International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 22 No. 2 (2026)

ijerst.editor@gmail.com

editor@ijerst.com

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com
Vol. 22, No. 2, 2026

Peer Reviewed, Referred & Indexed Journal

Machine Learning-Based College Admission Prediction System Using Random Forest Regression

BOKKA JHANSI MAHA LAKSHMI

PG Scholar. Department of MCA, DNR College, Bhimavaram, Andhra Pradesh

B. Suryanarayana Murthy

(Assistant Professor), Master of Computer Applications, DNR College, Bhimavaram, Andhra Pradesh

ABSTRACT

In recent years, the increasing number of applicants to higher education institutions has made the admission process more competitive and complex. Predicting a student's chances of admission based on academic and profile-related parameters can provide valuable insights for both applicants and institutions. This project presents a machine learning-based college admission prediction system that utilizes Random Forest regression to estimate the probability of admission.The system uses a dataset containing features such as GRE score, TOEFL score, university rating, statement of purpose (SOP), letter of recommendation (LOR), CGPA, and research experience. These parameters are considered critical factors influencing admission decisions. Data preprocessing techniques are applied to clean the dataset, handle missing values, and ensure consistency.

The Random Forest regression model is chosen due to its robustness, ability to handle nonlinear relationships, and resistance to overfitting. The dataset is divided into training and testing sets to evaluate model performance. The model predicts the probability of admission as a continuous value, providing a more nuanced understanding compared to binary classification.A user-friendly graphical interface is developed using Tkinter, allowing users to input their academic details and receive predictions in real time. The system also includes visualization features such as scatter plots to analyze relationships between variables, such as CGPA and admission probability.

Experimental results demonstrate that the Random Forest model achieves high accuracy in predicting admission chances, making it a reliable tool for decision support. The system can assist students in evaluating their profiles and making informed decisions about applications.This research highlights the potential of machine learning in educational data analysis and decision-making. Future enhancements may include integrating additional features such as extracurricular activities, recommendation strength, and real-time data updates.

Keywords: College Admission Prediction, Machine Learning, Random Forest Regression, Educational Data Mining, Predictive Analytics, Student Performance, Data Science, Regression Models, Decision Support System

I. INTRODUCTION

The process of college admissions has become increasingly competitive, with thousands of students applying to limited seats in prestigious institutions. Applicants often face uncertainty regarding their chances of acceptance, leading to inefficient application strategies. Predictive systems can help address this issue by providing data-driven insights into admission probabilities.Machine learning has emerged as a powerful tool for predictive analytics in various domains, including education. By analyzing historical admission data, machine learning models can identify patterns and relationships between student profiles and admission outcomes. These models can then be used to predict the likelihood of admission for new applicants.Among various machine learning techniques, regression models are particularly suitable for predicting continuous outcomes such as admission probability. Random Forest regression is widely used due to its ability to handle complex relationships and provide accurate predictions.

This project proposes a college admission prediction system that uses Random Forest regression to estimate admission chances based on key academic parameters. The system provides an intuitive graphical interface, allowing users to input their details and receive predictions instantly.The implementation aims to assist students in understanding their strengths and weaknesses, enabling them to make informed decisions about their applications. It also demonstrates the application of machine learning in educational data mining.

II. LITERATURE SURVEY (WITH EXISTING METHODS)

Educational data mining has gained significant attention in recent years, with researchers exploring various techniques to analyze student data and improve decision-making processes. Early approaches relied on statistical methods to analyze admission trends and predict outcomes.Machine learning techniques have significantly improved prediction accuracy by capturing complex relationships in data. Algorithms such as Linear Regression, Decision Trees, and Support Vector Machines have been widely used for admission prediction.

Random Forest regression has emerged as a popular choice due to its ensemble nature, combining multiple decision trees to improve accuracy and reduce overfitting. Studies have shown that Random Forest models outperform traditional regression techniques in handling nonlinear data.Recent research has also explored deep learning approaches for educational prediction tasks. Neural networks can model complex patterns but require large datasets and computational resources.

