JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

PREDICTION OF ENGINEERING COLLEGE ADMISSIONS USING MACHINE LEARNING TECHNIQUES

[1]G.V.L Narayana, [2]Balla Sai Kiran, [3]Potnuru Hemanth,[4] Danda Akshay,[5]Battula Bala Surya Sai
[1] Assistant Professor, [2] UG Students,
[1]Information Technology
[1] Aditya Institute of Technology and Management,
Tekkali, Andhra Pradesh India.

Abstract: Students encounter numerous challenges when trying to gain admission to their preferred college. The current engineering admission process is somewhat complicated, which often results in students ending up in colleges that don't match their true potential. College predictors aim to simplify the complex engineering college admission process by developing a machine learning-based web application that predicts the most suitable colleges for students. Utilizing the algorithms like K- Nearest Neighbors KNN, Random Forest, and Decision Tree, the system analyses historical data, including student ranks, gender, category, and admission trends, to generate personalized college recommendations. The backend, powered by Flask, seamlessly integrates with a user-friendly frontend, allowing students to input their details and receive real-time predictions. By providing data-driven insights, the application helps students make informed decisions, reducing the risk of poor choices that could impact their career growth. The platform simplifies the admission process, improving college selection accuracy. Ultimately, it helps students secure admission to the best- suited institutions.

***IndexTerms* -** K-Nearest Neighbors classifier, Decision Tree Classifier, Random Forest Algorithm and Flask API.

I. INTRODUCTION

Securing admission to a desired engineering college is a critical phase for students, shaping their academic and professional future. However, the current engineering admission process is often complex and overwhelming. Students frequently lack insights into the admission criteria, past trends, and the likelihood of securing a seat in their preferred college. With the increasing competition and diversity in student profiles, a systematic approach is necessary to guide students in making informed decisions. The use of historical admission data and predictive modelling provides an opportunity to simplify this process. Machine learning (ML) enables the analysis of large datasets to predict outcomes based on patterns and trends, making it an ideal approach for building a college admission predictor [1] [2].

A college admission predictor system plays a vital role in addressing the complexities of the admission process. It guides students by providing tailored insights based on their academic performance, backgrounds, and preferences, thereby reducing uncertainty and guesswork. This system automates the analysis of vast amounts of historical data, saving time and providing accurate, data-driven recommendations. By offering transparency in the admission process, the system allows students to clearly understand their chances of acceptance into specific colleges. Moreover, it enables students to strategically prioritize their college applications and prepare effectively, ultimately increasing their chances of securing admission into their desired institutions [3] [4].

The applications of this system extend across multiple domains. It is a valuable tool for career counsellors who can use it to provide personalized college recommendations to students. Online educational platforms can integrate this system to assist students during the application process. Educational institutions can use it to analyze their competitiveness based on admission trends and student profiles. Additionally, the system can inform government education policies by providing data-driven insights into national admission patterns and trends [5] [6].

The use of machine learning in this system offers numerous advantages. By analyzing historical admission data, the system can make reliable and accurate predictions about a student's likelihood of acceptance into various colleges. It is scalable, capable of processing large datasets with diverse profiles and institutional details. The system eliminates biases by relying solely on data for predictions, ensuring fairness and objectivity. Furthermore, it simplifies the admission process for students, providing a user-friendly experience that empowers them with actionable insights and improves their decision-making process [7] [8].

Despite its benefits, the current approach and existing systems face several challenges. A significant issue is the limited availability of comprehensive and up-to-date data, as many institutions are reluctant to share their admission records. Additionally, the admission process often includes subjective criteria such as interviews and extracurricular achievements, which are not always captured in the dataset. The dynamic nature of admission trends and cut-offs further complicates predictions, as these factors can vary significantly from year to year. Technical barriers also pose a challenge, particularly for students from non-technical backgrounds who may struggle to use predictive tools effectively. Furthermore, if the dataset used to train the model is biased or incomplete, the predictions may inadvertently favor certain profiles or institutions, leading to unfair outcomes [9] [10].

