Resume Sorter using NLP

Project Synopsis Report

Submitted in partial fulfillment of the requirement of the degree of

Master of Computer Application

to

K.R. Mangalam University

by

Kajal Garg (2401560029)

Rekha Kumayan(2401560008)

Under the supervision of
Dr. Tanu Gupta

Department of Computer Science and Engineering

School of Engineering and Technology

K.R Mangalam University, Gurugram-122001, India

January 2025

INDEX

ABSTRACT

The Resume Sorter is designed to automate and streamline the recruitment process by efficiently analyzing and ranking resumes. It enables quick and accurate evaluation of candidates based on job-specific requirements, reducing manual effort and ensuring fair and effective hiring decisions. This system aims to save time, enhance accuracy, and improve overall recruitment efficiency.

KEYWORDS : Resume Sorter, NLP, Machine Learning , Recruitment Automation, Candidate Classification.

INTRODUCTION

In the recruitment process, evaluating a large number of resumes is often a tedious and time-consuming task for organizations. Identifying the most suitable candidates manually can lead to inefficiencies, inconsistencies, and delays in hiring. To address these challenges, we are working on a project that utilizes *Natural Language Processing (NLP)* to create a framework for an efficient *Resume Sorter*.

The aim of this project is to analyze and categorize resumes by extracting critical information such as skills, experience, and qualifications, and aligning it with the job requirements. By automating this process, the Resume Sorter helps reduce manual effort, ensures accurate candidate evaluation, and speeds up recruitment.

This project highlights the potential of NLP in revolutionizing traditional resume screening practices. Through our efforts, we aim to develop a scalable, effective, and unbiased approach to enhance the efficiency of the recruitment process, ultimately benefiting both recruiters and job seekers.

MOTIVATION

The recruitment process is one of the most critical functions within any organization, as it directly impacts the quality and success of the workforce. However, with the increasing volume of job applications, manual resume screening has become a significant bottleneck in the hiring process. Recruiters are often overwhelmed with hundreds or even thousands of resumes for each job opening, making it difficult to quickly and effectively identify the best candidates. As a result, hiring decisions can be delayed, candidates may be overlooked, and human bias may influence the selection process. This highlights the need for a more efficient and automated solution to optimize resume screening.

Manual Resume Screening Challenges:

1. **Time-Consuming Process**: Reviewing each resume individually takes a significant amount of time and effort. On average, recruiters spend only a few seconds reviewing each resume, leading to rushed decisions and missed opportunities.
2. **Bias and Inconsistencies:** Human recruiters may unintentionally exhibit biases based on factors like gender, age, or educational background. These biases can affect the fairness of the hiring process and lead to unbalanced candidate selection.
3. **High Volume of Applications:** With the rise of online job portals, companies receive a massive influx of resumes. This high volume makes it nearly impossible for recruiters to thoroughly review each resume and compare them against the job requirements.
4. **Quality of Matching:** Identifying the most relevant candidates based on skills and qualifications is often challenging due to the variation in resume formats and the lack of standardization.

Efficiency, Accuracy, and Fairness:

The ultimate motivation for this system is to create an AI-driven, scalable solution that minimizes the time and effort required to screen resumes while improving the accuracy and fairness of the process. By automating the tedious tasks of sorting, categorizing, and ranking resumes, recruiters can focus on high-level decision-making and interacting with candidates, which enhances their overall productivity.

Additionally, the system significantly reduces human bias, ensuring that all candidates are evaluated based solely on their qualifications and experience, rather than subjective opinions or unconscious prejudices. It also improves the consistency of the selection process, ensuring that all resumes are evaluated according to the same criteria and standards.

A Future-Oriented Solution:

The motivation behind this project lies in creating a future-oriented solution that transforms traditional recruitment practices. By leveraging Natural Language Processing (NLP), we aim to automate resume evaluation, ensuring faster and more accurate shortlisting of candidates. This solution aligns with the evolving demands of modern organizations, which require scalable, unbiased, and technology-driven approaches to recruitment.

