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

ADMISSION PREDICTION ANALYSIS USING MACHINE LEARNING

[1]Aditi Patil, [2]Mrudul Kshirsagar, [3]Sanika Sawant, [4]Soham Kale and [5]Arti Devmane
Department of Information Technology,
SSJCOE, Dombivli, India

Abstract : This research suggests a Machine Learning (ML)-driven admission prediction system to assist students in determining their likelihood of gaining admission into B.Tech courses in Indian universities. The model analyzes major parameters like entrance test scores (CET, JEE, SSC, HSC), academic grades (10th and 12th-grade marks), and co-curricular activities. Based on past admission records, it learns patterns to make individualized predictions. The system utilizes ML algorithms such as Random Forest in processing student data and is incorporated into an online platform to give real-time estimates of admission probabilities. Through the simplification of the admission process, this tool facilitates data-based decision-making, maximizes university choice, and enhances rates of admission success.

IndexTerms **- Admission Prediction, Machine Learning, Higher Education, Random Forest, Data-Driven Decision Making, University Selection, Academic Performance.**

I. Introduction

Higher education institution admission processes tend to be complicated and competitive, prompting students to make strategic choices involving a range of academic and non-academic determinants. Admission Prediction Analysis is a sophisticated analysis model that was created to predict and evaluate student admissions based on the analysis of determinants such as academic results, standardized examination scores, and extracurricular activities. This research delves into the methods, data analysis processes, and forecast models used by Admission Prediction Analysis-Add to improve admission predictions. Admission Prediction Analysis-Add brings a data-based decision-making method through the incorporation of machine learning algorithms and statistical analysis, supporting students in making informed decisions and allowing educational institutions to effectively pick potential candidates. With functionalities like real-time predictions, interactive data visualization, and ongoing adaptability, Admission Prediction Analysis-Add seeks to make the admission process more streamlined, enhance transparency, and enable evidence-based university choice.

II. Literature Survey

Researchers in recent years have been actively researching the application of AI and machine learning to forecast undergraduate admissions. Amisha Priyadarshini (2024) has conducted one such study using Deep Neural Networks with the LIME model to detect influential features determining admission. While this approach provided rich information regarding decision-making, it was accompanied by drawbacks—it mostly dealt with undergraduate programs, had extensive computing power requirements, and was challenging to interpret. Likewise, Kelly Van Busum (2023) emphasized

significant issues with AI models, which included prediction bias, excessive dependence on data, restricted generalization, and no updates at regular intervals. These conclusions

demonstrate that while there is promise in AI, much remains to be achieved in terms of transparency and fairness. Other research has been more practical as it opted for a specific use case. For instance, Chithra Apoorva D A (2020) developed machine learning models with Naive Bayes and Linear Regression to assist students who wish to study in the U.S. Still, her model had a narrow coverage, being confined to specific universities and courses, and was unrepresentative of students in terms of diversity. In another paper, George Raftopoulos (2024) employed Logistic Regression and Decision Trees to forecast admissions based on predictor variables such as GRE scores, GPA, and research experience. Although his model sought to ensure greater transparency and fairness in its predictions, it too had pitfalls such as bias, ethical implications, and lower adaptability. Collectively, these studies present both the future and present prospects and limitations of AI in the admission of students and suggest a necessity for broader and more balanced models.

III. Methodology

Problem Understanding:

The first phase involves understanding the most important challenges of students while choosing appropriate universities from entrance examinations such as CET, JEE, and more. Misinformed application decisions arise due to unclear information and non-reliable advice. Admission Prediction Analysis-Add bridges this gap through its data-driven approach to predicting admission.

Data Collection:

Information is collected from a wide range of sources, such as past student admissions, entrance test scores (CET, JEE), academic performance (SSC, HSC), cut-off marks, and co-curricular achievements. This diverse data set ensures thorough analysis of admission criteria.

Data Pre-processing:

The data collected is properly cleaned and standardized. This involves missing value handling, outlier removal, feature normalization, and the generation of new derived features to improve model accuracy. This process prepares the dataset for machine learning use.

