# College Admission Prediction System

**ECE 143 Group 9 final project**

**Kendrick Nguyen**
**Chengzhan Gao**
**Varun Pawar**
**Weirong Dong**
**Houtianfu Wang**

# Motivation

- Time-saving: students can focus on schools most likely to accept them
- Strategic planning: knowing their chances of getting accepted can help students plan their future accordingly
- Peace of mind: a prediction system can alleviate the stress associated with waiting for acceptance letters.

# Dataset exploration

https://www.kaggle.com/datasets/mohansacharya/graduate-admissions

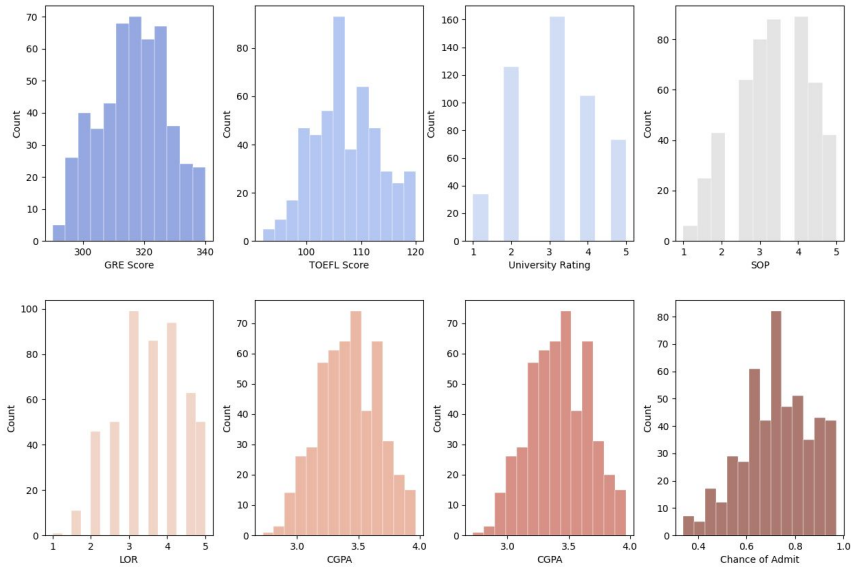
| | GRE Score | TOEFL Score | University Rating | SOP | LOR | CGPA | Research | Chance of Admit |
|---|---|---|---|---|---|---|---|---|
| 0 | 337 | 118 | 4 | 4.5 | 4.5 | 3.860 | 1 | 0.92 |
| 1 | 324 | 107 | 4 | 4.0 | 4.5 | 3.548 | 1 | 0.76 |
| 2 | 316 | 104 | 3 | 3.0 | 3.5 | 3.200 | 1 | 0.72 |
| 3 | 322 | 110 | 3 | 3.5 | 2.5 | 3.468 | 1 | 0.80 |
| 4 | 314 | 103 | 2 | 2.0 | 3.0 | 3.284 | 0 | 0.65 |

