EC601 Project:
Automated Admission System

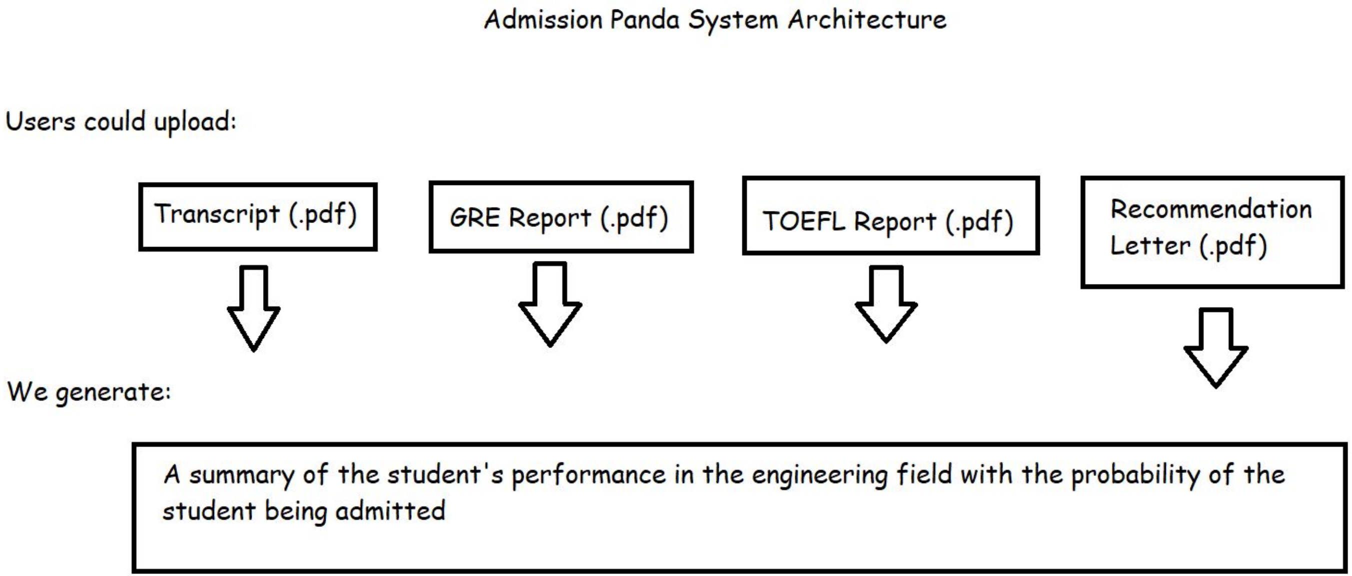
Sprint 2

Yu Liu
Shengyao Shao
Jing Song

Project Components

- Transcript Processing
  - Summary of the student’s performance in engineering field
  - Cloudconvert API: pdf to txt
  - Google Natural Language API: Analyzing Entities
- GRE score report, TOEFL Report processing
  - Cloudconvert API: pdf to txt
  - Google Natural Language API: Analyzing Entities
- Recommendation Letter Sentiment Analysis
  - Google Natural Language API: Sentiment Analysis
- Probability of admission
  - Machine Learning Algorithm
- Generate a pdf file of all parameters above for each student

System Architecture Pt.1



Admission Panda System Architecture

Users could upload:

Transcript (.pdf)

GRE Report (.pdf)

TOEFL Report (.pdf)

Recommendation Letter (.pdf)

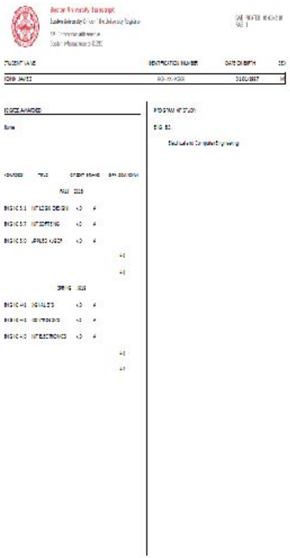
We generate:

A summary of the student's performance in the engineering field with the probability of the student being admitted

System Architecture Pt.2

Transcript Processing:

Transcript pdfs



[illegible] University Transcript

[illegible]

STUDENT NAME | IDENTIFICATION NUMBER | DATE OF BIRTH | SEX

JOHN JAMES | [illegible] | [illegible] | M

DEGREE AWARDED

None

PROGRAM OF STUDY

ENG BS

[illegible] Computer Engineering

COURSE(S) TITLE CREDIT GRADE [illegible]

