SKILLS JOB ADVISOR

Bhuvan Chopra – Btara Truhandarien – Grace Kingsly – Mohammad Ullah
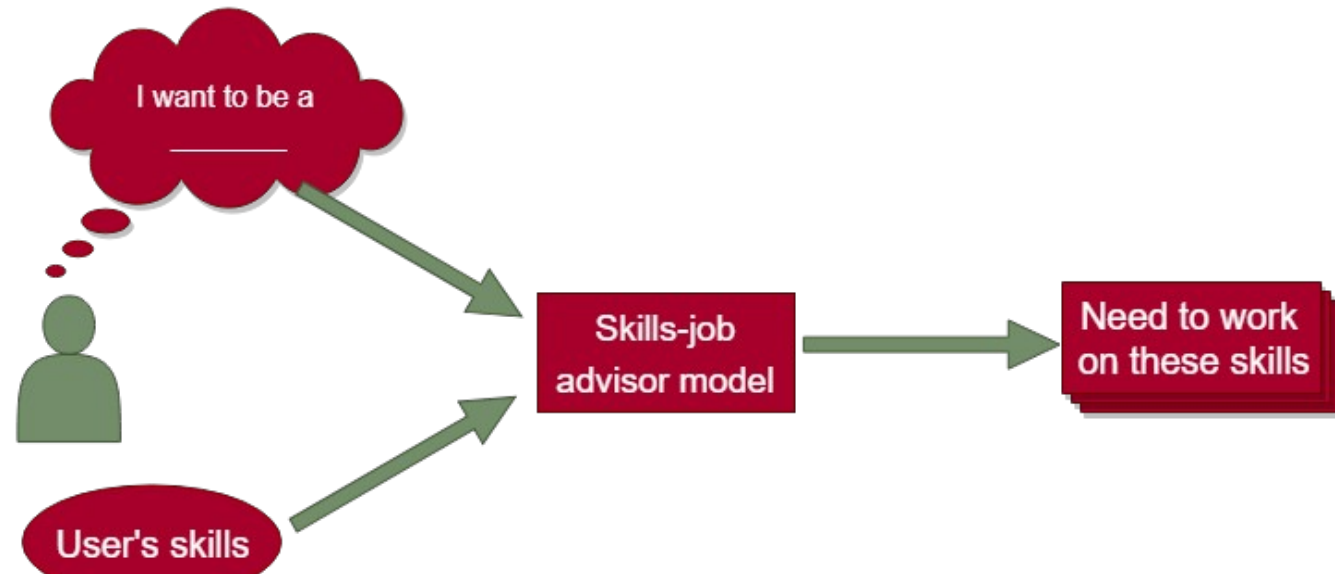
ABSTRACT

