Job recommendor system - DA 331 - Project Proposal

Nabeel Khan Lakshya Borra

***Abstract*—In an era characterized by vast amounts of online data, job recommendation systems have emerged as a critical tool in reshaping the job recruitment landscape.Drawing inspiration from machine learning, we investigate how recommendation systems have evolved to address these challenges. This research paper delves into the intricacies of recommender systems, highlighting their critical role in both assisting recruiters in identifying the most suitable candidates and guiding job seekers towards opportunities aligned with their skills and interests.These systems pave the way for a more efficient, informed, and fulfilling job search experience.**

***Index Terms*—Job Recommendor systems, Recommendation systems, Candidate Hiring, HR Research**

I. Introduction

The landscape of job recruitment has undergone a profound transformation in recent years, driven by the exponential growth of data available online. This vast volume of information, predominantly found on internet-based platforms, has made the evaluation and extraction of relevant information a daunting challenge due to its sheer scale. Whether it's an individual job seeker or a hiring organization, the quest for timely and accurate results can be overwhelming.

The recruitment process itself is intricate, involving the creation of job profiles, screening applicants, and selecting the best-suited candidates. It begins with defining the job's requirements, including mandatory and preferred skills, experience criteria, and location preferences. Applicants submit their profiles through various channels, often with incomplete data. Screening is a critical step where only candidates meeting minimum criteria progress, but the final selection is subjective and reliant on a recruiter's judgment.

This is where Recommender Systems (RS) emerge as a game-changer. RS, a subset of machine learning, harnesses the power of algorithms to analyze user profiles and provide personalized recommendations. While RS is traditionally associated with suggesting products or services, in the context of job recruitment, it takes on a unique role. It not only assists recruiters in identifying the most eligible candidates but also aids job seekers by recommending positions that align with their skills and interests.

The COVID-19 pandemic further underscored the importance of job recommendation systems, especially for unskilled workers and job seekers looking to transition into the IT field. In this new normal, where many individuals are seeking employment realignment, such systems can bridge the gap by analyzing skills and interests to recommend suitable job opportunities. Moreover, these systems have the potential to enhance the overall job search experience, foster trust, and foster customer loyalty. By suggesting the right job matches based on preferences, ratings, experience, and location, they empower individuals to make informed career choices and contribute to a more efficient and fulfilling job market.

In a world where college graduates often struggle to choose their career paths, job recommendation systems offer a lifeline. Many engineering graduates, for example, aspire to enter the IT domain but are unsure of their fit. These systems, by analyzing resumes and skills, provide a data-driven solution. Using preprocessing techniques we suggest jobs that meet the candidate's requirements like preferred location,salary and further selection will be based on his technical and soft skills.We use TF-IDF vectorization to match resumes with job descriptions and employ cosine similarity to score the matches,presenting job options hierarchically. Notably, it not only recommends jobs but also suggests areas for skill improvement, enabling candidates to enhance their qualifications proactively.

In conclusion, job recommendation systems have become indispensable in navigating the ever-expanding landscape of employment opportunities, especially in industries like IT services experiencing explosive growth. They provide a personalized and data-driven approach to both recruiters and job seekers, ensuring better job matches and streamlined career transitions. As the job market continues to evolve, these systems play a pivotal role in shaping the future of recruitment, offering efficiency, personalization, and growth opportunities for all stakeholders.

II. DATA SOURCES AND SITES

In the pursuit of constructing a comprehensive Job Dataset for research purposes, we combined multiple resources from Kaggle and conducted supplementary Google searches. Additionally, we collected an array of resume data from our network of acquaintances. It is noteworthy that certain datasets encompass supplementary information, notably location and salary columns, which serve the purpose of facilitating tailored job recommendations aligned with applicant preferences. In cases where these particular columns remain unpopulated, our model is designed to employ a hierarchical approach, thus prioritizing job recommendations based on inherent job attributes rather than specific location and salary requirements.

