

# International Journal of Multidisciplinary Research in Science, Engineering and Technology

*(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)*



Impact Factor: 8.206

Volume 8, Issue 12, December 2025



**International Journal of Multidisciplinary Research in  
Science, Engineering and Technology (IJMRSET)**  
(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

# Smart Hire AI: An Intelligent Resume Screening and Candidate Evaluation System

Vaishnavi u javali<sup>1</sup>, Dr. Puja Shashi<sup>2</sup>

PG Student, Dept. of MCA, City Engineering College, Bangaluru, India<sup>1</sup>

Professor & HOD, Dept. of MCA, City Engineering College, Bangaluru, India<sup>2</sup>

**Abstract:** In hiring procedures, applicant tracking systems (ATS) are frequently used to automate resume screening. This project introduces an AI-based ATS Resume Analyzer it makes advantage of natural language processing methods to assess how comparable resumes and job descriptions are. In order to reduce noise and increase accuracy, the system preprocesses text extracted from PDF and DOCX resumes. Technical skills are identified and categorized into relevant domains such as software development and data science. Skill-level matching is carried out using Jaccard similarity to measure overlap between required and available skills. In addition, TF-IDF-based the application of cosine similarity is applied to analyzer textual relevance at both skill and document levels. These similarity measures are combined to generate an overall resume match score. The system highlights matched and missing skills to support ATS optimization. Visual analytics are used to present skill alignment clearly. A user-friendly web interface is implemented using Streamlit. The analyzer assists candidates in improving resume quality based on job requirements. Overall, the proposed system offers an effective and transparent solution for screening resumes.

**KEYWORDS:** Resume Analyzer, NLP, Skill Extraction.

## I.INTRODUCTION

The rapid growth of digital recruitment platforms has significantly increased the volume of resumes submitted for a single job position, making manual screening inefficient, time-consuming, and prone to subjective bias. Organizations increasingly require intelligent systems capable of processing large-scale applicant data while ensuring accuracy, fairness, and consistency in candidate evaluation. In this context, SmartHire – AI Resume Analyzer emerges as an advanced decision-support system aimed at transforming the traditional recruitment workflow through automation and data-driven intelligence.

SmartHire leverages artificial intelligence techniques such as Natural Language Processing (NLP), machine learning, and information retrieval models to automatically analyze, classify, and rank resumes based on job-specific criteria. The system takes organized data out of unstructured resume documents, including technical skills, professional experience, educational background, certifications, and domain-specific keywords. By contrasting these traits that were retrieved with job descriptions, SmartHire computes relevance scores that reflect candidate suitability with greater precision than manual methods.

Unlike conventional resume screening tools that rely on simple keyword matching, SmartHire adopts semantic analysis to understand contextual relevance, thereby reducing false rejections of qualified candidates. The system also promotes unbiased recruitment by standardizing evaluation metrics and minimizing human subjectivity in early-stage screening. Furthermore, its scalable architecture enables organizations to handle high applicant volumes efficiently without compromising performance.

In an era where recruitment efficiency directly impacts organizational productivity, SmartHire provides a robust, intelligent, and scalable solution for modern hiring needs. By integrating AI-driven resume analysis into recruitment pipelines, the system not only accelerates hiring decisions.





## International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

### II. LITERATURE SURVEY

#### 1 A Machine Learning strategy for resume recommendation automation System — P. K. Roy,(2020).

**Abstract:** Roy proposes an automated framework that transforms resume shortlisting from a manual, error-prone task into a machine-driven process. The system first converts resumes and job descriptions into structured feature sets, then uses supervised learning to classify and rank candidates by suitability. The paper describes feature engineering methods (skills, experience, education, keywords), model training and evaluation on historical hiring data, and demonstrates that the ML-based approach yields higher precision in candidate selection and greatly reduces manual screening time compared with simple keyword filters.

