Forecasting Graduate Admission Outcomes: A Deep Learning Approach for Predictive Modeling

1st Ahanab Hafiz Anab
dept. of Computer Science and Engineering
Ahsanullah University of Science and Technology
Dhaka, Bangladesh
190204028@aust.edu

2nd Mariah Tasnim Molla
dept. of Computer Science and Engineering
Ahsanullah University of Science and Technology
Dhaka, Bangladesh
190204030@aust.edu

3rd Arif Hossen
dept. of Computer Science and Engineering
Ahsanullah University of Science and Technology
Dhaka, Bangladesh
190204034@aust.edu

4th Shamsia Haque Sristy
dept. of Computer Science and Engineering
Ahsanullah University of Science and Technology
Dhaka, Bangladesh
190204038@aust.edu

I. Introduction

Nowadays, a lot of students show interest in studying abroad rather than their home countries. This is mainly due to the hope of getting better education and having a better lifestyle eventually. Generally, students target universities in countries like Canada, USA, UK, Malaysia. There are some prerequisites,systematic processes for applying in these universities.It is a hard process to go through on their own by the students hence they often take help from education consultancies. This requires the students to pay a certain amount of charges to these consultancies. We have made this project to find the ratings of the universities by utilizing the attributes of the students GRE, TOFL, IELTS scores etc. In this paper, we tackle the crucial issue of student ad- mission in educational institutions, focusing on predicting a student's likelihood of being accepted into a master's program using machine learning models. Our primary goal is to em- power students by providing them with a tool to assess in advance whether they stand a chance of admission.

This paper endeavors to construct a comprehensive machine learning model, incorporating various influential factors including GRE Score, TOEFL Score, University Ranking, Research Experience, Undergraduate GPA, and Work Experience. By assimilating these diverse inputs, the model is designed to forecast the probability of admission. The integration of these specific elements aims to provide a nuanced and thorough analysis, offering a predictive tool that encompasses the multifaceted aspects contributing to the likelihood of acceptance into a given academic program. Through the amalgamation of key parameters, the model seeks to capture the intricate interplay between academic achievements, research background, and professional experience, thereby contributing to a more holistic understanding of the admission process.

II. LITERATURE REVIEW

In their study the authors Addressed the challenge faced by graduate students in selecting universities for their master's programs.Author utilized machine learning algorithms (Linear Regression, Decision Tree, Logistic Regression) to predict university ratings for student admission based on Kaggle data,split the dataset for training and testing, trained models, and evaluated their performance using RMSE (Root Mean Squared Error),concluded that Logistic Regression showed the best predictive accuracy (minimal error) for admission chances compared to other models.Author suggested Logistic Regression's potential in aiding students' admission decisions and proposed future work using machine learning to assist in the admission process.[1]

The author aimed to predict admission chances for master's programs using machine learning.The dataset used contained information such as GRE scores, TOEFL scores, university ratings, SOP, LOR, CGPA, and research experience.The author conducted descriptive statistics, normality tests, outlier detection, feature selection, and multicollinearity analysis on the dataset.They developed models using Multiple Linear Regression, K-Nearest Neighbor, Random Forest, and Multilayer Perceptron algorithms.The dataset was divided into training and testing sets to train and evaluate the models.[2]

In this study author addresses the limitations of traditional educational counseling methods, particularly in providing personalized guidance to high school seniors for their tertiary education choices.Data collected from 500 graduates in Morocco, gathering academic history, preferences, and eligibility criteria from 31 institutions.Discusses the analysis of collected data, identifying correlations between academic performance, subject grades, and admission chances, showcasing the relationship between various academic features and the

likelihood of admission.Presents the process of developing machine learning models for admission prediction, testing 13 algorithms, including Decision Trees, Regression models, Boosting techniques, and identifying the Huber Regressor as the most accurate model.[3] in this study ,Author helping students predict their chances of university .Built an ANN model using Keras with 11 layers and appropriate activation functions.Split the data, trained the model, and evaluated its accuracy (94%) using MSE.Used a dataset from Kaggle with attributes like GRE score, TOEFL, etc., to predict "Chance of Admit."[4]

