==========

MANUAL

==========

📁 Project Directory Structure

BERT/

```
│
├── BERT.py                                  # BERT classifier (HuggingFace + PyTorch)
├── NB+TF-IDF.py                             # Naive Bayes classifier with TF-IDF
├── Sensitivity Analysis of Parameters.py    # Sensitivity analysis on parameter impact
├── Title+Body.csv                           # Input dataset (text + sentiment)
│
├── requirements.txt                         # Python dependencies (plain text)
├── requirements.pdf                         # Python dependencies
├── manual.pdf                               # User manual for using the project
├── replication.pdf                          # Step-by-step replication guide
│
├── README for NB+TF-IDF.md                  # Markdown guide for NB+TF-IDF
│
├── results/                                 # Folder for data tested by the author
└── datasets/                                # Folder for datasets
```

📌 What Each File Does

- `BERT.py`:
  Trains and evaluates a BERT-based text classifier using Transformers.
  Input: Title+Body.csv
  Output: Evaluation metrics and result CSV

- `NB+TF-IDF.py`:
  Implements a baseline Naive Bayes model using TF-IDF vectorization.

- `Sensitivity Analysis of Parameters.py`:
  Tests how different parameters (e.g., learning rate, batch size) affect model performance.

📄 Dataset Format

File: `Title+Body.csv`

Must include at least:

- `text` column (string): bug report content (title + body)
- `sentiment` column (int): label (e.g., 0 or 1)

🧪 Running Each Script
---
1. Install dependencies (once):
    pip install -r requirements.txt

2. Run BERT classification:
    python BERT.py

3. Run Naive Bayes + TF-IDF:
    python "NB+TF-IDF.py"

4. Run sensitivity analysis:
    python "Sensitivity Analysis of Parameters.py"

📊 Outputs

- Result files will be saved under `BERT/`
- Outputs metrics like Accuracy, Precision, Recall, F1, AUC

✅ Notes

- Python ≥ 3.8 recommended
- GPU is optional but improves BERT training speed
- You can modify hyperparameters directly in each `.py` script
- All scripts assume `Title+Body.csv` is in the root directory

📬 Need Help?

If anything goes wrong, double-check:
- File names (especially `"NB+TF-IDF.py"`)
- Required columns in the dataset
- Python environment / dependencies

Happy experimenting! 🔬