Comprehensive Journal of Science
Volume (10), Issue (38), (Dec, 2025)
ISSN: 3014-6266

مجلة العلوم الشاملة
المجلد (10) ملحق العدد 38 ديسمبر 2025
ردمد: 3014-6266

An Intelligent Resume Analysis and Job Recommendation System Using NLP

A. M. Al-Aswad[1], H. M. Al-Qamati[2],H.S.Abu.Alhaji[3]
[1] Dept. of Electrical and Computer Engineering
School of Applied Science and Engineering
Libyan Academy
Tripoli, Libya
anasdid@gmail.com
[2] Information Technology Dept.
College of Engineering Technology – Janzour
Tripoli, Libya
hus_7119@yahoo.com
3 Dept. of Electrical and Computer Engineering
School of Applied Science and Engineering
Libyan Academy
Tripoli, Libya
hananalhaji73@gmail.com

تاريخ الاستلام:2025/11/10 –تاريخ المراجعة: 2025/12/2– تاريخ القبول: 2025/12/25– تاريخ للنشر:2026 /1/28

***Abstract*— Recruitment systems face increasing challenges due to the diversity of résumé formats, lack of personalized guidance for job seekers, and the persistence of human bias in candidate evaluation. Traditional systems often fail to accurately extract essential information, assess résumé quality, and provide actionable recommendations aligned with labor market requirements.**

This paper presents an intelligent résumé analysis and job recommendation system based on Natural Language Processing (NLP) and transparent rule-based logic. The system extracts structured information from text-based PDF résumés using PDFMiner3, PyResParser, and spaCy. Résumé quality is evaluated using a rule-based scoring mechanism, while skill gap analysis is performed through explicit comparison between extracted skills and predefined domain requirements. Based on this analysis, the system generates personalized job and training recommendations.

The system was evaluated through laboratory testing and real-world field evaluation in collaboration with Human Resources specialists. Experimental results show 90% skill extraction accuracy, 92% job domain classification accuracy, 89% skill gap detection accuracy, and over 91% relevance in job and training recommendations. The findings demonstrate that the proposed system provides an effective, interpretable, and fair solution for intelligent recruitment and career development.

***Keywords*— *Resume Analysis, Natural Language Processing (NLP), Skill Gap Analysis, Job Recommendation System, Rule-Based Decision Making, Recruitment Automation, Career Development, Human Resource Management.*.**

INTRODUCTION

In recent decades, recruitment processes have undergone a significant transformation driven by rapid advances in information technology and digitalization. Organizations increasingly rely

on automated systems to manage human resources and support hiring decisions, particularly as the number of job applicants continues to grow. As a result, résumé screening has become one of the most time-consuming and challenging stages of the recruitment cycle, often requiring substantial human effort and being prone to subjectivity and bias.

The résumé represents the primary document through which a candidate communicates qualifications, skills, and professional experience to potential employers. However, résumés are typically submitted in diverse formats, layouts, and writing styles, which complicates their automated processing. The absence of standardized résumé structures makes it difficult for traditional systems to accurately extract essential information and assess candidate suitability in a consistent and fair manner. Consequently, many qualified applicants may be overlooked, while recruiters face inefficiencies in handling large volumes of applications.

From the job seeker's perspective, rapidly evolving labor market requirements present additional challenges. Technological progress has led to continuous changes in required skills, particularly in technical and digital fields. Many candidates lack clear insight into the competencies demanded by employers and are often unaware of the skill gaps that limit their employability. Existing recruitment platforms typically focus on candidate selection rather than providing personalized guidance or recommendations that support professional development.

Natural Language Processing (NLP) has emerged as a powerful tool for addressing these challenges by enabling automated analysis of unstructured textual data. NLP techniques, such as Named Entity Recognition (NER), have been widely adopted to extract structured information from résumés, including skills, education, and work experience. Despite these advances, many state-of-the-art systems rely on complex machine learning and deep learning models that operate as "black boxes," offering limited transparency and explainability. This lack of interpretability raises ethical and practical concerns, particularly in recruitment contexts where fairness, accountability, and trust are essential.

