Journal of Mathematics and Statistics Studies
ISSN: 2709-4200
DOI: 10.32996/jmss
Journal Homepage: www.al-kindipublisher.com/index.php/jmss

JMSS
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

Predicting the Possibility of Student Admission into Graduate Admission by Regression Model: A Statistical Analysis

Ashiqul Haque Ahmed[1], Sabbir Ahmad[2], Md Abu Sayed[3] ✉ Malay sarkar[4], Eftekhar Hossain Ayon[5], Md Tuhin Mia[6], Tahera Koli[7] and Rumana Shahid[8]

[1]*Economics & Decision Science, University of South Dakota, South Dakota, USA*
[2]*Department of Mathematical Science University of South Dakota South Dakota, USA*
[3]*Department of Professional Security Studies, New Jersey City University, Jersey City, New Jersey, USA*
[4]*Department of Management Science and Quantitative Methods, Gannon University, USA*
[5]*Department of Computer & Info Science, Gannon University, Erie, Pennsylvania, USA*
[6]*School of Business, International American University, Los Angeles, California, USA*
[7]*Department of Computer & Info Science, Gannon University, Erie, Pennsylvania, USA*
[8]*Department of Management of Science and Quantitative Methods, Gannon University, USA*

Corresponding Author: Md Abu Sayed, **E-mail**: msayed@njcu.edu

| ABSTRACT

This study aims to alleviate the uncertainties faced by prospective students during the application process by developing a predictive model for admission probabilities based on CGPA and GRE scores. The research investigates the significance of these predictor variables about the response variable, "Chance of Admit." Employing linear regression analysis, the model is thoroughly examined to evaluate its adequacy, predictive accuracy, and the need for interaction terms. The findings indicate that both CGPA and GRE scores play a crucial role in forecasting admission chances, with an adjusted R2 value of 0.0835, suggesting an 80% reduction in variance around the regression compared to the main line. The diagnostic plot of the model confirms its precision, revealing minimal deviations from linearity and normality in residuals. Furthermore, the study addresses concerns about multicollinearity using the Variable Inflation Factor (VIF) and finds no significant correlation between GRE Scores and CGPA. In summary, this research presents a robust predictive model for student admission probabilities, offering valuable insights for both prospective applicants and educational institutions.

| KEYWORDS

Linear Regression, Predictive Model, Variable Inflation Factor, Adjusted R^2

| ARTICLE INFORMATION

ACCEPTED: 02 November 2023 **PUBLISHED:** 27 November 2023 **DOI:** 10.32996/jmss.2023.4.4.10

1. Introduction

In the realm of higher education, the journey towards pursuing advanced degrees often involves a pivotal decision-making process for prospective students. The uncertainty surrounding the admission criteria and the competitiveness of educational programs can be a source of anxiety for applicants. To address this challenge, our research focuses on developing a robust predictive model to forecast the likelihood of a student gaining admission to a graduate program. The primary predictors considered in our model are the Cumulative Grade Point Average (CGPA) and Graduate Record Examination (GRE) scores, two pivotal factors widely assessed by universities.

This study aims to contribute a valuable tool for both prospective students and educational institutions, offering insights into the intricate dynamics of the admission process. Through the lens of statistical analysis and regression modeling, we delve into the significance of CGPA and GRE scores as predictors for admission chances. The response variable, termed "Chance of Admit," serves as the focal point for our predictive model.

Our research builds upon the existing body of literature, incorporating elements of machine learning models previously explored by other scholars. While acknowledging the diverse methodologies and datasets used in previous studies, our approach stands out by prioritizing a sustainable and universally applicable statistical model. Unlike some existing works that lack a sustainable model, our research aims to provide a versatile and reliable predictive tool.

The predictive model is constructed using a linear regression framework, with meticulous attention given to the assessment of fitness, interaction terms, and diagnostic plots. We examine the adjusted R2 value, a measure of the model's explanatory power, indicating approximately 80% less variance around the regression than a mean line (Khan and Kayyum, 2023,2021). Furthermore, we explore the necessity of interaction terms through statistical analysis, ensuring the inclusion of influential factors in our model.

In this introduction, we have provided a glimpse into the core objectives and methodologies of our project. Subsequent sections will delve into the model's specifications, the significance of predictor variables, fitness evaluation, predictive results, and more. Ultimately, our model aims to empower students with valuable insights into their admission prospects, helping them make informed decisions about their academic journeys.

