Smart Resume Screening and Matching System

Siddharth Dogra
Department of Computer Science and Engineering
B. M. Institute of Engineering and Technology
Haryana, India
siddharthdogra03@gmail.com

Ayush Mittal
Department of Computer Science and Engineering
B. M. Institute of Engineering and Technology
Haryana, India
mittalayush989@gmail.com

Shivam Chaudhary
Department of Computer Science and Engineering
B. M. Institute of Engineering and Technology
Haryana, India
Chaudharys1704@gmail.com

Sonika Vasesi
Assistant Professor
Department of Computer Science and Engineering
B. M. Institute of Engineering and Technology
Haryana, India
sonikavasesi@gmail.com

Vishal Jain
Dean Academics
B. M. Institute of Engineering and Technology
Haryana, India
vishal_dehradun@yahoo.com

***Abstract*— The recruitment is a crucial factor in the development and achievement of an organization. Nevertheless, the conventional resume screening approaches are ineffective, time consuming and subject to biasness. Being AI-based, the Smart Resume Screening and Matching System described in this paper automates the processes of resumes parsing, matching with job-candidates, and ranking through the application of Natural Language Processing (NLP) and machine learning algorithms. The system picks out the structured information of unstructured resumes, matches them with job descriptions and generates compatibility scores. It tries to make the recruitment more efficient, less human, and fairer when assessing the candidates. The system makes use of Python-based structures, and NLP extractors spaCy and NLTK to extract and compute similarities. Findings show a high degree of screening accuracy and speed compared to manual methods and provides scalable and objective recruitment solution.**

This model is tested on 300 resumes of 10 IT job roles with an average screening accuracy of 90.3% and a shortened processing time by 65% and reduction in bias by 35 percent with anonymization. The operationally described notion of fairness is applied in cases of consistent ranking of anonymized profiles against non-anonymized ones. The latency of a typical processing was median = 0.12 s per resume (95th percentile = 0.35 s), which allowed shortlisting at very high throughput in batch mode.

Keywords*— *Recruitment Automation, NLP, Resume Parsing, AI Fairness, Machine Learning, Candidate Ranking.

I. INTRODUCTION

Recruiting talented employees has become a significant problem to contemporary organizations. The old-fashioned recruitment methods are manually screening the resumes that take time and cause inconsistency and bias [1]. Recruiters have to sift through hundreds of applications in a single job advert, which makes it more likely that they will miss out on skilled workers [13].

Recent progress in Artificial Intelligence (AI) and Natural Language Processing (NLP) allows now to automate such text analysis tasks as document classification and skill extraction [2]. The use of these methods in the employment procedures can significantly enhance efficiency, fairness, and transparency. The Smart Resume Screening and Matching System (SRSMS) is meant to overcome the shortcomings of traditional approaches by automating the parsing and ranking of resumes through machine learning.

The system is able to reduce the human factor and increase the objectivity of the hiring process by utilizing NLP to extract data on skills, experience, and education and similarity algorithms to match jobs and candidates [3][4]. The given system will solve the problem of ranking the job applicants according to the semantic similarity between the resume and the job description of the specific job. This is the solution that is best applicable in human resource departments, recruitment agencies and academic placement cells.

II. LITERATURE REVIEW

Automated recruiting systems have developed as basic keyword filters to high-level AI-based systems. Tayal et al. came up with a resume screening model based on machine learning and feature selection to enhance shortlisting of candidates [2]. Harsha et al. presented a resume screener based on NLP that can read unstructured text to extract essential skills and come up with a relevance score [1].

Although things have improved, research indicates that there are still many challenges related to bias minimization and algorithm transparency [11]. The approaches to explainable AI (XAI) that discuss the ability of recruiters to interpret model predictions and, therefore, become more likely to trust automated systems are described by Mikolajczyk et al. [9]. Similarly, Frissen et al. described the necessity to identify hiring advertisements that present discrimination as the use of machine learning methods [6].

The ethical concerns in the use of algorithms in the hiring process are highlighted by other studies. Zaker Ul Oman et al. demonstrated that AI may minimize as well as increase the bias relative to the quality of the training information [7][11]. Roslin et al. also examined fairness metrics of AI-based hiring, including continuous model auditing and accountability [8].