With increasing advancements in AI and NLP, this project offers a forward-thinking approach that not only addresses present challenges but also anticipates the needs of a rapidly changing job market. It enables organizations to adapt to high application volumes, improve decision-making, and foster fair hiring practices, ultimately paving the way for a smarter and more efficient recruitment process.

LITERATURE REVIEW

SNo.	Research Paper Title & Year	Journal / Conference	Dataset	Methodology	Results
1.	Resume Screening With Natural Language Processing 2024 [1]	Alphanumeric Journal	Custom Resume Dataset	NLP-based resume screening using BERT and RankNet for improved job matching and skill extraction accuracy.	NLP-based resume screening systems achieving up to 99% accuracy in classification and over 90% in parsing resumes.
2.	Application of LLM agents in recruitment 2024 [2]	Journal of Information Processing	Job-Salary Prediction Dataset	It utilizes LLMs to make decisions, identifying candidates to invite for interviews or extend job offers.	LLM agents for recruitment achieved an F1 score of 87.73% for resume classification.
3.	Resume Ranking Using NLP 2024 [3]	Journal of Computers and Intelligent System	Revisiting Resume Classification with Large-Scale Datasets	The researchers curated a dataset of 13,389 resumes from diverse sources and used LLMs like BERT and Gemma1.1 2B for classification.	The best model achieving a top-1 accuracy of 92% and a top-5 accuracy of 97.5%.
4.	Applicant Screening System Using NLP 2023 [4]	International Journal of Advanced Computer Science and	DataHack Resume Dataset	The system utilized NLP techniques to extract and rank candidate information from resumes,	It proposed an automated system that achieved an average text parsing accuracy of 85% and a ranking accuracy of 92%.

		Applications (IJACSA).		enhancing the efficiency and consistency of the recruitment process.	
5.	National Origin Discrimination in Deep/learning/powered Automated Resume Screening 2023 [5]	arXiv Journal	The "English Resume Dataset" by DataHack	The researchers developed a bias mitigation method and conducted extensive experiments on real candidate resumes to validate their findings.	National origin discrimination in deep learning-powered automated resume screening did not provide specific accuracy percentages.
6.	A Prototype to Enhance Recruitment Process with NLP based Resume Parsing 2023 [6]	International Conference on Intelligent Engineering and Management (ICIEM).	Job Candidate Dataset from GitHub	It converts resumes into a standard text format, parses them to extract necessary details.	The study reports that the system achieved an accuracy of 85% in parsing and extracting relevant information from resumes
7.	Resume Classification System using Natural Language Processing and Machine Learning Techniques 2022 [7]	Journal of Engineering and Technology	DataHack Resume Dataset	It proposed an automated system that classifies resumes into 25 job categories using supervised machine learning algorithms	The dataset comprised 962 labeled resumes, achieving an accuracy of 85% in classification.
8.	CV Analysis Using Machine Learning 2022 [8]	International Journal for Research in Applied Science and Engineerin	Custom Resume Dataset	The methodology involved parsing resumes to extract relevant information,	The system achieved an accuracy of 30.92% for detecting educational institutions and 19.68% for

		g Technology		followed by classification using machine learning algorithms.	company names, outperforming single-module approaches.

GAP ANALYSIS

The recruitment process faces multiple challenges, including inefficiencies in manually sorting resumes, human bias in candidate selection, and inconsistency in resume formatting. The Resume Sorter using NLP aims to address these gaps by automating resume classification and improving the hiring process . However, while the technology offers significant benefits, there are several gaps between the current practices in recruitment and the full potential of an automated resume screening system.

This GAP analysis will evaluate the existing practices in resume screening against the capabilities of the proposed system, identifying areas for improvement . Points are given:

1) Traditional resume screening is highly manual and time-consuming, leading to delays in candidate selection.

2) Human recruiters may unintentionally exhibit biases (e.g., based on gender, age, education, etc.) when reviewing resumes.

3)Resumes come in diverse formats and styles, making it difficult for recruiters to efficiently extract relevant information.

4) Traditional resume sorting methods often rely on keyword matching, which may not fully capture the nuances of a candidate's qualifications.

5) Current recruitment tools may not be standardized in their use of AI and NLP models, leading to inconsistent results.

