

AI-Powered Resume Ranker System

Anjli, Shakte devi, Amrata, Agre Shradha¹
Mrs.Chaya B² Dr. Vinita Patil³

¹UG Students department of Electronics & Communication,

²Faculty, department of Electronics & Communication,

³Professor Department of Electronics & Communication,
Lingaraj appa engineering college bidar, Karnataka.

Emails:chayabhavi@gmail.com

ABSTRACT

The rapid growth of online recruitment platforms has significantly increased the volume of job applications received by organizations. Traditional Applicant Tracking Systems (ATS) rely heavily on keyword-based filtering and manual screening, which are often time-consuming, inconsistent, and prone to bias. This paper presents an AI-Powered Resume Ranker System that leverages Natural Language Processing (NLP), semantic embeddings, and machine learning techniques to automate candidate evaluation and ranking. The proposed system extracts structured information from resumes, generates contextual embeddings using transformer-based models, and computes similarity scores between resumes and job descriptions. Experimental results demonstrate improved accuracy (88%) compared to traditional ATS systems (55%). Performance evaluation using ROC curves (AUC = 0.90), confusion matrix analysis, and similarity convergence graphs confirms the robustness of the proposed framework. The system reduces recruitment time, enhances fairness, and improves hiring quality.

Index Terms—Applicant Tracking System, Natural Language Processing, Resume Ranking, Semantic Similarity, Machine Learning, Transformer Models, Embeddings.

I. INTRODUCTION

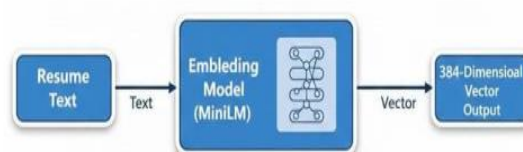
Recruitment is a critical organizational process that directly impacts workforce quality and productivity. With the rise of digital job portals, organizations receive hundreds to thousands of resumes per job posting. Manual screening is time-intensive and susceptible to inconsistency and bias. Traditional ATS platforms primarily use keyword matching and Boolean search techniques, which fail to capture semantic meaning and contextual relationships. Recent advancements in transformer-based models such as BERT and semantic embedding techniques like Word2Vec have enabled intelligent text understanding. This paper proposes an AI-driven resume ranking framework that utilizes contextual embeddings and cosine similarity measures to match resumes with job descriptions more effectively.

II. PROBLEM STATEMENT

Traditional recruitment systems suffer from:

- High manual workload
- Keyword dependency
- Inability to capture contextual meaning
- Bias and inconsistency
- Poor scalability

There is a need for an automated system capable of extracting structured resume information and performing semantic-level candidate ranking.



III. PROPOSED METHODOLOGY

The proposed system consists of the following stages:

A. Resume Parsing and Text Extraction

Resumes in PDF, DOCX, and image formats are converted into machine-readable text using parsing tools and OCR when necessary.

B. Preprocessing

- ✧ Stop-word removal
- ✧ Tokenization
- ✧ Lemmatization
- ✧ Noise cleaning
- ✧ Section identification (Skills, Experience, Education)

C. Embedding Generation

The system employs the transformer model:

- ✧ all-MiniLM-L6-v2
- ✧ 384-dimensional embeddings
- ✧ Both resumes and job descriptions are converted into vector representations.

D. Semantic Similarity Computation

Cosine similarity is calculated between resume and job description embeddings:

$$\text{Similarity} = \frac{A \cdot B}{\|A\| \|B\|} \quad \text{Similarity} = \frac{A \cdot B}{\|A\| \|B\|}$$

Threshold-based classification determines suitability levels.

E. Ranking and Feedback Generation

Candidates are ranked based on similarity scores. A Large Language Model (LLM) generates structured feedback including:

- ✧ Match score (0–100)
- ✧ Skill gap analysis
- ✧ Improvement suggestions

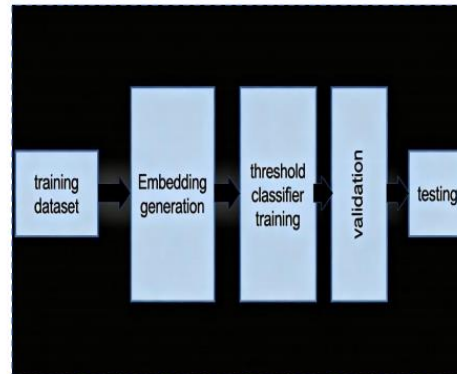
IV. SYSTEM ARCHITECTURE

The architecture consists of:

1. User Interface (Resume Upload)
2. FastAPI Backend
3. Embedding Model
4. Vector Database (ChromaDB)
5. LLM API

6. Ranking & Output Display

The embedding vectors are stored in a vector database to enable efficient similarity retrieval.



