Journal Publication of International Research for Engineering and Management (JOIREM)
Volume: 04 Issue: 06 | June-2026
ISSN (O) 3107-6696

SkillVector - An AI Based Resume Evaluation and Recruitment Management Platform For HR

Shreyash Kokane[1], Varsha Chavan[2], Rohan Patil[3], Nishant Raut[4], Om Bagle[5]

Department of Artificial Intelligence and Data Science, Anantrao Pawar College of Engineering and Research,Pune.

sherankanon@gmail.com, varsha.chavan@abmspcoerpune.org, rohanpatil1809@gmail.com, nishantraut225@gmail.com, bagleom0@gmail.com

**Correspondence: Shreyash Kokane – sherankanon@gmail.com*

---***---

Abstract - This paper introduces SkillVector, an AI-based platform designed to evaluate resumes and manage job descriptions more effectively. The system simplifies the screening process by automatically parsing resumes and comparing them with job descriptions using advanced NLP and machine learning methods. Through cosine similarity and BERT-based models, it generates a similarity score that ranks candidates based on how well they fit a given role. The backend uses PostgreSQL with the pgvector extension to store vectorized data and perform semantic similarity searches efficiently. A web-based interface allows recruiters to upload, review, and manage profiles while providing detailed analytics such as skill-gap insights, missing qualifications, and ranking visualization.

KEYWORDS: AI, Resume Evaluation, Job Recommendation, Natural Language Processing (NLP), Machine Learning, Cosine Similarity, PostgreSQL, Recruitment Automation.

1. INTRODUCTION

Hiring the right talent is crucial for any organization's success. Traditional recruitment processes are time-consuming, inconsistent, and prone to human bias. Recruiters spend hours manually reviewing resumes, relying on keyword searches and personal judgment, which frequently causes qualified candidates to be overlooked simply because their resumes do not match specific keyword patterns.

With the rapid growth of Artificial Intelligence (AI), Machine Learning (ML), and Natural Language Processing (NLP), recruitment practices are shifting toward greater automation and objectivity. These technologies enable systems to better interpret text, evaluate candidate profiles accurately, and match them to job descriptions based on data-driven analysis.

This paper presents SkillVector, an AI-powered recruitment management platform designed to simplify and improve the resume evaluation process. The system uses NLP to analyze resumes and job descriptions meaningfully, while ML models calculate similarity scores to determine how well a candidate fits a particular role. Unlike traditional keyword-based tools, SkillVector emphasizes meaning and context by using advanced Sentence-BERT embeddings, ensuring a more accurate and fair evaluation. The backend architecture combines PostgreSQL with the pgvector extension to store and retrieve high-dimensional embeddings, enabling scalable and fast semantic similarity searches. Recruiters interact with the system through a web-based interface offering dynamic dashboards, candidate rankings, and skill-gap analysis to support informed hiring decisions.

2. LITERATURE REVIEW

Recent AI and NLP advancements have transformed recruitment and resume evaluation into automated, data-centric methodologies. Numerous studies have explored strategies to improve the efficiency, accuracy, and fairness of hiring systems through machine learning and semantic analysis.

Ransing et al. [1] proposed a hybrid model integrating KNN, Linear SVC, and XGBoost for automated resume screening. Their approach improved classification accuracy and tackled challenges related to imbalanced data, establishing a foundation for ensemble AI solutions in recruitment.

Brindha et al. [2] embedded resume screening within a larger AI ecosystem that incorporated resume optimization, job matching, and recruiter analytics. Their NLP and ML implementation enhanced ATS compliance and enabled tailored job recommendations, significantly improving recruiter efficiency.

Fareed et al. [3] presented a hybrid technique merging KNN with cosine similarity for contextual classification. This strategy captured underlying semantic meanings of candidate profiles and job descriptions, yielding more accurate rankings compared to keyword-based approaches.

Journal Publication of International Research for Engineering and Management (JOIREM)
Volume: 04 Issue: 06 | June-2026
ISSN (O) 3107-6696

Abhishek et al. [5] introduced an AI-powered framework emphasizing clarity and bias reduction. Their research highlighted the importance of incorporating interpretability into automated screening to ensure transparent decision-making for recruiters.

Wailthare et al. [9] developed a ranking model based on cosine similarity, affirming it as a straightforward yet effective technique for evaluating text compatibility between resumes and job descriptions.

Ugale et al. [8] integrated unsupervised clustering with semantic matching to group similar resumes and enhance candidate filtering efficiency. Kinger et al. [10] demonstrated that combining YOLOv5 for layout detection with DistilBERT for semantic understanding surpasses traditional NLP models in capturing context.

