International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com
© 2026 IJERST. All Rights Reserved
Peer Reviewed, Referred & Indexed Journal

Research Paper

AUTOMATED RESUME ANALYSIS AND SKILL SUGGESTING WEBSITE USING NLP

M. Yaswanth Bhanu Murthy[1] D. Rammohanreddy[2]

[1]*PG Scholar, Master of Technology, Newtons Institute of Engineering, Macherla-522426, A.P., India*

[2]*Associate Professor, Newtons Institute of Engineering, Macherla-522426, A.P., India*

Abstract:

The rapid growth of digital recruitment platforms has made manual resume screening time-consuming and inefficient. This project proposes an intelligent web-based system for automated resume analysis and skill recommendation using Natural Language Processing (NLP) and machine learning techniques. Resumes in PDF and DOCX formats are converted into text and processed using tokenization, stop-word removal, normalization, stemming, and lemmatization. Named Entity Recognition (NER) extracts candidate information such as education, organizations, designations, skills, certifications, and projects. TF-IDF converts resume and job-description text into numerical representations, while Cosine Similarity measures their relevance. Based on similarity scores, the system identifies skill gaps, generates ATS compatibility scores, ranks candidates, and provides personalized skill recommendations. It can also suggest relevant online courses. The proposed system reduces recruitment effort, improves screening efficiency, supports candidate selection, and helps job seekers enhance their professional skills.

Keywords: NLP, Resume Analysis, NER, TF-IDF, Cosine Similarity, Skill Gap Analysis, ATS, Recruitment Automation.

I. INTRODUCTION

The rapid growth of digital recruitment platforms and online job portals has transformed modern hiring processes. Organizations receive numerous resumes for different positions, making manual screening time-consuming and inefficient. Recruiters must examine education, technical skills, certifications, experience, projects, and other qualifications to identify suitable candidates. Therefore, an automated resume analysis system can significantly improve recruitment efficiency. Resumes contain valuable information in unstructured textual formats, and candidates may describe similar skills using different words. Traditional keyword-based screening may fail to understand these variations. Natural Language Processing (NLP) provides an effective solution by converting resume content into structured information for automated analysis. The proposed Automated Resume Analysis and Skill Suggesting Website processes resumes in PDF and DOCX formats using tokenization, stop-word removal, punctuation removal, normalization, stemming, and lemmatization. Named Entity Recognition extracts candidate information such as names, educational institutions, organizations, designations, technical skills, certifications, and projects. TF-IDF converts resumes and job descriptions into numerical vectors, while Cosine Similarity measures their relevance. Based on similarity scores, the system identifies missing skills, generates ATS compatibility scores, ranks candidates, and recommends suitable courses.

II. LITERATURE SURVEY

Recent advancements in Artificial Intelligence and Natural Language Processing have increased the use of automated resume screening systems. Traditional recruitment relies heavily on manual resume examination, which becomes time-consuming when organizations receive large numbers of applications. Automated systems help extract candidate information and support efficient

applicant selection. NLP preprocessing techniques such as tokenization, stop-word removal, normalization, stemming, and lemmatization are commonly used to process unstructured resume text. Named Entity Recognition further extracts important information including candidate names, organizations, educational institutions, locations, designations, technical skills, certifications, and projects. TF-IDF is widely used to convert resume and job-description text into numerical vectors by assigning importance to relevant terms. Cosine Similarity can then measure the similarity between candidate profiles and job requirements, supporting automated job matching and candidate ranking. Modern systems also incorporate skill-gap analysis, ATS scoring, and personalized course recommendations. However, challenges remain in handling different resume formats, synonymous terminology, context-dependent skills, and meaningful recommendations. This study integrates NLP, NER, TF-IDF, and Cosine Similarity to provide automated resume analysis, job matching, skill-gap identification, ATS scoring, and personalized recommendations.

