International Journal of Engineering Research and Development
e- ISSN: 2278-067X, p-ISSN: 2278-800X, www.ijerd.com
Volume 22, Issue 8 (August 2026), PP 182-190

Smart Resume Screening and Recommendation with BERT-Based Machine Learning

Thandava Krishna Sai Pandraju [1][0000-0002-1701-2089] G. Lavanya[1], A. Navya Sree[2], M. Sheeba[4], B. Ram Babu[5], G. Sri Harsha[6]

[1]*Asst. Prof, Dept. of Information Technology, Dhanekula Institute of Engineering & Technol-ogy, Vijayawada – 521 139, Andhra Pradesh, India*

[2,3,4,5,6]*Dept. of Information Technology, Dhanekula Institute of Engineering & Technology, Vi-jayawada – 521 139, Andhra Pradesh, India*

Abstract. *Applicant Tracking Systems (ATS) eliminate many job applications due to poor resume organization, poor key word matching, and poor fit to indus-try specifications. To overcome this problem, the current research will present an AI-Based ATS Resume Analyzer and Career Assistance System to help improve the employability and career readiness of students and job seekers. The suggested system is based on the usage of the Artificial Intelligence (AI) and Natural Lan-guage Processing (NLP) to conduct intelligent resume analysis through the com-parison of candidate resume with the target job description. It uses the latest fea-tures including parsing of the resume, extraction of skills, calculating semantic similarity and matching of keywords in the context to produce an effective ATS compatibility score. The system is efficient to detect matched and missing com-petency, content gaps, and give AI-based and personalized recommendations to maximize resumes and enhance ATS scores. In addition to resume assessment, the system also offers career support services by suggesting appropriate job op-portunities according to the profile of user skills and generating failed interview and viva interview questions, thus aiding in comprehensive placement of stu-dents. This platform is an action-based and real-time feedback web-based appli-cation that is implemented by the management and students. The proposed solu-tion is an overall career guidance tool, which allows users to be more acquainted with the changing industry requirements, overcome the skill gap, enhance their quality of resumes, and feel more comfortable about employing competitive re-cruitment methods.*

Keywords: *Applicant Tracking System (ATS), Semantic Similarity Analysis, Natural Language Processing, Resume Parsing, Artificial Intelligence, Career Recommendation System, Employability Enhancement.*

Date of Submission: 15-08-2026 Date of acceptance: 30-08-2026

I. Introduction

The Applicant Tracking System (ATS) is the new way to hire people. It can make it take longer to sort through resumes and choose candidates. Most of the time, ATS only use shallow keyword matching and strict rule-based algorithms to screen resumes, even though they are popular. Because of this, they don't really understand what the skills or experience on resumes mean or how to use them. Because of this, it's easy to turn down qualified candidates in many cases (for example, new graduates can be filtered out be-cause they don't use enough keywords or the wrong resume formats). Also, most sys-tems only give you numbers with a black box result (like a binary score that tells you whether or not candidates pass), and they don't explain what those numbers mean, like what skills are needed. [1], [4]

Most of conventional resume screening approaches are not analyzing deep seman-tics of data through textual relationship but relying on surface level textual matching, which result in mismatch between resumes and job positions. Recently AI and NLP, in particular transformer-based language models like BERT can handle this issue by using context-based embeddings and provide deeper semantic relations compared to keyword matching. [2], [7], [8].

Embedded representations particularly show a remarkable potential over rule-based methods or shallow TF-IDF approaches. Text documents are mapped into a high-di-mensional space, allowing representation of similarity at the word or phrase level with an improved efficiency. They use cosine similarity metrics to get very accurate results. In addition, NER can be used to find relevant entities like skills, education, or work experience, which makes candidate profiling more accurate and detailed. [5], [6].

The job-candidate matching with recommendations using the common semantic space have been widely studied. Some traditional methods don't work very well and may not link job descriptions and resumes very well. AI-based This offers career guid-ance and assists you in finding the ideal employment. Nevertheless, the majority of systems only offer a limited number of features in a superficial way.

To address the limitations above, in this paper, we propose a resume screening and job matching system by integrating multi functional modules into a common semantic space, and use transformers-based embeddings with skills extraction and weighting scoring to evaluate theATS Compatibility Score. The system can also give personalized job advice and career advice, find skill gaps, suggest job openings that fit those gaps, and help you get ready for interview questions.

