ISSN (Online): 2583-648X

IJIRID

International Journal of Ingenious Research, Invention and Development

An International, High Impact Factor, Double,Blind Peer,Reviewed, Open,Access, Multidisciplinary Online Journal

www.ijirid.in

Volume 4 | Issue 2 | April 2025

AI-Driven System for Engineering College Admission Prediction & Allocation

Sanket Mandavgane[1], Saurabh Vaidya[2], Harsh Rahate[3], Varun Khatake[4], Prof. Vishal Naikwadi[5]

[1,2,3,4]*Student, Zeal College of Engineering and Research, Pune, India*

[5]*Assistant Professor, Zeal College of Engineering and Research, Pune, India*

Abstract: *The college admission process for engineering and technology colleges is an important decision-making process for students, usually determined by factors like CET percentile, location of the college, and college ratings. This paper suggests a machin e learning-based system for predicting and assigning colleges to students on the basis of these factors. The system seeks to simplify the admission process by automatically predicting and allocating colleges through a mixture of supervised and unsupervised machine learning algorithms. The model proposed in this work uses CET percentile, location of college, and college ratings as primary features for predicting the best-fit colleges for students. The system is aimed at helping academic institutions and students make proper choices based on their requirements, hence making the overall process more efficient. The suggested system exploits supervised models such as Decision Trees, Logistic Regression, and K-Nearest Neighbors (KNN) for predicting the chances of admission, and unsupervised models such as K-Means clustering for clustering students according to needs and performance. By making the process automated, the system minimizes manual labor and guarantees fairness and transparency. For students, it gives them individual college recommendations, and for institutions, it provides a data-centric model for seat distribution. The system also uses feature engineering to choose the appropriate features, guaranteeing true predictions. In the future, more features such as extracurricular activities and alumni opinions will be incorporated to make the system more user-friendly.*

Keywords: Admission Prediction, College Allocation, CET Percentile, College Location, College Ratings, Machine Learning, Decision Tree, K-Nearest Neighbors, Logistic Regression.

I. INTRODUCTION

The admission process for engineering and technology colleges is a complex and multi-faceted procedure that involves various factors such as entrance exam scores (e.g., CET percentile), college location, college ratings, and other personal preferences. With the increasing number of students applying for engineering programs each year, the admission process has become more competitive and challenging. Students often face difficulties in selecting the right college that aligns with their academic performance, geographical preferences, and career aspirations. On the other hand, academic institutions struggle to allocate seats efficiently based on student profiles and preferences, often leading to inefficiencies and dissatisfaction among students. This paper proposes a machine learning-based system to predict and allocate colleges to students based on their CET percentile, preferred college location, and college ratings, aiming to streamline the admission process and make it more efficient and student-centric.

 DOI: 10.5281/zenodo.15640665

The main goal of this research is to offer a predictive model that provides students with guidance on which colleges they would be suited to based on their CET percentile and preferences. The proposed system utilizes a combination of supervised and unsupervised machine learning techniques to provide a prediction on the admission likelihood for students into each college and to allocate colleges for students based on their profile. By automating the prediction and allocation, the system is focused on minimizing the complexity and manual effort in the admission process, therefore benefiting both students and educational institutions.

For students, this system provides personalized college recommendations according to their academic performance and preferences, which allows them to make an informed choice. For educational institutions, this system presents a more data reliant way of allocating seats efficiently to ensure the right students are admitted to the right colleges. The proposed model incorporates key features that include the student's CET percentile (indicative of academic performance); preferred location of college (geographical preference); college ratings (almost akin to quality and or reputation); and though not in detail, space availability; cut off percentiles for previous years; and the student's preferred branch of engineering. These features offered a comprehensive perspective to completion and construction of the model which would exclusively benefit colleges and students a like.

The proposed system is intended to consider a number of concerns related to the admission process. First, it will reduce the high amount of information overload for students by offering personalized college recommendations according to profiles students submit. Second, it will reduce the manual workload by academic institutions for processing applications and allocating seats during an admission process, improving the process and speed of admission, for students. Third, it will provide objectivity and transparency into the allocation from data-driven algorithms without human bias. Fourth, because the system will build on algorithms that can take enormous amounts of data, the system will have the potential to scale to the amount of data needed from the admission process of many colleges and thousands of students.