6) Sensitive candidate data (personal information, work history, etc.) needs to be protected in AI-driven systems, which may not always adhere to privacy standards.

PROBLEM STATEMENT

The traditional process of manually screening resumes is time-consuming, inefficient, and prone to human bias, especially when dealing with a large volume of applications. Organizations struggle to identify the most suitable candidates due to the inability to consistently extract and evaluate key information such as skills, experience, and qualifications from unstructured resume data.

Existing automated tools often fail to handle diverse resume formats, align candidate profiles accurately with job descriptions, and provide unbiased evaluations, leading to missed opportunities and delays in the recruitment process.

There is a need for a robust, scalable, and intelligent solution that can efficiently analyze resumes, extract relevant information, and rank candidates based on job-specific requirements, thereby streamlining the recruitment workflow and ensuring fair and accurate hiring decisions

OBJECTIVES

The objective of this project is to develop a framework for a Resume Sorter using Natural Language Processing (NLP) to automate and enhance the recruitment process. The system aims to:

1. **Accurately Extract Information:** Identify and extract key details such as skills, experience, education, and certifications from resumes.

2. **Streamline Candidate Matching:** Compare candidate profiles with job descriptions to rank and shortlist resumes based on job-specific criteria.

3. **Reduce Manual Effort:** Minimize the time and resources required for resume screening while improving the overall efficiency of the recruitment process.

4. **Ensure Fair Evaluation:** Provide unbiased and consistent candidate assessments to foster inclusivity and fairness in hiring decisions.

5. **Handle Diverse Formats:** Adapt to various resume styles and formats to ensure scalability and reliability.

Tools/Technologies Used

1.Programming Language: Python

2.NLP Libraries:

(i)NLTK: Text preprocessing (tokenization, stopword removal)

(ii)SpaCy: Advanced NLP tasks (NER, dependency parsing)

(iii)Gensim: Topic modeling with LDA

3.Machine Learning Libraries:

(i)Scikit-learn: Classification, model evaluation

(ii)TensorFlow : Deep learning, word embeddings (e.g.Word2Vec)

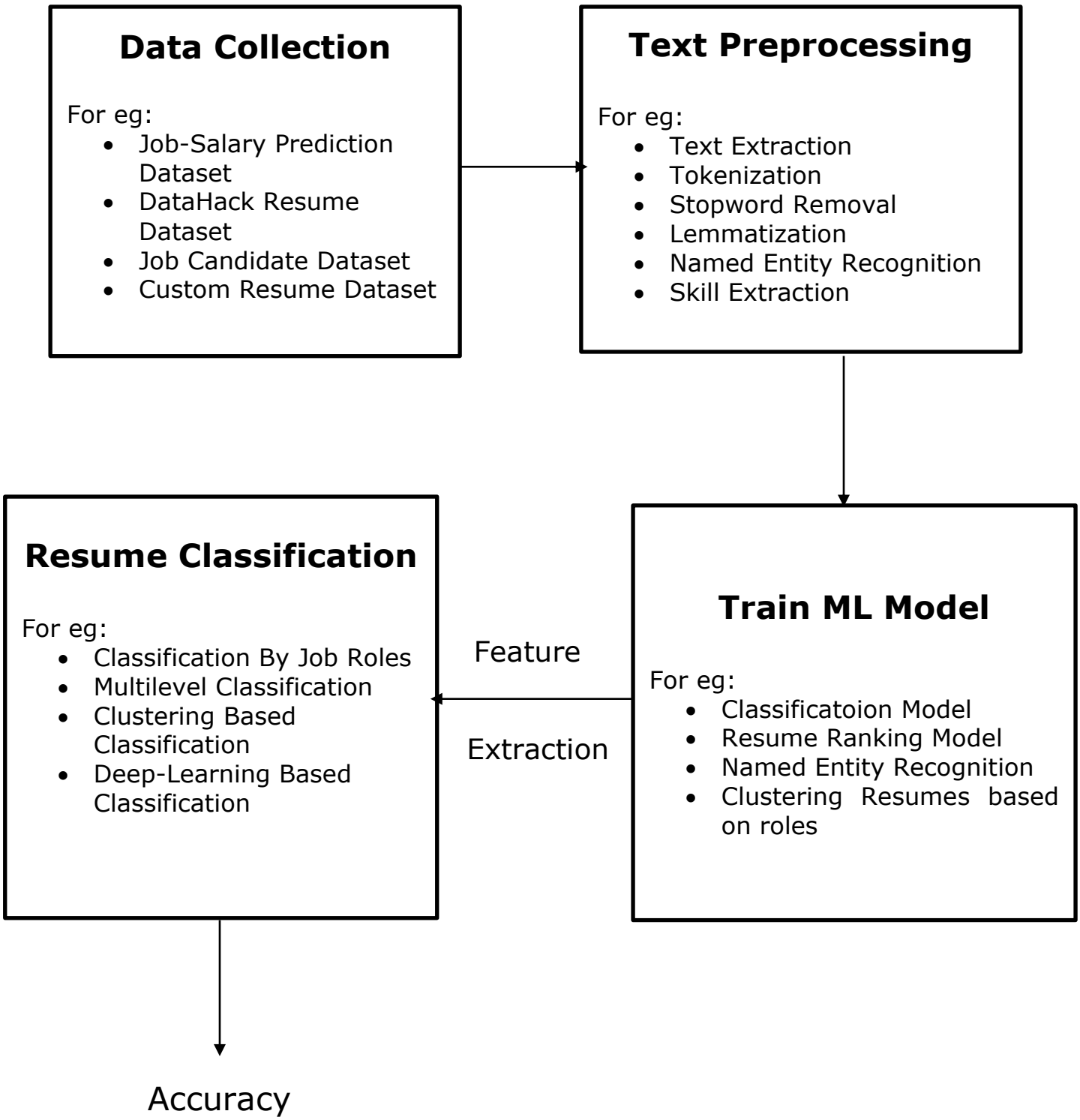
4.Text Vectorization: TF-IDF, Word2Vec, BERT

