

ResuMatch: An Interpretable NLP-Based Framework for Semantic Resume Job Description Matching in Automated Recruitment Systems

Subhranshu Pathak¹, Aman Kumar¹, Vishal Gupta¹, Vishwas Chakrawarti¹, and Dr. Mohd Sadim¹

Department of Computer Science, KCC Institute of Technology and Management, Greater Noida, India
namanpathak777@gmail.com, hii.vishalgupta@gmail.com, amanchakrawarti88@gmail.com, vs206242@gmail.com, sadim4u@gmail.com

Abstract. ResuMatch is a fair and smart recruitment aid system which uses sophisticated artificial intelligence to automatically map applicant rsums to relevant job descriptions semantically. The main purpose of the system is twofold: firstly, to make selection decisions more accurate and secondly, to significantly reduce the manual evaluation work of human resource staff. The approach uses semantic text representation models, detailed section wise information extraction, and differentially weighted algorithmic matching to measure conformity of candidate skills, work history, and educational qualifications to separate job requirements. While typical keyword based filtering only checks for the occurrence of certain lexical tokens in the text, ResuMatch performs contextual reasoning to locate relevant skills even if the terms are different or syntactically altered.

It is very important to us that our work is transparent, interpretable, and that the assessment is free from bias. The system designed in this paper can generate structured analytical outputs, such as numerical similarity scores, exact identification of significant competency gaps, categorization of strengths, etc. which will ensure fairness in recruitment evaluation. Meanwhile, candidates get constructive feedback that helps them improve their application materials and increase their chances of getting a job. The implementation workflow consists of the following sequential steps: automated document collection, text normalization by means of natural language processing (NLP), generation of vector embeddings, heuristic scoring, and supply of visual analytics dashboards customized for corporate human resources departments and academic career centers.

Keywords: AI resume matching, hiring automation, NLP, matching algorithm, skill extraction

1 Introduction

Recruitment strategies have changed so much that there has been a movement away entirely from people studying resumes to automatically doing stuff based on the data as a response to the huge numbers of applicants and the constantly changing skills [1][2]. At first, the hiring relied only on the judgment of human experts and very basic keyword searches which caused biases in the results, long cycle times and the exclusion of some qualified candidates without the intention. Later, with the rise of digital platforms, it became practically impossible to carry out the traditional screening without a lot of difficulties. Experiments show that just checking the text for matching words is not enough and that deep understanding of semantics, resolving ambiguities in the context and mapping of skills by using the knowledge of the domain are few of the main reasons why nowadays people strongly favor these approaches quite often. Therefore, AI is now a key part of the smart recruitment as it exploits human language understanding, document representations and measuring similarities among others for better matching candidate profiles to job descriptions. Besides delivering high accuracy and consistency among raters, these AI based tools also reduce the cognitive load on human judges by automatically doing the extraction of the information, identifying the criteria specific to that role and quantitatively scoring the suitability of a person[8][9]. On the other hand, a large part of the existing pieces of software do not adequately reveal their inner working mechanisms; in addition, they do not notify the users of the skills they lack and also offer interpretations only to the recruiters and not to the applicants[11][12].

ResuMatch fills in these gaps very well because it is not only an AI based method for extracting structured rsum data, but it also normalizes job requirements and measures semantic similarity among skills experience education, and the overall contextual meaning. It can even pinpoint candidate strengths, mark competency gaps, and provide actionable recommendationsthings that mere keyword filtering could never achieve [16][17]. The processing pipeline consists of text preprocessing, natural language processing, embedding based assessment, and interactive visualization dashboards for human resources, academics, and consulting stakeholders [18]. Here, the authors share a tool that not only makes recruitment more productive, but also enables fair assessment and makes it easier to understand real time job searches which are always changing very fast. By using AI profiling which is combined to matching that is right in the context, ResuMatch is changing the whole hiring process by both doing the work faster

in the screening stage and making selection outcomes more transparent and more easily justified [19][20]. The paper first of all extensively reviews the related literature, then explains research methods and describes implementation and experimental results, also finally talks about the future aims of recruitment intelligence development.