The machine learning approaches in this system utilize supervised learning models such as Decision Trees, Logistic Regression and K-Nearest Neighbors (KNN) as an indicator of likelihood of acceptance. The unsupervised learning approaches include K-Means clustering, which will group students to determine peer preferences and academic performance. The system also encapsulates feature engineering to determine features and extracting attributes of interest to the classifiers and that affect and influence the predictive behaviors on the students' profiles. The models will also be evaluated using traditional measures of accuracy, precision, recall, and F1-score. The clustering will be revealed through silhouette score and within-cluster sum of squares.

The proposed system described in this article has a number of potential applications in the education sector. The system acts as a decision-support tool for students that identifies best-fit colleges based on their academic performance and preferences. The system offers academic institutions a data-driven means of managing admission, enabling them to allocate seats and improve the quality of students admitted to the institution. The system is capable of being connected to existing admission portals and other admission platforms, thereby providing a seamless experience for both students and academic institution.

 DOI: 10.5281/zenodo.15640665

ISSN (Online): 2583-648X
IJIRID
International Journal of Ingenious Research, Invention and Development
An International, High Impact Factor, Double,Blind Peer,Reviewed, Open,Access, Multidisciplinary Online Journal
Volume 4 | Issue 2 | April 2025
www.ijirid.in

In summary, this paper has proposed a machine learning-based system that can predict and allocate colleges to students based on their CET percentile, desired location of college, and rating of colleges. The system would streamline the admission process by automating the prediction and allocation of colleges, thereby helping students and academic institutions. We believe the application of data-driven technology using machine learning algorithms are capable of providing a data driven, fair, transparent, and efficient means of solving the problems facing engineering and technology admissions. Future work on the project will focus on further improving the accuracy and usability features of the system. A number of ideas have been floated to further develop the system, including providing information on student extracurricular activities, college infrastructure, and college alumni feedback.

II. RELATED WORK

The use of machine learning techniques in the education sector, particularly in the area of admission prediction and college allocation, has been the subject of considerable research. Past research has usually focused on predictive algorithms (decision trees, logistic regression, clustering) based upon predictive factors such as academic performance, and scores from entrance exams, as well as several other aspects. Most research papers have not considered the combination of CET percentile, college location, and college rating as primary factors for prediction and allocation. This study will deal with this omission by adopting these essential features into a complete machine learning admission prediction and college allocation system.

In [1] "Admission Prediction in Undergraduate Applications: An Interpretable Deep Learning Approach," they proposed a model that uses Multiple Linear Regression with PCA, Random Forest Regression, and Support Vector Machine (SVM) to predict undergraduate admissions as an area of academic performance. entrance exam scores-type scores but didn't take into account elements like college location or ratings. According to the findings Random Forest Regression and SVM had higher accuracy predicting admissions but without the inclusion of location and rating features ,there is a limit to the model's personalized recommendations for users.

In [2], "College Admission Prediction using Machine Learning", the authors used a model based on Linear Regression and Ridge Regression for predicting college admissions. The authors focused on academic metrics like GPA and entrance test scores (verbal and quantitative). The models did reasonably well in terms of accuracy, but the lack of college location and rating data in the features meant that the model could only provide a limited overview of adoptions. Including additional features, particularly geographical preferences and a rating about the school could have made the results more holistic.

In [3], "Prediction of Graduate Admission Using Machine Learning", the authors used Stacked Ensemble Learning, K- Nearest Neighbour (KNN), and Logistic Regression (LR) to predict graduate admissions decision. The study used metrics like academic performance, entrance scores in exams,

 DOI: 10.5281/zenodo.15640665

ISSN (Online): 2583-648X

IJIRID

International Journal of Ingenious Research, Invention and Development

An International, High Impact Factor, Double,Blind Peer,Reviewed, Open,Access, Multidisciplinary Online Journal

Volume 4 | Issue 2 | April 2025

www.ijirid.in

and recommendation letters as the key features. Although the models provided useful results, similarly to the previous studies, the model did not include a college location or ratings; this is vital for students considering the decision of which institution to attend. The authors were able to show that ensemble learning methods performed better than traditional models, however, again with the limitations of location and ratings for the model in real-world applicability.

