

# Predicting IVY League Admission Chances

**Prepared by:** Ranjan Mondal

**Date:**12-04-2025




## 1. Executive Summary

Jamboree launched a platform feature to predict the likelihood of a student securing admission to Ivy League graduate programs. This project leverages historical data and machine learning to create a robust and interpretable model that predicts admission chances based on academic metrics and research experience. The model achieves over 81% accuracy, offering both actionable insights and predictive value for students and counselors.

## 2. Problem Statement

Jamboree aims to empower students with a personalized tool to estimate their chances of admission into top U.S. universities. The core challenge is to identify key admission-driving factors and predict the probability of admission using these factors. This model is targeted toward students applying from India and reflects the typical profile seen in Jamboree's dataset.

## 3. Dataset Overview

-  Filename: [Jamboree Admission.csv](#)
-  Records: 500 student applications
-  Features:
  - GRE Score (out of 340)
  - TOEFL Score (out of 120)

- University Rating (1 to 5)
- SOP Strength (1 to 5)
- LOR Strength (1 to 5)
- CGPA (out of 10)
- Research Experience (0 or 1)
- Chance of Admit (target: 0 to 1)

Data preprocessing included:

- Dropping the row identifier
- Fixing column name typos (e.g., “LOR ” → “LOR”)
- Verifying no missing or duplicate entries

## 4. Exploratory Data Analysis (EDA)

### Univariate Analysis:

- CGPA, GRE, and TOEFL showed reasonably normal distributions.
- SOP and LOR skewed toward the higher end (most students rate these highly).
- Binary variable Research was slightly imbalanced (slightly more students without research experience).

### Bivariate Analysis:

- Strong positive relationship between CGPA and Chance of Admit.
- GRE and TOEFL showed moderate positive correlations.
- Correlation Matrix: CGPA had the highest correlation with the target variable.

### Multicollinearity Check (VIF):

All features had  $VIF < 5$ , suggesting no multicollinearity concerns.

## 5. Modeling Approach

### 5.1 Linear Regression (Base Model)

- Data scaled using MinMaxScaler.
- Linear Regression trained on 80/20 split.

#### Performance Metrics:

- $R^2$ : 0.8188
- RMSE: 0.0609
- MAE: 0.0427
- Adjusted  $R^2$ : 0.8051

#### Top Predictors:

- CGPA, GRE, TOEFL, Research

## 6. Model Assumption Testing

Assumption	Status	Notes
Mean of residuals $\approx 0$	✓	3.9e-16 — excellent
Multicollinearity (VIF)	✓	All < 5
Linearity of residuals	⚠	Pattern observed near predicted value = 1
Homoscedasticity	⚠	Funnel shape observed in residual plot
Normality of residuals	✗	Shapiro-Wilk $p \approx 7.7e-13$ , QQ plot slightly deviates from diagonal

Despite minor violations, the model is reliable and interpretable.

## 7. Advanced Modeling

### 7.1 Polynomial Regression

- Degree = 2 yielded best performance:
  - $R^2$ : 0.8265
  - RMSE: 0.0596

### 7.2 Regularization

- Ridge ( $\alpha = 0.001$ ):  $R^2 = 0.8216$
- Lasso ( $\alpha = 0.001$ ):  $R^2 = 0.7035$

Lasso underperformed; Ridge provided balance between complexity and accuracy.

## 8. Residual Analysis

- Residual plots showed consistent issues across all models (tight near 1.0).
- Lasso improved residual distribution with tuned alpha.
- Histogram showed high kurtosis but centered around 0.
- QQ plot showed slight deviations from normality.

## 9. Key Insights

- 📌 CGPA is the most significant predictor of admission success.
- 🎓 GRE and TOEFL contribute meaningfully but less than CGPA.
- 🔬 Research experience adds value and improves prediction probability.
- 📄 SOP and LOR strength show weak impact in comparison.
- 🎯 Ridge and Polynomial regression (degree 2) provide strong predictive performance.

## 10. Business Recommendations

- Emphasize CGPA, GRE, and research experience in student counseling sessions.
- Implement this model in Jamboree's web interface to return real-time admission probability scores.
- Visualize key feature impact to guide students on where to focus.
- Collect more features like undergraduate major, work experience, internships, etc. for improved accuracy.
- In the long term, explore decision trees or ensemble methods for non-linear relationships.