2 Literature Review

This part is a summary of research related to AI-augmented recruitment, automated rsum parsing, semantic alignment of candidates with occupational specifications, fair practices in algorithmic hiring, and feedback-based recommendation systems. The aim of this literature review is to pinpoint what ResuMatch is designed to solve by placing the proposed system in the larger research context and pointing out the weaknesses of earlier methods.

2.1 Traditional Resume Screening and ATS Systems

Most traditional recruiting processes are heavily reliant on manual rsum review, often supported by simple Applicant Tracking Systems (ATS) that mainly perform basic keyword matching. Studies reveal that over-reliance on keywords is a very fragile way of assessment that can be easily manipulated by merely inserting keywords without any meaningful context. At the same time, it is very difficult for the system to recognize different ways the same skills can be described [1][2]. Early ATSs, for the most part, limited themselves to picking up superficial features such as whether a certain skill was mentioned, the level of academic qualification, or the number of years worked, etc. They did not have the ability to understand the meaning of the content [3].

Besides that, the literature also shows that high-volume recruitment scenarios such as campus placements or large-scale job posts in digital portals can be very taxing both cognitively and temporally for the human resource professionals. The practice of manual screening has an inherent tendency to give priority to candidates who have perfectly formatted CVs and have prestigious educational backgrounds, even if they Msitigating possess the skills required. This results in the loss of talented individuals, and also, the explanation given to unsuccessful candidates is often not sufficient [5].

2.2 Job Recommendation and Candidate Ranking Systems

Previous studies have shown that ranking and recommendation systems could be used to match candidates with the most suitable job openings through embedding based similarity computations and hybrid scoring heuristics [9][10]. Some methods provide ranked lists of candidates based on a single suitability metric, while others use learning to rank models to predict both the probability of a match and the level of interview readiness [11]. Nonetheless, many of such systems remain black boxes; they do not display the reasons behind the rank order, and thus recruitment experts are left wondering why certain profiles are given priority over others [12]. ResuMatch tackles this major issue by assigning and explaining scores for each segment, as well as by explanation based matching, which together greatly increase the openness of the assessment process.

2.3 Fairness, Bias, and Ethical Issues in Hiring Systems

The raising of recruitment tasks automatization raises the serious issue of algorithmic bias, mainly when the models used to predict end up reproducing sext historical hiring patterns without being aware of that. The research highlights that the neutral choice of features, the implementation of fairness metrics, and the usage of explainable criteria are key to the prevention of discrimination. Even though most people agree that fairness in the methods used is important, there have not been many cases where explicit architectural features have been provided to disclose or reduce bias. ResuMatch examines this issue directly since the use of demographic proxies is completely eliminated in favor of the use of signals corresponding to only competencies and the elements of the decision-making process are made fully transparent.

2.4 Job Description Understanding and Semantic Matching

Job descriptions (JDs) are the main source of skills duties competencies, and experience requirements. Initially, some methods treated JDs as just unstructured lists of keywords and evaluated their matching with resumes by using term frequency or TF IDF scoring heuristics [4]. These methods completely ignored semantic context and therefore gave negative marks to candidates whose CVs have terminologically different but conceptually equivalent

expressions. Nowadays, embedding based representational frameworks are used to realize a substantial increase in the accuracy of the jobsum alignment [3][10].

Semantic matching architectures based on models like BERT, sentence transformers, or domain adapted embeddings help to uncover deeper relational structures among text segments. In this model, both resumes and JDs are converted to dense, low dimensional vector spaces, and similarity is measured by cosine or other similar distance measures. This allows the recognition of conceptual similarities, for instance, the fact that "backend development" and "REST API implementation" are equivalent, despite different words at the surface level [4][11].