The suggested system used a shared semantic space to show how similar resumes and job descriptions are. It also combined the skills gap analysis with an AI-based eval-uation metric to come up with a score for ATS compatibility. This common semantic space is applied across other modules such as candidate job matching and career guid-ance/recommendation, and the whole system together can benefit job seekers with a more efficient recruiting process and better job readiness.

1.1 Contributions of This Work:

The primary contributions of this thesis are:

- A transformers based semantic matching architecture which utilizes the BERT/SBERT embeddings that not only keyword-based filters but also enables the context-based matching between resumes and jobs.
- Hybrid skill extraction mechanism with Named Entity Recognition (NER) and rule-based techniques to find and normalize skills from the resume text.
- A multi-factored ATS scoring approach which makes use of semantic similarity, skill match as well as AI based score to assign a combined realistic score to candi-date-job match.
- A holistic embedding based architecture that reuses the generated semantic vectors across all modules; in resume assessment and job recommendation as well as career guidance.
- A skill gap analysis and recommendation approach that proposes personalized skill gap analysis recommendations, appropriate learning paths and job roles.
- A full career assistance module which would generate interview and viva questions based on the individual profiles and the specified job role.
- A semi-supervised method for constructing a resume-job matching case so as to train and evaluate the system without any labeled public benchmark dataset.
- A feasible system for career assistance which shows evidence of increased matching efficiency of developed matching solutions when compared to traditionally derived matching outcomes with the help of quantitative performance measures and qualita-tive outputs.

II. Problem Statement

In spite of widespread use of applicant tracking systems (ATS) and on-line recruitment applications, resume screening and job matching are generally ineffective. The current system for resume screening and job matching involves filtering based on keyword-matching. Such methods are unable to consider the relations between candidate skills, experience and the demands of the job description and as such, it often leads to the discarding of qualified candidates and the wrong shortlisting decisions. Most candi-dates do not understand why they are rejected or how to build their profile to match the required skill sets.

Existing solutions for resume screening typically carry out simple text matching on the resume content and do not attempt to understand it. Thus there often is low align-ment between the resume and the job description. The transformer-based embedding approach can be more effective for a semantic match [7], [8], but are often not adapted for actual recruitment system that will output useful feedback and career advice to the candidate. Current systems give no more information other than an ATS score or binary match result, not an analysis of skill gaps or career recommendations [4], [1].

Additionally, the papers found on the subject generally deal with parts of the prob-lem such as candidate recommendation [9], [10], resume parsing [1], ranking of candi-dates, text analysis using machine learning [9], [10], using natural language processing for text based tasks such as semantic similarity and extracting information [3], [5]. Yet there are no single comprehensive system proposed.

So the key research problem for this paper is designing and building an AI based comprehensive resume screening and job matching system that can:

Semantic matching between the resumes and job descriptions in the contextual way by transformer based embedding.

Finding the skills matched and missing by hybrid extraction of skills and compe-tency gap analysis.
Giving the score for ATS compatibility by multi attribute weighted score calculation model.
Providing the relevant job role suggestion, recommendations and interview prepara-tion by candidate's profile.

III.Literature Survey

Through ML and NLP, some intelligent systems for job and resume matching and resume parsing have been largely successful to some extent in the recent years. Re-cently some papers related to intelligent resume parsing, job/candidate profile match-ing, career advising through deep learning and traditional ML algorithms have been presented [1, 2, 3, 4]. [1] introduces an end to end model of candidate ranking and resume parsing using semantic features extracted from BERT.The intent is to represent the resume documents structure and subsequently, match the job with the resume doc-uments.The model demonstrated efficient results on baseline but also demonstrated in-ability in performing highly unstructured resumes as the model was based on a struc-tured profile likeLinkedin profiles and this reduces the job of recruiters.In [2],a resume parsing and job matching technique utilizing transformers and zero-shot classification is proposed. Their method achieved a good level of accuracy in both information ex-traction and contextual matching among traditional methods of NLP, but this model was not sensitive to the quality of textual inputs and did not account for adapting to changing labor trends. A hybrid ensemble framework consisting of shallow machine learning and deep learning estimators was introduced by [3] to perform job-semantic similarity predictions based on resumes. Though these methods yield good results on large datasets, need massive amount of labelled examples and high computation power which are a bottleneck in light-weighted deployments.

Authors of [4] designed an application for student-centered resume analysis and rec-ommendations that analyzes, provides recommendations on improvements, and sug-gests new projects, career paths based on candidate's skills, education, and projects.