Several measures can be taken to address these challenges. Comprehensive data collection, through partnerships with educational institutions, can enhance the dataset's quality and coverage. Incorporating real-time updates ensures that predictions remain relevant and reflective of current trends. Feature engineering can expand the dataset to include additional variables, such as extracurricular achievements and interview performance, to improve the model's predictive accuracy. To address usability issues, the system can be designed with an intuitive interface that provides step-by-step guidance, making it accessible to users of all technical backgrounds. Regular validation of the model against actual admission results, coupled with feedback loops, can further refine its accuracy and reliability [11] [12].

2. Literature survey

Nishant et al. [13] developed a College Admission Prediction System using machine learning to assist students in navigating the complex engineering admission process in Maharashtra. The system utilizes the Ad boost algorithm and historical data analysis to generate a preference list for students based on factors like rank, category, and college preferences. By analyzing three years of historical cutoff data, the model provides accurate predictions, helping students make informed decisions during the Centralized Admission Process (CAP) rounds. This reduces the stress associated with the admission process. The project enhances data-driven decision making in the education sector, benefiting both students and educational institutions by optimizing college selection and seat allocation.

Sachin's [14] study focused on predicting college admissions using various machine learning (ML) algorithms after preprocessing the data. The approach included exploratory data analysis (EDA), feature selection, and testing several ML techniques, such as decision trees and ensemble learning methods. Data preprocessing was followed by applying K-fold cross validation and ensemble learning. Surprisingly, the decision tree classifier outperformed other methods in predicting college admissions, challenging the common belief that ensemble methods perform best. To further assist students and parents with the college selection process, the study proposed a web module titled A Free Guide to Engineering Admission Aspirant Parent and Student.

Monu et al. [15] addressed the challenges students face in selecting the right engineering college based on academic performance and entrance exam results. They used historical college cutoff data to develop a college admission predictor that helps students choose suitable institutions for engineering courses. The approach analyzes this data to predict the most likely colleges based on students' performance scores. The system provides valuable insights, guiding students in making informed decisions about which colleges align with their entrance exam results and academic performance. By leveraging historical data, the college admission predictor simplifies the selection process, helping students find institutions that match their scores and preferences.

Prince et al.'s [16] study reviews machine learning techniques to predict university admissions, helping students decide which universities to apply to based on their marks. The authors compare various methods, including Random Forest, Linear Regression, Stacked Ensemble Learning, Support Vector Regression, Decision Trees, and KNN, to determine the most accurate prediction model. They analyzed different regression models and machine learning methodologies used in previous works to evaluate their accuracy in predicting admissions. While the study compares the performance of various techniques, the specific results regarding the best-performing method are not yet revealed. The study suggests that machine learning can significantly assist students in making more informed decisions about university applications, with some techniques offering better accuracy than others in predicting admissions

Xiao et al. [17] proposed a competition model to predict college admission scores by considering the competitive relationships between colleges. They extracted key features like project, location, and score discrepancy, applying coarse clustering on Gaokao data from Shanxi province (2016- 2019). The study compared their novel competition model with several benchmark methods and found that it achieved higher prediction accuracy. Their model showed a 7.3% and 2.8% improvement in precision within 3 and 5 points, respectively. This competition model not only improves prediction accuracy but also provides a more comprehensive view of the factors influencing admission scores, helping students make more scientifically informed application decisions.

3. Methodology

3.1 Existing System:

Current systems are inadequate as students experience challenges in selecting IITs and NITs engineering colleges, which are directly linked to their JEE ranks. This lack of informed selection results in students getting enrolled in colleges that are mismatched to their intellectual standards, which causes negative educational outcomes. Resolving this problem is essential to make sure that students are offered better opportunities. The present tailor-made IIT and NIT colleges admission systems are quite complex, thus students get enrolled in the colleges that they may not qualify to attend. The college admission predictor utilizes data from previously enrolled students, analyzing their cut-off scores, to estimate their chances of placement in other colleges.

