International Journal of Scientific Research and Engineering Development-– Volume 8 Issue 5, Sep-Oct 2025

Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

Enhancing CV limitations using NLP

T. Amalraj Victoire[1], M. Vasuki[2], M. Aruljothi[3*]

[1](*Associate Professor, Department of Master of Computer Applications, Sri Manakula Vinayagar Engineering College, Pondicherry, India.*

Email:amalrajvictoire@gmail.com)

[2](*Associate Professor, Department of Master of Computer Applications, Sri Manakula Vinayagar Engineering College, Pondicherry, India.*

Email:dheshna@gmail.com)

[3](*PG Student, Department of Master of Computer Applications, Sri Manakula Vinayagar Engineering College, Pondicherry, India.*

Email:aruljothi1603@gmail.com)

----------**********----------

Abstract:

Conventional CV formats often do not adequately represent candidates and their abilities because of their structured formats, inconsistency of design, and issues of compatibility with automated hiring approaches & screens. As ATS (applicant tracking systems) gain popularity, many qualified resumes are filtered out by these systems before they ever reach the review stage of hiring. This work presents a NLP (natural language processing) based framework to address these barriers by providing semantic analysis, skill identification, enhancing keywords, and customizing resumes automatically based on a job description. The framework integrates advanced techniques, such as contextual embeddings, contrastive learning, and skill graph construction, to enhance the accuracy of matching candidates to job openings while providing an equitable and transparent hiring process. Experimental evaluations of the framework demonstrate that using the automated resume screening process improves effectiveness of screening, decreases bias, and allows candidates to obtain feedback to help them align. resume with job openings. This work supports enhanced recruitment processes that promote equitable AI hiring processes.

Keywords — Natural Language Processing, Resume Screening, Semantic Analysis, Skill Extraction, Fairness in Hiring.

----------**********----------

I. INTRODUCTION

The scope of recruitment is changing, due to new computerized technologies and automated systems to help organizations organize job applicants. Automated systems typically apply keyword-based filtering methods, which pose a problem because, in many instances, they can eliminate qualified candidates from the pool because their resumes do not contain a word-for-word matching or have the formatting that adheres to best practice. This creates not only efficiency but also exclusionary hiring practices that promote bias and reduce diversity in the workplace. In this scenario, Natural Language Processing (NLP) appears to be technology that adopts a sophisticated understanding application of resumes and job description. New advances from transformer-based models, like BERT and Sentence-BERT, and graph-based skill representations can understand the semantic level and contextual nuances of individual profiles beyond keyword matching. This results in greater resume screening accuracy, stating implicit abilities, and better alignment of professional experience with the job description.

Automated NLP-based enhancement systems can also support job seekers identify missing keywords, areas of expertise shortage, and opportunities for tailoring their resumes more effectively toward the targeted jobs selection process.Beyond these efficiency benefits, these intelligent systems use explainability-in-AI techniques such as SHAP and LIME for greater fairness and transparency in hiring decisions to disclose the rationale behind candidate assessors' evaluations. The proposed framework in this paper is a comprehensive, ethical, and scalable response to resume enhancement and candidate-job matching based on recent advances in NLP with fair-representation algorithms, and skill graph inference. This is geared towards reducing biases, increasing recruiter trust, and supporting candidates and employers in improving hiring practices.

II. LITERATURE REVIEW

People talk a lot these days about how Artificial Intelligence and Natural Language Processing are reshaping the way hiring works. Over the last few years, research from 2019 to 2022 has really transformed how resumes are understood and how candidates are matched with the right jobs. Instead of just keyword hunting, newer systems now look deeper into meaning, fairness, and even explainability.

Back in 2021, Chakraborty and his team pulled together a big survey that looked at how NLP fits into recruiting — things like resume parsing and skill extraction.They observed that most systems continue to depend on keyword matching and lack personalization and deeper semantic understanding. It's like they can see the words but not the real story that conveys them.

Then came Reimers and Gurevych in 2019 with something that changed the game -Sentence-BERT. It's a smart setup using a Siamese network on top of BERT that helps computers understand how similar two sentences really are. That kind of semantic matching made a big difference for resume and job description alignment.

