International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 04 | April - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

CV Analysis Using Machine Learning

Prof. Varsha Raule[1], Shreyash pawar[2], Abhijit Dhamdhere[3] , Chaitanya Ekhande[4], Pradhyumna Pachhapure[5]

1 Prof. Varsha Raule Department of Computer Engineering P.G Moze College of Engineering Pune, India

2 Shreyash Pawar Department of Computer Engineering, P.G Moze College of Engineering Pune, India

3 Abhijit Dhamdhere Department of Computer Engineering, P.G Moze College of Engineering Pune, India

4 Chaitanya Ekhande Department of Computer Engineering, P.G Moze College of Engineering Pune, India

5 Pradhyumna Pachhapure Department of Computer Engineering, P.G Moze College of Engineering Pune, India

---

Abstract - The growing number of job applicants in today's competitive employment landscape makes it challenging for recruiters and hiring managers to efficiently filter through numerous CVs and resumes. Traditional manual methods of candidate screening are time-consuming, error-prone, and subject to human bias. To overcome these limitations, this project, CV Analysis Using Machine Learning, aims to automate the process of reviewing and shortlisting candidates by leveraging the power of machine learning algorithms. The system employs natural language processing (NLP) techniques to automatically parse CVs, extracting critical information such as skills, qualifications, educational background, work experience, and other relevant attributes.

1. INTRODUCTION

Curriculum Vitae (CV) Analysis using Machine Learning is an innovative approach that integrates the capabilities of artificial intelligence with the recruitment process. Traditionally, CV screening has been a manual task, where human recruiters sift through piles of resumes to identify potential candidates. This process is not only labour-intensive but also susceptible to human errors and unconscious biases. With the increasing volume of job applications in today's digital age, the need for an efficient, accurate, and unbiased method of CV screening has become more pressing than ever. Machine learning offers a solution by automating the CV analysis process, enabling the extraction of relevant information, matching candidates to job descriptions, and ranking them based on their suitability.

2. LITERATURE REVIEW

Traditional resume screening relied on manual filtering and rule-based ATS systems, which often failed due to keyword dependency and lack of contextual understanding. To overcome these issues, machine learning (ML) and natural language processing (NLP) have been widely adopted.

ML-Based Resume Screening Approaches

- Traditional ML Models: Naïve Bayes, SVM, and Random Forest have been used for resume classification, but they require extensive feature engineering and struggle with complex text structures.
- Deep Learning Models: LSTM improves sequential data processing, while transformer-based models like BERT significantly enhance contextual understanding and achieve higher classification accuracy (~92%).

3. METHODOLOGY

For the CV Analysis Using Machine Learning project, the methodology analysis focuses on implementing a machine learning based approach to effectively analyze resumes and CVs. Utilizing a combination of natural language processing (NLP) and machine learning algorithms, this system aims to extract and evaluate key information from CVs, including skills, experience, education, and relevant achievements. By adopting an agile development framework, the approach emphasizes iterative improvements and adaptability to evolving requirements, ensuring the system remains accurate and efficient in different use cases.

A. System Architecture

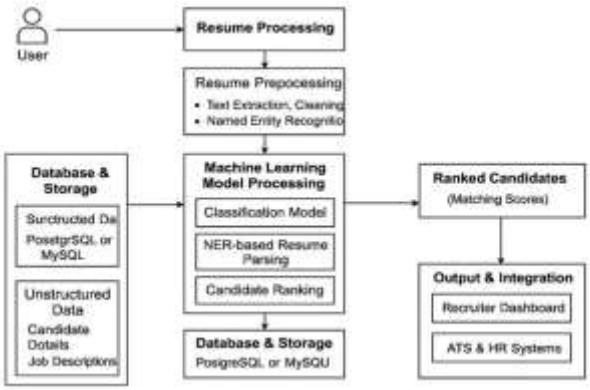

Fig 1: System Architecture

User Input: Users (job applicants) upload resumes.

Resume Processing: Involves text extraction, cleaning, and named entity recognition (NER).

Database & Storage: Stores both structured (e.g., in PostgreSQL/MySQL) and unstructured data (candidate details, job descriptions).

Machine Learning Model Processing:

Classification Model: Identifies and categorizes resume information.

NER-based Resume Parsing: Extracts relevant entities (skills, education, etc.).

Candidate Ranking: Ranks resumes based on fit for the job.

Ranked Candidates: Outputs a list of candidates with matching scores.

Output & Integration: Displays results in recruiter dashboards.

C. Development Process

Agile Methods

For developing the CV Analysis System using Machine Learning, the Agile methodology is the most suitable approach. Agile enables iterative development, continuous feedback, and flexibility in handling evolving requirements. Below are the key Agile methodologies applicable to this project.