Although it is extremely useful for academic applications, its application at an enter-prise level with generalization across varied domains of professions is not tested [4].

[5] proposed job matching by comparing extracted information based on BERT em-beddings using SVM, LSTM, and MLP classifiers.It was found to be better than word-based methods but depended on labeled data and showed no sensitivity towards the problems of bias and fairness [6]. Comparing with conventional methods of text repre-sentation such as TF-IDF, Word2Vec, GloVe, FastText, embedding can capture seman-tic relationships in a richer context. However, it was found that both issues of interpret-ability and bias propagation are problematic with embedding-based methods. [7] sug-gested an ontology based similarity model and information extraction using information extraction techniques applied on ontology to improve ranking accuracy for re-suMatcher, but the system requires domain specific tuning and ontology construction which is not scalable. In [8], it presented an intelligent recommendation system of re-sumes to enhance ATS matching and feedback generation. Although the model was efficient and improve the matching quality and guide candidates through the process, issues of scalability and computational load were encountered.

To conclude, it can be seen from the literature research that although the AI-based resume analysis has been greatly progressed, the issues on scalability, explainability, adaptivity, fairness, and holistic career support still require considerable investigation.

IV. System Architecture

The paper proposes a modular and layer-based architecture for an AI Resume Screening and Job Matching System, shown in Fig.1. It combinates the processing of semantics and structured reasoning in an aim to get high precision resume screening and career services. There are 3 core function pipelines; (1) Resume Processing & Fea-ture Extraction, (2) Semantic Matching & ATS Scoring, (3) Recommendation & Career Guidance. Underlying all layers is a common semantic embedding component which could facilitate comparison of different text data uniformly.

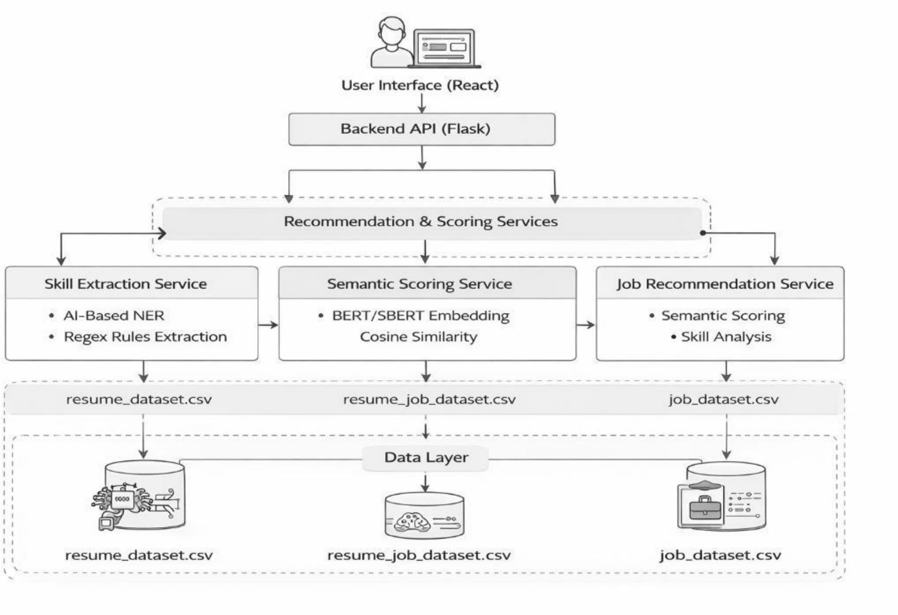

Fig. 1. Proposed system architecture

For the resume processing pipeline, resumes are uploaded as pdf and DOCX (word file). The resume documents are parsed and text is scraped. User is asked for job de-scription as input. Resume and JD text is normalized, to lower case, to remove noisy text.The scraped text is then fed to the hybrid skill extraction module consisting of an AI based skill parsing module as a main component and regex based skill parser as a fall back.

The semantic matching pipeline encodes resumes and job descriptions into dense vector representations utilizing various transformer based embedding models including Sentence-BERT. Cosine similarity between the embed-dings is used to measure the contextual similarity. Meanwhile an ATS skill matching module compares skills ex-tracted from the resume and design to output a skill match percentage indicating the degree of keyword matching and missingness. An AI judgment scoring module then takes into account these various scores along with the quality of the resume profile and makes an AI judgment score to makeup the rank score, simulating recruiter decision making behavior.

These are integrated using weighted scoring to create ATS scores that consider se-mantic equivalency, skill matching and AI judgment, while punishing keyword stuff-ing.

