

Enhancing the Accuracy of Candidate Selection for Hiring Using Natural Language Processing In Artificial Intelligence

Assistant Professor Shanmuga Priya K, Karan Singh D, Bharanitharan S, Manikandan S

Department of Information Technology K. S. R College of Engineering
Tiruchengode, India

Abstract- To leverage towards better hiring and recruiting methods the traditional review of resumes is often carried out via simple keyword matching. It records only superficial abilities and credentials. Although candidates are shown the basic skills and qualifications, it misses the deeper competencies such as solving problems, personal development and contributions to projects. This paper proposes an artificially intelligent approach which combines machine learning (ML) for assessing abilities and potential, natural language processing (NLP) for reviewing resumes and job descriptions, and graph-based career mapping to visualize career progression. Compared with traditional resume scoring models, this proposed methodology presents more informed candidate evaluation on all factors including context, experience and growth potential. Professional network analysis, accuracy and quality of input data, as well as candidate skill alignment is one of the important aspects of the proposed model. The graph-based method presents some career paths and the practical contributions to the study via mapping abilities over time. By using our proposed technology approach, hiring decisions can be improved and position matching can be performed optimally. Moreover, context-aware analysis can provide an accurate evaluation of candidate potential. In the field of HR technology, this innovative method has a new standard in fair and intelligent talent evaluation.

Keywords- Artificial Intelligence (AI), Machine Learning, Graph- Based Approach, Human Resource.

I. INTRODUCTION

Through the use of advanced NLP methods, the proposed solution, "Improving Candidate Selection Accuracy for Hiring through Natural Language Processing in AI," brings a groundbreaking perspective to modern recruitment. Conventional keyword-driven systems often struggle to understand the intricate context of candidates' skills and experiences, leading to inaccurate matches and missed opportunities. Contrarily, our system adopts an even more holistic view of resume evaluation: it involves sophisticated text processing techniques

such as contextual understanding, semantic analysis and cosine similarity, leveraging both keyword ingestion and context with which it occurs. By observing both the presence of keywords and the context in which they occur, this system can detect differences between trivial mentions and relevant hands-on experience, or can obtain a higher level of detailed prediction of technical competence by identifying candidates who have "built scalable APIs using Python" from one candidate who only mentions "Python" as a skill. Lastly, this system can also perform career progression assessment using graph-based career mapping, that represents

changes in skill development, job position and project impact over time, thereby reducing match mismatches, expediting the hiring process, and ensuring accurate assessment of high potential candidates (even in cases where they employ unconventional terminology). A continuous improvement of machine learning models with a candidate-job match added to the model serves to enable incremental improvements on the platform for the system to be more adaptable to changes in the labour market and emerging skills; in general, this will increase the efficiency of hiring, hence more reliable match between the candidate and the job, which will in turn reduce the cost of hiring and time to hire.

II. RELATED WORKS

[Arwa Najjar, Belal Amro, Mário Macedo [2020] – “A Smart Decision Support System for Hiring: Resume Filtering and Candidate Evaluation” In this research we provide a smart decision support system (SDSS) aimed at more automatic filtering of resumes and assessing the candidates using Machine Learning (ML) and Natural Language Processing (NLP). The system consists of three main elements: a Training Block (which learns from historical resumes), a Matching Block (which aligns resumes with job postings), and an Extracting Block (which evaluates candidates based on semantic relevance). Experimental findings demonstrate high precision and effectiveness, establishing I-Recruiter as a beneficial tool for hiring professionals. Further improvement is possible, however, because biases in training datasets and management of unstructured resumes are observed.

Mayuri Gund et al. ([2024]) – “Analysis and Interpretation of Resumes using Language Models (LLM)” Exploring the challenges of traditional recruitment approaches, which are lengthy, biased, and heavily manual work, this paper describes an application for optimizing HR processes using LLMs which leverages Natural Language Processing (NLP), Machine Learning (ML), and analytics to help improve hiring effectiveness. The system proposed automates necessary HR tasks such as job description assessment, ranking CVs, creating

screening questions, sending automated email notifications, and providing interview support. Through integration with existing HR tools, the new system enhances decision making and workflows while improving employee experience. Although the proposed system reduces the burden of manual tasks, decreases biases in hiring, and relieves the burden on managers, the system has drawbacks such as concerns about data privacy, complexity in integration, and reliance on high quality training data.

