Automated Job Resume Screening using NLP

By: Anusha Reddy Aerva

ID: B00118232

Project Overview

Problem:

- Manually screening resumes is time-consuming and prone to bias.
- Automating the process using machine learning improves efficiency and fairness.

Data:

- Resume dataset from Kaggle with text-based resumes categorized into different job roles.
- Preprocessed using tokenization, stopword removal, and stemming.

Model & Algorithm:

- TF-IDF vectorization for feature extraction.
- Implemented models: Naive Bayes, Logistic Regression, SVM, and Random Forest.
- Hyperparameter tuning for SVM and Random Forest.

Methodology & Implementation

Data Preprocessing:

- Lowercasing, punctuation removal, stopword filtering, stemming.
- TF-IDF applied to extract numerical features.

Model Training & Evaluation:

- Naive Bayes: Custom implementation using log probabilities.
- Logistic Regression: Gradient descent optimization.
- SVM: Hyperparameter tuning using GridSearchCV.
- Random Forest: Feature importance analysis.

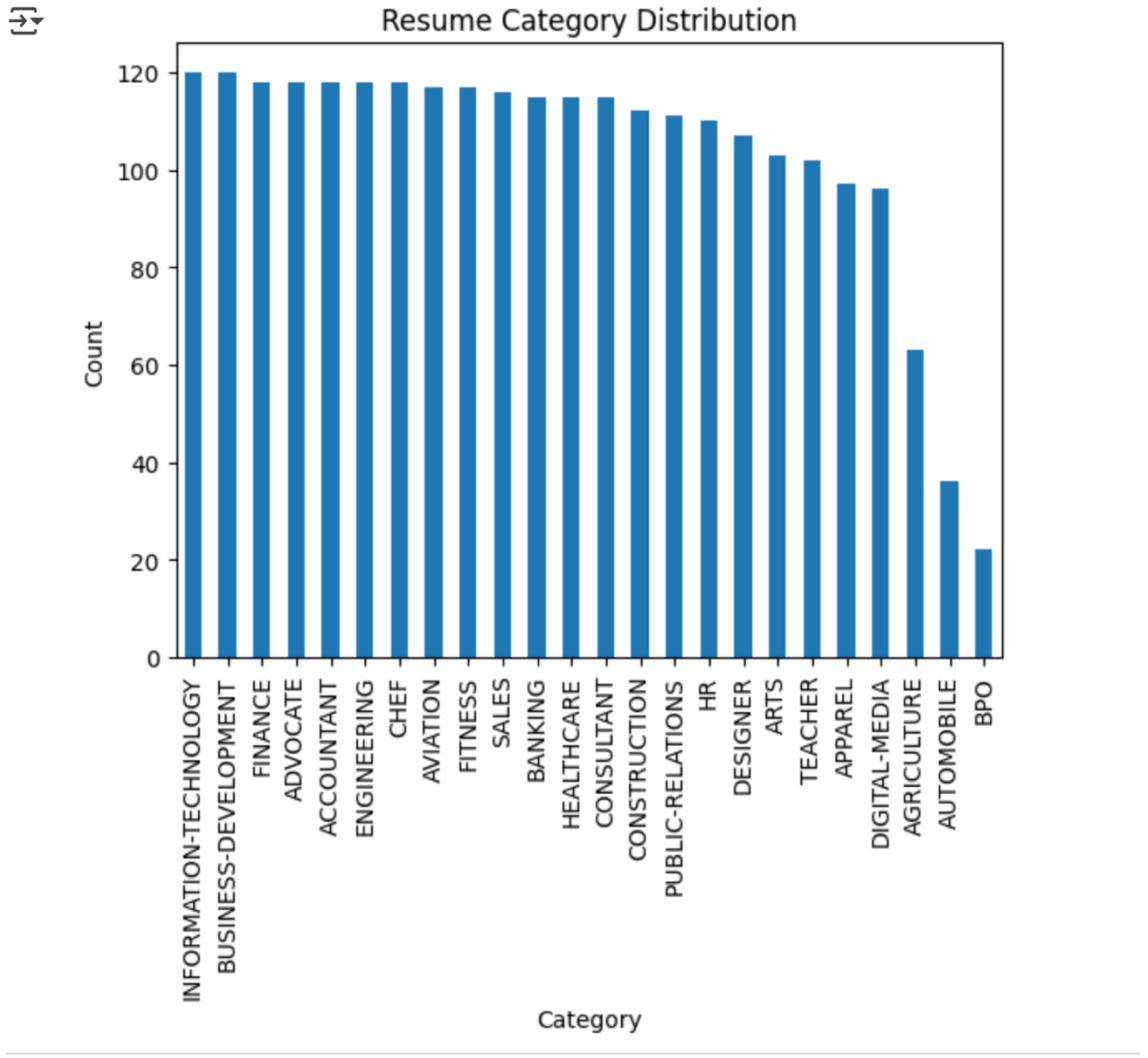
Performance Metrics:

- Accuracy, classification reports used for evaluation.

Results & Insights

•The dataset contains resumes from various job categories.
•The distribution of categories is visualized in the following bar chart, highlighting the class imbalance.
•Feature importance charts show key words influencing predictions.

Future work: Exploring deep learning models like BERT for better context understanding.



Thank you