

Project Name: Job Hunting Strategy - Career Development Recommendation

Yuanyuan Xiao: yuanyuan_xiao@berkeley.edu

Cheng-Chun Chao: kevin19960524@berkeley.edu

Yijun Liang: joanna_liana@berkeley.edu

Sean Chuang: sean_chuang@berkeley.edu

Xinyu Wang: xinyu_wang@berkeley.edu



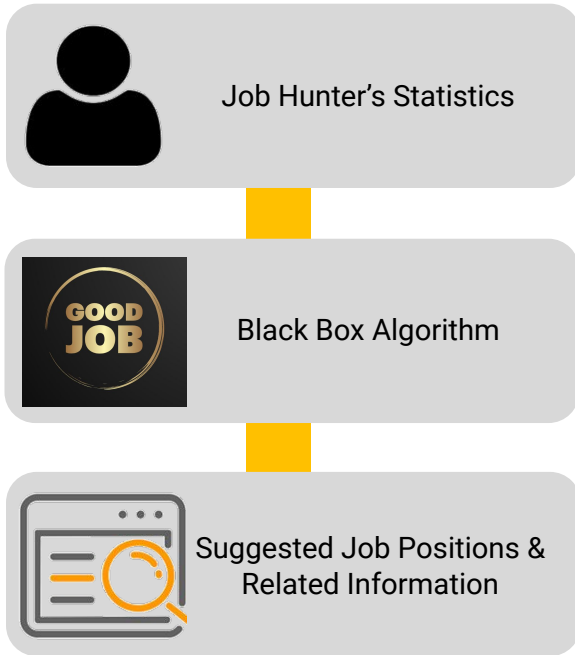
Job Hunting Strategy

- Career Development Recommendation

Project Overview

Cheng-chun Chao, Sean Chuang, Yijun Liang, Xinyu Wang, Yuanyuan Xiao

Job Hunting Strategy - Career Development Recommendation



Potential Users

- Anyone who is looking for a job, especially fresh graduate.
- Company HR or Staffing Agents

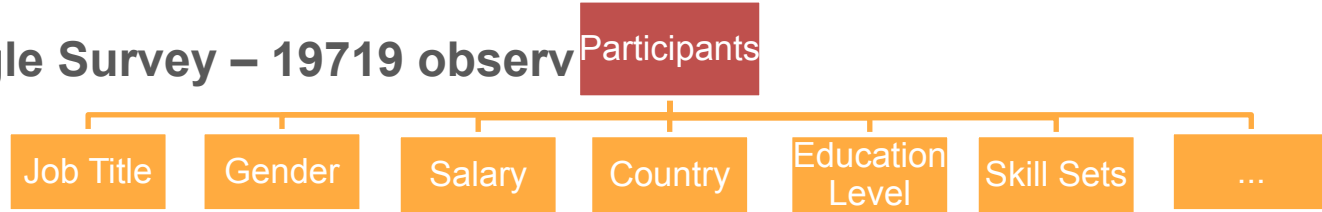
Interests/Benefits

- For Job Hunters:
 - Find suitable jobs based on their current skill sets.
 - How their majors and knowledge can fit the need of industrial
- For HR:
 - to better allocate talents in company

Data Collection and Data Cleaning

Data Source:

1. Kaggle Survey – 19719 observ



2. **Berkeley Academic Guides** – Courses recommendation for Berkeley student

3. **Amazon Books** – Books that help corresponding skills and salary level

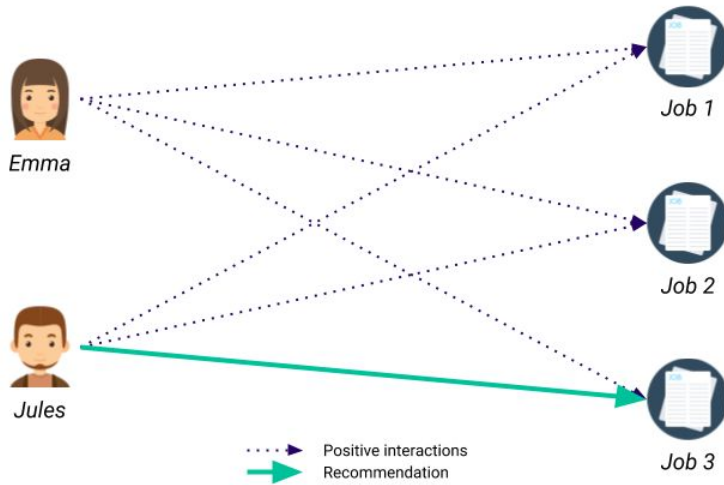
Other: PayScale, LinkedIn, Glassdoor

Data Cleaning:

- Combining data set from different sources
- Select important featured related to our work

Recommendation System Algorithm

Collaborative



Neural Network (NN)

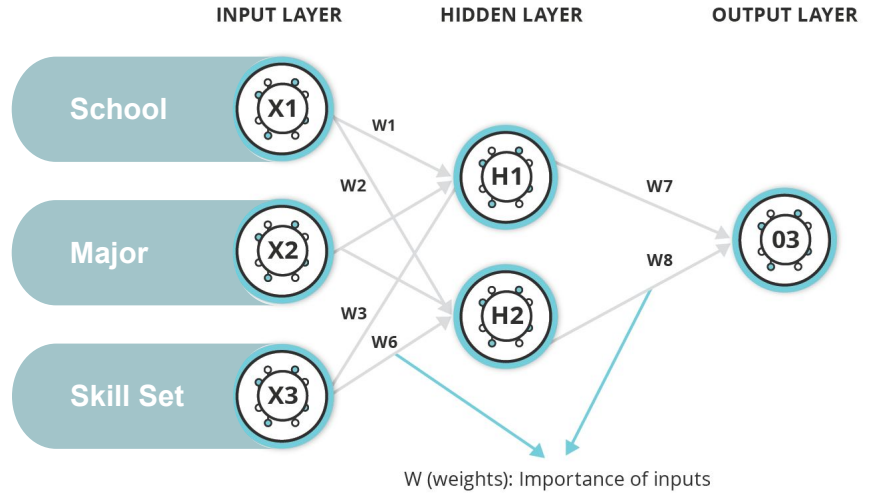


Image: Accessed on Feb. 21st, <https://www.welcometothejungle.com/en/articles/collaborative-filtering-job-recommendations>

Image: Accessed on Feb. 21st,

<https://www.getsmarter.com/blog/career-advice/how-artificial-neural-networks-can-be-used-for-data-mining/>

User Perspective. Illustrate UI or Input /Output

GoodJob Website

www.GoodJob.com

Major: Industrial Engineering
Skillset: Python
Experience: 0-1year
GO

} User information input

Recommendations:
Job:
Salary Range:
Skills to Master:

} Recommendations

Skills sets

V ☐

V ☐

X ☐

X ☐

Salary Range

Minimum Viable Product Prototype

GoodJob Website

www.GoodJob.com

Course Recommendation:

Course Number	Course Name	Instructor	Schedule	Description
1 27845	Special Topics in Industrial Engineering and Operation Research Applied Data Science with Venture Applications	Arash Nourian	TU 6:00 pm - 8:59 pm	Lectures and appropriate assignments
2 28490	Applications in Data Analysis	George Y Ng	W 6:00 pm - 8:59 pm	This course applies foundational concepts in....

Book Recommendation:

IKHLAQ SIDHU

INNOVATION
ENGINEERING

A Practical Guide to
Creating Anything New

Mirzadeh Rahati

An Introduction
to Machine
Learning

Second Edition

Springer

Gareth James,
David Witten,
Robert Kuhn,
Robert Iyer

An Introduction
to Statistical
Learning

with Applications in R

Springer

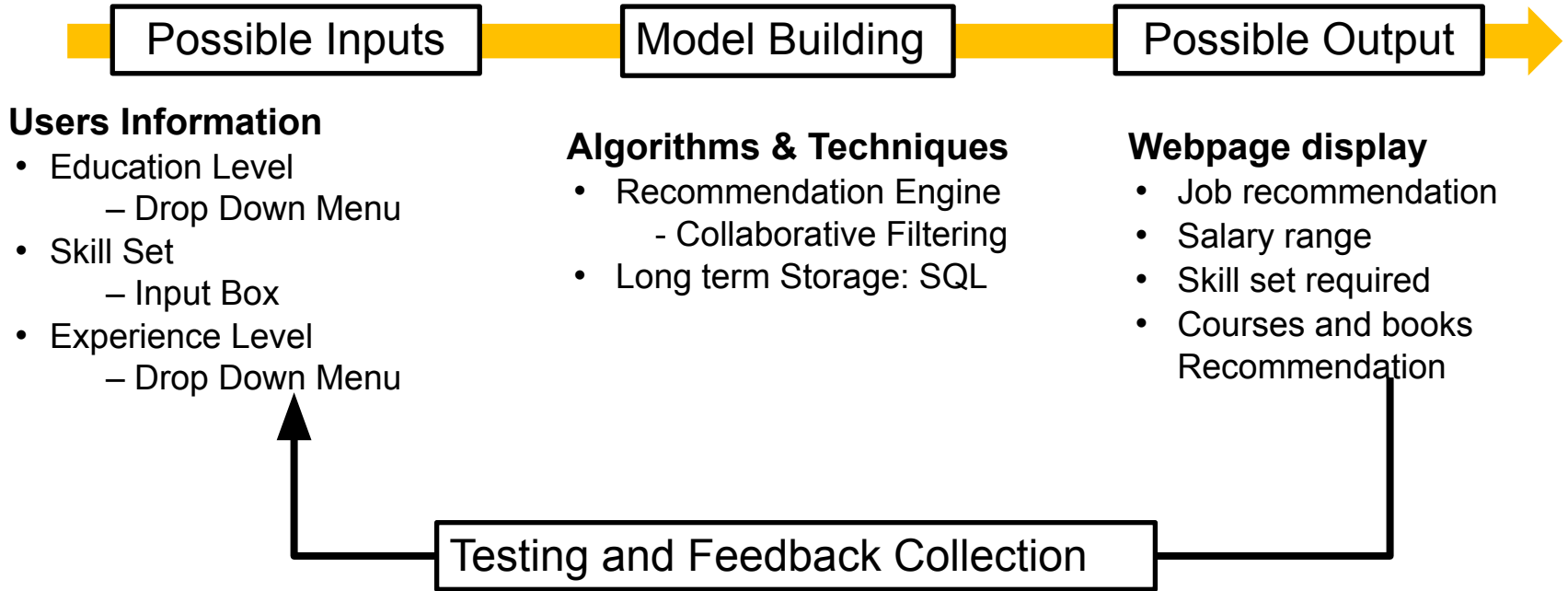
Technical Components of Project

Top Components in order of Importance:

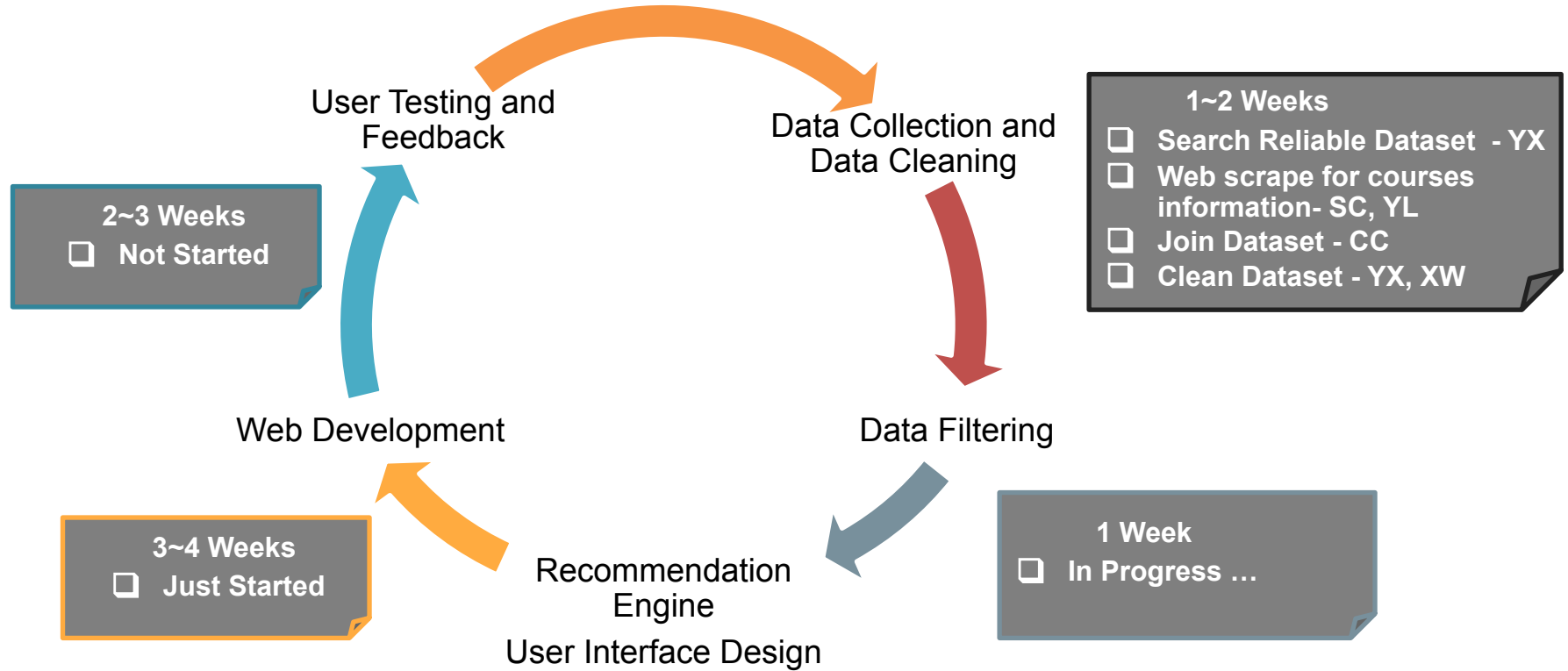
- Data Accessibility
- Python web data scraping
- Data cleaning
- Deep Learning based Recommendation System
- Use of Neural Network Algorithms
- User Interface Design
- Web Integration



Architecture Design



Project Plan – Hybrid Agile Management



Thank you !

Need	Approach	Benefit	Competition
Who needs it and why? Anyone who is looking for a job, especially new graduate student who doesn't know how their majors and knowledge can fit in the needs of Industries. It also can be used by HR to better allocate talents in company.	What approach will be used? We will apply web scrape and analytical skills to classify the job, salary and skills, then present the visualized result. (build a website if applicable)	What is the benefit? Our system can work as a career counselor that help people find the right jobs. The relative information about salary scales, needed skills can also support people to build their career paths.	What is the next best alternative if this project was not done? (edit) Career counselors Glassdoor

Why is this a good space?	Online project folders, documents:
How will others in the industry react? We believe that our product could be a good complementary to online job boards or physical career center appointment.	https://drive.google.com/open?id=1Ld7jE-pRI_rzck2o1iI9lnpwIIEvRAIO https://www.linkedin.com/pulse/how-easy-scraping-data-from-linkedin-profiles-david-craven/

How will you win?

• Answer:

1. Prepare all the required data including online data source and web scraping (Completed!)
2. Data cleaning of online data source (Completed!)
3. Adopting collaborative filtering (In progress)
4. Web UI design (In Progress)

What is Working/Known:

• Answer:

We have successfully collected data from online dataset, and we also leverage web scraping technique to acquire some useful data to be features of our model. Initial data cleaning and website UI design are in process.

What is Not Working/Unknown:

• Answer:

There are two main concerns at current stage. First, we are discussing about the scope of project. Should we focus on a specialized area of job recommendation and extend our goal afterwards, or build a prototype with broader function? Second, we are still learning about the recommendation model and other potential skills we might use.

Reflection 1:

• Answer:

Although it seems that we have already completed lots of task, we find that there are still many things to conquer in the future. Thus, time management and milestone timeline might be important.

Reflection 2:

• Answer:

Our team is devoted to have our first MVP ready by the first two week checkpoint. Sometimes it is challenging to decide at what extent will our data be enough to have a viable collaborative filtering result.