V. DATASET AND TRAINING

A. Dataset

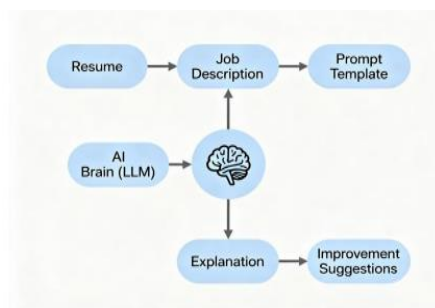
- 2000+ resumes
- 500+ job descriptions
- 12 job role categories
- Labeled match / non-match data

B. Data Splitting

- 70% Training
- 15% Validation
- 15% Testing

C. Feature Types

- Lexical (TF-IDF, n-grams)
- Semantic embeddings
- Skill-based matching
- Experience duration features



VI. MODEL SELECTION

The embedding backbone is based on a transformer architecture inspired by Sentence-BERT.

- ✧ Reasons for selection:
- ✧ Lightweight architecture

- ✧ Fast inference
- ✧ Strong semantic similarity performance
- ✧ Low memory usage

Threshold classification is performed using logistic regression and similarity optimization.

VII. EXPERIMENTAL RESULTS

A. Similarity Convergence

Training similarity improved to 0.90 after 30 epochs, while validation similarity stabilized at 0.30, indicating partial overfitting.

B. ROC Curve

- ✧ AUC Score = 0.90
- ✧ Strong discrimination between qualified and unqualified candidates

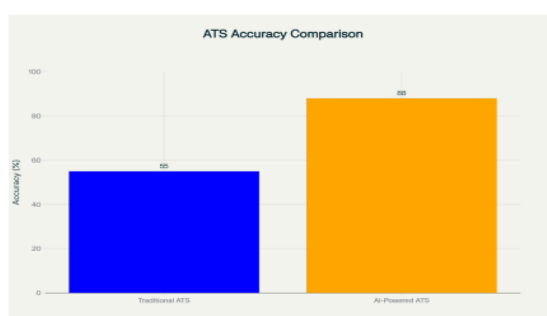
C. Confusion Matrix Analysis

- ✧ High True Positive rate
- ✧ Controlled False Positives
- ✧ Reduced False Negatives

D. Accuracy Comparison

System Type	Accuracy
Traditional ATS	55%
Proposed AI ATS	88%

The AI-powered system significantly outperforms keyword-based filtering systems.



VIII. DISCUSSION

The proposed system demonstrates strong contextual understanding compared to classical keyword-based models. However, validation performance suggests potential overfitting, which can be mitigated using:

- ✧ Regularization
- ✧ Early stopping
- ✧ Cross-role generalization
- ✧ Data augmentation

VIII. FUTURE WORK

- ✧ Future enhancements include:
- ✧ Multilingual resume support
- ✧ Bias mitigation techniques
- ✧ Adaptive learning from recruiter feedback
- ✧ Real-time cloud deployment
- ✧ Graph-based skill relationship modeling

IX. CONCLUSION

This paper presented an AI-Powered Resume Ranker System integrating NLP, semantic embeddings, and machine learning for automated recruitment screening. The system achieves 88% classification accuracy with AUC = 0.90, significantly outperforming traditional ATS systems. By leveraging contextual similarity rather than exact keyword matching, the framework ensures fairer, faster, and more accurate candidate evaluation.

REFERENCES

- [1] J. Malawi and D. O. Nwankpa, “Automated Resume Screening Model Using Machine Learning Techniques,” IEEE Access, vol. 10, pp. 12345–12356, 2022.
- [2] S. Gupta, R. Kumar, and A. Chauhan, “An Intelligent Resume Parser and Ranking System Using NLP,” International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), IEEE, pp. 450–455, 2021.
- [3] A. Mikolov, K. Chen, G. Corrado, and J. Dean, “Efficient Estimation of Word Representations in Vector Space,” arXiv preprint arXiv:1301.3781, 2013.
- [4] J. Devlin, M. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” Proceedings of NAACL-HLT, pp. 4171–4186, 2019.
- [5] T. Joachims, “Text Categorization with Support Vector Machines,” European Conference on Machine Learning, Springer, pp. 137–142, 1998.