INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD)
© 2026 IJNRD | Volume 11, Issue 5, May 2026 | ISSN: 2456-4184 | IJNRD.ORG

AI-BASED COLLEGE ENQUIRY CHATBOT USING NLP AND INTENT RECOGNITION

Arjun. S [1]
(M.Sc. Data Science and Business Analytics)
Department of Advanced Computing & Analytics
Vels Institute of Science, Technology & Advanced Studies
Chennai, India
Email: sankararjun349@gmail.com

Dr. K. Abirami [2]
Assistant Professor
Department of Advanced Computing & Analytics
Vels Institute of Science, Technology & Advanced Studies
Chennai, India
Email: abiramidharmarajan@gmail.com

Abstract- ***There has been an increase in the use of automated systems due to the expansion of digital technology. One example is the high volume of inquiries received at institutions of higher education from prospective and current students regarding the admissions process, courses, tuition, placement, and lodging, among other topics. As these inquiries continue to rise in number, institutions seek a means to assist them in answering such inquiries by the development of an AI-based College Enquiry Chatbot. The purpose of this project is to design and create the AI college inquiry chatbot with the objective of resolving the issues originated by the large volume of inquiries. The College Enquiry Chatbot will employ some combination of Natural Language Processing and intent recognition technologies to process and respond to a user's inquiry efficiently. The chatbot will be created using three steps: processing the inquiry using Natural Language Processing techniques; normalizing the user's inquiry; and converting the user's inquiry into a set of numerical features using TF-IDF vectorization; and testing six (6) machine learning classification algorithms to determine the best algorithm for use in the creation of the College Enquiry Chatbot.***

Keywords - Artificial Intelligence, Natural Language Processing, College Enquiry Chatbot, Intent Recognition, Machine Learning, TF-IDF, React, Flask API.

I. INTRODUCTION

Currently in an increasingly electronic and rapid paced world there continues to be an ongoing challenge for schools and universities with responding to many requests from students regarding admission to programs (courses), tuition fees, placement opportunities, dormitory (hostel) services and scholarships. All of the methods used by schools/colleges to respond to these requests consisted of making telephone calls, sending e-mail, meeting face-to-face at the school or university) prior to COVID 19. The aforementioned methods of communication have historically worked for smaller institutions. However, these methods are inefficient during peak seasons for admissions. The advent of Artificial Intelligence and Natural Language Processing technologies enables the efficient handling of these queries through the development of an intelligent chatbot. The traditional chatbots were designed to respond to limited queries and were rendered useless if the query was paraphrased. The modern-day machine learning-based chatbots are capable of understanding the meaning of the query irrespective of the wording. This project intends to develop an AI-Based College Enquiry Chatbot through TF-IDF Vectorization and six machine learning algorithms.

A. Background and Motivation

With an increase in the need for higher education, there is a direct increase in the volume of queries to such institutions. This poses a problem to administrative staff, whose numbers are limited, to respond to queries in a consistent and accurate manner. The objective of this venture is to create a smart chatbot capable of giving immediate answers to students' questions anytime they ask them. This will lead to a decrease in the number of hours the administration works on the average day.

B. Problem Statement

Problem Definition Manual response to student inquiries is both difficult and laborious. However, for many institutions, this difficulty is magnified during admissions time. Furthermore, there are likely to be delays due to the amount of queued inquiries. Therefore, an automated system is necessary to allow institutions to respond to a large volume of inquiries as consistently as possible and be able to do so 24 hours a day, seven days a week.

II. LITERATURE SURVEY

Some researchers have also explored the use of chatbot systems in educational institutions for the automation of queries from students. The early research indicated the use of cloud-based NLP services, such as IBM Watson, for handling queries from students regarding admissions, and the accuracy of these systems was above 95%. This indicated that these systems can be used for replacing humans in handling queries from students. In the later research, the introduction of a 24/7 web-based system for handling queries from students was explored, which indicated a reduction in calls for help desk queries regarding academic issues from students. Some of these methods used direct connections to institutional databases via keyword matching. However, these methods faced considerable difficulty when students used informal language to pose their queries. To address these challenges, later research used sentiment analysis and active learning methods to improve the system progressively.In later research on chatbots, attempts were made to improve their accessibility by incorporating both text and speech-based interfaces. This was done to ensure greater accessibility for a wider audience. Surveys conducted on various frameworks revealed that machine learning-based methods for query intent classification were superior to rule-based dialogue systems when there was considerable variation in language used by different users.

