Published by :
https://www.ijert.org/
An International Peer-Reviewed Journal

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

AI-Based Student Assistance Chatbot Using Natural Language Processing and Machine Learning

Varunraj Sachin Gaikwad
Independent Researcher
Cygnet Public School
Pune, India

Abstract **- This study presents the development of an AI-based student assistance chatbot using Natural Language Processing and machine learning techniques. The system is designed to classify student queries into four categories: academics, examinations, mental health, and time management. A dataset of 200 queries was manually created to ensure originality and relevance.**

Text preprocessing was performed using normalization and stop word removal, and feature extraction was carried out using TF-IDF vectorization with n-gram features. Three machine learning models—Logistic Regression, Naive Bayes, and Support Vector Machine—were implemented and evaluated. Experimental results show that Logistic Regression and Support Vector Machine achieved the highest accuracy of 82.9 percent, outperforming Naive Bayes.

The system was further enhanced with a web-based interface using Streamlit and AI-based response generation for dynamic interaction. The results demonstrate that the proposed approach is effective for real-time student query classification and support.

Keywords - Artificial Intelligence, Natural Language Processing, Machine Learning, Chatbot, Text Classification, Student Support System, TF-IDF, Support Vector Machine, Logistic Regression, Naive Bayes, Conversational AI, Educational Technology

I. INTRODUCTION

In recent years, the rapid growth of Artificial Intelligence has led to the creation of systems that can assist users in real time across various areas. In education, students often face challenges related to academic workload, exam pressure, time management, and mental health. Traditional support like teachers, counselors, and peer networks, while effective, may not always be available for immediate help. This creates a need for intelligent solutions that can provide instant support and guidance.

Natural Language Processing, a branch of Artificial Intelligence, allows machines to understand and respond to human language. When combined with machine learning, it becomes possible to create systems that can classify user questions and generate meaningful answers. Chatbots are one popular application that has grown significantly due to their ability to interact in real-time and provide scalable support. In educational settings, chatbot systems can help students by answering questions, offering study advice, and providing basic mental health support. However, many existing systems depend on rule-based methods, which are not flexible, or on complex deep learning models that need large datasets and significant computational power. Therefore, there is a need for a straightforward, effective solution that balances performance and ease of use.

This study proposes developing an AI-based student assistance chatbot that uses machine learning to categorize student questions into four areas: academics, exams, mental health, and time management. The system uses TF-IDF vectorization with n-gram features for text representation and applies several machine learning algorithms, including Logistic Regression, Naive Bayes, and Support Vector Machine, for classification. The main goal of this research is to evaluate the performance of different machine learning models for text classification and to build a chatbot that can provide accurate and relevant responses to student questions. Additionally, the study includes an AI-based response generation system to improve conversational quality and the overall user experience. The rest of this paper is organized as follows. Section II outlines the methodology, including dataset preparation, preprocessing, and model development. Section III presents the results and performance evaluation of the models. Section IV wraps up the study and discusses future directions.

II. METHODOLOGY

The development of the proposed chatbot system was carried out through a structured and systematic approach, involving several key stages such as dataset creation, preprocessing, feature extraction, model training, and system integration. Each of these stages plays a critical role in ensuring that the overall system functions effectively and produces reliable results.

The process began with the careful design and creation of a dataset that reflects real-world student queries. This step was particularly important, as the quality and relevance of the data directly influence the performance of the machine learning models. Once the dataset was prepared, preprocessing techniques were applied to clean and standardize the text data.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

This helped in removing unnecessary noise and improving the consistency of the input.

Following preprocessing, feature extraction was performed to convert textual data into a numerical format that can be understood by machine learning algorithms. This stage is essential because it determines how well the model can interpret and learn from the data. The extracted features were then used to train multiple classification models, allowing for comparison and selection of the most effective approach.

Model training and evaluation were conducted in a controlled manner to ensure that the system could generalize well to unseen data. Different algorithms were tested to understand their strengths and limitations in handling natural language queries. Finally, all components were integrated into a unified system, where the trained model works in combination with a user interface to provide real-time responses.