Khan et al. [11] compared BERT, Gemini, and LLaMA 3.1 for resume-to-job-description matching, finding that recent transformer models offer superior semantic understanding and greater generalization capabilities and observation that supports SkillVector's use of Sentence-BERT embeddings.

3. METHODOLOGY

SkillVector follows a hybrid AI architecture encompassing data input, preprocessing, feature extraction, similarity computation, ML-based evaluation, and final ranking. The algorithm takes a resume document R_{doc} and job description J_{desc} as inputs, and outputs a compatibility score S, a skill-gap report G_r, and personalized recommendations R_{ec}.

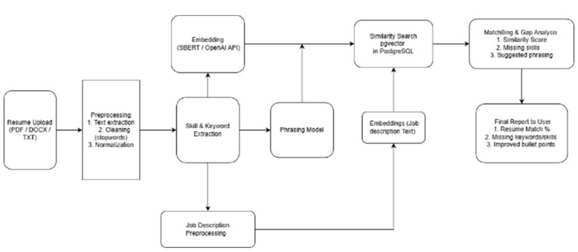

Figure 1: System Architecture Diagram

3.1 Text Extraction and Preprocessing

The process begins when a user submits a resume in PDF, DOCX, or TXT format. Text is extracted using

parsing libraries and then cleaned through tokenization, lemmatization, and stop-word removal to standardize the data for downstream analysis:

$$T_R \leftarrow ExtractText(R_{desc}),\ T_R \leftarrow Preprocess(T_R) \quad (3.1)$$

3.2 Skill Extraction

Named Entity Recognition (NER) and skill extraction modules identify candidate skills from the resume and required skills from the job description respectively:

$$S_R \leftarrow ExtractSkills(T_R),\ S_J \leftarrow ExtractSkills(J_{desc}) \quad (3.2)$$

3.3 Embedding Generation

Both documents are transformed into high-dimensional semantic vector representations using Sentence-BERT:

$$v_R \leftarrow Embed(T_R),\ v_J \leftarrow Embed(J_{desc}) \quad (3.3)$$

3.4 Similarity Computation

Cosine similarity measures the contextual alignment between the resume and job description embeddings, ignoring magnitude differences and focusing on directional similarity:

$$SkillSim = \frac{\{v_R \cdot v_J\}}{||v_R||\,||v_J||} \quad (3.4)$$

3.5 Keyword Match Ratio

The keyword overlaps ratio measures what fraction of the required job skills appear in the resume:

$$KeywordScore = \frac{\{|S_R \cap S_J|\}}{\{|S_J|\}} \quad (3.5)$$

3.6 Experience Match Scoring

An experience match score is computed based on the candidate's years of experience and job roles, normalized between 0 and 1:

$$ExpScore = MatchExperience(T_R, J_{desc}) \quad (3.6)$$

3.7 Compatibility Score Calculation

The final compatibility score is a weighted combination of the three sub-scores. Weights α, β, and γ are adjusted according to the relative importance of skill similarity, keyword matching, and experience for a given job role:

$$S = \alpha \times SkillSim + \beta \times KeywordScore + \gamma \times ExpScore \quad (3.7)$$

Subject to: $\alpha + \beta + \gamma = 1,\ \alpha, \beta, \gamma \in [0,1]$ (3.8)

3.8 Skill-Gap and Recommendation Generation

The skill gap is defined as the set of required skills absent from the resume. Based on this, the recommendation module generates personalized improvement suggestions:

$$G_r \leftarrow S_J \setminus S_R \quad (3.9)$$

$$R_{\{ec\}} \leftarrow Recommend(G_r, S) \quad (3.10)$$

Journal Publication of International Research for Engineering and Management (JOIREM)
Volume: 04 Issue: 06 | June-2026
ISSN (O) 3107-6696

The complete pipeline connects job requirements with candidate profiles through contextual language understanding and semantic similarity matching, ensuring accurate and fair evaluation.

Figure 2: AI Processing Pipeline

4. RESULT AND ANALYSIS

SkillVector's performance was evaluated across three dimensions: accuracy of information extraction, quality of candidate ranking, and computational efficiency. Results are summarized in Table 1, and the following subsections detail each metric, its formula, and the achieved outcome.

