Survey on Resume and Job Profile Matching System

Dipti Pawade
Information Technology,
KJ Somaiya College of Engineering
Mumbai, India
diptipawade@somaiya.edu

Tanuj Joshi
Information Technology,
KJ Somaiya College of Engineering
Mumbai, India
tanuj.joshi@somaiya.edu

Siddhesh Parkhe
Information Technology,
KJ Somaiya College of Engineering
Mumbai, India
siddhesh.parkhe@somaiya.edu

***Abstract*—Finding the correct job profile that matches the candidate's resume and eventually their skillset is a very tedious and time task. The employee as well as the aspirant candidates needs to continuously dig into the information sources to check whether a particular resume is apt for the given job profile and vice versa. Thus, many researchers have contributed to this area. This paper discussed various studies done to give the most up-to-date design methodology for the system that matches job profiles with resumes. For the study, we have considered research work published from year 2018 to 2023. The study shows that recently many researchers have considered such a system as the natural language-based classification problem. Still, the use of similarity measure is the most dominant technique used in this case.**

Keywords—Resume ranking, job profile matching, Job recommendation, candidate selection

I. Introduction

Employers are continuously searching for efficient approaches to hook up with proper opportunities and candidates. This has caused the development and implementation of superior technology, which includes Resume and Job Description Matching Systems. These structures function as a critical bridge between process seekers and employers by streamlining the recruitment method, saving time, and ensuring higher health between candidates and task openings.

A Job Profile Matching System is an advanced software utility designed to automate and optimize the initial levels of the hiring method. It accomplishes this by analyzing resumes submitted by using job seekers and job descriptions published by employers to perceive maximum appropriate fits. The goal of these systems is to shorten the time and effort needed for job seekers to find relevant process openings and for recruiters and hiring managers to learn about certified candidates. Such systems have numerous treasured blessings from the candidate's factor of view, enhancing the task search revel in and growing the likelihood of finding an appropriate position. The advantages of Job Profile Matching Systems are :

1. Time Efficiency: These systems extensively lessen the time required to sift through numerous resumes and activity postings, streamlining the recruitment technique.

2. Improved Quality of Matches: By using advanced algorithms and system-gaining knowledge of, those systems decorate the likelihood of an excellent fit between candidates and job openings.

3. Enhanced Candidate Experience: Job seekers acquire greater focused task recommendations, increasing their probability of finding a position that suits their qualifications and interests.

4. Data-Driven Decision-Making: Employers could make extra informed hiring selections based totally on facts and analytics furnished by way of the device.

In conclusion, Resume and Job Description Matching Systems have revolutionized the recruitment process by leveraging technology to connect job seekers with employers more effectively. These systems enhance efficiency, reduce human bias, and ultimately lead to better outcomes for both job seekers and employers in today's competitive job market.

The general architecture used by the majority of the resume and job profile matching system is shown in Figure 1. The ranking algorithm is used to determine each applicant's ranking after comparing their resume to the job description. The most relevant resume will have a higher ranking than others. The job description can be in any format such as a doc, pdf file or even a csv file. The JD file be parsed so that the required skill set is extracted. Consequently, the CV uploaded by candidates will also be parsed for extracting their skill set. Both the skill sets will be stored in a database.

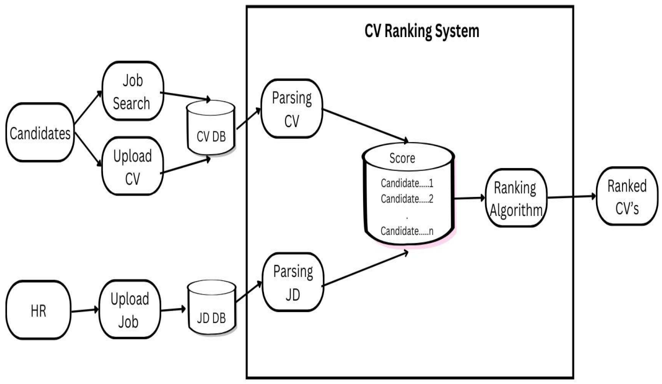

Fig.1. General Resume (CV) and Job Profile Matching System

This paper discusses about the study done by various researchers to design a system to match the resume with the appropriate job profile through various available algorithms.