Avisor/Manager:

• Answer:

Log Date : 3/9/2020

How will you win?

- Answer:
 1. Initial Web UI design (Completed!)
 2. First MVP Development (In Progress)
 3. Collecting extra data to support job recommendation for other majors to explore final deliverables market(Not started)

What is Working/Known:

- Answer:

We are currently exploring analysis models that could be used in the recommendation system, starting with collaborative filtering.

Besides, we starts to connect database with backend systems (analysis programs) and front-end system (website).

What is Not Working/Unknown:

- Answer:

Current model is limited to IEOR major. If we are expanding our model to other majors, we might need additional data to train our model.

Reflection 1:

- Answer:

Since datasets have been obtained and cleaned, we are prepared to input data into the algorithm. Collaborative filtering algorithm is commonly used for recommendation systems. We plan to do more research on the topic and at the same time work on the UI design.

Reflection 2:

- Answer:

We will need more technology knowledge to support us on web page connection with backend modeling and database.

Avisor/Manager:

- Answer:

Log Date: 3/20/2020

How will you win?

- Answer:
 1. Allow user input information and pass on to result page (Completed!)
 2. Feature selection and further dataset cleaning (Completed)
 3. Build 6 machine learning models, improve accuracy from 36% to 53% (Completed)
 4. First MVP Development (80% completed)

What is Working/Known:

- Answer:

A baseline model has been completed, and we will continuing explore and improve our models. Also, we might start building our another book/online course recommend system.

Beside, we are working on passing user inputs from HTML to Python API with a simple way.

What is Not Working/Unknown:

- Answer:

In model building and MVP integration part, we face some difficulties. First, we are finding new way to improve our model, including different feature selection and models.

The method of incorporating user input and pass on to our models and display results requires both html and Javascript knowledge which requires more investigation.

Reflection 1:

- Answer: Currently we are working on passing multiple variable from HTML pages to python so that we can run local models

Reflection 2:

- Answer: In order to give more accurate and supportive recommendations, we will need to webscrape more information from sources pages like LinkedIn or Payscale to provide more skill sets to help job hunters learn about what to learn,

Avisor/Manager:

- Answer:

Log Date: 4/6/2020

How will you win?

- Answer:
 - Can successfully get user input information and pass on to result page (Completed!)
 - Can pass user information as input to local model and return model's results to result page (Completed)
 - Explored logistic, decision tree, KNN, random forest, SVM, Perceptron, XGBoost (Completed)
 - First MVP Development (90% completed)

What is Working/Known:

- Answer:
 1. We are able to record user input and assign different input to parameters and display in the result page.
 2. Compared different models and will choose the best one to blend into our recommendation display.
 3. Using web scraping to get basic skill sets of different job positions from PayScale.

What is Not Working/Unknown:

- Answer:
 1. For the modeling part, we have optimized performance, however, it's still unsatisfactory. After discussion, we think the main problem is the limited data source
 2. For the web page, we are not able to handle user input errors (e.g. missing values, wrong format). Will add another input verification to avoid crashing result page.

Reflection 1:

- Answer: To choose the best classification model to use for job recommendation, and the best prediction model to use for recommended salary range .

Reflection 2:

- Answer: Currently, we can only recommend one job at a time due to the limitation of our model's output. Based on Prof. feedback, we will try to increase the number of jobs we that can recommend from 1 to 3. Also, we are currently rearranging the questions displayed on the web page to make it more reasonable.

Avisor/Manager:

- Answer:

Log Date : 4/17/2020

End of Project Checklist:

To do list / assigned name:

- Finding data source (All)
- Web scraping (Yijun, Sean)
- Data cleaning/wrangling (Kevin, Xinyu)
- Model selection (Kevin, Xinyu)
- Job classification model (Kevin, Xinyu)
- Salary classification model (Kevin, Xinyu)

To do list / assigned name:

- UI design (Yuanyuan, Sean, Yijun)
- Database connection (Yuanyuan, Sean, Yijun)
- Web page connection (Yuanyuan, Sean, Yijun)
- Final combination (Yuanyuan, Sean, Yijun)

