Beyond Keywords: Intelligent Resume Shortlisting Using Semantic Analysis

Jayashree Rajesh Prasad[1], Aayush Kholay[2], Rutuja Kshirsagar[3], Vedant Nandanwar[4], Vedant Sinare[5]
[1,2,3,4,5]Department of Computer Science Engineering, MIT Art, Design and Technology, Pune, Maharashtra, India.

***Abstract*—This Keyword-based filtering is the most used method by Traditional ATS for resume filtering. This results in strict textual matching which keeps eligible applicants out. The proposed solution for overcoming this limitation is the Smart Resume Shortlisting Tool: an intelligent candidate screener which helps enhance recruitment efficiency without completely replacing ATS platforms. To determine the contextual meaning of job descriptions and resumes this system makes the use of semantic analysis. Sophisticated NLP models are used to represent candidate profiles and job requirements as semantic embeddings. These semantic embeddings capture skills, experience, and role similarities that are beyond precise keyword matching. As a relevance indicator JD Matching Percentage, and for weighted similarity scores across multiple dimensions (skills, experience, and education) a candidate ranking module is applied. Additionally, the system offers side by side comparisons of candidates leading to a speeded process of informed decision making which ensures dynamic interpretation and context aware shortlisting against ATS. This model fits within the current hiring procedures and is an efficient way to assess the applicants altogether instead of the frequency of keywords.**

***Index Terms*—Candidate Screening, Semantic Analysis, NLP, Resume Ranking, Job Description Matching.**

I. INTRODUCTION

Finding the perfect candidate is challenging, even though recruitment is one of the most important processes. For automation of initial screening, the traditional ATS system is dependent on keyword-based filtering only, missing out on qualified candidates due to differences in terminology, phrasing, or context. Use of such strict measures can result in false negatives. Due to the absence of a few keywords, relevant candidates are rejected. This creates non-ideal hiring outcomes. We propose the Smart Resume Shortlisting Tool, an intelligent candidate screener designed to extend rather than replace the current applicant tracking system, which over- comes the drawbacks mentioned above. The system utilizes semantic analysis through NLP models, which transform job descriptions and resumes into high-dimensional embeddings. It depicts the contextual meaning of skills, experience, and education, due to which the system evaluates candidate applications based on relations and not on just matching texts. A weighted similarity metric with various factors like technical competence, educational history, and work experience is ap- plied for evaluating profiles by the candidate ranking module. The recruiters receive well-defined indicators of similarity through the job description matching percentage, which makes it easier to find the proper applicants faster. Along with that, the system supports side-by-side profile comparison to make conclusions and includes the integration of metrics, dynamic ranking, semantic understanding, and direct candidate comparison. Fairness and accuracy in the hiring process are thus achieved by addressing the drawbacks of conventional ATS.

A. Existing System and its Effects

Methods in use for recruitment process faces challenges for finding the right candidate and they ignores many deserving applicants because of strict keyword-based screening and lack of contextual assessing inefficient processing. Many qualified candidates are overlooked as their resumes do not match word- to-word with the job description. Traditional ATS systems are opaque, which means recruiters have no idea about the reason for a candidate's rejection or shortlisting. This can be damaging to effective and credible recruitment. Along with that, very minimal support is offered for side-by-side comparison of candidates, which requires manual evaluation, a process that is time consuming and error prone.

II. LITERATURE SURVEY

People have been doing a lot of research on how to make the process of looking at resumes. They want to use computers to help with looking at resumes. So they have been looking into using Artificial Intelligence and Natural Language Processing and Machine Learning to make the process of hiring people efficient and fair. They think Artificial Intelligence and Natural Language Processing and Machine Learning can really help with looking at resumes. The goal is to make hiring people more efficient and fair by using Artificial Intelligence and Natural Language Processing and Machine Learning to look at resumes Some studies have come up with systems that use Artificial Intelligence to look at resumes and rank them. This helps to reduce the amount of work that people have to do. It also helps to make the process more accurate.

