Stony Brook University

Graduate Admission Analysis

Name: Yifan Zhai

Department: Technology & Society

May 6, 2019

FAR BEYOND

Introduction

- Dataset: inspired by UCLA Graduate Dataset, open source on Kaggle
- Features: Serial No., GRE Score, TOEFL Score, University Rating, Statement of Purpose, Letter of Recommendation, Undergraduate GPA, Research Experience, Chance of Admission
- Objective: Influential factors, distribution of scores and make prediction

Tools

Application: Jupyter Notebook

Kenel: Python

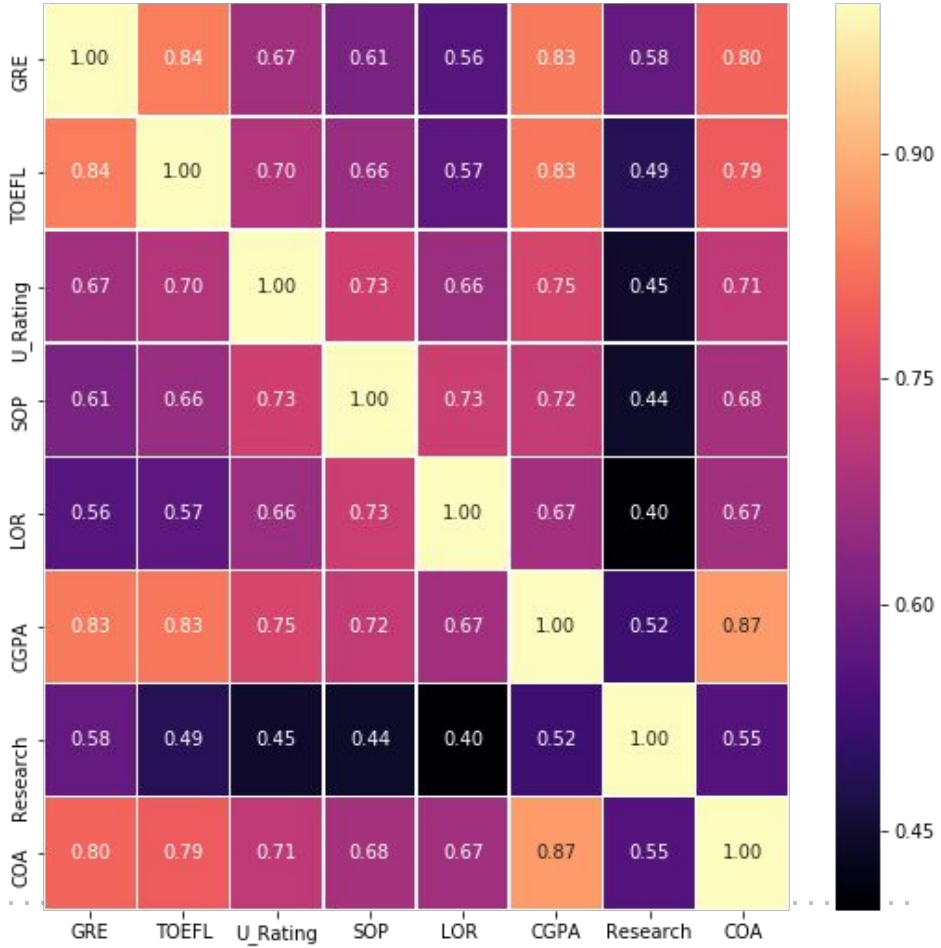
Package: Numpy, Pandas, Matplotlib, Seaborn, Sklearn

Models: Linear Regression, Logistic Regression, Decision Tree, Random Forest, SVM, KNN, Naive Bayes

Correlation

Three most important features:

- GRE Score (0.80)
- TOEFL Score (0.79)
- CGPA (0.87)



