verification of data quality [8]. The version of Fathul Qorib used is the Indonesian-translated edition by Muhammad bin Qasim Al-Ghazi. The collected data is then analyzed to assess its quality and characteristics, as well as to identify potential patterns that can be leveraged by the chatbot.

C. Data Preparation

The collected data is processed and prepared for integration into the system. This includes cleaning the data to eliminate errors or inconsistencies, and converting it into a format compatible with NLP algorithms and chatbot frameworks. The Fathul Qorib text data may require annotation to facilitate more advanced analysis. Table 1 presents examples of several data entries that are utilized in the development of the chatbot application.

Table 1. Shows sample data used in chatbot development.

Tag	Pattern	Response
sapa	"Hai", "Apa kabarmu?", "Apakah ada orang di sana?", "Halo!", "Selamat tinggal", "apa kabar", "hei"	"Senang bertemu denganmu lagi!", "Halo, ada yang bisa saya bantu?"
nama	"nama", "namamu", "apakah kamu punya nama", "apa yang kamu panggil", "Siapa namamu", "aku harus memanggilmu"	"Saya El-Huda Bot, Saya adalah sebuah chatbot. saya akan membantu ada kenal lebih dekat dengan Aplikasi ini"

Tag	Pattern	Response
	apa", "siapa namamu?", "apa yang kamu", "siapa kamu"	
Kitab	"apa itu fathul qorib", "apa itu fathul qarib", "apa itu kitab fathul qorib", "kitab fathul qorib menjelaskan tentang apa", "jelaskan tentang kitab fathul qorib", "fathul qorib kitab apa", "fathul qorib kitab tentang apa"	"Kitab Fathul Qarib adalah salah satu kitab berbahasa Arab tanpa menggunakan harakat dan terjemahan. Di dalam pesantren, kitab ini lebih dikenal dengan istilah kitab kuning atau kitab gundul. Dalam penyusunannya, kitab Fathul Qarib ini disusun secara ringkas dan sistematis, serta bermahzab Syafi'i. Kitab ini merupakan penjelasan dari kitab yang dikarang oleh Al Qadhi Abu Syuja, yaitu Al-Ghayah wa At-Taqrīb..."
bersuci	"apa itu bab bersuci", "apa isi dari bab bersuci", "berikan penjelasan tentang bab bersuci", "bab bersuci"	"Bab Bersuci atau Thaharah adalah bab yang menerangkan tentang melakukan sesuatu dengan tujuan untuk membersihkan diri dari hadats kecil dan besar yang bertujuan untuk menjadi syarat sah dalam beribadah"
sholat	"Jelaskan tentang bab sholat", "apa itu bab sholat", "apa isi dari bab sholat", "berikan penjelasan tentang bab"	"Bab Sholat adalah bab yang menerangkan tentang Syarat Wajib Sholat, Syarat Sah Sholat, Rukun Sholat, Perbedaan Antara Laki-laki dan

Tag	Pattern	Response
	sholat", "bab sholat"	Perempuan dalam Sholat, Hal-Hal yang membatalkan sholat, Jumlah Rakaat Shalat, Qoshor dan Jama' Shalat, Shalat Jum'at, Shalat Ied, Shalat Gerhana, Shalat Istisqo, Shalat Khauf, Hukum Sutra dan Perhiasan, Hal-hal yang berhubungan dengan jenazah"
Zakat	"Jelaskan tentang bab zakat", "apa itu bab zakat", "apa isi dari bab zakat", "berikan penjelasan tentang bab zakat", "bab zakat"	"Bab Zakat adalah bab yang membahas tentang Nishab dan kadar Zakat Unta, Nishab dan Kadar Zakat Sapi, Nishab dan Kadar Zakat Kambing, Zakat Persekutuan, Zakat Emas dan Perak, Zakat Pertanian, Zakat Dagangan, Syarat Wajib Zakat Fitrah, Penerima Zakat."
Puasa	"Jelaskan tentang bab puasa", "apa itu bab puasa", "apa isi dari bab puasa", "berikan penjelasan tentang bab puasa", "bab puasa"	"Bab Puasa menjelaskan tentang Syarat dan Rukun Puasa, Perkara yang membatalkan Puasa, Sunah Puasa, Hari-hari yang diharamkan Puasa, Kafarat dan Uzur meninggalkan Puasa, I'tikaf"