Table 1: Evaluation Matrix

Category	Metric	Definition	Target	Achieved
1. Accuracy	Precision (Overall)	% of extracted fields correct	>90%	92%
	Recall (Overall)	% of relevant items retrieved	>85%	87%
	Skill Extraction F1	Harmonic mean of Precision/Recall	>0.80	0.85
	Contact Info Accuracy	Phone/Email extraction accuracy	100%	97%
2. Ranking	NDCG@10	Ranking quality vs. human judgment	>0.85	0.88
	Precision@5	% relevant candidates in top 5	>80%	80%
	Human Agr. (SrS)	Spearman correlation with human scores	>0.70	0.76
3. Performance	OCR & Preprocessing	Text parse and clean time	<2.0s	0.01s
	Vector Gen. Latency	Embedding generation time	<1.0s	0.40s
	Skill Extraction	SpaCy NLP runtime (warm)	<0.5s	0.15s
	Scoring Logic	Math computation time	<0.05s	0.006s
	Total Latency	End-to-end per resume	<5.0s	0.60s

4.1 Accuracy Evaluation

Three metrics measured the quality of information extraction from resumes.

4.1.1 Precision, Recall, and F1-Score

Precision measures the proportion of correctly identified elements among all elements identified by the system. Recall measures the system's ability to retrieve all relevant information. The F1-score combines both into a single balanced metric:

$$Precision = TP / (TP + FP) \quad (4.1)$$

$$Recall = TP / (TP + FN) \quad (4.2)$$

$$F1 = 2 \times (Precision \times Recall) / (Precision + Recall) \quad (4.3)$$

The system achieved 92% precision, 87% recall, and an F1-score of 0.85 for skill extraction – all exceeding the defined targets. Contact information accuracy reached 97%, demonstrating robust parsing of structured fields like phone numbers and email addresses.

4.2 Ranking Evaluation

4.2.1 Normalized Discounted Cumulative Gain (NDCG@10):

NDCG measures the quality of the ranking order produced by the system compared to an ideal human-generated ranking, giving higher importance to candidates near the top of the list:

$$NDCG@K = \frac{DCG@K}{IDCG@K} \quad (4.4)$$

where rel_i is the relevance of the i^{th} candidate and IDCG@K is the ideal DCG from perfect human ranking. The system achieved NDCG@10 of 0.88, exceeding the 0.85 target.

$$\mathrm{DCG@K} = \sum_{i=1}^{K} \frac{\mathrm{rel_i}}{\log_2(i+1)} \quad (4.5)$$

4.2.2 Precision@5 and Human Agreement Score:

Precision@5 evaluates relevant candidates among the top five system results:

$$Precision@5 = \frac{Number\ of\ relevant\ candidates\ in\ the\ top\ 5}{5} \quad (4.6)$$

The Human Agreement Score uses Spearman Rank Correlation to compare system rankings against human recruiter evaluations:

$$\rho = 1 - \frac{6 \sum di^2}{(n\,(n^2 - 1))} \quad (4.7)$$

where d_i is the rank difference between system and human judgments and n is the number of candidates. The achieved SrS of 0.76 confirms strong alignment with human recruiter decision-making.

Journal Publication of International Research for Engineering and Management (JOIREM)
Volume: 04 Issue: 06 | June-2026
ISSN (O) 3107-6696

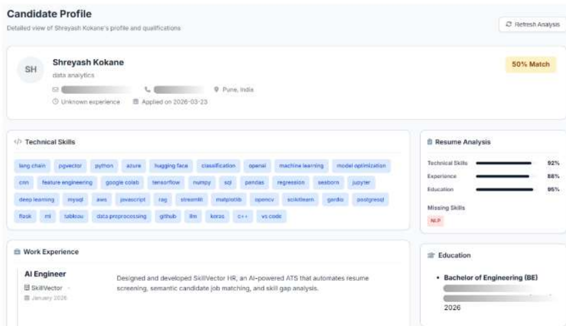

Figure 3: Candidate Result Report

4.3 System Performance

SkillVector operates at sub-second latency across all pipeline stages. The preprocessing is completed in 0.01 s, embedding generation in 0.40 s using Sentence-BERT, skill extraction via SpaCy in 0.15 s, and scoring logic computation in 0.006s. The total end-to-end latency of 0.60 s is far below the 5-second target, confirming that the platform can support real-time candidate evaluation at scale. These results validate the efficiency of the lightweight preprocessing pipeline and the effective utilization of the pgvector database extension for similarity operations.

The experimental results demonstrate that SkillVector performs strongly in all evaluation areas. High precision, good recall, and a balanced F1-score reflect the reliability of its semantic similarity and skill extraction mechanisms, while strong NDCG and SrS scores confirm that the system's ranking logic closely mirrors human recruiter judgment.

5. FUTURE SCOPE

SkillVector can be expanded with real-time analytical dashboards for tracking candidate pools, reviewing hiring trends, and visualizing recruitment performance. Predictive analytics could anticipate future talent demands and identify potential skill shortages within the workforce. Intelligent automation modules including AI-driven interview scheduling, automatic shortliting, and adaptive ranking mechanisms would further streamline the recruitment process and reduce manual workload.