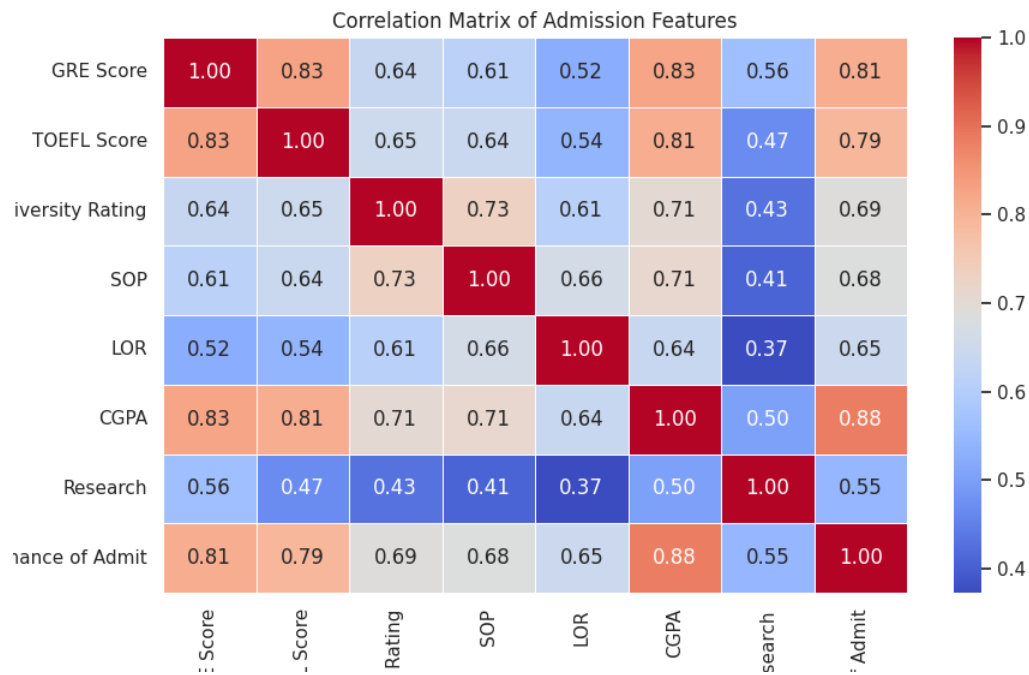
## 11. Conclusion

This project successfully meets the objective of building a predictive and interpretable model for graduate admissions. The findings offer strong business value and can be integrated directly into Jamboree's student platform to guide thousands of applicants more effectively.