Furthermore, most existing résumé analysis and job matching systems emphasize keyword-based similarity measures or classification accuracy, often neglecting résumé quality assessment and explicit skill gap identification. These systems rarely provide actionable feedback that helps candidates understand how to improve their résumés or acquire missing skills aligned with labor market demands. As a result, their practical value for career development remains limited.

Motivated by these limitations, there is a growing need for intelligent recruitment systems that go beyond simple résumé parsing and job matching. Such systems should be capable of evaluating résumé quality, identifying missing competencies, and generating personalized job and training recommendations in a transparent and interpretable manner. Combining NLP techniques with rule-based decision logic offers a promising approach to achieving this goal, as it balances analytical accuracy with explainability and ease of implementation.

This paper proposes an intelligent résumé analysis and job recommendation system that integrates Natural Language Processing with transparent rule-based logic. The proposed system aims to enhance recruitment efficiency, reduce human bias, and support professional development by providing clear and actionable recommendations. Additionally, the system is designed to support multilingual environments, with particular attention to Arabic language processing and right-to-left text representation, making it suitable for real-world deployment in diverse regional contexts.

LITERATURE REVIEW

Research on résumé analysis and recruitment automation has attracted growing attention in recent years due to the increasing volume of job applications and the need for efficient, fair, and scalable hiring processes. The existing literature can be broadly categorized into two major research directions: (1) résumé information extraction systems and (2) job recommendation and

candidate–job matching systems. Each direction addresses specific challenges in recruitment automation but also exhibits notable limitations..

Résumé Information Extraction Systems

Résumé information extraction systems primarily focus on converting unstructured résumé documents into structured data that can be processed by automated systems. These approaches commonly employ Natural Language Processing (NLP) techniques, including tokenization, part-of-speech tagging, and Named Entity Recognition (NER), to identify key entities such as skills, educational background, work experience, and personal information..

Several studies have demonstrated the effectiveness of tools such as PyResParser, spaCy, and Stanford NLP in extracting structured résumé content from text-based documents. These tools typically rely on predefined skill dictionaries, pattern matching, and statistical language models. While they achieve satisfactory extraction accuracy, their functionality is largely limited to information retrieval. Most existing systems do not assess the quality, completeness, or professional structure of résumés, which is a critical factor in recruitment decision-making. Moreover, résumé extraction systems often struggle with inconsistencies in résumé formats, variations in terminology, and domain-specific skill representation. Although some studies attempt to address these challenges through customized ontologies or domain-specific vocabularies, such solutions require extensive manual effort and are difficult to generalize across different job domains and industries

B. Job Recommendation and Candidate–Job Matching Job recommendation systems aim to match candidates with suitable job opportunities based on their skills, experience, and qualifications. Traditional approaches in this area frequently rely on similarity-based techniques such as Term Frequency–Inverse Document Frequency (TF-IDF), Cosine Similarity, and Jaccard Similarity to compute the textual similarity between résumés and job descriptions. These methods are computationally efficient and relatively easy to implement, making them popular in early recruitment systems.

However, similarity-based approaches suffer from several limitations. They often depend heavily on exact keyword overlap and fail to capture semantic relationships between skills and job requirements. As a result, candidates with relevant but differently phrased skills may receive low matching scores. Additionally, these systems typically provide only a matching score without offering explanations or actionable feedback, limiting their usefulness for job seekers seeking career guidance.

Recent research has explored the use of deep learning models, particularly transformer-based architectures such as BERT, to improve semantic understanding in résumé–job matching. These models have demonstrated higher accuracy in capturing contextual relationships between skills and job requirements. Nevertheless, their high computational cost, dependence on large labeled datasets, and lack of interpretability pose significant challenges for real-world HR applications, where transparency and trust are essential

C. Skill Gap Analysis and Explainability

A limited number of studies have addressed the concept of skill gap analysis, which focuses on identifying missing competencies that prevent candidates from qualifying for specific job roles. Existing approaches often integrate skill gap analysis implicitly within recommendation models, without explicitly highlighting missing skills or providing clear improvement paths. Consequently, candidates receive limited guidance on how to enhance their employability.