Shubham Bhor et al. [2024] – “Resume Parser: Use of Natural Language Processing Techniques” Examining the challenges faced in recruiting a candidate for a job, and proposing an NLP-based resume parser to automate the evaluation of candidates, the system takes structured data from resumes and organizes relevant sections, such as education, work experience and projects into one of the four dimensions:

(1) organization (1) matching (2) evaluation (3) ranking (4) highlighting A post's relevance to the company and job specifications The job portal consists of three components— employee, employer and administrator—with the administration headed by the administrator. Though, the proposed model considerably improves screen of resume and reduces the difficulties encountered in hiring, it also presents challenges like data extraction accuracy (5) biases in ranking (6) compatibility with corporate HR systems (AI has the potential to revolutionize the hiring process).

According to Chidinma A. Nwafor et al (2021), an Automated Multiple-Choice Question Generation system was developed using NLP strategies. The proposed system starts by extracting key concepts of the text using Named Entity Recognition (NER) and dependency parsing. Then using BERT model is used to check semantic similarity and the generated questions match closely with the main ideas in the text. The distractions are created using WordNet and TF-IDF ranking to select words that are contextually similar but distinct from the correct answer. However, difficulties in selecting distractors were observed. Some of the incorrect option

appeared too obvious and/or irrelevant to the contextual context. Also, the system requires the human inspection to fine-tune the generation of the questions and validate them. To mitigate these problems, the proposed system will be improved by using context aware models such as GPT for better distractor selection, and by applying reinforcement learning techniques to improve question relevancy over time.

Lino Mathew & Nithin C George [2020] – “ATS BREAKER” - A System for Comparing Candidate Resume and Company Requirements This paper presents an NLP-driven system aimed at helping job seekers optimize their resumes for Applicant Tracking Systems (ATS). The system identifies both technical and soft skills from resumes whether in text, PDF, or DOCX formats, assesses them against job specifications, and offers suggestions for improvement. Furthermore, it utilizes NLP methods such as Named Entity Recognition (NER) and text segmentation for grammar correction. The system is built using React and Flask, facilitating efficient processing and user engagement. Though this approach increases the likelihood of success for job seekers, it faces challenges in adjusting to various ATS algorithms and enhancing context-aware resume evaluation.

III. EXISTING SYSTEM

Contemporary recruitment software often employs a technology known as keyword screening. Such systems look for keywords in resumes that match the job specifications, ignoring information such as skills, education, and work experience. Although this process can greatly speed up the candidate screening process, it tends to miss the entire picture of experience, context, and problem-solving skills. Therefore, more qualified applicants can be rejected if their resumes do not contain appropriate keywords, whereas less qualified candidates could get through on the basis of including the correct words in their applications. A common tactic to avoid such problems is for the applicants to fill their resumes with key words—this approach tends to bring better ranks even though essential aspects of the resume are missing. Some of the other

significant features of this strategy are utilizing non-traditional formats to enhance effectiveness. In addition, as opposed to previous systems, keyword stuffing means filling a resume with many keywords (e.g., "customer service," "job interviewer," and "failsafe technician") so that less qualified applicants can look more skilled than actually competent candidates. In addition, some applicants use popular resume formats to circumvent the restrictions of computerized screening systems. For example, elaborate designs or innovative layouts might make parsing algorithms scratch their heads, potentially granting such resumes an unintended edge. That is why greater complexity screening systems beyond mere keyword matching, employing context-sensitive testing, machine learning, and NLP are essential to properly test the actual significance of a candidate's experience.