Tag	Pattern	Response
peradilan	"Jelaskan tentang bab Peradilan dan Persaksian", "apa itu bab Peradilan dan Persaksian", "apa isi dari bab Peradilan dan Persaksian", "berikan penjelasan tentang bab Peradilan dan Persaksian", "bab Peradilan dan Persaksian"	"Bab Peradilan dan Persaksian adalah bab yang menjelaskan tentang Syarat hakim, Larangan Bagi Hakim, Pembagi Harta, Dakwaan, Persaksian, Macam-macam Hak"

The data will be structured in JSON format prior to the modeling phase. Figure 2. illustrates an example of a dataset in JSON format.

```

},
{
  "tag": "Kitab",
  "patterns": [
    "apa itu fathul qorib",
    "apa itu fathul qorib",
    "apa itu kitab fathul qorib",
    "kitab fathul qorib menjelaskan tentang apa",
    "jelaskan tentang kitab fathul qorib",
    "fathul qorib kitab apa",
    "fathul qorib kitab tentang apa"
  ],
  "responses": [
    "Kitab Fathul Qarib adalah salah satu kitab berbahasa Arab tanpa menggunakan harakat dan terjemahan. Di dalam pesantren, kitab ini lebih dikenal dengan istilah kitab kuning atau kitab gundul. Dalam penyusunannya, kitab Fathul Qarib ini disusun secara ringkas dan sistematis, serta bermahzab Syafi'i. Kitab ini merupakan penjelasan dari kitab yang dikarang oleh Al Qadhi Abu Syuja, yaitu Al-Ghayah wa At-Ta'arib. Penggunaan kitab Fathul Qarib sendiri, dijadikan sebagai sumber primer dan pegangan wajib di sebuah madrasah diniyah atau lembaga pendidikan Islam yang bersifat salaf, yaitu pendidikan yang bercorak tradisional."
  ],
  "context set": ""
},
{
  "tag": "Kitab2",
  "patterns": [

```

Figure 2. Example of Chatbot Dataset in JSON Format

D. Modelling

- Programming Library Preparation.** In this phase, preparing the required programming libraries is a critical step in developing any machine-learning-based application or system, such as a chatbot employing a BERT model. First, ensure that Python is installed on the host machine and create a virtual environment to isolate all dependencies. Next, install the core libraries—numpy, pandas, and scikit-learn—followed by the principal NLP and BERT libraries, namely transformers and torch (or TensorFlow, if preferred).
- Dataset Loading.** During this stage, the curated dataset is loaded into the system to support chatbot development, guaranteeing that the data are in a compatible and ready-to-use format. The process entails reading the data from files or databases and verifying their accuracy and consistency. This step is essential to ensure the data can be exploited effectively when training the chatbot model and improving its responses.
- Data Inspection.** Here, a thorough examination of the loaded dataset is performed to confirm that the data contain valid and coherent information. The inspection includes verifying formats and data types, as well as identifying missing values or anomalies that could compromise subsequent analyses. The objective is to ensure the data are reliably prepared for the next phases.
- Sorting Data into Tag Classes.** In this stage, tag data are sorted to group and organise the tags into relevant categories. The procedure involves converting tags into appropriate labels to streamline downstream processing and analysis. The aim is to guarantee that each tag possesses a consistent, well-structured label, thereby facilitating model training and further analytical tasks.
- Assigning Numerical Labels.** The previously sorted labels are now assigned unique numeric identifiers for every category or class. This allocation of numerical labels enables more efficient processing and analysis of the data. The purpose is to simplify the identification and handling of labels within the model or system to be developed.
- Splitting Data into Training and Test Sets.** At this point, the dataset is divided into two subsets: a training set and a test set. The training set is used to fit the model, while the test set—kept separate—serves to evaluate the model's performance. This partitioning is crucial to ensure that the model can generalise effectively rather than merely memorising the training data.
- Loading a Pre-trained BERT Model and Applying Tokenisation.** In this process, a pre-trained BERT model is loaded into the system for analytical use. A tokenizer is then applied to convert textual data into the format required by the model, breaking the text into tokens that correspond to BERT's vocabulary. This step is vital to ensure that the textual data can be interpreted and processed efficiently by the model.
- Converting Data into Numerical Form.** The processed data are transformed into numerical representations so they can be ingested by the machine-learning model. This transformation converts text or other non-numeric data into numerical vectors or matrices required for model training and evaluation, thereby enabling effective analysis and predictive modelling.