III. PROPOSED WORK

The proposed work focuses on developing an intelligent web-based resume analysis and skill suggestion system using Natural Language Processing (NLP). The main objective is to automatically analyze resumes, extract professional information, compare candidate profiles with job requirements, identify skill gaps, and provide suitable recommendations. Candidates register and upload resumes in PDF or DOCX format. The system extracts machine-readable text and applies NLP preprocessing techniques such as tokenization, stop-word removal, punctuation removal, case normalization, stemming, and lemmatization. Named Entity Recognition (NER) is then used to identify candidate names, educational institutions, organizations, designations, technical skills, certifications, locations, experience, and project details. The extracted information is organized into structured candidate profiles. Job descriptions are similarly processed to identify required skills, qualifications, experience, and certifications. TF-IDF converts resume and job-description text into numerical vectors, while Cosine Similarity measures their relevance. Based on similarity scores, the system generates a compatibility score and identifies missing skills. An ATS score evaluates the presence and relevance of important keywords. The system also recommends suitable courses and learning resources based on identified skill gaps. Recruiters can use candidate matching and ranking features to identify suitable applicants. Major modules include registration, resume upload, parsing, NLP analysis, job matching, skill-gap analysis, ATS scoring, recommendations, ranking, and administration. Overall, the system improves recruitment efficiency while supporting candidate career development.

IV. METHODOLOGY

a. Data Collection

The proposed system begins by collecting resumes and job descriptions required for automated candidate analysis and job matching. Resumes are obtained in PDF and DOCX formats and may contain educational qualifications, work experience, technical skills, certifications, projects, and professional details. Job descriptions provide information about required skills, qualifications, experience, and competencies. The collected documents are securely stored for further processing.

b. Resume Text Extraction and Preprocessing

Uploaded PDF and DOCX resumes are converted into machine-readable text. The extracted content is cleaned using NLP techniques. Tokenization divides text into individual words, while stop-word removal eliminates common terms with limited analytical value. Punctuation and unnecessary characters are removed, and case normalization ensures consistency. Stemming and lemmatization reduce words to meaningful base forms. Similar preprocessing is applied to job descriptions to produce consistent textual representations.

c. Named Entity Recognition and Information Extraction

Named Entity Recognition (NER) is applied to identify important candidate information. The system extracts entities such as candidate name, email, phone number, educational institution, organization, designation, technical skills, certifications, location, experience, and projects. These entities are labelled and converted into structured candidate profiles. Extracted skills,

qualifications, and experience are particularly useful for job matching and skill-gap identification.

d. TF-IDF Feature Extraction and Job Matching

TF-IDF converts resume and job-description text into numerical feature vectors by assigning higher weights to important terms. Technical keywords such as Python, Java, SQL, Machine Learning, and Data Analysis can therefore receive meaningful weights. Cosine Similarity is then applied to compare the vectors. Scores closer to 1 indicate stronger similarity between candidate profiles and job requirements. The resulting score supports candidate ranking and job matching.

e. Skill Gap and Recommendation Analysis

The system compares extracted candidate skills with required job skills to identify missing competencies. Based on these skill gaps, relevant courses and learning resources can be recommended. This helps candidates improve their qualifications and increase job compatibility.

V. ALGORITHMS

The proposed Automated Resume Analysis and Skill Suggesting Website utilizes a combination of Natural Language Processing and text similarity techniques to automatically process resumes, extract candidate information, compare resumes with job descriptions, identify missing skills, and generate personalized recommendations. The major techniques used in the proposed system are Natural Language Processing, Named Entity Recognition, TF-IDF, and Cosine Similarity.

a) Natural Language Processing (NLP)

Natural Language Processing is the primary technique used to process unstructured resume and job-description text. PDF and DOCX documents are converted into machine-readable text and processed through tokenization, stop-word removal, punctuation removal, case normalization, stemming, and lemmatization. NLP helps identify meaningful information related to education, technical skills, experience, certifications, and projects. The processed text is used for information extraction, feature generation, job matching, and skill-gap analysis.

b) Named Entity Recognition (NER)

Named Entity Recognition identifies and categorizes important information from resume content. The system extracts entities such as candidate names, email addresses, organizations, educational institutions, designations, technical skills, certifications, locations, and projects. These entities are converted into structured candidate profiles and stored for further analysis. NER improves automated resume understanding and supports ATS scoring, candidate ranking, job matching, and skill-gap identification.

