NLP Chatbot with ANN and SGD for University Admissions Website

Atha An Naufal[1], Arif Tirtana[2*]

[1]*Informatic, Jl. Raya Tidar No.100, Karangbesuki, Kec. Sukun, Kota Malang, Jawa Timur 65146, Indonesia*

Keywords
Chatbot; NLP; ANN; SGD; TF-IDF; New Student Admissions Services

Correspondence Email:
arif.tirtana@ubhinus.ac.id

Abstract
Universities today face challenges in providing real-time and responsive information for New Student Admissions. This research aims to implement a chatbot based on Natural Language Processing (NLP) using an Artificial Neural Network (ANN) model optimized with Stochastic Gradient Descent (SGD) to address these challenges. The goal of the implementation is to enhance information services through the use of chatbots. The analysis was conducted on frequently asked questions (FAQs) from prospective students, followed by system design, ANN architecture development, and model performance testing. During the ANN model development, Term Frequency–Inverse Document Frequency (TF-IDF) was employed for numerical representation, enabling the training process to be processed effectively by the ANN model. The implementation results show that the chatbot deployed on the UBHINUS admission website successfully provides real-time and relevant information. Testing results demonstrate an accuracy of 92.5% and an F1-Score of 91.56%, indicating strong classification performance. It can be concluded that the chatbot is effective in facilitating access to information regarding New Student Admissions at UBHINUS.

1. Introduction

In today's digital era, the way people access and manage information has undergone a significant transformation. Higher education institutions, as providers of education services, are no exception to these changes. To attract prospective students, universities are required to provide innovative and responsive services. Among the most important aspects of this effort are the speed of obtaining information, the accuracy of the information provided, and the ease of access to such information [1].

The rapid development of digital technology has introduced innovative solutions to address challenges in information services. One such solution is the implementation of chatbots, which utilize Natural Language Processing (NLP) supported by Artificial Neural Network (ANN) models and optimized through Stochastic Gradient Descent (SGD). A chatbot is a computer-based system designed to simulate human dialogue, enabling real-time interaction with users as though conversing with another person. Through the integration of NLP, chatbots have become highly valuable tools for a wide range of applications [2].

To ensure accurate text classification in chatbot systems, selecting an appropriate model is crucial [3]. One of the widely adopted approaches is the use of Artificial Neural Networks (ANN), which are inspired by the structure of the human brain and consist of interconnected layers of artificial neurons [4]. For the ANN model to process user text input effectively, the Term Frequency–Inverse Document Frequency (TF-IDF) method is applied. This method extracts features from textual data, enabling the model to perform accurate classification tasks [5].

This research aims to implement an NLP-based chatbot using an ANN model with SGD optimization on the New Student Admissions (PMB) portal of Universitas Bhinneka Nusantara (UBHINUS). The chatbot is

expected to provide prospective students with fast, accurate, and relevant information, thereby improving the overall quality of admission services.

1.1 Literature Review

The adoption of digital technology in higher education has transformed service delivery, particularly in the area of admissions. Previous studies highlight that information accessibility, accuracy, and responsiveness are crucial factors in enhancing the user experience for prospective students [1].

Chatbots, as an innovative information service tool, have been widely discussed in academic and industrial domains. According to recent studies, chatbots leverage NLP to simulate human conversation, allowing users to interact seamlessly with computer systems [2]. These systems have been effectively applied in diverse fields, ranging from customer service to education, due to their ability to deliver real-time responses and personalized interactions.

In terms of text classification, choosing an appropriate computational model is a key factor. Prior works emphasize that the Artificial Neural Network (ANN) is among the most effective models for such tasks, as it mimics the structure and functionality of the human brain through interconnected artificial neurons [3][4]. ANN models are particularly suitable for handling complex textual data due to their ability to capture patterns and relationships in large datasets.

Furthermore, effective feature extraction methods are essential for improving model performance. Term Frequency–Inverse Document Frequency (TF-IDF) has been identified as a widely used technique in natural language processing to convert textual data into numerical representations. This transformation enables ANN models to process input effectively for classification purposes [5].

Building upon these findings, this research focuses on the integration of ANN with NLP and SGD optimization in designing a chatbot for UBHINUS's admission portal. The study contributes by demonstrating how such a system can enhance information services, streamline communication with prospective students, and ultimately strengthen institutional competitiveness.

2. Research Methods

This section elaborates on the scientific methodology applied in this study, covering the stages of research implementation, timeframe, and research location. The research process was structured into several systematic stages, as illustrated in the research flow diagram (Figure 1).