Author focused on high rates of attrition and low completion rates in nursing schools despite efforts to maximize enrollment.Collected data from 773 full-time undergraduate nursing students enrolled from 2004 to 2012 in a private urban Midwest university.Employed eight machine learning algorithms to predict graduation likelihood at various time points during the students' academic journey.Constructed 14 scenarios using different variable combinations and time points, employing machine learning models to predict graduation outcomes.Achieved prediction accuracy of over 80% after the first college year, reaching 90% after the second year and 99% after the third year.[5]

III. Dataset

A customized dataset is created , including GRE score, IELTS score, TOEFL score, CGPA, research experience, work experience, and projects. For individuals who provide an IELTS score, the corresponding TOEFL score column is filled with the equivalent value, and vice versa. All data is sourced from actual student profiles to ensure accuracy and authenticity.The target variable is the University Rating.

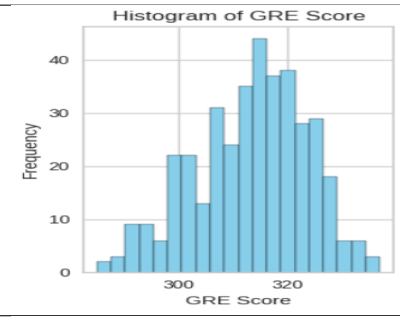
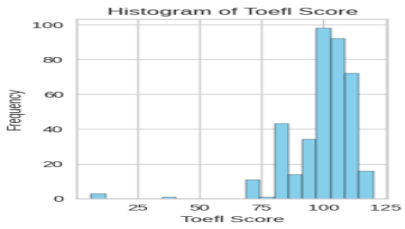
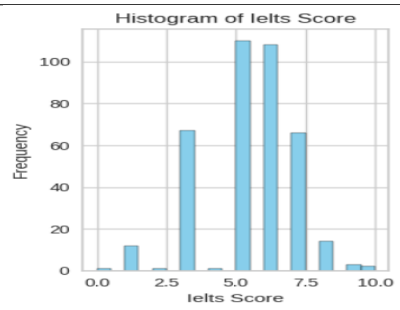
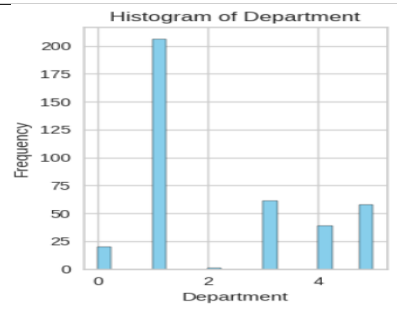
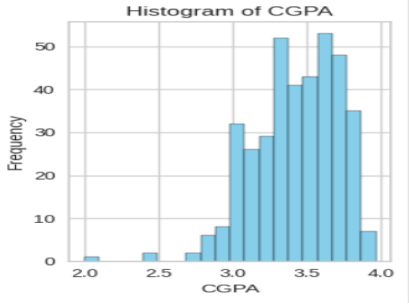
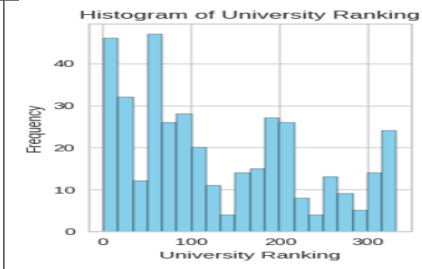
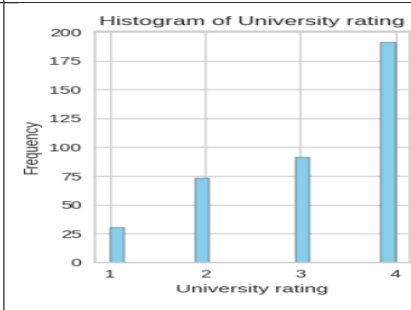
TABLE I
FIG:Admission Data

