ZhaoDong Wang   Feb 1st, 2020

<u>**Project Deliverable 1**</u>

1. Dataset

This project will attempt to predict what jobs people will apply given their previous applications, demographic information and work history. I will be using the Job Recommendation Challenge Dataset from Kaggle because it has all of the necessary components (i.e. job titles, users, users' background) to do a job prediction.

2. Methodology

I. Data Preprocessing

Since the users and its jobs are each linked with an id, each predication will have array of ids where the ids associated are job title. Each users' experience and skills will be lowercased, split into single word, and all punctuations will be removed.

II. Machine learning model

Since I want to predict what jobs people would want to apply for based on their background, this project will try to replicate two of Chen's recommendation system models: a hierarchical gating network (HGN) and a gated attentive-autoencoder (GATE) model. The HGN selects what item features to pass to the downstream layers and the GATE model is able to exploit neighboring relations between items to help infer users' preferences. These two models will be a challenge for me to test which model will give a better result, thus, pros and cons will be discussed in the future deliverables.

III. Final conceptualization

To present this model, I will be demoing a simple web/mobile application that will implement the model by taking inputs of a user's career background and then displaying the suggested jobs for him/her to apply for and saving the predictions to a database.

References:

- **Chen Ma**, Peng Kang and Xue Liu, "**Hierarchical Gating Networks for Sequential Recommendation**", in the 25th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (**KDD 2019** Research Track, acceptance rate: ~170/~1200=14%), Anchorage, USA, Aug. 2019.
- **Chen Ma**, Peng Kang, Bin Wu, Qinglong Wang and Xue Liu, "**Gated Attentive-Autoencoder for Content-Aware Recommendation**", in the 12th ACM International

Conference on Web Search and Data Mining (**WSDM 2019**, acceptance rate: 84/511=16%), Melbourne, Australia, Feb. 2019.