International Journal of Scientific Research in Computer Science, Engineering and Information Technology

ISSN : 2456-3307

OPEN ACCESS

Available Online at : www.ijsrcseit.com

doi : https://doi.org/10.32628/CSEIT2410275

Resume Screening using Machine Learning

Dr. Sandeep Tayal[1], Taniya Sharma[1], Shivansh Singhal[3], Anurag Kumar Thakur[4]

[1]Assistant Professor, Department of Computer Science, Maharaja Agrasen Institute of Technology, Delhi, New Delhi, India

[2,3,4]UG Student, Department of Computer Science, Maharaja Agrasen Institute of Technology, Delhi, New Delhi, India

ARTICLEINFO

Article History:

Accepted: 10 April 2024

Published: 19 April 2024

Publication Issue

Volume 10, Issue 2

March-April-2024

Page Number

602-606

ABSTRACT

This study explores the utilization of Machine Learning (ML) and Natural Language Processing (NLP) in automating the resume screening process. Traditional methods, often manual and subjective, fail to efficiently manage the volume and variety of resumes. By employing NLP techniques like named entity recognition and part-of-speech tagging, coupled with ML classifiers such as K-Nearest Neighbors and Support Vector Machines, we propose a system that enhances the precision of candidate selection while significantly reducing time and effort.

Keywords : Machine Learning, Natural Language Processing, Resume Screening, NLTK, K-Nearest Neighbors, Support Vector Machines.

I. INTRODUCTION

The modern job market is characterized by a dynamic and ever-evolving landscape where both employers and job seekers strive for the best possible matches. Employers seek to fill positions with candidates who not only possess the necessary skills and qualifications but also align with the company's culture and values. Conversely, job seekers aim to find roles that leverage their strengths and offer growth opportunities. This dual objective sets the stage for a complex matchmaking process. Traditional resume screening methods, heavily reliant on human judgment and manual effort, present significant challenges in meeting these objectives efficiently. Recruiters often sift through hundreds, if not thousands, of resumes for a single job posting, a task that is not only time-consuming but also fraught with the potential for oversight and bias. The traditional approach to resume screening involves a manual review of each application to assess the candidate's fit based on experience, education, skills, and other relevant criteria. This process is not only labour-intensive but also subjective, leading to inconsistencies in candidate selection. Furthermore, the manual screening process struggles to cope with the volume and variety of resumes, often resulting in qualified candidates being overlooked or the recruitment cycle being unnecessarily prolonged. In response to these challenges, our research proposes an automated system that harnesses the power of Machine Learning (ML) and Natural Language Processing (NLP) to revolutionize the resume

screening process. By automating the extraction and interpretation of information from resumes, this system aims to significantly reduce the time and effort involved in screening candidates while improving the accuracy and objectivity of the selection process.

ML algorithms can analyse vast amounts of data to identify patterns and make predictions. In the context of resume screening, ML models are trained to categorize candidates based on their suitability for a role, using historical hiring data and outcomes as a learning basis. This approach enables the system to evaluate candidates more accurately and consistently than human screeners, potentially uncovering strong candidates who might otherwise have been overlooked due to unconventional career paths or non-traditional skill sets. NLP techniques are utilized to interpret the textual content of resumes, extracting valuable information such as skills, work experience, education, and achievements. This technology allows the system to understand and process the natural language found in resumes, transforming unstructured text into structured data that can be easily analysed and compared across candidates.

II. LITERATURE

A. Resume Screening Using Machine Learning and NLP: A Proposed System

The model described processes student resumes to determine job role suitability and areas for resume improvement. It utilizes machine learning, with input being the resume and optional LinkedIn and GitHub links for additional insights. The architecture includes SQL databases for structured data storage and employs either K-Nearest Neighbour or Support Vector Machine for job role prediction. Cosine similarity is used in a recommendation system to suggest resume enhancements based on desired roles and existing qualifications. Data for model training comes from Kaggle.