Model Selection:

Different machine learning techniques were tested, such as K-Nearest Neighbors (KNN), Linear Regression, and Ridge Regression. Out of performance and fit, Random Forest was chosen for this project as it is powerful and can identify intricate, non-linear relationships between data.

Model Training & Testing:

The Random Forest model is then trained on the past data, such as exam marks and performance. Training and validation sets are divided from the dataset in order to test the prediction performance of the model and avoid overfitting. Performance metrics such as accuracy and F1-score are utilized to estimate effectiveness.

Development of User Interface

A friendly web interface is created where students can enter their academic information and get real-time predictions of admission probabilities. The interface has graphical representations like graphs and charts to facilitate easy interpretation of the results.

Evaluation & Improvement:

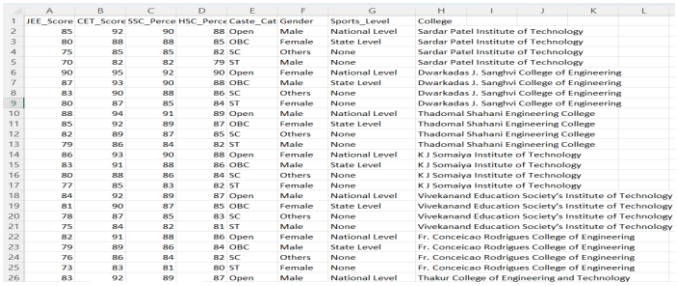
The performance of the model is regularly tracked and assessed. With every new availability of admission data, the system gets updated and retrained to achieve better accuracy and adapt to changing admission patterns. End-user feedback is also integrated to realize better functionality and usability.

Data Cleaning and Analysis

The data is initially preprocessed by eliminating missing values, duplicates and outliers, and feature normalization to maintain uniformity. Categorical variables are one-hot encoded, and additional features are created through feature engineering to improve model performance. Then exploratory data analysis (EDA) is conducted to reveal patterns and relationships, which inform model building and better predictiveness.

Data Visualization

Data visualization is the graphical display of data and information in the form of charts, graphs, and plots. It serves to simplify complex data sets, making trends, patterns, and insights more understandable. In the Admission Prediction Analysis-Add system, visualizations like bar charts, pie charts, and heatmaps are employed to represent admission probabilities and emphasize important factors affecting outcomes. This facilitates user understanding and aids in informed decision-making.



	A	B	C	D	E	F	G	H
1	JEE_Score	CET_Score	SSC_Perce	HSC_Perce	Caste_Cat	Gender	Sports_Level	College
2	85	92	90	88	Open	Male	National Level	Sardar Patel Institute of Technology
3	80	88	88	85	OBC	Female	State Level	Sardar Patel Institute of Technology
4	75	85	85	82	SC	Others	None	Sardar Patel Institute of Technology
5	70	82	82	79	ST	Male	None	Sardar Patel Institute of Technology
6	90	95	92	90	Open	Female	National Level	Dwarkadas J. Sanghvi College of Engineering
7	87	93	90	88	OBC	Male	State Level	Dwarkadas J. Sanghvi College of Engineering
8	83	90	88	86	SC	Others	None	Dwarkadas J. Sanghvi College of Engineering
9	80	87	85	84	ST	Female	None	Dwarkadas J. Sanghvi College of Engineering
10	88	94	91	89	Open	Male	National Level	Thadomal Shahani Engineering College
11	85	92	89	87	OBC	Female	State Level	Thadomal Shahani Engineering College
12	82	89	87	85	SC	Others	None	Thadomal Shahani Engineering College
13	79	86	84	82	ST	Male	None	Thadomal Shahani Engineering College
14	86	93	90	88	Open	Female	National Level	K J Somaiya Institute of Technology
15	83	91	88	86	OBC	Male	State Level	K J Somaiya Institute of Technology
16	80	88	86	84	SC	Others	None	K J Somaiya Institute of Technology
17	77	85	83	82	ST	Female	None	K J Somaiya Institute of Technology
18	84	92	89	87	Open	Male	National Level	Vivekanand Education Society's Institute of Technology
19	81	90	87	85	OBC	Female	State Level	Vivekanand Education Society's Institute of Technology
20	78	87	85	83	SC	Others	None	Vivekanand Education Society's Institute of Technology
21	75	84	82	81	ST	Male	None	Vivekanand Education Society's Institute of Technology
22	82	91	88	86	Open	Female	National Level	Fr. Conceicao Rodrigues College of Engineering
23	79	89	86	84	OBC	Male	State Level	Fr. Conceicao Rodrigues College of Engineering
24	76	86	84	82	SC	Others	None	Fr. Conceicao Rodrigues College of Engineering
25	73	83	81	80	ST	Female	None	Fr. Conceicao Rodrigues College of Engineering
26	83	92	89	87	Open	Male	National Level	Thakur College of Engineering and Technology