III. PROPOSED WORK

The proposed system is based on a pipeline system in which each stage of the system, from the ingestion of raw data to the final output response, is independently testable and replaceable. The proposed methodology has six stages.The first stage in the proposed methodology is the creation of a custom dataset of approximately 600 queries for students in 12 distinct intent categories: Admissions, Courses and Branches, Fee Structure, Scholarship Options, Hostel Facilities, Placement Statistics, Exam Schedule, Faculty Details, Transportation, Library Facilities, Contact Details, and Greetings. The queries in each of these categories are created with varying degrees of formality. Secondly, there is a series of preprocessing steps that are performed on the raw text data from the students, and this includes lowercase, punctuation, tokenization, stopwords, and Porter stemming. This is followed by tokenization to get a clean text string. This text string is then fed into a TF-IDF Vectorizer. This is followed by fitting the TF-IDF Vectorizer into the training data to avoid data leakage during validation and testing.Lastly, there are six classifiers that are utilized to train the data, and they include Naive Bayes, Logistic Regression, SVM, Random Forest, KNN, and Decision Tree classifiers. This is followed by an 80/20 split and cross-validation to get the best classifier from the classifiers. The classifiers are then serialized and hosted via a Flask REST API and a React front end with chat interface and timestamp bubbles and confidence badges.

IV. SYSTEM ARCHITECTURE

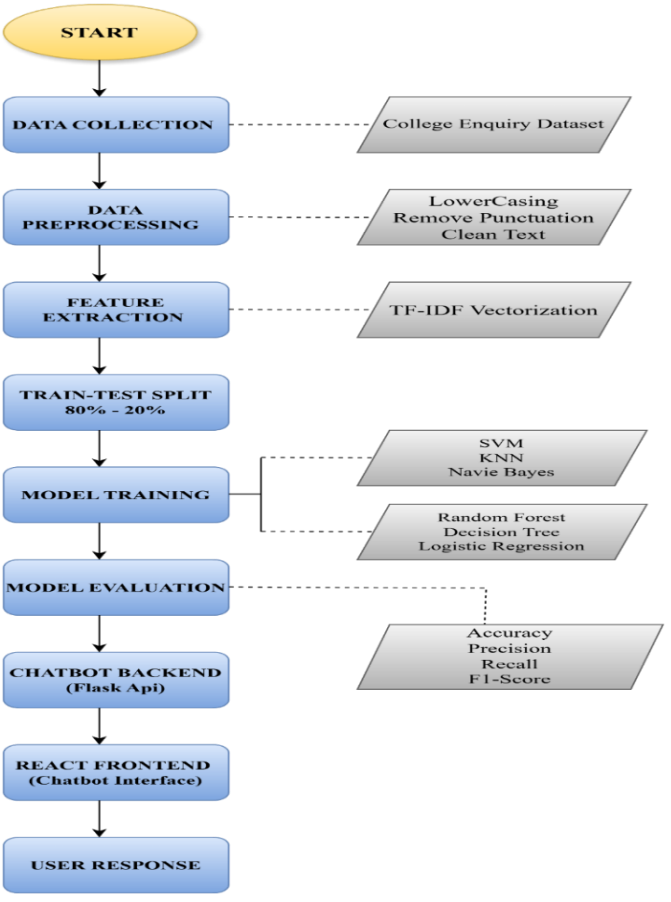

Fig 1: Architecture Diagram

The process begins by creating a College Enquiry Dataset. The data is then preprocessed by changing all the characters to lower case and removing the punctuation. Then, the text data is converted to numerical data by applying the TF-IDF Vectorization method. The data is then divided into training data and testing data, i.e., 80% and 20%, respectively. Then, six algorithms are trained, and the performance of the data is checked by evaluating the accuracy, precision, and F1 score to find the best algorithm. Then, the best algorithm is integrated with the chatbot, and immediate response is given to the students.

V. PROJECT PHASES /METHODOLOGIES

Developing the AI-Based College Enquiry Chatbot was broken into four phases; each focused on one area of development.

Phase 1 - Data Collection and Preparation

To create an entire dataset for the chatbots in the college inquiry system, it was not possible to use a previously-built dataset. There were a total of 12 different intent types that were used for creating queries of the most commonly asked query types by students. The query types include but are not limited to Admissions, Fee Structured, Courses, Scholarships, Hostel Facilities, Placement Statistics, Exam Schedules, and Contact info. Each of the query types had between 40-60 different queries. All queries in the dataset were written differently, some in proper complete sentences, some using slang, others using acronymized forms for some words and others containing misspelled words to enable the chatbot to recognize/query/perform the task regardless of how the user asked. JSON is the resulting output of the final dataset, each query contains an intent value.

