**Documentation and instructions:**

In the context of CV (Curriculum Vitae) domain classification, I began by acquiring the Cagle dataset through the provided link and subsequently set up the necessary libraries. Such as protobuf, pdf plumber and re, etc.

Then I initiated the process of extracting pertinent information from PDF files. Then extract specific sections such as Experience and Education, from the CVs. A CSV file was generated to accommodate this formatted data, and the extracted details—Experience Education, and Skills—were stored in another CSV file. Continuing the workflow, the formatted data was appended to the new CSV file.

The dataset was then partitioned into two segments: 80% for training and 20% for testing. This process included encoding the data, updating the CSV file, and performing the necessary data splitting, all in preparation for subsequent training and testing stages. With the tokenized training and testing data, along with corresponding labels, I proceeded to employ the BERT tokenizer.

The tokenized dataset was loaded, after which I compiled a model and commenced training using the 'distillery-base-uncased' model architecture. For the training process, a batch size of 16 and a learning rate of 1e-7 were utilized. Post-training, the model's performance was evaluated using metrics such as accuracy, precision, F1 score, and recall.

This evaluation step provided insights into the model's efficacy in classifying CV domains. Subsequently, the trained model was saved, ensuring its retention for future use and analysis.

In summation, this process encapsulated the complete workflow of CV domain classification—starting from data acquisition, extraction, and formatting, through training and evaluation, culminating in the creation of a trained model for accurate CV domain classification.