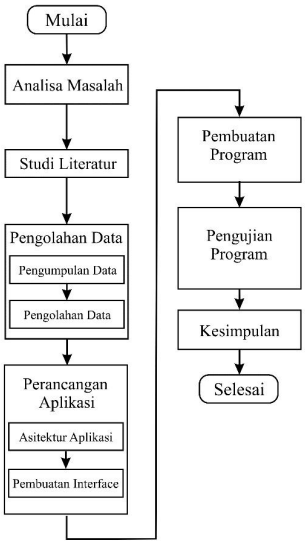

Figure 1. Research flow diagram

Research Flow

The research process was carried out through the following stages:

1. Problem Identification and Scoping

   The initial stage focused on developing a comprehensive understanding of the problem and user needs. The researcher identified, analyzed, and documented the existing conditions while clearly defining the research scope.

2. Literature Review

   The second stage involved reviewing previous studies relevant to the research topic. The findings from this stage were used as references to strengthen the research framework and methodology.

3. Data Collection

   The third stage consisted of collecting and managing both primary and secondary data. These data served as essential inputs for the analysis process and were used to support further research steps.

4. System Design

   At this stage, the researcher designed the system architecture and user interface for the chatbot. The design process included defining the components, workflow, and data processing techniques for the Natural Language Processing (NLP) module.

5. Model Development

   The chatbot development utilized Term Frequency–Inverse Document Frequency (TF-IDF) for text preprocessing and feature extraction. The TF-IDF method combines two key components. A term will have a high weight if it appears frequently in a specific document (high TF) but rarely across other documents (high IDF). This ensures that TF-IDF highlights words that are relevant for distinguishing topics.

6. Model Training

   After preprocessing, the extracted features were input into an Artificial Neural Network (ANN) model for training. The ANN was optimized using the Stochastic Gradient Descent (SGD) algorithm to improve classification performance.

7. System Testing and Evaluation

   The developed system was evaluated using accuracy testing and a Confusion Matrix to measure classification performance. If the results did not meet expectations, system refinement was conducted.

8. Conclusion Drawing

   Once the system achieved the desired level of performance, conclusions were drawn regarding the effectiveness of the implemented chatbot in addressing the identified research problem.

3. Result and Discussion

4.1 Problem Analysis

The problem analysis was carried out using the SWOT method (Table 2). The analysis highlights that the current system lacks interactivity and does not provide real-time responses outside operational hours. However, opportunities exist to integrate NLP-based chatbots using ANN optimized with SGD and TF-IDF

preprocessing. While this opens possibilities for adopting cutting-edge AI technology, continuous system updates and maintenance costs remain significant threats.

Table 1. SWOT Analysis

Strengths (Kekuatan)	Weaknesses (Kelemahan)
Sistem yang ada dapat dikembangkan lebih lanjut dengan fitur Chatbot NLP yang menggunakan model ANN dan optimasi SGD dengan preprocessing TF-IDF	Sistem saat ini kurang interaktif dan tidak memungkinkan calon mahasiswa untuk mendapatkan jawaban secara langsung terutama saat diluar jam operasional.
Opportunities (Peluang)	Threats (Ancaman)
Pengembangan sistem Chatbot NLP memungkinkan pemanfaatan teknologi terkini dalam bidang kecerdasan buatan serta dapat memberikan respon secara real-time kepada calon mahasiswa.	Teknologi terus berkembang, sehingga sistem yang dikembangkan perlu terus diupdate. Pengembangan dan pemeliharaan sistem chatbot membutuhkan biaya.

System Design and Implementation

The proposed system architecture was designed with four main components: NLP-based chatbot framework, ANN model design, user interface, and testing framework.

- NLP Workflow (Figure 4): Data preprocessing included case folding, tokenization, lemmatization, and TF-IDF transformation, ensuring textual inputs could be converted into numerical features.
- ANN Integration (Figure 2): After preprocessing, inputs were processed using an Artificial Neural Network optimized with Stochastic Gradient Descent (SGD). The ANN classified user queries into corresponding intents.
- Chatbot Flow (Figure 3): The system allowed users to interact via the website, where queries related to admissions were processed and responded to in real-time until the interaction ended.