In [4], "Post Graduate Admission Prediction System", the authors used Multiple Linear Regression with PCA, Random Forest Regression and Support Vector Machine (SVM) to predict postgraduate admission. The features that this study mainly used were academic performance, research experience, and entrance exam scores. Although these models had high accuracies, by not using college location and college ratings as features, the model had less ability to provide recommendations that would have been accurate for particular students. This study helps signify the importance of including college ratings and college location, to enhance the prediction systems and make more personalized recommendations.

In [5], "A Comparative Study on University Admission Predictions Using Machine Learning Techniques", the authors compared the results of Stacked Ensemble Learning, K-Nearest Neighbour (KNN), and Logistic Regression (LR) predictive models. The features included academic performance, entrance exam scores, and extracurricular activities. The absence of college location and college ratings for features can limit the ability of the model to provide a comprehensive prediction. The authors claimed the results showed Stacked Ensemble Learning and Ensemble Learning were provided the best results, though without the college location and college ratings it was a significant limitation.

Sr. No.	Methodology/ Algorithm Used	Accuracy Achieved
1	Artificial Neural Network	Feed Forward: 60%
2	ICNN with PCA	70.56%
3	Feed-Forward Neural Network with PCA	70.67%
4	Linear Regression	69%
5	Ridge Regression	68%
6	Artificial Neural Network	74.6%
7	Random Forests	68%
8	Linear Regression	60.2%
9	Multiple Linear Regression with PCA	MLR: 70%
10	Random Forest Regression	RFR: 66%

 DOI: 10.5281/zenodo.15640665

ISSN (Online): 2583-648X
IJIRID
International Journal of Ingenious Research, Invention and Development
An International, High Impact Factor, Double,Blind Peer,Reviewed, Open,Access, Multidisciplinary Online Journal
Volume 4 | Issue 2 | April 2025
www.ijirid.in

11	Support Vector Machine	SVM: 55%
12	Stacked Ensemble Learning	Stacked: 71%
13	K-Nearest Neighbour	K-NN: 62%
14	Logistic Regression	LR: 55%

III. METHODOLOGY

The methodology for developing the AI-powered Admission System consists of four primary stages: Data Collection, Pre-processing, AI/ML Model Development, and Dashboard Implementation. Each stage plays a crucial role in ensuring the system's accuracy and usability in predicting and recommending college admissions.

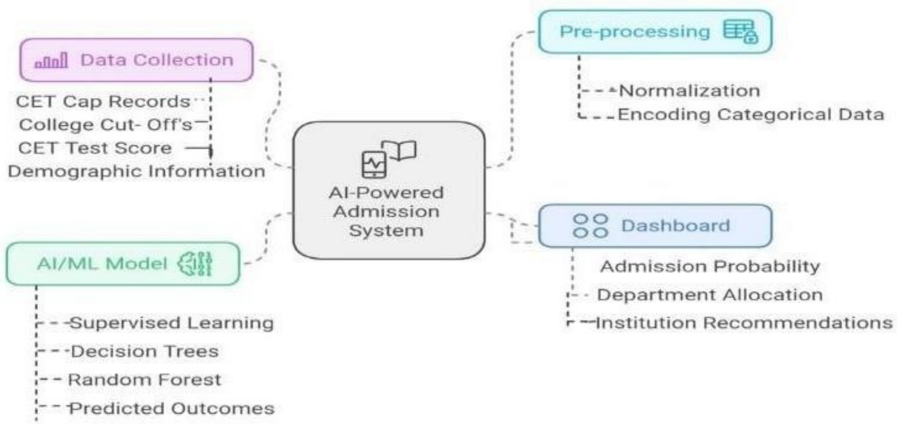

Figure 1: System Architecture

Data Collection:

The first part of this approach is collecting all the data needed to design the admission prediction system. The data includes CET Cap records, data that describes history in admission policies and allotment of seats; college cut off scores; cut off scores provide information about the ideal minimum marks required for attaining admission into different institutions and departments; CET test scores refer to the marks students achieved in the entrance test; as well as socio-economic and geographic factors from the students to be able to monitor trends and patterns that may have an impact on admission decisions. The data records will help us design a system that captures a complete picture and has the information available for making good predictions.