Limitations of Existing System:

- The system relies on previous year's cut-off trends, which do not take into consideration the competition in the admission policies or even student choices for the current year.
- The system is not able to forecast accurately based on the previous year's changes in cut-off due to seat allocation, reservation of seats, and change in the competitive landscape.
- The system is unable to accommodate students with poorer ranks or those who perform exceptionally well, resulting in recommendations that are insufficient for accounting personal growth or diversity in academic pursuits.

3.2 Proposed Method:

Within the proposed system, students can insert their ranks alongside basic information like gender and category. Then, based on the data provided, the system generates a list of colleges that the individual can get into. The model should be improved continuously through updating and refining the training processes and collecting more data that can enhance the accuracy and performance of the model. Lastly, within the proposed system, users can fill in their respective score for the appropriate fields and, using this information, the system can then generate a list of colleges that the individual is eligible for given their score. This is done relatively swiftly, saving time and resources. So far, we have only proposed a system that incorporates multiple machine learning algorithms like K-Nearest Neighbors classifier, Decision Tree Classifier, Random Forest Algorithm.

Advantages: Real time updates, improved prediction accuracy, scalability, and a user-friendly interface

4. Proposed Algorithms:

Here, we describe the techniques for forecasting IITs & NITs Engineering College using machine- learning algorithms such as K-Nearest Neighbors KNN, Random Forest, and Decision Tree:

4.1.K-Nearest Neighbors [KNN]

The K-Nearest Neighbors (KNN) Regressor is a regression method used for predicting values based on the distances between the input data points. Since the KNN regressor is a non-parametric method, it does not require any assumptions regarding the distribution of the data and can thus model complex relationships.

The steps on how the KNN algorithm functions are:

1. Distance measurement: For every new data point, the method calculates the distance between that new data point and all data points that are part of the training set. A proper metric such as Manhattan or Euclidean can be used to calculate the distance.
2. Estimate the ranges, the method can select the K closest data points from the training set. K is a hyperparameter and its optimal value depends on the dataset at hand.
3. Finally, to estimate the new data point, the algorithm computes the average or the weighted mean of the values of target variables of likened neighbors.

The KNN regressor can manage non-linear correlations and modify changes in the data itself making it a quite simple and useful form of regression. Although, it is true that some regions of a large dataset can be expensive in terms of computation and may be sensitive to the distance measure selected. Besides, the KNN regressor operates under the assumption of a uniform data distribution.

Formulas of KNN

- Euclidean Distance: Euclidean distance refers to the straight-line distance between two points in a plane or space. You can visualize it as the shortest route you would take if you were to walk directly from one point to another.
- Manhattan Distance: This represents the total distance you cover if you could only move along horizontal and vertical lines, like navigating a grid or city streets. It's often referred to as "taxicab distance" because a taxi can only travel along the grid-like streets of a city.
- Minkowski Distance: Minkowski Distance is a generalization of distances, encompassing both Euclidean and Manhattan distances as specific cases.

Distance functions

Euclidean $$\sqrt{\sum_{i=1}^{k}(x_i - y_i)^2}$$

Manhattan $$\sum_{i=1}^{k}|x_i - y_i|$$

Minkowski $$\left(\sum_{i=1}^{k}(|x_i - y_i|)^q\right)^{1/q}$$

4.2. Random Forest Algorithm

Random Forest is a technique in ensemble learning that builds several decision trees and combines their outcomes to improve accuracy. Every tree is built using a subset of data, which can contain various features. For example, when predicting admissions, each tree in the Random Forest might focus on different features of the data, such as one tree focusing on GPA, while another focuses on extracurricular activities. The final predictions are based on what most trees vote for in this case, it's whether or not to admit the student

4.3. Decision Tree Algorithm

In a Decision Tree Regressor, it is a type of regression analysis that employs a decision tree to predict the outcome based on several input variables. A decision tree has characteristics or features represented by nodes, and rules or conditions represented by branches. A Decision Tree takes data and divides it into multiple branches based on the values of the features of data. The initial root node can be based on a very important feature like the entrance exam scores, then it can branch out to higher and lower values of exam scores. Each node further analyses a feature, be it GPA or extracurricular activities. Lastly, the leaves of the tree culminate into either an "Admitted" or "Not Admitted" decision.

