STEVENS
INSTITUTE of TECHNOLOGY
THE INNOVATION UNIVERSITY®

BIA 679 – Team 2

Team Project Plan

For the scope of this project, we mainly planned to use 7 weeks/milestones to finish, in total 105 working hours, which is 15 total working hours per week/milestone, and for each team member 5 working hours per week/milestone.

We are planning to use GitHub (https://github.com/kaylaxzy/Big-Data-Practicum) to share code and Jira (https://kaylaxu.atlassian.net/jira/software/projects/T2/boards/1/roadmap) to track our project process.

1. Week 1 - getting raw data source for salaries

   Getting the raw data source ready for this project, preparing the whole plan and scope for the project.

2. Week 2 - compare data sources that we get

   Choose the dataset we are going to use and combine the useful dataset into one big dataset

3. Week 3 - get online Flatform ready for use

   Upload dataset to Flatform. Setup Flatform and report with Flatform weekly

4. Week 4 - start to check on and apply different models

   Choose the suitable models such as regression model to make salary prediction and classification model to make career forecast for data scientists. And search for other models that are reliable to fulfill our analytical purpose.

5. Week 5 - add performance measures

   Analyze the data by using linear regression model. Split the dataset into train set and test tests. Test the model and get the result of performance.

6. Week 6 - compare model performance

   Compare the accuracy result from each different regression models. Making the conclusion of which model has the best test performance.

7. Week 7 - conclusion and visualization

   Applying for Tableau to visualize the performance of each model and decide to choose the model which has the best performance

Dataset Website:

1. Data Analyst Analysis (7.57MB)

https://www.kaggle.com/datasets/andrewmvd/data-analyst-jobs

2. Data Scientist Analysis (3.12MB)

https://www.kaggle.com/datasets/nikhilbhathi/data-scientist-salary-us-glassdoor

3. 100+ Insights Data Science Jobs EDA

https://www.kaggle.com/code/nikhilbhathi/100-insights-data-science-jobs-eda/notebook

4. Simple Linear Regression for Salary Data

https://www.kaggle.com/code/vivinbarath/simple-linear-regression-for-salary-data/notebook