Pre-processing:

The data pre-processing stage is completed after collecting the data in order to prepare it for analysis and modelling. The first stage of pre-processing is normalisation. Normalisation scales numerical values into a common range, which will allow the machine learning algorithms to perform more efficiently. Encoding categorical values is also a critical part of any pre- processing exercise. Categorical

 DOI: 10.5281/zenodo.15640665

ISSN (Online): 2583-648X

IJIRID

International Journal of Ingenious Research, Invention and Development

An International, High Impact Factor, Double,Blind Peer,Reviewed, Open,Access, Multidisciplinary Online Journal

Volume 4 | Issue 2 | April 2025

www.ijirid.in

data, such as the name of a college and demographic characteristics, is converted into numerical formats because machine learning models typically only accept numerical input. After these procedures, you can be assured that the data is clear, orderly and ready to be used in training predictive models.

AI/ML Model Development:

The crux of this process is the making of the machine learning model. For this step, supervised learning techniques are used on the processed data to train the models. The algorithms use four decision-trees which classify the categorical output using decision-trees. Each tree has a number of branches split by the values of the features for the data collected. In order to increase the accuracy of predictions and increase the generalizability of the model, a Random Forest method is employed, that implements multiple decision-trees, thus allowing for multiple decision- trees to be integrated for more robustness concerning each student's characteristics / destination. The output for this step is the predicted results based on probability of admission, recommendations for the students departmental destination, and the recommended institutions for a specific student. The predictions are essentially the main role of the system.

IV. ALGORITHMS

Decision Tree (DT):

Decision Trees (DT) is a supervised machine learning algorithm that can be used for classification or regression tasks. They are employed in numerous fields because they are simple to use, easy to interpret, and can analyze both numerical and categorical data. A decision tree will first partition the data set into smaller subsets based on a parameter (or parameters), or make a sequence of decisions leading to all likely outcomes (for example past behaviour), which then creates a decision tree model structure. The decision tree model structure is composed of decision nodes and leaves. The decision nodes illustrate the features or attributes that are used to split the data, while the leaves show the ending outcome, or class labels.

$$\text{Entropy} = -\sum p(x) \log p(x)$$

Random Forest (RF):

Random Forest (RF) is one ensemble learning method that uses multiple decision trees either for classification and regression. Random Forest uses a procedure called Bootstrap Aggregation (Bagging), where we make several random subsets of the dataset by sampling rows and sampling features (with replacement) to build individual decision trees. Each tree in the forest makes a prediction independently, and then, for classification problems, we determine a single output label according to majority vote, or, for regression problems, we average the predicted outputs across all of the trees. The primary advantage of Random Forest is the ability to reduce overfitting, improve generalization, and mitigate errors over uncorrelated trees. Since Random Forest uses bags of data and features to create individual trees, the individual errors from those trees will likely cancel out. Therefore, we get more accurate and reliable predictions when we aggregate across trees.

