

Project Report: Resume Classification System

Project Overview

The Resume Classification System is a project aimed at automating the process of categorizing resumes into different job categories based on their content. This system utilizes Natural Language Processing (NLP) and Machine Learning techniques to analyze the text of resumes and predict the most suitable job category for each resume.

Project Components

The project consists of the following key components:

1. **Data Collection and Exploration:** The project starts with the collection of a dataset containing resumes from various job categories. In this case, a dataset named "UpdatedResumeDataSet.csv" was used, which contains resume text and their respective categories.
2. **Data Preprocessing:** Before training the machine learning model, the resume text data is preprocessed to clean it. The preprocessing steps include:
 - Removing URLs, hashtags, mentions, and special characters.
 - Converting all text to lowercase.
 - Tokenizing the text into words.
 - Removing stopwords.
3. **Data Visualization:** Data exploration is performed to understand the distribution of resumes across different job categories. Visualizations, such as bar charts and pie charts, are created to provide insights into the data.
4. **Feature Engineering:** Text data is transformed into numerical features using the Term Frequency-Inverse Document Frequency (TF-IDF) vectorization technique. This step converts the resume text into a numerical format suitable for machine learning.
5. **Model Building:** A machine learning model is trained to classify resumes into job categories. In this project, a K-Nearest Neighbors (KNN) classifier is used as the base estimator within a One-Vs-Rest (OvR) multi-class classification framework.
6. **Model Evaluation:** The model's performance is evaluated using accuracy as the evaluation metric. The accuracy score measures how well the model predicts the correct job category for each resume.

7. **Model Deployment:** The trained model, along with the TF-IDF vectorizer, is saved using the Pickle library for future use.
8. **Streamlit Web Application:** A web-based interface is developed using Streamlit, a Python library for creating web applications. Users can upload their resumes, and the system predicts the most suitable job category for the uploaded resume.

Project Results

The Resume Classification System achieved a high accuracy rate of approximately 99.17% on the test data, indicating that it effectively categorizes resumes into job categories.

The web application built using Streamlit provides a user-friendly interface for users to upload their resumes and obtain predictions for job categories. This makes it a practical tool for both job seekers and recruiters to streamline the resume screening process.

Future Enhancements

While the project has been successfully implemented, there are several potential enhancements and future directions:

1. **Improving Model Robustness:** Explore more advanced machine learning algorithms and deep learning techniques to improve model accuracy and robustness.
2. **Handling Imbalanced Data:** Address the issue of imbalanced data by using techniques such as oversampling or using more extensive datasets.
3. **NLP Techniques:** Experiment with advanced NLP techniques like word embeddings (Word2Vec, GloVe) and pre-trained language models (BERT, GPT) for better feature extraction.
4. **User Authentication:** Implement user authentication and authorization in the web application to ensure data privacy and security.
5. **Feedback Mechanism:** Add a feedback mechanism for users to provide feedback on the accuracy of predictions, allowing continuous model improvement.
6. **Mobile Application:** Develop a mobile application version of the system to make it more accessible to a broader user base.

Contact Information

- Project Completed by: Rakat Murshed
- Email: rakatmurshed13@gmail.com
- WhatsApp: +8801893395385

Conclusion

The Resume Classification System is a valuable tool for automating the process of categorizing resumes, making it easier for recruiters and job seekers to match candidates with suitable job positions. The high accuracy of the system, coupled with the user-friendly web interface, makes it a practical and efficient solution for resume screening in the job market.