Intelligent Resume Analysis and Job Fit Assessment System

[1]Mr Sandeep S Naik, [2] Anvitha G Rao, [3] Kripa K, [4] Krithi K, [5] Reethu K
[1] Assistant Professor, Department of CSE (IoT & Cyber Security with Blockchain Technology), Mangalore Institute of Technology and Engineering, India.
[2, 3, 4, 5] Student, Department of CSE (IoT & Cyber Security with Blockchain Technology), Mangalore Institute of Technology and Engineering, India.

Abstract: The "Intelligent Resume Analysis and Job Fit Assessment System" presents an opportunity to improve and automate the recruitment process through intelligent methods of resume analysis and evaluating job compatibility. Given the huge number of job applicants exponentially increasing, companies have encountered inefficiency in handling manually sifted resumes. We propose a system that uses Machine Learning and Natural Language Processing techniques to automate resume parsing and assess job fit. The system will use TF-IDF to extract meaningful keywords from resumes and job descriptions, and K-Nearest Neighbours to match resumes with the most suitable job roles based on the alignment of skills and experience. Moreover, Cosine Similarity is used in the measurement of similarity of resume and job description with which the prediction of job fit gets even more accurate.

Keywords: resume, machine learning, NLP, KNN, cosine similarity, feedback.

1. INTRODUCTION

With the dynamic and competitive nature of the job market, recruiters are often tasked with reviewing hundreds of resumes for each job opening. This can be very time-consuming and is prone to human error, potentially losing qualified candidates. Automated resume analysis systems have become popular in addressing these challenges. It will use Artificial Intelligence to scan a resume in little time, understand the most appropriate skills and experience, and identify the best matching candidates for any job role. The proposed system will automate the process of parsing resumes and making predictions about jobs using a hybrid approach of machine learning and NLP. This system will help organizations handle large-scale recruitment, reduce human bias, and improve the overall recruitment process by providing objective feedback for resume optimization.

2. LITERATURE SURVEY

2.1 Existing Systems

2.1.1 Resume Parsing Frameworks

Resume parsing is the first phase of resume analysis with AI technology, converting unstructured resume data into structured formats for easier processing. Sajid et al. (2022) proposed a Resume Parsing Framework for E-Recruitment which exploits NLP techniques to structure resumes effectively [1]. This has heavily lifted the efficiency of recruitment as their data extraction process relies on automated unstructured text. Channabasamma et al. (2021) have also developed a Contextual Model for Information Extraction based on the capabilities of SpaCy's NLP, solving the complications arising from different layouts of resumes. This contextual model enhances the precision by taking into account the different resume structures during the parsing process [2].

2.1.2 Automatic Resume Segmentation

The most important elements to be extracted from resumes include education, skills, and work experience. In this regard, Gunaseelan et al. (2020) proposed a Machine Learning Approach for Automatic Extraction of Resume Segments, focusing on these key areas. Their work shows how ML techniques can classify and extract relevant data efficiently, while preserving the integrity of the extracted information [5].

2.1.3 Resume Screening and Matching Algorithms

Advanced algorithms are needed to compare resumes with job descriptions. Sinha et al. (2020) discussed the models like Support Vector Machines (SVM) and K-Nearest Neighbors (KNN) to rank resumes on the basis of skills and experience. The authors demonstrated the possibility of ML algorithms in resume evaluation without human intervention, as it is faster and more accurate [9]. Pudasaini et al. (2022) presented a new approach by integrating Word2Vec and the Gale-Shapley Algorithm. Word2Vec captures semantic and contextual relationships between words, while the Gale-Shapley Algorithm optimizes job-candidate matching, ensuring the best-fit results for recruitment needs [8].

2.1.4 Entity Extraction for Resume Analysis

Entity extraction plays a vital role in efficient resume screening. Jivtode et al. (2023) applied ML and NLP techniques to extract entities like job titles, skills, and education. The unstructured resume data is thus converted into structured formats for further analysis [6]. Das et al. (2018) developed a CV Parser Model integrating Big Data tools with entity extraction techniques. This model processes large-scale resume data in real time, which is appropriate for the organizations managing huge volumes of applicants [12]. Addressing Contextual Challenges in Resume Analysis Despite the developments, contextual challenges continue to persist for AI-based resume analysis. Tikhonova and Gavrishchuk (2019) found segmentation errors to be one significant cause of the problem that is concerned where improper segmentation of resume sections may cause misclassification and inaccurate results. Proper segmentation is significant for accurate resume evaluation [4]. Jivtode et al. (2023) also pointed out the other challenge called contextual limitations whereby AI models fail to infer implicit information such as career gaps, or non-conventional career paths. Enhancing contextual understanding is necessary for fighting these limitations and proper analysis [6].

