Resume Analyzer with Job Recommendation

V. Mageshwari [1], M. Kanishca [2]
[1]*Assistant professor, Department of Master of Computer Applications, Christ College of Engineering and Technology*
[2]*MCA, Christ College of Engineering and Technology Moolakulam, Oulgaret Municipality, Puducherry*

***Abstract*—Despite the upsurge in recruitment platforms now present online for job recruitment tools, job seekers find it difficult to find a job that suits their skills, whereas employers are overwhelmed by piles of CVs on their desks. Manual screening is a frustrating task that calls for astute judgment on the part of the employer/employee. The Resume Analyzer cum Job Recommendation System bridges that gap through Natural Language Processing.**

When the resume is uploaded, the major details are extricated, which include skills, qualifications, and work experience, and these are then matched with options that are predefined in job categories. It relies on TF-IDF to derive the attributes from the resume, with the help of which the exact job categories are located using the technique of cosine similarity. Client-side services are based on HTML/CSS, whereas the server is based on Flask/SQLite.

***Index Terms*—Resume Analyzer, Job Recommendation System, Natural Language Processing, Skill Extraction, Flask (Python)**

I. INTRODUCTION

Online recruitment is also utilized in the current recruitment process; thus, the organizations are able to submit their applications electronically via the job portals [5]. However, the system has its limitations with respect to unstructured resume formats and large numbers of applications that are further processed for evaluation [6]. The manual recruitment process for resume screening is an ineffective process that may also include biases and overlooking potential candidates in the recruitment process [7].

Resume analysis by automated systems via NLP is a practical method that can be used to extract the relevant information from the resume and match the appropriate candidates to the correct jobs [8][9]. In effect, the proposed project is to develop a resume analysis automated system that has high accuracy and efficiency in the process of recruiting [10].

II. MAIN OBJECT

The main aim of carrying out this project work is to develop an efficient Resume Analyzer with Job Recommendation system that could recommend appropriate job roles for candidates based on their resume using text processing techniques [1], [4]. The secondary aim of carrying out this project work is to extract the main skills and keywords from the resume using TF-IDF in order to optimize the resume.

III. SYSTEM OVERVIEW

The system allows uploading resumes that are further processed for jobs based on various NLP techniques such as tokenization, stop word removal, and normalization [16]. Further, the regularized text is manifested into numerical vectors using TF-IDF. Vectors are then measured for similarity with stored jobs in the database using a cosine similarity function [17] [18].

IV. SYSTEM ARCHITECTURE

This modular and sequential architecture ensures efficient processing, scalability, and ease of use in the proposed Resume Analyzer with a Job Recommendation System. First, it involves the user interface layer at which candidates upload resumes and then get job recommendations via an interactive, user-friendly interface easily [1], [4]. This layer serves as the main interaction point, allowing seamless data input and result visualization. Once the resumes and job descriptions are retrieved, the system forwards the textual to the preprocessing module, where unwanted

features such as stop words, punctuation, and special characters are eliminated. This phase increases the quality and consistency of the processed data and promotes the quality of the analysis results, as raw and unstructured data is refined and transformed to a clean and standardized form [5], [10]. The refined data is next transferred to the feature extraction module, where the TF-IDF algorithm is used to extract key terms and measure the significance of these terms in the resume dataset [17].

In the matching module, the TF-IDF vectors of the resumes are matched with the vectors of the job descriptions on the basis of the cosine similarity measure, which gives the measure of relevance of candidate profiles with respect to the requirements of the jobs [18]. On the basis of these measures, the system suggests the best-likely openings for the candidates, which is an imperative aspect for appropriate and reliable job matching scenarios [14], [20]. Finally, in the presentation layer, the results of the matching and suggestions for employment are

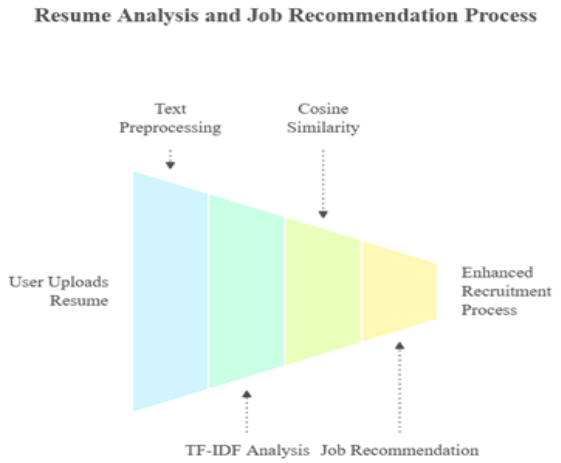

Figure 1. System Architecture

presented in an easy-to-understand form, which helps in proper analysis and decisions for candidates [2] [21] (Figure 1).

