NATIONAL UNIVERSITY OF COMPUTER AND EMERGING SCIENCES

Project Report

COURSE: RECOMMENDER SYSTEM (CS-4053)

PROJECT TITILE: JOB RECOMMENDER SYSTEM

GROUP MEMBERS:

- MUJTABA AHMED   20K-1801
- ALI IQBAL   20K-0141
- BILAL MAMJI   20k-1702
- AHMED AZAAN   20K-1099

Acknowledgement:

We came up with this idea with the help of our course teacher Sir Zain ul Hassan along with some web search. Stack overflowwebsite had many different ideas related to the project that we could work on and basically provided us with some core Recommender system ideas that we could modify and then implement so we as a group decided to go for it since it was quite comprehensive. Apart from efforts of all the team members, the section of this project report topic depends largely on the encouragement and guidance provided by our teachers. We take this opportunity to express our gratitude to the teachers who have been instrumental in the approval of this topic.

Introduction:

In today's rapidly evolving job market, finding the right employment opportunity that aligns with one's skills and interests can be a daunting task. With the ever-increasing volume of job listings, both job seekers and employers face the challenge of effectively matching talent with the right positions.

This project presents a comprehensive analysis and implementation of a job recommender system that aims to streamline the job search process by leveraging user-entered resumes and skills. By harnessing the power of advanced machine learning techniques and natural language processing, the system recommends jobs or similar positions based on the user's skillset and qualifications.

Features:

Content-Based Filtering with TF-IDF and Cosine Similarity:
In this project, a content-based filtering approach is employed to develop a job recommender system that utilizes the user's entered resume and skills to recommend relevant job opportunities. Unlike collaborative filtering methods that rely on user-item interactions, content-based filtering leverages the characteristics of the items themselves. In this case, the job postings are treated as items, and the user's resume acts as the basis for personalized recommendations.

To begin, the user's resume is pre-processed to extract relevant information such as skills, educational background, and work experience. One of the key techniques used in this project is the term frequency-inverse document frequency (TF-IDF) algorithm. TF-IDF assigns a

weight to each term in the resume, capturing its importance in the context of the entire document. This process helps in capturing the salient features of the resume and quantifying the significance of each skill mentioned.

After vectorizing the user's resume using TF-IDF, cosine similarity is applied to measure the distance between the user's resume and the job postings. Cosine similarity calculates the cosine of the angle between the two vectors, providing a measure of similarity between the user's skill profile and the job requirements. The closer the cosine similarity value is to 1, the more similar the job posting is to the user's skills, indicating a higher likelihood of a good job match.

By utilizing content-based filtering with TF-IDF and cosine similarity, this job recommender system offers personalized job recommendations based on the user's skill set. This approach enables the system to recommend jobs that closely align with the user's expertise, increasing the chances of finding a suitable employment opportunity. Furthermore, this approach mitigates the cold-start problem faced by collaborative filtering methods since it relies solely on the content of the user's resume rather than historical user-item interactions.

Dataset:

We have used dataset from Kaggle in our project. It contains 4 csv files which we have worked on:

➢ **Combined_Jobs_Final.csv**
➢ **Experience.csv**
➢ **Job_views.csv**
➢ **Position_of_interest.csv**

From the above 4 datasets we have worked on the following columns:

Combined_Jobs_Final:

Job.ID	Title	Position	Company	City	Industry

Experience:

Applicant.ID	Position.Name	Employer.Name	City	State.Name

Job_views:

Applicant.ID	Position	City	State.Name

Position_of_Interest:

Applicant.ID	Position.Of.Interest

Conclusion & Scope:

In conclusion, the job recommender system developed in this project showcases the effectiveness of content-based filtering using TF-IDF and cosine similarity. By considering the user's entered resume and skills, the system successfully generates personalized job recommendations that align with the user's expertise and requirements.

The scope of this project extends beyond the implementation of the job recommender system itself. The developed system can be further enhanced by incorporating additional features such as location preferences, industry preferences, and salary expectations, to provide more refined and accurate job recommendations. Moreover, integrating feedback mechanisms to gather user ratings and preferences can enhance the system's ability to learn and improve its recommendations over time. The system can also be extended to include features like resume parsing and real-time job data updates to ensure up-to-date recommendations for users.

