ABSTRACT

Automated resume categorization streamlines the recruitment process by leveraging machine learning and natural language processing techniques. Uploaded PDF resumes are processed to extract and clean textual content, enabling accurate feature extraction using the TF-IDF method. A pre-trained Logistic Regression classifier predicts the most relevant job category for each resume based on its content. The application, developed with Streamlit, offers a user-friendly interface for batch uploading and categorization. Each resume is sorted into a corresponding category folder, and years of experience are extracted for additional insights. Results are presented in a tabular format and can be exported as a CSV file for further analysis. The system handles multiple files efficiently and provides robust error management for unsupported or ambiguous inputs. Integration of machine learning models with a web interface demonstrates practical HR automation. Organizations benefit from reduced manual effort and improved consistency in resume screening. The solution is scalable and adaptable to various recruitment needs. Customizable category mappings allow the system to be tailored for different industries or job roles. The modular design supports easy updates to the classification model as new data becomes available. Security and privacy considerations are addressed by processing files locally without uploading sensitive information to external servers. Comprehensive test cases ensure reliability across a range of input scenarios, including edge cases and error conditions. The project exemplifies the effective application of data science to real-world human resources challenges, offering a foundation for further enhancements such as skill extraction or candidate ranking