Of course, none of that would've been possible without Devlin and his team's 2019 work on BERT, the foundation of it all. Their model reads text in both directions, giving it context and meaning. Since then, it's been the baseline for just about every advanced NLP system — even the ones fine-tuned for resume screening today.

As 2022 approached, Zhang et al. conducted a study of a heightened level of explainability in HR systems. They claimed that any hiring decision mediated by AI must be explainable so we can place trust in these systems. The more explainable an algorithm is in its decision-making process, the greater the fairness in hiring. Continuing down the same pathway of inquiry, Kenthapadi et al. (2020) reviewed algorithms that measure fairness, endorsing unbiased statistical practices and fair screening practices in AI hiring processes and systems.

While fairness may warrant concern in the implementation of AI in hiring, it is not the only consideration. Suresh and Guttag (2019) developed a framework depicting how bias can enter machine learning both through data collection, model design, and use of the system. After reviewing their analysis, they encouraged both researchers and practitioners to thoughtfully re-consider the implementation of AI systems in hiring, to avoid continuing systematic inequality.

In 2020, Papageorgiou and his team turned their attention to skill extraction. They used Named Entity Recognition and dependency parsing to make sense of how resumes are written. It works great for structured data, but they noted that it struggles with less formal or inconsistent information — the kind you often see in real resumes.

Then Malik et al. (2021) introduced a hybrid model using TF-IDF and BERT embeddings. By combining traditional keyword techniques with deep learning, they improved accuracy and produced more trustworthy job-document matching.

Gupta and Singh (2022) also created a resume enhancement system that offers candidates additional keywords they missed when compared to job descriptions. This was a clever idea, but they state it still lacks semantic richness and explainability.

III. NLP Fundamentals

NLP, or Natural Language Processing, is a component of artificial intelligence that allows computers to read, comprehend, and write human langage. You can use NLP for resumes and job processes to extract the most important information out of the text, such as resumes and job descriptions. The first step in the process is to clean up the information you are analyzing, to remove any common words, and to reduce the original words to their root words, if possible, to ensure it is prepared for analysis. Text is converted through key word evaluation and importance towards words or phrases within a document.

More recently, contextual embeddings from models have taken their place by taking the words into context by critically influencing the meaning of a specific word. This has been useful in resumes as it has accounted for minor differences in phrasing and industry terms. The core process of improving your resume with NLP includes noting structural components in qualifications of skills and organizations. Measuring relationship alignment on the profile against job requirements. Text is also categorized by aligning resumes against job roles or industry domains. Following 2nd order relationships it maps grammatical structure and relationships in resumes by including information add to the resume. Computer learning now enhances matching with mapping programs with the skill to spot missing skillsets. Training models alongside skill frameworks improve matching and pinpoint modifying contexts.

This makes the job matching system more correct. Ethical reasons in recruitment is important for applications. This shows algorithms in work and process of fair employment. For example to make everything clear, these AI systems show how a resume was matched and used ethically. In closing NLP contributes to building great resume technology by interpretable techniques that limit applicant biases. Ultimately this supports equal opportunity in employment.

IV. Methodology

This section describes the design, implementation, and testing of the proposed NLP-powered framework to improve CVs and semantic alignment with job descriptions. System Architecture The framework consists of 6 modules that work together as an integrated system:

A. *Preprocessing Module:*

This cleans and tokenizes resumes and job descriptions, using traditional NLP preprocessing approaches including tokenization, stopword removal and lemmatization, which we describe later in this section.

B. *Embedding Layer:*

This generates contextual embeddings with Sentence-BERT, which was -fine-tuned with resume--job pairs to capture semantic relationships.

C. *Semantic Matcher:*

This implements contrastive learning with triplet loss to maximize similarity scoring, focused on the candidates' profiles with respect to the job descriptions.

D. *Skill Graph Generator:*

This creates a graph-based representation of skills that infers implicit competencies using semantic expansion, and creates embeddings using Node2Vec.