IV. PROPOSED SYSTEM

The intended AI-based resume analysis system attempts to break the limitations of traditional keyword matching by incorporating Natural Language Processing (NLP) and Machine Learning (ML) techniques for more contextual and comprehensive assessment of candidates. Instead of mere existence of keywords, this system takes into account the usage of skills and identifies the difference between genuine expertise and mere keyword listing. In another scenario, a resume that has "Python" in sentence stating "developing machine learning models" will score higher than the one that just states the programming language. It also applies graph-based career mapping to evaluate career development, project complexity and industry applicability so it can be more successful at identifying high-potential applicants. The algorithms used are Keyword Matching to identify terms, TF-IDF to rank terms, Cosine Similarity to check overall compatibility between job advertisements and CVs, Rule-Based Filtering for eliminating unqualified candidates and Weighted Scoring for bringing to the fore essential skills. Metrics available that shall be utilized for assessing the usefulness of the system like data accuracy of extraction, context recognition, and data enrichment via professional networks to

ensure immaculate assessment of candidates. The required software stack consists of React.js for the frontend UI, Node.js with Express.js for the backend, MongoDB for data storage, and AI libraries such as TensorFlow for NLP, with all of them running on Linux servers and version controlled using GitHub.

V. SYSTEM DESIGN

Architecture Overview

The architecture makes use of a modular and flexible framework, which includes frontend, backend, and data components for processing efficiently. The frontend is developed with React.js to give an interactive user experience, whereas the backend runs on Node.js and Express.js to handle the application logic. For data storage of the needed amount of data, the application makes use of MongoDB. This architecture enables the deployment of NLP and machine learning algorithms, which are critical to thoroughly scrutinize resumes. It also provides data exchange APIs to integrate with different HR systems and platforms, like LinkedIn, to further enhance the candidate profiling process. It has real-time capabilities, high availability, and secure storage. The system accommodates highly available and reliable operation with load balancing, horizontal scaling, and database replication approaches and could utilize caching processes for read-intensive data to reduce latency and response time. While this process can significantly accelerate the process of screening candidates, it often misses the big picture of experience, context, and problem-solving abilities. So, more capable candidates may be disqualified if their resumes lack suitable keywords, while less capable candidates might pass based on the fact that they've used the right words in their applications. This underscores the necessity for more advanced screening techniques that transcend simple keyword matching, including context-aware analysis, machine learning, and natural language processing (NLP) to properly evaluate the actual relevance of a candidate's experiences.

Data Flow and Processing

The data flow of this system starts by collecting resumes from various sources including PDF / DOCX files, LinkedIn profiles and various others. Next, the collected data is pre-processed to remove all extra information, standardize the body of the text, and break down the contents of documents into different sections such as Education, Work History, Skills and finally, Natural Language Processing (NLP) algorithms are used to extract entities, evaluate context and uncover connections between skills and job roles. Additional topics include the development of machine learning models which combine data from candidates with prior knowledge of their skills and experiences to generate a profile of the candidate for ranking and matching applications.

Scalability and Performance Optimization

Aside from the data processing pipelines and load balancing techniques, efficient data processing pipelines and load balancing techniques are used. Caching is used to reduce response time and database optimization is used to retrieve data faster. Distributed architectures and microservices make it possible to handle high traffic and large volume of resume data without loss of resources and compromising speed or reliability. This approach supports real-time candidate screening, even during peak load periods, ensuring a seamless user experience.

Performance Metrics

Security is a critical component of the system design, focusing on data protection, encryption, and user privacy. It employs secure data transmission protocols, role-based access control, and multi-factor authentication to prevent unauthorized access. Sensitive data, including personal and professional information, is encrypted both at rest and in transit, aligning with data protection regulations like GDPR. Regular security audits and vulnerability assessments further enhance the system's resilience against cyber threats.

