

AI – BASED MULTIMODEL RESUME RANKING WEB APPLICATION FOR LARGE SCALE JOB RECRUITMENT

¹DULLA RAVI TEJA, ²Dr.GADI. HARITHA RANI

¹Student, ²Associate Professor

¹²DEPARTMENT OF COMPUTER APPLICATIONS

¹²RAJAHMAHENDRI INSTITUTE OF ENGINEERING AND TECHNOLOGY, Rajahmahendry

Abstract—This paper presents an AI-driven resume-ranking web application tailored to enhance large-scale job recruitment processes by leveraging cutting-edge deep learning and natural language processing (NLP) techniques. The system integrates multiple advanced models, including YOLOv9 for layout-aware segment detection, EasyOCR for high-accuracy text recognition, GLiNER for named entity recognition, and a fine-tuned multilingual BERT (mBERT) model for text classification. These components operate in tandem to extract and semantically interpret candidate information from resumes submitted in various formats such as PDF and DOCX. A custom dataset was developed specifically for training the object detection model, where YOLOv9 achieved a superior mean Average Precision (mAP) of 0.84. For semantic matching, the system adopts a hybrid approach combining dense vector similarity via gte-large-en-v1.5 embeddings and cosine similarity, alongside BM25-based keyword matching for relevance scoring. This dual strategy ensures accurate and contextually aware resume-to-job matching. The user-friendly web application allows HR professionals to upload resumes, define job criteria, and receive ranked results, thereby significantly reducing manual effort. Despite challenges in text extraction from irregular layouts and zero-shot NER limitations, the system exhibits robust performance, showcasing the promise of AI-powered recruitment tools in automating and improving candidate selection.

Index Terms—Ai-Based Multimodel, Resume Ranking , Web ,Application, For Large Scale Job Recruitment

I. Introduction

INTRODUCTION OF THE PROJECT

In today's competitive job market, recruitment teams face the daunting challenge of evaluating thousands of resumes with diverse formats, varying layouts, and inconsistent information structures. Manual screening not only consumes valuable time but also risks introducing biases and errors in judgment. Moreover, the lack of standardization in resume design further complicates the automation of the candidate shortlisting process. To address these issues, we propose a novel web-based application that employs state-of-the-art artificial intelligence techniques to rank resumes based on their relevance to predefined job descriptions. Central to our approach is the deployment of a fine-tuned YOLOv9 model for detecting and segmenting structured regions in resumes, paired with EasyOCR for extracting textual data even from noisy or unconventional layouts. We further enhance the semantic understanding of resumes using GLiNER for named entity recognition, regular expressions for structured parsing, and a multilingual BERT (mBERT) model to classify and categorize content. These components are integrated into a pipeline that extracts, matches, and ranks candidate profiles using a hybrid method of embedding-based semantic matching (gte-large-en-v1.5 with cosine similarity) and keyword relevance scoring (BM25). Our application provides a scalable and intelligent solution for HR professionals, aiming to automate the

resume evaluation process, reduce human error, and ensure that the most suitable candidates are identified efficiently.

ORGANIZATION PROFILE

Datapoint IT & Hardware Tech Private Limited is a Hyderabad-based company that provides complete IT solutions and skill development programs. It works with both businesses and educational institutions, helping them with technology needs as well as training. The company focuses on IT infrastructure, hardware and networking, software and web development, technical consulting, and corporate training. It also provides staffing and recruitment services, connecting skilled professionals with companies. Datapoint offers quality-driven and cost-effective services. It helps organizations with IT infrastructure setup, networking solutions, hardware design and support, server management, and software development. The company has expertise in different technologies like Java, Python, Embedded Systems, VLSI, MATLAB, and web development. Because of this wide knowledge, it can handle both simple and complex projects for businesses, academic institutions, and individuals. One of the main strengths of Datapoint is its focus on skill development. It conducts workshops, internships, and academic project support programs that gi

students and professionals practical exposure. Instead of just theory, Datapoint focuses on hands-on learning. Students can work on live projects like Java-based software, Python AI/ML models, IoT devices, embedded systems, and VLSI design. They also learn MATLAB for automation and signal processing, and full-stack web development. This experience helps them gain confidence and prepares them for real-world challenges. For businesses, Datapoint provides corporate training programs that improve employee skills. Many companies face the challenge of adopting new technologies, and Datapoint helps them by conducting customized training sessions. These sessions are designed based on the exact requirements of the company, so the employees learn what they really need for their work. The company also supports educational institutions. It organizes technical seminars, certification courses, and short-term workshops that help students go beyond regular classroom learning. Datapoint acts as a bridge between colleges and industries, making sure that students learn the skills that are actually in demand in the job market. Another important service is staffing and recruitment. Datapoint helps organizations find the right talent for their projects. It also conducts campus recruitment drives to identify fresh graduates who have potential. This helps companies save time in hiring and also helps students find good career opportunities.