2.5 Comparative performance of embedding models in recruitment contexts

This paper presents results from a research study that compared static embeddings, i.e. Word2Vec and GloVe, with contextual models, such as BERT and RoBERTa, on the task of jobsum matching. The results indicate that contextual models are better at capturing domain specific meanings primarily by disambiguating polysemous words and hence, bringing together semantically equivalent phrases that are worded differently. But this higher level of accuracy comes with the greater computational cost, which domain adapted embeddings fine tuned on occupational corpora can only partly solve. Without giving up too much accuracy you cannot just lower the amount or dimensionality of embeddings. ResuMatch is able to go around these compromise situations because it uses section wise embedding methods which make a trade off between the necessary representational accuracy and the potential scalability for production.

2.6 XAI in HR: Interpretable Models for Personnel Decision Support

Transparent score decomposition allows candidates to pinpoint exactly which competencies they are lacking and focus on those areas, a capability that totally matches what ResuMatch does: scoring candidates by sections and alerting them about missing skills. Giving justifications at the feature level for example, explaining how a certain skill affects the candidate's fit rating builds more trust of recruiters in the system. This change of mode is in line with changing regulatory requirements and societal pressures for explainable AI decision making. Nowadays, tools for post hoc explanations like SHAP and LIME provide realistic ways to break down scores without causing too much of a computational or architectural complexity lead.

2.7 Systems for Skill Gap Recognition and Constructive Feedback Provision

Current research recognizes that recruitment systems need to clearly point out competency gaps so that candidates can effectively improve their professional readiness by doing targeted remedial actions [12][19]. Identifying functional gaps through suitable systems can also help in the creation of training curricula, schooling programs, and developmental/professional pathways through projects [18]. Aligning with this pedagogical viewpoint, the ResuMatch model does more than just traditional selecting of candidates. It determines the competencies the candidates are lacking, lists the strengths of candidates that are already demonstrated, and provides helpful professional development steps.

2.8 Resume Format Variability

Research shows that reduced resume parsing accuracy is largely caused by some properties of the diverse formatted dictionaries, such as different document structures, non-standard section orders, and varying language expressions of the candidates. Unlike very clearly formatted technical papers, resumes often present information in quite innovative ways - they may have multi-columns, use icons, be in tables, or be template-based - all of these things together, from a rule-based ATS point of view, reduce the chance of successfully extracting the data [4][6][9]. Studies show that information about certifications, projects, or specialised skills can be very often found as part of narrative text, which means that finding such information using keyword-based heuristics is very difficult [10][11].

Therefore, the problem has only been solved with the help of semantic encoders and context-aware segmentation techniques that focus on understanding the content rather than relying on the position or formatting of the materials [14][15]. The ResuMatch system is based on a similar principle; it uses section-by-section similarity to measure and context-aware extraction to keep information retrieval working well even when the source documents show very different structural styles.

3 Proposed Methodology

The methodology we proposed has a multi stage pipeline including structured information extraction, semantic representation, and the weighted congruence computation aimed at automating the matching of potential candidate rsums with the related job descriptions. Our system design consists of four separate yet closely connected stages: parsing and preprocessing of rsums, normalizing and encoding job requirements, vectorial representation based on embeddings, and finally, calculating the total suitability score.

3.1 System Overview

The workflow starts with getting both the candidate resumes and the job description at the same time. Normally, an Applicant Tracking System(ATS) identifies discrete keyword occurrences for matching. In our case, a semantic comprehension model is adopted where even if two operators have different terminologies but convey the same meaning, their matching results will be similar and consistent. Additionally, the system provides a teardown level of areas of strengths and weaknesses uncovered, which increases interpretability for recruiters and applicants during the decision making process.

3.2 Evolution of Applicant Tracking Systems and their technological limitations

Applicant Tracking Systems (ATS) initially were nothing but simple repositories with a keyword search functionality but nowadays, they are capable of integrating machine learning based ranking. Early systems with first generation mainly relied on exact lexical matching of manually selected keywords; therefore, candidates using synonymous terms invariably were scored lower. Even top notch rule based parsers that can pull structured fields currently, they are still quite fragile when dealing with non standard formats like multi column layouts, tables, or graphical elements, causing them to make mistakes and have a higher number of false negatives. These enduring drawbacks provide the impetus behind the semantic, context aware extraction method that underlies the ResuMatch framework.