A. DATASETS

Unlike many studies that rely on publicly available datasets, the dataset used in this research was manually created by the author. This approach was taken to ensure that the data closely reflects real-life student queries and is relevant to the intended application.
A total of 200 text samples were collected, covering common concerns faced by students. These queries were categorized into four main areas: academics, examinations, mental health, and time management. While designing the dataset, special attention was given to maintaining balance across all categories so that the model would not become biased toward any particular class.
The queries were written in a natural conversational style, similar to how students typically express their problems. This makes the chatbot more realistic and improves its ability to handle actual user input.

B. DATA PROCESSING

Raw text data often contains noise and inconsistencies, which can negatively affect model performance. Therefore, preprocessing was an essential step in preparing the dataset for machine learning.
First, all text was converted to lowercase to ensure consistency across the dataset. This prevents the model from treating the same word differently due to capitalization. Next, common stop words such as “the,” “is,” and “and” were removed, as they do not contribute significant meaning to the text.
Tokenization was then performed, where each sentence was broken down into individual words or tokens. This step helps the model analyse the structure of the text more effectively. Overall, these preprocessing steps helped simplify the data while preserving the important information required for classification.

C. FEATURE EXTRACTION

Since machine learning models cannot directly understand text, the processed data needed to be converted into a numerical format. For this purpose, TF-IDF (Term Frequency–Inverse Document Frequency) vectorization was used.
TF-IDF assigns weights to words based on how important they are within a particular query compared to the entire dataset. Words that appear frequently in a specific query but not across all queries are given higher importance, making them more useful for classification.
To further improve performance, n-gram features were included. Instead of looking at single words only, the model also considers combinations of words. This helps capture context and improves the system’s ability to understand phrases such as “exam stress” or “study schedule,” which carry more meaning than individual words.

D. MODEL DEVELOPMENT

Three different machine learning models were selected and implemented in this study: Logistic Regression, Naive Bayes, and Support Vector Machine.
Logistic Regression and Support Vector Machine were chosen because they are well-suited for handling high-dimensional data, which is common in text classification tasks. These models are known for their reliability and performance when used with TF-IDF features. Naive Bayes was included as a baseline model due to its simplicity and efficiency, allowing for a fair comparison between different approaches.
The dataset was divided into training and testing sets using an 80:20 split. The models were trained on the training data and then evaluated on the testing data to measure their performance. This approach ensures that the results are unbiased and reflect the model’s ability to generalize to new data.

III. SYSTEM ARCHITECTURE

The system architecture of the proposed chatbot is designed as a structured pipeline that processes user queries step by step, ensuring efficient classification and response generation. The architecture integrates Natural Language Processing techniques with machine learning models to provide accurate and real-time assistance to users.
At the initial stage, the system receives input from the user in the form of a text query through the chatbot interface. This interface is developed using Streamlit, allowing users to interact with the system in a simple and intuitive manner. The input query can vary in length and structure, as it is designed to handle natural conversational language.
Once the input is received, it undergoes a preprocessing stage where the text is cleaned and standardized. This includes converting the text to lowercase, removing unnecessary words, and breaking the sentence into tokens. This step ensures that the input data is consistent with the format used during model training.
Following preprocessing, the cleaned text is passed through the feature extraction module. In this stage, TF-IDF vectorization is applied to convert the textual data into numerical form. The use of n-gram features allows the system to capture contextual relationships between words, improving its ability to understand user intent.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

The processed input is then fed into the trained machine learning model. Based on the learned patterns, the model predicts the most appropriate category for the given query. This classification step is crucial, as it determines the type of response that will be generated.

After the category is identified, the system moves to the response generation stage. Here, the chatbot produces a reply based on the predicted category. In addition to predefined responses, an AI-based mechanism is integrated to generate more natural and context-aware replies, enhancing the overall interaction experience.

Finally, the generated response is displayed to the user through the interface, completing the interaction cycle. The entire process occurs in real time, ensuring that users receive instant feedback to their queries.

The modular design of the system allows for easy scalability and future improvements. Additional categories, advanced models, or enhanced response mechanisms can be incorporated without significantly altering the existing architecture.

