

Final Project: Proposal

Btara Truhandarien (301360202), Mohammad Wali Ullah (301369145), Grace Vino Kingsly Kumaresan (301375997), Bhuvan Chopra (301377228)

Our project is related to the job market and people's skillsets. The ultimate goal is to be able to give advice to a person with skillset A and a desired job, the required skillsets that person must grow/improve/develop in order to be eligible for the job. The advice would be given by a machine learning model. For the purpose of this project we will scope our findings and research to the Canadian job market.

1. List 3 questions that you intend to answer (1 point)

Here are some of the questions that we wish to explore along the way:

- What are the common skillsets that most people have for some kind of jobs? Likewise, what are uncommon skillsets that some people do have holding that same job?
- How tightly correlated are people's skillset to their jobs? That is, is there statistical significance that certain skillsets are more or less important for certain jobs. Involves finding distribution and some sort of correlation measure
- Is there a good machine learning approach for the goal stated above? We will be trying some techniques and see if they will work out.

2. List all the datasets you intend to use (1 point)

We have a variety of data sources we are looking into. These data sources comprise of job boards, resume repository, APIs and openly available data:

- resume.indeed.com (resume repository)
- glassdoor.com (job boards)
- <http://dataatwork.org/data/> (API)
- <https://www.onetcenter.org/database.html#individual-files> (openly available data)
- <https://data.world/peopledatalabs/similar-skills-28935-unique-skills> (openly available data)

3. Give us a rough idea on how you plan to use the datasets to answer these questions. (2 points)

For our predictive model we would like to focus on resumes as much as possible with job boards as a second set of data points. This is because we want a model that is a reflection of actual workers, not an ideal description of a worker for a job. Due to the varied nature of job and skills descriptions we hope to be able to use the resources above and more to be able to come up with a limited set of skills description job titles. In some more detail we would like to:

- Data Collection: Using the above resources to gather necessary data points, particularly

resumes and job boards

- Data Cleaning and Integration: To help us create a set of common skillsets and linking different ways to describe the same descriptions to these common skillsets.
- Data Integration: We will use all available job portals (that we can extract data from) as our data source as a result, it is quite possible that same posting or resume will show up in multiple job portals. Hence, we need to have a data integration method to deal with these kinds of issues.
- EDA: To help us explore and answer the questions we have listed above and more
- Modeling and feature engineering: Using the cleaned and integrated data points we would use them as the baseline for our features and use them to train our model

4. Think about that once your project is complete, what impacts it can make. Pick up the greatest one and write it down. (1 point)

These days more and more jobs require a more sophisticated and intricate blend of various skills, these jobs include highly coveted ones like data scientists, programmers, researchers, engineers, startup founder and more. This is predicted to be the case as more and more of AI seep into the workforce, transforming the landscape for mundane, repeatable and highly repetitive work.

Many believe the only way to overcome skill redundancy is to move over to jobs that are more creative and believed to only be able to be done by humans. With this narrative, many EdTech companies are popping up, providing and spreading expert and practical knowledge on various areas. Universities are trying to stay relevant with the growth of EdTech. At the same time, companies are pushed even more to be more innovative and to be ready for the upcoming technological shift.

The ability to give sound and meaningful advice to a person (or groups of people) in a quick and efficient way for their future career planning is monumental for the upcoming skill-oriented workforce. Company HRs will be better able to plan and construct training programs for their current workforce for a targeted goal. EdTech companies can better supply quality consulting service. Universities will be able to reassess strategically their future course offerings according to students' interests and state of the job market. All of this done with data powering them.