B. Resume Screening Classification using Artificial Intelligence and Natural Language Processing

The paper "Resume Screening Classification using Artificial Intelligence and Natural Language Processing" introduces the Prospect model, a machine learning-based system for automating resume screening. This model achieves a remarkable accuracy of 93.5%, significantly outperforming traditional convolutional neural network models by 19.5%. It employs a two-phase approach, starting with the pre-processing and feature extraction from a unique dataset called the Prospect dataset, which includes around 5,000 resumes. This setup ensures an unbiased classification of resumes into "selected" or "rejected" categories based on a sophisticated matching score algorithm and custom logic. The integration of artificial intelligence and machine learning techniques in this model offers a promising direction for enhancing the efficiency and fairness of the resume screening process.

C. A Machine Learning approach for automation of Resume Recommendation System

The paper evaluates four machine learning models for classifying resumes to suggest job roles and improvements. The models and their accuracies are as follows: Random Forest at 38.99%, Multinomial Naive Bayes at 44.39%, Logistic Regression at 62.40%, and Linear Support Vector Machine Classifier at 78.53%. The SVM classifier outperforms others, highlighting its effectiveness in resume classification and job recommendation tasks.

D. Automatic Resume Filtering Using Machine Learning

The paper details a study on the automation of resume screening through the application of Machine Learning (ML) and Natural Language

Processing (NLP). The literature review highlights significant contributions to the field, including works by Nandhini S, Gomathi S, Lavanya S, Kondapalli Sai Pranay, Shweta Agrawal, and Sumit Gupta, among others. These studies collectively explore various methodologies for extracting and ranking data from resumes using NLP techniques and matching them with job descriptions through ML algorithms. The emphasis across the research is on enhancing the efficiency and accuracy of the resume screening process, which is crucial for streamlining recruitment and ensuring optimal job-candidate matches.

III.PROPOSED SYSTEM

A. Problem Statement

The current manual process of resume screening is labor-intensive, time-consuming, and susceptible to bias, failing to efficiently handle the volume and diversity of job applications. This necessitates an innovative approach to automate and enhance the screening process. Leveraging Natural Language Processing (NLP) and Machine Learning (ML) technologies, this study proposes a system aimed at improving the accuracy, efficiency, and fairness of candidate selection, addressing the pressing need for a scalable and unbiased recruitment solution in the digital age.

B. Solution

To address the inefficiencies and limitations of traditional manual resume screening, this research introduces a cutting-edge automated system, leveraging the synergy of Natural Language Processing (NLP) and Machine Learning (ML) technologies. The solution encompasses a comprehensive strategy beginning with the collection and preprocessing of a diverse dataset of resumes, ensuring readiness for detailed analysis through techniques such as text normalization, tokenization, and cleaning. Through the application of advanced NLP techniques like named entity recognition and part-of-speech tagging, the system adeptly extracts critical data from resumes, such as skills, education, and work experiences, while accommodating various resume formats and languages. This foundational work facilitates the transition to feature extraction, where key data points are transformed into a numerical format suitable for ML model training, employing methodologies like TF-IDF and word embeddings. Subsequent stages involve the deployment of ML classifiers—K-Nearest Neighbours, Support Vector Machines, and One v/s Rest among them—to categorize resumes effectively. This process is refined through rigorous training and testing phases, employing metrics such as accuracy, precision, recall, and F1 score for evaluation. The culmination of the system's development sees the integration of the NLP-based resume parser with ML classifiers, resulting in a comprehensive automated screening system. This system is meticulously tested and refined with industry feedback, ensuring it not only meets but exceeds the requirements of modern recruitment processes by delivering a solution that is both scalable and unbiased, significantly enhancing the efficiency, accuracy, and fairness of candidate selection in the recruitment landscape.

C. System Architecture

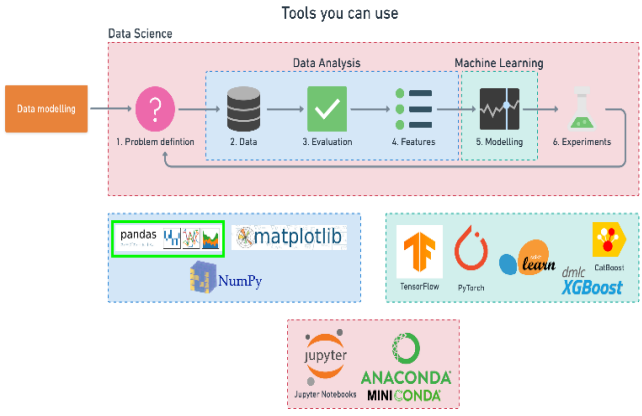

Image source:

https://dev.mrdbourke.com/zero-to-mastery-ml/introduction-to-pandas/

Figure 1. System Architecture

At its core, the system intends to handle the vast diversity of resumes across different job roles and industries by initially collecting a wide-ranging dataset of resumes. This dataset undergoes a comprehensive preprocessing phase where text normalization, tokenization, and cleaning are performed to prepare the data for subsequent analysis. Crucial to enhancing the dataset's utility, resumes are annotated with labels that mark relevant categories such as skills and experience levels, paving the way for a more structured and informative dataset. The utilization of NLP techniques, including named entity recognition and part-of-speech tagging, forms the backbone of the resume parsing process. This enables the system to accurately extract and organize key information such as skills, educational background, and work experience from the resumes. Special algorithms are developed to accommodate the variations in resume formats and languages, ensuring a versatile and robust parsing mechanism. With the foundation laid by NLP, the system then transitions to feature extraction, where a specific set of features is identified from the parsed data. This step is crucial for feeding well-structured input data into the ML classifiers.

The next phase involves the selection and implementation of ML classifiers, including but not limited to K-Nearest Neighbours, Support Vector Machines, and One v/s Rest. The system employs a strategic split of the dataset into training and testing sets, allowing for the comprehensive training and subsequent evaluation of the models. Performance metrics such as accuracy, precision, recall, and F1 score serve as benchmarks for evaluating the models' effectiveness in categorizing resumes into relevant job roles or skill categories. The culmination of the system's development is the integration phase, where the NLP-based resume parser is seamlessly combined with the ML classifiers, giving rise to a unified automated resume screening system. This system undergoes rigorous end-to-end testing with a distinct set of resumes to validate its accuracy and operational efficiency. Feedback from recruiters and industry professionals plays an instrumental role in refining the system, ensuring it meets the practical demands of the recruitment process.

D. Results and Accuracy Table

Table 1: Accuracy for Different Model

Model Used	Accuracy
KNN	98.44%
One v/s Rest	92.23%
SVM	94.76%

The table demonstrates the KNN model's superior performance, achieving an accuracy of 98.44%, making it the most effective model among those tested for the purpose of resume screening.

IV. CONCLUSION

The research embarked upon the ambitious goal of automating the resume screening process, harnessing the power of Natural Language Processing (NLP) and Machine Learning (ML) to introduce a solution capable of transforming the recruitment landscape. Through diligent exploration and implementation of various ML models, including K-Nearest Neighbors (KNN), One vs Rest, and Support Vector Machine (SVM), the study demonstrated not just the feasibility but the remarkable efficiency and accuracy of automating resume classification. The results underscored the superior performance of the KNN model, achieving an impressive accuracy of 98.44%, thereby showcasing its potential as a leading tool for resume screening. While the One vs Rest classifier and the SVM also performed commendably, with accuracies of 92.23% and 94.76% respectively, the KNN model stood out for its ability to navigate the complexities of resume data, offering a

scalable and highly reliable method for candidate classification.

IV. REFERENCES

[1]. B. Kinge, S. Mandhare, P. Chavan, and S. M. Chaware, "Resume Screening Using Machine Learning and NLP: A Proposed System," in International Journal of Scientific Research in Computer Science, Engineering and Information Technology, vol. 253, doi: 10.32628/CSEIT228240

[2]. A. K. Sinha, M. A. K. Akhtar, M. Kumar, and S. Upadhyay, "Resume Screening Classification using Artificial Intelligence and Natural Language Processing," Name of Journal, doi: 10.48047/ecb/2023.12.si4.130

[3]. P. K. Roy, S. S. Chowdhary, and R. Bhatia, "A Machine Learning approach for automation of Resume Recommendation system," in Procedia Computer Science, vol. 167, pp. 2318-2327, Jan. 2020, doi: 10.1016/j.procs.2020.03.284

[4]. D. L. Padmaja, Ch. Vishnuvardhan, G. Rajeev, and K. N. S. Kumar, "Automated Resume Screening Using Natural Language Processing," in Journal of Emerging Technologies and Innovative Research (JETIR), vol. 10, no. 3, Mar. 2023