Enhancing Stock Market Prediction using Extended News and Transformer Models

Team Members:

Muhammed Mustafa 22101336

Ramez Ezzat 22100506

Noureen Muhammed 22101109

Rokaia Islam 22101084

Introduction

- Why predict stock market trends?
- Traditional methods:
  - Technical analysis: Based on past stock prices/volumes.
  - Fundamental analysis: Based on company data.
- **Problems with Traditional Methods:**
  - Manual, slow.
  - Miss important external (news) events.

Problem with Previous Work

- Headline sentiment is **too simplistic**.
- Manual feature engineering **limits model learning**.
- Cannot fully capture **market complexity**.

Research Paper Solution

- **Full article embeddings** (FinBERT + TinyBERT, instead of headlines only).
- **1D convolutions** on raw financial metrics (no manual indicators).
- **LoRA** fine-tuning for lightweight adaptation.

Baseline vs TinyBERT+LoRA

Metric	StockFormer (Baseline)	TinyBERT+LoRA (Research Paper)
R^2	0.9956	0.9986
Directional Accuracy	57.66%	62.50%
Normalized MSE	0.004659	0.000113

Our Project Plan

- **Re-implement** TinyBERT+LoRA baseline.
- **Expand dataset**: Use 12+ months of news and stock data. (if possible)
- **Improve models**: Try DeBERTa / DistilBERT / Ensemble models.
- **Feature fusion**: Smarter integration of text and stock signals.
- **Adding reinforcement learning (RL)** for dynamic decision-making.

Our Innovation

- Extend history window: **20 days** instead of 9.
- Use **larger pre-trained models** (DeBERTa / DistilBERT).
- Optional: **Ensemble** multiple transformer models.
- Optional: Add new technical indicators (MACD, RSI, etc.)

Planned Steps

1. Rebuild original TinyBERT+LoRA model.
2. Train and evaluate using 5-month dataset (fair comparison).
3. Extend dataset to 12 months.
4. Train improved models.
5. Analyze and compare results.

Dataset Details

- **News Sources:** Bloomberg, CNBC, Yahoo Finance.
- **Stocks:** FAANG (Meta, Amazon, Netflix, Google, Apple).
- **Stock Metrics:** Open, High, Low, Close, Volume from Yahoo Finance.

Model Architecture

- **Baseline:**
  - TinyBERT + 1D Convolutions + LoRA.
- **Innovation:**
  - Longer news history input.
  - DeBERTa or DistilBERT.(for speed and more accuracy)
  - Potential ensemble learning.
- **RL agent for decision-making to decide.**
  - Buy/Hold/Sell based on model predictions.
  - Risk management (stop-loss, position sizing).

Evaluation Metrics

- MSE (Mean Squared Error)
- R^2 Score (Coefficient of Determination)

- Directional Accuracy (price up/down)
- Optional: Mean Absolute Error (MAE)

Team Roles

Team Member	Task
Member 1	Rebuild baseline TinyBERT+LoRA model
Member 2	Extend dataset (download extra articles, stock data)
Member 3	Build and train improved models (DeBERTa, Ensemble)
Member 4	Write and format project report (LaTeX)

Timeline

Week	Task
Week 1	Baseline model + dataset extension
Week 2	Train innovative models + experiment tracking
Week 3	Analysis, visualizations, error analysis
Week 4	Finalize report and code submission

Expected Results

- Higher prediction accuracy vs baseline.
- More stable performance over longer periods.
- Potential opportunity for conference publication (top teams).

Conclusion

- We combine **data expansion**, **stronger models**, and **smarter feature engineering**.
- Targeting **maximum marks** in innovation, analysis, and execution.
- Clear experiments, effective teamwork, and professional reporting.