All these studies indicate the increasing need to have fair, explainable, and efficient systems of hiring. The Smart Resume Screening and Matching System combines NLP and scoring models with bias reduction to build an unbiased and data-driven recruiting process.

TABLE 1: COMPARATIVE SUMMARY OF RELATED WORK

Study	Approach	Dataset / Scope	Limitation	Contribution over this work
T. M. Harsha et al. (2022) [1]	NLP-based resume screener using entity extraction and similarity scoring	100 resumes	No fairness or bias-reduction metric	Our model integrates anonymization and fairness evaluation
S. Tayal et al. (2024) [2]	Machine-learning-based keyword filtering and feature selection	200 resumes	Limited explainability and semantic context	Adds XAI-based scoring transparency and semantic embeddings
R. K. Magham (2024) [3]	Bias mitigation and Explainable AI (XAI) framework for recruitment	Conceptual / theoretical	No quantitative validation or dataset testing	Provides measurable fairness results and bias-reduction metrics
R. Frissen et al. (2023) [6]	ML model to recognize bias in job advertisements	Text-only corpus	Not an end-to-end candidate ranking system	Integrates full pipeline from parsing to ranking with fairness reporting

III. OBJECTIVES

The Smart Resume Screening and Matching System (SRSMS) was proposed to enhance efficiency and equality in the selection of the candidates. The main objectives are:

1. **Automated Resume Parsing**: Build a powerful resume parsing algorithm that can handle unstructured resume data and identify important information (skills, qualifications, experience, and contact details) on such documents (PDF, DOCX) [13].
2. **Job Description Matching**: This will require running NLP algorithms to process job descriptions and semantically match them with resume data.
3. **Compatibility Scoring**: Implement an objective quantifiable method of scoring that gives HR professionals objective rankings of applicants.
4. **Bias Reduction**: Guarantee unbiased recruitment, anonymize identifiable candidate information (name, gender, photo) and assess them.
5. **Scalability and Usability**: Develop a scalable architecture with a modular design in order to handle bulk resume uploads, which would qualify it to work with enterprise, campus placement and online hiring portal.
6. **Explainability and Transparency**: Include explainable AI capabilities to allow the HR staff to comprehend the score generation mechanism and what resume characteristics impacted rankings.
7. **Quality:** Achieve more than 85% ranking correlation with expert HR judgments.
8. **Speed:** Process at least 10 resumes per second in batch mode.
9. **Fairness:** Maintain at most 10% variation between anonymized and non-anonymized ranking results.

IV. EXISTING SYSTEM

Traditional Applicant Tracking Systems (ATS) are also mostly keyword based and rely on fixed rule sets to sieve resumes. Although they computerize the rudimentary sorting, they do not comprehend the contextual meaning of job descriptions and resumes [5].

An example is that an ATS can miss a candidate who excels in data analysis, but the job description consists of the term analytical modelling and would not get included because of the absence of an overlap between the keywords. More so, ATS systems do not detect bias or offer any form of fairness and in most cases, they reproduce historical discrimination inherent in training data.

Weakness of the current systems include:

- **Contextual Misinterpretation**: The absence of semantic comprehension between job descriptions and resumes.
- **Manual Intervention**: Recruiters will be required to manually screen shortlists because of false positives and negatives.
- **Bias and Subjectivity**: The systems are not anonymized to control human bias: the identities of the applicants are not blinded.
- **Small Analytics**: Current platforms are unlikely to provide information on the trends in hiring or skills gaps.
- **Scalability Problems**: The traditional systems are not able to effectively deal with multi-purpose or large-volume recruitments.

These shortcomings point to the necessity of an AI-based system that combines semantic analysis, automation, explainability, and fairness which are core components of the proposed Smart Resume Screening and Matching System.

V. PROPOSED SYSTEM

The Smart Resume Screening and Matching System (SRSMS) will present a multi-layered system that will combine NLP, machine learning, and explainable AI to improve the accuracy and fairness of recruitment.

A. System Architecture

The architecture is based on a three-tier design:

- **Presentation Layer**: A web interface built with HTML, CSS and JavaScript which will enable the HR manager to post resumes, job descriptions and see ranked results.
- **Application Layer**: This layer is developed with Python Flask or Django and it handles NLP processing, scoring code, and report-generation.
- **Data Layer**: A relational database (MySQL/PostgreSQL) with information on user credentials, job descriptions, resumes which are to be parsed, and match scores which have been computed.

