Data Scientist Jobs Analytics

Project Milestone Project Proposal

Student Name Zhao Zhang

Email [zhang.zhao2@northeastern.edu](mailto:zhang.zhao2@northeastern.edu)

Contribution Zhao Zhang – 100%

Submission Date June 10th, 2021

Proposal setting: Data scientist is said to be the sexist job in 21st century. More and more companies start to have increased number of opening data scientist positions. Different fields such as IT, finance, medicine all consider data scientist now is an indispensable position. As data scientist is a well-rounded job which combine machine learning knowledge with statistics and math, sometimes also business together, it enjoys the reputation of job of 21st century of high salary, meanwhile most positions have very high requirements for their candidates. Many new graduated students from different background are pursing to find a data scientist position but facing challenging time to find an ideal job. Identify the key features and qualifications for data scientist positions could be helpful for those students’ career.

Problem Definition: Find out what skills and education levels and other background are helpful to be a data scientist. Also compare different types of data scientist positions for the qualifications and treatments. Specific objectives are listed below:

Explore skills required for data scientist jobs from job descriptions.

Predict salary based on industry, location, company revenue, job title

Predict rating based on industry, location, company revenue, job title, salary

Figure out which type of jobs are easy to apply

Data Sources: Data Scientist Jobs Datasets on Kaggle. The license is not specified, and it is open for conducting tasks and doing date analytics.

reference link: <https://www.kaggle.com/andrewmvd/data-scientist-jobs>

Data Description: It is a dataset consists of 3909 jobs listing for data scientist roles scraped from glassdoor. It is a csv file with 3909 rows and 17 columns. The variables include important features that we need for analyzing such as job title, company name, location, rating, salary estimate and so on