- i) Creating a Data Loader. A data loader is constructed to prepare and manage the data used during the training and evaluation of the BERT model. The data loader efficiently supplies mini-batches, performs shuffling, and regulates the flow of data into the model. This stage ensures that data are presented in an optimal format and with maximal efficiency throughout training.
- j) Model Training. During this phase, the prepared data are used to train the model by applying the chosen learning algorithm to the training set. The process teaches the model to recognise pertinent patterns and features, with the goal of optimising its parameters so that it can achieve high accuracy in prediction or classification when applied to unseen data.
- k) Model Saving. Finally, the trained model is saved for integration into the chatbot application, preserving its learned parameters for future inference and deployment.

E. Evaluation

The developed model is tested to ensure that it meets business requirements established in the first phase (business understanding) and produces accurate outputs [9]. Testing is conducted by evaluating the chatbot using a set of question scenarios and comparing the responses it provides against the expected answers. This input-output experiment aims to assess the performance of the BERT model employed within the chatbot application. The primary focus is on the model's ability to comprehend user queries and deliver relevant and accurate responses. In this experiment, the chatbot application will be presented with 25 questions related to the Fathul Qorib text to evaluate its performance. Table 2. presents the list of questions submitted by santri (students) for testing the chatbot application.

Table 2: Students' Questions for Chatbot Evaluation

No	Q/A	R
1	Q: Apa definisi thaharah? A: Bab Bersuci atau Thaharah adalah bab yang menerangkan tentang melakukan sesuatu dengan tujuan untuk membersihkan diri dari hadats kecil dan besar...	Successful
2	Q: Sebutkan dan jelaskan macam-macam air yang suci dan menyucikan! A: Maaf saya tidak tau karena tidak ada dalam data saya	Unsuccessful
3	Q: Apa syarat-syarat sah wudhu? A: Fasal menjelaskan tentang fardhu-fardhu (kewajiban)	Successful

	dalam wudhu dan kesunahan dalam wudhu. Fardhu wudhu ada 6 yaitu...	
13	Q: Apa hukum fiqh mengenai jual beli barang dengan sistem cicilan, dan apa saja syarat-syarat yang harus dipenuhi agar transaksi tersebut sah? A: Maaf saya tidak tau karena tidak ada dalam data saya	Unsuccessful
14	Q: Bagaimana cara menentukan arah kiblat ketika berada di tempat yang tidak memiliki tanda atau penunjuk arah kiblat, menurut fiqh Islam? A: Maaf saya tidak tau karena tidak ada dalam data saya	Unsuccessful
15	Q: Apa hukum fiqh mengenai meng-qadha (mengganti) shalat yang terlewat, dan bagaimana tata cara pelaksanaannya? A: Maaf saya tidak tau karena tidak ada dalam data saya	Unsuccessful
16	Q: Apa itu fiqh? A: Maaf saya tidak tau karena tidak ada dalam data saya	Unsuccessful
17	Q: Apa perbedaan antara fiqh dan syariah? A: Maaf saya tidak tau karena tidak ada dalam data saya	Unsuccessful
18	Q: Apa yang dimaksud dengan fiqh muamalah? A: Maaf saya tidak tau karena tidak ada dalam data saya	Unsuccessful
19	Q: Bagaimana proses fatwa dikeluarkan? A: Maaf saya tidak tau karena tidak ada dalam data saya	Unsuccessful
20	Q: Bagaimana fiqh mengatur warisan? A: Maaf saya tidak tau karena tidak ada dalam data saya	Unsuccessful
21	Q: ada berapa rukun wudhu? A: Fasal menjelaskan tentang fardhu-fardhu (kewajiban) dalam wudhu dan kesunahan dalam wudhu. Fardhu wudhu	Successful