Explainability has also emerged as a critical concern in intelligent recruitment systems. Black-box models, although accurate, raise ethical and legal concerns related to bias, fairness, and accountability. Several researchers have emphasized the need for interpretable and rule-based approaches that allow HR professionals and candidates to understand the rationale behind system decisions.

D. Research Gap and Contribution

Despite the progress achieved in résumé parsing and job recommendation, there remains a clear research gap in developing integrated systems that combine résumé quality evaluation, explicit skill gap detection, and interpretable recommendation mechanisms. Most existing solutions address these components in isolation or overlook one or more of them entirely.

Unlike previous work, the proposed system integrates NLP-based résumé information extraction with rule-based résumé quality assessment and explicit skill gap analysis. By employing transparent decision rules rather than black-box models, the system provides explainable job and training recommendations that address both technical limitations and practical Human Resource management requirements.

SYSTEM ARCHITECTURE

The proposed system is based on a modular and scalable architecture designed to address challenges related to résumé format diversity, unstructured information extraction, skill gap analysis, and the generation of interpretable job and training recommendations. The architecture consists of five interconnected main modules that operate sequentially,.

Figure 1. Overall workflow of the proposed résumé analysis and job recommendation system.

A. Text Extraction Module

The Text Extraction Module represents the first stage of the system. It receives résumés submitted by users in text-based PDF format. This module relies on the PDFMiner3 library to extract raw textual content from PDF files while preserving the original structure as much as possible.

The importance of this module lies in converting unstructured résumé documents into machine-readable text suitable for Natural Language Processing (NLP) techniques. PDFMiner3 was selected due to its high accuracy in handling text-based PDF files compared to alternative solutions, as well as its flexibility in dealing with different document layouts and formatting styles.

B. Structured Data Extraction Module

After extracting the raw text, it is passed to the Structured Data Extraction Module, which applies NLP techniques and Named Entity Recognition (NER) to identify key résumé entities. This module utilizes PyResParser and spaCy to extract essential information, including:

- Name
- Email address

- Phone number
- Skills
- Educational qualifications
- Work experience

This module relies on pretrained language models and integrated skill dictionaries, enabling accurate extraction even when résumés differ in writing style or section ordering. The Structured Data Extraction Module serves as the foundation upon which all subsequent system processes are built.

C. Résumé Quality Evaluation Module

The Résumé Quality Evaluation Module aims to assess the completeness and organization of résumés, an aspect that is often overlooked in traditional recruitment systems. The system employs a rule-based scoring algorithm that checks for the presence of essential professional sections, such as:

- Objective
- Skills
- Projects
- Achievements
- Work Experience

Each section is assigned a predefined weight, and the final résumé quality score is calculated as the sum of these weights. This approach ensures transparency and interpretability, allowing the evaluation results to be easily explained to users or Human Resources specialists.

D. Skill Gap Analysis and Recommendation Engine

The Skill Gap Analysis and Recommendation Engine represents the core component of the proposed system. It performs skill gap analysis by comparing the skills extracted from the résumé with predefined job-domain skill requirements stored in the system database.

The skill gap analysis process relies on set difference operations rather than similarity-based metrics, ensuring clarity and ease of interpretation. Based on the identified missing skills, the system performs the following actions:

Recommends suitable job roles that match the user’s current skill level

Suggests relevant training courses to help bridge detected skill gaps

This approach enables the system to play a dual role: supporting recruitment decision-making on one hand, and guiding job seekers toward professional development on the other.

E. User Interface and Database Module

This module provides an interactive user interface developed using Streamlit, allowing users to upload résumés and view analysis results in a simple and intuitive manner.