Phase 2 - Model Development

It starts with cleaning the query data. This phase included the conversion of all text to lower case; removal of punctuation; removal of all stop words; and applying Porter stemming. After cleaning is complete, the cleaned queries will be changed to numeric feature vectors using TF-IDF vectorization. The data will then be split into two sets to be used in this study for training and testing. Data will be split into 80% for training and 20% for testing using stratified sampling. The classifiers used to model will include Naive Bayes, Logistic Regression, SVM, Random Forest, KNN, and Decision Tree. A 5-fold cross-validation method will be used to provide each of the classifiers with consistent training and similarly created datasets for each classifier will use the 5-fold validation process. The final classifier model was selected based upon its overall performance; however, each of the classifiers had potential. Ultimately, the Support Vector Machine Classifier was employed due to its potential demonstrated by each and every classifier.

Phase 3 - Interface Development

A chat interface using React has been created. The chat interface has been given a simple layout. The chat interface has been given a timestamp-based conversation bubble effect. A confidence level badge has been placed at the bottom of all bot responses. A 'Clear Chat' button has been included in the chat interface to clear all the conversations.The interface communicates using HTTP requests via Axios.

Phase 4 - Testing and Going Live

Unit tests showed that pre-processing was performed accurately with all possible exceptions; integration testing confirmed that the format of responses returned from the Flask API was valid. A load test demonstrated to us that the system could withstand large volumes of traffic, regardless of how many requests were made simultaneously. The backend has been Dockerized and deployed onto a cloud server; the React front-end has been built using Node.js, allowing access to ChatBot via any web browser at any time.

VI. INPUT

The chatbot receives questions in natural language. Students enter questions through a web interface. Questions may be long or short. It shows examples of questions from all categories.

Admissions	How do I apply for admission?
Fee Structure	How much is the annual tuition?
Courses	Which engineering branch are available?
Scholarships	Any scholarship for low-income students?
Hostel	Is there a separate hostel for girls?

Placement	What is the average package offered?
Greeting	Hello! Can you help me?
Contact	How can I reach the admission office?

VII. OUTPUT

The student will ask the system for an answer to their question. The system will provide an answer in JSON format with three attributes. The first attribute provided will give a free-text explanation of how the system arrived at the given answer. The second attribute provided will give an intent label that indicates what the system believes the student was attempting to do through their original question. The third attribute provided will give a confidence score, ranging from zero (no certainty) to one (absolute certainty), indicating how certain the system is of its answer. The front-end system will use this information to generate a response to the student question, as well as provide the student with a badge indicating the level of confidence assigned by the system to its answer to their original question.

Query	Intent	Confi
What is the B.Tech fee?	fee_structure	96%
Hostel available for boys?	hostel_facilities	93%
Which companies came last year?	placement_records	91%
How many seats in CSE?	courses_offered	89%
Hello	greeting	99%
Any merit scholarship?	scholarships	88%

VIII. RESULT AND DISCUSSION

Performance of the classifiers is evaluated using 20% of data that was withheld as test data. The evaluation includes overall accuracy, macro averaged precision, recall, and F1 scores for each of the six classifiers. Macro averaging assumes that each intent is equally important regardless of its support count. This makes macro averaging the best method for evaluation in this case because of the balance in the class distributions across the data sets.The SVM classifier was the best performing of the classifiers, achieving an accuracy of 94.7%. Linear SVMs generally perform well in high-dimensional sparse feature spaces, like TF-IDF feature space, because they have the property of margin maximization which guarantees a stable decision boundary will be found, even in sparse feature space.Logistic regression was a close second with an accuracy of 92.1%. The performance of logistic regression as a baseline classifier in text-based classification supports its use as an all-around reliable classifier for this type of classification. Its strong performance, compared to other more complex models such as Random Forest, is in line with existing literature that states simpler models generalize well in feature spaces where the features are already meaningful.