Use of the Proposed System

A number of essential factors were taken into account when The proposed AI based resume analysis system has been specifically developed in such a way that it is both precise, scalable and fair enough to be used today's recruiting process. One of the many factors that are considered is the accurate mapping of technical and soft skills on a resume. This is done by means of Natural Language Processing (NLP) and machine learning algorithms that go beyond simply matching keyword matches and use contextual information about skills mentioned within a resume to get a more complete picture of the skills mentioned by the candidate. Scalability and performance were also considered in the design to allow for efficient processing of thousands of resumes in real-time. Through this process the time-to-hire is significantly shortened as well as overall recruitment speeds being improved making the system ideally suited to enterprises of all sizes – from startups to multinational corporations. We also designed the system to be flexible and can take a wide variety of resume formats, which means that no information is lost due to differences in resume format resulting in a more comprehensive view of a candidate's potential suitability.

F. Flow Diagram

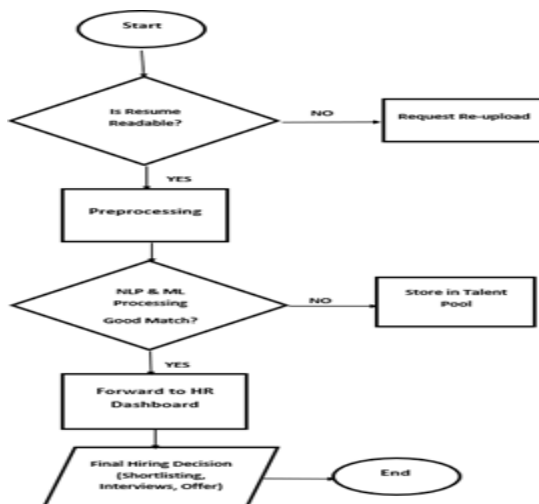


Figure 1. Block diagram

The system architecture proposed is designed to accommodate cutting edge Natural Language Processing (NLP), machine learning (ML) and graph based analytics to provide a candidate evaluation platform. In the system at the heart there are three major modules: Data Ingestion and Preprocessing, Contextual Skill Analysis and Career Trajectory Mapping. All are connected to a Knowledge Graph for data integration.

The Data Ingestion and Preprocessing module applies a standard set of extraction methods to extract resume data from various formats like PDF, DOCX, TXT. Also, it integrates external data sources like LinkedIn profiles, GitHub repositories and industry specific platforms. It uses NLP techniques to clean and normalize the data to ensure accurate analysis. This is then consumed by the Contextual Skill Analysis component to cross-examine all that information and determine the context and depth of each skill, going far beyond a simple keyword match. The Contextual Skill Analysis utilizes deep learning algorithms like BERT or RoBERTa to decipher how skills, projects, and job functions are related. And, it also uses entity recognition to identify technical proficiencies, leadership positions, and problem-solving experiences.

The Career Trajectory Mapping component tracks a candidate's professional milestones over time. The tracking recognizes key milestones such as promotions, management positions, and substantial contribution to projects. It primarily uses graph-based analytics to identify patterns in career progress, which gives a much more accurate representation of a candidate's experience and potential. Finally, the Knowledge Graph acts as a bridge between the data from various sources for the system to understand the full context of a candidate's career path and it captures connections between skills, roles, industries and career paths to provide a more complete and well-connected view of each candidate.

All the process details are provided in these 7 Steps.

Step 1: Uploading/Initializing – This involves creating an account and an individual Upload their

resume in the system. The system will review whether the format of the resume is acceptable, if so, continue on the preprocessing stage, if not, prompt them to upload again.

Step 2: Preprocessing – The system gathers and cleans the text from the resume.

Step 3: The system scores the qualifications of the candidate using natural language processing and machine learning algorithms.

Step 4: When the candidate fulfills the criteria the resume is forwarded to the HR dashboard.

Step 5: If the candidate does not qualify, the CV is kept with more possibilities.

Step 6: Once the final hiring decision has been made, shortlisting, interviewing and job offers are made. **Step 7:** Finalization – The process concludes here.