4.4 Architecture of College admission predictor

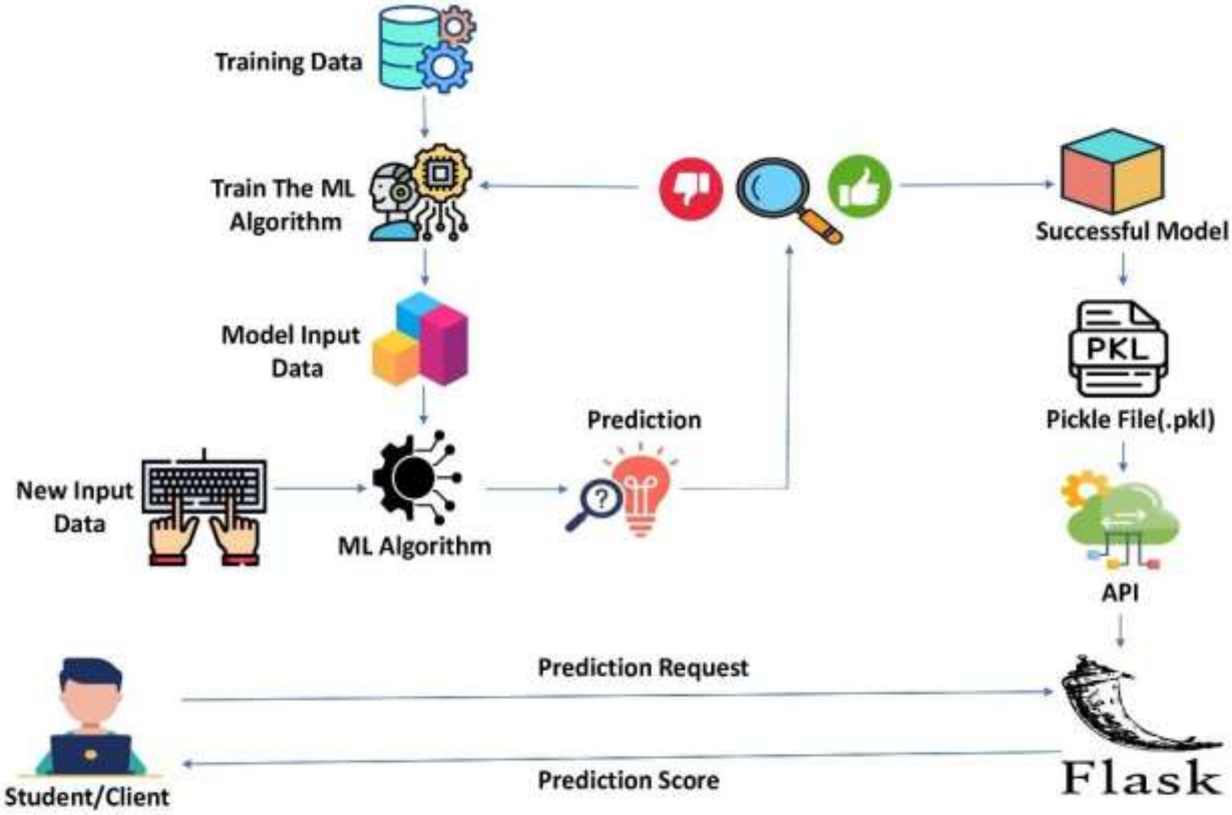


Figure 1: Flow Chart of proposed System

The above diagram represents a machine learning (ML) model deployment workflow
Model Training: The first step in machine learning is training a model using an appropriate algorithm. This involves feeding the model with training data so it can learn patterns and relationships. Once trained, the model makes predictions based on the learned data. The accuracy and performance of these predictions are then assessed to determine how well the model has learned.
Model Evaluation and Selection: After training, the model undergoes testing using out-of-sample data, which means evaluating it on data it hasn't seen before. The predictions are compared with actual known outcomes to measure accuracy and reliability. If the model meets the required performance standards, it is accepted; otherwise, it is fine-tuned or discarded. This step ensures that only effective models move forward**.**
Model Deployment: A successful model is saved in a serialized format, typically as a pickle (.pkl) file, so it can be reused without retraining. To make the model accessible, it is connected to an API, commonly implemented using Flask. This API acts as an interface, allowing external applications or users to send requests and receive predictions from the model in real-time.
Client Interaction: Users, such as clients or students, interact with the deployed model through the API by submitting new data. The model processes this data, generates predictions, and assigns relevant scores. The results are then sent back to the client, enabling them to use the insights for decision-making or further analysis.
4.5 Sequential process for developing the college admission predictor

