AI-DRIVEN AUTOMATED RESUME GRADING AND CANDIDATE SCREENING

Dr.V.K Shandilya, Anuradha Vidhale, Darshan Rahate, Samrudhi Tatte, Sarvesh Moharil, Amit Tayade, Department of Computer Science and Engineering, SIPNA COET, Amravati

Abstract: The project aims to revolutionize recruitment processes by introducing an automated system that leverages machine learning and natural language processing (NLP) technologies. Despite the rapid advancement of digital technology, many companies still rely on manual recruitment methods or are in the initial stages of adopting online tools. This manual or partially manual approach leads to time and cost overheads, especially when processing a large volume of applicants. To address these challenges, the project proposes an automated recruitment system that streamlines the hiring workflow. By utilizing machine learning algorithms, the system can efficiently evaluate applicant documents such as resumes, educational backgrounds, job experiences. It assigns relative grades to applicants based on these criteria, creating a hierarchical ranking for employers to make informed hiring decisions. Additionally, the system incorporates NLP techniques using libraries like NLTK and Spacy to extract key skills from resumes and match them with job descriptions. This ensures a more accurate assessment of candidate suitability while reducing manual effort and ensuring fair and standardized applicant grading. The ultimate goal of the "Automate Recruitment System" is to standardize applicant grading, reduce complexity, save time in the hiring process, and improve the likelihood of hiring qualified candidates efficiently.

Keywords : Machine learning , Natural language Processing , Natural Language Tool Kit(NLTK) , Resume feature Extraction

I.   Introduction:

Automated recruitment, including resume extraction and test generation, complements existing systems while reducing human intervention. It enhances hiring by automating tasks like screening and evaluation, leveraging algorithms for efficient talent identification and selection, crucial for organizational success and building a strong workforce.

Problem Statement:

Traditional recruitment methods face challenges due to their resource-intensive nature. HR teams invest considerable time and effort in manual processes like resume screening and interviews, leading to delays in hiring. Biases, including age or educational background, can also affect fairness. Managing high application volumes manually results in inefficiency and missed opportunities.

1.1   Need: -

Automated recruitment is an important topic to explore since the recruitment is manually or partial manually processed. In case manual recruitment, important parameters like resume verification, matching eligibility criteria of applicants, tests and interview of candidates are done by HR of the company. Automated recruitment process using machine learning is a technique that uses various algorithms and models to automate the process of extracting resume, selection of candidate on the basis of resume rating

The emergence of automated recruitment systems addresses these pressing needs in the following ways:

- **Efficiency**: Automation reduces the time and effort required to process applications, allowing HR teams to focus on strategic tasks rather than administrative duties.

- **Bias Reduction:** By using predefined criteria and AI algorithms, automated systems can minimize unconscious biases in candidate evaluation, ensuring fair treatment for all applicants.

- **Cost Savings:** Streamlining recruitment processes through automation leads to cost savings by reducing manual labour, minimizing errors, and optimizing resource utilization.

1.2   Objectives: -

The "Automated Recruitment Process using Resume Rating " project aims to achieve the following

objectives:

• Develop and implement machine learning algorithms for automated resume extraction, job profile identification, and categorization of resumes based on skills and experience alignment with job requirements.

• Utilize advanced Natural Language Processing (NLP) techniques to extract relevant information from resumes, match candidate skills with job descriptions, and assess eligibility based on predetermined criteria.

• Create a seamless and efficient recruitment workflow that reduces manual intervention, accelerates the hiring process, and improves overall recruitment outcomes.

By addressing these objectives, the project aims to revolutionize the recruitment process, making it more efficient, fair, and aligned with organizational goals and industry best practices.

II.   Literature Survey

Recruitment processes have evolved significantly over the years, transitioning from traditional manual methods to sophisticated automated systems driven by advancements in technology, particularly machine learning (ML) and artificial intelligence (AI). This literature review explores the evolution of recruitment methodologies, the implementation of automated systems, and various research endeavours in this domain.