At times we often want to know what is required to achieve a position in an organization. Our project advises people the skills they require in order to get their desired job. For this, we collected data of 15 diverse job titles specific to Canadian job market to train a machine learning model. These job titles are chosen based on the number of resumes yielded in our preliminary search.
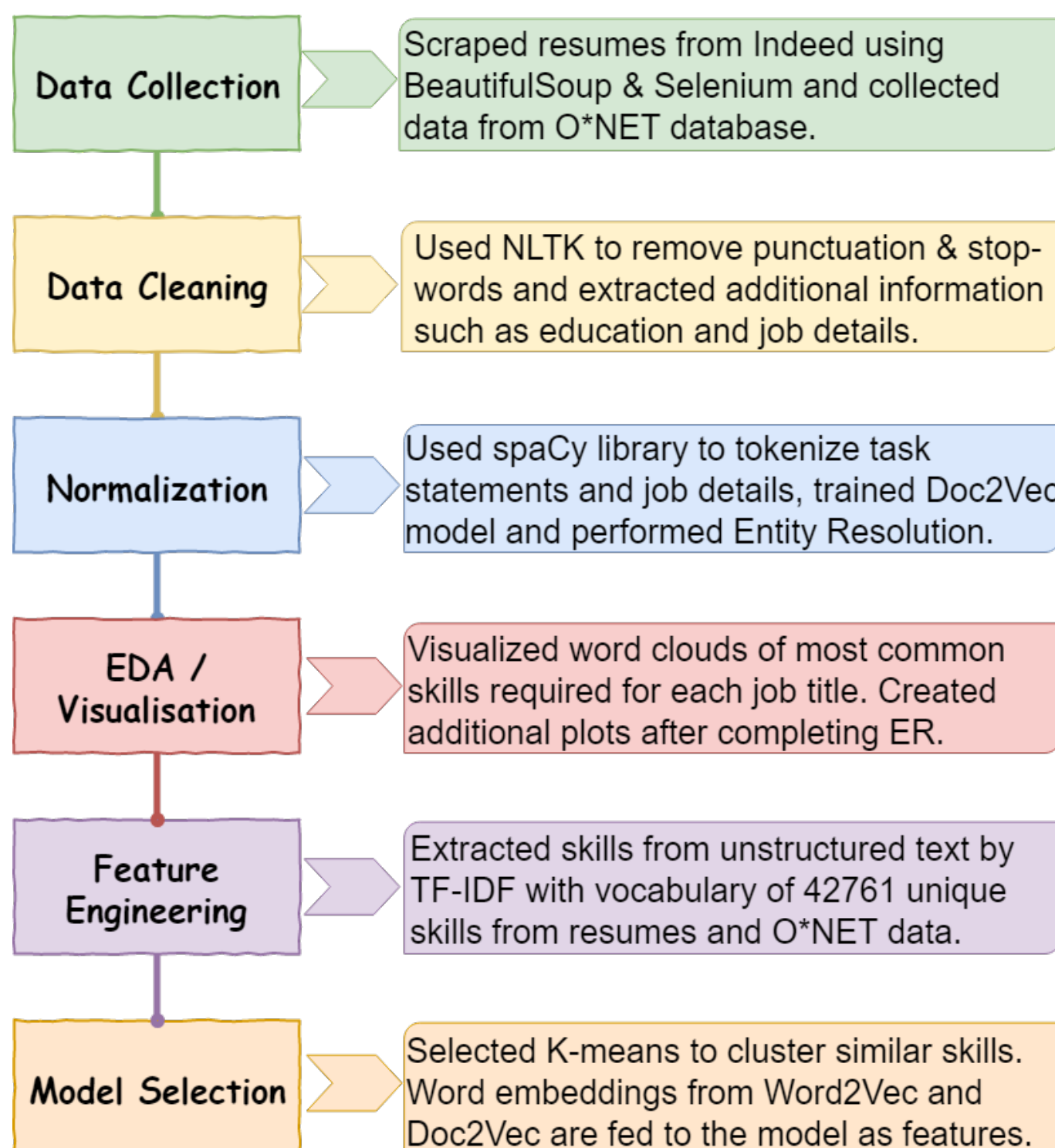
IMPACT

The ability to give sound and meaningful advice to people in a quick and efficient way for their future career planning is monumental for the upcoming skill-oriented workforce. With this project:

- Organization's HR managers can improve their recruitment by selecting the right candidate for the job. They can also improve their employee training programs.