2.1.5 Qualitative Analysis of Non-Technical Attributes

The AI system has a lot of trouble in assessing non-technical attributes such as communication skills, leadership, and teamwork. The following is an approach to qualitative appraisal by Kudatarkar et al. (2015) for reviewing candidates with other than conventional careers and unearthing unique characteristics [13]. Khan et al. (2023) opined that grading qualitative attributes of a candidate could be challenging. Qualitative aspects are often present in resumes via achievements and rich descriptive narratives rather than declaration [11]. Future systems, therefore, are expected to identify some methods to be able to successfully analyse such traits.

3. METHODOLOGY

3.1 Dataset

The dataset used for this work consists of resumes and job postings from public repositories and recruiter sites. Resumes often contain the following fields of information: candidate's name, contact information, academic history, skills, work experience, and projects. Job postings are comprehensive documents describing the role itself, its responsibilities, mandatory skills, and qualifications relevant to a specific job position. Such a structured dataset, therefore, serves as the premise for training and testing of the system.

3.2 Data Pre-processing

Ensures consistency and readies it for analysis through several pre-processing steps. Resumes, job descriptions are tokenized to break text into the individual words followed by deletion of stop words like "the" and "is", the meaningful content being left inside them. The reduction of each word back into its base form is an outcome that keeps various types consistent to form lemmas. Finally, the resume format and job description format becomes standardized to transform unstructured data into a structured form that is amenable to analysis by the system.

3.3 Features Extracted

It identifies and extracts specific features from resumes and job descriptions to help predict an accurate job-role. Skills are identified using predefined skill lists and contextual analysis, while education is extracted using keyword patterns that signify degrees or certifications. Work experience is parsed through the analysis of temporal phrases and action verbs. Keywords from resumes and job descriptions are then retrieved using the statistical TF-IDF method, which aims at highlighting the most important terms in the text.

3.4. Job Role Prediction Models

The system predicts the best-suited job role for candidate by using the K-Nearest Neighbours algorithm. Feature extraction from resumes is done against feature extraction from job descriptions and accordingly, the algorithm tries to find the closest matches based on the similarity of those features. This way, it can ensure that the candidates are matched against suitable roles that fit them perfectly in terms of skills and experience. The system predicts the most apt job role for a candidate by combining NLP techniques with the KNN algorithm. pre-processing of textual data extracted from resumes and job descriptions using tokenization, removal of stop words, and lemmatization allows for clean and uniform representation.

3.5 Similarity Calculation

Cosine similarity calculates the alignment between resume's keyword and keywords in job descriptions. It's a kind of similarity measure that uses cosine to compute the cosine value between the angle two vectors possess where those two vectors are created in response to two keywords such as in the resume and in the job description. Therefore, an individual can tell when there's good match with a person's qualification or requirement while trying to rank the resume in effective means. There is one mathematical formula by which cos similarity is obtained, i.e.

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

A and B be the resume and job description feature vector, respectively. $A \cdot B$ is the Dot product of two vectors. $\|A\|$ and $\|B\|$ denotes the magnitudes (norms) of the respective vectors. By this method, the resemblance between resumes and job descriptions can be calculated precisely, using all the features drawn.

3.6 Scoring of Resume

Resumes are scored through an evaluation process. Several factors, including a skill match, experience, and educational qualifications, constitute the scores assigned. This scoring will involve weights applied to the relevant skills so that matching of relevant experience and education against job requirements contributes more points to the scores. Thus, an objective final score, representing a weighted sum of all factors, evaluates whether the resume is appropriate for any specific job.

3.7 Feedback System

A feedback system is incorporated to assist candidates in improving their resumes. The system identifies deficiencies, such as missing relevant skills or projects, and provides actionable suggestions for enhancement. These recommendations include adding industry-specific keywords, highlighting achievements, and restructuring the resume layout for better clarity and impact. This feedback not only helps candidates improve their job prospects but also ensures that resumes are more tailored to meet employer’s expectations.