DOI: 10.5281/zenodo.15640665

In essence, Random Forest is robust, even when any individual tree makes an incorrect prediction. Specifically, Random Forest has good accuracy performance when working with high-dimensional data, can account for non- linear relationships, and provides feature importance rankings. Therefore, it has become a popular method for tasks like admission prediction, like CET percentiles, college location, and college rankings matter. Using multiple decision trees, Random Forest outperforms individual models, making it a great method for predictive analysis in education and more.
The equation (2), states, that our input x gets assigned to the class of the largest probability.
y^=T1t=1∑Tht(x)

Gradient Boosting Classifier:
In this research, the Gradient Boosting Classifier (GBC) was selected as the core machine learning algorithm for predicting the most probable engineering college allocations based on student profile data. Gradient Boosting is a sequential ensemble learning method that combines multiple weak learners—typically decision trees—into a strong predictive model by minimizing error in a stage-wise fashion. Unlike Random Forest, which builds trees in parallel and combines them via majority voting, Gradient Boosting builds trees sequentially, where each new tree corrects the residuals (errors) of the previous tree. This makes GBC particularly effective for handling structured data with complex interactions and nonlinear decision boundaries.
Technically, Gradient Boosting aims to minimize a loss function L(y,F(x))L(y, F(x))L(y,F(x)) by taking the gradient of the loss and fitting new learners to the negative gradient at each stage. For a regression problem, the process can be described as:
Fm(x)=Fm−1(x)+γmhm(x)
where:

- Fm(x)F_{m}(x)Fm(x) is the prediction at stage m,
- hm(x)h_m(x)hm(x) is the new weak learner (e.g., a decision tree),
- γm\gamma_mγm is the learning rate (step size), and hm(x)≈−∂L(y,F(x))∂F(x)h_m(x)\approx-\frac{\partial L(y, F(x))}{\partial F(x)}hm (x)≈−∂F(x)∂L(y,F(x)) represents the gradient of the loss function.

In contrast to Random Forest, which is more resistant to overfitting and faster to train, Gradient Boosting often yields superior predictive accuracy, especially when the model is well-tuned and the dataset size is moderate. Since our project aims to rank the top 5 colleges with the highest probability of allocation, GBC's ability to produce calibrated probabilities makes it more suitable than Random Forest for this probabilistic ranking task.
Overall, Gradient Boosting proved to be the optimal algorithm for our system, delivering high accuracy and interpretable probability estimates for top-ranked college predictions while maintaining a lightweight model footprint suitable for deployment on limited-resource systems.

Dashboard Development:
The final stage of the AI-powered Admission System will be the interactive dashboard, using Flutter technology, which will provide the front-end interface with which users will interact with the system's

 DOI: 10.5281/zenodo.15640665

ISSN (Online): 2583-648X
IJIRID
International Journal of Ingenious Research, Invention and Development
An International, High Impact Factor, Double,Blind Peer,Reviewed, Open,Access, Multidisciplinary Online Journal
Volume 4 | Issue 2 | April 2025
www.ijirid.in

predictions and recommendations. Designed with a user-friendly approach, the dashboard will provide comprehensive results, such as the admission probability for a student based on his or her CET scores and other attributes. It also provides the most detailed department-wise allocation guidelines so that a student can pick out academic streams suitable to their performance and interest. Moreover, it also indicates which institutions suit a student's profile, hence equipping users to make appropriate admissions decisions. With Flutter, users get the benefits of the strongest UI framework in terms of aesthetically beautiful design with effortless navigation while preserving performance and responsiveness across different devices.

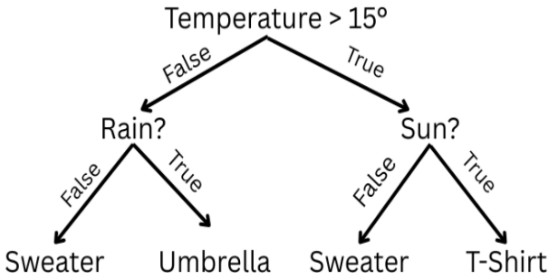

Figure 2 : Decision Tree

Model Evaluation:

We evaluated the performance of a machine learning algorithm we developed to predict college admission for engineering colleges using different classification measures. The accuracy overall was 81%, which shows that the model predicts near accurately this multi-class classification problem, with 1098 classes (colleges).