III. PREVIOUS INVESTIGATIONS

In the realm of online job recommendations, the system known as iHR, developed by Hong et al. in 2013 [1], stands as a noteworthy model. iHR categorizes users into distinct groups based on personal information and historical user behavior. While this approach exhibits promise, it presents challenges in scenarios where a one-size-fits-all recommendation approach may not be suitable due to variations in user attributes.

Although straightforward matching approaches exist, their limitations become apparent when considering their lack of individualization, which may not offer substantial assistance to job seekers. Personalized recommendations, on the other hand, derived from collaborative filtering and content-based methods, prove to be more beneficial to job seekers [2]. It's worth noting, however, that collaborative filtering encounters difficulties in addressing cold start issues, and content-based recommendations can sometimes yield overly specific outcomes.

Collaborative filtering, a method reliant on ample data for effective operation, has been explored by other researchers as a distinct recommender system. However, it may not adequately serve individuals seeking the ideal job fit. In some academic circles, content-based recommenders have been championed as the superior choice in the context of job recommendations [3].

A recommender system extends beyond mere prediction algorithms; it encompasses various components, including word vectorization and similarity functions. For instance, researchers like Li-Ping Jing [4] have advocated for the adoption of tf-idf (Term Frequency-Inverse Document Frequency) as an optimal strategy for text feature selection in the realm of text mining.

Furthermore, research by Mohammad Alobed, as demonstrated in ”A Comparative Analysis of Euclidean, Jaccard, and Cosine Similarity Measure and Arabic Wordnet for Automated Arabic Essay Scoring” [5], has indicated that Cosine similarity, in conjunction with various stemming techniques, exhibits the lowest error rates when compared to Jaccard and Euclidean similarity measures.

After a comprehensive review and thoughtful consideration of the intricacies and challenges associated with recommender systems, it has been determined that the most suitable approach for our research involves the utilization of cosine similarity and TF-IDF (Term Frequency-Inverse Document Frequency) methodologies.

IV. PROPOSED RESEARCH

The realm of research in job recommendation systems has traditionally been centered on either providing job recommendations to applicants or suggesting suitable candidates to hiring teams. These approaches have undoubtedly brought significant value to the recruitment process. However, our recommendation system takes a distinctive and holistic approach that transcends the conventional boundaries of this domain.

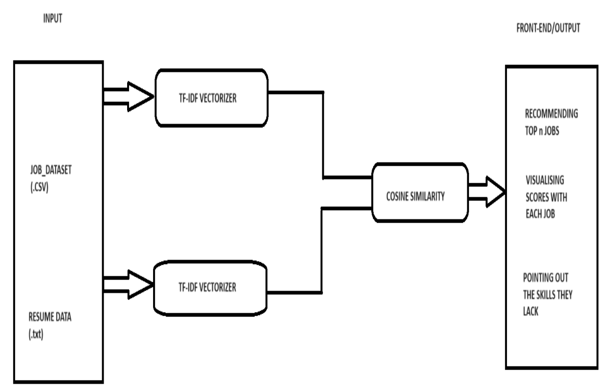

Fig. 1. Job recommendation architecture

Rather than exclusively catering to one side of the equation, our system is designed to address both the needs of job seekers and hiring teams in a synergistic manner. It acts as a bridge, connecting job seekers with opportunities and assisting recruiters in identifying the most promising candidates. This dual-pronged approach ensures that the recruitment process is not just efficient but also mutually beneficial.

One of the standout features of our recommendation system is its commitment to providing constructive feedback to applicants when their applications are rejected. This feedback mechanism represents a paradigm shift in the way applicants interact with the recruitment process. Instead of facing rejection without insights or guidance, applicants are empowered with valuable feedback that can be instrumental in their personal and professional growth.

Pre-processing is being applied to the resumes to improve the design.Porter Stemmer will make each word into its root word for pre-processing top words and stop words,while stop words will eliminate any unnecessary terms. The system is improved as a result. Methods of which are

Stop Word : The algorithm just reads or selects those features which are relevant, because it doesn't need words like ”is” or ”are” that don't award a score or points for comparison. Stop words are therefore used to get rid of all the unnecessary terms. The word frequency was then reduced by Porter Stemmer.

