

# Develop a Website for Career Guidance in Academics

Harsh Roy<sup>1</sup>, Prashanth Rao<sup>1</sup>, Arpit Rupauliha<sup>1</sup>, Harsh Sharma<sup>1</sup>, Swati Singh<sup>1\*</sup>

1. MIT Art, Design and Technology, Rajbaug, Loni Kalbhor, 412201

## ABSTRACT

Our "Predictive Chatbot System for Academic Performance Assessment" utilizes advanced machine learning and natural language processing techniques to predict students' future academic performance based on their previous marks and areas of academic interest. By analyzing data on academic marks, course preferences, and extracurricular activities, the system provides personalized recommendations for each student. This includes study strategies, reading materials, and extracurricular opportunities tailored to their individual goals and interests. Real-time interaction with the chatbot allows students to seek guidance, track their progress, and make informed decisions. The system continuously refines its predictions and recommendations, ensuring that the advice remains relevant and beneficial. This innovative system empowers students to take control of their academic journey, fostering a proactive approach to education and ultimately enhancing academic performance and satisfaction.

## Keywords

Chatbot System; Predictive Modeling, Academic Performance; Recommendation System.

\* Corresponding Author

Email Id: swati.singh@mituniversity.edu.in

## 1. INTRODUCTION

In the contemporary educational landscape, choosing the right college has never been more critical for students. It is a decision influenced by a multitude of factors, including academic performance and personal interests. This introduction unveils a groundbreaking "College Prediction Chatbot System" designed to assist students in making well-informed choices by predicting suitable colleges based on their 12th-grade marks and individual interests. This innovative chatbot system harnesses the power of data analytics and machine learning to offer students a personalized college selection experience. By analyzing their 12th-grade marks and specific areas of interest, the system can accurately match students with colleges that align with their academic aspirations and personal passions.

## 2. Applicant Ranking System

The software application which uses machine learning algorithms for the analysis and evaluation of resumes based on specific criteria is a resume ranking system. The system shall assign scores or ranking to each resume in order that the recruiter and hiring manager can select the most qualified candidates for a given job. The importance of the candidate's skills and experience to the requirements of the job, the quality and readability of his or her resume may be part of the criteria used in deciding which resumes are taken into account. By automating the resume ranking process, the system helps recruiters save time and improve the efficiency

of the recruitment process. This includes the following subsections in its processes.

### 2.1 Parser System

The tool extracts the relevant information from a given input and transforms it into a structured format. In the context of our project, a parser system dissects the incoming data, separating the 12th-grade marks from the interests, and then processes each element independently. By breaking down this information, we can accurately match students with colleges that align with their academic achievements and personal passions.

This parsing capability simplifies the complexity of the data and ensures a precise and personalized college recommendation process, ultimately assisting students in making informed decisions about their academic journey.

### 2.2 Candidate Skill Set Database

"Candidate Skill Set Database" could be a valuable resource. It could store information about students, including their 12th-grade marks, academic interests, extracurricular activities, and any additional skills or qualifications they possess. This database would enable us to enhance the college prediction process by considering a more comprehensive set of attributes, allowing for a more accurate and personalized matching of students with colleges based on their holistic skills and academic achievements.

In our project, the "Candidate Skill Set Database" emerges as a vital asset. This repository is designed to store a wide array of student information, encompassing 12th-grade marks, academic interests, extracurricular involvements, and a host of additional skills and qualifications. By housing this comprehensive dataset, our project is poised to revolutionize the college prediction process. With a holistic understanding of each student's unique skillset and academic achievements, we can deliver more accurate and personalized college recommendations. This resource becomes a cornerstone in empowering students to embark on academic journeys that align seamlessly with their individual profiles, ensuring a fulfilling and successful higher education experience.

### 2.3 Pre-Processing Module

In our college selection project, the "Pre-Processing Module" plays a crucial role in refining user data. It optimizes academic marks and interests, cleanses the data, and extracts essential information, all of which are pivotal for the accurate matching of students with suitable colleges. By converting raw inputs into standardized text and identifying key factors, this module simplifies data, ensuring its accuracy and relevance. This streamlined process enables precise and personalized college recommendations, enhancing the college selection experience for students.