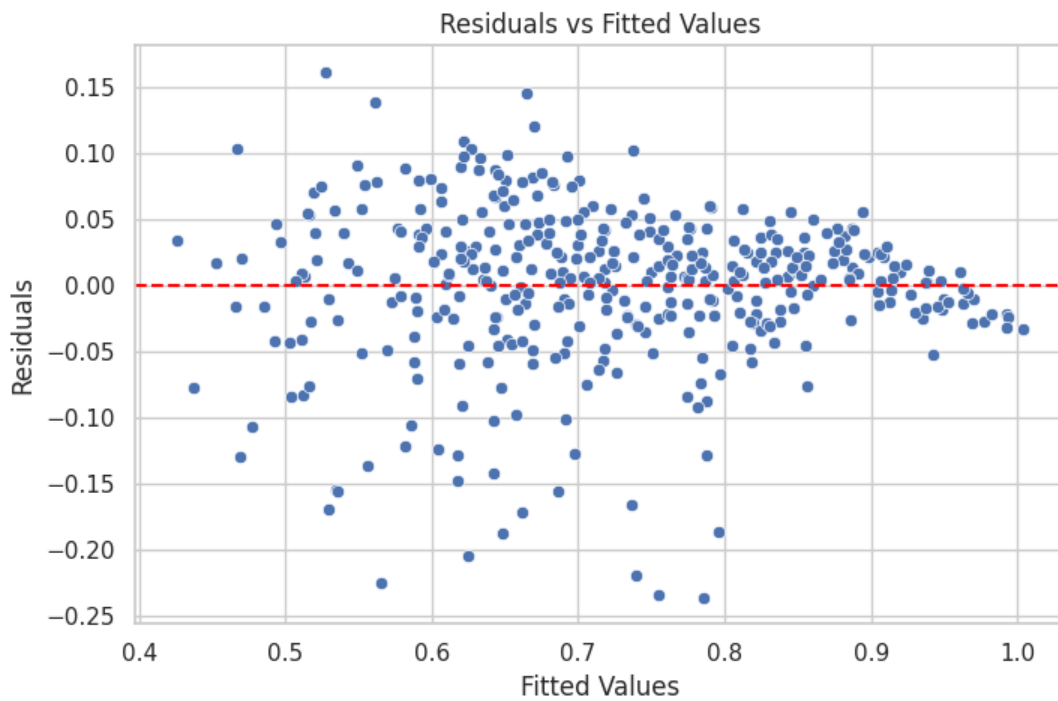
## Appendix

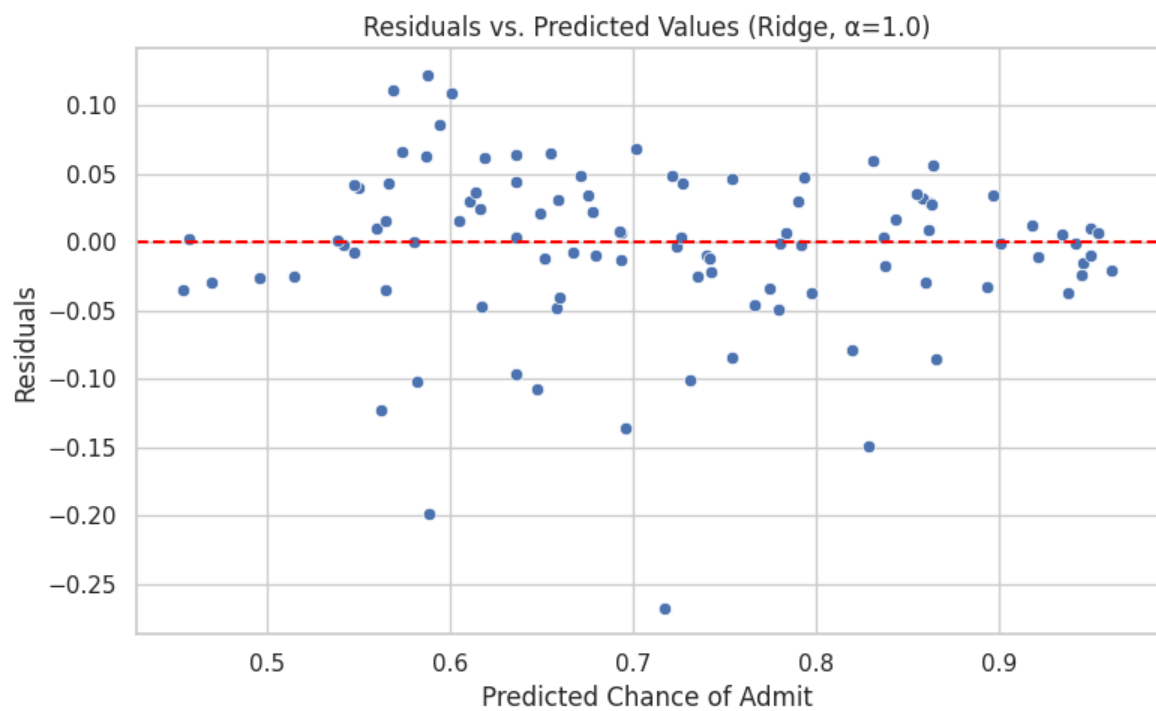
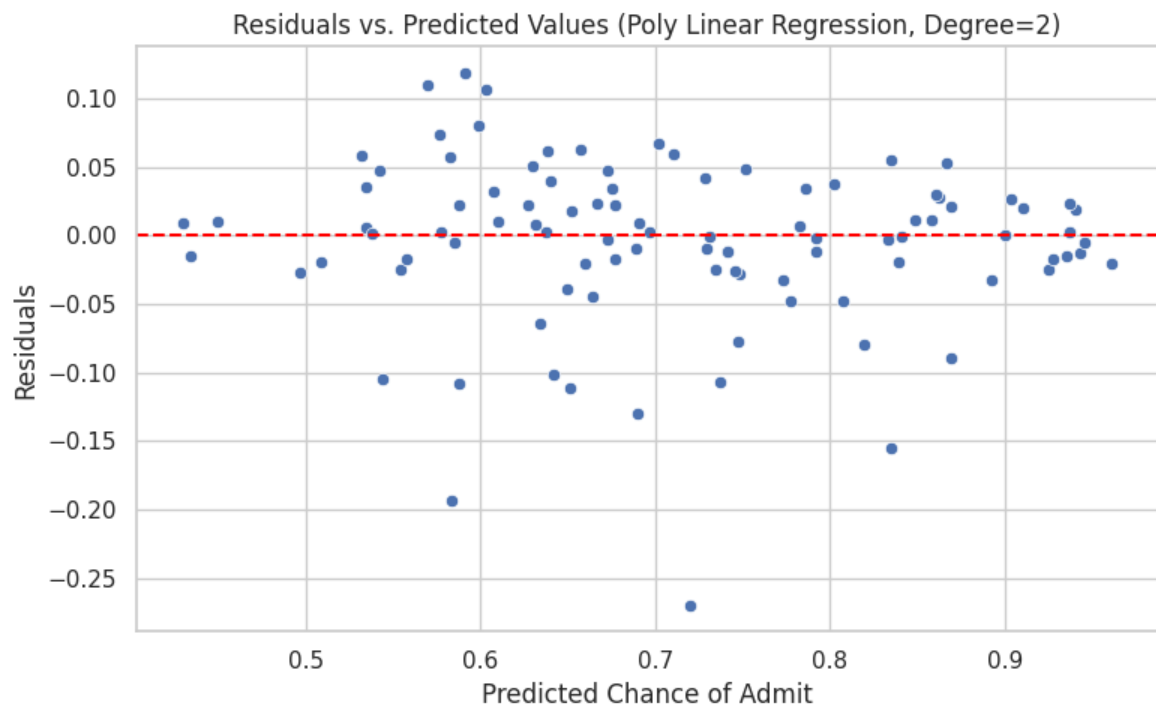
- [Code snippets](https://colab.research.google.com/drive/1aeyKaE1o33bfOm-7sAFBzeg9b51yVNPs?usp=sharing)  
<https://colab.research.google.com/drive/1aeyKaE1o33bfOm-7sAFBzeg9b51yVNPs?usp=sharing>