There are some ways of understanding what is written in resumes that are very good at getting the meaning of the words. For example a model called BERT is very good at understanding what the words in a resume mean. This helps to match the person to the right job and it helps to get the right information from the resume. Other techniques like Named Entity Recognition and TF-IDF are also used to get the information from resumes and to make the process of looking at resumes better.

Some people are working on making resumes that're better for the computers to read. They are using something called Large Language Models to do this. This helps the people who are looking for jobs to have a chance of being selected. There are also computers that can help people make resumes that are customized and easy to read.

There are still some problems, with these systems. One problem is that they need a lot of data to work well. If the data is not good then the system may not work well. It may also have some biases. It may not work well with different types of resumes or languages. Another problem is that these systems need a lot of power to work and they can have trouble understanding the meaning of some words.

Overall the research shows that using Artificial Intelligence to look at resumes is an idea.. It also shows that we need to make these systems better so they can work with all types of resumes and languages. We need to make them stronger and more able to understand the meaning of the words.

A. *Motivation*

The need for precise effective and clear candidate screening system inspires the creation of smart resume shortlisting tool. This system make sure that the hiring decisions are based on context relevance and complete evaluation and not just text matching. It uses semantic analysis, candidate ranking, job description matching, and side-by- side comparisons. The result is improved hiring quality and fairness.

B. *Objectives*

- Build a tool that does not just uses simple keyword matching to identify the best candidates.
- Use semantic analysis to understand the context of re- sumes for job descriptions.
- Apply a similarity score and rank candidates more accu- rately. It takes skills, experience, and education into account.
- Show job description matching percentages. This way recruiters can easily see how relevant each candidate is.
- Side by side candidate comparison can be used to help recruiters to make faster and effective hiring decision.

III. PROJECT BACKGROUND

Recruitment directly influences team performance; hence, it is one of the most important organizational functions. The increasing number of applications for each job sector leads organizations to rely on Applicant Tracking Systems for automation of the initial screening process.

A. *Gap Identification*

Keyword matching is used mostly by existing systems for filtering candidates. Even though Applicant Tracking Systems can handle a large volume of resumes, they ignore the context, semantics, or relevance of skills or experience unless candi- dates use the exact keywords present in the job description. Also, these traditional ATS systems are opaque, which means recruiters have no idea about the reason for a candidate's rejection or shortlisting. This can be damaging to effective and credible recruitment. Along with that, very minimal support is offered for side-by-side comparison of candidates, which requires manual evaluation, a process that is time-consuming and error-prone. To overcome such issues, the Smart Resume Shortlisting Tool offers an intelligent candidate screening sys- tem. It ranks applications based on multidimensional similarity metrics, interpreting resumes and job descriptions semanti- cally, and presents a JD matching percentage. This offers transparency. Easy side-by-side candidate comparison helps recruiters make quicker, better, and fairer choices. The system increases the efficiency and quality of recruitment process. It acts as a thin pre-screening layer. It enhances and does not replace existing Applicant Tracking Systems.

B. Key Findings

1) Semantic Analysis and NLP Foundations: Goddard [1] offers a practical introduction to semantic analysis method- ologies that underpin automated text processing. Goddard and Schalley [2] further elaborate on semantic analysis techniques, establishing frameworks for interpreting textual meaning in computational contexts. Early applications of NLP to text analysis include sentiment extraction techniques. Nasukawa and Yi [3] demonstrated methods for capturing favorability in text using natural language processing. Yi et al. [4] developed sentiment analyzers that extract sentiments about specific topics using NLP techniques, presenting their work at the Third IEEE international conference on data mining. Miller et al. [5] presented a novel application of statistical parsing for information extraction from text, demonstrating techniques applicable to structured data extraction from unstructured documents.