4. SYSTEM ARCHITECTURE

The system architecture of a Resume Analyzer is drawn in such a way using multi-layer and modular architectures for effective processing and analyzing resumes. The entire process begins with the Input Layer, which is meant to receive resumes in different formats, PDF, Word, plain text, among others. This would basically create a File Upload Interface for recruiters to upload resumes together or even separately. The Document Parser documents the files into the raw text with retaining relevant structure elements, which might be sections such as "Education," "Skills," and "Experience."; this ensures that no critical information is lost through the text extraction phase. Afterward, the data is sent to the pre-processing layer, where it gets cleaned to remove noise (e.g., special characters, unnecessary formatting) and normalized. That way, it ensures that all data is consistent and at that point ready for analysis. It is then possible to apply advanced NLP on the Feature Extraction Module to draw and categorize professional skills, academic qualifications, years of experience, and job titles. Such features were extracted in structures meant to machine readable formats to be stored into the Central Relational Database. This particular database does not just contain filled resume data but also consists of the different job descriptions and its predefined requirements for comparative consideration with respect to compatibility. The whole concept of the system is contained within the Matching and Scoring Engine, which uses machine learning algorithms and similarity metrics (TF- IDF, cosine similarity) to evaluate resumes against job requirement criteria for determining each resume's compatibility score and ranking them in accordance with their fit to the job description.

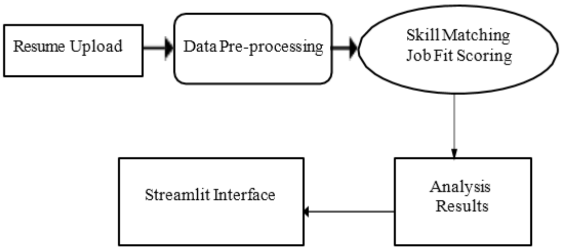

Fig 4.1 System Architecture

The processed results are made available at that stage which is termed as Visualization and Reporting Stage, and that stage directly forwards the results to recruiters on an intuitive dashboard where recruiters can see the key insights which include ranked leaders, matched skills with this list, compatibility scores, areas of improvement, etc. All this information can be presented in a user-friendly format. It also generates reports to be exported for external usage. Another aspect of making the model more accurate and adaptable at the same time is the presence of a Feedback Module in the architecture, where the recruiter gives a feedback recommendation based on the recommendations of the system regarding a candidate. This feedback is fed back into the machine learning models to improve their predictions and make future analyses even better. The Resume Analyzer combines automation with intelligent data-driven decision making and continuous learning to make a highly tough-cut process for recruitment with less manual effort at the same time improving recruitment outcomes.

5. TESTING AND SNAPSHOTS

The Intelligent Resume Analysis and Job Fit Assessment System was experimentally validated for effectiveness through a test-run using 500 resumes as well as 100 descriptions of the job which obtained from the public database along with different job portals so that it analyses the resumes correctly, forecasts the chances that it suits the type of job, and provides adequate feedback. All resumes and job descriptions went through pre-processing, including tokenization, stop-word removal, lemmatization, and format standardization. Major features like skills, education, work experience, and certifications were extracted from resumes using the TF-IDF method to match them up with job descriptions.

Fig 5.1 Home Page

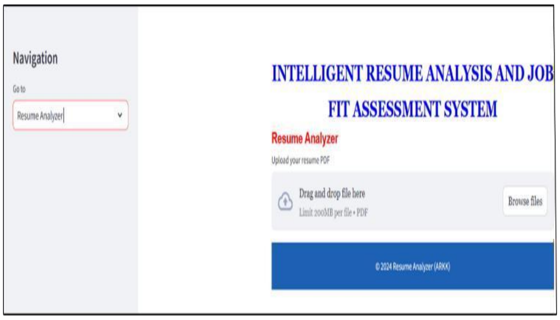

Fig 5.2 Browse and upload files

Fig 5.3 Top Matching Job titles

The K-Nearest Neighbours (KNN) algorithm and Cosine Similarity were used to predict job-role suitability by measuring the alignment between resume and job description vectors. Each resume was scored based on factors such as skill match, experience, and education, which were weighted according to their relevance to the job. The system also provided feedback for resume improvement. The performance of the system was evaluated in terms of precision, recall, and F1-score. Precision is 85.7%, and recall is 82.4%, while F1- score was 84.0%, implying high accuracy. The average cosine similarity score is 0.76, which strongly indicates alignment between resumes and job descriptions. The feedback system worked: 70% of respondents said that the suggestions improved their resumes. Pie charts showed that 65% of resumes had strong skill matches, 72% met educational requirements, and 80% had relevant work experience.