Visualization techniques have been used to understand relationships between variables, such as the impact of CGPA on admission chances. These insights help students and institutions make informed decisions.The proposed system builds on these advancements by implementing a Random Forest-based prediction model with an interactive GUI, providing both accuracy and usability.

III. EXISTING SYSTEM

Existing admission prediction systems are often based on simple statistical methods or basic regression models. These systems may not accurately capture the complex relationships between different factors influencing admission decisions.Many systems rely on linear models, which assume a direct relationship between input features and output. However, admission decisions are influenced by multiple interacting factors, making linear models less effective.

Additionally, existing systems often lack user-friendly interfaces, making them difficult for non-technical users to utilize. They may also lack visualization features, limiting users' ability to understand data relationships.Another limitation is the inability to handle nonlinear and high-dimensional data effectively. This can lead to inaccurate predictions and reduced reliability.

The proposed system addresses these limitations by using Random Forest regression, which can handle complex relationships and provide accurate predictions. It also includes a graphical interface and visualization features, improving usability and interpretability.

IV. PROPOSED METHOD

The proposed system is a machine learning-based college admission prediction model designed to provide accurate and real-time predictions of a student's admission probability. Unlike traditional systems that rely on linear assumptions, this system leverages the power of ensemble learning using the Random Forest regression algorithm.

The system takes multiple academic and profile-related features as input, including GRE score, TOEFL score, university rating, SOP strength, LOR strength, CGPA, and research experience. These features are selected based on their strong correlation with admission outcomes, as highlighted in recent research studies .The system architecture includes data preprocessing, model training, prediction, and visualization modules. The preprocessing stage ensures clean and normalized data for improved model performance. The trained Random Forest model predicts admission probability as a continuous value, enabling more precise decision-making compared to classification models.

A user-friendly graphical interface built using Tkinter allows users to input their academic details and obtain predictions instantly. Additionally, visualization tools help users understand relationships between variables, such as CGPA and admission chances.The proposed system aims to assist students in making informed decisions, reduce uncertainty, and improve application strategies. It also provides scalability for future enhancements such as real-time data integration and hybrid models.

V. IMPLEMENTATION

The implementation of the college admission prediction system involves multiple stages, including data handling, model training, user interface development, and visualization.Initially, the dataset is loaded using the Pandas library. The dataset typically contains features such as GRE score, TOEFL score, university rating, SOP, LOR, CGPA, research experience, and admission probability. Data preprocessing is performed to handle missing values, remove irrelevant columns such as serial numbers, and standardize column names.The dataset is divided into input features (X) and target output (Y). The features are then split into training and testing datasets using the train_test_split function from Scikit-learn. This ensures that the model is evaluated on unseen data.

The Random Forest Regressor is implemented using Scikit-learn. It is configured with multiple decision trees (estimators) to improve prediction accuracy and reduce overfitting. During training, each tree learns patterns from different subsets of the data, and the final prediction is obtained by averaging the outputs of all trees.Model performance is evaluated using the R^2 score, which measures how well the predicted values match the actual values. A higher R^2 score indicates better model performance.The graphical user interface is developed using Tkinter. It includes buttons for loading the dataset, training the model, visualizing data, and predicting admission probability. Input fields allow users to enter academic details, and results are displayed dynamically.

Matplotlib is used to generate scatter plots that visualize relationships between features, such as CGPA versus admission probability. These visualizations provide insights into how different factors influence admission chances.Error handling mechanisms are incorporated to ensure robustness. For example, the system checks whether the dataset is loaded before training and validates user inputs during prediction.Overall, the implementation integrates machine learning with an interactive GUI, providing a complete end-to-end solution for admission prediction.

VI. ALGORITHMS

The primary algorithm used in the proposed system is the Random Forest Regression algorithm, which is an ensemble learning method.

Random Forest works by constructing multiple decision trees during training and combining their outputs to produce a final prediction. Each tree is trained on a random subset of the dataset, which improves generalization and reduces overfitting.