Stony Brook University

Pair Plot

Blue Points: No Research

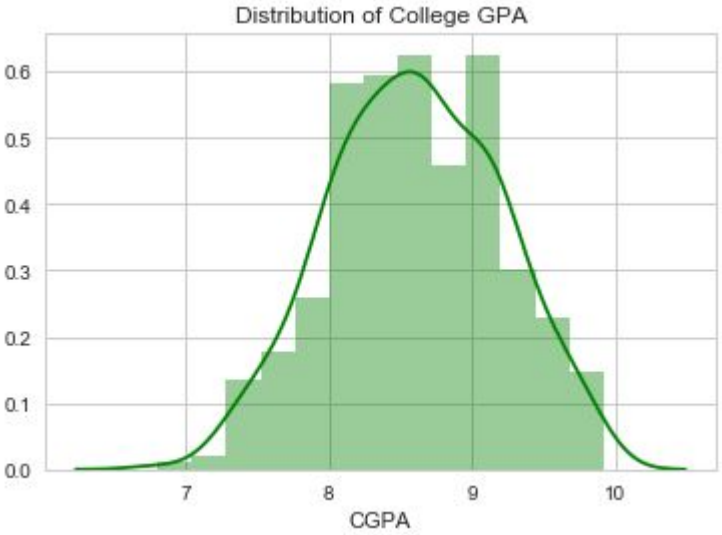
Green Points: Research Experience

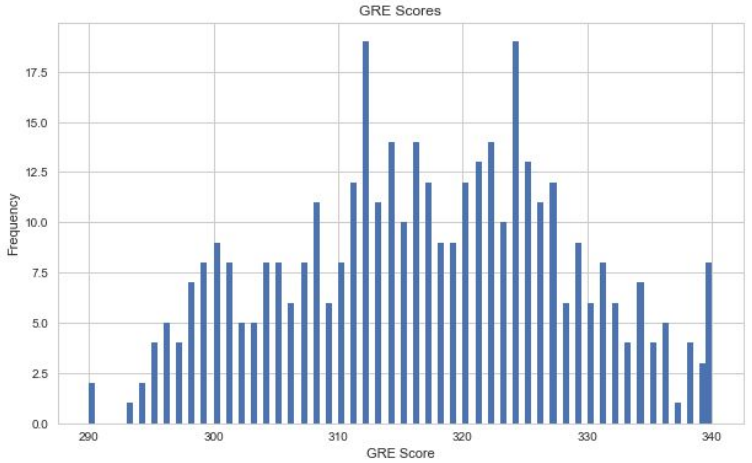
Relationship: Linear Dependence

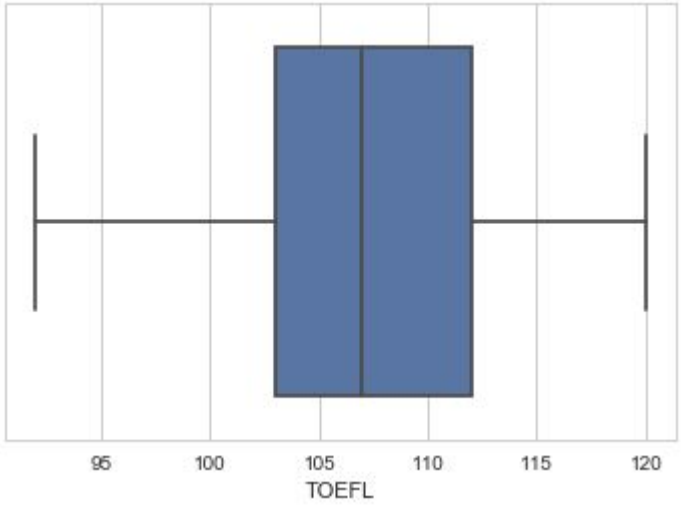
FAR BEYOND

Stony Brook University

Distribution







FAR BEYOND

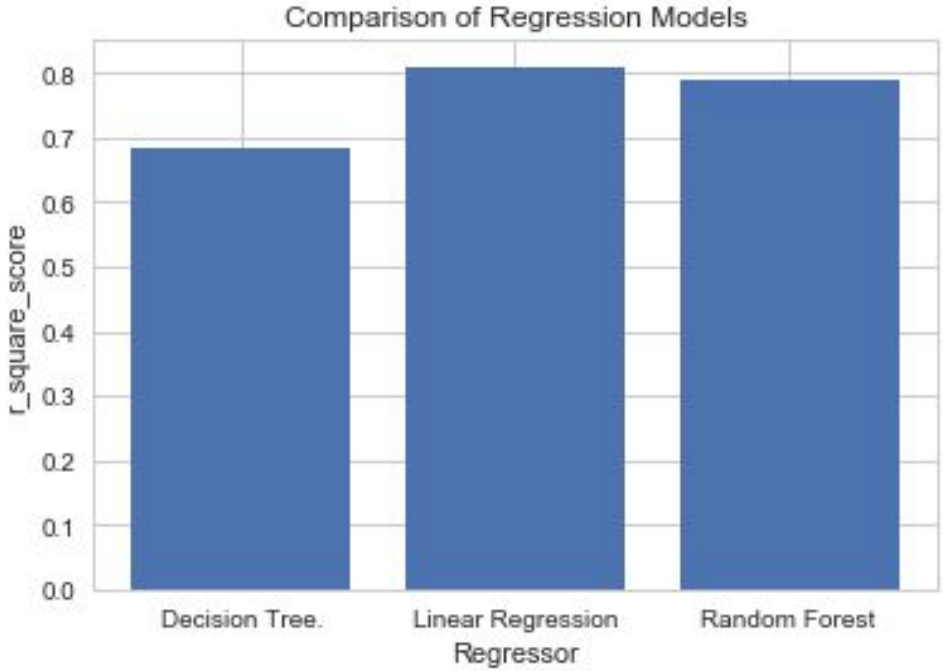
Regression Models