Table 1. Tasks in a Pre-Processing System

Pre-Processing Task	Description
Data Cleaning	Removal of irrelevant and redundant data, such as headers, footers, and graphics.
Text Extraction	Extraction of relevant text such as contact information, academic background, area of interest and more.
Tokenization	Breaking down the text into individual tokens or words, which can be analyzed and processed further.
Stop words Removal	Elimination of commonly used words that do not add meaning to the resume, such as "a", "an", "the", and "in".
Lemmatization	Reduction of words to their root form to simplify analysis and comparison, such as converting "running", "runs", and "ran" to "run".
Normalization	Conversion of text to a standardized format to ensure consistency and accuracy in analysis and processing.
Entity Recognition	Identification of named entities, such as colleges, courses, and areas of interest, to provide context for accurate recommendations

## 2.4 Database Management Module

In the context of our project, the "Database Management Module" plays a central role in constructing and maintaining a database of student profiles. This database is pivotal to our college recommendation system, enabling efficient storage and retrieval of user inputs, such as marks and interests. The module's functions encompass importing, structuring, and optimizing data for swift searches, ensuring precise college recommendations. It also performs validation and verification checks to maintain data accuracy, while data cleaning and standardization enhance consistency and analysis. Furthermore, the module is committed to data privacy and security, featuring access control, encryption, and data backup functions to safeguard sensitive user information. In this way, the Database Management Module empowers our project to provide personalized and secure college recommendations, delivering a seamless and reliable user experience

## 2.5 College Recommendation Algorithms for Personalized Higher Education Choices

In this section, we delve into the algorithms and methodologies utilized to enhance the selection of colleges for students based on their academic performance and personal interests. The algorithms considered for this project aim to provide personalized college recommendations, making the college selection process more streamlined and informed

### 2.5.1 Marks-Interest Matching Algorithm:

The primary algorithm in our project involves matching a student's 12th-grade marks and interests with college profiles. This approach assigns scores to colleges based on how well they align with the student's academic achievements and areas of interest.

### 2.5.2 Collaborative Filtering for College Selection:

A collaborative filtering algorithm is explored, which takes into account the preferences and selections of similar students. By leveraging user behavior and choices, it helps students discover colleges that match their profile based on the decisions of others with similar backgrounds and interests.

### 2.5.3 Collaborative Filtering for College Selection:

A collaborative filtering algorithm is explored, which takes into account the preferences and selections of similar students. By leveraging user behavior and choices, it helps students discover colleges that match their profile based on the decisions of others with similar backgrounds and interests.

### 2.5.4 Content-Based Filtering for Personalized College Recommendations:

Another technique involves content-based filtering, where colleges are recommended based on a comprehensive analysis of their attributes, such as courses offered, extracurricular activities, and reputation. This algorithm tailors recommendations to a student's specific interests and preferences..

### 2.5.5 Machine Learning for College Prediction:

Machine learning models are employed to analyze historical data and discover patterns that influence college choices. This algorithm assigns scores to colleges based on their fit with a student's marks and interests, providing data-driven recommendations.

### 2.5.6 User-Driven College Selection:

In this approach, the algorithm focuses on empowering the user to make informed decisions by presenting a range of college options based on their inputs. The user's feedback and interactions with the system play a significant role in fine-tuning recommendations.

## 2.6 User Interface

The User Interface (UI) is a critical component of our project, enhancing the college selection experience for users. It serves as the gateway for students to interact with the system, enabling them to input their academic data, explore college recommendations, and access parsed information. The UI offers user-friendly navigation, filtering options, and a dashboard to track college choices efficiently. Furthermore, it ensures that parsed data, including marks, interests, and key attributes, is presented in an organized and comprehensible manner, aiding students in making informed decisions. Notably, our system provides valuable insights, such as the resume score and skill assessment, categorized into different proficiency levels (intermediate, beginner, and experienced). These insights empower users for self-analysis and informed college selection, elevating their overall experience within the system candidate

## 3. Methodology

The methodology for a resume parsing system is a complex and iterative process that involves careful data collection, pre-processing, feature extraction, algorithm development, and evaluation. By following a systematic and rigorous approach, developers can ensure that the resulting system is accurate, effective, and able to identify the most qualified candidates for a specific job opening.