II. Literature Survey

There are insufficient resources available in the traditional hiring process to adequately screen each candidate. To address this issue, Ahmad Cucus et. al. [1] developed a profile matching method using specific criteria like skills, job titles, location, and years of experience. The proposed algorithm employed the following steps: it began by conducting gap mapping to measure the variations in the profiles according to attribute values, then proceeded to calculate the weights of the primary and secondary factors, and finally computed the percentage of the match. This algorithm efficiently selected the most suitable workers from a large dataset, significantly improving the match between workers and the tasks requested by the employer.

Irawan et al. [2] used the profile matching method to choose qualified applicants for the position of laboratory electrical machinery operators. The method uses criteria data to assign individual profile weights, resulting in a final score for each candidate. This profile matching algorithm is similar to the algorithm stated in [1]. The system performs a gap mapping procedure to gauge disparities in the profiles. Subsequently, it engaged in the computation of both primary and secondary factor weights before ultimately deriving the match percentage.

Rasika Ransing et. al. [3] developed an effective application for candidate screening and ranking according to specific criteria. This application enables recruiters to sort resumes by relevance, while individuals can evaluate their resume's suitability for particular job titles. The approach involved a stacked classifier that combined Linear SVC, XGBoost, and KNN algorithms. Resumes underwent preprocessing, followed by conversion into vector form using the TF-IDF model. Utilizing machine learning methods to ascertain the expected class probabilities, the resume filtering and ranking procedure was finally finished.

M.Alamelu et. al. [4] proposed a system to streamline the resume screening process by extracting essential data from uploaded resumes, utilizing Natural Language Processing. It employs data extraction, preprocessing, and segmentation, allowing candidates to review and adjust the extracted information. The system then trains datasets for specific job postings and calculates scores based on the match between resumes and job descriptions, generating a ranked list for recruiters' consideration using Cosine Similarity.

Narendra et. al. [5] provided a solution to screen a large number of resumes. This method focuses on extracting specific entities from resumes to comprehensively assess candidates using name entity recognition. It maintains a record of relevant terms during parsing and stores them for later classification. The study also involves document ranking, presenting the resumes in descending order of their scores, and simplifying the shortlisting process for recruiters. The system used a dataset of 200 resumes available online.

Thapanee et al. [6] selected the resumes of the best candidates using a machine learning-based method.. The methodology they employed for Natural Language Processing comprised multiple preprocessing stages, including lemmatization, stop word removal, and summarization. Then the performance of various machine learning classifiers, including SVM, CatBoost, Radial Basis Function (RBF) Decision Tree (DT) and KNN are evaluated. The study indicated that the RBF SVM classifier outperformed over all the methods.

Tallapragada et al. [7] proposed a resume ranking method that involves pre-processing resume data, including tokenization, stop word removal, and TF-IDF analysis to determine word significance. The text is then contextually identified and the CV is given a rank after using Bidirectional Encoder Representation Transformer (BERT) vectorization..

The study done by Niti Khamker et. al. [8] presented a system that shows the candidates' closeness results to the JD, assisting HR in identifying the most qualified applicants. with the best scores. Initially, the uploaded resume is tokenized and normalized with the help of a countvectorizer. The percentage that two documents match is then displayed using the cosine similarity method.

Leah G. Rodriguez et. al. [9] discussed an approach to identify key attributes for a profile matching model in a job matching system. These attributes are derived from more than two thousand resumes and job requirements available in the Public Employment Service Office (PESO) in Pangasinan. Data mining in WEKA software is used for analysis, and rankings are presented statistically. Once attributes are identified, the system used a clustering algorithm to match job seekers' profiles with job requirements.

Ahmad Alsharef et. al. [10] in his study explored the use of the system based on text similarity measures as an alternative to hiring managers proficient in the recruitment process. Three measures of texture similarity were considered: Cosine, Sqrt-Cosine, and Improved Sqrt-Cosine (ISC). The findings show that ISC and Sqrt-Cosine were closer to experts and human judgment than Cosine similarity.