E. Result Analysis

Candidate Resume (PDF Name)	Existing System	Proposed System	Improvement (%)
SAMPLE1.PDF	0.42	0.491	+7.1%
SAMPLE2.PDF	0.39	0.472	+8.2%
SAMPLE3.PDF	0.28	0.331	+5.1%
SAMPLE4.PDF	0.35	0.420	+7%
SAMPLE5.PDF	0.26	0.331	+5.1%
SAMPLE6.PDF	0.24	0.300	+6%
SAMPLE7.PDF	0.18	0.290	+11%
SAMPLE8.PDF	0.00	0.00	NO Match

Table 1. Cosine Similarity Improvement in Resume Matching

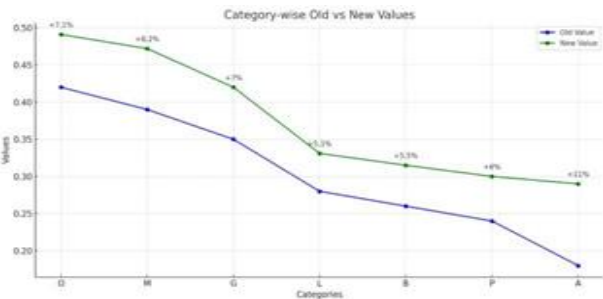


Figure 2. Comparison graph

Fig. 2 shows the comparison of the accuracy improvement in candidate and job match between the existing system and proposed systems (in this case, resume matching accuracy) using the accuracy measure cosine similarity. The accuracy value is calculated based on a widely used metric that measures textual similarity between candidate resumes and the target job description. Cosine similarity measures the angle between representations of the text; therefore, it is an approximate quantitative measure of how closely two documents are aligned. The accuracy results shown by this table are consistent across the various samples evaluated. For example, from Fig. 2 the proposed system had a gain in accuracy of 5.1% through 11%. The consistent gains in accuracy are attributed to the system's ability to capture more relevant features, such as keywords context, semantic meaning and skills relevant for the role. These steady gains on average suggest that the new system can accurately interpret and match resumes based on the nuanced requirements of each job profile, and thus reduce the likelihood of false positives in candidate selection, thereby improving candidate targeting. Fig. 2 also shows further evidence that a large increase in the accuracy results can be seen in some samples, suggesting that the proposed system has the capacity to improve candidate and job alignment; i. e. more precise candidate screening and a higher match rate.

VI. CONCLUSION

The proposed system, called Enhancing the Accuracy of Candidate Selection for Hiring Using Natural Language Processing in Artificial Intelligence, is a breakthrough in modern recruitment process and uses state-of-the-art NLP techniques. Keyword-based systems are usually mismatched between applicant and employer with few relevant hands-on experience, because candidates that can not handle the task in question have non-relevant hands-on experience. In contrast, this novel system employs a powerful text processing approach (combined with cosine similarity, semantic analysis, and contextual

understanding) to interpret resumes quantitatively. With the combination of keyword detection — by analyzing the representation of keywords and contextual context — the system can identify both superficial mentions and the more meaningful hands-on experience, which ensures a more accurate estimation of technical expertise. Resume natural language processing is made better; a career progression over time is also evaluated in terms of graph-based career mapping. For example, the system can identify patterns in candidate acquisition of skills over time, job roles and project impact. So the whole package reduces mismatches in comparison process, streamlines the recruitment process and ensures that high potential candidates get accurately identified even when applicants use non-standard terminology. A lot of features, like machine learning models, allow the system to continuously improve its understanding of current market trends and emerging skills. This allows matching candidates with specific jobs and reducing hiring costs and time to fill the positions.

Future Work

To overcome current shortcomings of career growth monitoring, problem solving assessment and keyword-based resume screening, future improvements must be based on AI- based contextual evaluation and natural language processing (NLP), technologies that provide excellent chances to improve accuracy and fairness of talent assessment over static keyword- focused approaches. AI-based systems can be used to quantify problem-solving abilities more efficiently than before. They could not only tell if candidates have certain skills, but also look for how they have been using those skills in real life situations. This would help you better understand a candidate's critical thinking, creativity, and flexibility – and therefore more accurately define their professional skills. Also, to prevent people with experience from being excessively ranked based only on tenure it is possible to use career trajectory analysis. This analyzes promotions, leadership positions, and the progression in overall responsibility leading up to the person's current level of success in their field.