### 3.1 Research design and approach

We adopt a quantitative research design with a comparative approach to assess the performance of various student profiles. This methodology allows us to pinpoint the unique needs and challenges students face when selecting the right colleges based on their academic achievements and interests. Our research journey involves conducting an extensive literature review to understand the landscape of higher education, defining system requirements, developing and rigorously testing prototypes, gathering and analyzing user data, iteratively enhancing the system based on user feedback and data insights, and ultimately validating the system's real-world effectiveness.

### 3.2 Data Collection and analysis methods

Within our college selection project, the data collection and analysis methods are pivotal to ensuring the system's accuracy and effectiveness. Here's an outline of the steps within this context:

1. **User Inputs:** Data collection begins with users' academic marks and areas of interest, which serve as the foundational information for the system.
2. **Data Pre-Processing:** The collected data undergoes pre-processing to cleanse and standardize it, removing inconsistencies and irrelevant details to ensure its accuracy.
3. **Algorithmic Analysis:** The system utilizes advanced algorithms, including natural language processing and machine learning, to analyze the pre-processed data. These algorithms identify patterns, relationships, and user preferences, contributing to personalized college recommendations.
4. **User Feedback:** Data collection extends to user feedback and interactions with the system. This feedback helps in refining the system and ensuring that it aligns with users' expectations.
5. **Data Validation:** The data analysis process involves validating the accuracy of recommendations by comparing them with real-world outcomes and user feedback.
6. **Iterative Enhancement:** Based on the results of the analysis, the

system is refined and improved to provide even more accurate and tailored college recommendations.

7. **Data Privacy:** Throughout these steps, data privacy and security measures are upheld to protect the sensitive user information.

8. **Historical Data Integration:** In order to enhance the accuracy of recommendations, historical data on college admissions, student outcomes, and past choices are integrated into the analysis. This allows the system to learn from previous patterns and better predict the most suitable colleges for current users.

9. **A/B Testing:** The system may implement A/B testing methodologies to compare different algorithms or recommendation strategies, collecting data on how users respond to each variant. This data helps in fine-tuning the recommendation engine.

10. **Machine Learning Training:** Machine learning models are continually trained with updated data to adapt to changing user preferences and evolving academic institutions. Regular model retraining ensures that the system remains up-to-date and relevant for users.

## 3.3 Sample size and selection criteria

### 3.3.1 Sample size

The sample size in our project depends on the scope and diversity of colleges and academic profiles we aim to analyze. While a larger sample size is desirable for accuracy, it's essential to balance it with the resource constraints of data collection and analysis. In our context, a minimum sample size of 100-200 student profiles, reflective of various academic backgrounds and areas of interest, is recommended. This size allows the system to train on a broad spectrum of college and student profiles, ensuring the recommendations are pertinent to a wide range of students.

### 3.3.2 Selection Criteria

The selection criteria for student profiles should align with the unique dynamics of the higher education landscape. Some of the key criteria for our project include:

**Diversity:** The student profiles should represent a diverse range of academic disciplines, colleges, and student preferences. This diversity ensures that the system can make relevant recommendations for students pursuing a wide array of academic paths.

**Quality:** The student profiles should consist of high-quality data with clear and well-structured information. Well-constructed academic profiles enable the system to accurately extract and analyze pertinent data for recommendations.

**Relevance:** The student profiles in our sample should be closely aligned with the specific academic context we're analyzing. For instance, if our system is designed for college selection in computer science programs, the sample should primarily consist of student profiles pursuing computer science and related fields.

## 3.4 Variables and measurements

In our college selection project, variables and measurements serve as the bedrock for creating an accurate and effective recommendation system. The variables represent the critical factors used to evaluate and rank students' academic profiles for college recommendations. These encompass academic marks, interests, extracurricular activities, and additional qualifications, all meticulously extracted during the pre-processing stage. These variables are the building blocks upon which our system constructs algorithms to ensure that it accurately pairs students with colleges that match their academic achievements and personal passions and we also have the measurements where

Complementing this, measurements serve as the criteria by which we

evaluate the system's performance. This includes assessing the accuracy and relevance of the recommendations to the user's academic interests, the efficiency of data processing, the precision of data extraction, and, most crucially, user satisfaction. As we constantly refine our system, these measurements enable us to enhance the accuracy and effectiveness of our college recommendations, ultimately empowering students in their higher education journey.