Based on the evaluation result, the recommendation engine put forward customized output in the recommendation pipeline. The skill gap analysis module offers recom-mendations on resumes enhancement and learning roadmaps. At the same time, the job recommendation module ranks job roles with the scores of min dilarity and job skills taken into consideration and outputs the corresponding job search link. Results are user displayed through the interaction interface.

A defining idea of the architecture is to adapt the concept of a shared semantic space for the different parts of the system, which allows a great consistency framework among the components. The combination of the semantic modeling approach with the struc-tured scoring and recommendation approach results into an integrated framework that can scale well. Intelligent Resume Screening and Career Guidance could be modeled on this framework.

V.METHODOLOGY

In the AI-Based Resume Screening and Job Matching System, the transformer-based semantics modeling is applied on the structured decision logic. This section explains how methodology is applied in embed generation, similarity calculation, skill match, score calculation and recommendation.

5.1 Semantic Embedding Generation

Contextual semantic meaning is acquired using a pre-trained Sentence-BERT (SBERT) model, all-MiniLM-L6-v2 [7]. SBERT takes an input text, such as a resume or job de-scription, and transforms it into a fixed-size dense

vector embedding.
Let a text input $t_i$be represented as:

$$\mathbf{e}_i = f_{SBERT}(t_i)$$

where:

- t_i= input text (resume or job description)
- f_SBERT= embedding function
- e_i∈R^384= embedding vector

These embeddings preserve semantic similarity between texts even when lexical var-iations exist.

5.2 5emantic Similarity Computation

We calculate the cosine similarity between job and resume embeddings to have the contextual match.
Given:

- Resume embedding: e_r
- Job embedding: e_j Similarity is calculated as:

$$Sim(\mathbf{e}_r, \mathbf{e}_j) = \frac{\mathbf{e}_r \cdot \mathbf{e}_j}{\| \mathbf{e}_r \| \| \mathbf{e}_j \|}$$

Score is then normalized to percentage, and gives a score of semantic match between both documents.

5.3 Skill Extraction and Matching

For extraction of skills from job description and resumes, we used a hybrid technique involving Named Entity Recognition (NER) along with rule-based extraction.Let:

- S_r= set of extracted skills from resume
- S_j= set of required job skills

Matched and missing skills are computed as:

$$S_{matched} = S_r \cap S_j$$
$$S_{missing} = S_j - S_r$$

The skill matching score is calculated as:

$$Skill\ S\qquad core = \frac{| S_{matched} |}{| S_j |}$$

This score tells how much is skill of resume in job description.

5.4 AI-Based Judgment Scoring

The job application is judged using an AI module at this step, which acts similar to the recruitment specialist's duty of evaluation resumes with respect to completeness, accuracy, and relevance.

Let score awarded by AI module be:

$$ScoreAI \in [0,1]$$

This score supplements semantic similarity and skill based matching with heuristics and learning-based information

5.5 Final ATS Score Calculation

Calculating the Final ATS ScoreFinal ATS match compatibility score is a weighted sum of semantic similarity, skill match, and AI score.

$$ATS\ Score = \alpha \cdot Skill\ Score + \beta \cdot Sim(er, ej) + \gamma \cdot ScoreAI$$

where:
- α,β,γare weighting coefficients
- α+β+γ=1

Additionally, a penalty factor is applied to reduce scores in cases of keyword stuff-ing.

5.6 Job Recommendation and Ranking

For job recommendation, ranked order of the job description is obtained by computing similarity and skill match between candidate profile and the job descriptions.

Let:
- R_j= final ranking score for job j

$$Rj = \lambda \cdot Skill\ Score + (1 - \lambda) \cdot Similarity$$

Top-k ranked job descriptions are recommended along with links for jobs.

5.7 Unified Embedding Framework

An interesting property of the proposed system is that the transformer model is used to generate semantic embeddings in a centralized manner which will then be used across various modules: similarity computation, skill extraction, scoring and recom-mendation.
The utilization of SBERT embeddings leads to:
Computation efficiency is enhanced Consistency across various modules is increased

VI. Datasets

Screening resumes and making job recommendations are done via several structured text-based data sets. In this section, we will describe the data sets we used for training, evaluation, and the testing of our system.