https://github.com/AlpAribal/gradcafestats/blob/master/data/submissions.csv

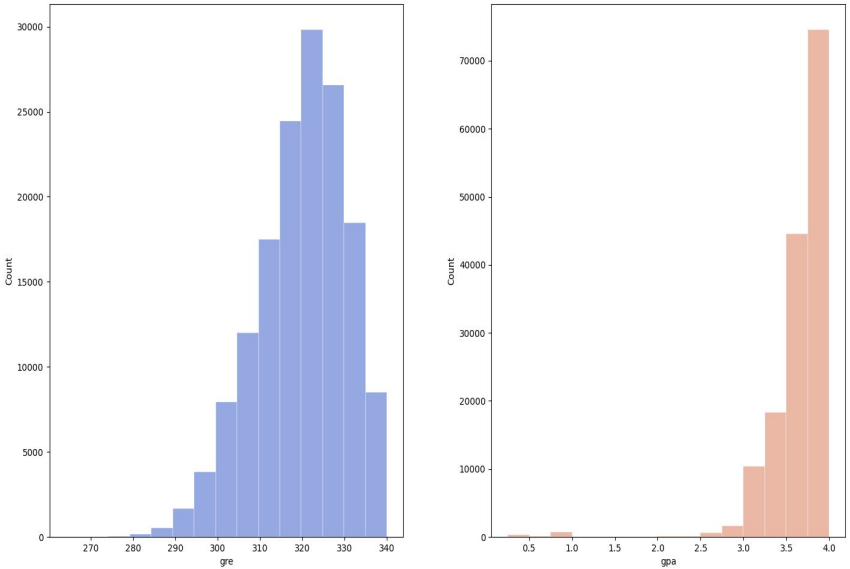
| | submissionDate | institution | major | degree | notif_result | notif_date | gpa | gre |
|---|---|---|---|---|---|---|---|---|
| 0 | 2010-01-12 | University Of California, San Francisco (UCSF) | Neuroscience | PhD | Interview | 2010-01-12 | 3.80 | 325.0 |
| 1 | 2010-01-12 | University Of California, San Francisco (UCSF) | Neuroscience | PhD | Interview | 2010-01-12 | 3.86 | 326.0 |
| 2 | 2010-01-12 | University College London | International Public Policy | Masters | Accepted | 2009-11-26 | 3.60 | 309.0 |
| 3 | 2010-01-12 | London School Of Economics (LSE) | Master Of Public Administration (MPA) | Masters | Accepted | 2010-01-11 | 3.60 | 309.0 |
| 4 | 2010-01-12 | Carnegie Mellon | Biological Sciences (Neuroscience) | PhD | Interview | 2010-01-12 | 3.40 | 331.0 |

# EDA of Kaggle and GradCafe Datasets

Distribution of Features (Kaggle Dataset, 500 samples)



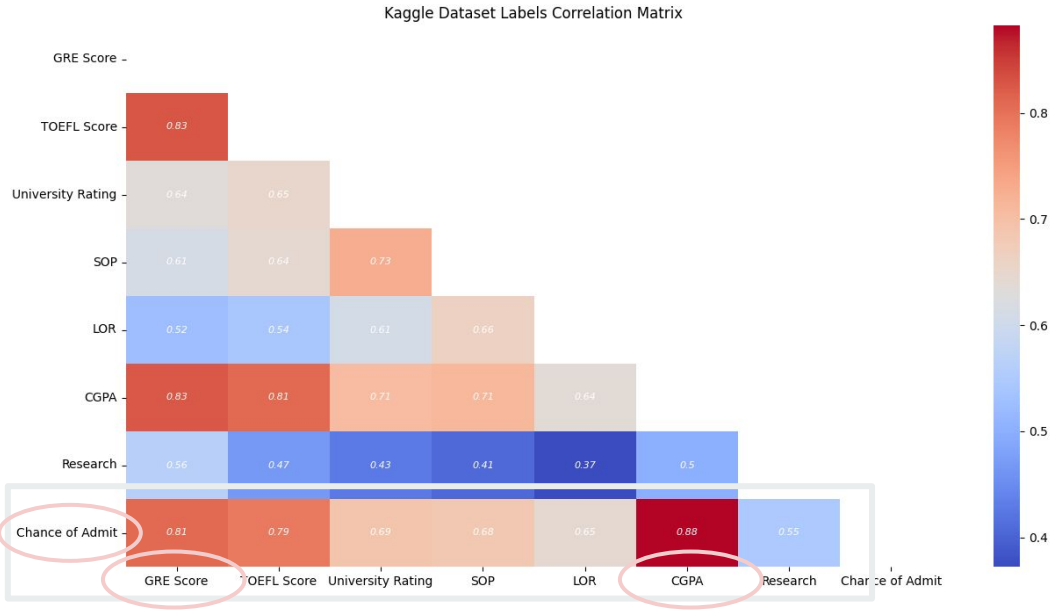

Distribution of Features (GradCafe Dataset, 152, 256 samples)