2) Traditional Approaches to Resume Screening: Daryani et al. [6] developed an automated resume screening system utilizing natural language processing and similarity measures. Their approach demonstrated the feasibility of automating initial candidate screening processes. Pimpalkar et al. [7] conducted a detailed examination of resume sorting techniques using NLP and machine learning, analyzing various method- ologies and their comparative effectiveness in recruitment con- texts. Tijare [8] discussed practical implementations of resume sorting tools designed for custom shortlisting requirements, addressing organizational-specific needs in the recruitment process.

3) Transformer-Based Models for Recruitment: The ap- plication of BERT based models has significantly advanced resume screening capabilities. Deshmukh and Raut [9] devel- oped an enhanced resume screening system using Sentence Bidirectional Encoder Representations from Transformers (S- BERT), demonstrating improved matching accuracy for smart hiring. Bocharova et al. [10] introduced VacancySBERT, an approach specifically designed for representing job titles and skills for semantic similarity search in the recruitment do- main. Lavi et al. [11] developed ConsultantBERT, a fine- tuned Siamese Sentence-BERT architecture for matching jobs and job seekers. This work demonstrated the effectiveness of Siamese architectures for recruitment matching tasks. Decorte et al. [12] presented JobBERT, focusing on understanding job titles through skills analysis, establishing relationships between job roles and required competencies.

4) Comprehensive AI Driven Frameworks: Ajjam and Al- Raweshidy [13] present an AI-driven semantic similarity- based job matching framework for recruitment systems, repre- senting current state of the art approaches integrating semantic understanding with matching algorithms. Gangoda et al. [14] developed Resume Ranker, an AI-based skill analysis and skill matching system, demonstrating practical implementation of automated skill extraction and comparison.

5) Challenges in E-Recruitment Systems: Mashayekhi et al. [15] provided a comprehensive analysis of e-recruitment systems based on implementation challenges, identifying key obstacles in system design and deployment.

6) Text Summarization Techniques: Zhang et al. [16] con- ducted a comprehensive survey on process-oriented automatic text summarization with exploration of LLM based methods, examining techniques relevant to resume and job description processing. Fang et al. [17] investigated multi LLM text summarization approaches, demonstrating ensemble methods for improved summarization quality.

IV. METHODOLOGY

This tool is a smart approach for screening a candidate. It increases the efficiency of the recruitment process by assessing the candidates holistically.

A. Resume parsing

Here, the necessary data is being ex- tracted from the resume. This data is structural. It can be in any format like Word, PDF or text. This data can be per- sonal profiles, education, experience, skills and certifications. Parsing finds all relevant data in a format which can be read by a machine. Natural language processing (NLP) is used for identifying the main entities and to set the foundation for analysis in the future.

B. Semantic analysis

Most traditional ATS works on keyword matches. This method poorly captures the meaning of the text. The Smart Resume Shortlisting tools use semantic analysis such as transformer-based embeddings, Bert or Sentence- Bert. Here, the resumes and job descriptions are represented as high dimensional vectors. After that the contextual meaning of education, experience and skills are found. It enables the system to deeply understand the capabilities of the candidates. This works even after certain keywords are missing in JD.

C. Candidate Ranking

The candidate profiles are ranked in multiple dimensions. This task is carried out by the rank- ing module. A weighted similarity metric is used for this. After skill alignment, work experience, previous education, additional relevant compo- nents like certifications, projects are considered. A ranked list of applicants is being made by assigning them similarity scores for each. This helps the recruiters to give priority to candidates who are most relevant to the position.

$$cos(\theta) = (A \cdot B) / (||A||\ ||B||)$$

D. Best Match with Job Description

Each candidate's job description matching percentage is calculated by the system. This score gives information about the candidate's profile fit for the given job requirement. It helps the recruiters to understand why a candidate ranks higher. The job description matching percentage addresses common flaws in traditional ATS. The shortlisting process is made clear.

$$S = w_1S_k + w_2S_e + w_3S_{edu} + w_4S_c$$

where S_k denotes the skill match score, S_e denotes the experience score, S_{edu} denotes the education score, and S_c denotes the cosine similarity score.