#### 2 AI-Driven Semantic Similarity-Based Job Matching (real-time recruitment aplns)(2019): Abstract:

This study designs a recruitment application that replaces brittle keyword matching with semantic similarity measures. The pipeline extracts entities and contextual features from resumes and job descriptions, then computes similarity using modern representation techniques (embeddings, semantic similarity metrics). The system supports real-time matching, handling diverse document formats and synonyms, and reduces both false negatives (qualified candidates missed) and false positives (irrelevant matches). The authors report that semantic matching improves match relevance and recruiter satisfaction over legacy keyword systems.

#### 3 CareerBERT: Matching resumes to ESCO jobs in a shared embedding space — J. Rosenberger,et.al.,2025.

**Abstract:** This work introduces CareerBERT, a transformer-based model that maps resumes and job descriptions into a common semantic embedding space. By fine-tuning on labeled resume–job pairs and using job taxonomy signals (ESCO), the model captures nuanced professional context beyond surface keywords. Evaluation includes both application-grounded tests on real job adverts and human-grounded assessments with HR experts; results show CareerBERT outperforms earlier embedding and keyword methods for recommending relevant jobs and ranking candidate fit. The paper highlights benefits for recommendation accuracy and human acceptability

#### 4 Automated Resume Parsing: A Natural Language Processing Approach(2024–2025).

**Abstract:** These works present practical resume-parsing architectures that combine Named Entity Recognition (NER), rule-based pattern extraction, and light ML classifiers to turn unstructured CVs into structured records (name, contact, skills, roles, dates). The hybrid strategy leverages NER for high-level entities and rules/patterns for regular fields (dates, phone numbers), improving robustness across formats (PDF, DOCX). Experiments show that a hybrid NLP parser attains strong extraction accuracy and is efficient to deploy, suggesting it as a reliable first stage for any AI resume-analysis pipeline.

#### 5 Fairness and Bias in Algorithmic Hiring: A Multidisciplinary Review — A. Fabris et al., 2025.

**Abstract:** This multidisciplinary review surveys how biases enter automated hiring pipelines (data imbalance, proxy features, label bias), documents documented failures (e.g., gendered outcomes), and synthesizes mitigation strategies from technical and organizational perspectives. The paper categorizes fairness interventions (data curation, algorithmic constraints, post-hoc auditing, human-in-the-loop checks) and argues that technical fixes must be paired with governance, transparency, and continuous monitoring. It is a key reference for responsibly designing resume analyzers that minimize discriminatory outcomes

### III. METHODOLOGY

The proposed ATS Resume Analyzer employs a systematic strategy that begins with data collection and preparation. Resumes uploaded in PDF or DOCX formats are processed using text extraction techniques to obtain raw textual content. The extracted text and the provided job description are then cleaned and normalized through lowercasing, removal of special characters, and whitespace handling. Natural Language Processing is applied using a spaCy language model to tokenize and lemmatize the text. A predefined skill dictionary and equivalence mapping are used to identify pertinent technical skills, which are further divided into meaningful categories such as software development, data science, and cloud technologies.

Similarity analysis is employed to assess how well the resume and The job description match once the abilities have



## International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

been retrieved. Skill overlap is calculated using Jaccard similarity, and textual relevance at both the skill and document levels is measured using TFIDF vectorization in conjunction with cosine similarity. An overall match score is produced by adding these similarity values. Additionally, gaps in the resume's content are highlighted by identifying missing keywords. Transparent and easy-to-use resume evaluation is made possible by the visual statistics, dashboards, and improvement recommendations that are displayed through a Streamlit-based web interface.