Datapoint supports many types of projects, such as:

- Java-based desktop and web applications
- Python programming for data science, AI, and machine learning
- Embedded systems and IoT-based solutions
- VLSI design and simulation projects
- MATLAB-based projects for signal processing and automation
- Full-stack web development and cloud-based solutions
- Hardware and networking implementation

Through these projects, students and professionals get real-time exposure to emerging technologies. For example, a student working on an IoT project will learn how sensors, microcontrollers, and cloud platforms work together. A Python AI/ML project will teach them how to process data, build models, and use them in real applications. These experiences make learners industry-ready.

Datapoint's work approach is simple yet effective. It not only delivers IT solutions to clients but also focuses on creating skilled professionals. By combining services and training, it ensures that businesses get the right technology support while students and professionals get the right

knowledge. This balance of technology and skill development makes Datapoint different from many other companies.

The company also believes in continuous learning and innovation. Technology is changing fast, so Datapoint keeps updating its training programs and services to match the latest trends. It introduces new workshops on emerging fields like cloud computing, data analytics, and cybersecurity. This helps both companies and students stay updated with modern technologies. In the long run, Datapoint IT & Hardware Tech Private Limited aims to create a strong connection between academics and industry. It wants students to have more than just theoretical knowledge and be able to work confidently on practical projects. It also wants businesses to easily adopt new technologies with proper training and support.

With its commitment to quality, learning, and innovation, Datapoint continues to support businesses, educational institutions, and individuals. Whether it is setting up IT systems, training corporate teams, guiding students on academic projects, or developing customized software, Datapoint remains a trusted partner in technology and skill development. By focusing on both solutions and people, it helps build a better and more skilled technology ecosystem.

II. OBJECTIVES OF THE PROJECT

1. To develop an intelligent web-based resume ranking system for automated candidate evaluation.
2. To reduce manual effort and time consumption in the recruitment and resume screening process.
3. To detect and segment resume sections accurately using the YOLOv9 deep learning model.
4. To extract textual information from resumes using EasyOCR, even from complex or noisy layouts.
5. To identify important entities such as skills, education, experience, and contact details using GLiNER and regular expressions.
6. To classify and categorize resume content effectively using the multilingual BERT (mBERT) model.
7. To compare resumes with job descriptions using semantic similarity and keyword-based matching techniques.

III. SYSTEM ANALYSIS

EXISTING SYSTEM

Traditional resume ranking systems typically rely on keyword-based matching techniques, rule-based filtering, and basic natural language processing methods to evaluate and rank candidate resumes against job descriptions. These systems often use simple text extraction tools to convert PDF resumes into raw text and then attempt to identify the presence of predefined keywords related to the job role. The ranking is generally computed by counting keyword occurrences or matching them to a manually constructed ontology or skill set. While this method can work to a limited extent for structured resumes with standardized formatting, it falls short when dealing with the wide variety of resume formats commonly encountered in real-world recruitment. Unstructured or visually complex resumes, such as those with multiple columns, graphics, or non-standard sectioning, are particularly problematic, as conventional systems struggle to accurately parse and interpret the content. Furthermore, keyword matching lacks semantic understanding, meaning it fails to account for synonyms, contextual relevance, or the actual competence reflected by the presence of a term. As a result, many potentially qualified candidates are overlooked due to mismatches in wording, while irrelevant resumes may be ranked higher simply because they contain more matching terms. Additionally, these legacy systems offer limited adaptability and customization, making them inefficient for recruitment scenarios that require deep contextual analysis or tailored evaluation criteria. Overall, while traditional systems are easy to deploy and computationally inexpensive, they are not sufficiently accurate or intelligent to meet the growing demands of modern, data-driven recruitment processes.

Disadvantages:

1. Lack of Semantic Understanding:

Existing systems rely heavily on exact keyword matching and do not understand the semantic meaning of words, leading to inaccurate resume-job matches. Synonyms or relevant skills phrased differently may be ignored.