The modular architecture provides ease of scale, security and maintenance.

B. Functional Workflow

- **Resume Upload and Parsing**: HR uploads the resume which are then processed by NLP-based Named Entity Recognition (NER) to pick features like skills, experience, education and certifications.
- **Job Description Analysis**: Job descriptions are tokenized and vectorized with the help of TF-IDF or Word2Vec embeddings.
- **Similarity Computation**: Cosine similarity is used to determine the extent of match between candidate resumes and job profile.

- **Score Calculation and Ranking**: A weighted scoring model is a scoring model that uses a combination of changes in skills, experience, education and so on to generate the overall compatibility score.

- **Bias Reporting and Filtering**: The personal identifiable information is obfuscated prior to the presentation of the results to enable reasonable evaluation. Reporting is then done with results in ranked candidates and their match percentages.

C. Technology Used

- **Programming Language**: Python 3.11.
- **Frameworks**: Flask/Django
- **NLP Libraries**: spaCy, NLTK
- **Database**: MySQL / PostgreSQL
- **Front-end**: JavaScript, CSS3, HTML5.
- **Tools**: Scikit-learn to compute similarity; Matplotlib to generate analytics dashboard.

D. Key Innovations

- Addition of semantics similarity to establish deeper contextual meaning [12].
- In- architecture ethical AI mitigation layer.
- Elaborate dashboards which realize candidate matching logic.
- Handling of hundreds of resumes at once in batch mode.

Anonymity of candidate profiles is employed in the evaluation in accordance with structures suggested by Magham [3] and Frissen et al. [6] so that gender or demographic bias is reduced.

VI. METHODOLOGY

The project has embraced the Agile Software Development Life Cycle (SDLC) which enables the design and testing to occur through the process.

A. Phases of Development

- **Requirement Analysis**: Consisted of a close work with HR professionals to get pain points and feature requirements.

- **Design**: Data flow Diagrams (DFDs) were created to map the data flow and system logic with the help of entity Relationship (ER) and UML.

- **Application**: Python spaCy and NLTK libraries were used to create NLP tools to analyze the resume and job description.

- **Testing**: Unit, integration and usability testing had been carried out to ascertain that data extraction accuracy and similarity computation were correct.

- **Deployment**: It is hosted on a cloud service to enable access by multiple users and scalability.

- **Maintenance**: Added performance optimization and retraining of the models.

It also used explainable AI frameworks to explain their decisions in the ranking process. [4][9]

B. Algorithmic Design

- **Vectorization:** Text data transformed to numerical vectors with TF-IDF (embedding) and Word2Vec (embedding).
- **Cosine Similarity:** Similarity = (A.B) / (||A|| * ||B||) used for job-resume matching.
- **Weight Optimization:** Aggregated compatibility score weights were chosen through grid search through the plausible weights (skill, experience, education, certification) in order to maximize Spearman correlation with HR gold rankings on a validation fold. The weights that were chosen were Skills = 0.40, Experience = 0.30, Education = 0.20, Certifications = 0.10.
- **Entity Recognition:** spaCy pre-trained models used to recognize and categorize entities (skills, organization, education).
- **Bias Reduction:** Anonymization algorithms automatically eliminate identifiable personal information before the scoring.

C. Scoring Pipeline

- **Representation:** Sentence embeddings (spaCy transformer model) are used as a representation of both resumes and job descriptions.
- **Skill Normalization:** Extraction skills are normalized against a standardized taxonomy.
- **Weight Assignment:** Skills = 0.4, Experience = 0.3, Education = 0.2, Certifications = 0.1 - also tuned empirically, to give a perfect HR correlation.
- **Explainability:** Visualization of the top-weighted features used to obtain the final score is presented in a dashboard.