Historically, recruitment relied heavily on manual processes such as resume screening, candidate interviews, and assessment by human recruiters. These methods were time- consuming, prone to biases, and often lacked scalability, especially when dealing with large volumes of job applications.

With the advent of digital technologies, there has been a paradigm shift towards automation in recruitment processes. Automated systems leverage ML and AI algorithms to streamline various aspects of recruitment, including resume parsing, candidate matching, skill assessment, and even initial interviews.

Numerous research studies have contributed to the development and enhancement of automated recruitment systems. For instance, (Roy et al., 2020) proposed a system that could work with a large number of resumes for first classifying the right categories using different classifier, once classification has been done then as per the job description, top candidates could be ranked using content-based Recommendation, using cosine similarity and by using K-NN to identify the CVs that are nearest to the provided job description

One of the research focuses on developing an intelligent system using NLP and ML to rank resumes based on client requirements, emphasizing data extraction from diverse resumes, grading based on skills and job descriptions, and integrating social media information for better assessment. It addresses challenges in parsing varied resume formats, ensuring accuracy, and aims to efficiently process large volumes of resumes, extract structured data, and streamline recruitment processes through resume parsing, indexing, and social profile analysis.(Tiwari & Jain Choudhary, n.d.)

Another notable research endeavour by (Sinha et al., n.d.), this paper proposes a two-phase model named "Prospect" based on feature extraction and matching using machine learning. The first phase pre-processes the dataset and extracts resume content by using feature extraction. The second phase applies "selection" and "rejection" classification by applying a matching score algorithm and custom logic. To validate its approach, this paper also designs a unique Prospect dataset with approximately 5,000(thousand) resumes, which incorporates different data sets to generate an unbiased classification output. Experimental result shows that the Prospect model categorizes the resume in "selected" and "rejected" categories with a 93.5% accuracy which improves the overall accuracy by 19.5% compared to convolutional neural network models

III.   Methodology

The methodology for this project revolves around developing an automated resume grading and candidate screening system using machine learning (ML) techniques. The process involves several

critical stages executed sequentially.

[A]   Dataset:

The project uses a dataset of 962 resumes from various professions across 25 categories. The data is in Excel format, with three columns: ID, Category, and Resume. ID is the Resume's sequence number, the Category is the industry sector to which the Resume belongs, and the Resume is the candidate's complete CV. The distribution of domain wise resume is shown in below diagram 3.1

Fig. 3.1 Distribution of Resume Categories
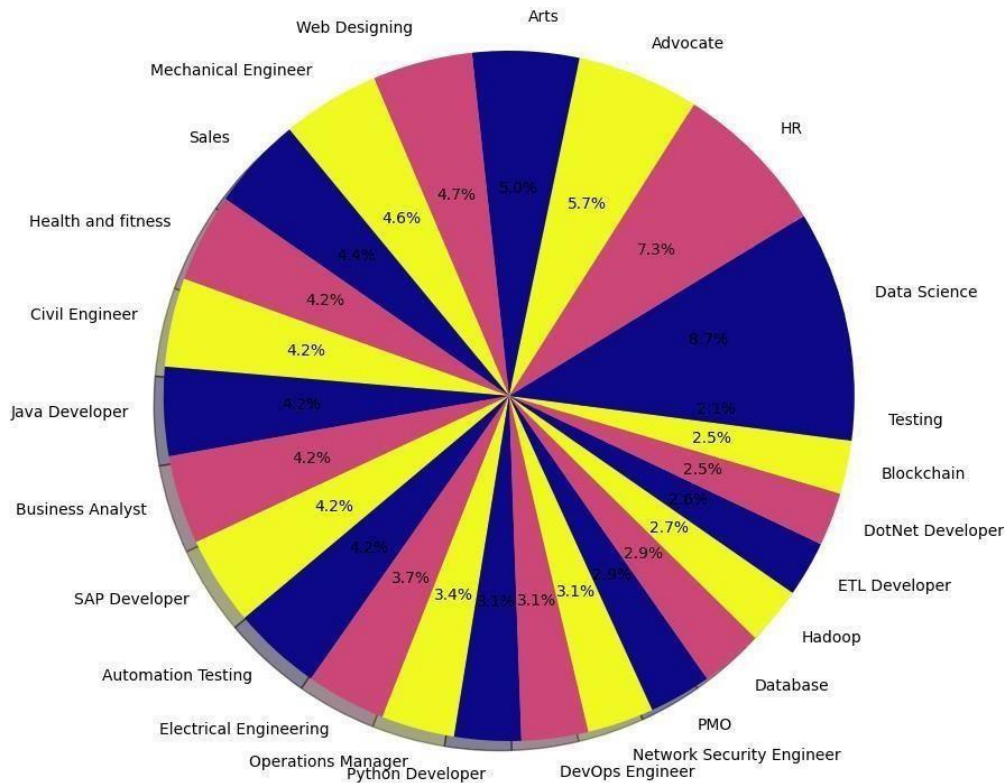
[B]   Phase 01: Prediction of Job Profile/Domain:

We trained a machine learning model to predict the job profile, which will help create a pool of candidates based on the job profile or domain. It helps in:

•   The first stage of screening resumes.

•   Job recommendation feature for the user, which helps him find out about specific job posts

Step 1: Data Preprocessing

The initial stage of our proposed system focuses on utilizing natural language processing to extract relevant information from resumes. However, the information in resumes is often unstructured and may contain extraneous, inconsistent, or irrelevant data that is useless to recruiters. To address this issue, the preprocessing techniques we employed are:

i)   Remove URLs, spaces, hashtags, special characters, and numbers using Python RegEx:

ii)   NLTK in Python: removing stop words

This preprocessing generates a more condensed version of each Resume that can be efficiently utilized in the subsequent phases of the resume screening system. This approach allows for further processing and analysis of the resumes, enhancing the system's effectiveness.

Step 2: Vectorization

Machine learning algorithms often use numerical data, so when dealing with textual data or any natural language processing (NLP) task, a sub-field of ML/AI dealing with text, that data must first be converted to a vector of numerical data by a process known as vectorization.

i)   We encode category columns using label encoding.

Label encoding in Python can be achieved using the scikit-learn library. Scikit-learn provides a very

efficient tool for encoding the levels of categorical features into numeric values. Label Encoder encodes labels with a value between 0 and n_classes-1, where n is the number of distinct labels.

ii)   Resume data vectorization using TF-IDF.

TF-IDF (Term Frequency—Inverse Document Frequency) is a handy algorithm that uses word frequency to determine how relevant a word is to a given document.

Step 3: Building a Model

i)   The dataset is divided into a training set (80%) and a testing set (20%) to ensure the models are trained on a sufficient amount of data and evaluated on a different set of data.

ii)   Here, we are dealing with the supervised multilevel classification problem in machine learning. The algorithms we are using to train the model are:

- K-Nearest Neighbours:
- Logistic Regression:
- Support Vector Machine:
- Random Forest:
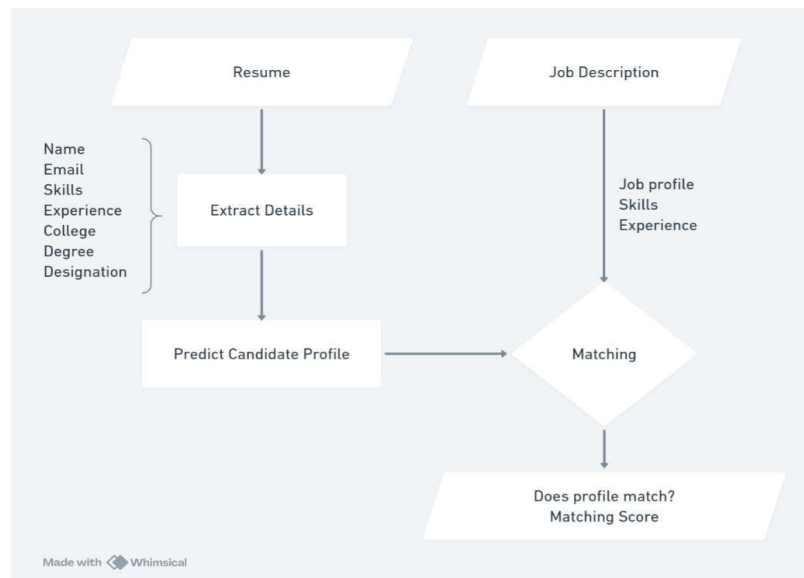
Fig. 3.2 System Work Flow Diagram

[C]   Phase 02: Resume Scoring:

Step 1: Extracting Text from PDF:

The system utilizes pdfminer3 to extract text from candidate resumes in PDF format, ensuring data accessibility and processing.

Step 2: Extracting Skills:

NLP libraries NLTK and Spacy are leveraged to extract relevant skills from resumes, enhancing the accuracy of candidate evaluation.

Step 3: Calculating the Score

A scoring mechanism is implemented based on the number of matching skills between candidate profiles and job descriptions. This scoring system assigns a percentage score, reflecting the candidate's qualifications and experience relevant to the job requirements.

IV.   Result Analysis

The results from each model are mentioned in Table 1. so that they can be compared and useful conclusions may be drawn from them. An 80-20 split was used to create training and testing sets through a stratified division of the data

The methodology outlined above ensures a systematic approach to developing an automated resume grading and candidate screening system. It incorporates data preprocessing, model building, and scoring mechanisms to identify suitable candidates efficiently and accurately from the provided

dataset.

Table. 4.1 Comparison of categorization models

Algorithm	Accuracy on training data (%)	Accuracy on test data (%)
KNeighborsClassifier	98.56	98.44
LogisticRegresssion	99.73	99.48
SVM	100	99.48
RandomForestClassifier	100	99.48

V.   Conclusion

The development of an automated Resume Grading and candidate screening system using machine learning (ML) techniques is a significant leap in recruitment and talent acquisition. This project's methodology, structured into two phases, effectively utilizes NLP and ML algorithms to streamline candidate selection. In the first phase, focusing on data preprocessing and model building with algorithms, SVM with 99.48 accuracy yielded accurate predictions of job profiles and suitable grades for candidate resumes based on qualifications.

The second phase, implementing resume scoring mechanisms and NLP libraries, enhanced candidate evaluation against job requirements. The system's outcomes demonstrate its potential to revolutionize recruitment, reducing manual efforts, and improving efficiency by streamlining candidate screening and offering job vacancy recommendations.

Overall, this project signifies a step forward in leveraging technology to optimize talent acquisition processes, offering a scalable and efficient solution for organizations seeking to enhance their recruitment strategies. The methodology employed serves as a robust framework for future research and development in the field of automated recruitment systems.

References:

[1]   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. https://doi.org/10.1016/j.procs.2020.03.284

[2]   Sinha, A. K., Amir, M., Akhtar, K., Kumar, M., & Upadhyay, S. (n.d.). Resume Screening Classification using Artificial Intelligence and Natural Language Processing Section A-Research paper ISSN. *Eur. Chem. Bull*, *2023*, 14473–14480. https://doi.org/10.48047/ecb/2023.12.si4.1309

[3]   Tiwari, V., & Jain Choudhary, S. (n.d.). Intelligent Hiring with Resume Parser and Ranking Using Machine Learning and Natural Language Processing. In *International Journal of Innovative Research in Computer and Communication Engineering*.