Jamboree Admission Prediction — Business Report

Executive Summary

This project aims to predict students' chances of admission to graduate programs using the Jamboree Education dataset. By applying Linear and Ridge Regression, the model identifies which academic and qualitative factors most strongly influence admission outcomes. Key insights reveal that CGPA is the dominant predictor, followed by GRE, TOEFL, LOR, and Research experience.

Business Objective

Assist students and education consultants to understand which factors most strongly influence graduate admission chances and how these factors interact. Predictive modeling translates data into actionable recommendations, enabling personalized guidance.

Dataset Overview

- **Dataset:** Jamboree Education Dataset (500 rows)
- **Target Variable:** Chance of Admit (0–1)
- **Features:** GRE, TOEFL, University Rating, SOP, LOR, CGPA, Research

Methodology

- **EDA(Exploratory Data Analysis)**
  - Understanding feature distribution and skewness to fit in model
  - Outlier Handling
- **Data Preprocessing**
  - Handled missing values appropriately.
  - Verify assumption of linearity
  - Standardization and Split data into train and test (80/20).
- **Model Building**
  - Developed baseline **Linear Regression** model.
  - Applied **Ridge and Lasso Regression** for regularization and coefficient stability.
  - Tuned hyperparameter alpha using cross-validation (RidgeCV, LassoCV).
- **Model Evaluation**
  - Metrics used:
    - **R^2 (Coefficient of Determination)**
    - **RMSE (Root Mean Squared Error)**
    - **MEA(Mean Absolute Error)**
    - **Residual Analysis**

- **Visual diagnostics:**
    - Actual vs Predicted plot
    - Residual distribution(Homoscadacity)
    - Q-Q plot for residuals

Feature Importance

Feature	Coefficient	Relative Influence (%)
CGPA	0.0417	49.6
GRE Score	0.0116	13.8
TOEFL Score	0.0096	11.5
LOR	0.0091	10.8
Research	0.0060	7.2
SOP	0.0039	4.7
University Rating	0.0021	2.5

❖ Interpretation:

CGPA is the single most important factor, followed by GRE and TOEFL.

Qualitative factors (LOR, SOP) and Research experience also add measurable value to admission probability.

Insights & Recommendations

• For Students

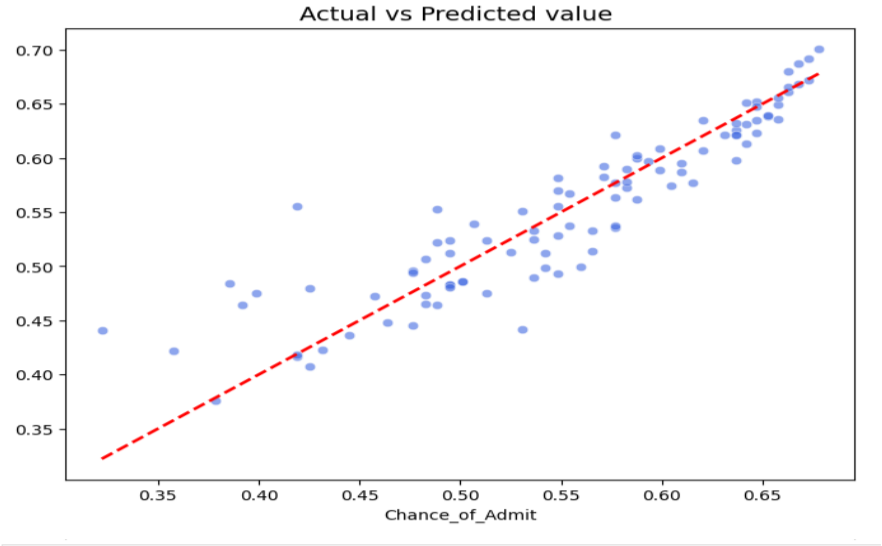
1. **Focus on Academic Excellence (CGPA ~50 %)**
    - The most influential factor — consistent academic performance significantly raises admission chances.
2. **Prepare Strategically for GRE & TOEFL**
    - Strong test scores complement academic performance together by ~**25%** and improve overall competitiveness.
3. **Invest in Strong LORs & Research Exposure**
    - Quality recommendations and practical research work enhance profile credibility.
    - Students with research experience have **16%** more chance of admission than inexperienced one
4. **Craft a Strong SOP**
    - While moderate in quantitative impact, a compelling SOP can differentiate similar applicants.
5. **University Rating Has Limited Effect**
    - Focus on your profile rather than perceived prestige alone as academic and test performance dominate the outcome.