# EDA of Kaggle and GradCafe Datasets

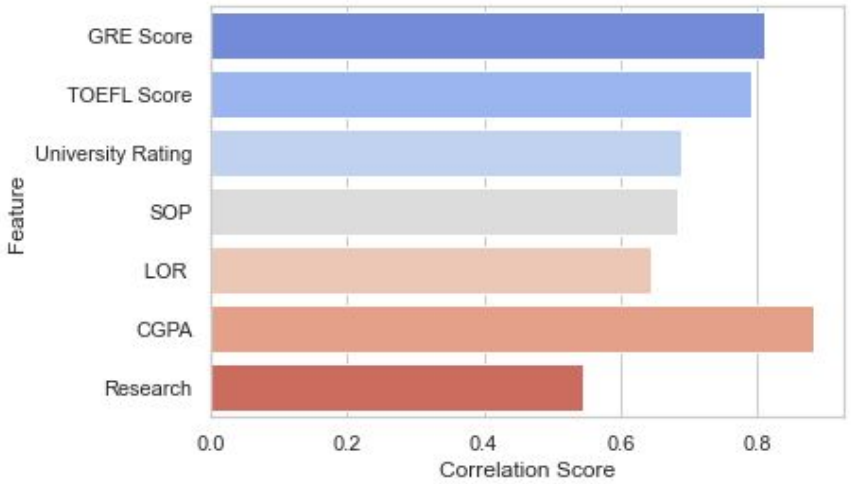
Our strongest labels as observed against the Chances of Admission are GRE scores and CGPA.

They are strongly correlated with correlation coefficients of *0.81* and *0.88* respectively

