International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 05 | May - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

AI-Driven College Admissions and Counselling *

1st Dr.Radhakrishna Naik
Department of Computer Engineering, Sanjivani College of Engineering,
Kopargaon, India
radhakrishnanaikcomp@sanjivani.org.in

2nd Swapnil Jadhav
Department of Computer Engineering, Sanjivani College of Engineering,
Kopargaon, India
swapniljadhav3181@gmail.com

3rd Rajwardhan jadhav
Department of Computer Engineering, Sanjivani College of Engineering,
Kopargaon, India
rajwardhanjadhav7666@gmail.com

***Abstract*—Over the past few years, the integration of Artificial Intelligence (AI) in education has significantly improved student support systems. However, the college admissions and counseling process remains largely manual, leading to inefficiencies and limited accessibility for students seeking personalized guidance. Traditional methods often fail to provide real-time, data-driven assistance, making it challenging for students to make informed decisions about college selection and admission criteria.This paper presents an AI-Driven College Admissions and Counseling Assistant, designed to streamline the admissions process by offering personalized, real-time responses to student inquiries. The system utilizes machine learning and Natural Language Processing (NLP) techniques to analyze structured data from official admission portals, including cutoffs, rankings, and seat availability. The model is built using a Retrieval-Augmented Generation (RAG) approach, where student queries are processed using a Chroma vector database, and responses are generated by OpenAI's GPT model. The predictive component of the system was evaluated using multiple classifiers, including Support Vector Machine (SVM), Random Forest, and Logistic Regression, achieving an accuracy of 87.5% with an AUC of 0.94%[1]. Further optimization with feature selection techniques improved accuracy to 90%, ensuring high precision in query classification and response generation.The counseling assistant is deployed on scalable cloud platforms such as Azure, Render, and Hostinger, providing seamless access through a user-friendly web interface. By automating admissions counseling, the system enhances accessibility, minimizes student uncertainty, and empowers them with relevant, data-backed insights to make well-informed academic decisions.**

Keywords -*1. AI in education, 2. College admission counseling, 3. Machine learning in student guidance, 4. AI chatbot for admissions, 5. NLP-based admission query classification, 6. Predictive analytics in education.*

I. INTRODUCTION

College admissions can be a complex and overwhelming process for students, requiring them to navigate various eligibility criteria, cutoffs, rankings, and seat availability across multiple institutions[1]. Traditional counseling methods, such as in-person group sessions, often fail to provide personalized guidance, leaving students with limited resources to make well-informed decisions about their academic futures. With advancements in Artificial Intelligence (AI) and Machine Learning (ML), there is a growing need for automated, data-driven solutions that can offer accurate, real-time counseling tailored to individual student needs.To address these challenges, this project introduces the AI-Driven College Admissions and Counseling Assistant, a system designed to enhance and automate the college counseling experience[6]. The system leverages Natural Language Processing (NLP) and machine learning models to analyze structured data from official admission portals, providing students with personalized insights on admission cutoffs, eligibility, and college selection. The Retrieval-Augmented Generation (RAG) approach, which combines Chroma vector databases with OpenAI's GPT model, ensures context-aware and precise responses to student inquiries. Additionally, the predictive component employs Support Vector Machine (SVM), Random Forest, and Logistic Regression, achieving high accuracy in query classification and response generation[2].Beyond information retrieval, the system also includes a recommendation engine that suggests alternative colleges based on a student's academic background and preferences. This approach empowers students by expanding their options beyond their first-choice institutions and helping them explore viable alternatives. The platform is deployed on scalable cloud services like Azure, Render, and Hostinger, ensuring 24/7 accessibility through a user- friendly web interface.By integrating AI into college admis- sions counseling, this project bridges the gap between students and valuable admission resources, reducing uncertainty and improving decision-making. The system not only enhances accessibility and efficiency but also sets the foundation for future advancements in AI-driven educational support, making college admissions a more transparent and informed process for students worldwide.

II. LITERATURE REVIEW

Majjate et al. (2024) propose an AI-powered academic guidance system that personalizes student recommendations based on their profiles and interests. By integrating machine learning models, the system enhances decision-making and provides tailored career and course suggestions, improving student engagement compared to traditional counseling methods[1]. Agrawal et al. (2024) present an AI-driven college admissions system that automates the evaluation of student applications using predictive analytics and NLP. Their research highlights how AI improves efficiency, reduces manual workload, and