In their study, Suleiman Ali Alsaif et al. [11] offered a system to help job searchers locate recommends the pinnacle jobs based on their resumes. By utilizing content-based filtering to analyze and quantify the percentage of similarity between a candidate's talents and specific job ad elements, the suggested approach helps job searchers find the best positions. To finally match job recommendations, the proposed process included steps like site scraping, tokenization, Jaccard Coefficient, and cosine similarity.

Asrar et al. [12] presented a three-step method for candidate ranking and selection using NLP tools to extract semantic information for 13 qualities, creating comparison descriptions by comparing pairs of resumes for each attribute, and using the SVM algorithm to score resumes.

Moving forward B. Lalitha et.al.[13] research focuses on making job application process efficient and very simple by developing a web application. In this instance, the degree of correspondence between the JD and the CV is assessed, and the results are displayed as a percentage. Firstly, Tokenization is used for separating individual tokens, or words, then measure the relevance using cosine similarity.

Painem et. al. [14] built a profile matching system using PHP and used MySQL for backend. Here the core factor and secondary factor are calculated, gap mapping is done and then the ranking is performed. The dataset is created using the information from the Budi Luhur University HR.

Then shifting to the advanced technologies, Sujatha Krishnamoorthy et.al.[15] focused on active resume scanning and resume classification. The dataset is collected from various colleges and there are one lac resumes from more than 100 colleges and universities. The dataset is divided into 25:75 percent as the testing and training dataset. For analysis purpose, three steps are carried out which were segmentation of words, deactivation, and the substitution of unique words. The weights produced by the analysis are then applied to different models, and the resumes with the highest score are then matched to appropriate jobs for each applicant. The authors used various methods such as CNN, BiLSTM, CRNN, and BPNN but CRNN has better accuracy than others which is 96% respectively.

Further Belal Amro et.al.[16] also discussed the system to rank the candidate based on their expertise score. For their study, they have used the dataset comprised of 101 unstructured resumes from the IT domain which are being gathered from various resources. The first training block, trains the domain-specific word embedding from resumes using the efficient Skip-Gram model. The second matching block uses Cosine similarity matching to match the text of the JD with the text of each candidate CV. Lastly, candidate details are extracted using pattern matching and NLP gazetteer approaches.

Dor et. al. [17] used fine-tuning a Siamese SentenceBERT (SBERT) a model to find the match between job description and candidate The dataset used for the study comprises of 274,407 resume-vacancy pairs (negative and positive) gathered from various sources. This study compares different methods for creating embeddings from job vacancies and resumes to construct useful textual feature representations. The proposed approach outperformed the TF-IDF features and a pre-trained BERT embedding approach.

Chirag et. al.[18] describe a solution for shortlisting the resume according to the job specification. The system uses NER to extract the named entity which specifies the relevant information. After that, TF-IDF, cosine similarity, and the vectorization model are used to compare each resume to the job description.

Sridevi G.M. et al. [19] created an AI system to make the process of locating a suitable candidate easier. Adjectives, adverbs, and primary and secondary skills are the four fields that are formed for each candidate CV and JD. Next, the Jaccard similarity between the relevant clusters for the resume of the candidate and the job description is computed. Candidate suitability prediction is then carried out using classifiers like Adaboost, XGBoost linear regression and decision tree. The XGBoost classifier attains a top average accuracy of 95.14%.

Pradeep et. al. [20] suggested the creation of an automated machine learning model that uses job descriptions to suggest resumes from qualified candidates to HR. The program purges single-letter words, digits, and special characters from resumes. NLTK is used to tokenize the dataset, and additional preprocessing steps include lemmatization, stemming, and stop word removal. A wide range of methods have been studied, such as Random Forest, Logistic Regression, k-Nearest Neighbors, Linear Support Vector Machine Classifier, Multinomial Naive Bayes, and Content Based Recommendation using Cosine Similarity. The best resume is suggested based on the job description by using methods like content-based recommendation via cosine similarity and k-nearest neighbors.

Tejaswini et. al. [21] uses a content-based recommendation method to find the CV that, using cosine similarity, most closely resembles the job description. After text pre-processing, the TF-IDF value is calculated. Then Cosine similarity and KNN methods are used for recommendation.

Peng Xu and Denilson Barbosa [22] explored various models which included random forest, densely connected neural networks and gradient boosting tree, for comparing resumes to job specifications. They found that stacked models consistently outperformed standard methods, although they didn't address complex cases like candidates with experience in different technologies than those mentioned in their resumes.