Results:

Tags created for vectorization technique:

user

[559]:

	Applicant.ID	tags
0	2	volunteer writer for the uloop blog

Tags of Job:

new_data.head()

	Job.ID	tag
0	111	server tacolicious server tacolicious palo alto
1	113	kitchen staff chef claude lane kitchen staff c...
2	117	bartender machka restaurants corp. bartender m...
3	121	server teriyaki house server teriyaki house br...
4	127	kitchen staff chef rosa mexicano sunset kitche...

Tags of User's Item Features: (this tag is generated by combining last 3 csv files position's column value)

	Applicant.ID	tags
0	2	volunteer writer for the uloop blog
1	3	marketing intern server prep cook
2	6	project assistant
3	8	deli clerkserver cashier food prep order tak...
4	11	cashier

Vectorization Technique:

```python
from sklearn.feature_extraction.text import TfidfVectorizer
```

```python
cv = TfidfVectorizer()
```

```python
vectors = cv.fit_transform(new_data['tags'])
```

Cosine Similarity:

```python
from sklearn.metrics.pairwise import cosine_similarity
```

```python
user_vector = cv.transform(user['tags'])
```

```python
output = map(lambda x: cosine_similarity(user_vector, x),vectors)
```

```python
output = list(output)
```

Most Frequent Tags through cosine similarity:

	JobID	tag
13	243391	technical writer accountemps technical writer ...
16	255776	volunteer manager vitas healthcare volunteer m...
15	267063	volunteer coordinator the evangelical lutheran...
18	269751	specialized writer officeteam specialized writ...
19	271793	volunteer coordinator little sisters of the po...
11	271956	writer the boss group writer the boss group ad...
17	287632	volunteer hospice services bayada home health ...
12	293687	writer the boss group writer the boss group ad...
14	301358	technical writer technical writer saint cloud
10	315309	office volunteer coordinator office volunteer ...

ENTER APPLICANT ID (CV / Resume ID)

```python
yourinput = int(input("Enter Applicant ID: "))
while yourinput not in orig['Applicant.ID'].values:
    print("Desired Applicant ID Not Found")
    clear_output(wait=True)
    yourinput = int(input("Enter Applicant ID: "))

print("\n\nApplicant ID: ", yourinput)
```

Enter Applicant ID: `2`

Recommendations will be done on the basis of Position.Name

file1

	Applicant.ID	Position.Name	Employer.Name	City	State.Name
2762	2	Writer for the Uloop Blog	Cecilia Abate	San Francisco	California
2763	2	Volunteer	School for Self-Healing	San Francisco	California

Recommendations of the above query:

	Job.ID	Provider	Status	Slug	Title	Position	Company	City	State.Name	State.Code	...	Industry	Job.Description
13906	243391	2	open	phoenix-az-accountemps-technical-writer-243391	Technical Writer @ Accountemps	Technical Writer	Accountemps	Phoenix	Arizona	AZ	...	NaN	Ref ID:04730-006434Classification:Fin. Systems...
26245	255776	2	open	shrewsbury-nj-vitas-healthcare-volunteer-manager	Volunteer Manager @ VITAS Healthcare	Volunteer Manager	VITAS Healthcare	Shrewsbury	New Jersey	NJ	...	NaN	The Volunteer Services Manager is responsible ...
37525	267063	2	open	canton-sd-the-evangelical-lutheran-good-samari...	Volunteer Coordinator @ The Evangelical Luther...	Volunteer Coordinator	The Evangelical Lutheran Good Samaritan Society	Canton	South Dakota	SD	...	NaN	NaN
40214	269751	2	open	cedar-rapids-ia-officeteam-specialized-writer-...	Specialized Writer @ OfficeTeam	Specialized Writer	OfficeTeam	Cedar Rapids	Iowa	IA	...	NaN	Ref ID: 01520-9734786Classification: General O...
42255	271793	2	open	san-francisco-ca-little-sisters-of-the-poor-vo...	Volunteer Coordinator @ Little Sisters of the ...	Volunteer Coordinator	Little Sisters of the Poor	San Francisco	California	CA	...	NaN	\r\n\r\n\r\nVolunteer Coordinator\r\n\r\n\r\n\...
42418	271956	2	open	addison-tx-the-boss-group-writer	Writer @ The BOSS Group	Writer	The BOSS Group	Addison	Texas	TX	...	NaN	The BOSS Group is seeking a Creative Copywrite...