V. ALGORITHM

5.1 Text Preprocessing

Text preprocessing is a necessary step in the Resume Analyzer with Job Recommendation System. It cleans the raw content of resumes to make them ready for successful analysis. Usually, a resume includes lots of irrelevant symbols, special characters, formatting tags, and text structures that may badly affect feature extraction. Preprocessing cleans up the extracted text and homogenizes it in such a way that only relevant information is left for further processing.

It indicates the levels of detail that involve quantitative this encompasses conversion to lower case, removal of punctuations, numbers, special characters, and removal of common stop words that do not play an important role in comprehending the context of a resume. Other techniques involved include tokenization to provide proper formatting to the text. Text pre-processing will enable extraction of TF-IDF features along with calculation of cosine similarity to provide an appropriate skill extraction and job recommendation.

5.2 TF-IDF (Term Frequency-Inverse Document Frequency)

TF-IDF gives weights to the terms based on their relevance to a document compared to the whole corpus based on [17].

$$\mathrm{TF}(t, d) = \frac{\text{frequency of t in d}}{\text{total terms in d}}$$

$$\mathrm{IDF}(t) = \log\left(\frac{N}{\mathrm{df}(t)}\right)$$

$$\text{TF-IDF} = \mathrm{TF} \times \mathrm{IDF}$$

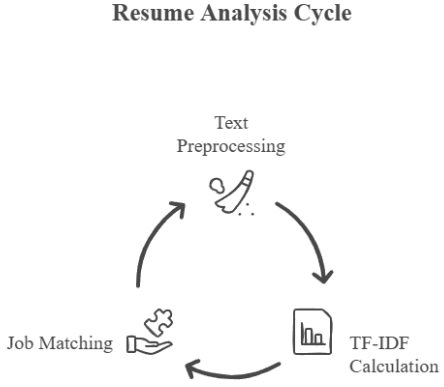

Figure 2. TF-IDF

5.3 Cosine

Cosine similarity is used for measuring the similarity between the vectors of the resume and the vectors of the job description.

$$\text{Similarity} = \frac{A \cdot B}{\| A \| \times \| B \|}$$

The greater the similarity measures, the better matching is between the skills and requirements [22].

VI. DESCRIPTION OF THE DATASET

The proposed Resume Analyzer with Job Recommendation System uses a dataset with a set of sample resumes and predefined job descriptions. Such information would include skills, educational qualification, work experience, and personal details in the resumes that would be represented both in structured and unstructured format. This variation allows the system to assess its effectiveness in handling real-world resume diversity commonly found in online recruitment platforms [6], [10].

The job descriptions in this dataset define certain job roles along with their corresponding required skills, qualifications, and expectations associated with the domain. These predefined job profiles serve as a reference corpus against which resumes of candidates are matched during any job-matching process. Such structured job descriptions help ensure that the relevance between resumes and requirements for jobs using techniques based on similarity is calculated consistently and accurately (Figure 2).

SQLite is a backend database that efficiently manages data, storing resumes, job descriptions, extracted features, and matching results. SQLite is selected for this work because it is light in weight, easy to integrate into any flask-based applications, and its query performance is fast, going well for small- to medium-scale recruitment systems [19]. The organized dataset structure will support reliable storage, quick retrieval, and smooth execution (Figure 3)s.

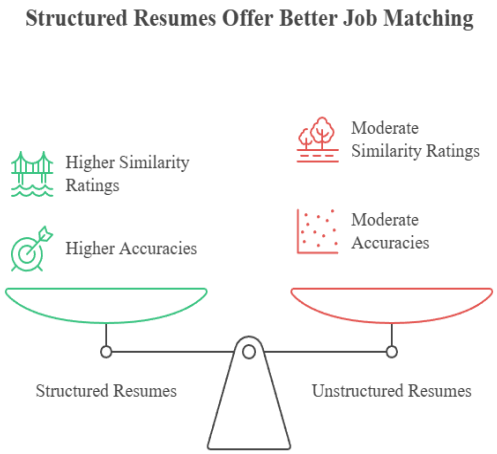

Figure 3. Comparison

VII. RESULTS AND DISCUSSION

T The system was tested with resumes of different skill sets and formats to see its performance in auto job matching.

The results were pretty accurate for the skills extracted and jobs recommended using NLP and similarity-based analysis [11], [17]. Structured resumes scored higher in similarities and achieved higher accuracy due to standardized formatting, but unstructured resumes showed moderate accuracy due to inconsistent text structure [6], [10]. Overall, the combination of TF-IDF with cosine similarity gives reliable performance for resume screening and job recommendation tasks, hence reducing manual efforts in recruitment processes [14],[18],[22].