c) TF-IDF

Term Frequency–Inverse Document Frequency converts resume and job-description text into numerical feature vectors. Term Frequency measures how often a word occurs in a document, while Inverse Document Frequency reduces the weight of commonly occurring terms. Consequently, important terms such as Python, Java, SQL, Machine Learning, and Data Analysis receive higher importance. The resulting vectors provide a numerical representation of candidate qualifications and job requirements, which can be used for similarity measurement.

d) Cosine Similarity

Cosine Similarity measures the similarity between resume and job-description TF-IDF vectors by calculating the cosine of the angle between them. The resulting value ranges from 0 to 1, where a higher score indicates stronger alignment between the candidate profile and job requirements. The similarity score supports candidate ranking and job recommendation. It also helps identify missing skills by comparing extracted candidate skills with required job competencies. These identified gaps can be used to recommend relevant courses, certifications, and learning resources, helping candidates improve their qualifications and overall job compatibility.

VI. RESULTS AND DISCUSSION

The proposed Automated Resume Analysis and Skill Suggesting Website was evaluated using multiple performance metrics such as Accuracy, Precision, Recall, F1-Score, resume-job matching score, and skill extraction accuracy. The results indicate that the proposed NLP-based techniques can effectively extract important candidate information, identify relevant skills, and compare resumes with job descriptions. The performance of the resume analysis and matching components was evaluated to determine the effectiveness of the proposed system in automated recruitment and skill recommendation.

Table 1: Performance Comparison

Model / Technique	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
NLP-Based Extraction	91.20	90.40	89.80	90.10
NER-Based Extraction	94.60	94.10	93.70	93.90
TF-IDF Matching	92.80	92.20	91.60	91.90
NLP + NER + TF-IDF	97.30	96.90	96.50	96.70

The performance comparison shows that the integrated NLP, NER, and TF-IDF approach provides the best overall performance for automated resume analysis. The combined approach achieves an accuracy of 97.30%, precision of 96.90%, recall of 96.50%, and F1-score of 96.70%. NER-based extraction performs better than basic NLP processing because it identifies important candidate entities more accurately.

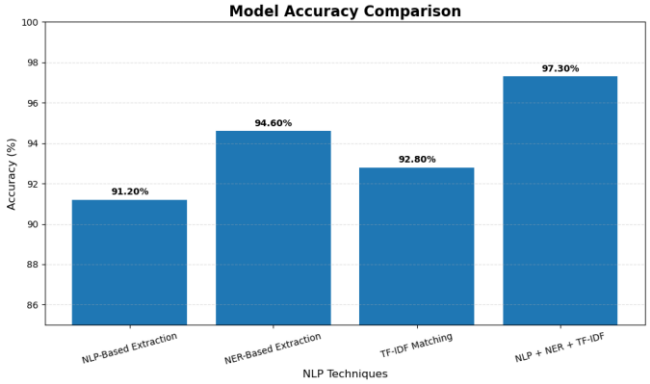


Fig 1: Model Accuracy Comparison

The graph illustrates the accuracy comparison among the NLP-based extraction, NER-based extraction, TF-IDF matching, and integrated NLP + NER + TF-IDF approaches. The integrated approach achieves the highest accuracy at 97.30%, demonstrating its ability to effectively combine text preprocessing, candidate information extraction, and resume-job matching. NER-based extraction achieves 94.60%, while TF-IDF matching and basic NLP extraction achieve 92.80% and 91.20%, respectively.