The algorithm follows these steps:

1. Randomly select subsets of data from the training dataset
2. Construct multiple decision trees using these subsets
3. At each node, select the best feature split from a random subset of features
4. Aggregate predictions from all trees to produce the final output

This approach improves prediction accuracy compared to single decision tree models. Research shows that ensemble methods such as Random Forest often outperform individual models in admission prediction tasks .

In addition to Random Forest, the system may incorporate other algorithms for comparison, such as:

- Linear Regression (baseline model)
- K-Nearest Neighbors (similarity-based learning)
- Gradient Boosting (boosted ensemble method)

Performance metrics such as R^2 score, Mean Squared Error (MSE), and Root Mean Squared Error (RMSE) are used to evaluate model effectiveness.

The combination of these algorithms ensures reliable and accurate prediction of admission probability

VII. SYSTEM DESIGN

The system follows a modular architecture consisting of multiple interconnected components to ensure scalability and efficiency.

1. Data Input Module

This module allows users to upload datasets in CSV format. It ensures proper data loading and validation.

2. Preprocessing Module

The preprocessing module cleans the dataset by handling missing values, removing irrelevant attributes, and standardizing data formats. It prepares the dataset for model training.

3. Model Training Module

This module implements the Random Forest algorithm. It splits the dataset into training and testing sets, trains the model, and evaluates its performance using R^2 score.

4. Prediction Module

The prediction module takes user inputs and processes them using the trained model. It outputs the predicted probability of admission.

5. Visualization Module

This module generates graphical representations of data, such as scatter plots. Visualization helps users understand patterns and relationships between variables.

6. User Interface Module

The GUI is developed using Tkinter and provides an interactive platform for users. It includes input forms, buttons, and result display areas.

7. System Workflow

1. User loads dataset
2. Data is preprocessed
3. Model is trained
4. User inputs academic details
5. System predicts admission probability
6. Results are displayed