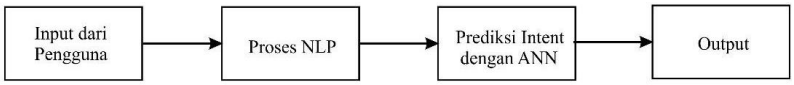

Figure 2. Chatbot Architecture

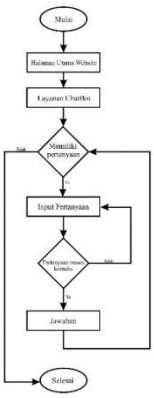

Figure 3. Chatbot Flowchart

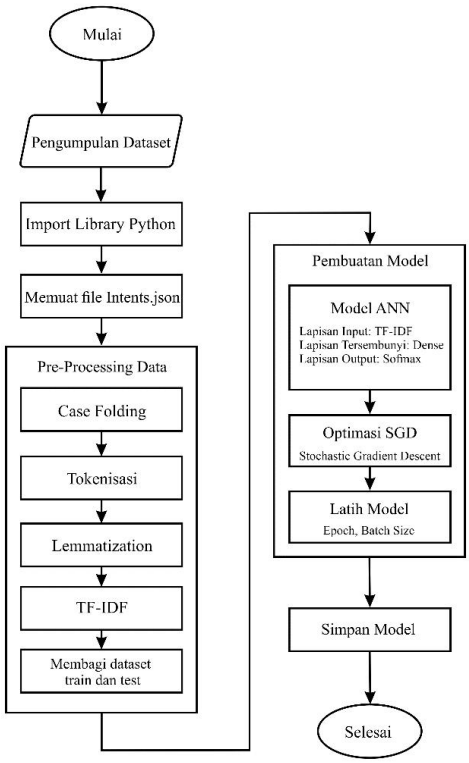

Figure 4. NLP Workflow

The chatbot was successfully implemented on the UBHINUS admissions portal, as shown in Figure 5. It was capable of interpreting informal queries, identifying relevant keywords, and classifying intents accurately.

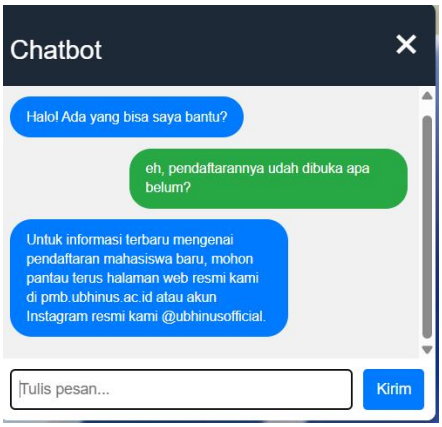

Figure 5. Chatbot User Interface

Classification Performance

The classification performance of the ANN model was evaluated using a classification report (Table 3).

- The model achieved an overall accuracy of 92% across 327 test samples.
- The macro average F1-score was 0.92, while the weighted average F1-score also reached 0.92.
- Several intents, such as *jadwal_daftar*, *tes_masuk*, and *informatika*, achieved perfect precision, recall, and F1-scores (1.00), indicating robust classification performance in specific categories.

Table 2. Classification Report

Intent	*Precision*	*Recall*	*F1-Score*	*Support*
sapaan	*1*	*0.6*	*0.75*	*5*
tanya	*0.8*	*0.67*	*0.73*	*6*
jadwal_daftar	*1*	*1*	*1*	*10*
cara_daftar	*0.89*	*1*	*0.94*	*8*
alur_daftar	*0.75*	*0.75*	*0.75*	*4*
tes_masuk	*1*	*1*	*1*	*9*
berkas	*1*	*0.94*	*0.97*	*17*
format_file	*1*	*1*	*1*	*8*
pasfoto	*1*	*1*	*1*	*3*
legalisasi_berkas	*0.89*	*1*	*0.94*	*8*
mengapa_ubhinus	*0.57*	*1*	*0.73*	*4*
prodi	*0.86*	*0.86*	*0.86*	*7*
informatika	*1*	*1*	*1*	*8*
lulusan_informatika	*1*	*1*	*1*	*3*
S1_SI	*0.83*	*1*	*0.91*	*5*
lulusan_SI_S1	*1*	*1*	*1*	*5*
D3_SI	*1*	*1*	*1*	*5*

This result demonstrates that the ANN model was able to generalize effectively across multiple classes, though some misclassifications occurred, especially for intents with overlapping linguistic patterns (e.g., *sapaan* misclassified as *unknown*).

Confusion Matrix Analysis

Further evaluation was conducted using a Confusion Matrix (Figure 6).

- The diagonal values indicate the number of correctly classified samples, which dominated across most intents.
- Misclassifications occurred in a few categories with semantically similar queries. For example, the intent *sapaan* was occasionally misclassified into *unknown* due to ambiguous phrasing by users.
- Despite minor errors, the confusion matrix shows that most user intents—especially admission-related queries (*jalur_daftar*, *jadwal_daftar*, *biaya_kuliah*)—were predicted accurately.