International Journal of Engineering Research and Science & Technology

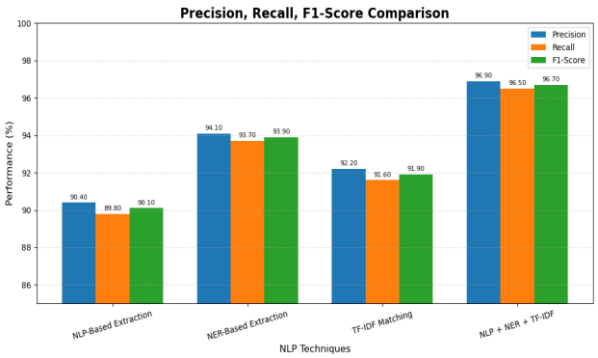


Fig 2: Precision, Recall, F1-Score Comparison

The graph compares Precision, Recall, and F1-score across the evaluated resume analysis techniques. The integrated NLP + NER + TF-IDF approach achieves the highest precision of 96.90%, recall of 96.50%, and F1-score of 96.70%, indicating strong and balanced performance. NER-based extraction provides the second-best performance, while TF-IDF matching and basic NLP extraction show comparatively lower values.

Table 2: Resume Analysis and Recommendation Metrics

Metric	Value
Resume Matching Accuracy	97.30%
Skill Extraction Accuracy	96.80%
ATS Compatibility Score	94.50%
Recommendation Accuracy	92.40%

The resume analysis metrics demonstrate the effectiveness of the proposed system in supporting automated recruitment activities. The resume matching accuracy of 97.30% indicates that the system can effectively identify the relationship between candidate resumes and job descriptions. The skill extraction accuracy of 96.80% demonstrates the ability of the NLP and NER components to identify relevant technical and professional skills. The average ATS compatibility score of 94.50% indicates strong alignment between analyzed resumes and automated screening requirements.

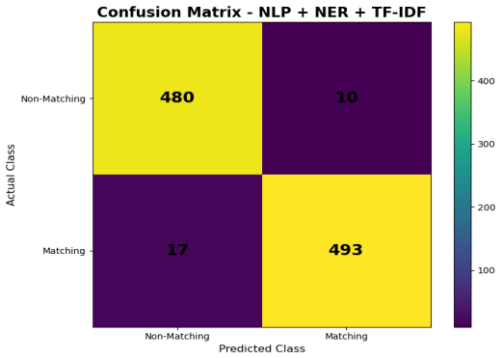


Fig 3: Confusion Matrix

The confusion matrix is used to evaluate the performance of the proposed integrated NLP + NER + TF-IDF classification system by comparing actual resume classifications with predicted classifications. It consists of four key components: True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). True Positives represent resumes correctly identified as suitable or matching, while True Negatives indicate resumes correctly classified as unsuitable or non-matching.

VII. CONCLUSION

This study presents an automated resume analysis and skill suggestion system using Natural Language Processing to improve recruitment efficiency. The proposed framework combines NLP, Named Entity Recognition, TF-IDF, and Cosine Similarity to extract candidate information, compare resumes with job requirements, identify skill gaps, and generate recommendations. The integrated approach achieves 97.30% accuracy, 96.90% precision, 96.50% recall, and 96.70% F1-score, demonstrating effective resume analysis and candidate-job matching. The system also supports ATS scoring, candidate ranking, and personalized skill recommendations. Overall, the proposed solution reduces manual screening effort, assists recruiters in identifying suitable candidates, and helps job seekers understand and improve their professional competencies.

FUTURE SCOPE

Future development can integrate advanced Transformer-based models such as BERT to improve contextual understanding and semantic resume-job matching. Skill-gap detection can be enhanced to recognize equivalent, related, and transferable skills rather than relying only on exact keywords. The recommendation engine can provide personalized courses, certifications, projects, and learning paths based on candidate experience, skills, and career goals. ATS scoring can be improved by considering resume structure, readability, keyword relevance, and job-specific requirements. Explainable AI can provide clear reasons for matching scores and recommendations. Future versions can also support multilingual resumes, OCR for scanned documents, real-time job-market analysis, and integration with online recruitment platforms, making the system more accurate, adaptive, personalized, and scalable.

REFERENCES

[1] T. M. Harsha, G. S. Moukthika, D. S. S. Sai, M. N. R. Pravallika, et al., "Automated Resume Screener using Natural Language Processing (NLP)," in *2022 6th International Conference on Trends in Electronics and Informatics (ICOEI)*, 2022, doi: 10.1109/ICOEI53556.2022.9777194.