Sushruta Mishra et al. [23] proposed a two-component approach. The Stacked KNN Learner Module employed different KNN variants to predict candidates' suitability for IT-based jobs. The Hard Voting Ensemble Predictive units, aggregated class votes to make final decisions.

Riza et al. [24] categorized resumes using KNN and assessed the degree of similarity between the CV and JD using the Cosine Similarity method. However, their model only accepts resumes as Common Separated Value format, potentially losing valuable information from other formats during text summarization.

TABLE I presents the summary of all the approaches discussed in this paper.

TABLE I. SUMMARY OF THE LITERATURE STUDIED

Ref No	Year	Dataset	Method
1	2022	Kaggle	Core Factor And Secondary Factor Calculation, Profile matching algorithm
2	2022	Data collected from student	Core Factor And Secondary Factor Calculation, Profile matching algorithm
3	2021	Kaggle	Tokenization, TF-IDF, Stacked model (KNN, Linear SVC, and XGBoost)
4	2021	Resume collected from the user	NER, Cosine Similarity Measure
5	2022	200 online resumes	Tokenization and Token normalization, NER
6	2022	2027 resumes	NLP pipeline for data preprocessing, DT, KKN, RBF SVM, CatBoost
7	2023	100 resumes	TF-IDF, BERT vectorization
8	2021	Resume collected from user	Cosine similarity
9	2019	Data from PESO	Data mining, similarity calculation, Clustering
10	2023	HR data collected from companies located in Europe and the US	TF-IDF Vectorization, Cosine, Sqrt-Cosine, and Improved Sqrt-Cosine (ISC).
11	2022	The dataset is created through web scraping	Tokenization, Lemmatization, Stop words removal, Cosine similarity, Jaccard Coefficient
12	2023	228 resumes from LinkedIn, Kaggle, Github	NER, SVM
13	2023	Resume collected from user	Tokenization, Stemming, Lemmatization, Cosine Similarity
14	2022	The Budi Luhur University HR	Profile matching using the calculation of the core and

Ref No	Year	Dataset	Method
		dataset	secondary factors
15	2021	1 lakh CV's from one hundred and twenty-three colleges and universities.	Word segmentation, deactivation, unique word substitution, CNN, BiLSTM, CRNN, BPNN
16	2022	101 resumes from different available resources.	Skip gram, cosine similarity, NLP gazetteer, and a pattern matching
17	2021	274,407 resume-vacancy pairs(negative and positive)	TF-IDF, BERT, SBERT
18	2020	Resume collected from user	Tokenization, Lemmatization, POS Tagging, NER, Vectorization, TF-IDF, Cosine similarity
19	2022	The dataset containing 14,906 job seeker resumes and 8 JDs collected from various sources	Clustering, Jaccard similarity, linear regression, decision tree, Adaboost, and XGBoost
20	2020	Kaggle	Logistic Regression, Linear Support Vector Machine Classifier, Random Forest, Multinomial Naive Bayes, TF-IDF, Cosine Similarity, and k-Nearest Neighbors
21	2022	Information from Kaggle	TF-IDF, Cosine Similarity, k-Nearest Neighbours
22	2022	Set of resumes	Stacked models
23	2021	Resume repository	Stacked KNN Learner Module
24	2021	Resume from user	KNN

III. Discussion

Researchers are continuously exploring various methodologies to screen out the most suitable resume for a given job description and vice versa. Through the literature survey, it is observed that there are three major tasks involved in this process.

1. Screening of resumes and extracting the key features (attributes). These features are normally the skill set depicted through fields like technologies known, areas of expertise, certification, project details, etc.
2. Screen the job description and extract the key skill set requirement for the given job profile.
3. Determining the resume score by contrasting the applicant's information with the requirements listed in the job description. Calculating the resume score by comparing the candidate's details against the job description requirement.

