

# Recommender System for Jobs

Venkatesh Tanniru (013744757)

## Abstract:

*With so many different IT-related job types in demand nowadays, it is hard for a fresh graduate or multi-skilled professional to determine where they would fit best. Several unwanted jobs appear in the sorted lists. There is a need for the recommendation system with more accuracy. The aim of this Project is to scrap the job postings from indeed.com and generate recommendations based on a set of skills. We use content-based filtering for generating recommendations.*

## Introduction:

With so many different IT-related job types in demand nowadays, it is hard for a fresh graduate or multi-skilled professional to determine where they would fit best. This problem is most apparent in today's data-driven economy, where the job descriptions and roles are closely connected and often overlapping. Several unwanted jobs appear in the sorted lists. There is a need for the recommendation system with more accuracy. In order to solve this problem, I have implemented a recommendation system. I have scrapped the data from indeed.com and stored the data in the CSV file using requests, Beautiful soup. From the scrapped data the recommendations using content-based filtering. I have used TF-IDF and cosine similarity in generating recommendations.

Our goal is to recommend the jobs based on a set of skills by filtering the jobs based on the skills and to recommend them.

## Related Work:

I used several libraries for this project. They are Requests, Beautiful soup, os, CSV and Pandas.

Requests allow you to send HTTP/1.1 requests. You can add headers, form data, multipart files, and parameters with simple Python dictionaries, and access the response data in the same way. It's powered by httplib and urllib3, but it does all the hard work and crazy hacks for you.

Beautiful Soup is a library that makes it easy to scrape information from web pages. It sits atop an HTML or XML parser, providing Pythonic idioms for iterating, searching, and modifying the parse tree.

OS module provides a portable way of using operating system dependent functionality. If we just want to read or write a file see open(), manipulate paths, see the os, path module, and lines in all the files on the command line see the file input module. For creating temporary files and directories see the tempfile module, and for high-level file and directory handling see the shutil module.

pandas is a Python package providing fast, flexible, and expressive data structures designed to make working with structured (tabular, multidimensional, potentially heterogeneous) and time series data both easy and intuitive. It aims to be the fundamental high-level building block for doing practical, real world data analysis in Python. Additionally, it has the broader goal of becoming the most powerful and flexible open source data analysis / manipulation tool available in any language.

## Implementation:

I have imported all the necessary libraries and then searched for the jobs in indeed and copied the link. I have aimed to scrap the data from 100 pages and then created a empty Jobs\_data.csv file for storing the scrapped data. Then I have scrapped the job title, company, location, salary, summary separately and stored the data in the csv file created initially. For all the 100 pages, I managed to scrap 1477 jobs. I used ‘try and except’ concept in scrapping the data. Requests, BeautifulSoup, os, CSV are the libraries used for scrapping the data. The scrapped data is stored in CSV file. Using pandas, I have read the CSV file and printed the Jobs and shape of CSV file. I then tokenize the summary and store them in another file. Then apply TF-IDF to it and store the vectors in a data frame. Then we find a set of skills for generating the recommendations. Then apply TF-IDF to it and store the vectors in a data frame. Then we find cosine similarity between the data frames and print similarity values. Then I filter

the similarity values and then print top 10 similar jobs.