## 4. Discussions

The research findings in our college selection project suggest that our system offers an efficient and accurate way for students to evaluate their academic profiles and identify the most suitable higher education institutions. By suggesting potential colleges and providing guidance on how to enhance their chances of admission, our system has the potential to significantly improve the educational process by empowering students to make informed decisions about their academic future..

### 4.1 Key Findings

In the context of our college selection project, the key findings may revolve around the effectiveness and precision of the system in matching students with appropriate higher education institutions. Some of the potential key findings could include

#### 4.1.1 Relevance of Academic Interests:

One key finding may be that the system's ability to accurately identify relevant keywords and match them to a job description is crucial to its overall effectiveness. This may be particularly true for technical roles in software development, where specific programming languages, tools, and technologies may be required.

#### 4.1.2 Impact of User Feedback:

Another key finding may underscore the significance of user feedback in refining the system. Regular feedback loops with students can lead to iterative improvements in the recommendation process, resulting in more personalized and effective guidance.

#### 4.1.3 Data Quality and Diversity:

The quality and diversity of the data used to train and test the system could emerge as a crucial finding. Ensuring that the system can handle a wide range of academic profiles, areas of interest, and college attributes is pivotal for providing accurate and comprehensive recommendations..

**Machine Learning Adaptability:** Another key finding may be the impact of machine learning algorithms on the overall effectiveness of the system. For example, using an algorithm that can learn from past parsing successes and failures, or that can adapt to new job postings and resume formats, may result in a more accurate and effective system overall.

### 4.2 Implications of Recruitment Practices

Our college selection practices hold significant implications for the educational landscape and the students who are navigating their academic journeys. By offering a system that tailors college recommendations based on individual academic profiles and interests, we empower students to make more informed decisions about their higher education. This not only leads to enhanced decision-making but also contributes to improved student outcomes. When students are matched with colleges that align with their aspirations, they are more likely to excel in their chosen academic paths

### 4.3 Limitations and future scope

#### 4.3.1 Limitations:

1. **Data Constraints:** One of the limitations is the quality and diversity of the available data. While we strive to provide accurate recommendations, the system's effectiveness may be constrained by the data it has been trained on. Expanding and improving the dataset could address this limitation

2. **Resource Dependency:** The system's performance is influenced by the availability of resources, both in terms of data collection and computational capabilities. These resource constraints may limit the scalability and accessibility of the system.

3. **Dynamic Educational Landscape:** The higher education landscape is continually evolving with new programs, colleges, and areas of study. The system may not capture these changes in real-time, and there is a limitation in adapting to such rapid developments.

#### 4.3.2 Future Scope:

1. **Enhanced Data Collection:** Expanding the dataset to include a more comprehensive and diverse set of academic profiles, colleges, and student feedback will improve the system's accuracy and relevance.

2. **Enhanced machine learning models:** With advancements in machine learning, future research can explore using more advanced algorithms like deep learning and neural networks to improve the accuracy of parsing systems.

3. **Contextual analysis:** Another direction is to incorporate contextual analysis by using natural language processing and sentiment analysis to extract additional information from resumes and job descriptions.

4. **Mobile Accessibility:** Developing a mobile application or platform for the system could make it more accessible to a broader range of students, further democratizing college selection

## 5. Conclusion:

In conclusion, our college selection system represents a significant advancement in the way students navigate their higher education journeys. By leveraging advanced algorithms, data-driven insights, and a user-centric design approach, we have created a platform that empowers students to make informed decisions about their academic futures.

The system's ability to provide personalized college recommendations based on academic achievements and interests holds the promise of transforming the college selection process. It not only streamlines the decision-making process but also enhances the likelihood of students thriving in environments that align with their passions and abilities.

## 6. REFERENCES

1. Smith, J. (2020). "The Future of Education: Trends and Innovations." Academic Press.
2. Johnson, M. A. (2019). "Machine Learning for Beginners." O'Reilly Media.
3. Brown, A., & Lee, C. (2018). "Impact of Artificial Intelligence on College Admissions." *Journal of Educational Technology*, 25(2), 143-158.
4. Patel, S., & Williams, L. (2020). "Enhancing Student Decision-Making in College Selection." *International Journal of Higher Education*, 7(3), 94-110.