As the complexity of job descriptions increases, it is important to consider the context rather than just relying on the keywords. Hence, CNN, RNN, and LSTM based approaches may give better results than the conventional similarity-based approaches [7][17]. There is a need for an intelligent system that can understand complex information relying on quality and not on quantity. For example, a single certification from a renowned institute/organization like AWS, Oracle, Microsoft, etc. is much better than two certifications from a non-renowned institute. The same is applicable for evaluating the experience. Weightage should be given to the experience in the said domain and the kind of projects done rather than the total number of working years

The job profile matching system generally ends up providing the suitability of the resume for a given job. A recommendation system that can give certain inputs on the improvisation of the candidate's skill set and his resume can be a good feature to include. To give a fair recommendation, a system should consider the local space as well as the global space. In the local space, resume improvisation recommendation is based on the given job description only while in the global space, the recommendation is given based on the analysis of all the similar job profiles available in the system (may be past or current). This will lead to a comprehensive analysis.

Our proposed methodology is to build a resume matching system which matches candidates resume with appropriate job descriptions and show the percentage of suitability between them to the HR. HR can also see the ranking of resumes in the descending order. There would be two ends one for the HR and other for the candidates.

IV. Conclusion

From this study, it is observed that majorly researchers have categorized the resume into primary and secondary attributes, and based on the final score the resumes are ranked. The higher the rank, the more is the resume suitable for the given job profile. This approach is dominantly dependent on the attributes and if they are not selected properly then ranking may not be reliable. Several academics have shown that the primary cosine similarity is a very effective method for resolving the problem of how well a CV matches a job profile. Even though the resume and job profile corpus is big this methodology may take more time to analyze the data and provide the results. There are a few shreds of evidence where a machine learning classification algorithm is used effectively to categorize the resume and job profile.

Even with the abundance of accessible techniques, it might be difficult to address the ambiguities in natural language included in both the JD and the CV. Researchers may also encounter difficulties when managing resumes with various layouts or unstructured content.

REFERENCES

[1] Cucus, Ahmad, Luhur Bayu Aji, Al-Fahim Bin Mubarak Ali, Afrig Aminuddin, and Lilis Dwi Farida. "Selection of Prospective Workers Using Profile Matching Algorithm on Crowdsourcing Platform." In 2022 5th International Conference on Information and Communications Technology (ICOIACT), pp. 122-126. IEEE, 2022.

[2] Wahyono, Irawan Dwi, Khoirudin Asfani, Hari Putranto, Djoko Saryono, Herri Akhmad Bukhori, Tiksno Widyatmoko, Mohd Shafie Rosli, Nurbiha A. Shukor, and Noor Dayana Abdul Halim. "Selection Modification of Model in Profile Matching Algorithm for Admission and Placement of Assistant Members." In 2022 IEEE 8th Information Technology International Seminar (ITIS), pp. 1-6. IEEE, 2022.

[3] Ransing, Rasika, Akshaya Mohan, Nikita Bhrugumaharshi Emberi, and Kailas Mahavarkar. "Screening and Ranking Resumes using Stacked Model." In 2021 5th International Conference on Electrical, Electronics, Communication, Computer Technologies and Optimization Techniques (ICEECCOT), pp. 643-648. IEEE, 2021.

[4] M. Alamelu, D. S. Kumar, R. Sanjana, J. S. Sree, A. S. Devi and D. Kavitha, "Resume Validation and Filtration using Natural Language Processing," 2021 10th International Conference on Internet of Everything, Microwave Engineering, Communication and Networks

(IEMECON), Jaipur, India, 2021, pp. 1-5, doi: 10.1109/IEMECON53809.2021.9689075.

[5] Narendra, G. O., and S. Hashwanth. "Named Entity Recognition based Resume Parser and Summarizer."

[6] Boonchob, Thapanee, Nuengwong Tuaycharoen, Santisook Limpeeticharoenchot, and Narongthat Thanyawet. "Job-Candidate Classifying and Ranking System-Based Machine Learning Method." In 2022 26th International Computer Science and Engineering Conference (ICSEC), pp. 94-99. IEEE, 2022.

[7] Tallapragada, VV Satyanarayana, V. Sushma Raj, U. Deepak, P. Divya Sai, and T. Mallikarjuna. "Improved Resume Parsing based on Contextual Meaning Extraction using BERT." In 2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS), pp. 1702-1708. IEEE, 2023.

[8] Khamker, Niti, Y. Khamker, and M. Butwall. "Resume match system." International Journal Of Innovative Science And Research Technology (2021).