2. Literature Review

Acharya et al. and Dawes et al. (2019, 1971) evaluate various regression algorithms, including Linear Regression, Support Vector Regression, Decision Trees, and Random Forest, based on student profiles. Subsequently, we calculate error metrics for each model, enabling a comparative analysis of their performance. The outcomes assist in determining whether the selected university is considered ambitious or a safer choice. They also present findings on the perspectives of potential students concerning the significance of identical admission criteria—an influential stakeholder group that holds substantial importance yet often lacks empowerment in the admissions procedure. We observe a notable consensus between students and faculty regarding the importance attributed to recommendation letters, undergraduate academic performance in math or physics, and standardized exam scores (GRE). Conversely, students assign greater significance to certain criteria, such as personal statements, previous research experiences, publications, and familiarity with the department, compared to faculty evaluations. An apparent "overemphasis" on certain criteria could adversely impact students' decision-making in the admissions context, potentially diminishing their chances of success. Consequently, these results underscore the imperative of incorporating students' viewpoints in the admissions process. The study evaluates regression algorithms based on student profiles and discusses the significance of admission criteria. However, the specific criteria considered and their importance may vary across different academic settings.

Wilson et al. (2019) investigated the impact of the Graduate Record Examination (GRE) on applicant demographics at the University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences using two review models: a metrics-based approach and a holistic approach. We assessed the correlation between the GRE scores of doctoral applicants and evaluations by the admissions committee. The metrics-based review method resulted in the exclusion of twice as many applicants who identified as historically underrepresented minorities compared to their counterparts. This study examines the impact of the GRE on applicant demographics using specific review models. The limitation could be the context-specific nature of GRE influence, which may differ in other academic institutions. While the study achieves high accuracy with the Logistic Regression classifier, the performance may be context dependent. The dataset and characteristics considered might not be universally applicable.

Jeganathan et al. (2021) utilized various classification methods, including Logistic Regression, KNN Classification, Support Vector Classification, Naive Bayes Classification, Decision Tree Classification, and Random Forest Classification, on a provided academic admission dataset. Upon evaluating the accuracy and mean absolute error of each model, the Logistic Regression classifier demonstrated superior performance, achieving an accuracy rate of 99%. The exploration of machine-learning regression models for university categorization and employability prediction is valuable. However, the effectiveness of the selected model may vary based on the characteristics and dynamics of different educational institutions.

Krishna et al.'s (2023) article explores a range of machine-learning regression models, including gradient boosting regression, support vector regression, random forest regression, decision tree regression, and ridge regression. The objective is to identify the most effective model, which will be employed to determine whether the university being considered by MS aspirants is categorized as ambitious or safe. Additionally, the selected model is used to predict the likelihood of students' employability in their academic placements. The paper also discusses the utilization of Streamlit, an open-source app framework, for creating a user-friendly web application interface that incorporates the chosen best-performing model.

Chakrabarty et al. (2020) assessed a student's academic accomplishments and university rating and provided the likelihood of admission to that university as an output. The implementation of the gradient-boosting regressor model yielded a $R2$-score of 0.84, ultimately surpassing the performance of the current leading model. Beyond the $R2$-score, various performance error metrics, including mean absolute error, mean square error, and root mean square error, are calculated and presented. Like other studies, the limitation lies in the specific context of the research. The effectiveness of the gradient-boosting regressor model with an $R2$-score of 0.84 may not be universally applicable.

All the following work focused on specific datasets and models, and nobody proposed any sustainable model of the statistical model. Our work fully focused on the model which is performing better for different datasets.

3. Methodology

3.1 Data Collection and Processing

Drawing from a dataset available on Kaggle titled "Graduate Admission 2", we have gathered the necessary data to construct the predictive model. Through analysis and mathematical formulations, we have established the foundation for our model's functionality. GRE score and CGPA are the vital elements that universities consider while deciding on a student's admission; we have selected GRE scores and CGPA as our predictor variables and the response variable is the Chance of Admit. Our approach utilizes a linear regression model, with a mathematical formula Y= F(x)+ ϵ, [12] where Y represents the chance of admission, F(x) denotes the function dependent on the predictor variables (GRE Scores & CGPA), and epsilon (represents the error term. Our model takes the form:

Y= -1.668 + .0030*GRE Score + .1651*CGPA

The intercept term, -1.668, is augmented by the effects of GRE Score and CGPA, each represented by their respective coefficients, to predict the chance of admission.

We have split data into training data (320 observations) and test data (80 observations). Our main objective is to reduce the test error rate. If the two variables that we have selected against the response variable are significant enough, we will make a prediction based on the training model, and then we will evaluate the test model. From the test model summary, we can see that the R^2 is .8047042, which is the same as the training model summary. This is the amount of variation in the data accounted for by the model. So, 80% of the data is accounted for by the model, which is the signal, and the rest 20% is random, and that is the noise.