	ada 6 yaitu; 1. Niat (ketika membasuh wajah), 2. Mebasuh wajah, 3. Membasuh kedua tangan sampai kedua siku, 4. Mengusap sebagian kepala atau rambut, 5. Membasuh kedua kaki sampai mata kaki, 6. Tertib atau berurutan. Adapun kesunahan dalam wudhu ada 10 yaitu; 1. Membaca basmallah, 2. Membasuk kedua telapak tangan, 3. Berkumur, 4. Menghisap air dengan hidung, 5. Mengusap seluruh kepala, 6. Mengusap kedua daun telinga, 7. Menyela-nyela jenggot, 8. Mendahulukan anggota kanan dari anggota kiri, 9. Mengulang tiga kali pada setiap basuhan, 10. Berkesinambungan.			ia tinggalkan dan meneruskan sholat dan melakukan sujud sahwi. 2. Apabila yang ditinggalkan adalah sesuatu yang sunah ab'at maka orang yang sholat tidak boleh kembali mengerjakan sesuatu yang tertinggal jika ia telah melakukan rukun yang selanjutnya akan tetapi ia disunahkan untuk mengganti dengan sujud sahwi. 3. Apabila yang tertinggal adalah sunnah hai'at maka orang yang sholat tidak diperbolehkan melakukan kembali rukun yang tertinggal dan tidak pula melakukan sujud sahwi.	
22	Q: Apakah kita harus sholat untuk menjadi baik atau tunggu menjadi baik dlu baru sholat? A: Maaf saya tidak tau karena tidak ada dalam data saya	Unsuccessful			
23	Q: Apa niat sholat subuh? A: Maaf saya tidak tau karena tidak ada dalam data saya	Unsuccessful			
24	Q: berapa rakaat shalat subuh itu? A: Jumlah rakaat dalam shalat wajib atau fardhu: Dhuhur sebanyak 4 rakaat, Ashar sebanyak 4 rakaat, Magrib sebanyak 3 rakaat, Isya sebanyak 4 rakaat, Subuh sebanyak 4 rakaat.	Successful			
25	Q: bagaimana jika kita tertinggal rukun shalat, apakah harus mengulang? A: Perkara yang tertinggal dalam sholat diperinci menjadi 3 aspek yaitu: 1. Apabila yang ditinggalkan adalah sesuatu yang fardhu (rukun) sholat, maka tidak bisa diganti dengan sujud sahwi, akan tetapi jika orang yang sholat ingat dalam waktu yang dekat maka ia harus melakukan rukun yang	Successful			

From a total of 25 (twenty-five) questions submitted by users to the chatbot, 11 (eleven) responses were aligned with the expected answers, while 14 (fourteen) responses did not meet expectations. The discrepancy in the latter responses is attributed to a mismatch with the labels present in the dataset.

The BERT model testing in the chatbot application was conducted through a single training session using 100 epochs. The chatbot's performance was evaluated using a confusion matrix, which produced values for recall, precision, accuracy, and F1-score. The confusion matrix or error matrix is a common mechanism for evaluating model performance [10]. This method provides a more detailed overview of the chatbot's ability to accurately understand and respond to user questions. The results and analysis of the chatbot model testing are explained as follows:

- An accuracy score of 90% indicates that the chatbot performs well in recognizing and responding to user queries.
- A precision score of 87% demonstrates that the system is effective in providing accurate and relevant responses to user questions.
- A recall score of 87% indicates that the chatbot is capable of correctly identifying most relevant questions, reflecting its ability to avoid missing important queries.
- An F1-score of 86% shows that the chatbot achieves an optimal balance between delivering correct answers and recognizing all relevant questions, indicating a well-balanced performance in terms of both precision and recall.

F. Deployment

The implementation results of the developed chatbot application are presented as follows:

a) Home Page Interface

The home page display is shown in Figure 3. This interface contains selection buttons that direct users to either the chatbot menu or the about page.

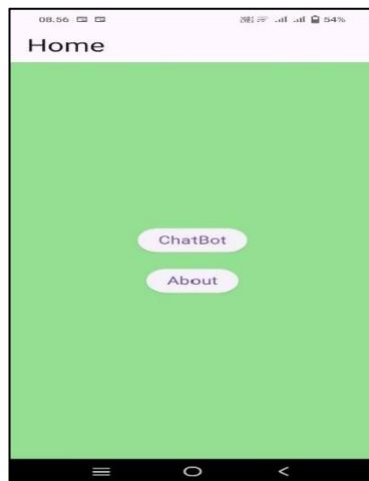


Figure 3. Home Page Interface

b) Chatbot Interface

The chatbot interface is presented in Figure 4. This platform enables users to interact with the chatbot application to inquire about the Fathul Qorib scripture.