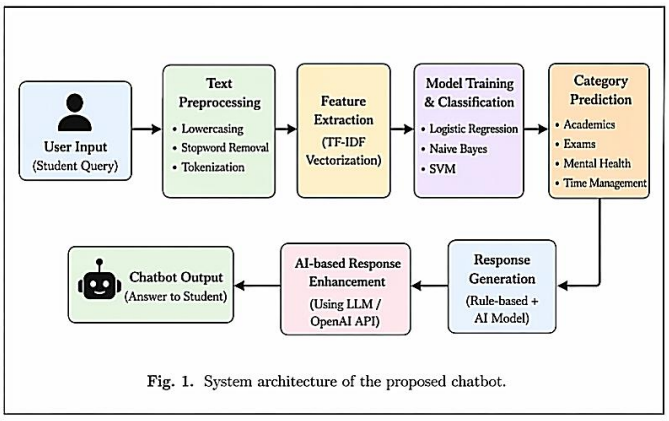

Fig. 1. System architecture of the proposed chatbot.

Fig. 1. System architecture of chatbot

IV. RESULTS AND DISCUSSION

This section presents a comprehensive evaluation of the proposed chatbot system, focusing on both quantitative performance metrics and qualitative interpretation of the results. The aim is not only to compare the effectiveness of different machine learning models but also to understand how well the system performs in handling realistic student queries in practical scenarios.

This work emphasizes applied performance—how accurately and reliably the chatbot can classify queries that resemble real student concerns. Therefore, multiple evaluation metrics were used, and the results were analysed from both a statistical and practical perspective.

A. MODEL PERFORMANCE

To evaluate the classification capability of the system, three machine learning algorithms—Logistic Regression, Naive Bayes, and Support Vector Machine—were implemented and tested on the dataset. These models were selected to represent different approaches to text classification, ranging from probabilistic methods to linear and margin-based classifiers.

The performance of each model was measured using accuracy, which indicates the proportion of correctly classified queries out of the total number of test samples. The results are summarized in Table I.

Model	Accuracy
Logistic Regression	82.9%
Naive Bayes	78.0%
Support Vector Machine	82.9%

Table I: Model Accuracy Comparison

From the results, it can be clearly observed that both Logistic Regression and Support Vector Machine achieved the highest accuracy of 82.9 percent. This indicates that these models are better suited for handling text-based classification problems, especially when combined with TF-IDF feature representation. Naive Bayes, while efficient and computationally less demanding, achieved a slightly lower accuracy of 78.0 percent. This difference highlights the limitations of the model, particularly its assumption that features are independent of each other. In natural language data, words often have contextual relationships, and ignoring these relationships can reduce classification performance.

Overall, the results suggest that models capable of capturing feature interactions and handling high-dimensional data are more effective for this type of problem.

B. ACCURACY COMPARISON

To provide a more intuitive understanding of the differences in model performance, the accuracy values were also visualized using a bar graph

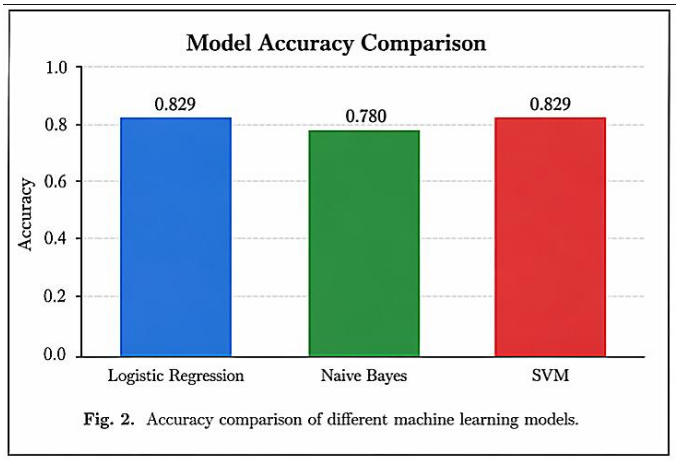

Fig. 2. Accuracy comparison of different machine learning models.

Fig. 2. Accuracy comparison of models

The graphical representation reinforces the numerical findings. Logistic Regression and Support Vector Machine show nearly identical performance, forming the top-performing group, while Naive Bayes lags slightly behind.

This trend can be explained by the nature of TF-IDF features, which produce high-dimensional sparse vectors. Linear models such as Logistic Regression and Support Vector Machine are particularly well-suited for such data, as they can efficiently separate classes using linear decision boundaries.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

Naive Bayes, on the other hand, relies on probability distributions and independence assumptions, which may not fully capture the complexity of natural language patterns.