5.Data Processing: Pandas, NumPy

6.Visualization: Matplotlib, Seaborn

These technologies enable efficient resume parsing, classification, and categorization based on job relevance.

METHODOLOGY



REFERENCES

[1] Saatci, M., Kaya, R., & Ünlü, R. (2024). Resume Screening With Natural Language Processing (NLP). Alphanumeric Journal, 12(2), 121-140

[2] Gan, C., Zhang, Q., & Mori, T. (2024). Application of LLM Agents in Recruitment: A Novel Framework for Automated Resume Screening. Journal of Information Processing, 32(0), 881–893.

[3] Naveed, Z., Nisar, B., Saifullah, D. M., & Iqbal Baig, J. (2024). Resume Ranking Using Natural Language Processing. Journal of Computers and Intelligent Systems, 2(1), 61– 66.

[4] Lalitha, B., Kadiyam, S., Kalidindi, R. V., Vemparala, S. M., Yarlagadda, K., & Chekuri, S. V. (2023). Applicant Screening System Using NLP. 2023 International Conference on Innovative Data Communication Technologies and Application (ICIDCA), 379–383.

[5] Li, S., Li, K., & Lu, H. (2023). National Origin Discrimination in Deep/learning/powered Automated Resume Screening. 1–15.

[6] Mohanty, S., Behera, A., Mishra, S., Alkhayyat, A., Gupta, D., & Sharma, V. (2023). Resumate: A Prototype to Enhance Recruitment Process with NLP based Resume Parsing. 2023 4th International Conference on Intelligent Resume Screening with Natural Language Processing (NLP) | Saatçı et al., 2024 neering and Management (ICIEM), 1–6

[7] Ali, I., Mughal, N., Khan, Z. H., Ahmed, J., & Mujtaba, G. (2022). Resume Classification System using Natural Language Processing and Machine Learning Techniques. Mehran University Research Journal of Engineering and Technology, 41(1), 65–79

[8] Anand, A., & Dubey, M. S. (2022). CV Analysis Using Machine Learning. International Journal for Research in Applied Science and Engineering Technology, 10(5), 1316–1322.