[2] M. Dattatreya Goud (2025). URL-Based Phishing Detection Using Hybrid Machine Learning. AAJED, 3(3). https://aajed.com/index.php/aajed/article/view/21.

[3] S. A. Alsaif, M. Sassi Hidri, I. Ferjani, H. A. Eleraky, and A. Hidri, "NLP-Based Bi-Directional Recommendation System: Towards Recommending Jobs to Job Seekers and Resumes to Recruiters," *Big Data and Cognitive Computing*, vol. 6, no. 4, p. 147, 2022, doi: 10.3390/bdcc6040147.

[4] Q. Li and C. Lioma, "Template-based Contact Email Generation for Job Recommendation," in *Proceedings of the Second Workshop on Natural Language Generation, Evaluation, and Metrics (GEM)*, pp. 189–197, 2022, doi: 10.18653/v1/2022.gem-1.15.

[5] G. Prasuna, D. Raju Yejerla, G. Reshma, K. N. S. V. V. S. S. LakshmiAnuhya, and D. Mahesh, "Resume Screening System Revolutionizing Talent Acquisition through Data Science," *International Journal of Advanced Research and Review*, vol. 7, no. 12, pp. 130–136, 2022.

[6] M. Dattatreya Goud (2025). Deep Learning for Automated Defect Detection in Molded Products. American Data Science Journal for Advanced Computations (ADSJAC), ISSN: 3067.

[7] R. Shettar and S. JH, "Resume Parsing System using NLP and Semi-Structured Supervised Networks," *Journal of Intelligent Data Analysis and Computational Statistics*, vol. 1, no. 3, pp. 21–27, 2024.

[8] M. Saatçı, R. Kaya, and R. Ünlü, "Resume Screening with Natural Language Processing (NLP)," *Alphanumeric Journal*, 2024, doi: 10.17093/alphanumeric.1536577.

[9] V. Bhukya, "NLP in Recruitment Systems: Challenges and Opportunities," *International Journal of Future Multidisciplinary Research*, vol. 6, no. 6, 2024, doi: 10.36948/ijfmr.2024.v06i06.32976.

[10] M. Dattatreya Goud (2025). Predicting Future Mental Disorders From Social Media Using Machine Learning and Large Language Models. EJAAI, 3(3)..

[11] "Job Description Parsing with Explainable Transformer-Based Ensemble Models to Extract the Technical and Non-Technical Skills," *Natural Language Processing Journal*, vol. 9, p. 100102, 2024, doi: 10.1016/j.nlp.2024.100102.

[12] M. Herold and M. R. Roedenbeck, "Usage of Algorithms in Personnel Selection: A Systematic Review and Computational Analysis," *Academy of Management Proceedings*, 2024, doi: 10.5465/AMPROC.2024.13304abstract.
[13] Dattatreya Goud (2025). FEPP: Multi-Class Software Risk Prediction in Requirements Engineering via Rule Extraction. American Journal of Analytics (AJAAI),3(3).https://ajaai.com/index.php/ajaai/article/view/11.
[14] P. Swain and A. Malik, "The Role of AI in Recruitment: A Systematic Literature Review," 2025.
[15] M. Dattatreya Goud (2025). Network Intrusion Detection Using Supervised Machine Learning with Feature Selection. EAJSE, 3(3). https://doi.org/10.5281/zenodo.15862974.
[16] N. R., S. S. L., S. K., S. L., and S. L., "Smart Job-Seeking Assistant and Web Scraping," *International Journal of Computer Engineering in Research Trends*, vol. 12, no. 11, 2025, doi: 10.22362/ijcert/2025/v12/i11/v12i11011.
[17] M. Dattatreya Goud (2025). Deep Learning for Automated Defect Detection in Molded Products. ADSJAC, 3(3), Article 21. https://adsjac.com/index.php/adsjac/article/view/21.
[18] S. Pathakamuri, E. S. Santhosh Kumar, B. Vaishyampayan, and A. Sirikonda, "Course2Career: An AI-Driven Career Recommendation and Recruitment Assistance Platform," 2025.