$$SSR = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$$

$$SSE = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

$$SSTO = \sum_{i=1}^{n}(y_i - \bar{y})^2$$

$$r^2 = \frac{SSR}{SSTO} = 1 - \frac{SSE}{SSTO}$$

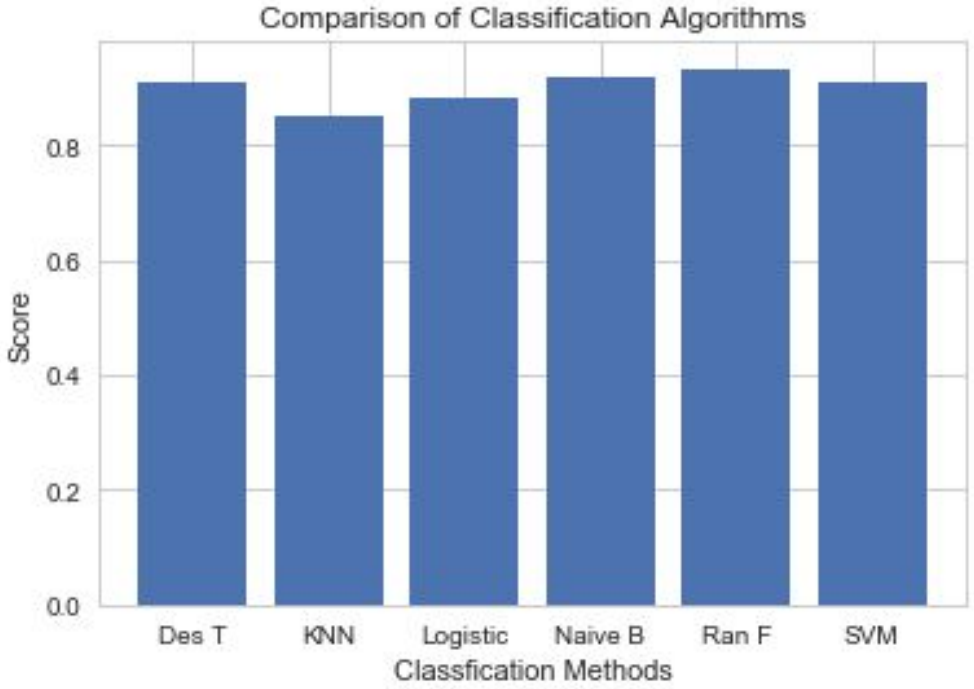


Classification Models

$$Precision = \frac{TP}{TP+FP}$$

$$Recall = \frac{TP}{TP+FN}$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$



