

ECE 143
Group 3
Shawn Winston
Jingyang Li
Pingchun Chiang
Chen Zhao

Group 3 Proposal: Predicting Job Salaries

Problem:

We want to learn from job data in order to predict job salaries based on job details and job descriptions. This would be useful for people looking for jobs and wanting to know what their potential salaries would be before applying. This would also be useful for prospective college students that want to determine what skill set they should acquire in order to maximize their future personal value.

Dataset:

We will be using this Kaggle dataset
(<https://www.kaggle.com/c/job-salary-prediction/data>).

This dataset includes over 240,000 training dataset entries and also includes a separate test and validation dataset. The dataset only includes data from the UK so it is unlikely that we would be able to predict salaries accurately for jobs outside of the UK after training. However, the data will still be useful to see how job titles and categories compare relatively on salary.

The most useful categories of data that are provided to us are:

Job title, job description, job location, company, category of job, salary, and website name where job was found.

Proposed Solution:

We plan on trying various machine learning techniques in order to find the best technique that can predict job salaries based on the data most accurately. We will try a few of the following machine learning techniques such as nearest neighbor, linear regression, logistic regression, SVR, random forest, PCA, SVD, etc. and pick the best predictor.

Project Steps:

Steps	Estimated completion time	Person(s) in charge (among the group of 4)
1. Extracting and cleaning up data	One week	
2. Data visualization (to obtain data statistics) and coalesce data to see trends	One week	
3. Salary prediction based on categories	Two weeks	
4. Aggregate previous steps to put into a formal presentation	A few days	