Admission Data Analysis

Overview

This analysis aims to explore the relationship between various admission parameters and the likelihood of admission to universities. The dataset used for this analysis is sourced from the `"Admission_Predict_Ver1.1.csv"` file.

Methadology

1. **Reading Data**:
   - The data was structured as a numpy array with named columns using `np.genfromtxt()`
2. **Linear Regression Model**:
   - I have assumed the admission chance depends linearly on the parameters
   - A linear regression model was used to predict the chance of admission based on various parameters.
   - The model was built using the least squares method by using the function `np.linalg.lsq()`
3. **Correlation Analysis**:
   - Correlation coefficients were calculated to determine the strength and direction of the relationships between each parameter and the chance of admission.
   - The function `np.corrcoef()` is used to find the correlation of the parameters.
   - The top 3 parameters with the highest correlation coefficients were identified.
4. **R-squared Value**:
   - The R-squared value, which represents the proportion of the variance in the dependent variable that is predictable from the independent variable(s), was calculated.
5. **Visualization**:
   - Three plots were generated to visualize the results:
     - A bar chart comparing the correlation coefficients of different parameters(for all universities).
     - A bar chart comparing the correlation coefficients of different parameters(for top universities).
     - A scatter plot comparing the actual and predicted chances of admission.

Analysis

1. **Linear vs Non-Linear Modelling**
   - I have used linear modelling because of simplicity and ease of use.
   - Also I have found that for increase in the powers of the parameters in the predicting model, the R-Squared value increases.
   - The R-Squared value for the linear modelling is `0.8219007395178417`
   - So I have concluded that the linear modelling is more appropriate for least square fitting method.
   - If the dependency is non-linear we can use machine learning alogorithms, polynomial regression, exponential models, or other specialized models (e.g., logistic regression for binary outcomes), may be more appropriate.

Main Conclusions

- **Function of parameters**: Approximately I have defined the admission chance as the linear function of given parameters:
  - GRE Score: 0.00185851
  - TOEFL Score: 0.00277797
  - University Rating: 0.00594137
  - SOP: 0.00158614
  - LOR: 0.01685874
  - CGPA: 0.11838505
  - Research: 0.02430748
  - Intercept: -1.27572508
- **R-squared Value**:

- The R-squared value indicates that the linear regression model explains a substantial portion of the variance in the chance of admission.
- For one of the run, `R-Squared value = 0.8219007395178417`
- More closer the value of R-Squared to 1 more accurate is the plot.The plot otained between predisted admisssion chance vs given chance is:

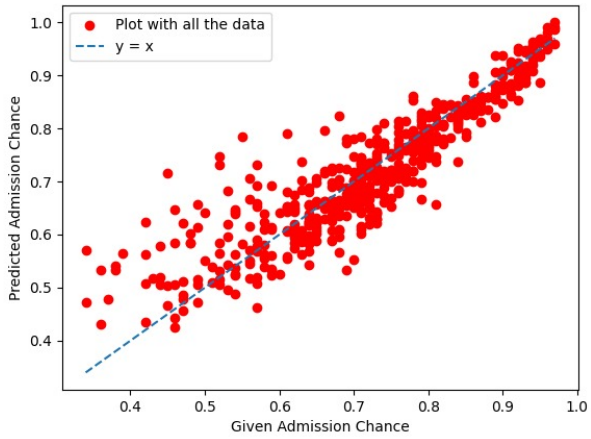
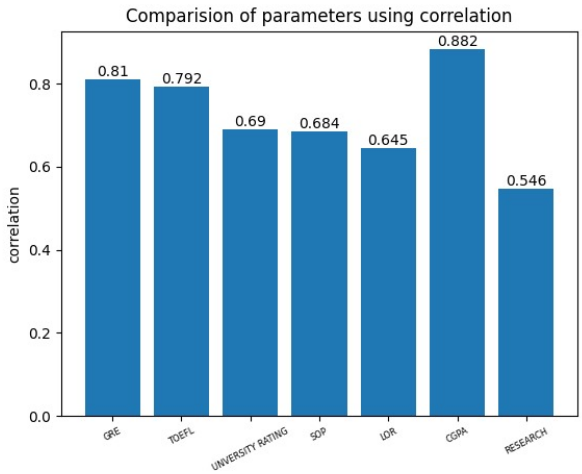
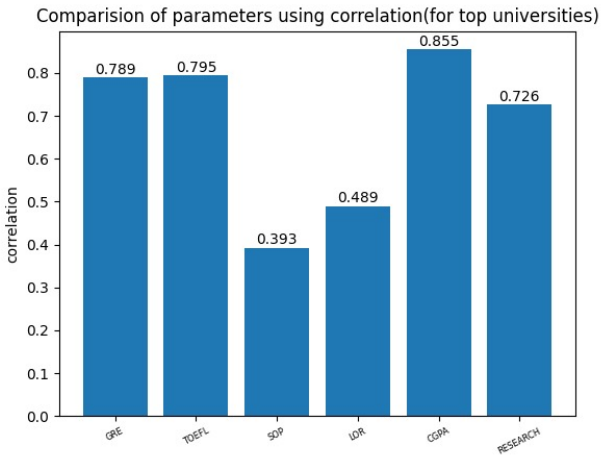


Figure 1: img

- **Top Parameters for Admission (All Universities)**:
  - CGPA(1st preference), GRE SCORE(2nd preference) and TOEFL Score(3rd preference) have the highest impact on the chance of admission across all universities.
- **Top Parameters for Admission (Top-Ranked Universities)**:
  - Among top-ranked universities (rating 5) - CGPA(1st preference),TOEFL SCORE(2nd preference), GRE SCORE(3rd preference) continue to have the highest impact on the chance of admission.





Recommendations for Admission to Top-Ranked Institutions

- **Focus Areas**:
  - Prospective students aiming for admission to top-ranked institutions should prioritize improving their GRE Score, CGPA, and TOEFL Score.