enhances transparency in admission processes by analyzing academic performance, entrance exam scores, and extracurricular activities[2]. Kuhail et al. (2024) investigate the role of AI chatbots and virtual assistants in student counseling. Their study demonstrates how personalized AI recommendations help students navigate academic decisions while reducing stress. However, they raise concerns about bias in AI models and data privacy issues, emphasizing the need for ethical AI frameworks in education[3].Majjate et al. (2023) focus on AI-based admission prediction, using machine learning classification models to estimate student acceptance rates. Their findings suggest that AI-driven predictions improve admission forecasting by 30 compared to traditional statistical models, reinforcing the importance of explainable AI (XAI) to ensure fairness in decision-making[4].Van Busum and Fang (2023) analyze different AI models used in student admissions, comparing machine learning classifiers such as random forests, neural networks, and deep learning models. They emphasize the importance of fair and transparent AI algorithms, highlighting challenges like algorithmic bias and fairness in AI-driven admissions[5]. Keir et al. (2023) explore AI's role in higher education, focusing on its impact on automated decision-making, personalized learning, and student retention. Their study discusses how AI-based recommendation systems improve engagement but also addresses concerns about data security, ethical AI governance, and regulatory compliance[6].

III. SCOPE

1) **Admission Prediction –** Estimates a student's probabil- ity of acceptance using Huber Regressor and other ML models.
2) **University Recommendations –** Suggests top 5 similar universities based on eligibility and preferences.
3) **User-Friendly Interface –** Developed using Streamlit (Frontend) and FastAPI (Backend) for seamless interac- tion.
4) **Data-Driven Insights –** Utilizes CET admission cutoffs and historical trends for accurate predictions.
5) **Scalability Security –** Hosted on Render/AWS with PostgreSQL/Firebase for data storage.

IV. OBJETIVES

1) **Predict Admission Chances –** Use ML models (Huber Regressor, XGBoost) to estimate student acceptance probabilities.
2) **Recommend Universities** – Suggest top 5 similar col- leges based on eligibility and preferences.
3) **User-Friendly Platform –** Develop an interactive web app (Streamlit FastAPI) for real-time guidance.
4) **Data-Driven Insights –** Analyze CET admission data for accurate predictions.
5) **Scalable Future-Ready –** Deploy on Render/AWS with plans for AI chatbots real-time updates.

V. CHALLENGES

1) **Limited Access to Real**-Time Data – Universities fre- quently update their cutoffs, making real-time admission predictions difficult.
2) **Inconsistent or Incomplete Data –** Extracting data from CET portals PDFs requires extensive cleaning, normalization, and handling missing values.
3) **Handling Outliers Unpredictable Cases –** Some students get admission despite low scores, affecting prediction accuracy.
4) **Balancing Precision Recall –** Ensuring the model doesn't overestimate or underestimate admission probabilities.
5) **Real-Time Processing –** The system must handle mul- tiple concurrent users while maintaining fast response times.
6) **Backend Frontend Integration –** Seamless communi- cation between FastAPI (backend) Streamlit (frontend) is necessary.
7) **Cold Start Problem –** New users lack historical data, making personalized recommendations challenging.
8) **Diverse Academic Backgrounds –** Students from dif- ferent education boards and entrance exams require tailored recommendations.
9) **Protecting Student Information –** The system must securely store and encrypt academic data.
10) **Compliance with Data Privacy Regulations –** Following laws like GDPR UGC guidelines is essential for user trust.
11) **Implementing AI Chatbots –** Developing an NLP-powered chatbot requires additional training and optimization.
12) **Expanding to International Universities –** Different admission processes and eligibility criteria make inte- gration complex.

VI. MATERIALS AND METHODS

A. Data Collection

Data collection is a crucial step in developing any AI-driven system, particularly for college admissions and counseling. Our study utilized admission-related data extracted from official CET portal documents. The data was gathered, processed, and structured to ensure consistency and accuracy.

- The dataset includes information on:
- Admission cutoffs, ranks, and seat availability
- Eligibility criteria from multiple universities
- Academic performance trends based on past student data

We employed Python-based libraries (PyPDF2 and Tabula-py) to extract and preprocess structured data from PDFs. Additionally, we used OpenAI embeddings to convert extracted textual data into vectorized representations for efficient retrieval and recommendations.

B. Data Analysis

After the raw data was extracted, it underwent several preprocessing steps, including:

- Translation and Cleansing: Handling missing values, re- moving special characters, and standardizing formats
- Feature Selection: Identifying the most impactful factors affecting admission, such as cutoff scores, rank distribu- tions, and eligibility criteria
- Normalization: Standardizing data to ensure consistency across different universities and admission categories

The final structured dataset was then used to develop a machine-learning model capable of predicting admission probability for students based on their profiles.