Figure 1. Admission Prediction CSV Data

Numeric Factors will generate histograms for JEE_Score, CET_Score, and SSC_Percentage to check how these values are distributed across the dataset.

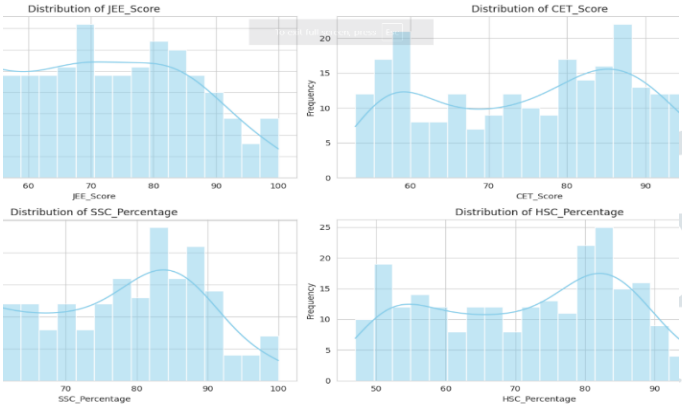


Figure 2. Distribution of Academic Scores Used in Admission Prediction

This map will show you the strength and direction of relationships between JEE_Score, CET_Score, SSC_Percentage, and HSC_Percentage.

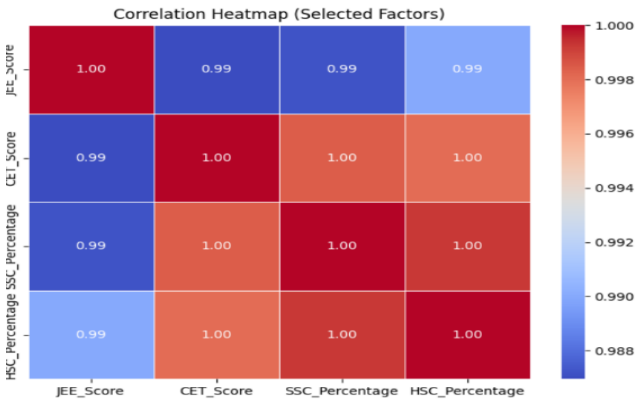


Figure 3. Correlation Heatmap of Academic Performance Metrics

Figure 4 displays the first five rows of the dataset, showcasing key features like scores, percentages, category, and college. These attributes form the input for the Admission Prediction Analysis-Add admission prediction model.

```
Displaying the first 5 rows of the dataset:
   JEE_Score  CET_Score  SSC_Percentage  HSC_Percentage Caste_Category  \
0         85         92              90              88          Open
1         80         88              88              85           OBC
2         75         85              85              82            SC
3         70         82              82              79            ST
4         90         95              92              90          Open

   Gender    Sports_Level                                College
0    Male  National Level          Sardar Patel Institute of Technology
1  Female     State Level          Sardar Patel Institute of Technology
2  Others             NaN          Sardar Patel Institute of Technology
3    Male             NaN          Sardar Patel Institute of Technology
4  Female  National Level  Dwarkadas J. Sanghvi College of Engineering
```

Figure 4. Sample Data Preview of Student Admission Dataset

Figure 5 highlights the relative contribution of each input feature in the prediction model. CET score, SSC percentage, HSC percentage, and JEE score also Gender and Sports Level are the most influential factors in determining admission chances.

