

Project Report: Automated Resume Screening System

The Organization/Department and Its Problem

The Human Resources (HR) department plays a vital role in an organization's recruitment process by evaluating candidates based on resumes. The challenge arises when HR departments receive a large volume of resumes, making it difficult to screen them efficiently. Manually reviewing thousands of resumes is time-consuming, subjective, and prone to human error, which can result in overlooking qualified candidates. The HR department aimed to address this issue by developing an automated system using Machine Learning (ML) to efficiently classify resumes into job roles.

Problem Statement

The problem this project addresses is the inefficiency and bias involved in manual resume screening. The goal is to develop a machine learning-powered web application that automatically classifies resumes into predefined job roles (e.g., Software Engineer, Data Scientist, HR). This system will help HR teams filter candidates more quickly and accurately, reducing human bias and streamlining the hiring process.

How Machine Learning (ML) Can Solve the Problem

Machine Learning and Natural Language Processing (NLP) techniques are used to automate the resume screening process. The resumes undergo text preprocessing, where stop words, punctuation, and inflected words are removed or normalized. Then, Term Frequency-Inverse Document Frequency (TF-IDF) is used to convert textual data into numerical features, emphasizing important terms in the resumes.

We use multi-class classification to train the model, enabling it to categorize resumes into specific job roles. This model learns from a labeled dataset and can predict job categories for new resumes uploaded by users. By automating this process, the system helps HR professionals save time and make data-driven, unbiased hiring decisions, allowing them to focus on critical tasks like interviews. We used Random Forest model and logistic regression

of which we found that Random Forest was more fitting for our project, and we used it as the primary model.

Challenges Faced During Data Collection/Preprocessing

A significant challenge in developing this system was acquiring a diverse and labeled dataset. Resumes are confidential, and publicly available datasets were limited or not properly formatted. Data cleaning and preprocessing were necessary to standardize resumes and extract meaningful information.

Another issue was dealing with unstructured resume formats, such as PDFs, which required conversion into machine-readable text. Additionally, the dataset was imbalanced, with certain job roles underrepresented, which could lead to biased predictions. To address this, the dataset was balanced, and overfitting was mitigated through cross-validation and regularization techniques.

Reflection on Ethical Considerations

The ethical implications of automating resume screening must be carefully considered. Data privacy is a primary concern, as resumes contain sensitive personal information. Ensuring that resumes are processed securely and with the consent of applicants is essential. Organizations must implement measures to protect this data and comply with privacy laws, such as GDPR.

Another ethical challenge is bias in the model. If the training data is biased such as underrepresenting certain groups, the model may perpetuate these biases, leading to unfair outcomes. To prevent this, balanced datasets were used, and model performance was evaluated on multiple metrics to ensure fairness and avoid overfitting.

Conclusion

The automated resume screening system using Machine Learning and Natural Language Processing provides an effective solution to the challenges faced by HR departments. It improves the efficiency of the recruitment process by automating resume categorization,

reducing bias, and saving time. Although challenges such as data collection and ethical concerns must be addressed, this system offers a scalable, cost-effective way to improve hiring decisions. By ensuring data privacy and fairness, this system can be responsibly implemented to benefit both recruiters and applicants.