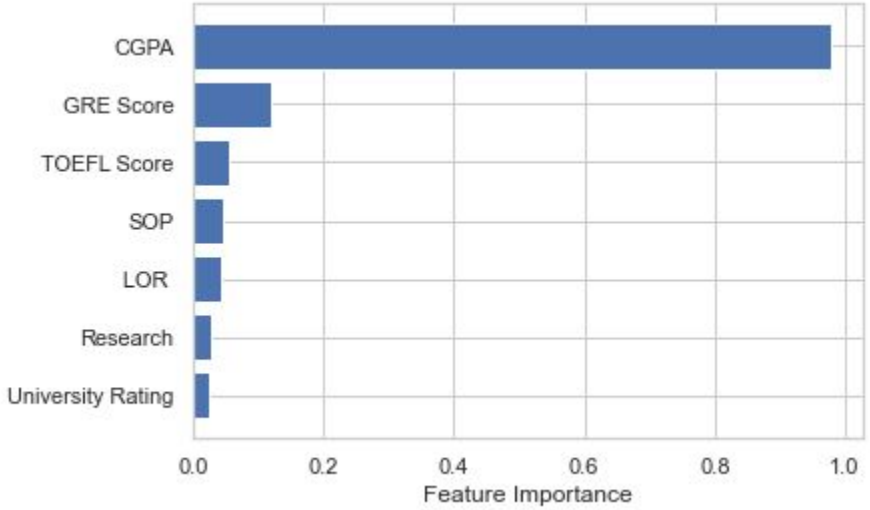


# Which features have greater impact on the results ?

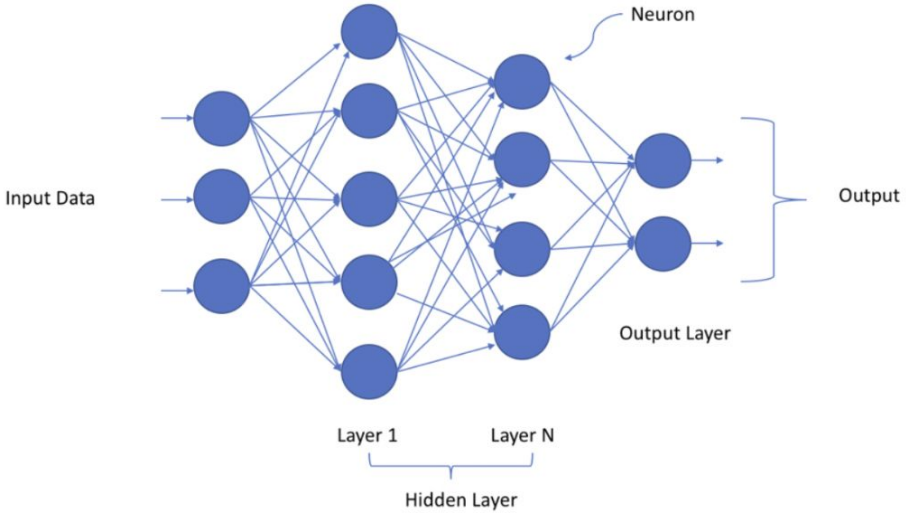
Correlation



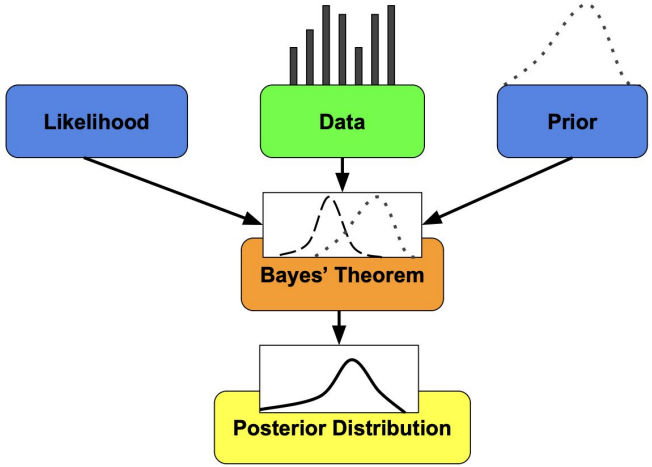

Feature permutation importance



# Methodology of prediction model




ANN Algorithm




Naive Bayes Theorem Algorithm

# Result and conclusion

## Part1 Models on Kaggle Dataset

Several different models were trained on the Kaggle Dataset.

Model with highest accuracy: Naive Bayes

Model with lowest accuracy: K-Neighbours Regressor

| GRE Score | TOEFL Score | University Rating | SOP | LOR | CGPA | Research |
|---|---|---|---|---|---|---|
| 0.0024 | 0.0030 | 0.0026 | 0.0018 | 0.017 | 0.28 | 0.024 |

Coefficients of features of Linear Regression Model

| Model | Accuracy |
|---|---|
| Linear Regressor | 0.94 |
| Decision Tree Regressor | 0.94 |
| Random Forest Regressor | 0.95 |
| XGB Regressor | 0.94 |
| K-Neighbours Regressor | 0.93 |
| ANN | 0.96 |
| Naive Bayes | 0.98 |
| **Average** | 0.95 |

# Result and conclusion

## Part2 Models on Grad-cafe Dataset

The results on the Grad-cafe dataset are not **satisfactory**.

**Reasons**:

1. Most universities (**>99%**) only has less than **10** pieces of data
2. **Inconsistency** of universities' name

Improvement ways:
Train models separately on different universities and programs

| General Model | Accuracy |
|---|---|
| ANN | 0.53 |
| Naive Bayes | 0.35 |

**Separately Training**

| Separate Model | ANN | Naive Bayes |
|---|---|---|
| UCLA | 0.60 | 1.00 |
| UPenn | 0.73 | 1.00 |
| Brown University | 0.80 | 0.90 |
| … | … | … |