E. Candidate Comparison

The system allows side-by-side graphical comparison. It also shows key metrics like education, experience, skills and JD percentage matching. Time for manual evaluation is reduced. It helps the recruiters to make quick fair and right decisions

F. Workflows and Integration

This tool is lightweight. It can be integrated easily with the existing applicant tracking systems. It is flexible, effective for companies of all sizes. It focuses on screening and ranking the applicants. It does not require you to replace the complete applicant tracking system.

V. WORKING

A. Recruiter Input

This process starts when recruitment provides the resumes of applicants, mostly in PDF or DOCX format, with compre- hensive details about the role. This system is dependent on this information because it takes the job specifics as a reference and matches the resumes against them. The input tool supports many types of resumes and provides a user-friendly interface to input the details

B. Parsing and Text Extraction

The uploaded documents are processed for text extraction, converting various PDF or DOCX files to readable lines. It uses tools like pdfminer and docx2txt to pull out the text while maintaining sections such as experience, education, and skills in order. Extracts details about the job to maintain consistency in data for comparison.

C. Natural Language Processing

Then, the extracted text will be processed and refined using NLP techniques. Skill identification is done by the library spaCy, which executes this function using named entities and rules that identify technical and interpersonal skills. Fur- ther, the transformer model MiniLM-l6-v2 will transform the treated text into numbers able to grasp word relationships. This approach transforms text into numbers for meaning-based matching.

D. Semantic Similarity Calculation

It makes use of cosine similarity to match resumes to a job description. In that way, the system can verify if the candidate's background fits the needs of the job by searching for logical coherence, not simple keyword matches. High cosine values indicate a good fit between the candidate and the role, though the results are sensitive to wording.

E. LLM Summarization

A large text model then summarizes each candidate's resume. It summarizes key skills, job history, and education in a way that can give recruiters a fast and helpful overview. In this way, review times are faster and less effortful.

F. Candidate Scoring and Ranking

Candidates are scored by the measure of similarity to the job description. Scores derive from the calculation of similarity. A scoring system sorts candidates according to their skills, experience, or education. The system shows skill gaps when some of the applicant's skills fail to match the requirements of the job. Candidates are listed, with top performers being the first.

G. Visualization and Result Export

Export Results are shown using charts or split views that allow side-by-side comparison of candidate matches. Results can be saved to a CSV format for review or record-keeping in Excel. Presentation of data based on the values calculated will help make informed hiring decisions.

H. Session Cleanup (Privacy)

With regard to ensuring data security, all the files uploaded and created in the process get deleted once the session is complete. Indeed, this is a common practice while handling data, instilling confidence in the security features of the system.

VI. SYSTEM ARCHITECTURE

The proposed Smart Resume Shortlisting System is de- signed as a multi-layered architecture that integrates Natural Language Processing (NLP) and semantic intelligence to auto- mate candidate evaluation and ranking. The system processes resumes and job descriptions, extracts meaningful information, computes similarity scores, and generates ranked outputs for recruitment decisions.

The architecture is divided into four major layers, each responsible for a specific stage in the pipeline:

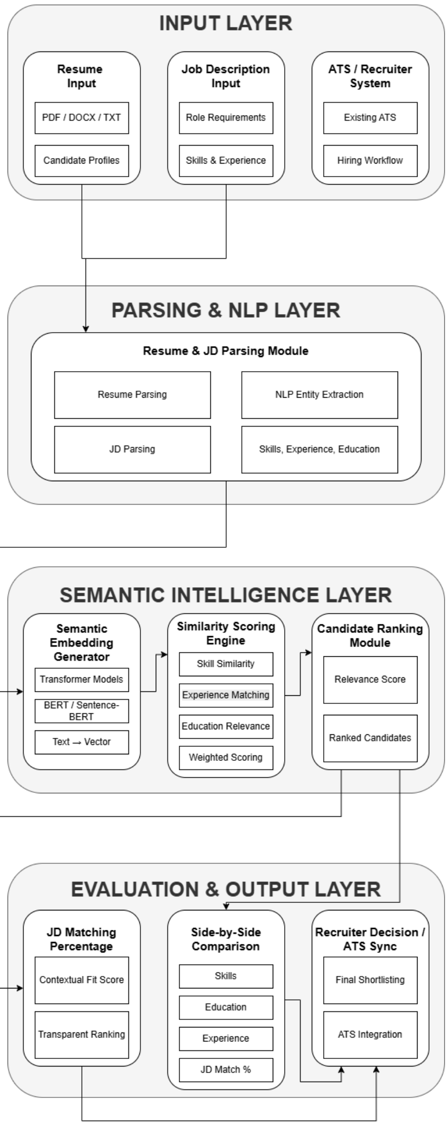

Fig. 1: System architecture of the proposed model

A. Input Layer

This layer handles all incoming data sources. It accepts resumes in multiple formats such as PDF, DOCX, and TXT, along with job descriptions containing role requirements, skills, and experience details. Additionally, it integrates with existing Applicant Tracking Systems (ATS) to utilize recruiter workflows and candidate databases.

B. Parsing and NLP Layer

In this layer, resumes and job descriptions are processed using NLP techniques. The system performs resume parsing and job description parsing to extract structured information. Name Entity Recognition (NER) and text processing methods identify key attributes such as skills, experience, and educa- tional qualifications.

C. Semantic Intelligence Layer

This is the core layer of the system. It generates semantic embeddings using the transformer based models such as BERT or Sentence BERT to capture contextual meaning. A similarity scoring engine evaluates candidates based on skill similarity, experience matching, and educational relevance. These scores are combined using weighted scoring techniques, and a rank- ing module produces a prioritized list of candidates based on relevance.

D. Evaluation and Output Layer

The final layer presents the results in a meaningful way. It computes job description matching percentages and provides side-by-side comparisons of candidates based on skills, educa- tion, and experience. The system enables transparent ranking and supports recruiter decision-making. The shortlisted candi- dates can be seamlessly integrated back into ATS platforms for final selection and hiring workflows.

The architecture makes sure the recruitment process is efficient and smart. It does this by using Natural Language Processing and other methods to understand what people are saying. The system also automatically ranks people to make things easier. This way the recruitment process becomes more efficient.

ACKNOWLEDGEMENT

The authors acknowledge the assigned project guide and mentor, Dr. Jayashree Rajesh Prasad, for their association with this work. We appreciate their role in overseeing the project as per institutional requirements.

We also extend our sincere gratitude to our institution for providing the necessary resources, infrastructure, and aca- demic environment that allowed us to successfully complete this project. Finally, we thank all those who directly or indirectly contributed to the completion of this work.

REFERENCES

[1] G. Navarro, "Fair and Ethical Resume Screening: Enhancing ATS with JustScreen the ResumeScreeningApp," Journal of Information Technology, Cybersecurity, & AI, vol. 2, no. 1, pp. 1–7, Jan. 2025

[2] M. V. Koroteev, "BERT: A Review of Applications in Natural Language Processing and Understanding," arXiv preprint arXiv:2103.11943, Mar. 2021.

[3] J. P. Jagadish, A. V. Abhishek, A. Shukla, A. V. and P. K. Reddy, "Resume Screening and Ranking with spaCy," Turkish Online Journal of Qualitative Inquiry, vol. 12, no. 7, 2021.