E. *Explainability Engine:*

This leverages the SHAP and LIME techniques to give the recruiter insight into how the model arrived at its decision.

F. *Fairness Module:*

The module builds upon fairness-aware training with demographic parity constraints, adversarial debiasing, and anonymized corpus to avoid bias and keep candidate screening equitable.

G. *Datasets:*

Three datasets were utilized: Resume Corpus: 5,000 anonymized resumes collected from open-access sources on the web, including Kaggle.

H. *Occupations:*

2000 postings from LinkedIn and Indeed that are ensured, for diversity of coincidental job postings and to consider domains

I. *Skills taxonomy:*

Combines ESCO, O*NET, and from manuals in the field of study it was described that logically informed recognition from skills.

All data was obtained in an ethical manner, all personal identifiers were removed, and the ethical collection of all data complied with appropriate provincial and federal privacy and ethical regulatory requirements for the protection of data. Embedding and Matching Workflow Contextual embeddings were created using Sentence-BERT, which created 768-dimensional vectors deep embeddings of resumes and job descriptions. The model was trained using contrastive learning with triplet loss, which used an anchor (resume), a positive (the job that matched), and a negative (the job that did not match).Cosine similarity is used to rank the job relevance, and to create top recommendation.

J. *Skill Graph Creation:*

Explicit skills are recognized and extracted using Named Entity Recognition (NER) and dependency parsing by spaCy. Implicit skills are defined utilizing semantic co-occurrence and graph expansion. Skill node embeddings are generated using Node2Vec resulting in improved candidate profiling and adaptable matching.

K. *Explainability and Fairness:*

SHAP and LIME are utilized at the final scoring layer to provide recruiters with interpretable and actionable insights into impactful resume features. Fairness-aware training is used to mitigate bias using demographic parity constraints and adversarial debiasing. In addition, the framework is multilingual and parses resumes in XLM-R embeddings optimized for English,Hindi,and Tamil.

L. *Metrics for Evaluation:*

The evaluation of the framework is evaluated by the following metrics:

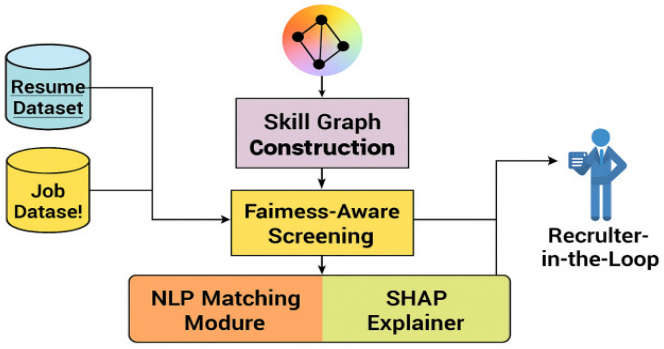
1) ***Matching Accuracy:*** Precision@K and Recall@K of top-K job matches.

2) ***Quality of Skill Inference:*** F1-score from human-annotated skill sets.

3) ***Fairness Metrics:*** Disparate Impact Ratio and Equal Opportunity Difference.

4) Explainability: Panel of 10 recruiters perform human evaluation of SHAP outputs.

Fig 1: The architecture of the NLP-powered resume improvement framework.



V. RESULTS AND EVALUATION

In this section, we provide empirical evaluation of our NLP-aided framework across three datasets: anonymized resumes, a selection of job postings, and blended skill taxonomy.The evaluation of the system was performed in four areas: match accuracy, quality of inferring skills, fairness, and explainability.

