One-Page Instruction Document

Improving Dynamic ESG: ESG Signal Detection from News Articles

Student: Lipika Basumatary | Student ID: 1127148
Teammate: Cissy | Student ID: 1127192

Short Project Summary

We built and compared models to classify ESG news articles into Environmental (E), Social (S), and Governance (G) using the ESG-FTSE corpus. We evaluated a TF-IDF + Logistic Regression baseline and transformer models (BERT/DistilBERT/XLM-RoBERTa). The best-performing model was tuned XLM-RoBERTa.

GitHub Repository Link:

<PASTE YOUR GITHUB LINK HERE>

Execution Instructions (Reproduce Results)

1) Clone the repo: git clone <repo-link>
2) Install requirements: pip install -r requirements.txt
3) Place dataset file at: data/esg_ftse_corpus.json (data/ is gitignored)
4) Run the notebook: notebooks/ESG_ratings_Final_Project.ipynb (Colab recommended)
5) Verify outputs: metrics printed + results/xlm_roberta_tuned_predictions.csv

Special Notes

- Tuned XLM-R settings: epochs=5, learning_rate=5e-5, weight_decay=0.001
- Use GPU for transformers (Colab T4 recommended).
- If dataset sharing is restricted, include the official download link in README (Section 4).