
ML4774 PROJECT PROPOSAL

A PREPRINT

 **Tianze Ren**

tr2bx@virginia.edu

 **Kaicheng Chu**

cbf6ah@virginia.edu

 **Haley Fitch**

hnf8bgt@virginia.edu

October 7th, 2023

1 Motivation

When it comes to their future moves after college, university students tend to have different plans. Some might decide to start their own business, some might want to work for big companies, while others choose to attend graduate programs to keep improving themselves. Information asymmetry would be a severe obstacle for graduate program applicants since the compatibility between students' status and school requirements is hard to measure. Every applicant has different educational backgrounds and professional experiences, and every program looks for different qualities in applicants. Thus, admission decisions are hard to make. We would like to build a graduate program recommendation system that provides individual applicants with the best possibility of admission to different schools. A system like this will provide applicants with an insightful reference for their application decisions.

2 Dataset

Data url: "<https://www.kaggle.com/datasets/nitishabharathi/university-recommendation/data>"

Data description: this dataset includes information about graduate program applicants and their applied universities admission results (admitted/ rejected). The dataset contains two major files:

1. Original.csv: profiles of students with admits/rejects to 45 different universities in USA.
2. score.csv contains score conversion for GRE old to GRE new score mechanism.

Potential Used Columns: major, researchExp, specialization, tofleScore, department, greV, greQ, Cgpa, univName, admit

Dataset size: 11.43 MB

Total attribute: 26

Total observation: 14798

3 Related Work

3.1 Example 1

University Classification by MOHAMMED RIZWAN MALIK

This project contains two main steps Data preprocessing and applying models.

Models Applied:

1. Random Forest Classifier,

2. Support Vector Classifier,
3. K Nearest Neighbours,
4. XGBoost Classifier,
5. Light GBM Classifier

3.2 Example 2

Project Classification by MOHAMMED AYAD

The author provides detailed data preprocessing methods to each single column of the dataset, and provides different plots to help visualize the data distribution.

Models Applied:

1. Logistic Regression,
2. KNeighborsClassifier,
3. Support Vector Classifier,
4. Decision Tree Classifier,
5. Random Forest Classifier,
6. XGB Classifier,
7. Naive Bayes Classifier

4 Intended Experiments

We plan to try different machine learning algorithms and compare the performance over the train set. A promising one would be K Nearest Neighbors algorithm, which gives each test set a closest successfully admitted example to different schools. Other algorithms we would like to try are Bayesian Networks, clustering models, etc.

For the test set, we plan to survey a group of UVA students with different attributes and give them recommendations for graduate programs. Through this test set sampling, we are using our model to contribute to our community.