This is the readme file for the “Audit Engineers & Brains“ team.

We are using the following model - <https://huggingface.co/distilbert/distilbert-base-uncased-finetuned-sst-2-english>.

This model reaches an accuracy of 91.3 on the dev set.

To use this code, you'll need to:

1. Install required packages:

```bash

pip install pdfplumber pandas transformers torch nltk regex

```

2. Download NLTK data:

```python

import nltk

nltk.download('punkt')

```

3. In the code update the following variables

Input\_dir - Put the directory of all the dataset at this location