C. PERFORMANCE EVALUATION

While accuracy provides a general overview, it does not capture the full behaviour of the model across different categories. Therefore, additional metrics such as precision, recall, and F1-score were used to evaluate performance in greater detail.

Classification Report (Logistic Regression)

Class	Precision	Recall	F1-Score	Support
academics	0.86	1.00	0.92	6
exams	0.91	0.62	0.74	16
mental_health	0.69	0.90	0.78	10
time_management	0.90	1.00	0.95	9
accuracy			0.83	41
macro avg	0.84	0.88	0.85	41
weighted avg	0.85	0.83	0.82	41

Fig. 3. Classification report of the Logistic Regression model.

Fig. 3. Classification Report

The overall accuracy of the model is 0.83, with a macro average F1-score of 0.85 and a weighted average F1-score of 0.82.

A closer look at the classification report reveals several important insights. The **academics** category demonstrates excellent performance, with perfect recall and high precision. This indicates that the model is highly effective at identifying academic-related queries without missing relevant instances.

Similarly, the **time management** category shows strong performance, achieving both high precision and recall. This suggests that queries related to scheduling, productivity, and planning are relatively distinct and easier for the model to classify.

The **mental health** category presents slightly lower precision but high recall. This means that while the model successfully identifies most mental health-related queries, it occasionally misclassifies queries from other categories as mental health-related. This can be attributed to overlapping language patterns, as emotional expressions may appear in multiple contexts.

The **examinations** category shows the lowest recall among all categories. This indicates that the model fails to identify some exam-related queries correctly. One possible reason is that exam-related language often overlaps with academic queries, making it more difficult for the model to distinguish between the two.

D. CONFUSION MATRIX ANALYSIS

For further analyse of classification performance, a confusion matrix was generated.

```
[[ 6  0  0  0]
 [ 1 10  4  1]
 [ 0  1  9  0]
 [ 0  0  0  9]]
```

The confusion matrix provides a detailed breakdown of correct and incorrect predictions. Most values are concentrated along the diagonal, indicating correct classifications, which confirms the overall effectiveness of the model.

However, some off-diagonal values can be observed, particularly in the examinations row. This shows that a few exam-related queries were misclassified into other categories such as academics or mental health. This behavior is expected in Natural Language Processing tasks, where different categories may share similar vocabulary or context.

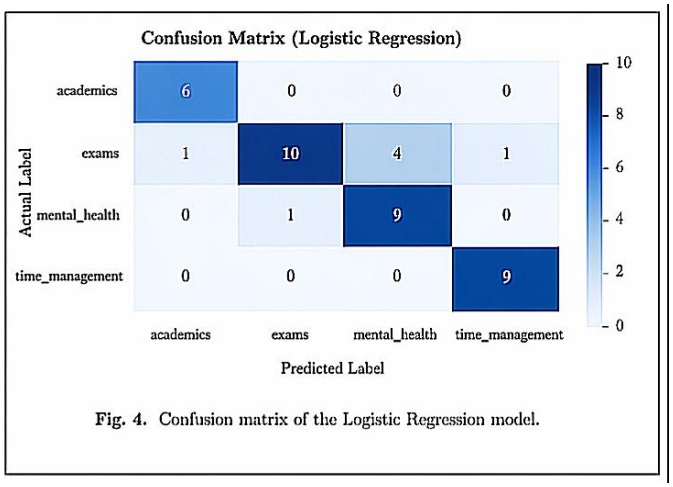

Fig. 4. Confusion matrix of the Logistic Regression model.

Fig. 4. Confusion Matrix

For example, a query like “I am stressed about my exams” could be interpreted as either an examination-related query or a mental health concern. Such ambiguity makes classification inherently challenging and explains the observed misclassifications.

E. DISCUSSION

The overall results demonstrate that Logistic Regression and Support Vector Machine are more effective for text classification tasks compared to Naive Bayes in this context. Their ability to handle high-dimensional feature spaces and capture relationships between features allows them to perform better when working with TF-IDF representations.