Porter Stemmer : Porter Stemmer is a stemming algorithm. Stemming is a method of reducing a word to its base (root word), which joins with other roots, suffixes, and prefixes to make a base word.

A. Recommendation System Model

The recommendation system model is meticulously structured into three distinct layers, each fulfilling a pivotal role in the recruitment process:

1) **Applicant-Centric Layer:**
Within this foundational layer, our recommendation system undertakes a crucial responsibility with precision and thoroughness. Its primary aim is to align job listings precisely with the specific requirements articulated by individual applicants. Within this layer, applicants are granted the privilege to seamlessly interact with a curated selection of job listings. This interaction transpires effortlessly through a streamlined application process or via search functionalities meticulously tailored to match their unique interests and aspirations. This is the very first segregational layer

2) **Non-Technical Skill Evaluation:**
Transitioning to the employer's vantage point, our system embarks on a comprehensive and intricate process to sift through the diverse pool of applicants. This discerning process identifies and distinguishes the upper echelon—the top 50 percent of candidates who exhibit an exceptional resonance with the roles in question. The criteria for selection within this distinguished group transcend mere technical competence. Instead, they encompass a comprehensive evaluation of non-technical skills, which are integral and directly pertinent to the intricate fabric of the job roles. The result is a meticulous filtering mechanism, ensuring that candidates possess the finesse, interpersonal communication skills, and soft skills that are indispensable for success in their respective positions. This effectively aims to provide a soft skills barrier

3) **Technical Skill Assessment:**
In the culminating phase of our model, the focus sharpens towards further refinement. The initial pool, comprising the top 50 percent of applicants, undergoes a judicious selection process, culminating in the ascension of the uppermost 25 percent. This elite group is meticulously chosen through a rigorous evaluation of their technical proficiencies, meticulously documented within their resumes. These candidates are extended a special invitation to participate in meticulously scheduled interviews. During these pivotal conversations, a deeper and more nuanced analysis unfolds—a comprehensive exploration of their aptitude, potential, and overall suitability for the esteemed positions that await them.
This structured model serves as the bedrock of our research, offering a comprehensive approach to enhance the efficiency and effectiveness of the recruitment process. It reflects our commitment to precision, fairness, and a holistic assessment of candidates, contributing to a more robust and refined recruitment ecosystem.

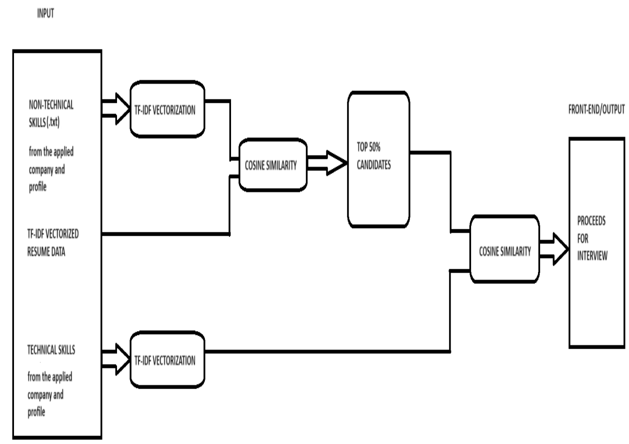

Fig. 2. Candidate recommendation architecture

B. Additional Details

To provide further context, it's important to highlight the significance of our approach. By addressing both the applicant-centric and employer-centric aspects of the recruitment process, we aim to enhance the overall efficiency and effectiveness of job placement. This holistic approach not only optimizes the matchmaking between job seekers and job opportunities but also streamlines the hiring process for organizations.