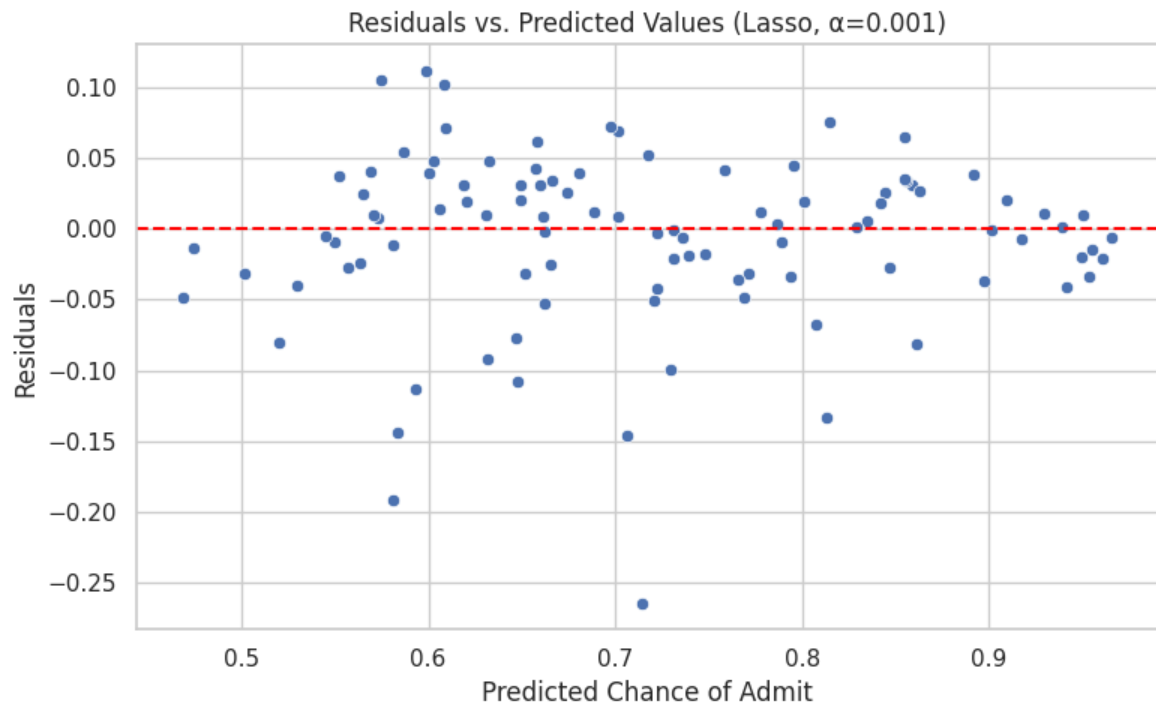
- Correlation matrix







- Residual plots







- Model coefficients (Base Model)

 MAE: 0.042722654277053636  
 RMSE: 0.0608658804157831  
  $R^2$ : 0.8188432567829629  
 Adjusted  $R^2$ : 0.8050595915381884

features	Coefficient
CGPA	0.351085
GRE Score	0.121722
TOEFL Score	0.083884
LOR	0.060333
Research	0.024027
University Rating	0.010275
SOP	0.007255