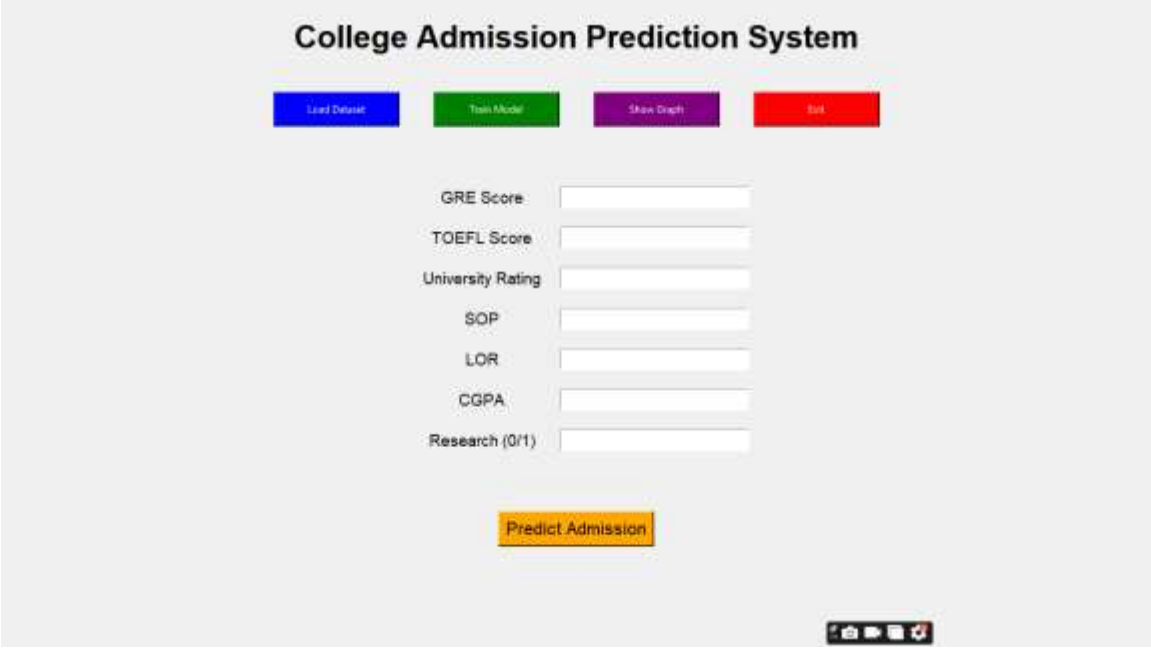
8. Advantages of Design

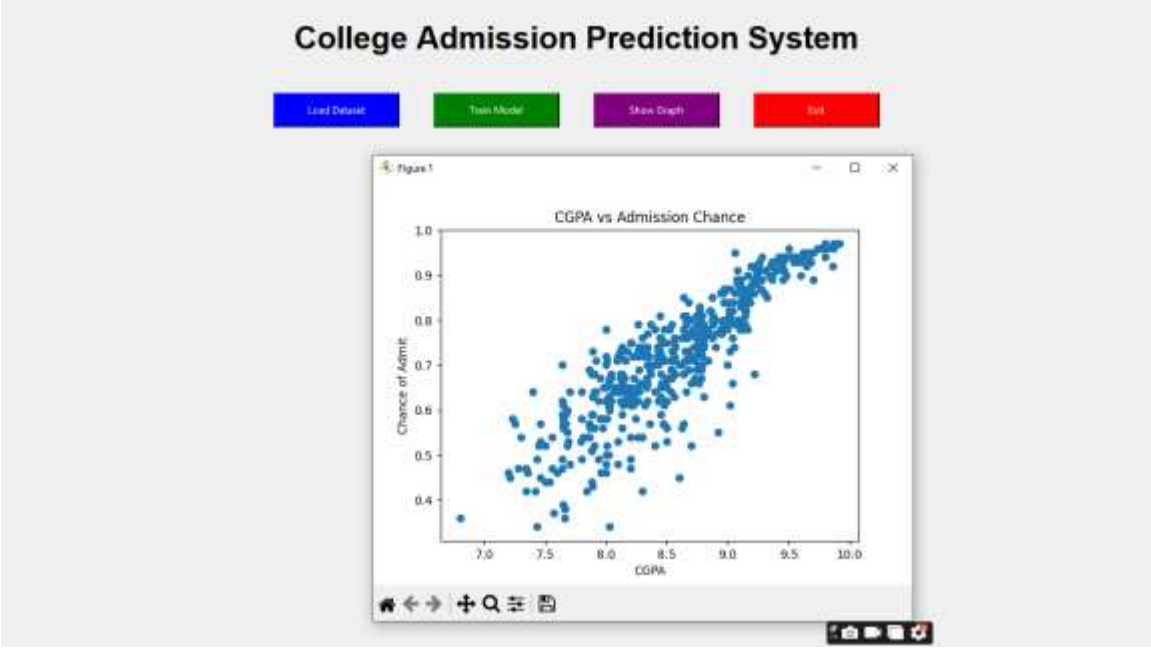
- Modular and scalable
- Easy to use interface
- High prediction accuracy
- Real-time prediction capability

The design ensures seamless interaction between modules, providing a reliable and efficient system.