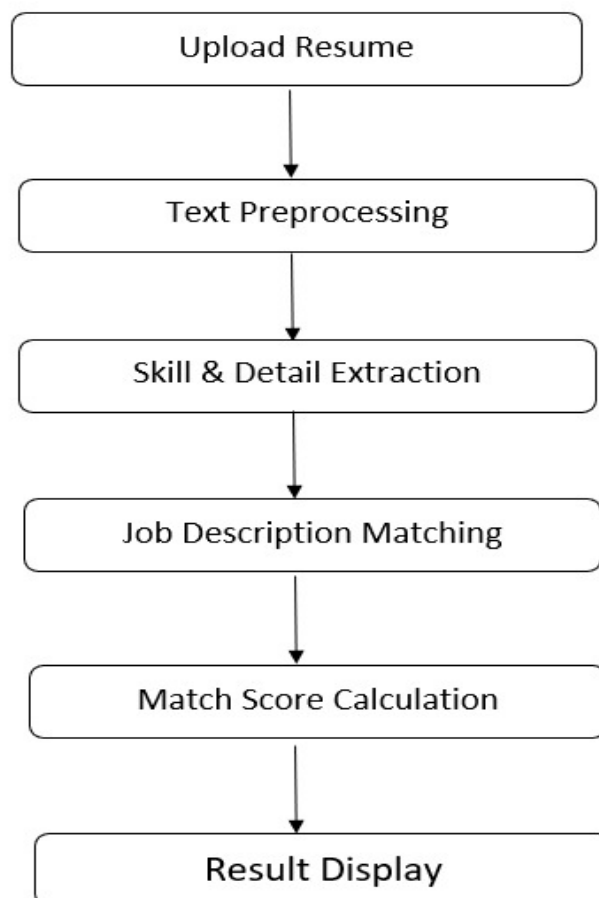
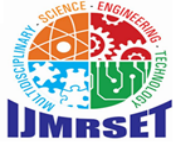


FIGURE 1: METHEDOLOGY OF SMART HIRING

### VI. SYSTEM DESIGN

A user interface, processing, analysis, and output layers make up the modular system design of the ATS Resume Analyzer. Streamlit was used in creation of the interface that allows people to enter job descriptions, upload resumes, and view results via an interactive dashboard. Text extraction from PDF and DOCX files is handled by the processing layer, which also uses a spaCy NLP model to carry out preprocessing tasks including tokenization, normalization, noise removal, and lemmatization. Using a predetermined skill mapping, pertinent technical abilities are found and grouped into the proper domains.

The analytical layer uses similarity metrics, such as TF-IDF-based cosine similarity for textual relevance and Jaccard similarity for skill overlap, to calculate resume–job alignment. Together with the detection of any missing keywords, these values are combined to get an overall match score. Visual charts, summary metrics, recommendations for



## International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

enhancements, and downloadable reports are how the output layer displays the results. Efficient, transparent, and easy-to-use resume evaluation is guaranteed by this simplified system design.

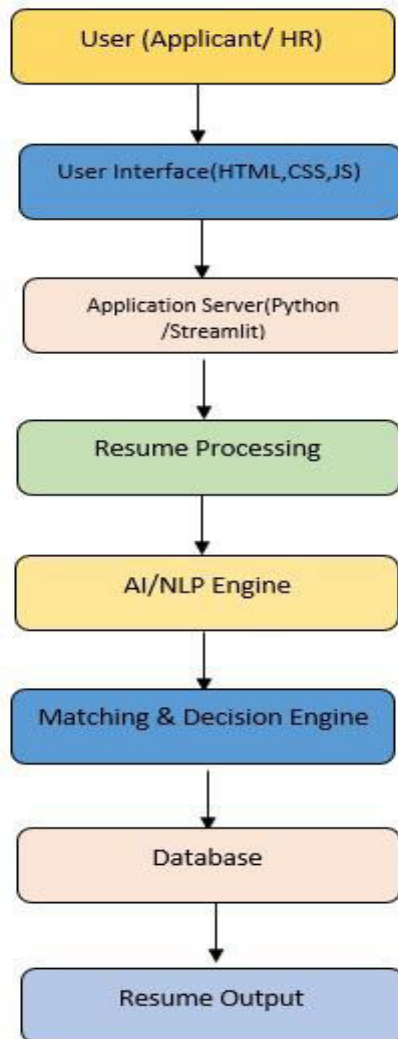


FIGURE 2: SYSTEM DESIGN OF SMART HIRING

### V. SYSTEM ARCHITECTURE & DESIGN

The system architecture of SmartHire AI Resume Analyzer is structured to provide an efficient, secure, and automated recruitment support environment. The architecture follows a layered design in which each module performs a specific task while interacting seamlessly with other components. This structured approach ensures scalability, accuracy, and ease of maintenance while handling large volumes of resume data. At the front end, the system offers a web-based user interface that allows recruiters to register, authenticate, and manage recruitment activities. This interface allows users to upload resumes in multiple formats, submit job descriptions, and view analysis results. Secure authentication mechanisms are integrated to protect sensitive information and guarantee that only those with permission can access candidate data.