## Results:

Initially the data is scrapped from 100 pages is stored in CSV format. The data is scrapped as job title, company, location, salary, summary.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	Job Title	Company	Location	Salary	Summary															
2	Software Engineer	Confidential	None	None	We have one large project that is in Go, using Google Cloud, and another project that is in PHP and also uses some Google Cloud features.															
3	Software Engineer	Honeywell	None	None	You will implement software projects (with coaching from the Lead Engineer) for an integrated system or sub-system. Minimum travel is 50 percent.															
4	Robotics	Top Flight	None	None	Develop firmware and embedded software in C/C++ and python. Java GUI development a plus. Work side by side with a world-class technology and business leadership[4]															
5	Software Engineer	Twitter	San Francisco	None	Pursuant to the San Francisco Fair Chance Ordinance, we will consider for employment qualified applicants with arrest and conviction records.															
6	Software Engineer	Perspecta	Campbell	None	Reviews and evaluates systems and software for adherence to government or commercial directives, standards, guidelines, and criteria concerning software safety[4]															
7	Software Engineer	Summyvale	None	None	Java, C/C++, Python, JavaScript, or Go. The candidate will participate in the design, development and operation of the KIR or NAC product.															
8	Software Engineer	Paypal	San Jose	None	Paypal is seeking accomplished and passionate software engineers to join our team in San Jose, CA, contributing to the exciting future of software technologies[4]															
9	Software Engineer	WorlData	Santa Clara	None	Experience with revision control software and management tools, such as git, Perforce, Jenkins, JIRA, Confluence and Maven. What we need to see:															
10	Software Engineer	Quantara	Santa Clara	None	Proficient with Python, JavaScript, C++/C, Java, HTML. Web application design is a plus but not required.															
11	Software Engineer	Facebook	Seattle	None	2+ years coding experience in C, C++, Java, C#, Perl, PHP, Haskell and/or Python. Code using primarily C/C++, Java, PHP and track. Conduct design and code reviews.															
12	Software Engineer	Apple	Santa Clara	None	Strong coding skills in one or more languages such as Java, C++, Go, or Python. At Apple, great ideas have a way of becoming great products, services, and[4]															
13	Software Engineer	University of Santa Cruz	None	None	The incumbent solves dynamic and challenging problems, learning new technologies as necessary and collaborating with colleagues and external partners.															
14	Software Engineer	VeriBit	Carlsbad	None	Previous experience or exposure in Python, Java, or Go/ing development skills and willingness to learn other languages. Interested in joining our mission?															
15	Software Engineer	Lenovo	Chennai	None	In the industry for over 25 years, Lenovo's customer needs are diverse with project requirements ranging from simple to extremely complex.															
16	Senior Software Engineer	Large	None	None	Software Engineer Experience and Qualifications: As a Software Engineer this position is responsible for testing the design and implementation of robotics and[4]															
17	Software Engineer	Confidential	None	None	We have one large project that is in Go, using Google Cloud, and another project that is in PHP and also uses some Google Cloud features.															
18	Software Engineer	None	None	None	You will implement software projects (with coaching from the Lead Engineer) for an integrated system or sub-system. Minimum travel is 50 percent.															
19	Robotics	Top Flight	None	None	Develop firmware and embedded software in C/C++ and python. Java GUI development a plus. Work side by side with a world-class technology and business leadership[4]															
20	Software Engineer	Netflix	Los Gatos	None	Strong in Java experience, proficiency with Ruby on rails OR another object oriented programming language. Proficient in HTML5, CSS, and JavaScript frameworks [4]															
21	Software Engineer	Proteus	Seattle	None	Experiences in developing backend large scale distributed services and data processing workflows in Java and Python.															
22	Embedded Software Engineer	Facebook	Menlo Park	None	We are seeking an Embedded Software (Firmware) Engineer to support development of ASICs custom-designed for computer vision and machine learning.															
23	Software Engineer	Intel	Intel	None	As the world's largest chip manufacturer, Intel strives to make every facet of semiconductor manufacturing state of the art - from semiconductor process[4]															
24	Software Engineer	Intel	San Jose	None	Is a leading provider of embedded control applications. Good oral and written communication skills. A broad technical background in relevant areas will be very[4]															
25	Software Engineer	Intel	San Jose	None	Java, C, C#, Perl, Python. In this position, the selected candidate will design and implement and deliver software to enable efficient project execution of [4]															

Then we load the CSV file using Pandas. Then print the head of the CSV file.

	Job Title	Company	Location	Salary	Summary
0	Software Engineer	Confidential	None	None	We have one large project that is in Go, using...
1	Software Engineer Level 1	Honeywell	None	None	You will implement software projects (with coa...
2	Robotics Software Engineer	Top Flight Technologies	None	None	Develop firmware and embedded software in C/C+...
3	Software Engineer	Twitter	San Francisco, CA 94103 (South Of Market area)	None	Pursuant to the San Francisco Fair Chance Ordi...
4	Software Developer I	Perspecta	Campbell, CA 95011	None	Reviews and evaluates systems and software for...

Then I have tokenized summary column and stored in a data frame.

	Job Title	Company	Location	Salary	Summary	words
0	Software Engineer	Confidential	None	None	We have one large project that is in Go, using...	[We, have, one, large, project, that, is, in, ...]
1	Software Engineer Level 1	Honeywell	None	None	You will implement software projects (with coa...	[You, will, implement, software, projects, wit...
2	Robotics Software Engineer	Top Flight Technologies	None	None	Develop firmware and embedded software in C/C+...	[Develop, firmware, and, embedded, software, i...
3	Software Engineer	Twitter	San Francisco, CA 94103 (South Of Market area)	None	Pursuant to the San Francisco Fair Chance Ordi...	[Pursuant, to, the, San, Francisco, Fair, Chan...
4	Software Developer I	Perspecta	Campbell, CA 95011	None	Reviews and evaluates systems and software for...	[Reviews, and, evaluates, systems, and, softwa...