- ## For Education Consultants / Jamboree

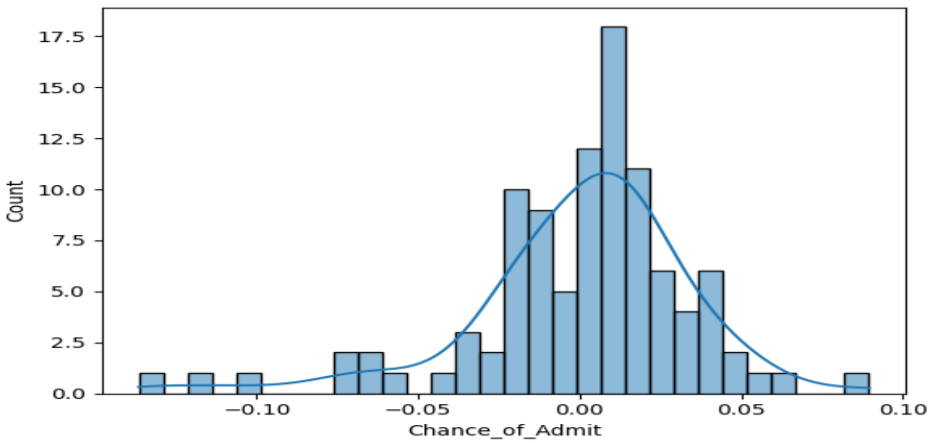
    - Prioritize training modules that improve academic performance and test preparation.
    - Offer personalized admission strategy sessions focusing on key predictors.
    - Integrate predictive modeling into student dashboards for real-time guidance.

Visualization Summary

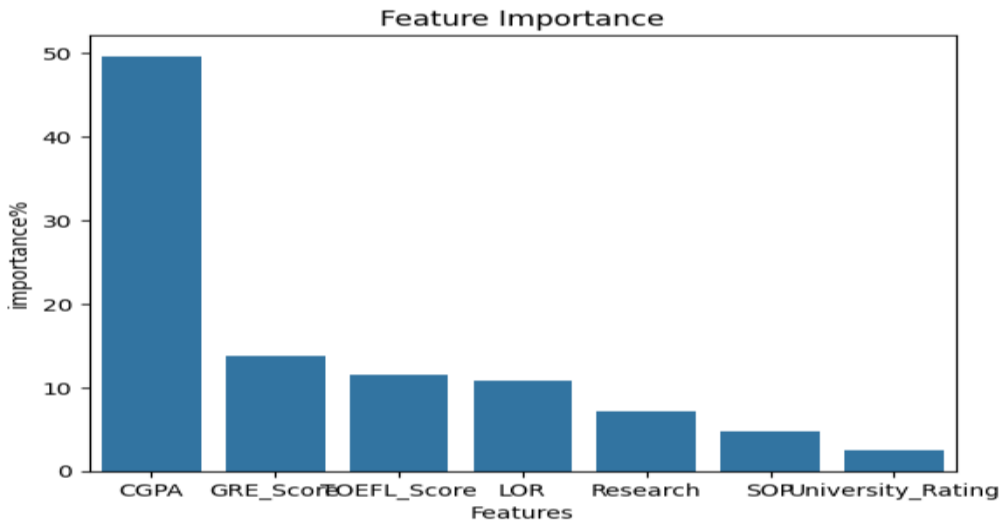
- **Actual vs Predicted Plot:** Most predictions lie close to the diagonal line, confirming a good model fit.



- **Residual Analysis:** Residuals are slightly left skewed showing models' overfiting or overestimating sometimes.



- **Feature Impact Chart:** Visualizes CGPA as the dominant driver, followed by GRE and TOEFL.



Conclusion

Academic consistency (CGPA) and test scores (GRE, TOEFL) drive admission success. Qualitative factors (LOR, SOP, Research) strengthen profiles but cannot substitute strong academics. This project demonstrates how analytics can support educational strategy and decision-making.

Future Enhancements

- Incorporate **non-linear models** (Random Forest, Gradient Boosting) for comparison.
- Add **SHAP/Feature Importance visualizations** for model explainability.

Project Summary

Author: Manish Nandi
Tools Used: Python, Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn
Techniques: Linear Regression, Ridge Regression, Model Evaluation
Environment: Jupyter Notebook
Type: Predictive Modeling & Business Analytics