Classification Report:

The classification report summarizes the precision, recall, and F1-score for individual classes as well as aggregated averages:

Metric	Precision	Reca ll	F1-Score	Support
Macro Avg	0.50	0.78	0.76	1098
Weighted Avg	0.55	0.81	0.81	1098

ROC Curve Analysis:

We used the Receiver Operating Characteristic (ROC) curve and computed the Area Under Curve (AUC) for Gradient Boosting Classifier (GBC) to ascertain how well our admission prediction model performed. The ROC curve highlights the balance between the True Positive Rate (or sensitively) versus the False Positive Rate to demonstrate the model's ability to separate the admitted and not admitted classes. As shown in the ROC curve, the model demonstrates strong discriminative capability for certain classes. Specifically, Class 0 and Class 1 exhibit high area under the curve (AUC) values of

 DOI: 10.5281/zenodo.15640665

ISSN (Online): 2583-648X
IJIRID
International Journal of Ingenious Research, Invention and Development
An International, High Impact Factor, Double,Blind Peer,Reviewed, Open,Access, Multidisciplinary Online Journal
www.ijirid.in
Volume 4 | Issue 2 | April 2025

0.89 and 0.87 respectively, indicating excellent predictive accuracy for these college categories. This suggests that the model is highly effective at identifying students likely to be admitted into these institutions.

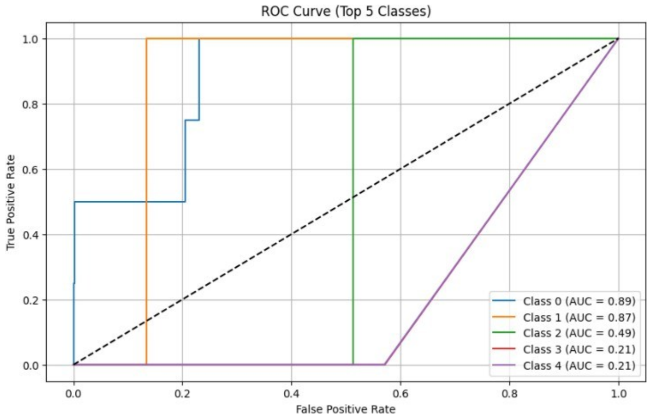

Figure 3: ROC Curve

Confusion Matrix:

The confusion matrix presented above offers a detailed view of the classification performance of the proposed AI- powered college admission prediction system for the top 10 most frequently predicted colleges. A strong diagonal dominance indicates that the model has successfully learned to classify most instances correctly, demonstrating effective predictive capabilities. For example, the model correctly classified 14 instances for College ID 230 and 11 for College ID 130, showcasing a high level of confidence and accuracy in these cases. Similarly, classes such as 34, 138, and 20 also show reliable performance with 7, 5, and 4 correct predictions respectively.

Overall, the results from the confusion matrix affirm that the developed model can provide meaningful and reasonably accurate college predictions for the majority of inputs. This reinforces the practical applicability of our system in assisting students with their college allocation decisions based on entrance exam percentile, category, branch preference, and location.

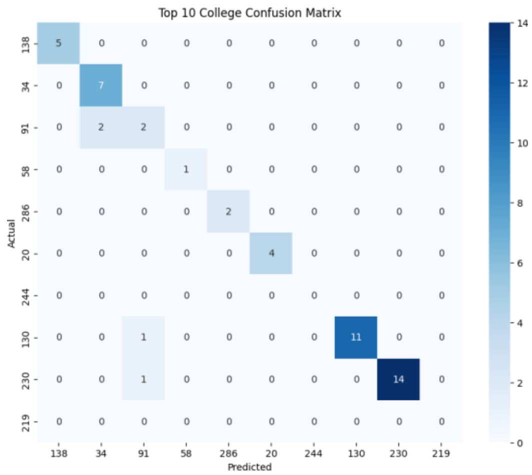

Figure 4: Confusion Matrix

ISSN (Online): 2583-648X
IJIRID
International Journal of Ingenious Research, Invention and Development
An International, High Impact Factor, Double,Blind Peer,Reviewed, Open,Access, Multidisciplinary Online Journal
www.ijirid.in
Volume 4 | Issue 2 | April 2025

V. RESULT AND DISCUSSION

The AI-based Engineering College Admission Prediction System uses state-of-the-art machine learning algorithms, namely the Gradient Boosting Classifier (GBC), to intelligently predict the suitable colleges for students, applicants, or candidates pursuing engineering degrees based on their CET percentile, category, gender, branch choice, and location. The dataset was preprocessed effectively (dealing with missing values, categorical variable encoding, and systematic train-test splitting). The GBC, once trained, is able to provide the top 5 most probable college choices along with reported prediction probabilities. The classification accuracy being 81% is promising given the multi- class and imbalanced classification problem.