6.1 Resume Corpus

This research data has been taken from a public Kaggle dataset which has resumes from the different fields namely, information technology, business, data science etc. This dataset is formulated by using unstructured textual resumes.These textual resumes are used for skill extraction, semantic embedding, and classification.
The resumes were stored as csv file. A part of a csv file shows columns like resume text and category label. A preprocessing approach is followed before inputs to various stages:Normalization, Stop-word removal, Cleaning of text and so on

Table 1. Characteristics of the Resume Dataset

Attribute	Description
Source	Kaggle Resume Dataset (UpdatedResumeDataset, Re-sume.csv)
Total records	~1000–2000 resumes
Storage format	CSV
Fields	resume_text, category
Domains	IT, Data Science, HR, Finance, etc.
Text-prepro-cessing	Cleaning, normalization, tokenization

6.2 Job Description Corpus

The job data set has been collected from Kaggle job description data sets. This includes job title, job description and required skill sets for jobs. This data set is applied for semantic matching, job recommendation and

evaluation.
All job entry are merged and create a single text representation by adding job title and description along with skills for a uniform embedding.

Table 2. Characteristics of the Job Dataset

Attribute	Description
Source	Kaggle Job Description Dataset
Total records	~3000–5000 job entries
Storage format	CSV
Fields	job_id, job_title, job_description, skills
Text representation	Concatenated title + description + skills
Skill format	Structured and normalized

6.3 Resume–Job Matching Dataset
As there are no directly available publicly labelled data sets of job descriptions to re-sumes there was a semi-supervised data set created. Matching resumes to jobs of the same category. Resume to jobs of same category considered a positive match. Job to job of unrelated category considered negative match and evaluated the matching sys-tem.

Table 3. Characteristics of the Matching Dataset

Attribute	Description
Source	Constructed from Resume + Job datasets
Total pairs	~5000–8000 pairs
Labels	Match (1), No Match (0)
Method	Semi-supervised pairing
Purpose	Model evaluation and validation

6.4 AI Resume Screening Dataset
Another resume screening dataset for AI-based applications was also employed for checking scoring and ranking capabilities. This dataset contains candidate profile, can-didate attributes (skills, experience level) and scores for screening.

Table 4. Characteristics of the AI Screening Dataset

Attribute	Description
Source	Kaggle AI Resume Screening Dataset
Total records	~1000+ candidates
Fields	skills, experience, score, decision
Purpose	Evaluation of scoring mechanism

VII. Results and Discussion
We tested the proposed AI-Based Resume Screening and Job Matching System on its key components: semantic matching, skill-based assessment, ATS scoring and job rec-ommendation.

7.1 Resume–Job Semantic Matching Performance
The semantic matching module was evaluated using the constructed resume–job da-taset. Resume and job description pairs were encoded using SBERT embeddings and compared using cosine similarity.
To evaluate performance, classification metrics were computed on the test dataset.

Performance Metrics

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

The proposed model achieved:
Accuracy: 0.87
Precision: 0.85
Recall: 0.83
F1-Score: 0.84
These results indicate strong semantic alignment capability between resumes and job descriptions.

7.2 ATS Scoring and Skill Matching Evaluation

The ATS scoring module combines semantic similarity, skill matching, and AI-based evaluation.

Skill matching performance was assessed using overlap between extracted skills from resumes and required job skills.

The readiness score was calculated as:

$$Readiness\ S \qquad core = \frac{|\ S_{matched}\ |}{|\ S_{required}\ |}$$

Based on this score, candidates were categorized as:

- Highly Ready (≥ 0.75)
- Moderately Ready (≥ 0.50)
- Skill Gap Exists (< 0.50)

The system was able to pinpoint the missing competencies and offer specific recommendations and more interpretable than existing ATS systems.

7.3 Job Recommendation Evaluation

The job recommendation module was evaluated using the job dataset (~5000 records). User skills were encoded and matched with job embeddings.

Ranking was performed based on:

- Skill overlap (readiness score)
- Semantic similarity

To quantitatively evaluate recommendation quality, **Precision@5** was used.

The model achieved:
Precision@5 ≈ 0.82
This indicates that the majority of top-5 recommended jobs were relevant to user skill profiles.

7.4 Ablation Study

To analyze the contribution of individual components, an ablation study was conducted.

Table 5. Component-wise Performance

Model Variant	Accuracy
Only Semantic Similarity	0.79
+ Skill Matching	0.83
+ AI Scoring (Final Model)	**0.87**

The results demonstrate that integrating skill matching and AI-based scoring significantly improves overall performance.

