



Shiny App

Guide to Picking a College Major

Team 9

Kai-Duan Chang, Li-Ci Chuang, Sri Manogna Gurijala, Ryan Egbert, Meghan Harris

Business Problem

**Students feel lost and have no direction
when deciding a college major.**

Business Problem

Stakeholders

High school
students

College
students

Students
looking to
change their
college major

Our Opportunity

Provide a guide to help students choose their major.



Our Opportunity

Provide a guide to help students choose their major.

- 1 Help students understand the earning difference and employment rate between majors.
- 2 Recommend majors based on students' interests.

Analytics Problem

**Find a relationship between major category
and the high school subjects.**

Analytics Problem

Drivers and relationships to outputs

1

Average Salary

2

Employment Rate

3

Number of students/ Male and Female

Key Assumptions

Students want to pursue higher salary.

Students care about the employment rate in the field they choose.

Students would like to know the class profile before making a decision.

Key metrics of success

Potential model performance measurements:

- Hit Ratio
- Accuracy Percentage
- AUC
- Utilizing `h2o.get_leaderboard/h2o.performance`

Data

Data PreProcessing Steps

1

Data Acquisition

2

Data Cleaning

3

**Data
Transformation**

1

Data Acquisition

- “College Majors” dataset was found on Kaggle.com

Data Cleaning

Data
Transformation

2

Data Acquisition

Data Cleaning

- Removed irrelevant major category and missing values
- Split combined major category into separate major categories
- Added unique features corresponding to each major category for predictive model

Data Transformation

3

Data Acquisition

Data Cleaning

Data Transformation

- Added dummies based on major category
- Checked and removed multicollinearity and perfect collinearity

Methodology Selection

Methodology Types

1

Descriptive Analysis

2

Predictive Analysis

1

Descriptive Analysis

Using the data of several majors list, understanding the statistics of each major preferences, and visualizing the results.

1

Descriptive Analysis

Using the data of several majors list, understanding the statistics of each major preferences, and visualizing the results.

Why?

1

Descriptive Analysis

Using the data of several majors list, understanding the statistics of each major preferences, and visualizing the results.

For the ease of exploring and comparing different variables within each major category.

Why?

Providing accurate data analysis through the model.

Effective exploratory data analysis using tables, bar graphs, scatter plots, and slope chart.

2

Predictive Analysis

Predicting the majors based on the selection of interests by the students.

2

Predictive Analysis

Predicting the majors based on the selection of interests by the students.

Why?

2

Predictive Analysis

Predicting the majors based on the selection of interests by the students.

Forecasting the major category selection and analyzing the prospects like salary , employment rate etc.

Why?

Projecting what is going to happen through the selections made by the students.

2

Predictive Analysis

Predictive Variables:

2 fields of
interest to
draw the
comparison

Salary range
selection

Three
favorite/least
favorite high
school subjects

R

Advantages of Using R

The **library(ggplot2)** is used for the aesthetic graphical representation, which is one of the key aspects in our project.

Using the **shiny package** was extremely easy to make our reports with data, plots and R scripts embedded within.

Regression, predictive, and descriptive models, being majorly used throughout our project, are facilitated through the R environment.



R

Advantages of Using R

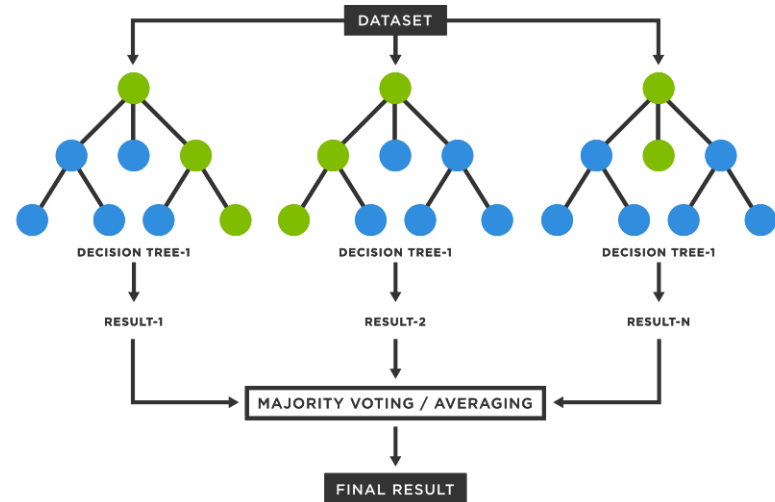
R is an open source programming language providing the freedom to work with in the environment with multiple packages.

The main packages we used in this project helped us transform messy data into a structured form.

Model Building

Model Building

- There are only a handful of predictors
- Use Random Forest model (H2O)



Method

1

Iterate through several different values

2

Final parameters: `ntrees = 50`, `max_depth = 5`

3

Model is loaded in when app starts

4

On the 'Recommend' page, the prediction is made

Functionality

1

Explore

2

Compare

3

Recommend

Future enhancements

1

**Include more
data and data
variables**

2

**Add more
predictive
variables for
predictive model**

References

References

We based our Shiny app UI on Zimin Luo's and Lasha Gochiashvili's project "Graduate Employment in Singapore".

https://gesurvey.shinyapps.io/Graduate-Employment-Survey/#1_brief_introduction



GRADUATE EMPLOYMENT IN SINGAPORE

Summary

Visual Comparison

Details By University

Data by Year

Select Year

2018

Ranking	No. of Graduates	
University	Employment Rate Overall (Avg %)	Gross Monthly Median Income (Avg)
Singapore University of Social Sciences	94	3133
Singapore Management University	93	3955
Singapore Institute of Technology	92	3350
National University of Singapore	91	3625
Nanyang Technological University	89	3472

Referenced UI example

Thank you!