IV. METHODOLOGY

The proposed methodology follows a structured and sequential pipeline designed to transform unstructured résumé documents into meaningful, structured information that can be evaluated, analyzed, and used to generate interpretable job and training recommendations. The methodology emphasizes accuracy, transparency, and practical applicability in real-world recruitment environments.FORECASTING METHODS

A. Data Input and Résumé Upload

The process begins with the user uploading a résumé in text-based PDF format through the system interface. Restricting the input to text-based PDFs ensures reliable text extraction and minimizes errors commonly associated with scanned or image-based documents. This design decision improves system robustness and consistency during subsequent processing stages.

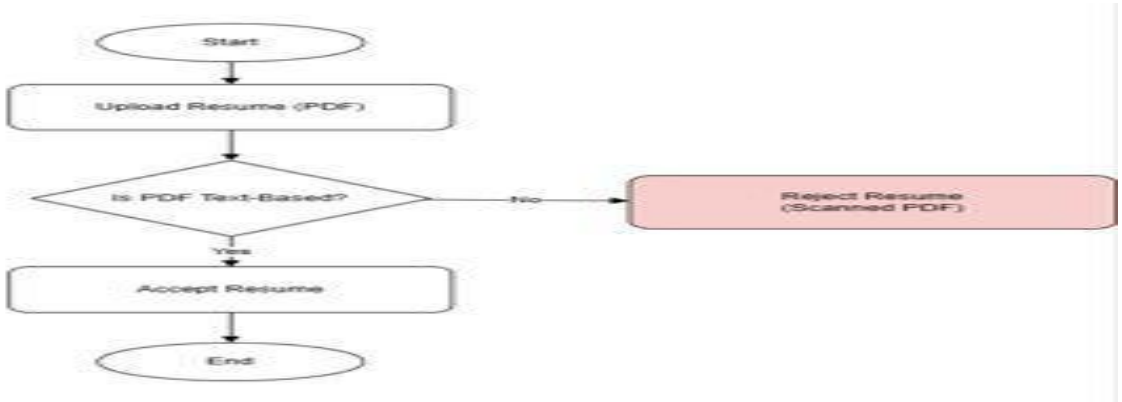

Figure 2. Flowchart of the data input and résumé upload process.

B. Text Extraction

Once the résumé is uploaded, it is passed to the text extraction stage, where the PDFMiner3 library is employed to extract the complete textual content from the PDF file. This step converts the unstructured document into raw text suitable for Natural Language Processing (NLP) techniques.

Special attention is given to preserving the original content order, such as headings and section boundaries, as this significantly enhances the accuracy of later processes including section detection and entity extraction.

C. Data Cleaning and Preprocessing

Data cleaning and preprocessing represent a critical stage of the methodology, as the quality of extracted information directly depends on the cleanliness of the input text. The preprocessing pipeline includes the following steps:

1. **Removal of Irrelevant Symbols and Noise**
   Special characters, duplicated tokens, excessive whitespace, and formatting artifacts introduced during PDF conversion are removed.
2. **Case Normalization**
   All text is converted to lowercase to reduce lexical variation and improve matching consistency.
3. **Stop Word Removal**
   Common non-informative words that do not contribute semantic value are eliminated to enhance extraction efficiency.

4. **Spelling Normalization**
   Minor spelling inconsistencies, particularly in frequently used technical skills, are corrected when possible.
5. **Skill Normalization**
   Different representations of the same skill (e.g., *"Machine Learning"*, *"ML"*) are unified into a standardized format to improve skill matching accuracy.

These preprocessing steps reduce textual noise and significantly improve the performance of downstream NLP and rule-based components.