Serial No	GRE Score	Toefl Score	Ielts Score	Department	CGPA	Research	Project	Work Experience	University Ranking	University Rating
1	299	105	7.5	cse	3.74	0	1	3	250	2
2	303	98	7	cse	3.68	0	0	0	250	2
3	300	98	7	cse	3.68	0	0	1	250	2
4	310	98	7	cse	3.78	1	1	2	133	3
5	306	98	7	cse	3	0	3	1	133	3
6	311	85	6.5	cse	3.36	0	0	0	133	3
7	306	101	7.5	cse	3.78	0	3	4	115	3
8	308	108	7.5	cse	3.84	0	8	1	115	3
9	309	106	7.5	cse	3.74	0	3	3	115	3
10	305	100	7	cse	3.68	1	0	0.6	299	2
11	301	85	6.5	cse	3.42	0	2	3.5	299	2
12	300	85	6.5	cse	3.3	0	0	1.8	299	2
13	295	85	7	cse	3.71	0	0	4	115	3
14	314	103	7.5	cse	3.81	0	0	1	115	3
15	304	112	8	cse	3.3	0	1	2	115	3
16	310	105	7.5	cse	3.78	0	2	0.2	115	3
17	321	99	7	cse	3.33	0	0	3.5	115	3
18	306	98	7	cse	3.14	0	0	0	219	2
19	306	105	7.5	cse	3.6	0	0	3.5	219	2
20	305	88	6.5	cse	3.48	0	0	0	219	2
21	306	95	6.5	cse	2	0	0	5	219	2

TABLE II
FIG:TABLE OF DESCRIPTION DATA

Column	Description
GRE Score (Out of 340)	The GRE (Graduate Record Examination) is a standardized test for graduate and business school admissions. Administered by ETS, it evaluates verbal reasoning, quantitative reasoning, critical thinking, and analytical writing. Scores aid in admission decisions. Similar to the GMAT and SAT, it assesses skills crucial for higher education success.
Toefl Score(Out of 120)	TOEFL (Test of English as a Foreign Language) is a standardized test assessing English language proficiency for non-native speakers. Required for admission to English-speaking universities and institutions globally.
IELTS Score(out of 8)	IELTS (International English Language Testing System) score measures English language proficiency for non-native speakers. Used for university admissions and immigration purposes, it assesses listening, reading, writing, and speaking skills.
Department	The Department column indicates from which department the candidates have been selected.
CGPA	The CGPA column displays scores out of 4. If a candidate's scale is out of 10, it is converted to a scale out of 4 for uniformity and comparison purposes.
Research	The research column displays the counts of research papers published in any journal or conference.
Project	The Project column indicates the count of projects a candidate has completed.
Work Experience	The Work experience column shows the years of work experience that a candidate has.
University rankings	The University ranking column denotes the ranking of universities selected from the US News website.
University Ratings	The university has been rated into four categories, with 4 indicating the highest rating.

TABLE III
HISTOGRAM OF FEATURES

GRE Score	
Toefl Score	
Ielts Score	
Department	
CGPA	
Research	
Project	
Work Experience	
University Ranking	
University Rating	

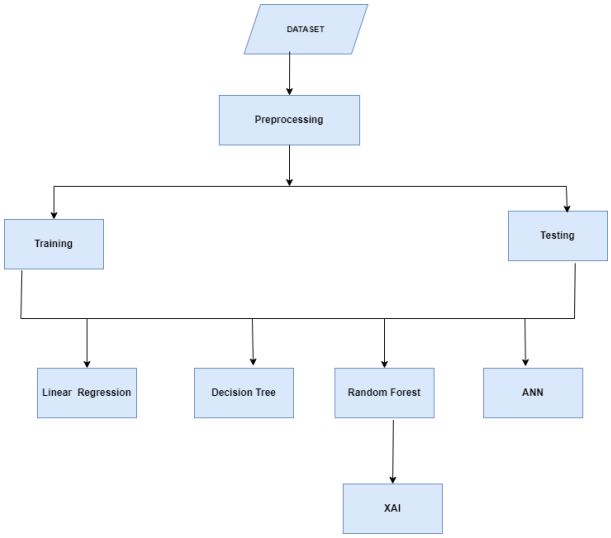


Fig. 1. Methodology

IV. METHODOLOGY

We sourced individual student data from the Yocket website, ensuring proper referencing for authenticity. As our dataset comprises students appearing in either the IELTS or TOEFL exams, we normalized scores for equivalency across exams.

For preprocessing, we employed Boxplot analysis to identify outliers. Using the interquartile range method, we effectively identified and removed outliers to maintain dataset consistency and reliability.

Histogram analysis provided insights into data variance and distribution, critical for model performance assessment.