The complete scoring pipeline consists of the following steps:

- **Preprocessing of input:** Extracts text from PDF and DOCX files while normalizing case and punctuation for consistent NLP processing.
- **Skill extraction:** Utilizes custom regular expressions and spaCy models to identify and extract skills, education, and experience entities from resumes.
- **Vector representation:** Compute resume sections and job description using spaCy transformer model sentence embeddings.
- **Component scoring:** Find similarity scores of Skills, Experience, Education and Certifications (cosine similarity between corresponding vectors or normalized numeric difference in experience).
- **Aggregation:** Add individual components (Weights 0.40, Experience 0.30, Education 0.20, Certifications 0.10) to acquire the overall compatibility score.
- **Output:** List of the ranked candidates and per-feature contribution breakdown and anonymization log.

D. Evaluation Metrics

- **Accuracy**: Percentage of the correct top candidates identified.
- **Precision and Recall:** To assess the false positives and negatives.
- **Processing Time:** Mean time which can be used to process a batch of resumes.
- **Fairness:** Our operationalization of fairness is the shift in the average rank in an anonymized and non-anonymized assessment. Reported Fairness Score is the average absolute rank difference (as a percentage of list length). The low percentage means that the demographic impact on ranking is smaller [7][8].

The comparative analysis was made on two baseline models:

(a) **Keyword-Matching Model:** A typical Applicant Tracking System (ATS);

(b) **Sentence-Embedding Average Model:** Uses Cosine similarity between document vectors.

The evaluation experiments were repeated with 300 resumes and 10 IT job roles, which is consistent and statistically reliable to guarantee the performance outcomes.

VII. RESULTS AND DISCUSSIONS

A total of 300 resumes and 10 job descriptions in the IT industry were used to test the Smart Resume Screening and Matching System.

A. Quantitative Results

1. **Accuracy of the Screening:** 90.3% (when compared to the expert HR assessment)
2. **Time Savings:** 65% savings in time on shortlisting time in relation to manual screening.
3. **Bias Minimization:** There is reduction of bias of 35 percent with anonymized profiles.
4. **Ranking Correlation:** Candidate rankings had a correlation of 88% with the HR judgments and this goes to confirm the reliability of the system.

TABLE 2: QUANTITATIVE SUMMARY

Job Role	Accuracy (%)	Ranking Corr. (%)	Avg. Time/Resume (s)
Data Analyst	91.2	89	0.11
Web Developer	90.5	88	0.13
AI Engineer	89.7	87	0.14

B. Qualitative Analysis

According to the HR professionals, the explainable visual dashboards of the most influential resume features in the rankings led to improved decision transparency. The multi-factor scoring of the system guaranteed the balanced weighting of the skills, experience, and education, which over-dependent on one, such as the degree titles.

C. Results by Seniority

To evaluate consistency across different professional levels, the dataset was divided into three seniority categories based on years of experience: Junior (0–2 years), Mid-level (3–6 years), and Senior (7 years and above). The system’s ranking accuracy and correlation with HR evaluations were computed for each group to assess robustness across experience ranges.

TABLE 3: RESULTS BY SENIORITY

Seniority	Accuracy (%)	Ranking Corr. (%)
Junior (0-2 yrs)	92	90
Mid (3-6 yrs)	89.8	87
Senior (7+ yrs)	88.2	85

D. Comparative Insights

The system compared to the traditional ATS tools showed:

- Increased transparency and accountability.
- Better fit of job requirements and shortlisted candidates.
- Ability to manage 10x bonus resumes per hour without compromised performance.

We have modelled the concurrent batch processing of 500 resumes using the same test hardware. The median latency also improved slightly to 0.18 s/resume, with 95th percentile less than 0.40 s/resume. The 500 resumes throughput was still high (above 8 resumes/sec), which demonstrates that the system can support large-scale batch loads.

These results confirm that the application of NLP and explainable AI in the context of recruiting contributes to the improvement of accuracy and fairness considerably [3][6][8].

VIII. FUTURE SCOPE

Even though the existing version does a good job at automation of screening, it can be improved with several additions:

- **Explainable AI Expansion:** Adoption of counterfactual explanation solutions [9] to present what-if scenarios that indicate how small resume changes can alter the ranking of the candidates.
- **Blockchain Addition:** Saving verified education and working records to eliminate fake data.
- **Multilingual Processing:** The inclusion of regional languages based on the multilingual NLP models to increase the accessibility [10].
- **Predictive Analytics:** Using machine learning models to forecast post-employment success, retention, and cultural fit.
- **Integration with LLMs:** Contextual interview question generation and intelligent chat-based interaction with candidates with Large Language Models (LLMs) [10].
- **Ethical AI Framework:** Establishing a system to keep identification of bias and adherence to data privacy legislation such as GDPR and India DPDP Act.