The GBC doesn't solely make binary decisions as a classifier but considers predicting probabilities and as an outcome providing students with a more data-driven decision perspective about admission possibilities. Given the described features of the model's outcome, it should help students form more realistic and contextually advisable college choices. While there are a number of variables influencing the output, the prediction model produces output in an interpretable format, as it ranks colleges rather than providing a single output, providing prospective applicants with a range of potential college options.

VI. CONCLUSION AND FUTURE SCOPE

In conclusion, this project effectively showcases the potential of machine learning in revolutionizing the college admission process. By intelligently analyzing inputs such as CET percentile, category, gender, preferred branch, and location, the system delivers personalized and data-driven recommendations. Despite the challenges posed by limited data and the complexity of diverse categorical inputs, the model functions as a practical and insightful recommendation engine. It successfully bridges the gap between raw entrance exam scores and actionable college-level insights, providing students with a reliable and intuitive tool to support informed decision-making.

Looking ahead, several avenues exist to enhance the system further:

- Integration of richer datasets, including historical cut-offs, fee structures, placement statistics, and college reviews.
- Application of more advanced models, such as deep learning or refined ensemble techniques, to boost prediction accuracy.
- Implementation of fairness-aware algorithms to ensure that recommendations remain unbiased and equitable across all demographic categories.

Overall, the system lays a strong foundation for a scalable and intelligent AI-powered admission counseling platform. With continuous improvements and user feedback, it holds the promise of making the admission process more transparent, accessible, and empowering for every aspiring engineering student.

REFERENCES

[1] Priyadarshini, A., Martinez, B., Gago Masague, N. S. (2024). Admission Prediction in Undergraduate Applications: An Interpretable Deep Learning Approach.

DOI: 10.5281/zenodo.15640665

[2] Wang, Z., & Shi, Y. (2016). Prediction of the Admission Lines of College Entrance Examination Based on Machine Learning. Methodologies: Random Forest, AdaBoost.

[3] Patel, S., Waghela, H., Gupta, P., & Rajgor, N. (2022). Prediction of Graduate Admission Using Machine Learning.

[4] Anil, B., Pasha, A., Singh, A. K., & Singh, A. K. (2019). Multiple Machine Learning Classifiers for Student's Admission to University Prediction.

[5] Apoorva, C., Malepati, D. A., Peta, C. N., & Shree, B. S. (2020). Prediction for University Admission Using Machine Learning. Methodologies: Ridge Regression, Random Forest, K-NN. Accuracy: Ridge Regression (80%), Random Forest (95%), K-NN (85%).

[6] "Prediction fora University Admission Using Machine Learning" by Chitra Apoorva D.A, Malepati Chandu Nath, Peta Rohith, Swaroop S, Bindushree S - Blue Eyes Intelligence Engineering & Sciences Publication. International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8, Issue-6 March 2020.

[7] "Machine Learning Basics with the K-Nearest Neighbors Algorithm"- https://towardsdatascience.com/ machine-learning basics-with-the-k- nearest-neighbors algorithm

[8] "Graduate Admission Prediction Using Machine Learning" by Sara Aljasmi, Ali Bou Nassif, Ismail Shahin, Ashraf Elnagar - ResearchGate Publication, December 2020.

[9] Sujay S "Supervised Machine Learning Modelling & Analysis For Graduate Admission Prediction" Published in International Journal of Trend in Research and Development (IJTRD), ISSN: 2394 9333, Volume-7 | Issue-4 , August 2020.

 DOI: 10.5281/zenodo.15640665