PREDICTING CHANCE OF GRADUATE ADMISSION USING LINEAR REGRESSION

Hazim Fitri Bin Ahmad Faudzi | P152419

Universiti Kebangsaan Malaysia
The National University of Malaysia

Fakulti Sains & Teknologi

INTRODUCTION

Graduate school admissions play a crucial role in shaping academic and professional careers, and admission decisions often rely on multiple factors such as standardized test scores, academic performance, and research experience. Understanding how these factors influence admission chances can help prospective applicants make informed decisions. This study analyzes a dataset created for predicting graduate admissions in India (Kaggle, 2016), includes variables such as GRE and TOEFL scores, CGPA, university rating, statement of purpose (SOP), letters of recommendation (LOR), and research experience. By leveraging multiple linear regression, this research aims to determine the most influential predictors of graduate admissions and develop a predictive model for better decision-making.

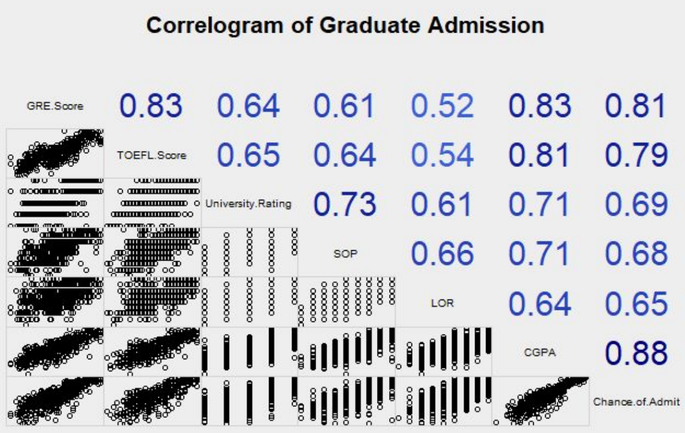
METHOD

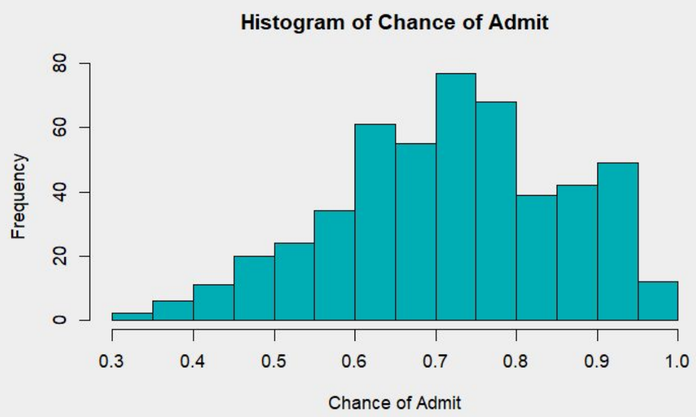
- **Data Collection:** Retrieve the graduate admission dataset from kaggle.
- **Data Preprocessing and EDA:** Check for missing value, variable correlation, and explore variable distribution using visualization.
- **Model Implementation**: Apply multiple linear regression to predict admission chances based on given features.
- **Model Evaluation**: Assess model performance using adjusted R^2 and p-value to ensure reliability.

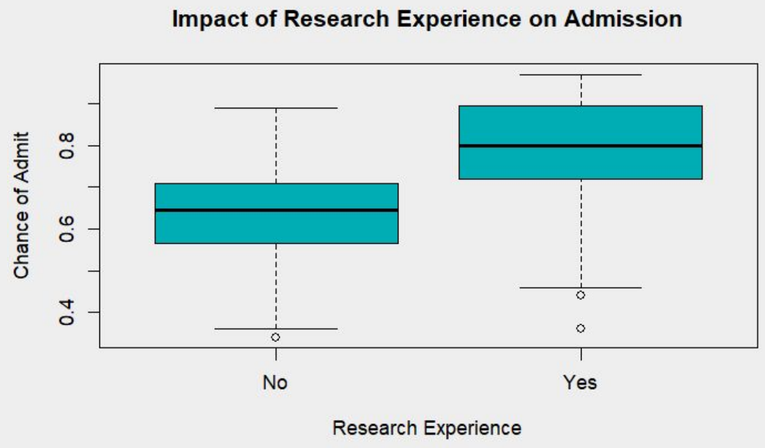
OBJECTIVES

- To study correlations among key factors to understand their influence
- To develop predictive model using multiple linear regression to estimate the chances of graduate admission based on several key factors
- To analyze the impact of individual predictors on admission chances

