FIT5145 Assignment – 2

Case Study on Data Science at Indeed

Jaimon Thyparambil Thomas
Student ID: 29566428
Email: jthy0001@monash.student.edu
Monash University
September 16, 2018

Contents

Project Information

Introduction

This report is an attempt at analysing the data models and the technologies employed in 'Indeed', a popular online job portal for handling huge amount of data.

What is Indeed?

Indeed, is one among the most popular site for job portal spread across 60 plus countries available in 28 different languages covering [1]. Their motto is helping people get jobs. Which involves gathering all available jobs worldwide in one website, by crawling over 100+k websites and based on the information provided by the employers. Through this site they help job seekers to search for jobs, post their resumes and research on various companies. Thus, every day they try to connect millions of people to new opportunities.

Relation to Data Science

In indeed about 9.8 jobs are added every passing second with around 200 million unique visitors. Given the enormity of data handled, data science is included at its very core. They use data science in how the rank search results, recommend positions and various skills for job seekers, in estimating salaries for jobs, in providing tips for employers for creating great job descriptions. [2]

Data Roles

Some main roles performed are listed below [3]

Data scientist

- Analyse, visualize and model job search related data.
- Will have to build machine learning models to improve performance
- Should have the skills and knowledge of a statistician and machine learning expert
- Should be proficient in Python/R, Spark etc and their scientific libraries.

Data Engineer

- Skilled in extracting, transforming, and loading data
- Integrate with diverse APIs
- Will have work along with analyst, data scientists as well as the data supply side
- Will have to work with relational database

Product Scientist

- Use data mining, and machine learning techniques to understand how job applicants and suppliers are interacting on Indeed, and how it is reflected in data.
- Execute statistically sound tactics for evaluating web pages across business related metrics.
- Work in expanding Indeed's set of tools and techniques used for manipulating and interpreting huge amount of product data.

Business Intelligence Analyst

- Should be able to create dashboards, visualize data, find algorithms, and business tools to use across the organization.
- Should Also have profound knowledge in Python (particularly Pandas and NumPy), R, SAS programming experience.
- Should have an aptitude to analyse trends, compile data into comprehensive reports, and making recommendation based on data

Business Model

Nowadays unemployment is an acute problem that societies face. In this scenario, job portals like Indeed makes transformative changes by recommending the most eligible job seeker to the best possible employers. From a business perspective, indeed aims to increase click rates in general and conversion of profile matches to successful placements. To achieve these aims, they decided to go beyond basic job searches and added functionalities like job recommendation as a new mode of interaction. As they found out that

- One fourth of the searches in Indeed specified only location without any keywords. As there are many job seekers who don't know what keywords to search for.
- We also know that as we provide recommendations users start feel more personalised experience and find more desired results

As per basic work flow, they create lots of classifiers based on jobs description and resume. For instance, one of the classifiers tries to find minimum years of experience. Also, to differentiate between skill sets like expert and beginner.

Based on these classifiers they try to find out matching cases. That is when job seekers' resume is consistent with the pre-requisite skill sets expected by an employer, a 'match' is obtained. In that case [4]

- For both employer and the job seeker a text 'match' (as a highlighted text) will be shown for matched applications.
- Similarly, the same recommendation system is used for suggesting recommended jobs for the job seeker based on his resume.
- Also, while sending mail for the job seeker suggesting that there is a job vacancy.

As the recommendations becomes better the click rates will increase. It will also help in providing more successful interviews

Some of the main factors that they took into consideration while processing the data are [5]:
- Recommendation set keeps on changing as new jobs keeps on coming.
- Everyday 200+ million unique users visit the site. As a result, they should be able to recommend with limited amount of user data.
- Contents freshness matters as compared to the newer jobs the older jobs has more chance of already been filled
- Jobs will have only a limited supply of seats. So, if we keep on recommending the same job for everyone. The employer will be flood with applications.

Challenges Faced

Scaling into different countries

Initially, indeed was available only for US job market. Now it is available in more than 60 countries and 28 languages. To achieve this, they provided different domain based on countries. Like indeed.com for US and indeed.ca for Canada. Within each server there is an application part and configuration part. In the application they make checks based on domain to show country specific data. Thus, they implemented a system such that only one product is there which is provided in all countries [6]

Job seekers might not know correct job description

To overcome this issue, indeed used the concept of stemming of the words that the job seeker used to search. Like rather than just searching for the exact word match they also search for synonyms etc.

References

[1]https://au.indeed.com/about/our-company?hl=en

[2] https://medium.com/indeed-data-science/introducing-the-indeed-data-science-blog-7e2985fe1e92

[3] https://www.indeed.jobs

[4] https://engineering.indeedblog.com/talks/data-to-deployment/

[5] https://www.oreilly.com/ideas/algorithms-and-architecture-for-job-recommendations

[6] https://engineering.indeedblog.com/talks/internationalize-success/