REFERENCES

1. Alamelu, M., Kumar, D., Sanjana, R., Sree, J., Devi, A., & Kavitha, D. (2021). Resume Validation and Filtra tion using Natural Language Processing. 2021 10th International Conference on Internet of Everything, Microwave Engineering, Communication and Networks (IEMECON), 1–
2. Ali, I., Mughal, N., Khan, Z. H., Ahmed, J., & Mujtaba, G. (2022). Resume Classification System using Natural Lan guage Processing and Machine Learning Techniques. *Mehran University Research Journal of Engineering and Technology*, 41(1), 65–79.
3. Bharadwaj, S., Varun, R., Aditya, P. S., Nikhil, M., & Babu, G. C. (2022). Resume Screening using NLP and LSTM. 2022 International Conference on Inventive Computation Technologies (ICICT), 238–241.
4. Cabrera-Diego, L. A., Durette, B., Lafon, M., Torres-Moreno, J.-M., & El-Bèze, M. (2015,). How Can We Measure the Similarity Between Résumés of SelectedCandidates for a Job?. *International Conference on Data Mining DMIN* 2015.
5. Chou, Y.-C., & Yu, H.-Y. (2020). Based on the application of AI technology in resume analysis and job recommen dation. 2020 IEEE International Conference on Compu tational Electromagnetics (ICCEM), 291–296.
6. Daryani, C., Chhabra, G. S., Patel, H., Chhabra, I. K., & Patel, R. (2020, January). An Automated Resume Screening Sys tem Using Natural Language Proc
7. FraiJ, J., & László, V. (2021). Literature Rev essing and Similarity. *Ethics and Information Technology*.
8. view: Artificial Intelligence Impact on the Recruitment Process. *International Journal of Engineering and Managemen Sciences*, 6(1), 108–119.
9. Gan, C., Zhang, Q., & Mori, T. (2024). Application of LLM Agents in Recruitment: A Novel Framework for Auto mated Resume Screening. *Journal of Information Pro cessing*, 32(0), 881–893. <https://doi.org/10.2197/ipsjjip.32.881>

10. Harsha, T. M., Moukthika, G. S., Sai, D. S., Pravallika, M. N. R., Anamalamudi, S., & Enduri, M. (2022). Automated Resume Screener using Natural Language Processing(NLP). 2022 6th International Conference on Trends in Electronics and Informatics (ICOEI), 1772– 1777.
- Kino, Y., Kuroki, H., Machida, T., Furuya, N., & Takano, K. (2017). Text Analysis for Job Matching Quality Improvement. *Procedia Computer Science*, 112, 1523–1530
11. Lad, A., Ghosalkar, S., Bane, B., Pagade, K., & Chaurasia, A. (2022). Machine Learning Based Resume Recommendation System.
12. International Lalitha, B., Kadiyam, S., Kalidindi, R. V., Vemparala, S. M., Yarlagadda, K., & Chekuri, S. V. (2023). Applicant Screening System Using NLP. 2023 International Conference on Innovative Data Communication Technologies and Application (ICIDCA), 379–383.
13. Li, C., Fisher, E., Thomas, R., Pittard, S., Hertzberg, V., & Choi, J. D. (2020). Competence-Level Prediction and Resume & Job Description Matching Using Context-Aware Transformer Models. (EMNLP), 8456– 8466
14. A., & Dubey, M. S. (2022). CV Analysis Using Machine Learning. *International Journal for Research in Applied Science and Engineering Technology*, 10(5), 1316–1322.
15. Aydin, O., Karaarslan, E., & Narin, N. G. (2024). Artificial Intelligence, VR, AR and Metaverse Technologies for Human Resources Management. <https://doi.org/10.48550/ARXIV.2406.15383>