3.2 Model Specification

The Summary for the Training model is listed below.

Table 01: Residuals (Admission Chance ~ Training Data)

Min	1Q	Median	3Q	Max
-0.245610	-0.025709	0.004788	0.039830	0.140026

Table 02: Coefficients Admission Chance ~ Training Data)

	Estimate	Standard Error	t-value	Pr. (>|t|)
Intercept	-1.6687448	0.1114161	-14.978	$< 2e^{-16}$
GRE Score	0.0030703	0.0005706	5.381	$1.44e^{-07}$
CGPA	0.1651157	0.0113219	14.584	$< 2e^{-16}$

To assess the connection between the predictor variables, GRE Score and CGPA, and their influence on predicting admission chances, we examine the F-statistic, which stands at 653.1, significantly surpassing 1. The p-values provide further insights into the predictors' significance, with a remarkably low p-value of 1.44e-07 for the GRE Score and an even smaller p-value of 2.2e-16 for the CGPA. A lower p-value indicates greater predictor significance, and in this case, both predictors are deemed highly significant based on the training model summary.

To gauge the model's fitness, we turn to the R2 value, ranging from 0 to 1. An R2 of 1 suggests perfect alignment between predictions and data, while an R2 close to 0 implies a weak linear relation and a feeble correlation. In our training model, the R2 value is 0.8047, signaling that approximately 80% less variance exists around the regression than the mean line. This implies that around 80% of the data aligns with the regression model, indicating a strong and effective fit. Overall, the statistical analysis reinforces the robust relationship between GRE Score, CGPA, and the prediction of admission chances in our model.

3.3 Predictive Results

In our interaction model, we can see the mean chance of admission is -2.3536 + .0053*GRE score + 0.2461*CGPA – 0.0002*GRE score*CGPA.

Table 03: Coefficients Table 05: Coefficients (Admission Chance ~ GRE Score * CGPA)

Min	1Q	Median	3Q	Max
-0.294889	-0.027517	0.008542	0.041708	0.153239

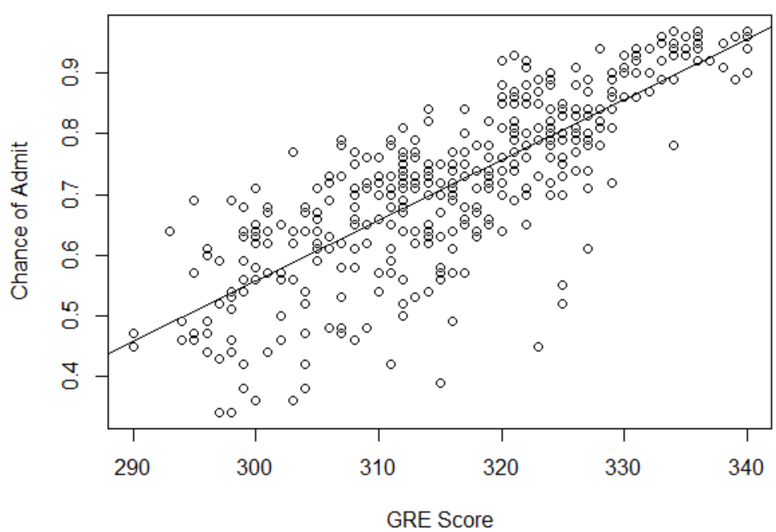
Table 04: Coefficients (Admission Chance ~ GRE Score * CGPA)

	Estimate	Standard Error	t-value	Pr. (>|t|)
Intercept	-2.3536850	1.2161115	-1.935	0.0537
GRE Score	0.0053695	0.0038522	1.394	0.1641
CGPA	0.24610099	0.1421528	1.731	0.0842
GRE Score: CGPA	-0.0002708	0.0004457	-0.608	0.5438

From F-statistics 471.5, which is much higher than 1, we can say the predictor variables CGPA, GRE score, and CGPA*GRE Score have a strong relationship with our response variable of a student's chance of admission. However, our main objective is to find out whether we need to establish an interaction term in our model, and from the p-value of .5438 of interaction between GRE score and CGPA, which is not statistically significant, we should not use an interaction term. Here, in this model, we can see the main effects of the GRE Score (.005369) and CGPA(.2461). We can also observe the interaction effect, which is -0.00027.

