Automated Job Resume Screening using NLP

By: Anusha Reddy Aerva

ID:B00118232

Problem, Data, and Model

Problem:

- Manual resume screening is time-consuming and subjective. Automating this process can enhance recruitment efficiency and consistency.

Dataset:

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

Approach:

- Preprocessing with NLTK & SpaCy
- TF-IDF vectorization
- Models used: Custom Naive Bayes, Custom Logistic Regression, SVM, Random Forest

Training & Model Performance

Training:

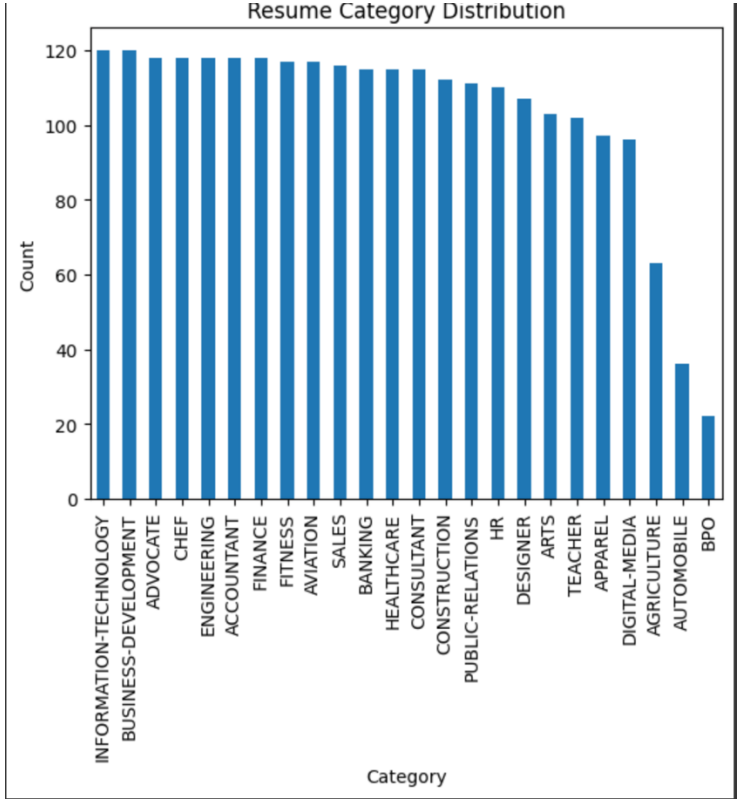
- Text preprocessing (tokenization, stemming, stopword removal)
- TF-IDF Vectorizer
- 80/20 train-test split

Performance Metrics

- Naive Bayes: Accuracy = 54%
- Logistic Regression: Accuracy = 60%
- SVM (best): Accuracy = 63%
- Random Forest: Accuracy = 67%

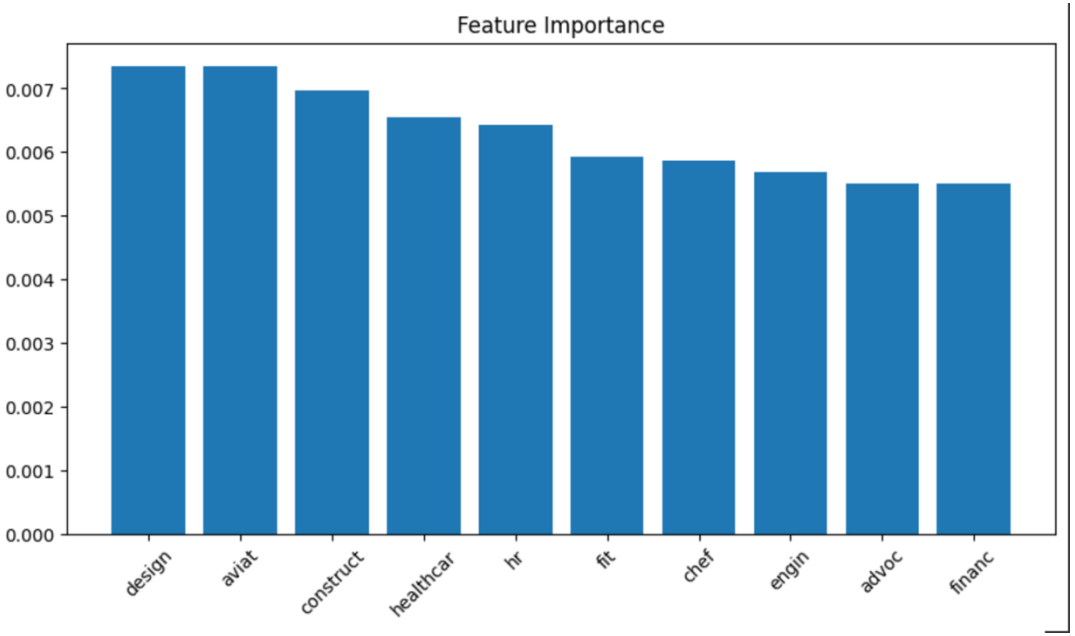
Hyperparameter Tuning:

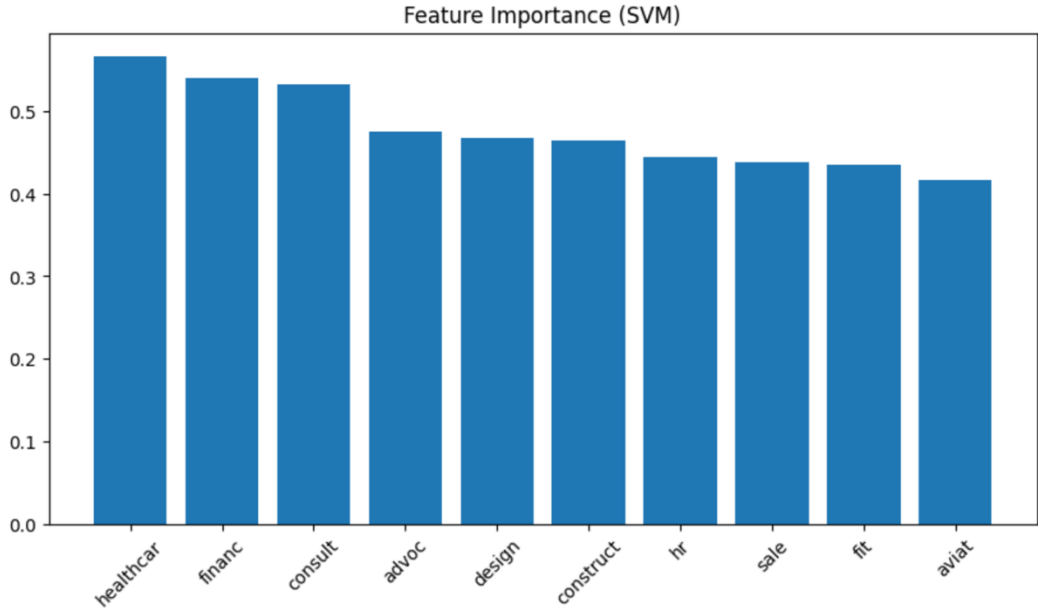
- GridSearchCV used for SVM and RF



Model Insights & Live Prediction

- Feature Importance: Top influential terms visualized from SVM & RF models





Conclusion

Key Takeaways:

- Automated classification can significantly improve resume screening
- RF outperformed other models in this setup
- Custom implementations allowed deeper model understanding

Future Work:

- Integrate with a web-based HR portal

Thank You