FALL [illegible]

[illegible] DIGITAL LOGIC DESIGN 4.0 A

[illegible] 4.0 A

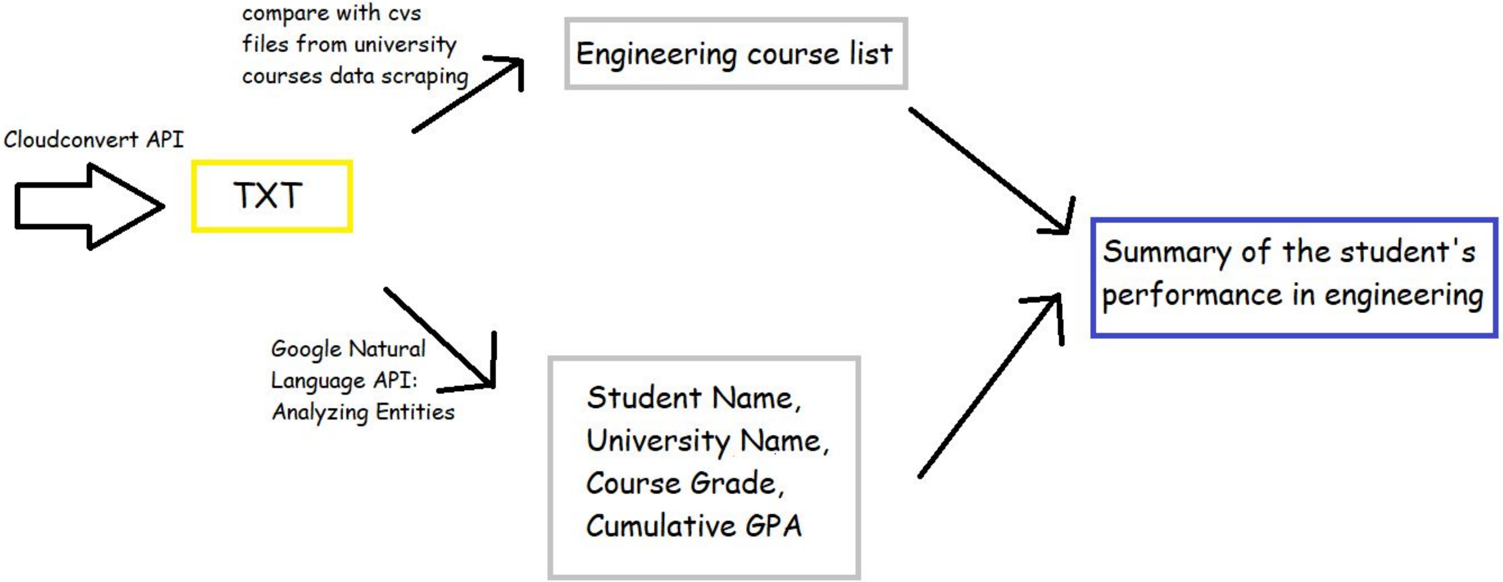
[illegible] 4.0 A

SPRING [illegible]