3.4 Evaluation Matrix

The adjusted R^2 [13] for our training model (summary depicted above) is .8035, and it shows the correlation between the predictors is good enough. For our coefficients, we can see that the GRE score has a standard error of .0005, and the CGPA has a standard error of .0113, which is small, and the magnitude is also smaller. So, the range is very narrow, and there should be a lot of stability. Here, we are trying to predict the admission rate given the GRE Score. The regression line is predicted or fitted Y values given X, which means as per the GRE Score the Chance of Admit is shown in the regression line. The residuals are the difference between the predicted Y value and fitted Y value (Chance of Admit). The plot for this predictor vs response variable is given below:



In the summary statistics below, the estimate for the model slope is .0099759, and the intercept is -2.4360842. The F statistics 720.6, which is far larger than 1, indicates that there is a strong relationship between the predictor and response variable. R^2 is .6442. but it is not bad as well. The residual standard error is .08517, which is the average or typical size of errors or residuals. The summary statistics for this model created above are given below:

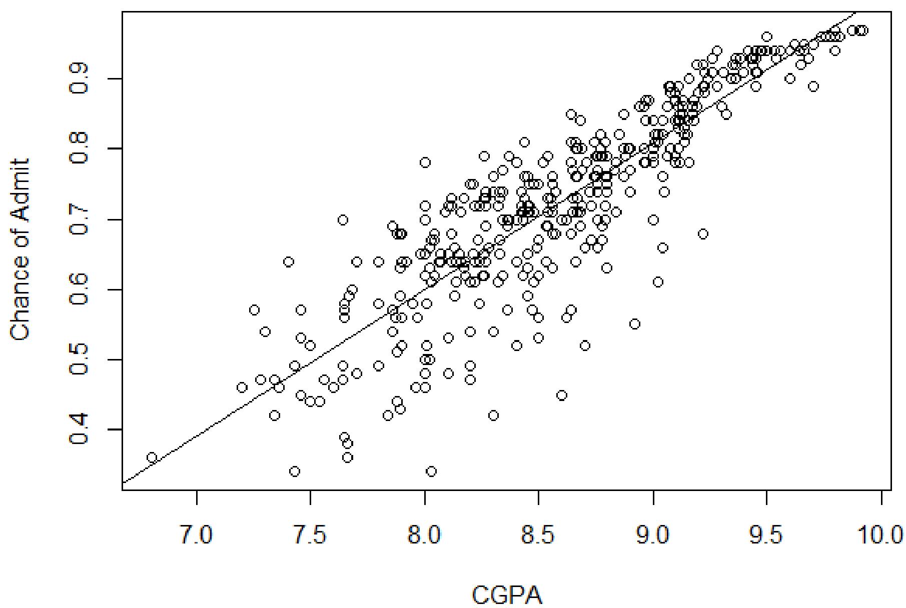
Table 05: Coefficients (Admission Chance ~ GRE Score)

Min	1Q	Median	3Q	Max
-0.33613	-0.04604	0.00408	0.05644	0.18339

Table 06: Coefficients (Admission Chance ~ GRE Score)

	Estimate	Standard Error	t-value	Pr. (>|t|)
Intercept	-2.4360842	0.1178141	-20.68	$< 2e^{-16}$
GRE Score	0.0099759	0.0003716	26.84	$<2e^{-16}$

Here, we are trying to predict the admission rate given CGPA. The regression line predicts the fitted model of the Chance of Admission given the CGPA. The plot for this predictor vs response variable is given below:



The summary statistics below the estimates for model slope(coefficient) is .20885, and the intercept is -1.07151. The F statistics 1279, which is far larger than 1, indicates that there is a strong relationship between predictor and response variables. R^2 is .7626, which captures this much variance within the model. The residuals are the difference between predicted and fitted Chance of Admit values. The residual standard error is .06957, which is the average or typical size of errors or residuals. The summary statistics for CGPA in predicting the admission score are given below.

Table 07: Coefficients (Admission Chance ~ CGPA)

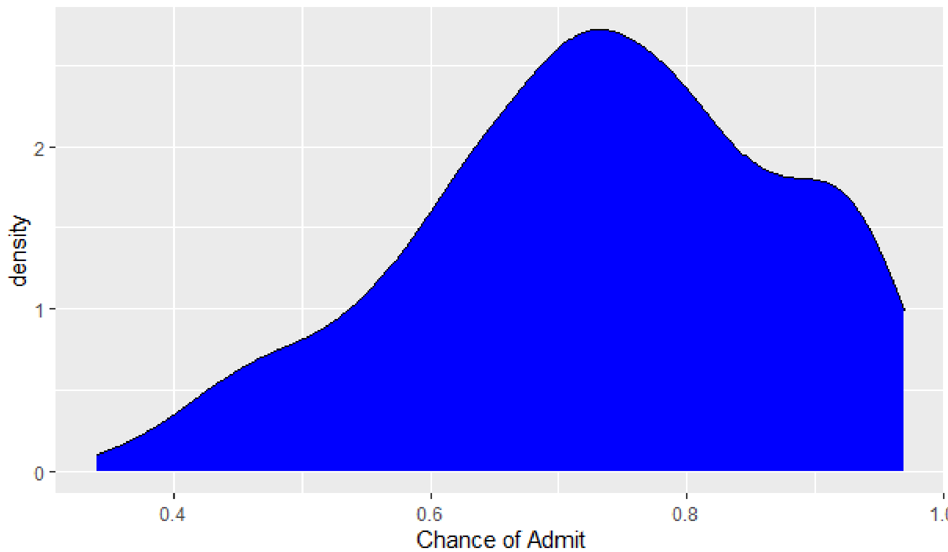
Min	1Q	Median	3Q	Max
-0.274575	-0.030084	0.009443	0.041954	0.180734