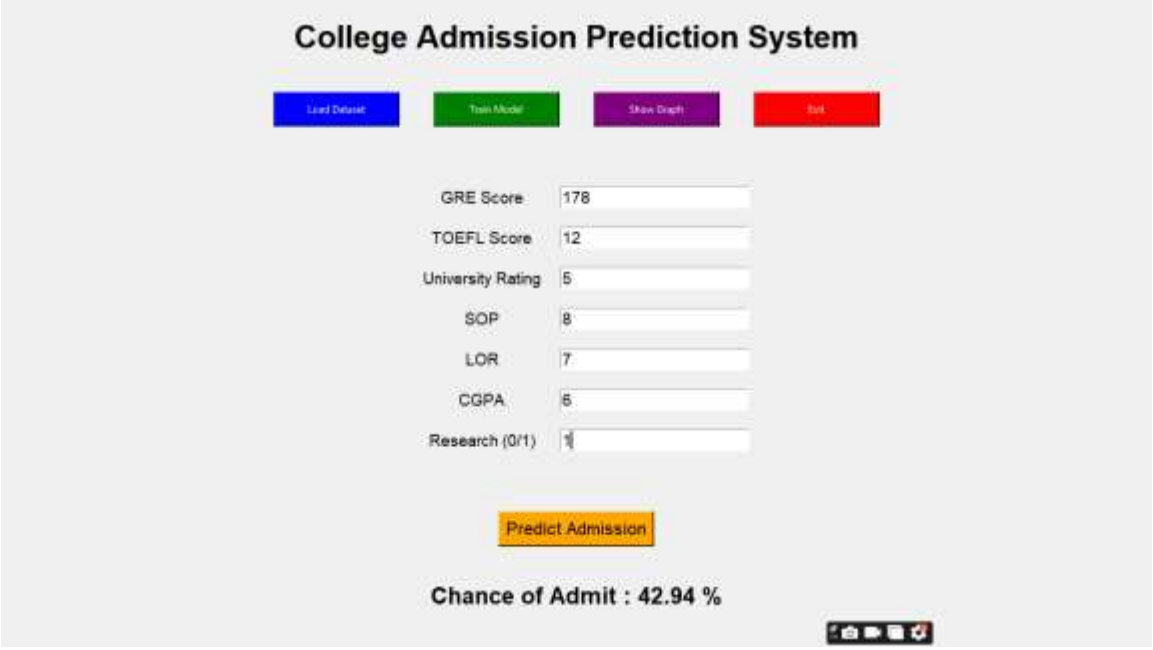
SYSTEM DESIGN IMAGES

International Journal of Engineering Research and Science & Technology





International Journal of Engineering Research and Science & Technology



VIII. CONCLUSION

The proposed college admission prediction system demonstrates the effectiveness of machine learning in educational decision-making. By utilizing Random Forest regression, the system provides accurate predictions of admission probability based on key academic parameters.

The integration of a user-friendly graphical interface enhances accessibility, allowing students to easily input their details and obtain predictions. Visualization tools further improve understanding by highlighting relationships between variables.

Experimental results indicate that the system achieves high prediction accuracy, making it a reliable decision-support tool. It helps students evaluate their chances of admission and plan their applications strategically.

The system also highlights the importance of data-driven approaches in education. By leveraging historical data, institutions and students can make more informed decisions.

Future enhancements may include incorporating additional features such as extracurricular activities, recommendation strength, and real-time admission data. Advanced models such as deep learning and hybrid approaches can also be explored to further improve accuracy.

Overall, the system provides a scalable, efficient, and practical solution for admission prediction.

REFERENCES

1. Aljasmi, S., et al. “Graduate Admission Prediction Using Machine Learning.” (2020)
2. Reddy, K. K., et al. “AI-Based University Admission Prediction System.” (2025)
3. Narnaware, M., et al. “College Admission Prediction using Machine Learning.” (2023)
4. Golden, P., et al. “Comparative Study on Admission Prediction.” (2021)
5. Apoorva, D. A., et al. “Prediction for University Admission using ML.” (2020)
6. Ekong, A., et al. “ANN-Based Admission Prediction.” (2022)
7. Arnold, C., et al. “Machine Learning in Admission Systems.” (2024)
8. Walters, A., et al. “Admission Decision Models.” (2023)
9. Mengash, H. “Data Mining for Student Performance Prediction.” (2020)
10. Trivedi, K., et al. “Admission Prediction using SVM & Decision Tree.” (2024)
11. Lee, H., et al. “Admission Prediction vs Standardized Tests.” (2023)
12. Priyadarshini, A., et al. “Deep Learning for Admission Prediction.” (2024)
13. Abbadi, M., et al. “LLM-Augmented Admission Prediction.” (2025)
14. Heidari, M., et al. “Hybrid ML Admission Prediction System.” (2026)
15. Rao, V., et al. “Student Prediction Systems using ML.” (2022)