Project Proposal

JobFitt

April 10th 2018

Team #5 - Athena

Anu Khandelwal

Aiyana Liz Mathew

Aruna Amaresan

Aswathy Mohan

Pallavi Donwad

Overview

Team Athena is looking to build a job fit recommender and/or classifier system to indicate top 5 highly likely position titles with mapping salary that would fit a given job seeker's current skills, education, education level, years of experience and other factors.

Business Problem

With the changing technology trends with automation, analytics, cloud computing and machine learning becoming more popular, it is challenging currently for a fresh grad or an experience job seeker to understand the kinds of titles and salaries that would actually map to their current skills. They end-up feeding the same info to multiple sites and apply on different sites. It would help:

a) The job seeker to understand the value of their skills more and target their job application process.

Eventually, this would help the entire recruiting ecosystem that includes company career sites, job boards, applicant tracking systems, and staffing agencies to improve site engagement and candidate conversion.

Goal

The goal is to provide an app where a user would input info on skills, major, college (last attended), education level. These would form the independent X variables. Give this info, we would provide highly likelihood of title and salaries that map the user's skills. This would be the dependent Y variables.

To do this, the plan would be to train a model using random forest algorithm and/or clustering with datasets found and/or using API calls to provide highly likelihood of title and salaries that map the user's skills.

High-level plan

1. Our goal is to first tackle sharing likely recommendation of title and mapping salaries for fresh Grads

If time permits, we would like to:

2. Apply a new model with training and test set to train for experienced job seekers.
3. Take this to the next step where:
    a. the app would suggest positions available right now in careers websites that map the title and salaries that were suggested and/or
    b. Find what skill gaps to be updated to achieve the job position needs for a job seeker.

Tasks to be done:

a) Cleansing Data
b) Training and testing model (machine learning) – Mostly 2 models (Random Forest and TBD)
c) Front end – Index HTML page (visualization and then form based input)
d) Front end - About Page
e) Front end - Visualizations page for historical data info and/or API data
f) Flask App -> to call on API
g) Logic to ensure form info is submitted back into testing set
h) Mongo DB backend

i) Visualization and plots to show on front end web pages using D3, plotly and/or leaflet
j) APIs to call from monster or some other website

Tools to be used:

a) Machine learning using Natural language processing classifier APIs in google or TBD and random forest algorithm
b) MongoDB for backend database
c) Python, Pandas and Flask, PyMongo for hosting app, routing endpoints and connecting with Mongo DB backend
d) For Front end – Bootstrap , CSS and visualization using D3, plotly and leaflet.
e) Connecting with monster APIs (TBD)

Dataset links and other links:

For college fresh grad:

a) https://www.kaggle.com/wsj/college-salaries
b) https://data.world/fivethirtyeight/college-majors/workspace/file?filename=women-stem.csv
c) https://data.world/fivethirtyeight/college-majors/workspace/file?filename=recent-grads.csv

For experienced job seeker dataset:

a) https://www.kaggle.com/madhab/jobposts/data
b) https://www.onetcenter.org/dictionary/22.2/excel/skills.html
c) https://www.kaggle.com/PromptCloudHQ/us-jobs-on-monstercom (TBD)

Natural language classifier:

a) From google: https://cloud.google.com/natural-language/docs/reference/rest/
b) From IBM: https://console.bluemix.net/docs/services/natural-language-classifier/getting-started.html#natural-language-classifier

Other useful links:

a) https://ikdd.acm.org/Site/CoDS2016/datachallenge.html
b) http://www.nltk.org/book/ch06.html