Table 08: Coefficients (Admission Chance ~ CGPA)

	Estimate	Standard Error	t-value	Pr. (>|t|)
Intercept	-1.07151	0.05034	-21.29	$< 2e^{-16}$
GRE Score	0.20885	0.00584	35.76	$< 2e^{-16}$

3.5 Exploring the Response Variable

We will explore the distribution of our response variable Chance to Admit, and from the graph, we can see that our response variable distribution is a bit skewed to the right, which is due to some of the outliers, e.g., points 125, 192, 231. The attached picture is below for reference.

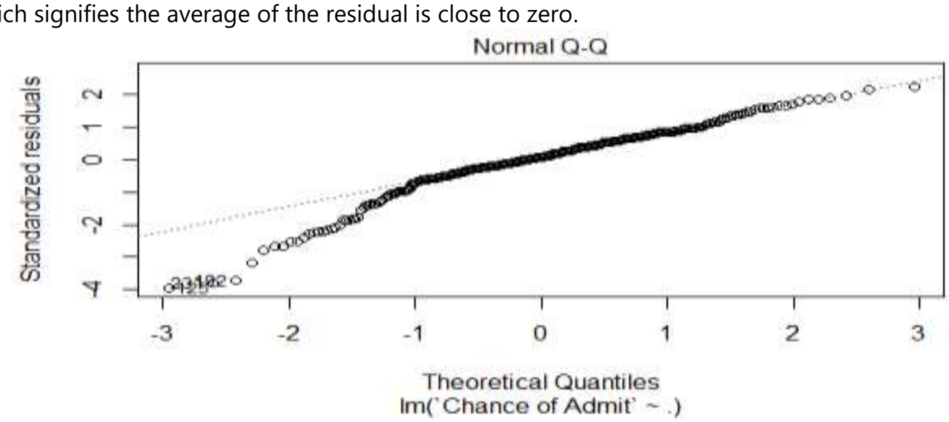


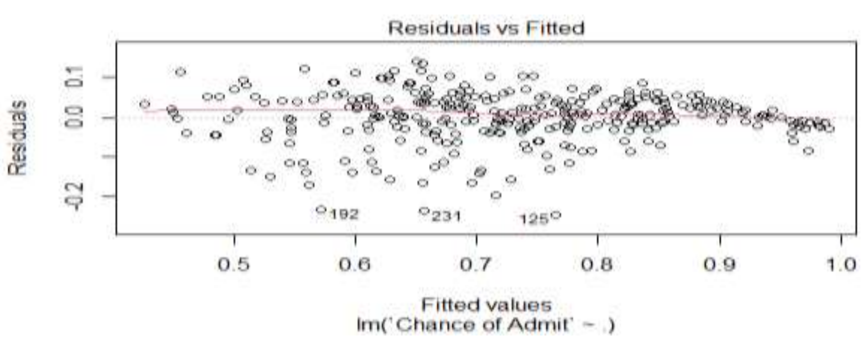
We are predicting the Chance of Students' Admission based on their CGPA and GRE scores. We will be splitting them into two sets. Training data set and Test Data set. Among a total of 400 observations, for training data set evaluation, we have selected 320 observations, and for the Test Data set, we have selected 80 observations as we have chosen to split the data based on 80-20.

