Resume Screening App: Project Documentation

Recruitment processes in modern organizations face growing challenges due to the overwhelming volume of job applications received for each open position. Manually reviewing and categorizing hundreds or thousands of resumes is both time-consuming and prone to human error or bias. To address this inefficiency, this project proposes a **Resume Screening Web Application** powered by **Machine Learning (ML)** and **Natural Language Processing (NLP)**. The system is designed to automatically analyze resumes and classify them into appropriate job categories such as Data Scientist, Software Developer, Web Designer, HR, and more.

The core of this project lies in the development of a machine learning model that can learn patterns from a labeled dataset of resumes and predict the job role that best fits a newly uploaded resume. The app allows users to upload their resumes in .pdf or .txt format. Once uploaded, the resume content is processed and passed through a trained classifier which returns the predicted job role.

Data Preprocessing and Feature Extraction

The solution begins with an essential **data preprocessing** step. Resume texts are cleaned by removing stopwords, punctuation, and irrelevant characters. Using the **Natural Language Toolkit (NLTK)**, lemmatization is applied to normalize words into their root forms. Then, **Term Frequency-Inverse Document Frequency (TF-IDF)** vectorization is used to convert the cleaned text into numerical features suitable for machine learning algorithms. After feature extraction, a multi-class classification model is trained using algorithms such as **Random Forest** and **Logistic Regression**. The model is evaluated using appropriate performance metrics and then saved for deployment in the web application.

Technologies Used

The technologies used in the development of this system include:

- **Python** as the primary programming language.
- **Scikit-learn** for machine learning.

- **NLTK** for NLP preprocessing.
- **Pandas** and **NumPy** for data handling.
- **Streamlit** for building the user interface.
  The trained model and the vectorizer are saved using Python's joblib module for efficient reuse during prediction.

User Interaction with the App

From the user's perspective, the application is simple to use. After downloading the project folder named Intro_AI_Project_Group4, the user navigates to the streamlit_app directory and runs the training script **train_models.py** to build the model. Once the model is trained, they can launch the application using the command:

python -m streamlit run resume_classifier_app.py

This is how the Streamlit app looks. The user can browse and upload the file
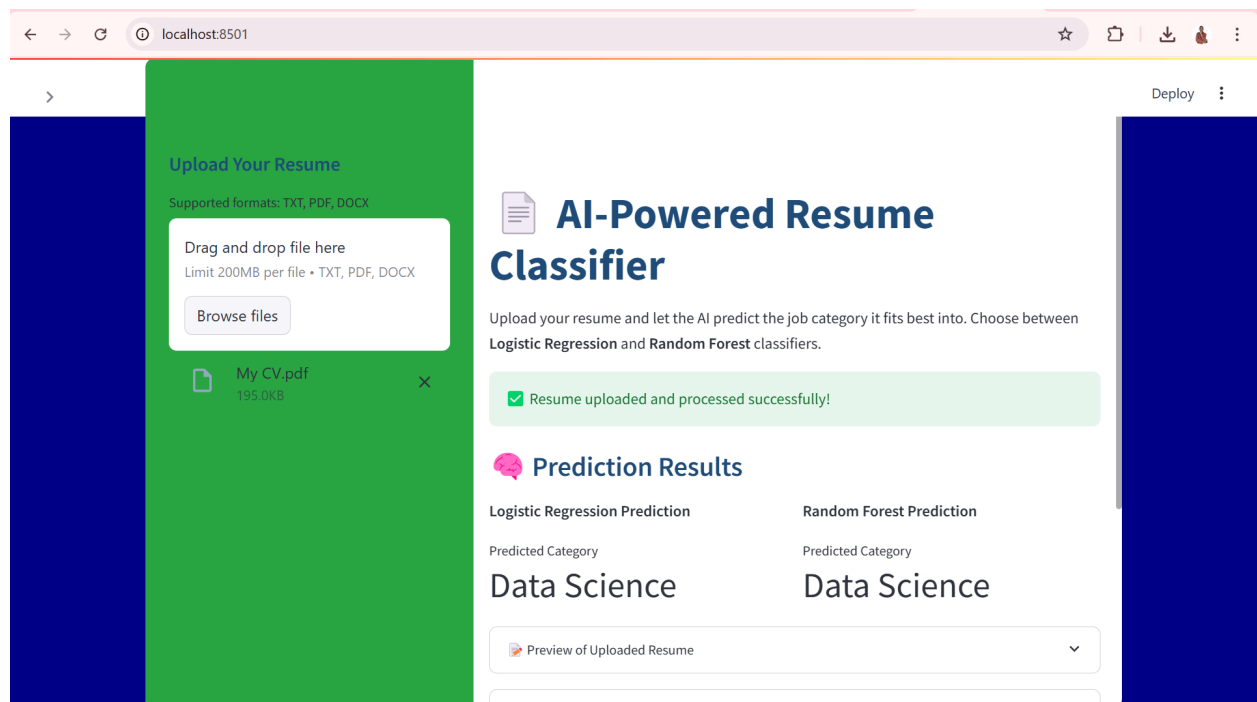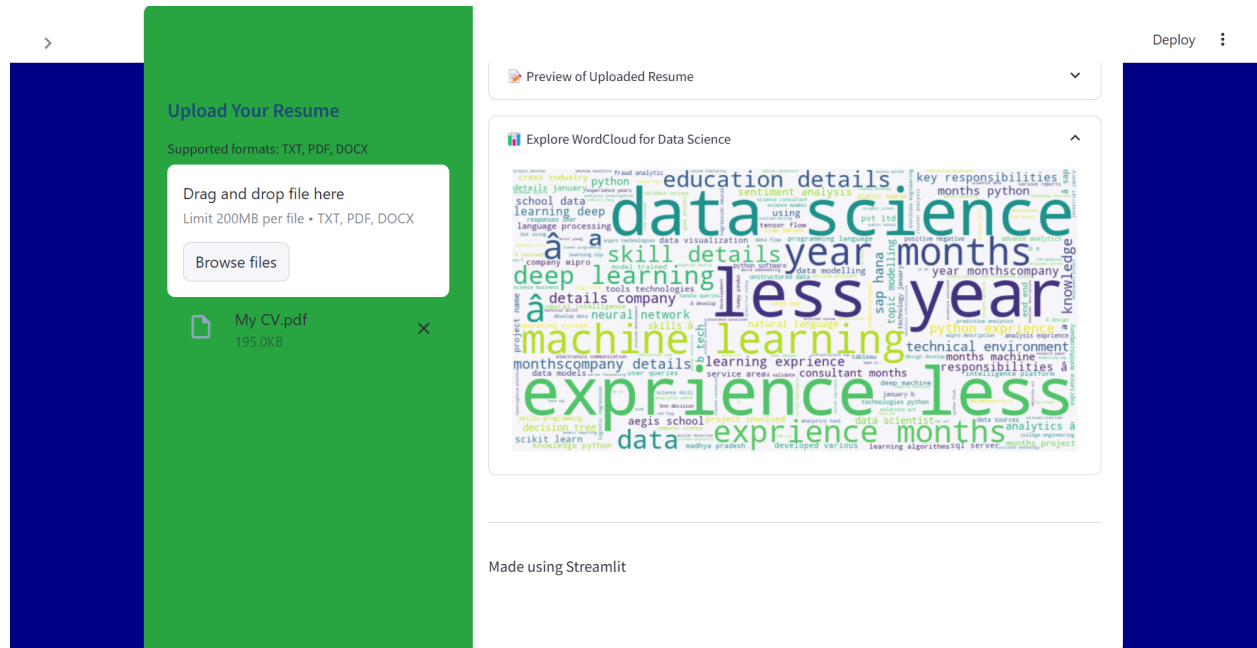
The app opens in the browser, allowing users to upload resumes and receive real-time predictions.

Challenges Faced

However, the project encountered some challenges. These included handling unstructured data in resumes, especially PDFs with formatting issues, and class imbalance in the dataset where certain job roles had significantly more samples than others. Additionally, fine-tuning the model to maintain high precision without overfitting was another critical hurdle.

Ethical Considerations

Ethical considerations were also taken into account during the design and development process. The app is designed not to store or misuse any personal data from uploaded resumes, ensuring **data privacy**. **Bias mitigation** was also considered by focusing the model purely on skill-related text, avoiding sensitive fields like gender or name. Furthermore, the model's predictions are designed to be explainable and transparent, aligning with **responsible AI practices**.

Conclusion

In conclusion, the **Resume Screening App** demonstrates how **machine learning** and **NLP** can be harnessed to solve real-world problems in **human resource management**. It not only improves the efficiency of resume screening but also adds consistency and objectivity to the recruitment process. As the system evolves, it can be expanded to include more job categories, integration with **applicant tracking systems (ATS)**, and additional features such as feedback or ranking mechanisms. This project marks a significant step towards **data-driven recruitment practices** that benefit both employers and job seekers.