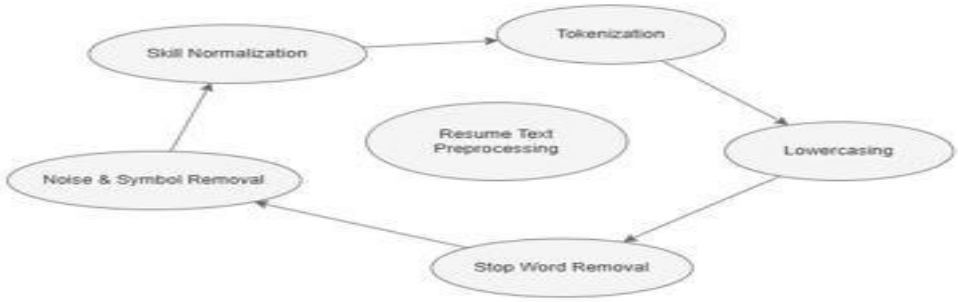

Figure 3Resume Text Preprocessing Stages

D. Structured Information Extraction Using NER

After preprocessing, the cleaned text is processed by the Structured Information Extraction Module. This stage employs **Named Entity Recognition (NER)** techniques using **PyResParser** and **spaCy** to extract key résumé entities, including:

- Personal information
- Technical and professional skills
- Educational background
- Work experience

This approach leverages pretrained language models and curated skill dictionaries, allowing the system to handle diverse résumé formats and writing styles effectively.

E. Résumé Quality Evaluation

The system evaluates résumé quality using a **rule-based scoring mechanism** that assesses résumé completeness and organization. The evaluation focuses on the presence of essential professional sections such as:

- Objective
- Skills
- Projects
- Achievements

- Work Experience

Each section is assigned a predefined weight, and the final résumé quality score is computed as the sum of these weights. This method ensures transparency and allows both users and Human Resources specialists to easily interpret evaluation results.

F. Skill Gap Analysis

Skill gap analysis is performed through a direct comparison between the extracted skill set from the résumé and predefined skill requirements associated with specific job domains stored in the system database.

Instead of using similarity-based metrics, the proposed system applies set difference operations to explicitly identify missing skills. This approach provides clear and interpretable results, making it particularly suitable for decision-making and career guidance applications.

G. Job and Training Recommendation Generation

Based on the outcomes of résumé quality evaluation and skill gap analysis, the system generates:

- Job recommendations aligned with the user's current skills and experience level
- Training and course recommendations aimed at addressing identified skill gaps

All recommendations are generated using predefined, transparent rules, ensuring explainability and reducing potential bias in the recommendation process.

H. Methodology Summary

The proposed methodology is characterized by the following strengths:

- A comprehensive end-to-end résumé processing pipeline
- Transparent, rule-based decision logic
- Practical applicability in real-world recruitment and career development scenarios

By combining NLP techniques with explicit rule-based analysis, the methodology enhances recruitment efficiency while providing meaningful, actionable guidance to job seekers.

V. EXPERIMENTAL EVALUATION

A. Evaluation Setup

The proposed system was evaluated using two complementary approaches to assess its performance and practical applicability.

First, a laboratory evaluation was conducted using a diverse set of text-based PDF résumés covering multiple technical domains. This phase aimed to validate the correctness and robustness of the information extraction, preprocessing, and recommendation modules under controlled conditions.

Second, a field evaluation was performed in collaboration with Human Resources (HR) specialists using 1,750 real-world résumés. This evaluation focused on assessing the system's effectiveness in realistic recruitment scenarios, as well as the relevance and usability of the generated job and training recommendations.

B. Evaluation Metrics

The system performance was assessed using several quantitative metrics, including:

- Skill extraction accuracy
- Job domain classification accuracy

- Skill gap analysis accuracy
- Job recommendation accuracy
- Training recommendation relevance
- Average processing time per résumé

These metrics provide a comprehensive evaluation of both the technical accuracy and the practical efficiency of the proposed system.

C. Results

Table X summarizes the experimental results obtained during the evaluation process.

Metric	Accuracy
Skill Extraction	90%
Job Domain Classification	92%
Skill Gap Analysis	89%
Job Recommendation	93%
Training Recommendation	91%

The experimental results demonstrate that the proposed system achieves high accuracy across all evaluation metrics. Additionally, the average processing time was approximately 1.5 seconds per résumé, indicating that the system is suitable for real-time or large-scale recruitment applications.

VI. DISCUSSION

The experimental results demonstrate that combining Natural Language Processing (NLP) techniques with transparent rule-based logic can achieve high accuracy while maintaining interpretability and fairness. Unlike purely data-driven or deep learning–based approaches, the proposed system provides clear and explainable decision-making mechanisms, which are particularly important in recruitment and Human Resources (HR) environments.