[illegible] 4.0 A

[illegible] 4.0 A

[illegible] INT ELECTRONICS 4.0 A

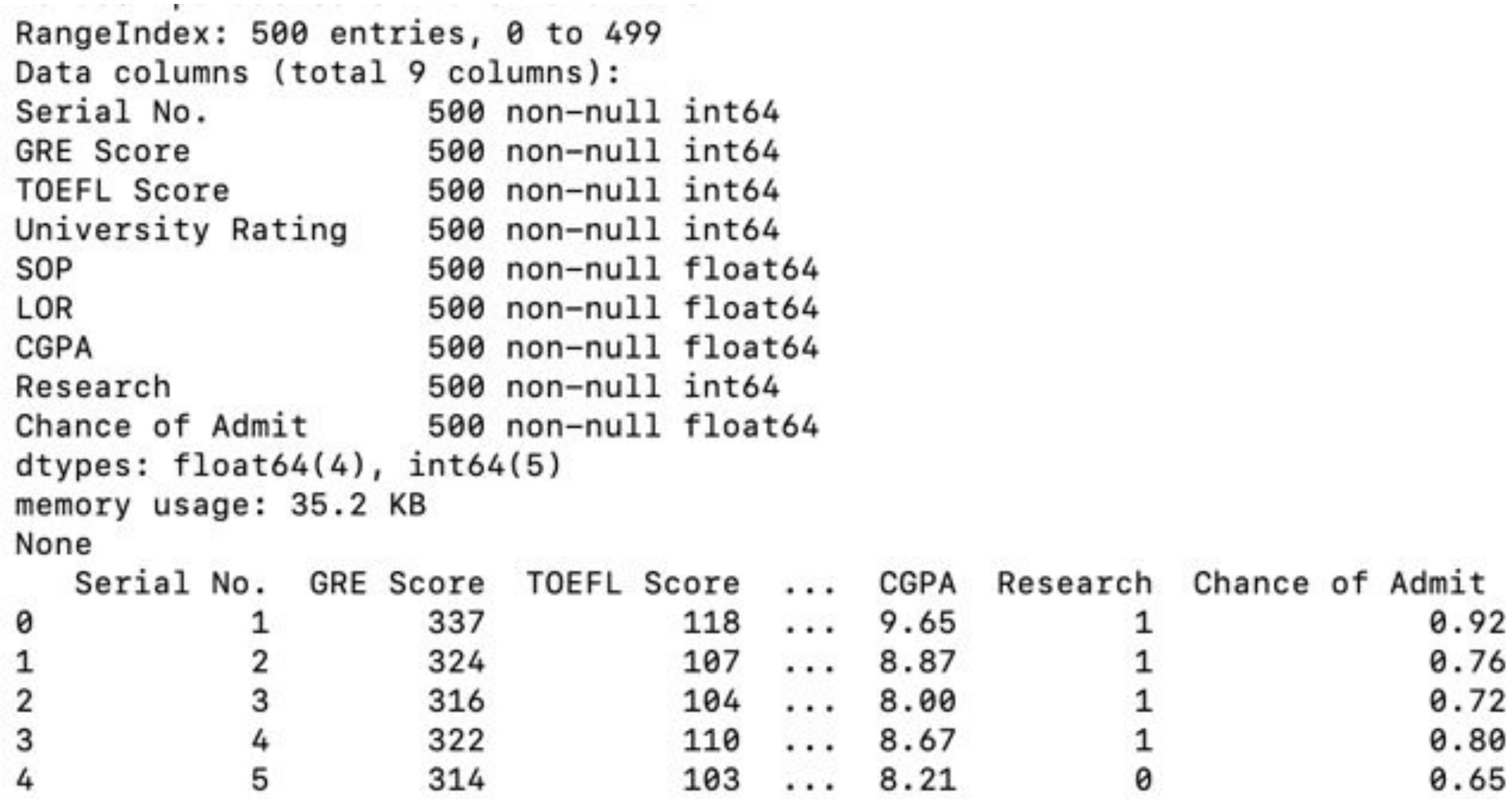


Probability of admission

Dataset: **Graduate Admissions Dataset** from Kaggle

This dataset is inspired by the UCLA Graduate Dataset. The GPAs are in the older format. The dataset is owned by Mohan S Acharya.