Data was split into 80% training and 20% testing sets. Linear regression, decision tree, random forest, and artificial neural network models were trained and evaluated.

To enhance interpretability, we integrated XAI. Leveraging Lime, especially with complex models like Random Forests, facilitated understandable explanations for individual predictions. This transparency fosters trust and comprehension in the model's decision-making process.

V. MODEL TRAINING AND EVALUATION

For artificial neural networks (ANN), the RMSE values for both the training and test sets serve as metrics to evaluate the model's predictive performance. Lower RMSE values indicate better predictive accuracy, hence are preferred. The accuracy scores are typically represented as percentages, denoting the proportion of variance captured by the model's predictions. In the context of ANN, higher accuracy values, such as 92%, are highly desirable as they signify the model's capability to explain a significant portion of the data's variance.

For linear regression the RMSE values for both the training and test sets provide a measure of the model's prediction error. Lower RMSE values are desirable.The accuracy scores are printed as percentages, representing the proportion of variance explained by the model. Higher accuracy values are preferable.It's important to compare the training and test scores to assess whether the model is overfitting or generalizing well to new data. If the training score is significantly higher than the test score, it may indicate overfitting.Regression Model provides an accuracy of 85%

The accuracy score is given as 92%, indicating that the Decision Tree Regressor is able to explain a substantial proportion of the variance in the target variable on the test set.The high accuracy score suggests that the Decision Tree model performs well on the given test data.

Random Forests are an ensemble of decision trees, and they are known for their robustness and ability to capture complex relationships in the data.The accuracy score is 92%, indicating the proportion of variance in the target variable that the RandomForestRegressor is able to explain on the test set

Lime (Local Interpretable Model-agnostic Explanations) is used with complex models like Random Forests to provide interpretable explanations for individual predictions. This helps in understanding the factors influencing a specific prediction, promoting transparency, and building trust in the model's decision-making process. Lime's model-agnostic nature allows it to be applied to a variety of machine learning models, providing local explanations that are easier to interpret for humans.

VI. RESULT ANALYSIS

TABLE IV
ACCURACY COMPARISON

Model	**Accuracy (%)**
Random Forest	96.09%
Decission Tree	96.14%
Linear Regression	90.61%
ANN	95.78%

VII. CONCLUSION

In this study, we successfully developed a predictive model to estimate the likelihood of admission for candidates with an impressive accuracy of 96% and minimal Mean Square Error. By leveraging attributes like exam scores, our model demonstrated robust predictive capabilities.

Looking ahead, our project holds promising avenues for expansion. Incorporating additional features such as internship details and other pertinent parameters could enrich the model's predictive accuracy and provide a more comprehensive understanding of admission determinants.

Moreover, exploring advanced neural network architectures or implementing ensemble techniques presents opportunities to further enhance the model's performance. These avenues warrant exploration in future iterations of the project, potentially refining predictive accuracy and broadening its applicability in

admissions decision-making processes. Overall, our study lays a solid foundation for future research endeavors in predictive modeling for admission processes.

REFERENCES

[1] Graduate admission prediction using Machine learning. (2020). International Journal of Computers and Communications, 14. https://doi.org/10.46300/91013.2020.14.13

[2] Al-Ghamdi, A. a. M., Barsheed, A., Almshjary, H., Alghamdi, H. S. (2020). A Machine Learning Approach for Graduate Admission Prediction. International Journal of Computers and Communications. https://doi.org/10.1145/3388818.3393716

[3] Majjate, H., Bellarhmouch, Y., Jeghal, A., Yahyaouy, A., Tairi, H., Zidani, K. A. (2023). AI-Powered academic guidance and counseling system based on student profile and interests. Applied System Innovation, 7(1), 6. https://doi.org/10.3390/asi7010006

[4] Ijraset. (n.d.). Post graduate admission prediction using ANN. IJRASET. https://www.ijraset.com/research-paper/post-graduate-admission-prediction-using-ann

[5] Hannaford, L., Cheng, X., Kunes-Connell, M. (2021). Predicting nursing baccalaureate program graduates using machine learning models: A quantitative research study. Nurse Education Today, 99, 104784. https://doi.org/10.1016/j.nedt.2021.104784

[6] https://www.ets.org/toefl/score-users/ibt/compare-scores.html