2. Poor Handling of Unstructured Formats:

Resumes come in various formats and layouts, including multi-column designs, infographics, and embedded tables. Traditional systems often fail to parse such structures accurately, leading to data loss or misinterpretation.

3.Inability to Identify Contextual Relevance:

Keyword-based methods cannot differentiate between superficial keyword presence and actual demonstrated expertise. A resume that merely lists keywords without real context may score higher than a genuinely qualified one.

4.Static and Non-Adaptive:

These systems lack the flexibility to adapt to different job descriptions or domain-specific requirements. Customization is limited, and manually updating keyword libraries is inefficient.

5.Low Accuracy and Recall:

Many relevant resumes are missed (low recall), and irrelevant ones may be ranked higher (low precision) due to the simplistic ranking mechanism based purely on keyword frequency.

6.No Support for Deep Candidate Insights:

Traditional systems do not perform deeper analysis such as entity recognition, skill-level inference, or experience validation, which are critical for effective candidate assessment.

7.Limited Multilingual Support:

Most existing systems are designed for English and struggle with resumes written in other languages or containing code-switched content.

Often, significant manual intervention is required to validate results. This can introduce human bias and reduce the efficiency gains the system is intended to provide.

Algorithms Used:

1. Upload resumes and store them in the database.
2. Detect and segment resume sections using YOLOv9.
3. Extract text from resumes using EasyOCR.
4. Preprocess extracted text by cleaning and tokenizing data.
5. Identify entities like skills, education, and experience using GLiNER and Regex.
6. Classify resume content using multilingual BERT (mBERT).
7. Process job descriptions and extract important keywords.

IV. PROPOSED SYSTEM

The proposed system introduces an intelligent, deep learning-based web application designed to revolutionize the resume ranking process by integrating multiple state-of-the-art AI technologies. Unlike traditional keyword-based systems, this solution uses YOLOv9, a cutting-edge object detection model, fine-tuned on a custom-built dataset specifically curated for diverse resume layouts, enabling

