Proposal for Final Capstone Project

Problem Statement

# While looking for jobs, one important aspect is whether the salary meets job seeker’s expectation. As a result, job seekers tend to spend time on researching salaries to get a general idea of how much their skills and experience worth. To ease this process, this project will build an algorithm for a job search application. Specifically, when a job seeker identifies target job title, employment length and location, this app generates an estimated annual income. In addition, this project will investigate the importance of job title for predicting salary and provide implications. This solution is valuable not only for job seekers to gauge their salary expectations, but also for recruiters to evaluate their job offers.

Data Source

# The Lending Club dataset (downloaded from [Kaggle](https://www.kaggle.com/wendykan/lending-club-loan-data)) will be used for this project. This dataset contains rich information related to job, location and annual income. Most recent records will be extracted from this dataset. Variables including job title, state, zip code, employment length and annual income will be used in the analysis.

Goals of Analysis

Data cleaning/processing, data visualization and machine learning algorithms will be applied to this project. In addition, natural language processing (NLP) techniques will be used to analyze the semantics of job title and its relationship with annual income. Supervised learning, unsupervised learning and neural networks will be implemented to find the best possible model for the solution.

Challenges

As features will be the numeric representations of natural language variable and categorical variables, a potential challenge is the computational power especially when fitting neural network models. In addition, this project aims to use limited information to predict annual income which could present a challenge to modelling and the outcome.