3.3 Machine Learning Models and Optimization

The framework that has been proposed makes use of three different models. The models are logistic regression, which is known for its simplicity and interpretability [16], Random Forest, which captures non linear feature interactions in a very complex way [15], and XGBoost, a model optimized using Optuna, which gives very good predictive results [14]. Optuna is a tool that very cleverly employs Bayesian optimization to automatically tune the hyperparameters such as learning rate, tree depth, and others, to prevent the model from overfitting [13]. It is a very attractive and logical idea to unite both interpretability and predictive power in one model. However, usually when the model is deployed to a production environment this may lead to less accuracy in explanations [14].

3.4 Resume Processing

Resumes that are uploaded are first turned into plain text from their original proprietary formats especially PDF or DOCX. After that, the document text is normalized in depth, e.g. by: The system is not fixated on templatebased parsing rules; it uses context managed processing to carry out these tasks.

3.5 Embedding Based Document Representation

Resumes of candidates and job descriptions are transformed into dense vector representations in order to facilitate a very thorough comparison of these two with each other. Such vector embeddings should reflect semantically very close concepts, hence the problem of lexical variability is eliminated. It should be noted that the vector representations of the text are created independently for several sections, e.g. skills, work experience education etc. which means that each of the aspects is measured separately rather than through one single aggregated scoring. This method of multi level basically layered analysis provides for a more in depth and detailed measurement of a job candidate fit while, at the same time, it helps to overcome the limitations of simple lexical similarity metrics that only focus on surface level term overlaps.

Dimension	Traditional hiring	ResuMatch (AI)
Screening method	Manual keyword search and subjective review	NLP + semantic embedding with cosine similarity
Bias risk	High — unconscious bias, inconsistent criteria	Reduced — objective scoring on skills and context
Screening time	Slow — days to weeks per job opening	Fast — automated pipeline processes instantly
Transparency	Low — decisions rarely explained to applicants	High — skill gap reports and match explanations
Candidate feedback	Rarely provided, generic rejections	Actionable — strengths, gaps, recommendations
Scalability	Limited by human reviewer capacity	Scales to thousands of resumes automatically



Fig. 1. Comparative Analysis of Traditional Hiring vs. ResuMatch AI-P

Fig. 2. ResuMatch AI-Powered Recruitment Pipeline

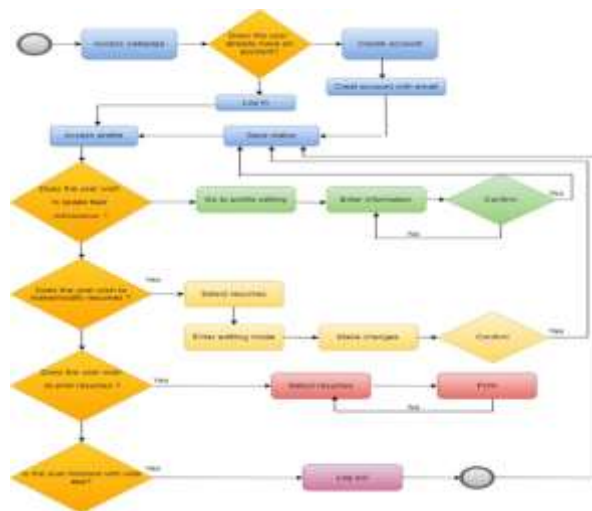


Fig. 3. User interaction and workflow process of the Re-sume Job Matching System

3.6 Multi Criteria Scoring Model

A weighted scoring system is used to calculate the overall suitability index, where each factor is assessed separately and then combined into a composite score. The model has different evaluation tables for skill matches, experience levels, education degrees equivalences and keyword or rolespecific relevances. In this model, competency based measures get the major weights, skill relevance being the highest priority, then experience. So professional core competences are the main basis for the recommendation result, not the surface or presentational features.