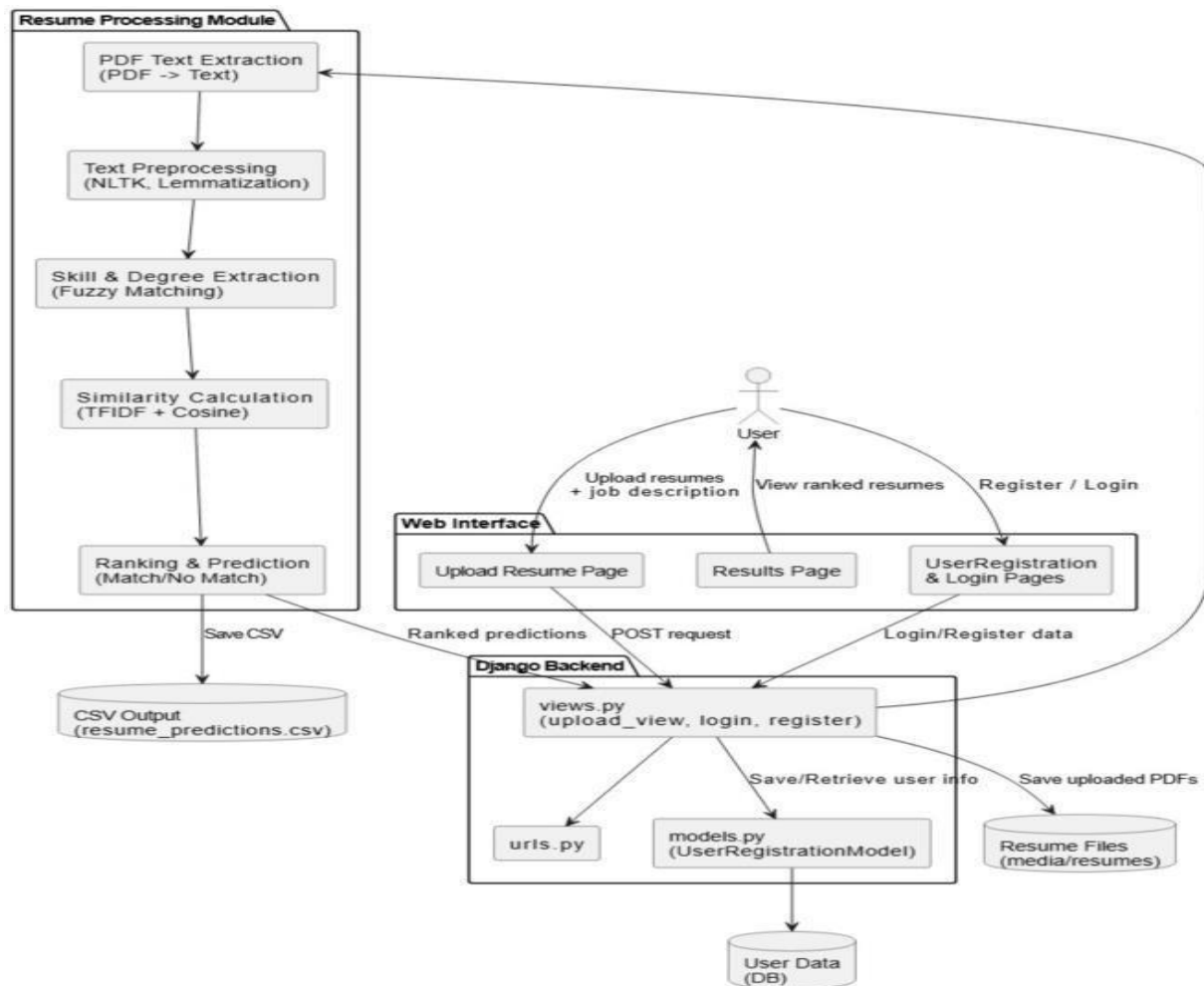
precise segmentation of sections such as education, experience, and skills. Following segmentation, EasyOCR is employed for robust and accurate text recognition from segmented regions, ensuring consistent extraction of content even from complex or unconventional resume formats. The extracted text is then passed through mBERT (multilingual BERT) fine-tuned for job-related classification tasks, enabling the system to categorize content into relevant job fields regardless of language or phrasing differences. To further enhance understanding, GLiNER, a powerful named entity recognition (NER) model integrated with regular expressions, is used to detect and extract important entities like degrees, skills, and locations. The system adopts a hybrid resume-to-job matching strategy that combines semantic similarity using the GTE-large embedding model with cosine similarity to capture contextual relevance, and BM25 for traditional keyword matching. This dual strategy ensures high precision and recall in ranking resumes based on both textual relevance and semantic alignment with job descriptions. The user-friendly web interface allows HR professionals to easily upload resumes in bulk, input job descriptions, and receive a ranked list of candidates complete with detailed metadata like email, extracted skills, and predicted job match confidence.

V. SYSTEM ARCHITECTURE

The system architecture of the AI-Based Resume Ranking and Candidate Matching System is designed to automate the resume screening and candidate ranking process using advanced machine learning and natural language processing techniques. The architecture consists of three major components: the Resume Processing Module, Web Interface, and Django Backend. Initially, users register and log in through the web interface, where HR professionals can upload resumes along with job descriptions. The uploaded resume files are stored securely in the system database and media storage. The Resume Processing Module then performs PDF text extraction to convert resumes into machine-readable text format. After extraction, text preprocessing techniques such as tokenization, stop-word removal, and lemmatization are applied using NLP libraries. The processed text is analyzed to extract important skills, degrees, and experience details through fuzzy matching and entity extraction methods. The system then calculates similarity scores between resumes and job descriptions using TF-IDF and cosine similarity algorithms to measure candidate relevance. Based on these scores, the Ranking and Prediction Module classifies resumes as matched or unmatched and generates ranked predictions. The processed results are saved as CSV files and stored in the database for future reference. The Django Backend manages all server-side operations including user authentication, resume uploads, database management, and

communication between modules using views, models, and URL routing. Finally, the ranked results are displayed on the Results Page through the web interface, enabling HR teams to efficiently shortlist

the most suitable candidates with improved accuracy, reduced manual effort, and faster recruitment processing.



VI. CONCLUSIONS

The proposed AI-based multimodal resume linking web application represents a significant advancement in automating and optimizing the recruitment pipeline. By leveraging deep learning techniques such as YOLOv9 for layout detection, EasyOCR for robust text extraction, mBERT for semantic understanding, and GLiNER for named entity recognition, the system accurately interprets and ranks candidate resumes against job descriptions. The hybrid approach of combining semantic similarity via dense embeddings and keyword relevance using BM25 ensures both contextual understanding and keyword-based filtering. Our custom dataset and model fine-tuning processes demonstrate the importance of domain-specific data in achieving high performance, as reflected in our YOLOv9 mAP score of 0.84. While challenges such as zero-shot NER limitations and extraction from non-standard layouts persist, the results affirm the potential of AI in transforming recruitment workflows. This web application not only accelerates candidate screening but also enhances the objectivity and scalability of the hiring process, providing HR professionals with an effective tool for identifying top talent.