4. Result

4.1 Diagnostic Plots of the Final Model

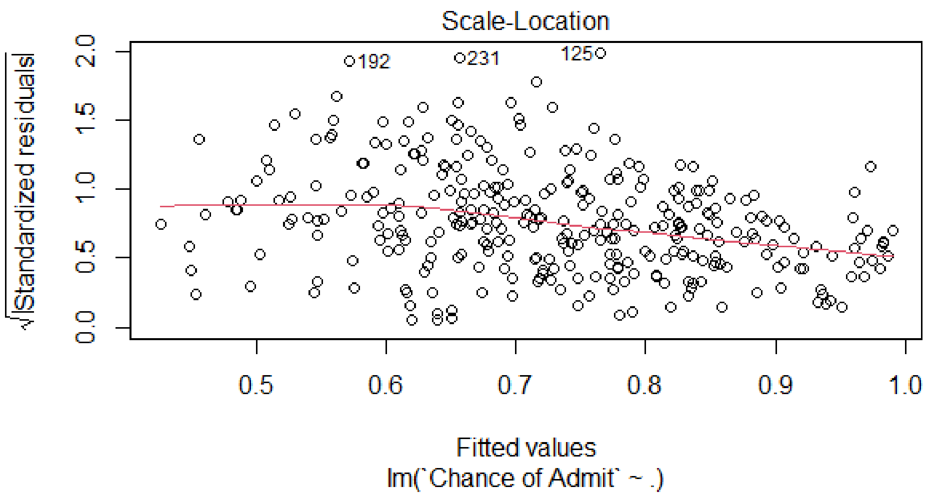
Here, the X-axis represents the predicted Y values; in this case, it represents the response variable, which is the chance of admission based on GRE score and CGPA. On the other hand, the Y-axis represents residuals or errors. If the linearity assumptions are met, we should see no pattern here, and the red line should be flat. From the plot above, we can see the red line is almost 0 except at the beginning, which signifies the average of the residual is close to zero.





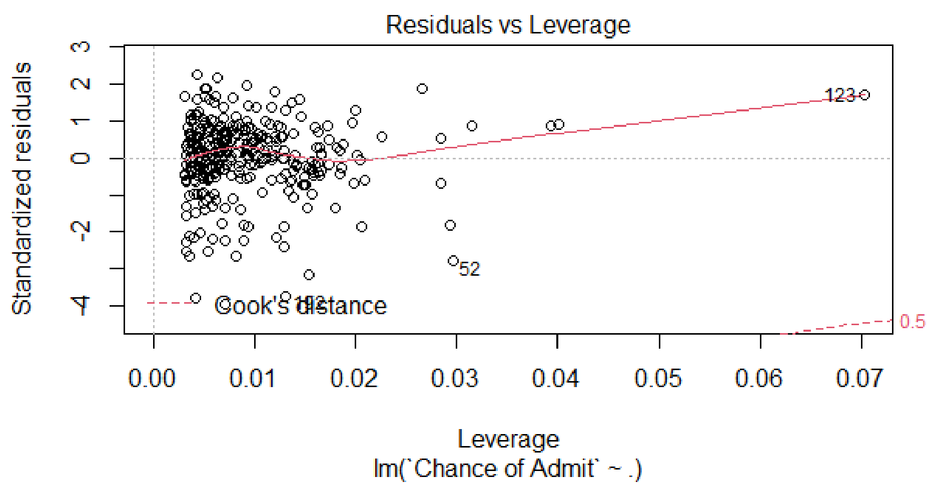
4.2 Quantile-Quantile Plot

The Q-Q plot demonstrates whether the residuals are normally distributed. The Y-axis represents the ordered Standard residuals, and the X-axis represents theoretical residuals, which are a student's chance to admit. If the error points or the residuals are normally distributed, the points should fit on the line. From the plot above, we can see that there are some deviations at the beginning or bottom left corner with some extreme points and the right top corner of the line. Most of the points almost fit the line, and by the bulk of the data, the model fits well.



4.3 Scale Location

For the scale location model, the X-axis represents the response variable, which is the fitted value. In Y, the actual value of the residuals is taken, and then they are scaled by the estimated error in the model and then the square root values of those residuals are taken. It shows how the residuals are spread and whether the variance of these residuals has equal variance or not. The average residuals should fall around 1 for a perfect model. Here in our scale location model, the residuals are almost evenly spread at the beginning, which has closer values to 1 and at the end, we can observe some deviation as the line leans downwards. Also, we have got some outliers at the top.