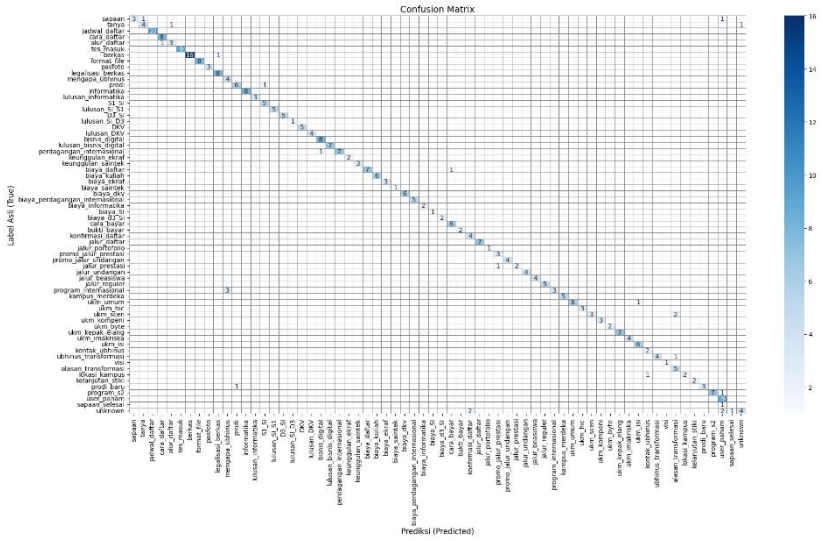

Figure 6. Confusion Matrix Results

Discussion

The results confirm that implementing an NLP-based chatbot with ANN and SGD optimization is effective in enhancing information services at UBHINUS. The chatbot achieved high accuracy and F1-scores, demonstrating strong classification performance.

- Strengths: High classification accuracy, real-time responsiveness, and the ability to handle informal queries.
- Weaknesses: Some misclassifications occurred due to overlapping language patterns.
- Implications: The chatbot improves user experience by providing instant, accurate, and relevant information outside operational hours, thereby increasing service accessibility.
- Future Improvements: Enhancing the dataset with more diverse examples and fine-tuning the ANN could further reduce classification errors, especially for ambiguous queries.

4. Conclusions

This study successfully developed and implemented a chatbot system for the New Student Admissions (PMB) website of Universitas Bhinneka Nusantara. The system was built by designing and training an Artificial Neural Network (ANN) model optimized using Stochastic Gradient Descent (SGD) with the intents.json dataset, and further integrating the chatbot interface into the PMB UBHINUS website.

The experimental results demonstrate that the proposed model performs effectively in classifying user queries, achieving an accuracy of 92.5% and an F1-score of 91.56%. These results indicate that the ANN-based chatbot is capable of correctly classifying most test data and providing relevant responses in real time.

Overall, the implementation of this chatbot proves to be a practical and reliable solution to improve the accessibility, responsiveness, and quality of information services in the university's admission process.

5. References

Furqan, M., Sriani, S., & Shidqi, M. N. (2023). Chatbot Telegram menggunakan Natural Language Processing. Walisongo Journal of Information Technology, 5(1), 15–26.

Hayat, M. A. (2024). Klasifikasi pengaduan pelayanan Fakultas Teknik Universitas Muhammadiyah Makassar menggunakan Natural Language Processing. Arus Jurnal Sains dan Teknologi, 2(2), 572–579.

Meilina, L., Kumara, I. N. S., & Setiawan, I. N. (2021). Literature review klasifikasi data menggunakan metode cosine similarity dan artificial neural network. Majalah Ilmiah Teknologi Elektro, 20(2), 307–314.

Putra, R. R., Putri, N. A., & Putra, A. D. (2024). Teknik cosine similarity dan TF-IDF dalam analisis data. Serasi Media Teknologi.

Ryansyah, A., & Andayani, S. (2017). Implementasi algoritma TF-IDF pada pengukuran kesamaan dokumen. Jurnal Sistem & Teknologi Informasi Komunikasi, 1(1), 1–10.

Sundari, E., & Nugroho, G. (2023). Pengaruh kualitas pelayanan akademik dan citra lembaga terhadap keputusan pemilihan mahasiswa perguruan tinggi Islam swasta di Kota Pekanbaru. Management Studies and Entrepreneurship Journal (MSEJ), 4(2), 1344–1352.

Zhafira, D. F., Rahayudi, B., & Indriati, I. (2021). Analisis sentimen kebijakan kampus merdeka menggunakan Naive Bayes dan pembobotan TF-IDF berdasarkan komentar pada Youtube. Jurnal Sistem Informasi, Teknologi Informasi, dan Edukasi Sistem Informasi, 2(1).