3.7 Output Generation

The last phase of the analysis generates a threefold output consisting of: a single suitability score given as a percentage, a skills comparison table showing matched skills to job requirements as well as gaps, and a scoring summary for each rsum section, thus evaluating candidates on various document parts. Besides these numerical evaluations, candidates get personalized pointers on their shortcomings that identify in what precise professional areas they need to improve for example, lack of knowledge of necessary software, outdated technology usage, or a shortfall in a specific functional domain. Therefore, the system is not just capable of sending a disqualification note to the candidates but also of providing them a broad informational base for the targeted remediation and getting ready for future careers.

4 Conclusion

This paper has introduced ResuMatch, an artificial intelligencebased system that helps in automating the semantic mapping of candidate rsums to job specifications based on structured extraction, contextual semantic comparison and multi dimensional scoring algorithm. The system identifies conceptually meaningful matching between documents instead of superficial lexical matching by using vector based similarity computations, thus providing candidate evaluations that are accurate and fair.

Besides ranking, ResuMatch also provides transparent and rationale based explanations by explicitly showing matched competencies, missing skills and the experiential evidence of proficiencies. Apart from aiding recruiters, candidates also get considerable advantages from receiving clear, actionable advice on how to increase their chances of being selected for a future job. The framework shortens screening cycles, increases inter rater ranking consistency, and decreases the dependence on subjective heuristic judgment.

Future work includes multilingual rsum processing, ATS integration, and the development of real time adaptive feedback modules. In sum, ResuMatch is a demonstration of how AI powered assessment tools can lead to increased trust, higher accuracy, and faster processing in the contemporary recruitment scenario.

Acknowledgment

KCC Institute of Technology and Management and the Department of Computer Science and Engineering are greatly thanked by the authors for providing the necessary infrastructure and support enabling the development of the ResuMatch framework. To the project guide and faculty mentors, the authors owe a great deal for their excellent guidance, continuous advice, and highly valuable intellectual contributions during the research.

Their deep knowledge in the areas of natural language processing, rsum parsing, recruitment automation, and AI-based matching systems was crucial not only to the theoretical development of this project but also to the realization of a tangible prototype from initial ideas. We also thank Dr. A.P.J. Abdul Kalam Technical University (AKTU) for consistently offering an academic environment that promotes innovation, deep exploration, and advancement in technology.

Besides, we would like to thank our colleagues, friends, and family members who, by giving us constant in-spiration, teamwork, and moral support, helped us to keep going even during tough times with the project. The authors gratefully acknowledge the joint efforts of the project team including Subhranshu Pathak, Aman Kumar, Vishal Gupta, and Vishwas Chakrawarti whose perseverance, excellent team spirit, and steadfast devotion played an integral role in the successful completion of this work.

References

1. Rigotti, C., Fosch-Villaronga, E.: Fairness, AI & recruitment. *Computer Law & Security Review* 53 (2024). <https://www.sciencedirect.com/science/article/pii/S0267364924000335>
2. Saatçi, M., Kaya, R., Ünlü, R.: Resume screening with natural language processing (NLP). *Alphanumeric Journal* 12(2), 121–140 (2024). <https://alphanumericjournal.com/article/resume-screening-with-natural-language-processing-nlp>
3. Bevara, R.V.K., et al.: Resume2Vec: Transforming applicant tracking systems with intelligent resume embeddings for precise candidate matching. *Electronics* 14(4), 794 (2025). <https://www.mdpi.com/2079-9292/14/4/794>
4. Singla, P., Verma, V.: An intelligent job recommendation system based on semantic embeddings and machine learning. *Journal of Information Systems Engineering and Management* 10(5s) (2025). <https://jisem-journal.com/index.php/journal/article/view/681>
5. Gondaliya, D., Kumar, A.S.: AI-based resume screening & ranking system. *International Journal of Sciences and Innovation Engineering* 2(8), 238–246 (2025). <https://ijsci.com/index.php/home/article/view/461>
6. Myneni, M.B., Quadhari, F.S.: Hybrid transformer-based resume parsing and job matching using TextRank, SBERT and DeBERTa. *SSRG International Journal of Electronics and Communication Engineering* 12(9), 63–71 (2025). <https://www.internationaljournalssrg.org/IJECE/paper-details?Id=989>
7. Mujtaba, D.F., Mahapatra, N.R.: Fairness in AI-driven recruitment: challenges, metrics, methods, and future directions. *arXiv preprint arXiv:2405.19699* (2025). <https://arxiv.org/abs/2405.19699>
8. Saxena, H., et al.: AI-powered resume analyzer and job matching system: a comprehensive review. *International Journal of Research and Innovation in Applied Science* 10(11) (2025). <https://rsisinternational.org/journals/ijrias/view/ai-powered-resume-analyzer-and-job-matching-system-a-comprehensive-review>
9. Yadav, S., et al.: Resume analysis using NLP and ATS algorithm. *International Journal of Latest Technology in Engineering, Management & Applied Science* 14(4), 761–767 (2025). <https://www.ijltemas.in/submission/index.php/online/article/view/1937>