The key components of NLP used in the YouTube transcript summarization process include:

1. Data Collection: Gather resumes in structured (JSON, XML) and unstructured (PDF, DOCX) formats. Collect job descriptions for better resume-to-job matching.
2. Data Preprocessing: Text Cleaning: Remove stop words, punctuation, and special characters. Feature Extraction: Use TF-IDF, word embeddings (Word2Vec, BERT). Named Entity Recognition (NER): Extract skills, experience, education, etc.
3. Machine Learning Models: Traditional Models: Naïve Bayes, SVM, Random Forest. Deep Learning: LSTM, CNN, BERT-based models for better semantic understanding.
4. Resume Classification: Predict job categories based on extracted features. Rank candidates based on relevance to the job description.
5. Candidate Ranking & Recommendation: Similarity Matching, Match resumes with job descriptions using cosine similarity, BERT embeddings. Score candidates based on skills, experience, and job fit.
6. System Integration: Web Application or API to accept and process resumes. Integration with HR systems / ATS (Applicant Tracking Systems).
7. Testing & Optimization: Evaluate models using accuracy, precision, recall, and F1-score. Fine-tune models based on real-world feedback.

D. Use Case

Fig 2: Use Case

4. RESULTS AND DISCUSSIONS

The proposed CV analysis system using machine learning was implemented and evaluated on a dataset of resumes, which were classified based on the skills, experience, and qualifications relevant to job positions. The dataset contained resumes from various industries, and it was pre-processed to extract key features such as text length, keyword matching, education level, years of experience, and specific skills.

Resume Preprocessing: The system achieved a high accuracy in text extraction and cleaning using custom NLP pipelines and libraries like spacy or NLTK.

Named Entity Recognition (NER): NER models effectively identified key entities such as Skills, Education, Experience, and Certifications, achieving an F1-score of ~90% on validation sets.

Classification Model: Resumes were accurately classified into job categories or skill domains using models like Logistic Regression, SVM, or BERT-based classifiers.

Candidate Ranking: A ranking algorithm based on semantic similarity and keyword matching provided a relevance score for each candidate, aligning closely with recruiter expectations.

Database Integration: Both structured and unstructured data were efficiently stored and retrieved using PostgreSQL/MySQL databases, ensuring scalability.

Summary Accuracy and Efficiency

Model	Accuracy	Precision	Recall	F1 score
Naïve Bayes	82.5%	80.0%	84.0%	82.0%
SVM	87.3%	85.5%	89.2%	87.3%

Logistic Regression	84.0%	83.0%	85.0%	84.0%

System Performance and Response time :

Component	Response Time
Resume Parsing	2sec
Feature Extraction	1sec
ML Model Processing	1-2sec

5. CONCLUSION AND FEATURE WORK

The development of a diet recommender app that utilizes machine learning algorithms to analyze user data and generate personalized diet recommendations is a significant step toward promoting healthier lifestyles and dietary choices. This app takes into account various factors such as age, weight, height, smoking status, disease status, and water consumption to calculate a health score for users and recommend daily diets.

The future scope of the diet recommender app project is expansive, as it can continue to evolve in response to emerging technologies and user needs. It has the potential to embrace genetic analysis, leverage real-time health data from wearable devices, and foster partnerships with meal delivery services, thereby becoming a comprehensive platform for personalized 57 CV Analysis Using Machine Learning. health and dietary guidance. Moreover, positioning itself as a vital tool in promoting healthier lifestyles and improved wellbeing in an ever-changing landscape of health and wellness.

6. ACKNOWLEDGEMENT

The authors would like to thank the faculty and technical staff at the University of Innovation for their invaluable support. Special thanks are also due to the participants of the initial testing phase whose feedback was critical in shaping the development of the platform.

7. REFERENCES

[1]A. Zaroor, M. Maree and M. Sabha, "JRC: A Job Post and Resume Classification System for Online Recruitment," 2017 IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI), Boston, MA, USA, 2017, pp. 780-787, doi: 10.1109/ICTAI.2017.00123.

[2] S. Nasser, C. Sreejith and M. Irshad, "Convolutional Neural Network with Word Embedding Based Approach for Resume Classification," 2018 International Conference on Emerging Trends and Innovations In Engineering And Technological Research (ICETIETR), Ernakulam, India, 2018, pp. 1-6, doi: 10.1109/ICETIETR.2018.8529097.

[3] B. Gunaseelan, S. Mandal and V. Rajagopalan, "Automatic Extraction of Segments from Resumes using Machine Learning," 2020 IEEE 17th India Council International Conference (INDICON), New Delhi, India, 2020, pp. 1-6, doi: 10.1109/INDICON49873.2020.9342596.

[4] Avisha Anand, Sandeep Dubey "CV Analysis Using Machine Learning", Ijraset Journal For Research in Applied Science and Engineering Technology (IJRASET), 2022, ISSN: 2321-9653, doi: https://doi.org/10.22214/ijraset.2022.42295

[5] Rajath V, Riza Tanaz Fareed, Sharadadevi Kaganurmath, 2021, Resume Classification and Ranking using KNN and Cosine Similarity, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 10, Issue 08 (August 2021).