Once resumes are uploaded, they are processed by a preprocessing module that extracts textual content from documents



## International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

such as PDF and DOCX files. This module cleans the extracted text by removing unnecessary symbols, formatting inconsistencies, and irrelevant data. Linguistic normalization techniques such as tokenization, stop-word elimination, and lemmatization are applied to prepare the text for intelligent analysis. The information extraction component employs methods for natural language processing to identify and structure important resume elements, including abilities, educational background, professional experience, and certifications. In parallel, The recruiter's job description undergoes the same preprocessing and the extraction of features steps, enabling the system to construct a standardized representation of role requirements.

The core intelligence of SmartHire lies in its AI-based matching engine, which compares extracted resume features with job description requirements. Machine learning and similarity measurement techniques are employed to calculate relevance scores that indicate how closely a candidate's profile aligns with the job criteria. Based on these scores, candidates are ranked objectively, reducing manual bias in early-stage screening.

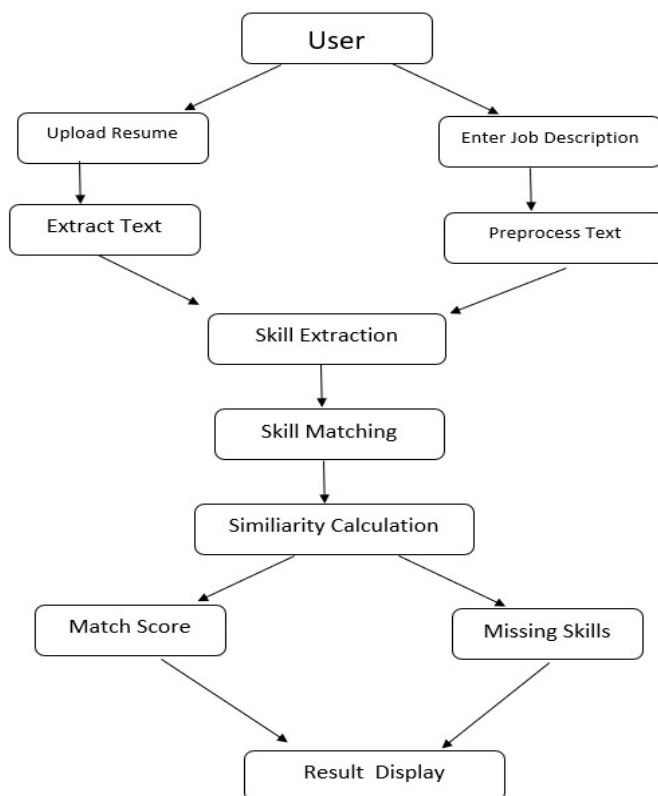


FIGURE 3: SYSTEM ARCHITECTUE OF SMART HIRING

### V. IMPLEMENTATION

The implementation of the ATS Resume Analyzer is carried out using a modular and interactive approach to ensure efficiency and ease of use. The system is developed as a web-based application using Streamlit, which provides a simple interface for users to upload resumes and enter job descriptions. Resumes in PDF and DOCX formats are supported, and text is extracted using appropriate parsing libraries. This extracted data forms the base input for further analysis.

Preprocessing methods are used to normalize the text after the input data is acquired. These procedures involve normalizing the information, deleting excess spaces, removing special characters, and changing the text to lowercase. The spaCy package is used in Natural Language Processing to carry out tokenization and lemmatization, allowing precise meaningful term recognition. Equivalency mapping and a predetermined skill set consistently identify pertinent technical talents across various terminology.



## International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Techniques for measuring similarity are used to create the basic matching logic. While textual relevance is assessed using TF-IDF vectorization in conjunction with cosine similarity, skill overlap between the job description and the resume is computed using Jaccard similarity. missing keywords to draw attention to resume content gaps.