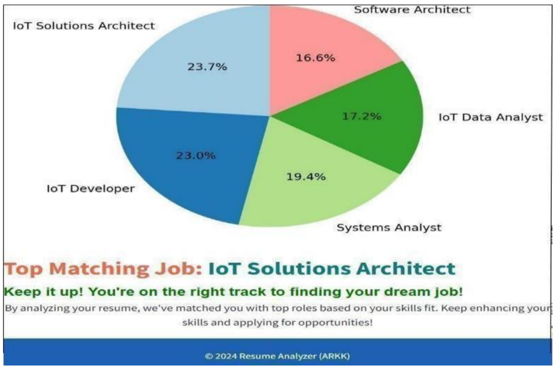

Fig 5.4 Pie chart visualization

Snapshots of the system's user interface demonstrate key features such as:

i. Resume Upload: Users can drag and drop resumes into the system for analysis.
ii. Compatibility Reports: The processed scores are presented to the user, along with rankings of candidates to the role.
iii. Detailed feedback for candidates that also gives actionable suggestions for their improvement through making their resume better in terms of aligning with the job requirements.

7. CONCLUSION

The "Intelligent Resume Analysis and Job Fit Assessment System" greatly improves the efficiency and effectiveness of recruitment processes by leveraging Machine Learning (ML) and Natural Language Processing (NLP) techniques. The system analyzes resumes and job descriptions using TF-IDF for keyword extraction, KNN for job-role prediction, and Cosine Similarity for precise matching. The system closes the gap between the applicant and the employer, allowing applicants to spend less time and efforts on the process of screening resumes by manually giving proper job-role predictions and providing useful feedback for candidates. It ensures objective compatibility scoring from the perspective of candidate-job compatibility. It empowers applicants with useful information from the feedback system, improving their resumes. The overall system promotes better hiring choices and improved recruiting effectiveness by an employer while improving the opportunities for jobs applicants.

REFERENCES

[1] H. Sajid, J. Kanwal, S. U. R. Bhatti et al., "Resume Parsing Framework for E-recruitment," in proceedings of the 16th International Conference on Ubiquitous Information Management and Communication (IMCOM), 2022.

[2] Channabasamma, Y. Suresh, and A. M. Reddy, "A Contextual Model for Information Extraction in Resume Analytics Using NLP's Spacy," International Journal of Artificial Intelligence and Soft Computing (IJAISC) in Volume 10, Issue 10, 2021.

[3] A. K. Sinha, M. A. K. Akhtar, and M. Kumar, "Break Down Resumes into Sections to Extract Data and Perform Text Analysis using Python," International Journal on Recent and Innovation Trends in Computing and Communication, 11(6s), 391–400, 2023.

[4] M. Tikhonova and A. Gavrishchuk, "NLP methods for automatic candidate's CV segmentation," International Conference on Engineering and Telecommunication, 2019.

[5] B. Gunaseelan, S. Mandal, and V. Rajagopalan, "Automatic Extraction of Segments from Resumes using Machine Learning," IEEE 17th India Council International Conference (INDICON), 2020.

[6] A. Jivtode, K. Jadhav, and D. Kandhare, "Resume Analysis using Machine Learning and NLP," in International Research Journal of Modernization in Engineering Technology and Science, 2023.

[7] U. Goyal, A. Negi, A. Adhikari et al., "Resume Data Extraction Using NLP," in Journal Innovations in Cyber Physical Systems 2020.[11:54 PM]

[8] S. Pudasaini, S. Shakya, S. Lamichhane et al., "Scoring of Resume and Job Description Using Word2vec and Matching Them Using Gale Shapley Algorithm," in Lecture Notes in Networks and Systems,vol 209, 2022.