The integration of résumé quality evaluation with skill gap analysis represents a significant contribution compared to existing systems. Most traditional résumé screening tools focus solely on candidate-job matching, whereas the proposed approach extends functionality by assessing résumé completeness and providing actionable feedback. This transformation allows the system to function not only as an automated screening tool but also as a professional development assistant for job seekers.

Furthermore, the use of explicit set-based skill comparison instead of similarity metrics enhances transparency and reduces ambiguity in identifying missing competencies. This design choice helps mitigate bias and increases trust among recruiters and applicants alike. The relatively low processing time confirms that the system is suitable for real-time and large-scale recruitment applications.

Despite its strengths, the system is currently limited to text-based PDF résumés and predefined domain skill sets. These limitations highlight opportunities for further enhancement and expansion.

VII. CONCLUSION AND FUTURE WORK

This paper presented an intelligent résumé analysis and job recommendation system that integrates NLP techniques with interpretable rule-based logic. The proposed system effectively addresses key challenges in modern recruitment, including résumé format diversity, skill mismatch, lack of transparency, and human bias.

Experimental evaluation, including both laboratory testing and real-world field evaluation with HR specialists, demonstrated high accuracy in skill extraction, job domain classification, skill gap analysis, and recommendation relevance. The results confirm that transparent and rule-driven approaches can achieve competitive performance while preserving explainability and fairness.

Future work will focus on several extensions. First, support for scanned and image-based résumés will be incorporated using Optical Character Recognition (OCR) techniques. Second, multilingual NLP support will be expanded, with particular emphasis on Arabic résumé parsing to better serve regional labor markets. Finally, adaptive recommendation techniques such as collaborative filtering and hybrid recommendation models will be explored **to enhance personalization and responsiveness to evolving labor market trends.**

REFERENCES

[1] J. R. Finkel, T. Grenager, and C. Manning, "Incorporating non-local information into information extraction systems by Gibbs sampling," Proc. ACL, pp. 363–370, 2005.

[2] M. Lops, M. de Gemmis, and G. Semeraro, "Content-based recommender systems: State of the art and trends," in Recommender Systems Handbook, Springer, 2011, pp. 73–105.

[3] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," Proc. NAACL-HLT, pp. 4171–4186, 2019.

[4] A. Jain and D. D. Zongker, "Feature selection: Evaluation, application, and small sample performance," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 19, no. 2, pp. 153–158, 1997.

[5] S. Russell and P. Norvig, Artificial Intelligence: A Modern Approach, 4th ed., Pearson, 2021.

[6] E. Riloff and R. Jones, "Learning dictionaries for information extraction by multi-level bootstrapping," Proc. AAAI, pp. 474–479, 1999.

[7] S. Bird, E. Klein, and E. Loper, Natural Language Processing with Python, O'Reilly Media, 2009.

[8] A. T. Al-Dmour, R. Alsudani, and S. Al-Btoush, "Automated résumé classification and job matching using NLP techniques," International Journal of Advanced Computer Science and Applications, vol. 11, no. 6, pp. 456–463, 2020.

[9] M. H. Hassan and A. Mahmoud, "An intelligent recruitment system based on text mining and semantic analysis," Journal of Information and Knowledge Management, vol. 18, no. 3, pp. 1–15, 2019.

[10] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," Proc. ICLR, 2013.

[11] F. Ricci, L. Rokach, and B. Shapira, Recommender Systems Handbook, 2nd ed., Springer, 2015.

[12] A. Caliskan, J. J. Bryson, and A. Narayanan, "Semantics derived automatically from language corpora contain human-like biases," Science, vol. 356, no. 6334, pp. 183–186, 2017.

[13] R. Feldman and J. Sanger, The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data, Cambridge University Press, 2007.

[14] H. Zhang, J. Zhao, and Y. LeCun, "Character-level convolutional networks for text classification," Proc. NIPS, pp. 649–657, 2015.