In addition to the three primary layers of our recommendation system, we also emphasize the importance of continuous data enrichment. To ensure the optimal performance of our model, it is imperative that the dataset remains up-to-date and continuously enriched with job offerings contributed by the company. This iterative process is essential for maintaining the model's accuracy and relevance, thereby enhancing its utility in the realm of job recommendation research.

This multifaceted approach not only optimizes the matchmaking between job seekers and job opportunities but also streamlines the hiring process by identifying the most promising candidates at each stage, ultimately enhancing the overall efficiency and effectiveness of the recruitment process.

In order to ensure the optimal performance of our model, it is imperative that the dataset remains up-to-date and continuously enriched with job offerings contributed by the company. This iterative process is essential for maintaining the model's accuracy and relevance.

*1) **Cosine Similarity**:* Moving on to this similarity metric, any two things from the input are taken as two items. Here, the angle idea is utilized to determine how comparable the items are.

$$\text{similarity} = \cos(\theta) = \frac{\mathbf{A}\cdot\mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}},$$

$$\text{TF(word, text)} = \frac{\text{number of times the word occurs in the text}}{\text{number of words in the text}}$$

$$\text{IDF(word)} = \mathbf{log}\left[\frac{\text{number of texts}}{\text{number of texts where the word occurs}}\right]$$

$$\text{TF-IDF(word, text)} = \text{TF(word, text)} \times \text{IDF(word)}$$

*2) **TF-IDF (Term Frequency-Inverse Document Frequency):**:* *Term Frequency:* TF of a term or word of a term or word is the ratio of the number of times the term appears to the total number of words in the document.

Inverse Document Frequency: A term's IDF reflects the percentage of corpus documents that use that term. More weight is given to terms that are specific to a small number of texts than to words that appear in all papers. TF-IDF scores a word by multiplying the word's Term Frequency (TF) with the Inverse Document Frequency (IDF)

V. Conclusion and Future Scope

Two pre-processing techniques, text mining technique, and similarity function are used here. Porter stemmer and stop words are the pre-processing techniques. Tf-idf is the text mining technique. A cosine similarity function is used to measure similarity. Pre-processing techniques are used with job descriptions and resumes to improve system efficiency by removing some unnecessary words. TF-IDF is used to transform text-based job descriptions and resumes into a matrix for comparison. The similarity between each job description and resume will be calculated using the cosine similarity method.

Finally, it will sort and display the scores for each work. Utilizing a list comparison method to evaluate the job abilities and resume, and then suggesting skills that should be improved.

Only career recommendations and skill suggestions for IT occupations are included in the planned effort. It can be enhanced by offering positions and qualifications for non-IT jobs.

Some people may discover a superior method of similarity detection in the future than cosine similarity. The recommendation becomes more precise as a result.

References

[1] Hong. W,Zheng. S,Wang. H, and Shi. J (2013) 'A job recommender system based on user clustering', Journal of Computers, Vol. 8, No. 8, pp.1960–1967.

[2] Roshan G. Belsare, Dr. V. M. Deshmukh ' Employment Recommendation System using Matching, Collaborative Filtering and Content Based Recommendation'

[3] Minh-Luan Tran, Anh-Tuyen Nguyen, Quoc-Dung Nguyen, Tin Huynh, "A comparison study for job recommendation", International Conference on Information and Communications (ICIC), pp. 199-204, 2017, doi:10.1109/infoc.2017.8001667

[4] Li-Ping Jing, Hou-Kuan Huang, Hong-Bo Shi, "Improved feature selection approach TFIDF in text mining", International Conference on Machine Learning and Cybernetics, pp. 944-946, 2002, doi:10.1109/icmlc.2002.1174522.

[5] Mohammad Alobed;Abdallah M M Altrad;Zainab Binti Abu Bakar, " A Comparative Analysis of Euclidean, Jaccard and Cosine Similarity Measure and Arabic Wordnet for Automated Arabic Essay Scoring", Fifth International Conference on Information Retrieval and Knowledge Management (CAMP), pp.70-74, 2021, doi:10.1109/camp51653.2021.9498119.