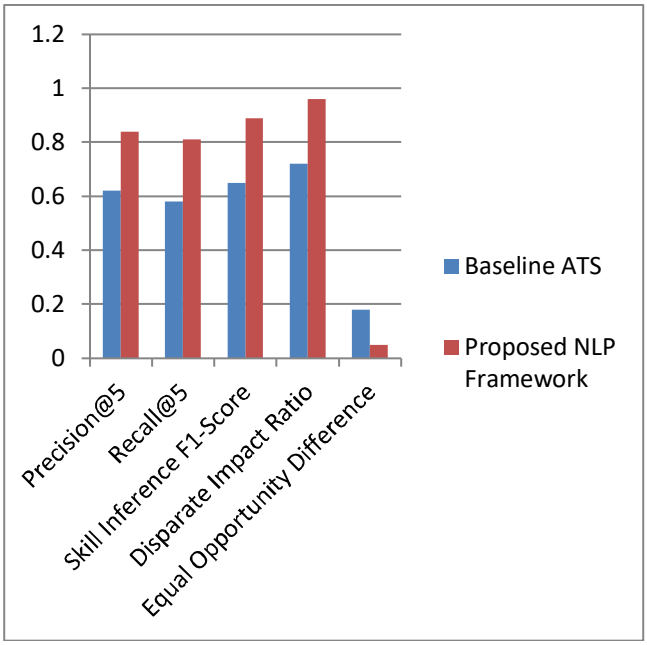
Table 1: Quantitative Metrics

Evaluation Metric	Baseline ATS	Proposed NLP Framework
Precision@5	0.62	0.84
Recall@5	0.58	0.81
Skill Inference F1-Score	0.65	0.89
Disparate Impact Ratio	0.72	0.96
Equal Opportunity Difference	0.18	0.05

A. *Evaluation of Explainability:*

A cohort of 10 different recruiters reviewed systems predictions using SHAP and LIME values. It was indicated that 87% of the jobs were considered "clear and justified" matches with resumes providing insights about the overall model's interpretability and trust in its recommendations.

Fig 2 :Performance Comparison of Baseline ATS and Projected NLP-Based Resume Screening Framework



VI. Applications of NLP in Resume Screening

This section discusses the application of Natural Language Processing (NLP) techniques for resume enhancement and candidate-job alignmen purposes. It discusses applications that include semanticsimilarity, skill matching, resume enhancement, explainable matching, and bias leverage. NLP the field is developing rapidly and could be a key driver of change in the recruitment and talent acquisition process. One of the most powerful applications NLP can provide for the recruitment process is with resume enhancement and job match accuracy.

NLP techniques can lead to the identification of missing keywords, inference of implicit skills, and matching of candidate profile to job descriptions. Previously executed only by recruiters, NLP techniques are now conducting parts of the recruitment process with scalable precision and performance.

A. *Semantic Similarity:*

Natural language processing (NLP) models, say BERT and Sentence-BERT, can use an approach called "semantic similarity" to evaluate resumes in relation to job descriptions. As an example, these models can determine the preferred candidates even when there are not exact keyword overlaps. A candidate who states "led agile sprints" could be matched to a job description that requires "project management experience", in this case there would be a semantic overlap but not an exact keyword overlap.

B. *Skill Extraction and Inference:*

Named Entity Recognition (NER) and dependency parsing are used to extract the structured information including skills, education, and experience. Advanced systems will also infer skills based on context. For example, if an applicant has skills related to programming in Python and is skilled in TensorFlow, the system can infer that this candidate has experience in machine learning.

C. *Resume Augmentation:*

QR systems using basic natural language processing (NLP) systems can find skills that are lacking or over-represented, while also suggesting to a candidate areas to strengthen or improve. This is particularly valuable because of potential increased matching when users are able to intently amend their resumes to meet the needs of the open position.

D. *Explainability matching:*

An example of systems designed to describe how and why a resume was matched to a job is SHAP and LIME. This can help bring transparency to a match and helps explain why the automated process decided on a particular match.

E. *Multilingual process:*

Rapidly developing NLP systems and programs that are trained on multilingual datasets will have the ability to process resumes submitted in multiple languages, which will inherently promote inclusion in hiring processes and allow organizations to access a global job market.

F. *Detecting and reducing bias:*

Natural language processing (NLP) systems can also be trained to detect biased language in job position descriptions and candidate resumes, and to create anonymized resumes that may help mitigate conscious bias in the resume screening process in the hiring process.These applications provide a meaningful understanding of how the use of NLP can increase accuracy, efficiency, and equity into hiring processes. As NLP technology continues to develop, its role in hiring will continue to extend.