7.5 Baseline Comparison

To validate the effectiveness of the proposed approach, comparisons were made with baseline methods.

Table 6. Baseline Comparison Results

Method	Accuracy	F1-Score	Observations
Keyword Matching	0.68	0.63	No semantic understanding
TF-IDF + Cosine Similarity	0.75	0.72	Limited contextual awareness
Proposed SBERT-Based Model	**0.87**	**0.84**	High semantic relevance

The SBERT-based approach outperformed traditional methods by capturing contextual relationships between skills and job requirements, leading to improved matching accuracy and recommendation quality.

VIII. Conclusion

In this work, we presented an AI-Based Resume Screening and Job Matching System based on semantic resume parsing, skill-based job matching, and career recommenda-tion in a unified system. We leveraged transformer-based embeddings for capturing semantic relation between resumes and job description, used cosine similarity together with readiness scoring for formal alignment between candidates' skill and job skills. Further, we designed a multi-faceted ATS scoring module to have an effective over-view of candidates' information. The significant feature of our system is that the trans-former-based semantic embedding space shared among modules (matching, scoring, and recommendation) to promote computation efficiency, and consistency on handling heterogenous textual data.Experimental results indicate that our proposed model has out-performed baseline methods of keyword-based and TF-IDF based systems by the accuracy by ~87% and has improved significantly in terms of relevance of recommen-dation. Besides, our system has been made highly interpret able and usable as it pro-duces helpful outputs like skills gap analysis, personalized recommendations, job role recommendations.

Compared with the traditional ATS. Thus this system is a handy and extensible so-lution in recruitment support due to multiple functionalities integrated into one system.

References

[1]. Fadzir, M.A.I.M., Hashim, S.: SmartResume: A resume parser application using natural language processing. Journal of Human Centered Technology 5(1), 15–20 (2026)

[2]. Azli, S.M.A.I.S., Syaerill, S., Zakaria, N.A., Asfi, M., Aini, Q.: AI-powered resume crafting and screening. International Journal on Perceptive and Cognitive Compu-ting 12(1), 125–130 (2026)

[3]. Kadhim, E.A., Feizi-Derakhshi, M.R., Aghdasi, H.S.: Advanced text summariza-tion model incorporating NLP techniques and feature-based scoring. IEEE Access 13, 19302–19319 (2025)

[4]. Aggarwal, A., Jain, S., Jha, S., Singh, V.P.: Resume screening. International Jour-nal for Research in Applied Science and Engineering Technology 134, 66–88 (2022)

[5]. Narendra, G.O., Hashwanth, S.: Named entity recognition-based resume parser and summarizer. International Journal of Advanced Research in Science, Communica-tion and Technology 2, 728–735 (2022). https://doi.org/10.48175/IJARSCT-3029

[6]. Mahato, M., Shah, S., Sheth, A., Doshi, P.: Content-based recommendation model for intelligent resume screening. Journal of Technical Education, 159

[7]. Deshmukh, A., Raut, A.: Enhanced resume screening for smart hiring using sen-tence-bidirectional encoder representations from transformers (SBERT). Interna-tional Journal of Advanced Computer Science & Applications 15(8) (2024)

[8]. Scientific, L.L.: AI-powered resume screening system for smart hiring: leveraging NLP and large language models for efficient and fair recruitment. Journal of Theo-retical and Applied Information Technology 103(23) (2025)

[9]. Pulavarthi, S., Reddy, B.R., Sairam, T., Bhattacherjee, A., Jallipalli, S.: Improving resume screening with NLP and machine learning: addressing efficiency and fair-ness. Grenze International Journal of Engineering & Technology (GIJET) 11 (2025)

[10]. Yadav, K.S., Matthew, B.H., Prakash, A.H., Harish, S., Hemanth, T., Hariharan, K.: AI-powered resume screening and job matching system for intelligent career guidance and recruitment optimization. IJSAT - International Journal on Science and Technology 16(4) (2025)

[11]. Fadzir, M.A.I.M., Hashim, S.: SmartResume: A resume parser application using natural language processing. Journal of Human Centered Technology 5(1), 15–20 (2026)

[12]. Anusha, S.B., Vismaya, V., Ali, K.A.W.T.: Resume screening and job matching system: a machine learning approach. Journal of Artificial Intelligence and Com-putational Technology 2(2) (2025)

[13]. Xie, J., Zhao, L.: Research on resume screening and career matching model based on fuzzy natural language processing. Journal of Computational Methods in Sci-ences and Engineering (2025)