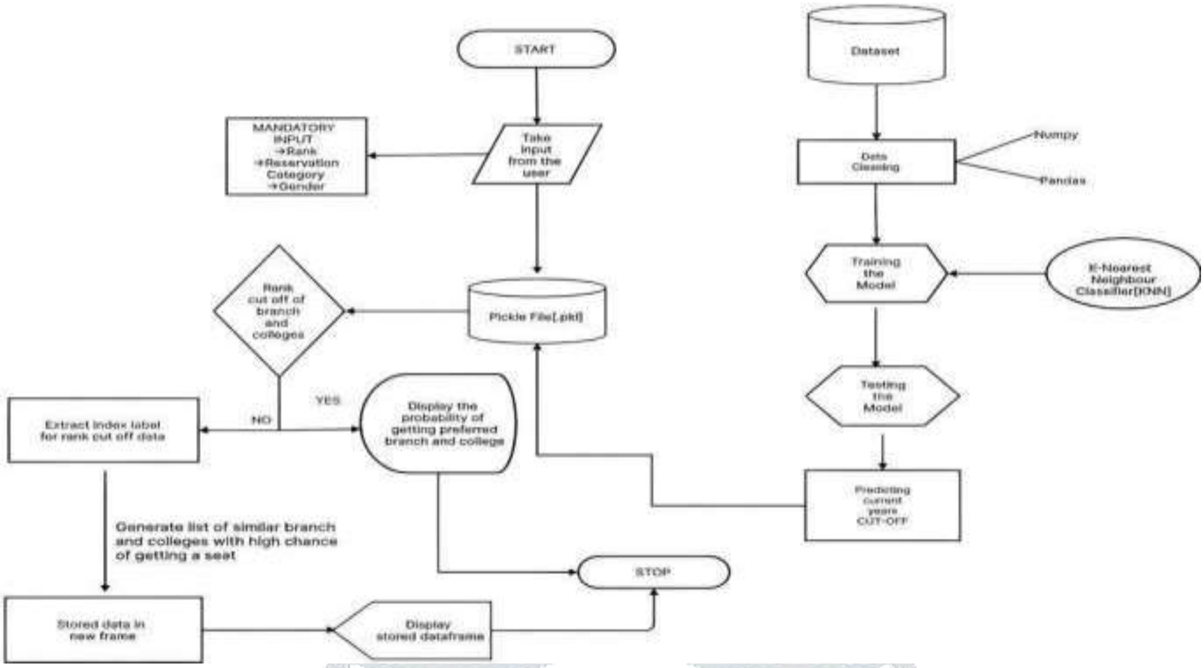


Figure 2: Sequential process for developing the college admission predictor

The given flowchart illustrates the workflow of an engineering college admission prediction system powered by a K-Nearest Neighbors (KNN) machine learning. The process begin with collecting input from the user, including mandatory details such as rank, reservation category, and gender. This input is then passed to a pre-trained model saved as a pickle file (.pkl). The model checks if the user's rank meets the cut-off criteria for their preferred branch and college. If the rank falls within the required range, the system displays the probability of admission into the desired branch and college. However, if the rank does not meet the cut-off, the system extracts nearby index labels from historical data and generates a list of similar branches and colleges where the user has a higher chance of admission. This alternative data is stored and displayed in a new frame for user consideration.