References

- [1] Baykal, Elif. (2020). Digital Era and New Methods for Employee Recruitment. 10.4018/978-1-7998-1125-1.ch018.
- [2] S. Solanki and Dr. Gujarati, "The Digital Revolution In Recruitment: Unraveling The Impact And Challenges Of E-Recruitment," Educational Administration Theory and Practices, vol. 30, Jun. 2024, doi: 10.53555/kuey.v30i6(S).5362.
- [3] Rozario, S. D., Venkatraman, S., & Abbas, A. (2019). Challenges in recruitment and selection process: An empirical study. *Challenges*, 10(2),35.
- [4] Palshikar, G. K., Pawar, S., Banerjee, A. S., Srivastava, R., Ramrakhiyani N., Patil, S., ... & Chalavadi, D. (2023). RINX: A system for information and knowledge extraction from resumes. *Data & Knowledge Engineering*, 147, 102202.
- [5] Komanduri, S., Roopa, Y. M., & Bala, M. M. (2019, March). Novel approach for image text recognition and translation. In 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC) (pp. 596-599). IEEE.
- [6] Subramanian, S., Kekatpure, V., Raymond, G., Parab, K., Dugad, S., & Shirke, A. (2022, January). TEYSuR-Text Extraction with YOLO and Super Resolution. In 2022 International Conference for Advancement in Technology (ICONAT) (pp. 1-7). IEEE.
- [7] M. Alamelu, D. S. Kumar, R. Sanjana, J. S. Sree, A. S. Devi and D.Kavitha, "Resume Validation and Filtration using Natural Language Processing," 2021 10th International Conference on Internet of Everything, Microwave Engineering, Communication and Networks (IEMECON), Jaipur, India, 2021, pp. 1-5, doi: 10.1109/IEMECON53809.2021.9689075
- [8] Roy, P. K., Chowdhary, S. S., & Bhatia, R. (2020). A Machine Learning approach for automation of Resume Recommendation system. *Procedia Computer Science*, 167, 2318-2327.
- [9] Kinge, B., Mandhare, S., Chavan, P., & Chaware, S. M. (2022). Resume Screening using Machine Learning and NLP: A proposed system. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, 8(2), 253-258.
- [10] Nisha, B., Manobharathi, V., Jeyarajanandhini, B., & Sivakamasundari, G. (2023, December). HR Tech Analyst: Automated Resume Parsing and Ranking System through Natural Language Processing. In 2023 2nd International Conference on Automation, Computing and Renewable Systems (ICACRS) (pp. 1681-1686). IEEE.

- [11] Amin, S., Jayakar, N., Sunny, S., Babu, P., Kiruthika, M., & Gurjar, A. (2019, January). Web application for screening resume. In 2019 International Conference on Nascent Technologies in Engineering (ICNTE) (pp.1-7). IEEE.
- [12] Artajaya, H., Giancarlos, J., Moniaga, J. V., & Chowanda, A. (2024, February). Job Recommendation System based on Resume using Natural Language Processing and Distance-based Algorithm. In 2024 IEEE International Conference on Artificial Intelligence and Mechatronics Systems (AIMS) (pp. 1-6). IEEE.
- [13] Mukherjee, A. (2024, February). Resume Ranking and Shortlisting with DistilBERT and XLM. In 2024 IEEE International Conference for Women in Innovation, Technology & Entrepreneurship (ICWITE) (pp. 301-304). IEEE.
- [14] Jayakumar, N., Maheshwaran, A. K., Arvind, P. S., & Vijayaragavan, G. (2023, April). On-Demand Job-Based Recruitment For Organisations Using Artificial Intelligence. In 2023 International Conference on Networking and Communications (ICNWC) (pp. 1-6). IEEE.
- [15] Mohanty, S., Behera, A., Mishra, S., Alkhayyat, A., Gupta, D., & Sharma, V. (2023, May). Resumate: A prototype to enhance recruitment process with NLP based resume parsing. In 2023 4th International Conference on Intelligent Engineering and Management (ICIEM) (pp.1-6). IEEE