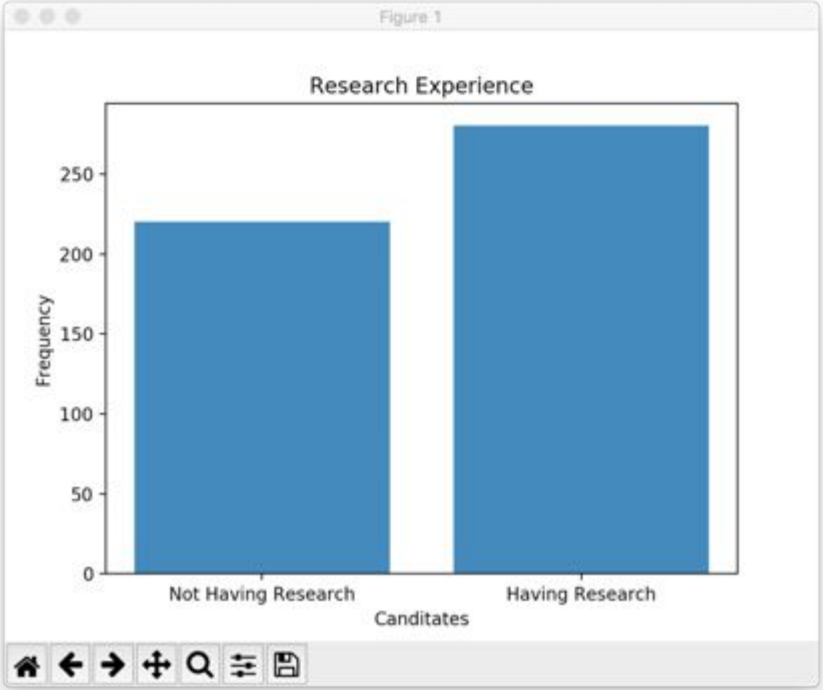
Probability of admission

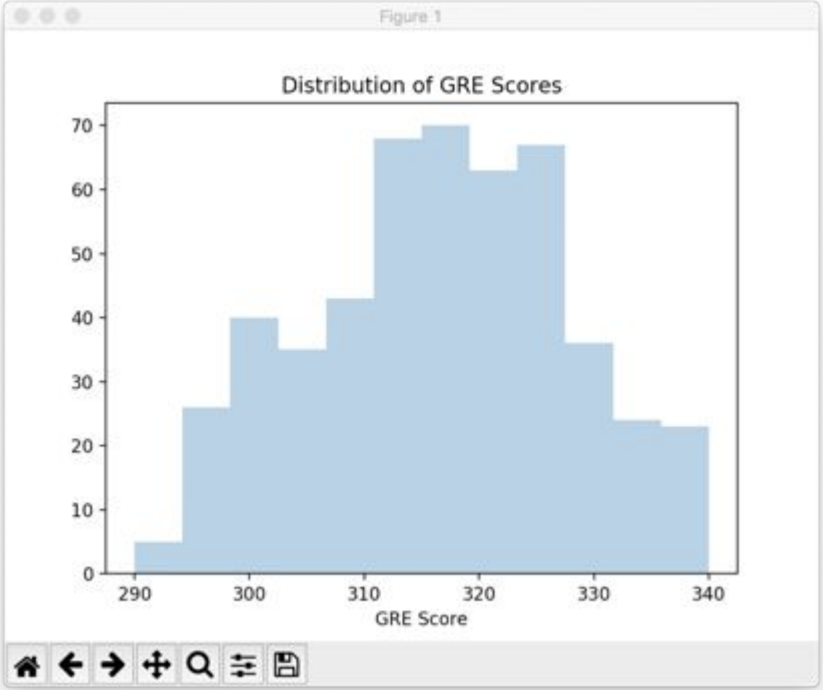


```
RangeIndex: 500 entries, 0 to 499
Data columns (total 9 columns):
Serial No.           500 non-null int64
GRE Score            500 non-null int64
TOEFL Score          500 non-null int64
University Rating    500 non-null int64
SOP                  500 non-null float64
LOR                  500 non-null float64
CGPA                 500 non-null float64
Research             500 non-null int64
Chance of Admit      500 non-null float64
dtypes: float64(4), int64(5)
memory usage: 35.2 KB
None
   Serial No.  GRE Score  TOEFL Score  ...  CGPA  Research  Chance of Admit
0           1        337          118  ...  9.65         1             0.92
1           2        324          107  ...  8.87         1             0.76
2           3        316          104  ...  8.00         1             0.72
3           4        322          110  ...  8.67         1             0.80
4           5        314          103  ...  8.21         0             0.65
```

Probability of admission

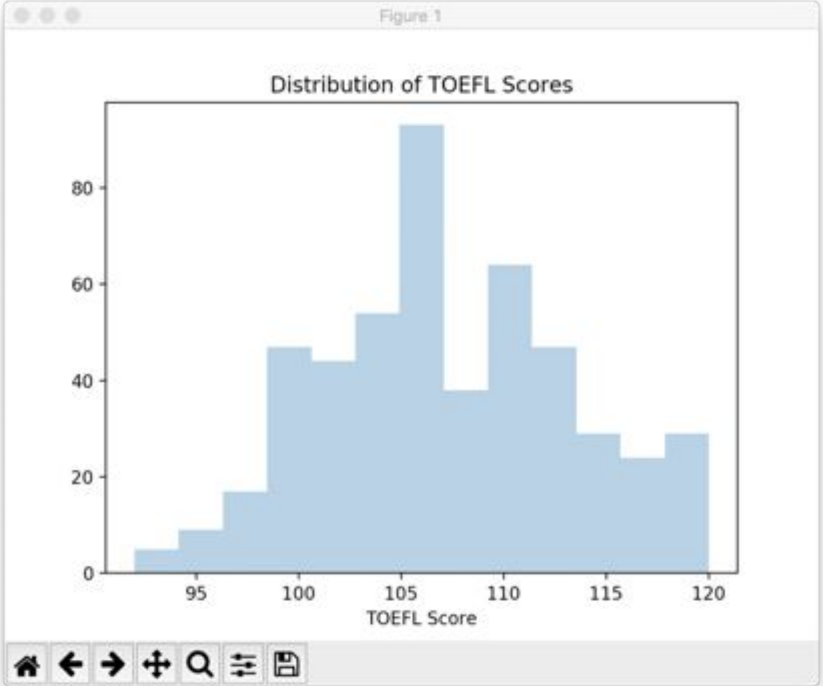
Data Visualization:

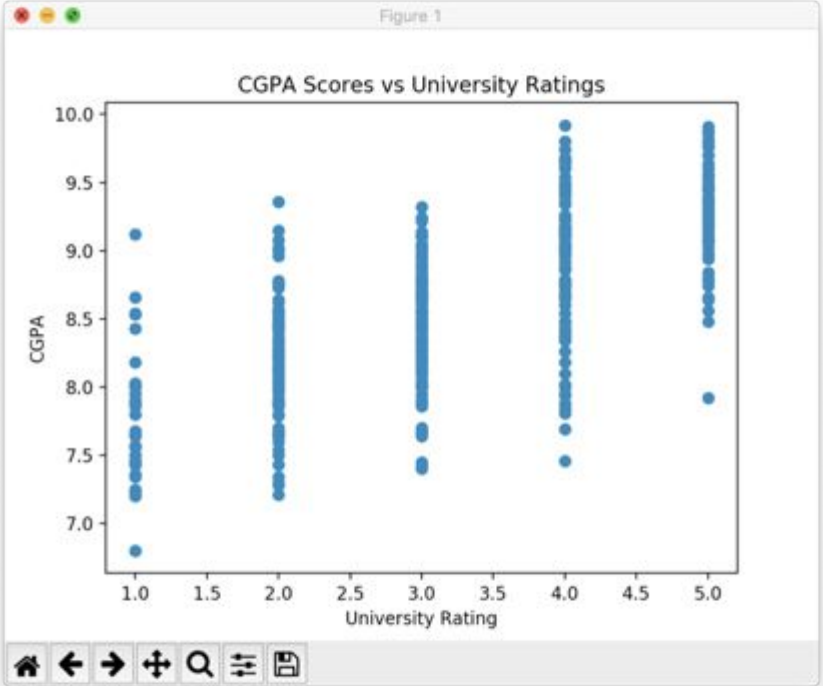




Probability of admission

Data Visualization:





Probability of admission

Feature importance:

```
            Features  Importance
5               CGPA    0.722230
0          GRE Score    0.164595
4                LOR    0.031082
3                SOP    0.029332
1        TOEFL Score    0.027604
2  University Rating    0.015509
6           Research    0.009648
```

Probability of admission

Mean squared error of probability prediction by using different Algorithms:

Decision Tree : 0.08841945487278237
Linear Regression : 0.05783397524904237
Random Forest : 0.05521014399546518
Lasso: 0.1148860969290504
Ridge: 0.05831519386729676
Bayesian Ridge: 0.059175262027732614

Probability of admission

Preprocessing:

```
data = pd.read_csv("/Users/liuyu/Downloads/Admission_Predict_ver1.1.csv", sep = ",")

scalar = preprocessing.MinMaxScaler()

scalar_fit = scalar.fit(data[['GRE Score', 'TOEFL Score']])

normalized_score = scalar_fit.transform(data[['GRE Score', 'TOEFL Score']])
```

Probability of admission

Results:

```
   GRE Score  TOEFL Score  ...  ChancesOfAdmit  PredictedChancesOfAdmit
0       0.94     0.928571  ...            0.92                  0.92980
1       0.68     0.535714  ...            0.76                  0.77490
2       0.52     0.428571  ...            0.72                  0.69135
3       0.64     0.642857  ...            0.80                  0.78755
4       0.48     0.392857  ...            0.65                  0.64575

[5 rows x 9 columns]
Mean Absolute Percent Error:  3 %
Average Accuracy of the model:  97 %
```

Demo

Sprint 3 Tasks

- Utilize database (MySQL)
- Expand data scraping scale, pass the module to Professor Trachtenberg for testing
- Add score report processing
- Add firewall protection to the file uploading process (utilize open-source security framework??)
- Work on website design (file uploading system, etc.)

Questions?

Thanks!