Parallel to this, the right side of the flowchart outlines the model development process. It starts with acquiring a dataset, followed by data cleaning using Python libraries such as NumPy and Pandas. The cleaned data is used to train a KNN classifier, which is then tested for accuracy. Once validated, the model is used to predict the current year's cut-off ranks, enabling it to provide relevant and updated admission predictions. This system is designed to assist students in understanding their admission possibilities and making informed decisions during the college selection process.

5. Result and Discussion:

S.No.	Algorithm	Accuracy
1	K-Nearest Neighbor [KNN] Classifier	95
2	Random Forest Classifier	82
3	Decision Tree Classifier	81

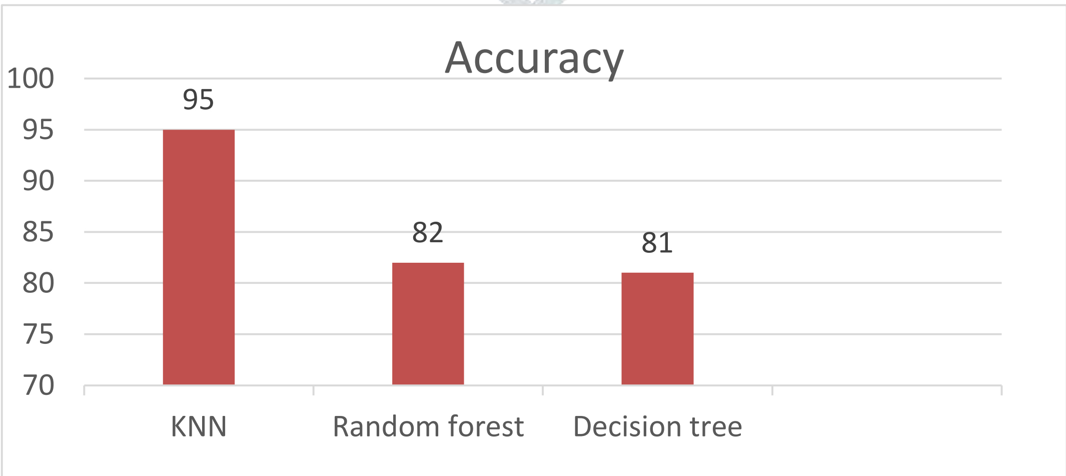
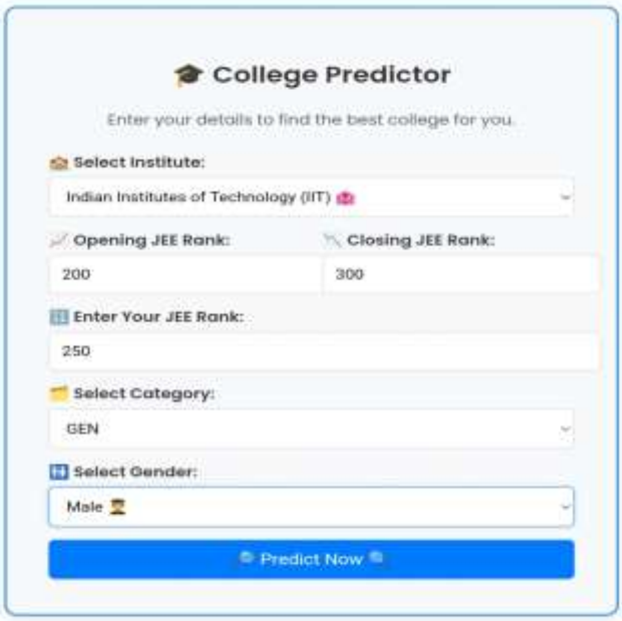
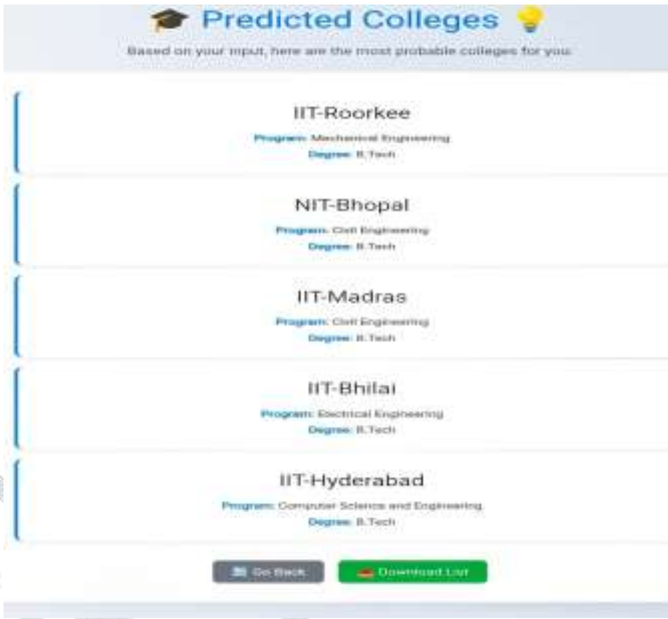