VIII. PROJECT'S BENEFITS

The Resume Analyzer with Job Recommendation system enables auto resume screening and matching for faster recruitment processing and lesser manual work involved in recruitment tasks [1], [4]. TF-IDF helps to find appropriate skills from resumes, and Cosine similarity allows the exact matching of jobs regarding relevance [17], [18]. This system promotes accurate recommendations, lesser human interference, and enhanced credibility among online recruitment platforms [8], [14]. It also helps job applicants with skill deficiency analysis and resume refinement for better opportunities and hiring [6], [15]. All this makes it an effective, reliable, scalable, and user-friendly approach towards a recruitment platform in a digital era [2] [21].

IX. CHALLENGES AND LIMITATIONS

One of the key challenges of the proposed system is dealing with unstructured and irregularly formatted resumes, which would considerably degrade the skill extraction and job matching performance of the system [6], [21]. This is highly dependent on terminologies used, synonyms, and context in which the information in a resume is provided, which reduces the effectiveness of the analyses based on NLP [11]. The proposed system also depends on predefined job descriptions; hence, it may not adapt well to emerging roles and ever-changing industrial requirements [14]. The TF-IDF-based feature extraction fails to capture

the semantic meaning properly, which may reduce the accuracy in complex matching scenarios [17], [22]. The performance is related to the size and quality of the dataset because poor or imbalanced datasets will affect generalization [10]. Moreover, currently, the system does not have any real-time learning, thus it cannot be continuously improved based on user feedback [20].

X. FUTURE ENHANCEMENTS

Future enhancements of the proposed Resume Analyzer with Job Recommendation System should go toward semantic understanding and adaptability. Advanced deep learning models like BERT or LSTM can be integrated for a better contextual relationship between skills, experiences, and job requirements to provide better accuracy in matching rather than keyword-based approaches [23]. Inclusion of semantic embeddings would serve the system better in synonyms, skill variations, and contextual meanings than the traditional TF-IDF methods [11].

This system can be further extended to include resume scoring and ranking mechanisms that will enable the recruiter to accord priority to candidates in terms of relevance scores and suitability metrics. Features such as skill gap analysis can enable candidates to determine their missing competencies for their job role aspirations, hence providing better career guidance and employability [9], [14]. Furthermore, support for multilingual resumes can be provided to enhance accessibility and usability across user groups [20]. Scalability can be considerably enhanced by integrating real-time data from the job market and adaptive learning mechanisms within the system, thus keeping it relevant to changing recruitment trends [2],[21].

XI. CONCLUSION

The Resume Analyzer with the Job Recommendation System shows the effective application of Natural Language Processing and Machine Learning techniques for the automation and improvement of the recruitment process. By extracting relevant skills and information from unstructured resumes and matching them against predefined job descriptions, the system enhances the accuracy, speed, and consistency of job recommendations [3], [9]. The use of TF-IDF and cosine similarity makes computation of relevance efficient, thereby decreasing manual screening efforts and recruitment bias [14], [17], [18]. Overall, the proposed system provides a scalable, reliable, and user-friendly solution for modern recruitment platforms and serves as a strong foundation for future intelligent hiring systems [2], [21].

REFERENCES

[1] Sharma, M., Automated Resume Screening Techniques, 2021.
[2] Rao, K., AI in Recruitment Systems, 2020.
[3] Gupta, S., NLP Applications in HR Analytics, 2019.
[4] Verma, P., Online Recruitment Portals, 2022.
[5] Jain et al., Digital Recruitment Trends, 2020.
[6] Kumar, D., Challenges in Resume Parsing, 2020
[7] Fernandez, L., Bias in Recruitment Systems, 2019.
[8] Thomas, B., Automation in Hiring, 2021.
[9] Mehta, A., Career Recommendation Systems, 2021
[10] Singh, A., Mining Text for Recruitment, 2020.
[11] Watson, J., Natural Language Processing Fundamentals, 2021.
[12] Brown, L., Methods of Skill Extraction, 2023.
[13] Wilson, K., Job Matching Algorithms, 2020.
[14] Chandra, S., Impact of Recruitment Automation, 2021.
[15] Tan, J., Ethics in AI Hiring, 2023.
[16] Microsoft Docs, Text Preprocessing Techniques, 2023.
[17] Wang & Liu, TF-IDF Based Text Similarity, 2020.
[18] More & Kadam, Cosine Similarity in Job Recommendation, 2021.
[19] SQLite Documentation, Lightweight Database Systems, 2022.
[20] Dawson, R., Scalable Web Architectures, 2022.
[21] Collins, A., Text Preprocessing Techniques. 2023.
[22] Philip, J., Similarity Measures in NLP, 2020.
[23] Das & Bhattacharya, Deep Learning for Resume Analysis, 2022.