Societe Generale Hackathon

Team Members: Divya Sivaraman, Anish Kumar K

Topic: Matching of Job Openings with Skill Database using Machine Learning

Dataset Creation

The first step towards going about the problem statement was to create a dataset. We had created a dataset with 3 columns namely - Job Title, Required Skills, Responsibilities.. A new column called skill_list is introduced which comprises all the skill sets as a list for a certain record.

Data Preprocessing

On which, various data preprocessing steps were performed using nltk library such as removing HTML tags in between, stopword removal, lemmatization etc.

Modeling

In order to perform classification, we made use of supervised learning algorithms as we are already well aware of the class labels i.e. Job title for the given skill set. After performing vectorisation and the train test split, the following are the supervised learning models that were made us of to determine the model that was best suited for our problem:

- Support Vector Machine
- Naive Bayes Classification
- K nearest neighbor
- Random forests

Evaluation

On the basis of accuracy, which was the evaluation metric that we chose, we could conclude that we attained the maximum accuracy for the Naive Bayes classifier model.

Resume Matching

Using the joblib library we are saving the model that produces the highest accuracy as a .pkl file. Through the UI, which was built using streamlit, the resume will be uploaded as a .pdf or .docx file from where the candidate's skill set is extracted and passed on to the model as input parameters, after which the model runs to predicted the best suited job title for the given skillset.