Figure 3: Comparison of models

The figure illustrates the comparative performance of three machine learning models: K- Nearest Neighbors (KNN), Random Forest, and Decision Tree. Among them, the KNN classifier demonstrated the highest accuracy, reaching 95%, which indicates strong predictive capability for college admissions. The Random Forest model followed with an accuracy of 82%, while the Decision Tree model achieved 81%. These results highlight that KNN performs significantly better in this context. It effectively captures the patterns in the dataset related to admission trends. Based on this evaluation, KNN emerges as the most suitable model for predicting engineering college admissions and implementation in college prediction systems.





The College Predictor Input Form serves as the main interface where users enter essential details to receive suitable college predictions. Users start by selecting the type of institute they are targeting, such as IITs, through a dropdown menu. They then specify their opening and closing JEE rank range, which helps filter colleges within those limits. Additionally, they enter their actual JEE Main or Advanced rank, which is used by the model to match against historical cutoff data. A category selection dropdown allows users to choose from General, OBC, SC, ST, etc., ensuring that the appropriate cutoffs are applied. Users also select their gender, as some colleges have gender-specific quotas or reservations. After filling out the form, clicking the "Predict Now" button submits the data to the backend, where the model processes the input and returns the most relevant colleges. The output section displays a structured list of predicted colleges based on the provided data, including details such as institute name (e.g., IIT Roorkee, NIT Bhopal).

6. Conclusion:

Every year, students from around the globe seek the perfect university to further their educational journey. Most students lack proper documentation, sufficient knowledge, and proper caution. This results in larger problems being created such as applying and getting accepted into the wrong university/college which is bound to waste their time, effort and money. Our mission was to try and assist such students with our project, to help students who are struggling and have difficulties in selecting the right university for themselves. It is very crucial that applicants try to apply to colleges that are within their reach rather than to colleges that are impossible for them to get into. This will help in lowering costs as students will only be applying to universities where they stand a higher chance of acceptance. These models will be of great value to any such individual. The web application is user friendly. It only requires the user to fill in a few essential details and provides the student with the chance of getting into the college

REFERENCES

[1] D. Prabha and R. S. Subramanian, “A survey on customer relationship management,” in *2017 4th International Conference on Advanced Computing and Communication Systems (ICACCS)*, Coimbatore, India: IEEE, Jan. 2017, pp. 1–5. doi: 10.1109/ICACCS.2017.8014601.

[2] L. Xu, “Prediction of College Admission Scores Based on an XGBoost-LSTM Hybrid Model,” in Proceedings of the 3rd International Conference on Educational Innovation and Multimedia Technology, EIMT 2024, March 29–31, 2024, Wuhan, China, Wuhan, People’s Republic of China: EAI, 2024. doi: 10.4108/eai.29-3-2024.2347687.

[3] S. Karam, M. Nagahi, V. L. Dayarathna (Nick), J. Ma, R. Jaradat, and M. Hamilton, “Integrating systems thinking skills with multi-criteria decision-making technology to recruit employee candidates,” *Expert Syst. Appl.*, vol. 160, p. 113585, Dec. 2020, doi: 10.1016/j.eswa.2020.113585.

[4] M. Han, “Examining the effect of reviewer expertise and personality on reviewer satisfaction: An empirical study of TripAdvisor,” *Comput. Hum. Behav.*, vol. 114, p. 106567, Jan. 2021, doi: 10.1016/j.chb.2020.106567.