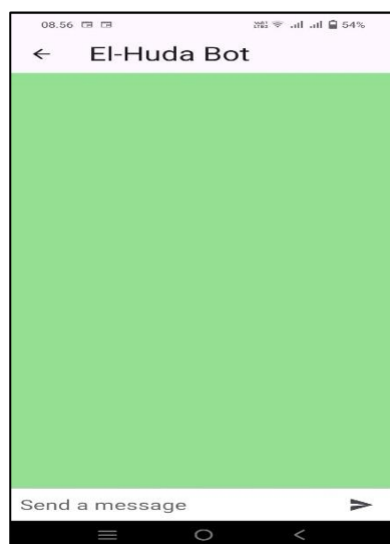


Figure 4. Chatbot Interface

c) About Page Interface

The about page display is illustrated in Figure 5. This section provides information regarding the chatbot application. The final stage involves the deployment of the chatbot system within the actual learning environment at Pondok Pesantren El-Huda

El-Islamy. The system was integrated into the platform used by the students and tested under real-world conditions. During this stage, an evaluation was conducted based on the previous tests, including input-output trials and model testing.



Figure 5. About Page Interface

In the input-output trial stage, the chatbot application was tested by submitting 25 questions. The chatbot successfully answered 11 questions and failed to answer 14 questions, resulting in a success rate of 44%. The chatbot's inability to provide accurate responses was primarily due to limitations in the dataset.

IV. CONCLUSION

Based on the design and evaluation results, it can be concluded that this study successfully implemented the BERT model to interpret and respond to user conversations. The evaluation results indicate an accuracy of 90%, a precision of 87%, a recall of 87%, and an F1-score of 86%. In the input-output trial phase, the chatbot successfully answered 11 out of 25 questions, resulting in a success rate of 44%. The limitations of the dataset prevented the chatbot from responding to several questions as expected by the users.

The chatbot has not yet demonstrated a balanced capability in providing accurate answers and recognizing all relevant queries. This underscores that the chatbot is still unable to deliver sufficiently accurate and relevant responses to users. Therefore, the chatbot is not yet considered an effective solution for delivering text-based services to users.

REFERENCES

- [1] A. Muchasan, N. Syam, and A. Humaidi, "Title: Use of Technology in Islamic Boarding Schools (Impact and Solutions in Educational Context)".

-
- [2] M. R. Astari, M. Sa'id, A. Kunta Mardlian, S. Bahri, and M. U. Siregar, "Rancangan Aplikasi Chatbot Telegram 'Tanya Zaid' Sebagai Media Pembelajaran Nahwu," 2023.
 - [3] A. Y. Chandra, D. Kurniawan, and R. Musa, "Perancangan Chatbot Menggunakan Dialogflow Natural Language Processing (Studi Kasus: Sistem Pemesanan pada Coffee Shop)," *Jurnal Media Informatika Budidarma*, vol. 4, no. 1, p. 208, 2020.
 - [4] Vidya Chandradev, I Made Agus Dwi Suarjaya, and I Putu Agung Bayupati, "Analisis Sentimen Review Hotel Menggunakan Metode Deep Learning BERT," *Jurnal Buana Informatika*, vol. 14, no. 02, pp. 107–116, 2023.
 - [5] Yudha Nurdin, Khairun Saddami, Nasaruddin "Pengenalan Praktis Supervised Machine Learning: Dengan Jupyter Notebook", Banda Aceh, 2025, pp. 16-17.
 - [6] Anita Wasilewska, Churn-Jung Liao, Tsau Young Lin, Ying Xie, "Data Mining: Foundations and Practice", Berlin, 2008, pp. 260.
 - [7] Nasir Ali Mirza, "Practitioner's Guide to Data Science. Streamlining Data Science Solutions Using Python, Scikit-Learn, and Azure ML Service Platform", Delhi, 2022, pp. 116
 - [8] Sreedhar, G, "Web Data Mining and the Development of Knowledge-Based Decision Support Systems", Hershey PA, 2016, pp. 209.
 - [9] David L. Olson, Dursun Delen, "Advanced Data Mining Techniques", Berlin, 2008, pp.10.
 - [10] Graham J. Williams, Graham Williams, "Data Mining with Rattle and R. The Art of Excavating Data for Knowledge Discovery", New York, 2011, pp. 36.