Classifier	Acc.	Prec.	Rec.	F1
SVM	94.7	94.3	94.1	94.2
Logistic Reg.	92.1	91.8	91.5	91.6
Random Forest	89.4	89.0	88.6	88.8
Naive Bayes	86.7	86.2	85.9	86.0
KNN	83.5	83.1	82.7	82.9
Decision Tree	80.2	79.7	79.3	79.5

IX. SCREENSHOTS

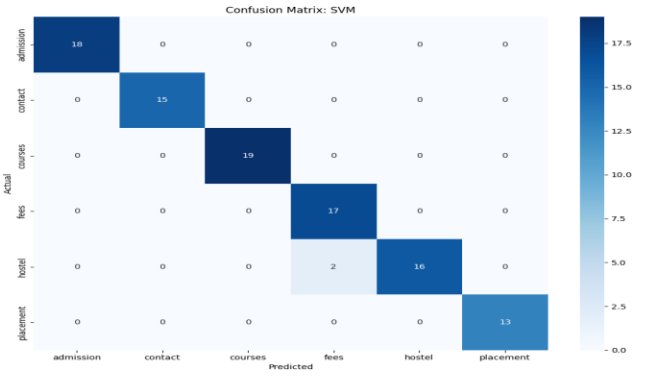

Fig2:Confusion Matrix of the SVM Model

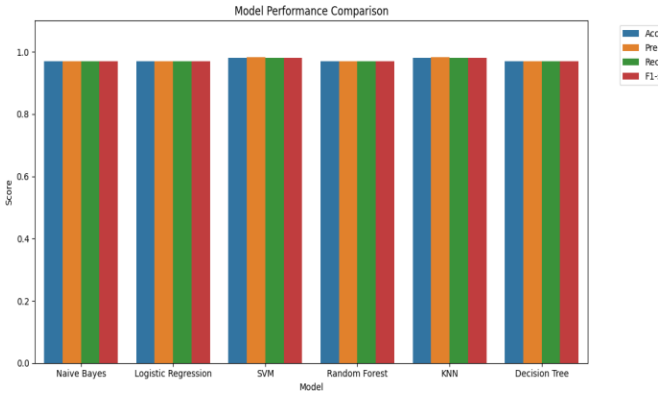

Fig3: Model Performance Comparison

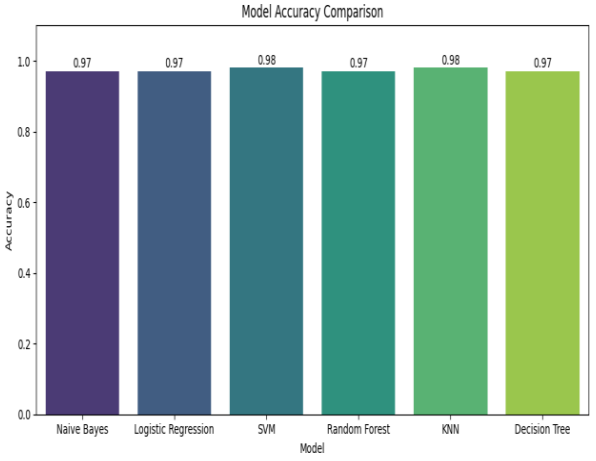

Fig4: Model Accuracy comparsion

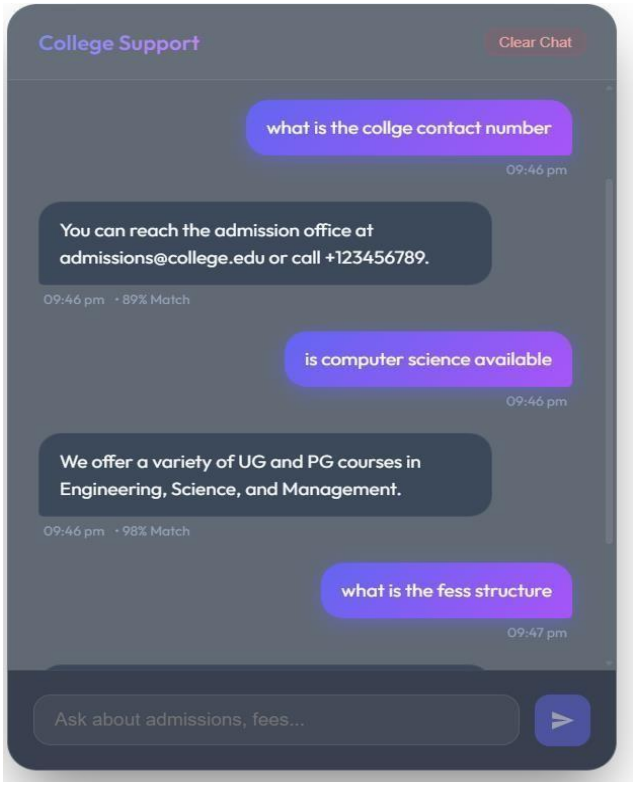

Fig5: Chatbot Interface Output

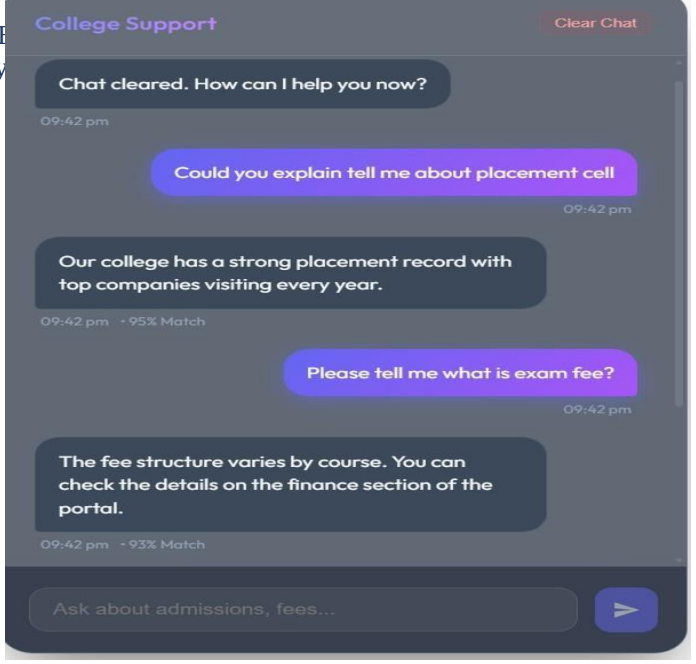

Fig6: Chatbot Interface Output

X. CONCLUSION

The AI-based College Enquiry Chatbot has been designed to address the ever-increasing problem of student queries in educational institutes effectively. The chatbot is capable of giving precise and accurate answers to queries without the need for human intervention. Out of the six classifiers used, the Support Vector Machine Classifier was found to be the best with an accuracy rate of 94.7%, including the intent classification too. Implementation of the project can now begin as the chatbot is available for use to reduce administrative staff workloads and provide accurate, timely information for students at all times. The development of the project demonstrates that artificial intelligence can be applied to solve real-world problems proficiently. The project can be enhanced to provide support for more languages and incorporate the use of deep learning techniques.

XI. REFERENCES

[1] L. O. Gbenga, T. O. Okedigba, and H. O. Aworinde, “An Improved Rapid Response Model for University Admission Enquiry System Using Chatbot,” International Journal of Computer (IJC), vol. 38, no. 1, pp. 123–131, 2020.

