Levels Forecaster
by Nicholas Farkas

Background & Overview

Levels.fyi is one of the most popular salary tracking web applications especially among engineers. The concept of the website is simple, individuals post their job title, company, compensation information, ect. Then Levels.fyi will display the compensation information to the world with anonymity. The one thing levels doesn't have, is a tool for is to predict the wages of a specific job.

Website being mined for data:
https://www.levels.fyi/

Example of Data Table
https://www.levels.fyi/t/software-engineer?countryId=254&country=254&offset=10

Data Collection and Aggregation

This year I want to try something new and web scrape this data from levels.fyi. Luckily levels.fyi's html is formatted in such a way that makes it perfect to scrape. I will scrape grouping by job. Starting with Software Engineer and tie in as many different jobs as time allows for. The features I will be selecting for the project are all contained within a row in the Salary table on each job's individual page. It also displays 1000 different postings under a category so I will collect 1000 rows for each job.

This will provide us with at least 7 predictors options for each job which include:

- Company (Binned, Categorical)
- Industry (Categorical)
- State (Categorical)
- Posting Date (Numeric)
- Years of Experience Total (Numeric)
- Years of Experience at hiring company (Numeric)
- Level Name (Binned, Categorical)

Models and Validation

I plan to use as many different data modeling techniques as possible and show the accuracy of each one. We will use the at least 7 predictors to predict total compensation. The two data

models are not what I am limiting myself too but are the two that I am committing to and will pick other models as I learn more about my data. (I know we haven't covered Neural Networks for Regression yet, but I will bring some prior experience from your data science class and self-study).

1. **Linear Regression**
2. **Neural Network**

Note to Professor: Please tell me any flaws you see in my plan, as well as things I have forgotten to plan for. I had a hard time covering everything that I need to be planning because there is a lot to plan for. Thank you!