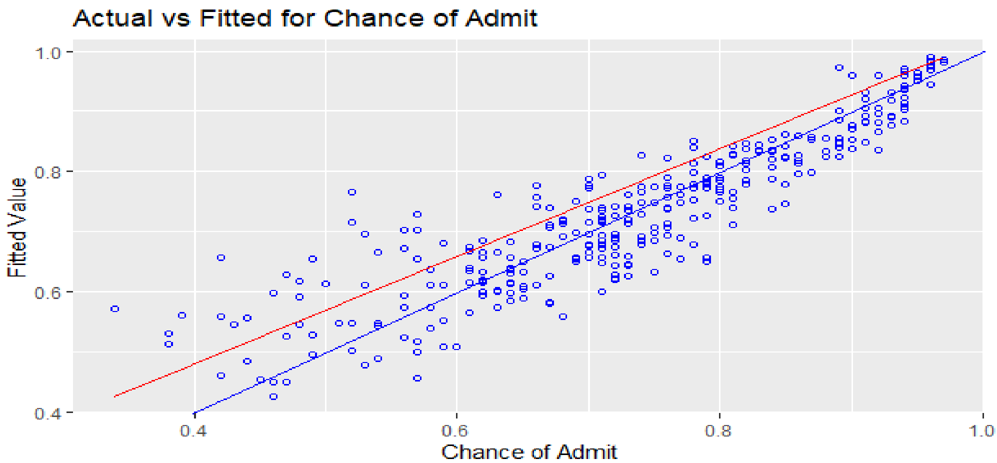
4.4 Residuals Vs Leverage

In the previously presented residuals leverage plot, the Y-axis represents the standardized residuals, not their absolute values, resulting in a mix of positive and negative values. On the X-axis, we find leverage, a measure of how much an individual observation influences the model fit. Observations with low leverage have minimal impact on model fitness, while those with high leverage, such as points 52 and 123, can significantly alter the model fit. For instance, removing these high-leverage points would lead to a substantial change in the model's overall fit. Point 123, highlighted in the plot, exhibits both high leverage and a large residual, indicating a poor fit by the model and a substantial impact on the overall result.

4.5 Variable Inflation Factor

The Variable Inflation Factor (VIF) is employed in our regression model to assess multicollinearity between variables, which refers to the correlation among variables adversely impacting regression estimates. While various opinions exist on the standard VIF result, a measurement below 5 is considered acceptable. In our established model, the VIF for predictor variables GRE Score and CGPA is 3.490459, indicating satisfactory levels of multicollinearity. This suggests that there is no significant risk of excessive correlation between the predictor variables nullifying each other.

The accompanying plot features a red line representing the fitted line without any modeling and a blue line depicting the fitted line based on our created model. Notably, the blue line lies below the fitted red line, and most observations cluster around it. This alignment indicates the effectiveness of our model, supporting the expectation that it is sufficiently accurate in its predictions.



5. Conclusion and Discussion

In conclusion, this developed model empowers both prospective students and educational institutions by offering a reliable tool for predicting graduate admission chances. By addressing uncertainties, this research facilitates informed decision-making in higher education's complex and competitive landscape. This research contributes to the field by providing a comprehensive analysis of a predictive model for graduate admission chances. The emphasis on statistical rigor, interaction effects, and diagnostic plots adds value to the existing literature. The study aimed to address the uncertainty faced by prospective graduate students by developing a predictive model for admission chances based on CGPA and GRE scores [16,17]. The linear regression model [18,19] was employed, and the results indicated that both CGPA and GRE scores significantly influence the prediction of admission chances.

The model's fitness was rigorously evaluated, considering the adjusted R2 value, diagnostic plots, and other statistical metrics. The adjusted R2 value of 0.0835 suggests that the model explains approximately 80% of the variance in the chance of admission, indicating a strong correlation. The developed model has practical implications for prospective graduate students, providing insights into their admission prospects based on CGPA and GRE scores. Educational institutions can also benefit from understanding the significance of these predictor variables in the admission process. This research contributes to the field by providing a comprehensive analysis of a predictive model for graduate admission chances. The emphasis on statistical rigor, interaction effects, and diagnostic plots adds value to the existing literature.

While the model demonstrated robustness, it is essential to acknowledge limitations such as dataset specificity. Future research could involve testing the model on diverse datasets and exploring additional predictors to enhance predictive accuracy.

Funding: This research received no external funding.
Conflicts of Interest: The authors declare no conflict of interest.
Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

References

[1] Acharya M. S., Armaan A. and Antony A. S., (2019) A Comparison of Regression Models for Prediction of Graduate Admissions, *2019 International Conference on Computational Intelligence in Data Science (ICCIDS)*, Chennai, India, 2019, pp. 1-5, doi: 10.1109/ICCIDS.2019.8862140.

[2] Asiah, M., Zulkarnaen, K. N., Safaai, D., Hafzan, M. Y. N. N., Saberi, M. M., & Syuhaida, S. S. (2019). A review on predictive modeling technique for student academic performance monitoring. In *MATEC Web of Conferences* (Vol. 255, p. 03004). EDP Sciences.