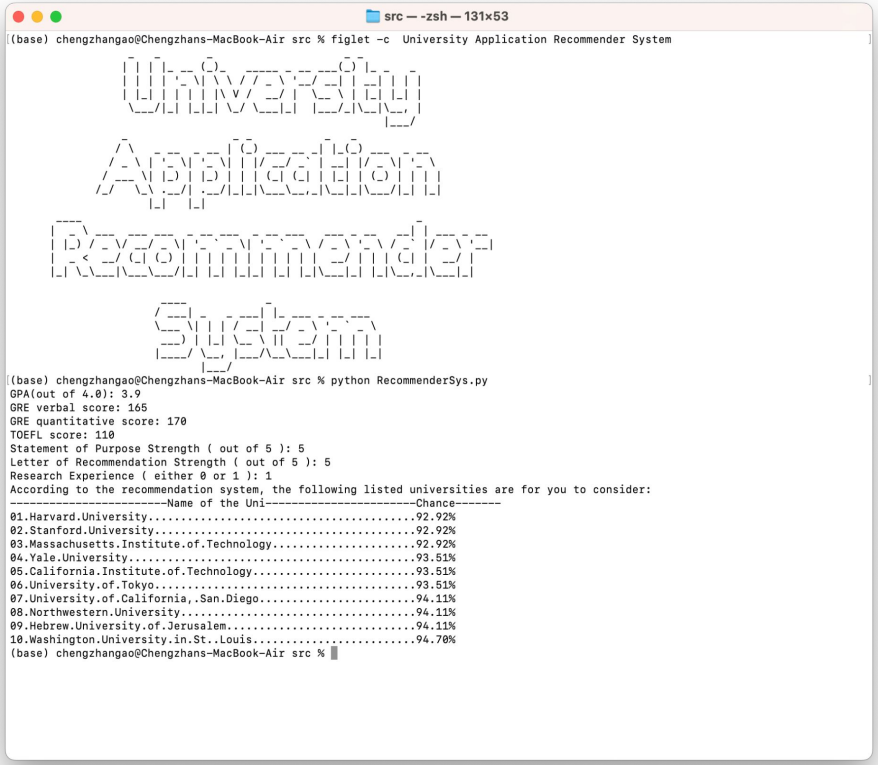
# Application in real world

Previous well-trained models can be used to build a recommender system for students base on their own profile.

**Benefits:**

1. **Personalized**
2. **Time-saving**
3. **Extendable**

Video Link:
https://youtu.be/vYbMVhXM0XA




Recommender System Demo on Terminal

# Q & A

[ECE 143] Project Schedule

Project Start Date 1/27/2023 (Friday) Display Week 1
Project Lead Christoffel

| WBS | TASK | START | END | DAYS | % DONE | WORK DAYS |
|---|---|---|---|---|---|---|
| 1 | **Project proposal** | | - | | | - |
| .1 | Review sample proposal | Fri 1/27/23 | Sun 1/29/23 | 3 | 100% | 1 |
| .1 | Choose topic | Sun 1/29/23 | Mon 1/30/23 | 2 | 100% | 1 |
| .1 | Write proposal draft | Mon 1/30/23 | Wed 2/01/23 | 3 | 100% | 3 |
| .1 | Complete project proposal | Wed 2/01/23 | Thu 2/02/23 | 2 | 100% | 2 |
| 1 | **Model training and visualization** | | - | | | - |
| .1 | Data preprossing and cleaning | Tue 2/07/23 | Tue 2/14/23 | 8 | 100% | 6 |
| .1 | Coding for prediction model | Wed 2/15/23 | Sat 3/04/23 | 18 | 100% | 13 |
| .1 | Data visualization | Wed 2/15/23 | Sat 3/04/23 | 18 | 100% | 13 |
| 1 | **Final project presentation** | | - | | | - |
| .1 | Complete coding | Sun 3/05/23 | Tue 3/14/23 | 10 | 100% | 7 |
| .1 | Complete presentation slides | Mon 3/06/23 | Tue 3/14/23 | 9 | 100% | 7 |
| .1 | Complete presentation script | Mon 3/06/23 | Tue 3/14/23 | 9 | 100% | 7 |
| .1 | Presentation | Wed 3/15/23 | Wed 3/15/23 | 1 | 100% | 1 |
| .1 | Document submission | Wed 3/15/23 | Fri 3/17/23 | 3 | 80% | 3 |