10. Soleimani, M., et al.: Reducing AI bias in recruitment and selection: an integrative grounded approach. *The International Journal of Human Resource Management* 36, 2480–2515 (2025). <https://www.tandfonline.com/doi/full/10.1080/09585192.2025.2480617>
11. Heakl, A., et al.: ResumeAtlas: Revisiting resume classification with large-scale datasets and large language models. *arXiv preprint arXiv:2406.18125* (2024). <https://arxiv.org/abs/2406.18125>
12. Walid, O., et al.: MSLEF: Multi-segment LLM ensemble finetuning in recruitment. *arXiv preprint arXiv:2509.06200* (2025). <https://arxiv.org/abs/2509.06200>
13. Iso, H., Pezeshkpour, P., Bhutani, N., Hruschka, E.: Evaluating bias in LLMs for job-resume matching: gender, race, and education. *arXiv preprint arXiv:2503.19182* (2025). <https://arxiv.org/abs/2503.19182>
14. Skondras, P., Zervas, P., Tzimas, G.: Zero-shot resume–job matching with LLMs via structured prompting and semantic embeddings. *Electronics* 14(24), 4960 (2025). <https://www.mdpi.com/2079-9292/14/24/4960>
15. Gutiérrez-Batista, K., et al.: A novel approach for job matching and skill recognition from resumes. *International Journal of Interactive Multimedia and Artificial Intelligence* (2025). <https://www.sciencedirect.com/science/article/pii/S2214579625000048>
16. Saatçı, M., Kaya, R., Ünlü, R.: Resume screening with natural language processing (NLP). *Alphanumeric Journal* 12(2), 121–140 (2024). <https://dergipark.org.tr/en/pub/alphanumeric/issue/89551/1536577>
17. Ashutosh, Kumar, V.: AI-powered applicant tracking system: an intelligent approach to modern recruitment. *International Journal for Research in Applied Science and Engineering Technology* 13(11), 1913–1918 (2025). <https://www.ijraset.com/research-paper/ai-powered-applicant-tracking-system-an-intelligent-approach-to-modern-recruitment>
18. Singla, P., Verma, V.: An intelligent job recommendation system based on semantic embeddings and machine learning. *Journal of Information Systems Engineering and Management* 10(5s) (2025). <https://jisem-journal.com/index.php/journal/article/view/681>
19. Dutta, P.M., et al.: Reimagining recruitment: traditional methods meet AI interventions – a 20-year assessment (2003–2023). *Cogent Business & Management* 12(1) (2025). <https://www.tandfonline.com/doi/full/10.1080/23311975.2025.2454319>
20. Heakl, A., et al.: ResumeAtlas: Revisiting resume classification with large-scale datasets and large language models. *Procedia Computer Science* (2024). <https://www.sciencedirect.com/science/article/pii/S1877050924029909>