G.*Benefits of NLP in Resume Enrichment and Job Matching:*

This section describes the benefits of using NLP for resume enrichment and job matching. This benefits include increased efficiency, better accuracy, cost efficiency, and the ability to perform complicated tasks that have typically been performed by humans.

NLP offers several advantages that facilitate automated resume screening and improved job matching:

1) *Increased Efficiency:* NLP systems can process thousands of resumes in seconds, resulting in shortened time-to-hire.

2) *Reduced Costs:* NLP can also lower hiring-related costs by decreasing manual processes so that staff can spend more time on strategic priorities.

3) *Increased Quality:* NLP will select the most qualified candidates, improving quality of matches, regardless of whether resumes contain appropriate words.

4) *Scalability:* NLP can provide effective efficiencies in hiring across various departments and locations.

5) *Equity:* Natural Language Processing (NLP) can help organizations diminish, if not outright remove, hiring bias, which ultimately allows for greater equity and reduced discrimination in a hiring process.

6) *Transparency:* AI classification has allowed for greater transparency around how decisions are made, which facilitates trust in machines.

7) *Candidate Empowerment:* Providing candidates with programs that assist them in enhancing their resumes, increases their chances of being identified.

Table 2: Comparison of Traditional Resume Screening vs. NLP-Based Resume Screening

Challenge	Traditional Screening	NLP-Based Screening
Keyword Dependency	Allows only for exact keyword match screening.Keywords don't allow a relevant candidate to screen in if it is just not a keyword (phrase).	Uses semantic similarity to match resumes (even if no exact match to a keyword).
Skill Identification	Identifies/extracts only the skills which have been overtly stated, with no capability to infer hidden skills.	Extracts and infers both overt and covert skills by using NER (named entity recognition) and skill graphs.
Scalability	Manual screening for a reasonably large applicant pool moves slowly and is therefore not scalable.	Fully automated screening of thousands of resumes can be provided and remain consistent with the accuracies.
Bias and Fairness	Screeners can exhibit unconscious bias based on their decision-making.	Reduces bias through the use of fairness-aware algorithms and anonymization.
Transparency	There is little to no record of decisions made when screening candidates which makes the process difficult to validate.	We can run decisions through Explainable AI (e.g., SHAP, LIME) for benefits of transparency.
Candidate Feedback	No real feedback for resumes processed (everything from no contact to boilerplate email).	Insight into the resumes with suggestions to improve the resume to suit the job specifications.

VII. Conclusion

This study concludes by discussing a novel NLP-powered framework that is capable of improving important weaknesses with respect to traditional resume screening. The framework is able to leverage semantic similarity, skill graph induction, and a fairness-aware approach to evaluation. Furthermore, the framework utilizes contextual embeddings, a contrastive learning framework, and explainable AI methods to result in a more accurate, equitable, candidate-job matching process.

Furthermore, unlike earlier work, our approach is not just concerned with ranking recommendations, but it also allowed for both recruiters and applicants to have transparency using a multilingual text representation approach, and given feedback on improvement recommendations for resume enhancements. Our experiences indicate a significant positive impact with new precision recall and fairness based metrics being utilized. Future work will include expanding our language coverage, incorporating adaptive recruiter learning, and the deployment of iterative feedback loops in hiring to quantify a given designated impact in the hiring process.

VIII. LIMITATIONS AND FUTURE WORK

While there is sufficient evidence of enhancements in semantic matching, skill inference, and fairness-aware screening in the proposed NLP-based framework, several aspects still require examination.

A. *Language Limitations:*

The model proposed is trained and optimized in English, Hindi, and Tamil. While multilingual embeddings (e.g., XLM-R) were used, the hiring platform is unlikely to generalize effectively to other languages or dialects or to formats of resumes that differ culturally. This is a limitation for its use in the global hiring contexts where such attributes differ in language and structure.

B. Real-Time Interaction:

As of now, this framework lacks real-time resume enhancement and real-time access to recruiter's feedback. These features are essential while using the tool in real-world, fast-moving recruitment situations, as iteration and interaction between recruiter and candidate population can significantly increase match quality.