[9] Rodriguez, Leah G., and Enrico P. Chavez. "Feature selection for job matching application using profile matching model." In 2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS), pp. 263-266. IEEE, 2019.

[10] A. Alsharef, Sonia, H. Nassour and J. Sharma, "Exploring the Efficiency of Text-Similarity Measures in Automated Resume Screening for Recruitment," 2023 10th International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 2023, pp. 36-42.

[11] Alsaif, Suleiman Ali, Minyar Sassi Hidri, Hassan Ahmed Eleraky, Imen Ferjani, and Rimah Amami. "Learning-based matched representation system for job recommendation." Computers 11, no. 11 (2022): 161.

[12] Alderham, Asrar Hussain, and Emad Sami Jaha. "Comparative Semantic Resume Analysis for Improving Candidate-Career Matching." In 2022 14th International Conference on Computational Intelligence and Communication Networks (CICN), pp. 313-321. IEEE, 2022.

[13] B. Lalitha, S. Kadiyam, R. V. Kalidindi, S. M. Vemparala, K. Yarlagadda and S. V. Chekuri, "Applicant Screening System Using NLP," 2023 International Conference on Innovative Data Communication Technologies and Application (ICIDCA), Uttarakhand, India, 2023, pp. 379-383, doi: 10.1109/ICIDCA56705.2023.10099953.

[14] Painem, H. Soetanto and U. Budiyanto, "Analysis of Job Placement Based on Employee Competency Using Profile Matching," *2022 9th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*, Jakarta, Indonesia, 2022, pp. 394-398, doi: 10.23919/EECSI56542.2022.9946448

[15] Liu, Jiahao, Yifan Shen, Yijie Zhang, and Sujatha Krishnamoorthy. "Resume parsing based on multi-label classification using neural network models." In Proceedings of the 6th International Conference on Big Data and Computing, pp. 177-185. 2021.

[16] Amro, Belal, Arwa Najjar, and Mario Macido. "An Intelligent Decision Support System For Recruitment: Resumes Screening and Applicants Ranking." (2022).

[17] Lavi, Dor, Volodymyr Medentsiy, and David Graus. "consultantbert: Fine-tuned siamese sentence-bert for matching jobs and job seekers." arXiv preprint arXiv:2109.06501 (2021).

[18] Daryani, Chirag, Gurneet Singh Chhabra, Harsh Patel, Indrajeet Kaur Chhabra, and Ruchi Patel. "An automated resume screening system using natural language processing and similarity." ETHICS AND INFORMATION TECHNOLOGY [Internet]. VOLKSON PRESS (2020): 99-103.

[19] Sridevi, G. M., and S. Kamala Suganthi. "AI based suitability measurement and prediction between job description and job seeker profiles." International Journal of Information Management Data Insights 2, no. 2 (2022): 100109.

[20] Roy, Pradeep Kumar, Sarabjeet Singh Chowdhary, and Rocky Bhatia. "A Machine Learning approach for automation of Resume Recommendation system." Procedia Computer Science 167 (2020): 2318-2327.

[21] Tejaswini, K., Umadevi, V., Kadiwal, S.M. and Revanna, S., 2022. Design and development of machine learning based resume ranking system. Global Transitions Proceedings, 3(2), pp.371-375.

[22] Xu P., Barbosa D. (2018) Matching Résumés to Job Descriptions with Stacked Models. In: Bagheri E., Cheung J. (eds) Advances in Artificial Intelligence. Canadian AI 2018. Lecture Notes in Computer Science, vol 10832. Springer, Cham. https://doi.org/10.1007/978-3-319-89656-4_31

[23] Mishra, Sushruta & Mallick, Pradeep Kumar & Tripathy, Hrudaya & Jena, Lambodar & Chae, Gyoo-Soo. (2021). Stacked KNN with hard voting predictive approach to assist hiring process in IT organizations. The International Journal of Electrical Engineering & Education. 002072092198901. 10.1177/0020720921989015.

[24] Rajath V , Riza Tanaz Fareed , Sharadadevi Kaganurmath, 2021, Resume Classification and Ranking using KNN and Cosine Similarity, international journal of engineering research & technology (IJERT) Volume 10, Issue 08 (August 2021)