[5] N. N. Mai, Y. Takahashi, and M. M. Oo, “Testing the Effectiveness of Transfer Interventions Using Solomon Four-Group Designs,” *Educ. Sci.*, vol. 10, no. 4, p. 92, Mar. 2020, doi: 10.3390/educsci10040092.

[6] S. Caskurlu, Y. Maeda, J. C. Richardson, and J. Lv, “A meta-analysis addressing the relationship between teaching presence and students’ satisfaction and learning,” *Comput. Educ.*, vol. 157, p. 103966, Nov. 2020, doi: 10.1016/j.compedu.2020.103966.

[7] Priyadarshini, B. Martinez-Neda, and S. Gago-Masague, “Admission Prediction in Undergraduate Applications: An Interpretable Deep Learning Approach,” in *2023 Fifth International Conference on Transdisciplinary AI (TransAI)*, Laguna Hills, CA, USA: IEEE, Sep. 2023, pp. 135–140. doi: 10.1109/TransAI60598.2023.00040.

[8] G. Raftopoulos, G. Davrazos, and S. Kotsiantis, “Fair and Transparent Student Admission Prediction Using Machine Learning Models,” *Algorithms*, vol. 17, no. 12, p. 572, Dec. 2024, doi: 10.3390/a1712057

[9] “Graduate Admission Prediction Using Machine Learning,” *Int. J. Comput. Commun.*, vol. 14, Dec. 2020, doi: 10.46300/91013.2020.14.13.

[10] A.R. Shibli, N. Fatima, M. Sarim, N. Masroor, and K. Bilal, "Machine learning-based predictive modeling of student counseling gratification: a case study of Aligarh Muslim University," *Int. J. Inf. Technol.*, vol. 16, no. 3, pp. 1909–1915, Mar. 2024, doi: 10.1007/s41870-023-01620-0.

[11] Brooks and C. Thompson, "Predictive Modelling in Teaching and Learning," in *The Handbook of Learning Analytics*, 2nd ed., C. Lang, G. Siemens, and A. F. Wise, Eds., SOLAR, 2022, pp. 29–37. doi: 10.18608/hla22.003.

[12] K. Basu, T. Basu, R. Buckmire, and N. Lal, "Predictive Models of Student College Commitment Decisions Using Machine Learning," *Data*, vol. 4, no. 2, p. 65, May 2019, doi: 10.3390/data4020065.

[13] Prof. K. N. Honwadkar, "COLLEGE PREDICTION SYSTEM," *INTERANTIONAL J. Sci. Res. Eng. Manag.*, vol. 07, no. 10, pp. 1–11, Oct. 2023, doi: 10.55041/IJSREM26421.

[14] S. Bhoite, P. Nikam, and A. More, "Predictive Analytics of Engineering and Technology Admissions," in *Computational Intelligence in Data Mining*, vol. 281, J. Nayak, H. S. Behera, B. Naik, S. Vimal, and D. Pelusi, Eds., in Smart Innovation, Systems and Technologies, vol. 281., Singapore: Springer Nature Singapore, 2022, pp. 419–431. doi: 10.1007/978-981-16-9447-9_33.

[15] M. Narnaware, "College Admission Prediction using Machine Learning," *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 11, no. 11, pp. 520–523, Nov. 2023, doi: 10.22214/ijraset.2023.56539.

[16] P. Golden, K. Mojesh, L. M. Devarapalli, P. N. S. Reddy, S. Rajesh, and A. Chawla, "A Comparative Study on University Admission Predictions Using Machine Learning Techniques," *Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol.*, pp. 537–548, Apr. 2021, doi: 10.32628/CSEIT2172107.

[17] X. Chen, Y. Peng, Y. Gao, and S. Cai, "A competition model for prediction of admission scores of colleges and universities in Chinese college entrance examination," *PLOS ONE*, vol. 17, no. 10, p. e0274221, Oct. 2022, doi: 10.1371/journal.pone.0274221