C. *Dataset Bias:*

Although the datasets used were ethically sourced and anonymized, those datasets may carry historical bias captured in job postings and convention in professional resumes which can impact model behaviors and fairness outcomes - the models might take some industries or jobs to be over-represented and therefore lead towards bias in skill inference or matching behavior. Future work will seek to curate more representative datasets (geography, industries, and demographic) and employ bias auditing methods to ensure equitable performance in bias models.

D. *Explainability Depth:*

While SHAP and LIME offer transparency vis-a-vis the decision making of the model, their interpretability may be varied among recruiter profiles; some users might prefer more intuitive visualizing of machine-driven recommendations or some more specific to the domain or industry to feel comfortable trusting machine-leveraged recommendations.

E. *Scalability and Deployment:*

The framework has yet to be tested in real enterprise settings, so its scalability, integration with existing ATS software, or enforced performance and constraints, (e.g., in cases of incomplete resumes, noisy data, etc.) are still challenges the framework will face.

IX. FUTURE DIRECTIONS

In future efforts to address limitations, we will focus on:

A. *Multilingual Expansion:*

Expanding to additional regional and global languages to support hiring across geographic areas.

B. *Adaptive Learning:*

Adding jobs about recruiter interaction data to update matching algorithms to improve algorithms to reflect shifted recruiter preferences and hiring trends.

C. *Live Deployment:*

Trialing the frameworks in live recruitment platforms to assess its usability, scalability, and impact on hiring outcomes.

D. *User-Centric Explainability:*

Creating recruiter-facing dashboards that enable transparency levels to be adjusted by the users and visual summaries of the model rationale. Through attention to attention these areas, the framework can become a robust, ethical, and scalable solution for intelligent hiring—empowering candidates and recruiters alike to engage in a data-driven and inclusive environment among stakeholders.

X. REFERENCES

[1] S. Chakraborty, A. Roy, and P. Das, “Applications of Natural Language Processing in Recruitment: A Survey on Resume Parsing and Job Recommendation,” *J. Intell. Inf. Syst.*, vol. 56, no. 4, pp. 1123–1140, 2021.

[2] N. Reimers and I. Gurevych, “Sentence-BERT: Sentence Embeddings Using Siamese BERT-Networks,” in *Proc. EMNLP 2019*, 2019. [Online]. Available: https://arxiv.org/pdf/1908.10084.pdf

[3] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” in *Proc. NAACL 2019*, 2019. [Online]. Available: https://arxiv.org/pdf/1810.04805.pdf

[4] B. Zhang, Y. Li, and Z. Chen, “Explainable Artificial Intelligence for Human Resource Systems: Enhancing Transparency and Trust in Automated Hiring,” *IEEE Trans. Artif. Intell.*, vol. 3, no. 2, pp. 145–157, 2022.

[5] K. Kenthapadi, S. Jain, and R. Sharma, “Fairness-Aware Algorithms for Automated Resume Screening,” in *Proc. Int. Conf. Fairness,*

Accountability, and Transparency (FAT), 2020. [Online]. Available: https://arxiv.org/pdf/2001.08813.pdf

[6] H. Suresh and J. Guttag, "A Framework for Understanding Unintended Consequences of Machine Learning," *Commun. ACM*, vol. 64, no. 11, pp. 62–71, 2019.

[7] T. Papageorgiou, P. Karampelas, and G. Tsoumakas, "Skill Extraction from Unstructured Resumes Using Named Entity Recognition and Dependency Parsing," *Expert Syst. Appl.*, vol. 162, 113751, 2020.

[8] R. Malik, N. Kaur, and P. Singh, "A Hybrid Job Recommendation System Using TF-IDF and BERT Embeddings," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 8, pp. 451–459, 2021.

[9] S. Gupta and A. Singh, "Resume Enhancement through Keyword Recommendation and Job Matching," *SN Comput. Sci.*, vol. 3, no. 4, pp. 327–336, 2022.