[2] O. L. S. S. Kumari, AI Based Student Enquiry System, Master’s Dissertation, University of Colombo School of Computing, Sri Lanka, 2020.

[3] R. Vijayakumar, B. Bhuvaneshwari, S. Adith, and M. Deepika, “AI Based Student Bot for

Academic Information System using Machine Learning," International Journal of Scientific Research in Computer Science, Engineering and Information Technology, vol. 5, no. 2, pp. 590–596, 2019.

[4] U. K. Bavishi, Implementing a College Enquiry Chatbot, Master's Project Report, Department of Computer Science, California State University, Sacramento, USA, 2019.

[5] A. Halvankar, A. Chaudhari, A. Wable, and P. Barkund, "College Enquiry for Student using AI ChatBot," International Journal of Engineering Research, 2021.

[6] A. Phalle, S. Kadam, S. Sonphule, and I. Savant, "AI and Web-Based Interactive College Enquiry Chatbot," International Research Journal of Engineering and Technology (IRJET), vol. 8, no. 11, pp. 1369–1373, 2021.

[7] G. Khandagale, M. Wagh, P. Patil, and H. Lad, "Intelligent Chatbot for College Enquiry System: A Survey," International Research Journal of Engineering and Technology (IRJET), vol. 8, no. 10, pp. 1998–2002, 2021.

[8] R. Tiwari, R. Khandelwal, Y. Agrawal, V. Tiwari, and W. H. Bisen, "AI Chatbot for College Enquiry," International Journal of Engineering and Management Research, vol. 13, no. 2, pp. 90–93, 2023.

[9] S. Singh and S. Choudhary, "Chatbot for University Enquiry Management System," Journal Publication of International Research for Engineering and Management (JOIREM), vol. 3, no. 12, 2025.

[10] S. Pawar, O. Rane, O. Wankhade, and P. Mehta, "A Web Based College Enquiry Chatbot with Results," International Journal of Computer Applications, 2018.