C. System Architecture

1) Recommendation Module: The recommendation module employs content-based filtering to suggest universities based on user preferences and historical data. Two recommendation approaches were implemented:

- Cold Start Recommendations: Displaying the most pop- ular universities based on overall student interest
- Personalized Recommendations: Using embeddings to match students' preferences with similar institutions

2) Prediction Module: The prediction module helps students estimate their likelihood of admission based on their academic scores and preferred universities. The system processes inputs through a trained Huber Regressor, which was found to be the most accurate model, achieving:

- MSE: 0.0017
- RMSE: 0.0422
- R-squared: 0.9306

We evaluated multiple models, including Linear Regression, Random Forest, XGBoost, and CatBoost, to determine the most suitable approach for admission prediction. The Huber Regressor was chosen due to its robustness against outliers and superior accuracy.

D. Implementation

To ensure accessibility and ease of use, the AI-driven system was developed using FastAPI for backend processing and Streamlit for frontend deployment. The application is hosted on Render and supports Docker-based containerization for efficient scaling.

Upon user input (academic details and preferred university), the system:

- Processes the query using vector embeddings
- Retrieves relevant admission data
- Generates an admission probability prediction
- Provides university recommendations

This structured approach enhances the accuracy and usabil- ity of the AI-driven counseling assistant, empowering students with real-time, data-backed decision-making tools.

VII. System Architecture

The system architecture, as shown in Fig. 1, outlines the structured flow of the AI-Driven College Admissions and Counseling Assistant. It processes student queries through Streamlit (Frontend), FastAPI (Backend), OpenAI LLM (AI Processing), and Chroma Vector Store (Database Retrieval) using Retrieval-Augmented Generation (RAG). Cloud deployment on Azure, Render, and Hostinger ensures scalability, speed, and real-time responses, making the system efficient and accessible.

A. Data Input Layer

The Data Input Layer is responsible for receiving queries from users. Students enter their college-related queries through a web-based interface. This input is then validated and preprocessed before being forwarded to the AI processing system. The layer ensures that queries are correctly formatted to improve response accuracy.

B. Excel File Extraction(output.xlsx)

The system extracts and stores college admission data from structured CET portal PDFs into an Excel file. This data includes cutoff marks, seat availability, ranking lists, and eligibility criteria. The Excel file acts as an intermediate storage format before being processed into a structured format for AI-based retrieval.

1) Preprocessing: Text Splitting: In the preprocessing phase, the Text Splitting process is applied to break down large PDF documents and queries into smaller, manageable segments. This step ensures that data is structured correctly and enables efficient vector embedding for retrieval. This preprocessing is crucial for improving the AI model's ability to retrieve relevant information.

C. Frontend: Streamlit

The Streamlit frontend serves as the user interface where students interact with the system. It allows users to enter queries related to college admissions, cutoffs, and rankings and displays AI-generated responses. The frontend is designed to be intuitive, interactive, and user-friendly, ensuring that students can easily navigate and obtain the required information.

D. Backend: FastAPI

The FastAPI backend acts as a communication bridge between the frontend, AI model, and database. It is responsible for handling API requests, processing user queries, and routing them to the AI model. Once the AI model generates a response, the backend ensures that it is formatted properly and sent back to the frontend. FastAPI enhances performance, speed, and scalability, ensuring real-time responses to user queries.

E. Vector Database: Chroma

The Chroma Vector Store is the database layer responsible for storing vector embeddings of processed admission-related data. Instead of storing raw text, the system converts admission records into vector representations, allowing for fast and accurate similarity-based searches. When a student submits a query,

the system searches for contextually relevant information from the Chroma vector database. This approach improves retrieval efficiency and ensures that students get precise answers.

F. AI Model: OpenAI LLM

The AI-driven processing is handled by OpenAI's Large Language Model (LLM), integrated with LangChain. The model follows the Retrieval-Augmented Generation (RAG) approach, which combines pre-trained knowledge with real-time document retrieval from Chroma. The AI model processes queries by:

- Converting them into vector embeddings.
- Searching for relevant admission data in the Chroma database.
- Generating intelligent responses based on retrieved data.

This method ensures context-aware and accurate responses for students.

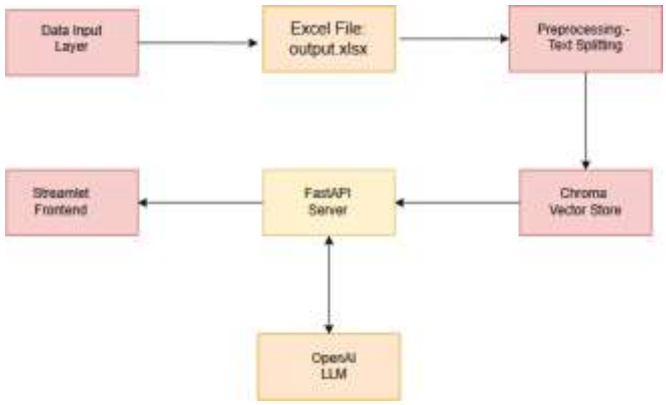

Fig. 1. System Architecture of AI-Driven College Admissions and Coun- selling

VIII. RESULT ANALYSIS

The front-end interface of our project is designed to provide an interactive and seamless experience for students seeking information about college admissions. The system features a chatbot-style interface where users can input queries related to admissions, counseling, and entrance exams. The user interface consists of the following key components:

- **Chat Interface:** The primary interaction area where students can enter their queries. Responses are generated dynamically by the AI model based on the user's input.
- **Automated Responses:** The assistant provides instant, AI-generated replies to user inquiries, ensuring quick and accurate responses.
- **Minimalist Design:** A clean and structured layout enhances user experience, with well-organized text bubbles differentiating user input from AI responses.
- **Navigation and Search Bar:** A designated input field allows users to ask specific questions regarding college admissions, eligibility criteria, and counseling processes.
- **Custom Branding:** The interface includes a header with the project title and institutional branding, along with important navigation elements such as an MHT-CET button for exam-specific queries.

This system ensures accessibility and efficiency in retriev- ing relevant admission-related information, enhancing the decision-making process for students.

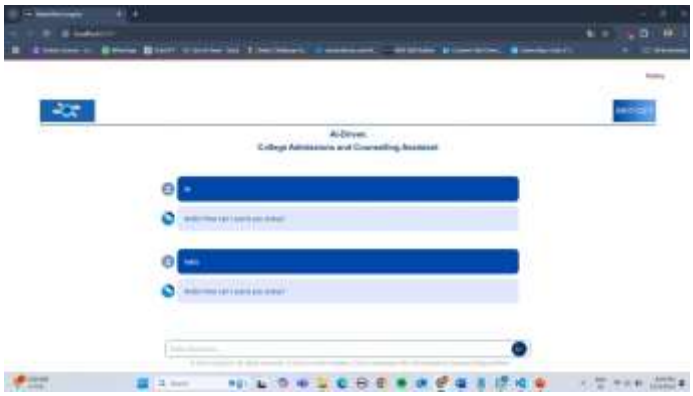

Fig. 2. User Interface of AI-Driven College Admission and Counseling Assistant

REFERENCES

[1] Majjate, H.; Bellarhmouch, Y.; Jeghal, A.; Yahyaouy, A.; Tairi, H.; Zidani, K.A. AI-Powered Academic Guidance and Counseling System Based on Student Profile and Interests. Appl. Syst. Innov. 2024, 7, 6. https://doi.org/10.3390/asi7010006

[2] Rachita Agrawal*1, Shrawani Kondawar*2, Achal Gupta*3, Durgesh Sawale*4, Prof. J.S. Pawar*5 Department Of Information Technology, Singhad College Of Engineering, Vadgaon, Pune, India. DOI: https://www.doi.org/10.56726/IRJMETS63607

[3] Mohammad Amin Kuhail a , Nazik Alturki b,* , Justin Thomas c , Amal
K. Alkhalifa d a College of Technological Innovation, Zayed University, Abu Dhabi, United Arab Emirates b Department of Information Systems, College of Computer and Information Sciences, Princess Nourah Bint Abdulrahman University, Riyadh, Saudi Arabia c King Abdulaziz Center for World Culture, SYNC, Digital Wellbeing Program, Dhahran, Saudi Arabia d Department of Computer Science and Information Technol- ogy, Applied College, Princess Nourah Bint Abdulrahman University,
Riyadh, Saudi Arabi

[4] Hajar Majjate 1,*ORCID,Youssra Bellarhmouch 1,Adil Jeghal 2ORCID,Ali Yahyaouy 1ORCID,Hamid Tairi 1 andKhalid Alaoui Zidani 1 1 LISAC Laboratory, Faculty of Science Dhar El Mahraz, Sidi Mohamed Ben Abdellah University, Fez 30030, Morocco 2 The National School of Applied Sciences of Fez (ENSA), Sidi Mohamed Ben Abdellah University, Fez 30030, Morocco * Author to whom correspondence should be addressed.

[5] Van Busum, K., Fang, S. (2023, March). Analysis of AI Models for Student Admissions: A Case Study. The 38th ACM/ SIGAPP Symposium on Applied Computing, Tallin, Estonia. https://doi.org/10.1145/3555776.3577743

[6] Graham Keir, Willie Hu, Christopher G Filippi, Lisa Ellenbogen, Rona Woldenberg JAMIA Open, Volume 6, Issue 1, April 2023, ooad011, https://doi.org/10.1093/jamiaopen/ooad011 Published: 17 February 2023