These advances would enable the system to become a fully developed AI-based recruitment system that can take control of the entire hiring pipeline - resume to final selection.

IX. CONCLUSION

The Smart Resume Screening and Matching System (SRSMS) is a revolutionary advancement in contemporary data-driven recruitment through a combination of artificial intelligence and natural language processing into the process of parsing and matching of resumes and candidate-job matches and rankings [1][2]. The system is more accurate, more efficient, and fairer because semantic similarity algorithms, explainable AI and bias mitigation mechanisms are applied to permit transparent and unbiased assessments [3][4][6]. Its scalability, cloud compatibility and modular architecture can be relied upon to perform reliably in diverse hiring settings. In addition to enhancing the accuracy of shortlisting and lowering screening time, SRSMS encourages responsible and ethical recruitment through the reduction of human bias and the focus of accountability in AI decision-making [7][8]. The system had a ranking accuracy of 90.3% and a 65 per cent faster screening and a 35 per cent reduction in bias which made an explainable fair and quantifiable AI model of candidate shortlisting. The system can form the basis of future advancements in intelligent hiring by integrating the future potential of advanced XAI frameworks, predictive analytics, and credential verification to blockchain-based systems, showing that AI can enhance rather than eliminate human judgment and promote fairness, inclusivity, and transparency in the hiring process worldwide [9][10].

REFERENCES

[1] T. M. Harsha et al., “Automated Resume Screener using Natural Language Processing (NLP),” Proc. 6th Int. Conf. on Trends in Electronics and Informatics (ICOEI), 2022, pp. 1772–1777.

[2] S. Tayal, T. Sharma, S. Singhal, and A. Thakur, “Resume Screening using Machine Learning,” Int. J. of Scientific Research in Computer Science, Engineering and Information Technology, vol. 10, no. 2, 2024.

[3] R. K. Magham, “Mitigating Bias in AI-Driven Recruitment: The Role of Explainable Machine Learning (XAI),” Int. J. of Scientific Research in Computer Science, Engineering and Information Technology, 2024.

[4] R. Avacharmal, “Explainable AI: Bridging the Gap between Machine Learning Models and Human Understanding,” Journal of Informatics Education and Research, vol. 4, no. 2, 2024.

[5] “A Systematic Review of Explainable Artificial Intelligence in Terms of Different Application Domains and Tasks,” Applied Sciences, vol. 12, no. 3, 2022.

[6] R. Frissen, K. J. Adebayo, and R. Nanda, “A Machine Learning Approach to Recognize Bias and Discrimination in Job Advertisements,” AI & Society, vol. 38, pp. 1025–1038, 2023.

[7] Z. U. Oman, A. Siddiqua, and R. Noorain, “Artificial Intelligence and Its Ability to Reduce Recruitment Bias,” World Journal of Advanced Research and Reviews, vol. 24, no. 1, pp. 551–564, 2024.

[8] J. Roslin and B. R. Mishra, “AI-powered Recruitment and Employee Selection: Evaluating Bias and Fairness in Hiring Practices,” European Economic Letters, vol. 14, no. 1, pp. 827–834, 2024.

[9] A. Mikołajczyk et al., “Towards Explainable Classifiers Using the Counterfactual Approach – Global Explanations for Discovering Bias in Data,” Journal of Artificial Intelligence and Soft Computing Research, vol. 11, no. 1, pp. 51–67, 2021.

[10] C. Gan, Q. Zhang, and T. Mori, “Application of LLM Agents in Recruitment: A Novel Framework for Resume Screening,” arXiv preprint arXiv:2401.08315, 2024.

[11] Pisanelli, E., et al. "Automated resume screening as a strategy to reduce gender gap in employment opportunities," ScienceDirect, 2022.

[12] Boubacar, Z.S., et al. "Automating Recruitment Process Using NLP and Deep Learning," IEEE Xplore, 2024.

[13] Abbasi, F.K., Ali, A., Bibi, N. "Analysis of skill gap for business graduates: managerial perspective from banking industry," Education and Training, 2018.