[3] Chakrabarty, N., Chowdhury, S., & Rana, S. (2020). A statistical approach to graduate admissions' chance prediction. In *Innovations in Computer Science and Engineering: Proceedings of 7th ICICSE* (333-340). Singapore: Springer Singapore.

[4] Dawes, R. M. (1971). A case study of graduate admissions: Application of three principles of human decision making. *American psychologist*, *26*(2), 180.

[5] Haque, M. S., Amin, M. S., & Miah, J. (2023). Retail demand forecasting: a comparative study for multivariate time series.. https://doi.org/10.21203/rs.3.rs-3280263/v1

[6] Haque, M. S., Amin, M. S., Miah, J., Cao, D. M., & Ahmed, A. H. (2023). Boosting Stock Price Prediction with Anticipated Macro Policy Changes. *Journal of Mathematics and Statistics Studies*, *4*(3), 29–34. https://doi.org/10.32996/jmss.2023.4.3.4

[7] Islam, M, N, Shah A A., Islam, M., Rakib R, Md A, Miah, J, Kayyum S, Shadaab, A . and Faisal, F A et al. (2020) An Empirical Study to Predict Myocardial Infarction Using K-Means and Hierarchical Clustering. International Conference on Machine Learning, Image Processing, Network Security and Data Sciences. Springer, Singapore, 2020.

[8] Jeganathan S, Parthasarathy S, Lakshminarayanan A. R., Ashok K P. M. and Khan M. K. A., (2021) Predicting the Post Graduate Admissions using Classification Techniques, 2021 International Conference on Emerging Smart Computing and Informatics (ESCI), Pune, India, 2021, 346-350, doi: 10.1109/ESCI50559.2021.9396815.

[9] Khan R. H., Miah J, Abed N S. A. and Islam M, (2023) A Comparative Study of Machine Learning classifiers to Analyze the Precision of Myocardial Infarction prediction, 2023 IEEE 13th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA. 0949-0954, Doi: 10.1109/CCWC57344.2023.10099059.

[10] Khan R.H., Miah J, Tayaba M, Rahman M.M. (2023) A Comparative Study of Machine Learning Algorithms for Detecting Breast Cancer, 2023 IEEE 13th Annual Computing and Communication Workshop and Conference (CCWC), USA.

[11] Kayyum S et al., (2020) Data Analysis on Myocardial Infarction with the help of Machine Learning Algorithms considering Distinctive or Non-Distinctive Features, 2020 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, 2020, pp. 1-7, Doi: 10.1109/ICCCI48352.2020.9104104.

[12] Krishna K G. S., Prithvi J, Divya M. and Kumari C. U., (2023) Predicting Employability and Admission for MS Students using ML Regression Models, 2023 IEEE 8th International Conference for Convergence in Technology (I2CT), Lonavla, India, 2023, 1-8, doi: 10.1109/I2CT57861.2023.10126208.

[13] Khan R. H., Miah J, Rahman M. M., Hasan M. M. and Mamun M., (2023). A study of forecasting stocks price by using deep Reinforcement Learning, 2023 IEEE World AI IoT Congress (AIIoT), Seattle, WA, USA, 2023, 0250-0255, doi: 10.1109/AIIoT58121.2023.10174358.

[14] Kayyum S. et al., (2020) Data Analysis on Myocardial Infarction with the help of Machine Learning Algorithms considering Distinctive or Non-Distinctive Features, 2020 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, 2020, pp. 1-7, doi: 10.1109/ICCCI48352.2020.9104104.

[15] Khan R. H., Miah J., Abed N S. A. and Islam M., (2023). A Comparative Study of Machine Learning Classifiers to analyze the precision of Myocardial Infarction prediction, 2023 IEEE 13th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA, 2023, 0949-0954, doi: 10.1109/CCWC57344.2023.10099059.

[16] Kumar, M., Singh, A. J., & Handa, D. (2017). Literature survey on student's performance prediction in education using data mining techniques. *International Journal of Education and Management Engineering*, *7*(6), 40-49.

[17] Ofori, F., Maina, E., & Gitonga, R. (2020). Using machine learning algorithms to predict students' performance and improve learning outcome: A literature-based review. *Journal of Information and Technology*, *4*(1), 33-55.

[18] Wilson, M. A., Odem, M. A., Walters, T., DePass, A. L., & Bean, A. J. (2019). A model for holistic review in graduate admissions that decouples the GRE from race, ethnicity, and gender. *CBE—Life Sciences Education*, *18*(1), ar7.

[19] Zohair, A., & Mahmoud, L. (2019). Prediction of Student's performance by modelling small dataset size. *International Journal of Educational Technology in Higher Education*, *16*(1), 1-18.