[9] A. K. Sinha, M. A. K. Akhtar, and A. Kumar, "Resume Screening Using Natural Language Processing and Machine Learning: A Systematic Review," Machine Learning and Information Processing. Advances in Intelligent Systems and Computing, vol 1311 , 2020.

[10] M. Harshada, A. N. Kshirsagar, N. Sonali, and A. Pimparkar, "NLP Based Resume Parser Analysis," Journal of Emerging Technologies and Innovative Research, vol. 1, no. 1, pp. 1-10, May 2023.

[11] [E. Khan, H. Patel, A. Shaikh et al., “Resume Parser and Summarizer,” International Journal of Advanced Research in Science, Communication and Technology, vol. 2, no. 2, pp. 11-20, Apr. 2023.

[12] P. Das, M. Pandey, and S. S. Rautaray, “A CV parser Model using Entity Extraction Process and Big Data Tools,” International Journal of Information Technology and Computer Science vol. 10, no. 9, pp. 21-31, Sep. 2018.

[13] V. R. Kudatarkar, M. M. Ramannavar, and N. Sidnal, “An Unstructured Text Analytics Approach for Qualitative Evaluation of Resumes,” International Journal of Information Technology & Computer Science, vol. 7, no. 4, pp. 49-80, 2015.

[14] SriUsha, K. Rasmitha Choudary, Kavitha C. R., T. Sasikala, Data Mining Techniques used in the Recommendation of E-commerce services, Proceedings of 2nd International Conference on Electronics , Communication and Aerospace Technology (ICECA 2018), March 2018, pp: 379 –382.

[15] Vidya Kumari K. R., Kavitha C. R., Data Mining for the Social Awareness of the Social Networks, 3rd International Conference on Computational System and Information Technology Sustainable Solutions (CSITSS 2018), December 2018, pp: 7 17.

[16] Kavitha C.R., Rajarajan S. J., R. Jothilakshmi, Kamal Alaskar, Mohammad Ishrat, V. Chithra, Study of Natural Language Processing for Sentiment Analysis, 2023 3rd International Conference on Pervasive Computing and Social Networking (ICPCSN), 19- 20 June 2023.

[17] Tamizharasan, M., Shahana, R.S., Subathra, P.“Topic modeling-based approach for word prediction using automata”,Journal of Critical Reviews,PP 744- 749,2020.

[18] Bindu K.R., Aakash C., Orlando B., Latha Parameswaran p, “An algorithm for text prediction using neural networks”, Lecture Notes in Computational Vision and Biomechanics, Springer Netherlands, Page No.186 – 192, ISSN: 22129391.

[19] Murali, S., Swapna, T.R.”An empirical evaluation of temporal convolutional network for offensive text classification” 2019,’International Journal of Innovative Technology Engineering,Volume8, 2184,Article,Scopus and ,Issue Exploring 8,PP-2179

[20] Radhika. G., Narasimha Prasath, K., Rajasundari, T., “Sentimental analysis on Tamil news feed,”Journal of Advanced Research in Dynamical and Control Systems,Journal of Advanced Research in Dynamical and Control Systems,Volume 9 ,Issue 11,Page no : 397- 399,Scopus – October 2017,ISSN:1943 023X,Publisher: Institute of Advanced Scientific Research.

[21] Sini Raj P,Raj Vikram R.,Swathi Lakshmi. B, “Sentiment analysis using deep learning technique CNN with K Means,”International Journal of Pure and Applied Mathematics,Volume 114, Issue 11,Page no: 47-57,Scopus -September 2017,ISSN: 13118080,Publisher: Academic Press.

[22] Vamsee Krishna Kiran., Vinodhini, R.E., Archanaa, R., Vimalkumar, K.,” User specific product recommendation and rating system by performing sentiment analysis on product reviews,” 2017 4th International Conference on Advanced Computing and Communication Systems, ICACCS 2017., 6-7 Jan. 2017, Venue: Coimbatore, India, Publisher: IEEE, Scopus – November 2017, ISBN: 978-1 5090- 4559-4, Article number :8014640.

[23] Dhanya N.M, Harish U.C., “Sentiment analysis of twitter data on demonetization using machine learning techniques”, Lecture Notes in Computational Vision and Biomechanics, Springer Netherlands, Page No.227 – 237, ISSN: 2212939