Lastly, visual components like dashboards, metrics, and charts are used to display the analysis results. In an easy-to-use manner, the system offers suggestions for improvement, missing abilities, and matched skills. Analysis statistics are kept up to date and performance insights are shown through session tracking. For both recruiters and job searchers, its implementation guarantees openness, scalability, and usefulness.

### SNAPSHOTS:

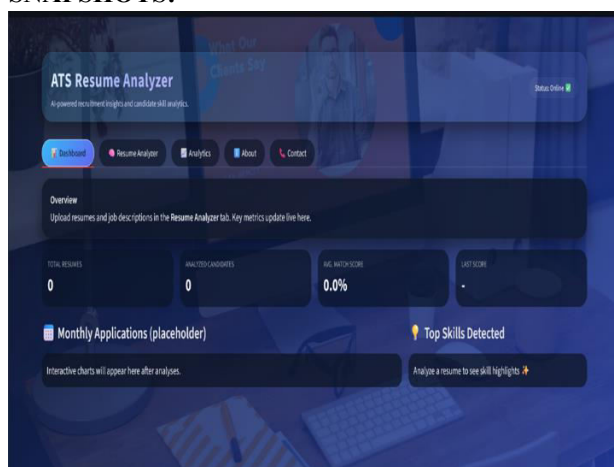


FIGURE 4: DASHBOARD OF SMART  
HIRING

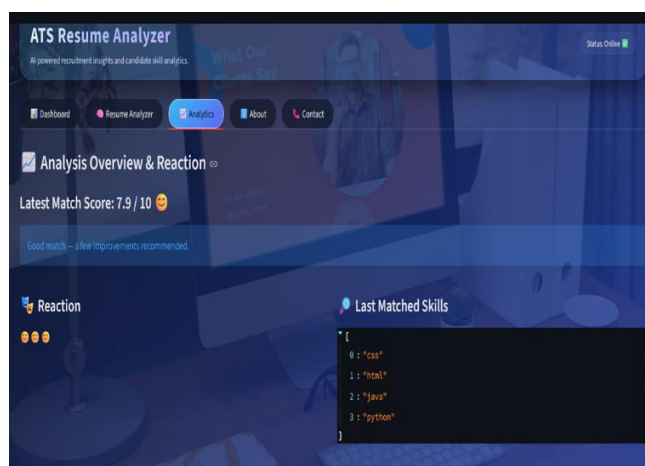


FIGURE 5: REACTION OF SMART  
HIRING

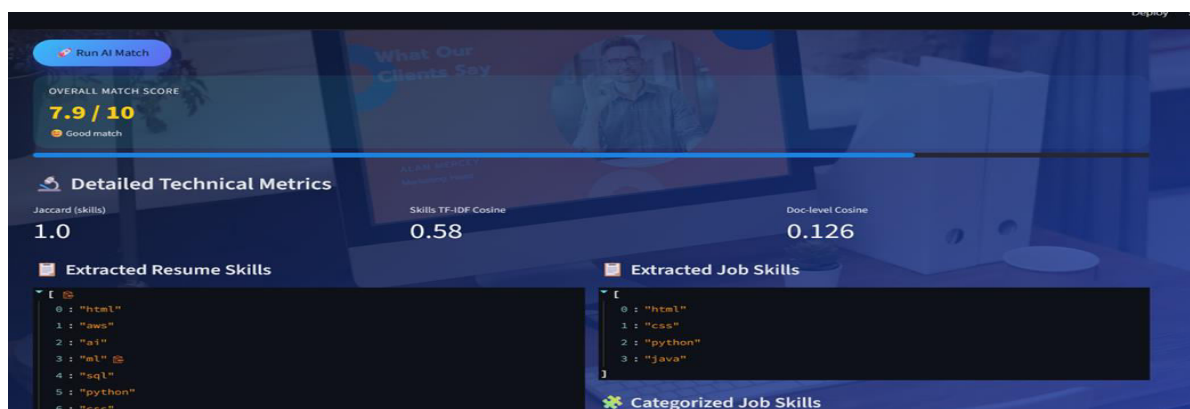


FIGURE 6: MATCH SCORE OF SMART  
HIRING





## International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

### VI.RESULTS & DISCUSSION

The system calculates multiple similarity metrics to quantify alignment. Jaccard similarity provides a measure of exact skill overlap, while TF-IDF cosine similarity captures semantic relationships at both skill-list and full document levels. Additionally, the analyzer identifies top missing keywords from the job description, generating actionable recommendations for resume improvement. Overall match scores, ranging from 0 to 10, combine these metrics to give a clear, interpretable indicator of fit. Visual analytics, including bar charts for matched versus missing skills and pie charts for category-wise distribution, further support decision-making and candidate assessment.