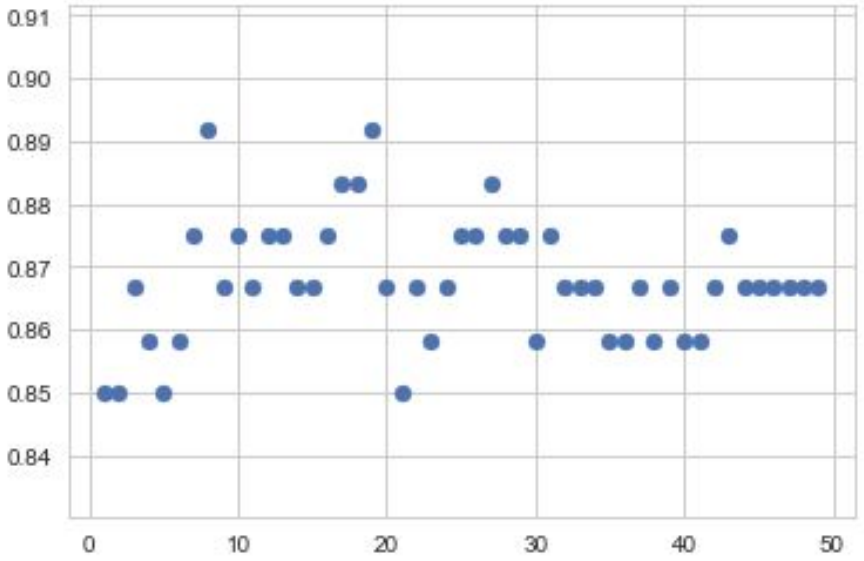
Something to Notice

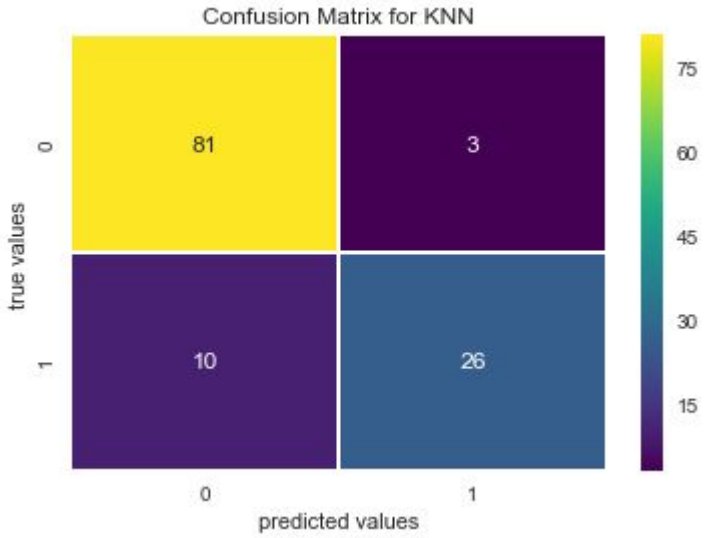
Decision Tree is easy to be overfitted (High Variance, Low Bias)

How to test: Make Prediction on both training and test dataset

Solution: Restricted the Max Depth of the tree

K-Value?


0.91
0.90
0.89
0.88
0.87
0.86
0.85
0.84
0
10
20
30
40
50


Confusion Matrix for KNN
81
3
10
26
true values
predicted values
0
1
0
1
75
60
45
30
15

Conclusion

- Best Regression Model: Linear Regression
- Best Classification: Random Forest Classification
- Future Work: Cross Validation, Larger Dataset

Reference

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