- Universities and EdTech companies can make their courses more relevant to industry needs.
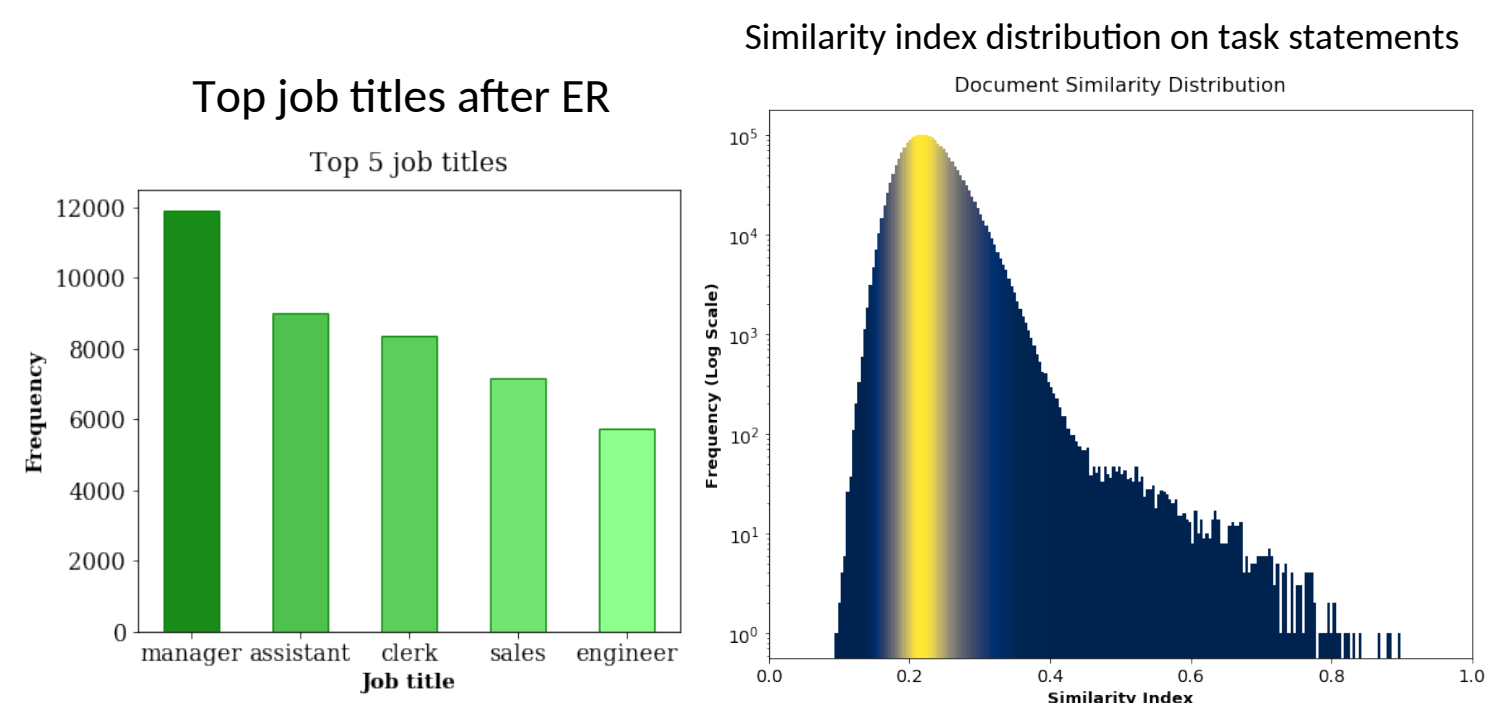
METHODOLOGY

We have focused on resumes instead of job descriptions because we wanted to get real information of people and not information of an ideal candidate.

Data Collection → Scraped resumes from Indeed using BeautifulSoup & Selenium and collected data from O*NET database.

Data Cleaning → Used NLTK to remove punctuation & stop-words and extracted additional information such as education and job details.

Normalization → Used spaCy library to tokenize task statements and job details, trained Doc2Vec model and performed Entity Resolution.

EDA / Visualisation → Visualized word clouds of most common skills required for each job title. Created additional plots after completing ER.

Feature Engineering → Extracted skills from unstructured text by TF-IDF with vocabulary of 42761 unique skills from resumes and O*NET data.

Model Selection → Selected K-means to cluster similar skills. Word embeddings from Word2Vec and Doc2Vec are fed to the model as features.

Common skills for each job title

Comparision against O*NET database

Top job titles after ER

Similarity index distribution on task statements

MODEL SELECTION

This is an unsupervised machine learning problem. We have selected K-means clustering algorithm to cluster the skills and job title vector embedding, with cosine similarity. Doing so we get clusters of jobs and related skills. With the help of t-SNE, the following plot shows similar skills (blue) and job (red) clustered together for Data Scientist and few random words (green) far from the cluster.

t-SNE visualization for Data Scientist