RESULTS & DISCUSSIONS





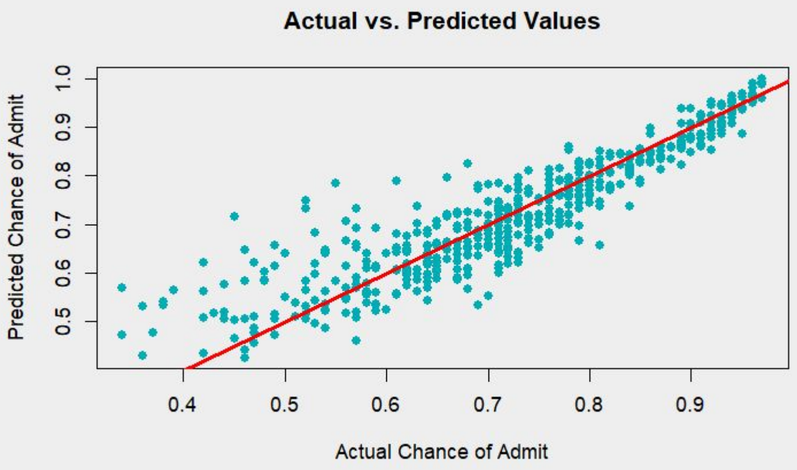


	Estimate	Significance Level
(Intercept)	**-1.2757251**	***
GRE.Score(x1)	**0.0018585**	***
TOEFL.Score(x2)	**0.0027780**	**
University.Rating(x3)	**0.0059414**	
SOP(x4)	**0.0015861**	
LOR(x5)	**0.0168587**	***
CGPA(x6)	**0.1183851**	***
Research1(x7)	**0.0243075**	***

Signif. codes: 0 '*' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1**

Adjusted R^2 : 0.8194
P-value : < 2.2e-16

- ★ The correlogram shows that **CGPA has the strongest correlation (0.88)** with the chance of admit, making it the most critical factor in graduate admissions.
- ★ The histograms shows **slightly right-skewed distribution**, meaning that most applicants have moderate chance of admission.
- ★ The boxplots show that applicants with **research experience generally have a higher chance** of admission compared to those without, as indicated by the higher median and overall distribution.



- ★ **GRE, LOR, CGPA, Research experience, and TOEFL are all significant** while university ranking and SOP has lower significance.
- ★ Adjusted R^2 shows that the regression model **explains 81.94% of the variance** in admission.
- ★ Extremely small p-value suggest that the overall model is **highly statistically significant**

Chance.of.Admit = -1.27573 + 0.00186x1 + 0.00278x2 + 0.00594x3 + 0.00159x4 + 0.01686x5 + 0.11839x6 + 0.02431x7

CONCLUSION

This study demonstrates the effectiveness of linear regression in predicting graduate admissions based on several key factors. The model achieved a **high adjusted R^2** score indicating strong relationship between chances of admission and predictors such as CGPA, research experience, and letter of recommendation strength. Among these factors, **CGPA has the highest impact** on the chance of admission.

REFERENCES

- Kaggle. (2016). Graduate Admission 2. Retrieved from https://www.kaggle.com/datasets/mohansacharya/graduate-admissions
- Mohan S Acharya, Asfia Armaan, Aneeta S Antony : A Comparison of Regression Models for Prediction of Graduate Admissions, IEEE International Conference on Computational Intelligence in Data Science 2019