For job seekers, the platform can be enhanced to deliver personalized skill-gap feedback, suggest relevant training and upskilling opportunities, and recommend job openings aligned with evolving skill profiles. Incorporating continuous learning algorithms would allow SkillVector to adapt to changing job market trends, improving its long-term accuracy and reliability. Collectively, these developments would transform SkillVector into a comprehensive, intelligent recruitment ecosystem bridging the gap between candidates and employers through automation, transparency, and data-driven insights.

6. CONCLUSIONS

SkillVector demonstrates the practical integration of artificial intelligence in modern recruitment. By leveraging Sentence-BERT embeddings, ML-based compatibility scoring, and pgvector powered similarity searches, the platform delivers consistent, accurate, and bias-reduced candidate evaluations. The hybrid architecture combining cosine similarity, keyword matching, and experience scoring provides recruiters with an objective and interpretable measure of candidate-job compatibility. Sub-second end-to-end processing latency confirms real-world scalability, while a Human Agreement Score of 0.76 validates that the system's ranking logic closely mirrors human recruiter judgment.

SkillVector offers interactive visualization tools and in-depth analytics enabling recruiters to identify top candidates and skill gaps efficiently. For job seekers, it provides actionable feedback on skill improvement and better visibility of role compatibility. In advancing toward a smarter, fairer, and more transparent digital hiring environment, SkillVector represents a significant step in transforming traditional recruitment workflows into objective, data-driven processes.

REFERENCES

[1] R. Ransing, A. Mohan, N. B. Emberi, and K. Mahavarkar, "Screening and Ranking Resumes using Stacked Model," in *Proc. ICEECCOT*, IEEE, 2021. DOI:10.1109/ICEECCOT52851.2021.9707977

[2] M. Brindha, M. Karthikeyan, T. Jessinth, A. S. Karthe, and A. Dhanush, "AI-Powered Resume Optimizer and Job Recommendation System," *JETIR*, Apr. 2025, ISSN: 2349-5162.

[3] R. T. Fareed, V. Rajath, and S. Kaganurmath, "Resume Classification and Ranking using KNN and Cosine Similarity," *IJERT*, vol. 10, no. 8, Aug. 2021.

[4] A. Baghbanzadeh, "Job-Resume Compatibility Scoring Using Graph Neural Networks and Large Language Models," M.S. thesis, Univ. of Windsor, Canada, 2025.

[5] K. L. Abhishek et al., "Developing an Intelligent Resume Screening Tool With AI-Driven Analysis and Recommendation Features," 2025. DOI: 10.1002/ail2.116

[6] S. Kulkarni et al., "AI-powered Resume Screening System," *World J. Adv. Eng. Technol. Sci.*, vol. 15, no. 1, pp. 2263–2277, 2025.

[7] V. Jagwani, S. Meghani, K. Pai, and S. Dhage, "Resume Evaluation through LDA and NLP for Effective Candidate Selection," *Sardar Patel Institute of Technology*, Mumbai.

[8] A. V. Ugale et al., “Resume Clustering and Job Description Matching,” Apr. 2025.
[9] K. Wailthare, A. Tamhane, V. Mulik, and K. Suryawanshi, “A Cosine Similarity-Based Resume Screening System for Job Recruitment,” Apr. 2023. DOI: 10.56726/IRJMETS35945
[10] S. Kinger, D. Kinger, S. Thakkar, and D. Bhake, “Towards Smarter Hiring: Resume Parsing and Ranking with YOLOv5 and DistilBERT,” *Multimedia Tools Appl.*, 2024. DOI: 10.1007/s11042-024-18778-9
[11] S. Khan, A. Shahi, A. Alam, A. Bafna, and S. Therese, “Comparison of Models for Resume-JD Matching: BERT, Gemini, and Llama 3.1,” 2025. DOI: 10.9790/0661-2702050110
[12] A. K. Jirjees et al., “Machine Learning for Recruitment: Analyzing Job Matching Algorithms,” 2023.
[13] A. Agarwal and Senthilkumar, “Resume Recommendation System Using Cosine Similarity,” *SRMIST*, Chennai, 2022.
[14] M. Mehboob et al., “Evaluating Automatic CV Shortlisting Tool for Job Recruitment Based on ML,” *MAJICC*, IEEE, 2022.
[15] T. T. Q. Trinh, Y. C. Chung, and R. J. Kuo, “A Domain Adaptation Approach for Resume Classification Using Graph Attention Networks,” Feb. 2023.