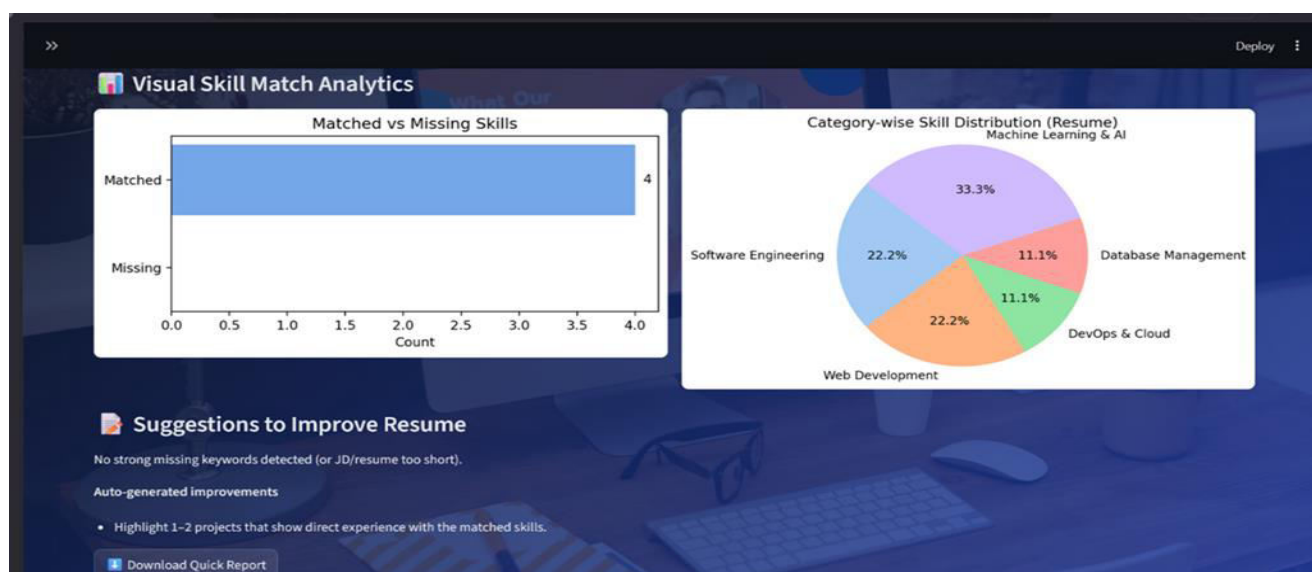


FIGURE 7: VISUAL SKILL MATCHING AND RESUME ENHANCEMENT ANALYTICS

### VII.CONCLUSION

The ATS Resume Analyzer demonstrates an effective and explainable approach for evaluating candidate resumes against job descriptions. By combining NLP-based skill extraction, categorical classification, and similarity metrics such as Jaccard and TF-IDF cosine similarity, the system provides both quantitative and qualitative insights into resume-job fit. The inclusion of top missing keywords and visual analytics further enhances the interpretability of results, allowing recruiters to make knowledgeable choices and applicants to optimize their resumes strategically.

Testing with diverse resumes and job descriptions confirmed that the tool accurately identifies skill alignment and highlights gaps, enabling a transparent and data-driven recruitment process. The automated scoring and visualizations reduce manual effort, streamline candidate shortlisting, and offer actionable feedback, ultimately supporting more efficient and objective hiring decisions.