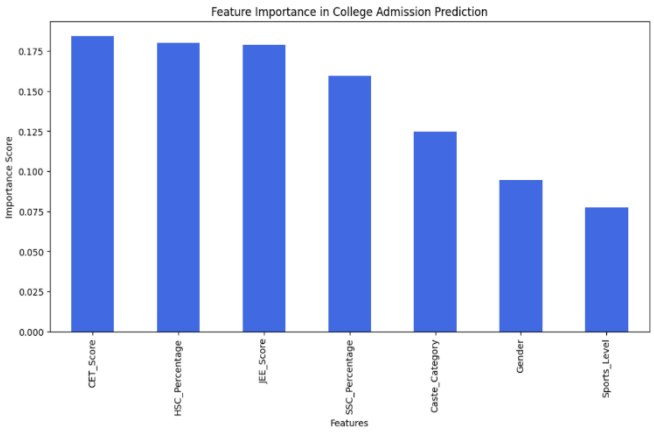


Figure 5. Feature Importance in College Admission Prediction

IV. Algorithm

The Admission Prediction Analysis-Add system uses the Random Forest algorithm, a powerful ensemble learning technique for classification tasks. It works by building multiple decision trees during training and outputs the mode of the classes (for classification) or the mean prediction (for regression) of the individual trees. This approach reduces overfitting and improves accuracy compared to a single decision tree.

Random Forest is well-suited for admission prediction as it handles large datasets with high dimensionality, manages missing values effectively, and captures complex interactions between features. It also provides feature importance scores, helping identify which parameters most influence admission outcomes. This makes it a reliable and interpretable choice for modeling college admission predictions.

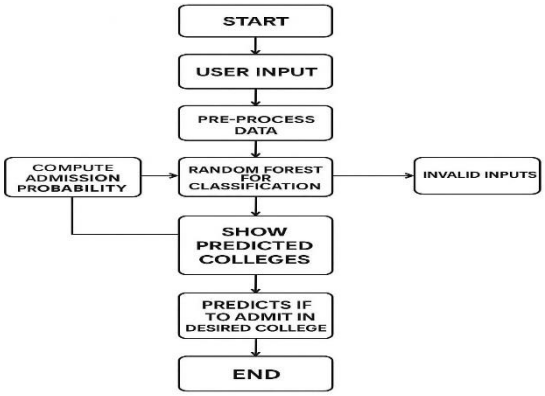


Figure 5. Flowchart

V. Conclusion

The Admission Prediction Analysis-Add: Admission Prediction Analysis system effectively resolves the problems of students while applying for college admissions through machine learning and statistical analysis. By evaluating different parameters like exam scores, academic percentages, caste category, gender, and extracurricular activities, the system gives precise predictions regarding admission opportunities. This enables students to make more informed application choices and avoids the possibility of losing out on good opportunities. The immediate feedback and friendly interface further enhance the overall experience, enabling students to know where they are and how they can get better.

On the institutional front, Admission Prediction Analysis-Add assists universities and colleges in optimizing their candidate assessment process by finding high-potential candidates at an efficient cost. The flexility of the system in responding to the varying requirements of the institution makes Admission Prediction Analysis-Add a useful tool in tailoring the predictions to the needs of particular admission criteria. With the continued learning from new data, the prediction becomes increasingly more precise and credible. In general, Admission Prediction Analysis-Add encourages a data-driven, open, and streamlined admissions process that works to the advantage of both students and educational institutions.

References

[1] Amisha Priyadarshini, Barbara Martinez-Neda, Sergio Gago-Masague, Admission Prediction in Undergraduate Applications: an Interpretable Deep Learning Approach, Irvine, CA, 2024.

[2] Kelly Van Busum, Shiaofen Fang, Bias Analysis of AI Models for Undergraduate Student Admissions, Indiana, United States, 2024.

[3] Chithra Apoorva D A, Malepati ChanduNath, Peta Rohith, Bindu Shree.S, Swaroop.S, Prediction for University Admission using Machine Learning, 2020.

[4] George Raftopoulos † , Gregory Davrazos † andSotiris Kotsiantis, Fair and Transparent Student Admission Prediction Using Machine Learning Models, Greece, 2024.