Then I applied TF-IDF for the data frame and stored the values in a data frame.

```

0 0.019649 0.000000 0.0 0.0 0.0 0.0 0.000000 0.000000 0.0 0.0 ...
1 0.018275 0.000000 0.0 0.0 0.0 0.0 0.000000 0.000000 0.0 0.0 ...
2 0.015187 0.000000 0.0 0.0 0.0 0.0 0.000000 0.000000 0.0 0.0 ...
3 0.013838 0.000000 0.0 0.0 0.0 0.0 0.000000 0.000000 0.0 0.0 ...
4 0.011591 0.000000 0.0 0.0 0.0 0.0 0.000000 0.000000 0.0 0.0 ...
5 0.014888 0.000000 0.0 0.0 0.0 0.0 0.000000 0.000000 0.0 0.0 ...
6 0.012718 0.000000 0.0 0.0 0.0 0.0 0.000000 0.000000 0.0 0.0 ...
7 0.011823 0.000000 0.0 0.0 0.0 0.0 0.000000 0.000000 0.0 0.0 ...
8 0.019767 0.000000 0.0 0.0 0.0 0.0 0.000000 0.000000 0.0 0.0 ...
9 0.009113 0.000000 0.0 0.0 0.0 0.0 0.000000 0.000000 0.0 0.0 ...
10 0.014896 0.000000 0.0 0.0 0.0 0.0 0.000000 0.000000 0.0 0.0 ...
11 0.012386 0.000000 0.0 0.0 0.0 0.0 0.000000 0.000000 0.0 0.0 ...
12 0.013882 0.000000 0.0 0.0 0.0 0.0 0.000000 0.000000 0.0 0.0 ...
13 0.014228 0.000000 0.0 0.0 0.0 0.0 0.000000 0.000000 0.0 0.0 ...
14 0.020125 0.000000 0.0 0.0 0.0 0.0 0.000000 0.000000 0.0 0.0 ...
15 0.019649 0.000000 0.0 0.0 0.0 0.0 0.000000 0.000000 0.0 0.0 ...
16 0.018275 0.000000 0.0 0.0 0.0 0.0 0.000000 0.000000 0.0 0.0 ...
17 0.015187 0.000000 0.0 0.0 0.0 0.0 0.000000 0.000000 0.0 0.0 ...

```

Then I have tokenized the skillset array.

```
['java', 'python', 'javascript', 'C++', 'C', 'R']
```

Then I applied TF-IDF for the skillset array and stored the values in a data frame.

```

0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 ... written mww year \
1 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 ... 0.0 0.0 0.0
2 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 ... 0.0 0.0 0.0
3 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 ... 0.0 0.0 0.0
4 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 ... 0.0 0.0 0.0
5 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 ... 0.0 0.0 0.0

yearn years yes yewno zullily équipe éventail
0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
1 0.0 0.0 0.0 0.0 0.0 0.0 0.0
2 0.0 0.0 0.0 0.0 0.0 0.0 0.0
3 0.0 0.0 0.0 0.0 0.0 0.0 0.0
4 0.0 0.0 0.0 0.0 0.0 0.0 0.0
5 0.0 0.0 0.0 0.0 0.0 0.0 0.0

```

[6 rows x 3338 columns]

Then I found similarity using Cosine similarity.

```

1 print(similarity)
[[0. 0. 0. ... 0. 0. 0. ]
 [0. 0. 0. ... 0. 0. 0. ]
 [0. 0. 0. ... 0. 0. 0. ]
 [0. 0. 0.36642144 ... 0. 0.16272871 0. ]
 [0. 0. 0.36642144 ... 0. 0.16272871 0. ]
 [0. 0. 0. 0. ... 0. 0. 0. ]]

```

Based on the similarity scores generated I found the first 10 Jobs most similar to the skill set and recommend them.

```

Job_Title Software Engineer
Company Confidential
Location None
Salary None
Summary We are based in San Francisco and would set yo...
Name: 239, dtype: object

Job_Title Entry Level Product Engineer
Company Locus Technologies
Location Mountain View, CA 94043
Salary None
Summary You will have the opportunity to utilize and h...
Name: 240, dtype: object

Job_Title Software Engineer - Back-end Development
Company Tynker
Location None
Salary None
Summary Our learning platform delivers courses that te...
Name: 1300, dtype: object

Job_Title Senior Software Engineer
Company Blue Sages
Location None
Salary None
Summary Software Engineer Experience and Qualification...
Name: 1299, dtype: object

Job_Title Software Quality Engineer
Company Apple
Location Austin, TX 78758 (North Austin area)
Salary None
Summary We are looking for a passionate and hardworkin...
Name: 1262, dtype: object

Job_Title Software Engineer
Company Capital Group
Location Irvine, CA 92618
Salary None
Summary The Solutions Engineer based in Irvine, CA is ...
Name: 1263, dtype: object

Job_Title Software Engineer Level 1
Company Honevwell

```

## Conclusion:

The recommended system for jobs is build by scrapping the indeed data and implemented it, by using content-based filtering, using TF-IDF and cosine similarity. The drawback of this system is that the dataset is too small. As the dataset is scrapped from the indeed.com. Future scope of this project is that this recommendation system can be developed by selecting multiple skillsets for recommending jobs with much better efficiency than the implemented recommender system.