Another important observation is that the performance of the model is influenced not only by the algorithm but also by the nature of the dataset. Since the dataset contains real-world-like queries, it includes variations, ambiguities, and overlapping language patterns. These factors introduce complexity but also make the system more realistic.

Despite these challenges, the chatbot achieves a strong overall performance, indicating that the methodology is effective. The integration of an AI-based response generation system further enhances usability by providing more natural and engaging responses. This makes the chatbot more suitable for practical

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

deployment, as it can interact with users in a conversational manner

From an application perspective, the system demonstrates the potential to serve as a supportive tool for students. It can provide immediate assistance, reduce dependency on human intervention for basic queries, and improve accessibility to guidance resources.

While the current system performs well, there is scope for improvement. Increasing the dataset size, incorporating more diverse queries, and using advanced deep learning models could further enhance accuracy and contextual understanding.

V. CONCLUSION AND FUTURE WORK

This study presented the design and implementation of an AI-based student assistance chatbot using Natural Language Processing and machine learning techniques. The primary objective was to develop a system capable of understanding and classifying student queries into meaningful categories and providing appropriate responses in real time. The results demonstrate that the proposed approach is both effective and practical for addressing common student concerns.

Through the experimental evaluation, it was observed that Logistic Regression and Support Vector Machine achieved the highest classification accuracy of 82.9 percent, outperforming Naive Bayes. This confirms that linear models are well-suited for handling high-dimensional text data, especially when combined with TF-IDF vectorization. The additional evaluation using precision, recall, and F1-score provided deeper insights into the model's performance across different categories, highlighting both strengths and areas for improvement.

One of the key strengths of this system is its ability to handle real-world style queries. By using a manually created dataset that reflects natural student language, the chatbot becomes more practical and relatable in actual usage scenarios. The integration of an AI-based response generation mechanism further enhances the system by allowing it to produce dynamic and context-aware replies, improving the overall user experience compared to static response systems.

Another important contribution of this work is the combination of simplicity and effectiveness. Instead of relying on complex deep learning models, the system uses relatively simple machine learning techniques that are easier to implement and require fewer computational resources, while still achieving strong performance. This makes the proposed solution accessible and scalable, especially in environments with limited resources.

However, like any system, the proposed chatbot has certain limitations. The dataset size is relatively small, which may restrict the model's ability to generalize to a wider range of queries. Additionally, some misclassifications were observed, particularly in categories where the language overlaps, such as academics and examinations. These limitations highlight the challenges associated with natural language understanding and indicate areas where further improvements can be made.

For future work, several enhancements can be considered. Expanding the dataset with more diverse and complex queries would improve the robustness of the model. Incorporating advanced techniques such as deep learning models or transformer-based architectures could further enhance classification accuracy and contextual understanding. Additionally, improving the response generation mechanism by integrating more sophisticated conversational AI models can make the chatbot more interactive and human-like.

In conclusion, the proposed chatbot system demonstrates that a well-designed combination of Natural Language Processing and machine learning techniques can effectively support students by providing real-time assistance. The system has strong potential for practical deployment and can be further developed into a more advanced intelligent support system for educational environments.

VI. REFERENCES

[1] [1] D. Jurafsky and J. H. Martin, *Speech and Language Processing*, 3rd ed. Stanford University, 2023.

[2] [2] F. Pedregosa *et al*., "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[3] [3] S. Bird, E. Klein, and E. Loper, *Natural Language Processing with Python*. Sebastopol, CA, USA: O'Reilly Media, 2009.

[4] [4] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*. Cambridge, U.K.: Cambridge University Press, 2008.

[5] [5] T. Mitchell, *Machine Learning*. New York, NY, USA: McGraw-Hill, 1997.

[6] [6] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. Cambridge, MA, USA: MIT Press, 2012.

[7] [7] A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, 2nd ed. Sebastopol, CA, USA: O'Reilly Media, 2019.

[8] [8] Streamlit Inc., "Streamlit Documentation," 2024. [Online].

[9] [9] OpenAI, "OpenAI API Documentation," 2024. [Online].

[10] [10] J. Brownlee, *Machine Learning Mastery with Python*. 2016.

[11] [11] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 3rd ed. Upper Saddle River, NJ, USA: Prentice Hall, 2010.