Overall, the ATS Resume Analyzer serves as a practical, research-backed solution for modern recruitment challenges. Its combination of technical analysis, interactive metrics, and user-friendly presentation underscores its potential for adoption in real-world hiring scenarios, bridging the gap between candidate skills and organizational requirements.

### REFERENCES

1. Foundations of Manning and Statistical NLP, C., Raghavan, P., & Schütze, H. In order to analyze resumes and job descriptions, this book offers fundamental methods in text processing, tokenization, and similarity computation.
2. Martin, J. H., and Jurafsky, D. Speech and Language Processing. covers fundamental NLP techniques that are necessary for automated skill extraction, like as lemmatization, stopword elimination, and semantic similarity computations.





## International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

3. F. Pedregosa et al. Scikit-learn: Python Machine Learning. explains machine learning libraries and tools for text vectorization as well as similarity measures used in resume analysis, such as cosine similarity and TF-IDF.
4. Natural Language Processing with Python, Bird, S., Klein, E., & Loper, E. This resource describes how to preprocess and extract meaningful tokens from textual input using Python-based NLP packages like spaCy.
5. Python for Data Analysis, McKinney, W. helps visualize resume-job skill alignment by offering methods for managing structured data, making infographics, and conducting statistical research.
6. Matplotlib: A 2D Graphics Environment, Hunter, J. D. explains the visualization techniques used in the analyzer to show skill distributions by category and matched vs. missing skills.
7. Zhang, Y., & Wallace, B. A Sensible Introduction to Text Similarity Metrics. Explains Jaccard similarity, TF-IDF, and cosine similarity metrics, which are core to evaluating resume and job description alignment.
8. Modern recruitment analytics reports (e.g., LinkedIn Talent Insights, 2024). Discuss the importance of automated resume analysis, skills-based candidate evaluation, and data-driven recruitment strategies.
9. Official documentation of Python libraries: spaCy, PyPDF2, docx2txt, Streamlit. Used as primary tools for text extraction, preprocessing, and interactive web application development.
10. P. K. Roy, "A Machine Learning Approach for Automation of Resume Recommendation System," International Journal of Computer Applications, vol. 975, no. 8887, pp. 1–8, 2020.
11. urafsky, D., & Martin, J. H. (2023). Speech and Language Processing (3rd ed.). This book provides an in-depth explanation of modern NLP approaches, including word embeddings and transformer-based models, which are useful for semantic resume–job matching and contextual skill extraction.
12. Manning, C. D., Raghavan, P., & Schütze, H. (2008). Introduction to Information Retrieval. This reference explains core information-retrieval techniques such as document ranking, similarity scoring, and vector space models that form the backbone of automated resume screening systems.
13. Aggarwal, C.C. (2018)  
This book discusses text classification and similarity algorithms that are applicable for categorizing resumes and ranking candidates based on job requirements.
14. Goldberg, Y. (2017). Neural Network Methods for Natural Language Processing. This resource focuses on neural-network-based NLP models that enhance resume analysis by capturing semantic relationships beyond traditional keyword matching.
15. Vaswani, A., et al. (2017). Attention Is All You Need. This paper introduces transformer architectures that enable advanced contextual understanding in AI systems, which can significantly improve resume parsing and intelligent candidate evaluation.
16. Salton, G., Wong, A., & Yang, C. S. (1975). A Vector Space Model for Automatic Indexing. This paper introduces the vector space model, which forms the basis for representing resumes and job descriptions as numerical vectors for similarity comparison in automated screening systems.
17. Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in VectorSpace.  
This work presents word embedding techniques that help capture semantic meaning in resumes, enabling better skill matching beyond simple keyword-based approaches.
18. Pennington, J., Socher, R., & Manning, C. D. (2014). GloVe: Global Vectors for Word Representation. This paper explains global word embedding methods that improve contextual understanding in resume analysis and job–candidate matching.
19. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019) This study introduces contextual language models that enhance resume parsing accuracy and semantic similarity computation between resumes and job descriptions.



INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



# INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | [ijmrset@gmail.com](mailto:ijmrset@gmail.com) |

[www.ijmrset.com](http://www.ijmrset.com)