[4] S. Anand and A. K. Giri, "Optimizing Resume Design for ATS Com- patibility: A Large Language Model Approach," International Journal on Smart & Sustainable Intelligent Computing, 2024

[5] R. Deeptha, A. Dey, G. S. G. R. Gowtham, and A. Tripathi, "Resume Screening using NLP," i-manager's Journal on Information Technology, vol. 13, no. 3, pp. 37–43, 2024

[6] A. Deshmukh and A. Raut, "Applying BERT-Based NLP for Automated Resume Screening and Candidate Ranking," Annals of Data Science, vol. 12, no. 2, pp. 591–603, 2025

[7] A. Kalpund, S. Jadhav, P. Chemate, and S. Pawar, "AI-Powered Resume Parsing using Django: A Smart Recruitment Solution," in *Proceedings of the 2025 International Conference on Data Science and Business Systems (ICDSBS)*, Chennai, India, 2025, pp. 1–5

[8] K. Poongothai, S. Aarthi, A. N. Loshiya, and S. Anusri, "INTEL- LIRECRUIT: An AI-Powered Resume Screening System Integrating Ma- chine Learning, Natural Language Processing, and Fair Recruitment Au- tomation," in *Proceedings of the 2025 5th International Conference on Evolutionary Computing and Mobile Sustainable Networks (ICECMSN)*, Coimbatore, India, 2025, pp. 1–4

[9] A. Rajput, A. Dubey, R. Thakur, D. Singh, and U. P. Singh, "Career Craft AI: A Personalized Resume Analysis and Job Recommendations System," in *Proceedings of the 2024 1st International Conference on Innovative Sustainable Technologies for Energy, Mechatronics, and Smart Systems (ISTEMS)*, Dehradun, India, 2024, pp. 1–6

[10] V. A. R., R. K. P. Kumar, S. P. V. V. N. S., V. K. V., and S. P., "An ML-Based Resume Screening and Ranking System," in *Proceedings of the 2024 International Conference on Signal Processing, Computation, Electronics, Power and Telecommunication (IConSCEPT)*, Karaikal, India, 2024, pp. 1–6.

[11] S. Ambareesh, N. K. Thakur, U. Bhattarai, S.

K. Yadav, J. N. Thakur, and A. K. Mahato, "Resume Shortlisting Using NLP," in *Proceedings of the 2024 4th International Conference on Data Engineering and Communication Systems (ICDECS)*, Bangalore, India, 2024, pp. 1–5
[12] R. Sampath, A. S., A. C., and P. K. S., "Automated Resume Production System: A Structured Approach to Efficient Resume Management and Classification," in *Proceedings of the 2025 International Conference on Computing and Communication Technologies (ICCCT)*, Chennai, India, 2025, pp. 1–6
[13] A. Mukherjee and U. S. M., "Resume Ranking and Shortlisting with DistilBERT and XLM," in *Proceedings of the 2024 IEEE International Conference for Women in Innovation, Technology & Entrepreneurship (ICWITE)*, Bangalore, India, 2024, pp. 301–304.
[14] S. Dhokane, C. Deshmukh, A. Bollabattin, S. Karande, B. Karangale, and P. S. Varade, "BM25 Implementation for Information Retrieval: Candidate Shortlister for Recruitment Process," in *Proceedings of the 2024 Intelligent Systems and Machine Learning Conference (ISML)*, Hyderabad, India, 2024, pp. 722–727.
[15] V. P. R., Y. S., M. Revathi, and J. G. Beulah Patturose, "Resume Ranker for Freshers Using NLP," in *Proceedings of the 2025 7th International Con- ference on Innovative Data Communication Technologies and Application (ICIDCA)*, Coimbatore, India, 2025, pp. 902–907.
[16] S. Bharadwaj, R. Varun, P. S. Aditya, M. Nikhil, and G. C. Babu, "Resume Screening Using NLP and LSTM," in *Proceedings of the 2022 International Conference on Inventive Computation Technologies (ICICT)*, Nepal, 2022, pp. 238–241