neoshare

Programming Task: Production-Ready Document Classification

Objective: Develop a production-ready document classification system in Python that not only classifies documents accurately but also emphasizes the ability to explain the chosen model and its structures. The solution should be invoked through a command-line interface (CLI) and must handle only one text file at a time. Utilize the provided dataset for training and testing: Document Classification Dataset.

Requirements:

1. Implement a **robust** document classification system using an advanced machine learning model of your choice. Choose the features and preprocessing steps thoughtfully.
2. Develop a command-line interface (CLI) using the `argparse` or `typer` library. The main script should be invoked with a command like the following:

```
python main.py --file "document.txt"
```

The script should take a text file (`document.txt`) as input and output the predicted category for the document.

3. Structure the code to be sort of production-ready, including proper error handling, logging, and adherence to best practices.
4. Include a file listing all necessary dependencies (e.g. project.toml, requirements.txt)

Submission: Submit the source code, any necessary files behind the document classification model. You can send it to us per zip file or via a private git repository.