

News Sentiment Analyzer for Stock Market Prediction

The **News Sentiment Analyzer** aims to interpret the sentiment of financial news headlines to understand market emotions and predict potential stock movements. Sentiment — whether **positive**, **neutral**, or **negative** — plays a key role in shaping investor behavior. This project combines transformer-based contextual embeddings with machine learning to form a robust sentiment classification pipeline. It builds on state-of-the-art NLP methods to represent financial text meaningfully and predict sentiment trends that can inform trading strategies.

Dataset Description

The dataset used is **mltrev23/financial-sentiment-analysis** from Hugging Face. It includes financial headlines labeled with sentiment polarity. Each record contains a sentence describing market or company-related news and its sentiment label among three classes — positive, neutral, and negative. A subset of **500 samples** was used for model development: **400** for training and **100** for testing.

Methodology

The methodology consists of three integrated components:

- Text Embedding using FinBERT:** Financial headlines were tokenized and converted into high-dimensional embeddings using the **ProsusAI/FinBERT** model — a domain-specific variant of BERT trained on financial texts. Each sentence is represented as a **768-dimensional vector** encoding its semantic and sentiment context.
- Feature Engineering:** The embeddings were aggregated using mean pooling to form uniform numeric representations. Data cleaning ensured the removal of missing or duplicate entries.
- Classification with XGBoost:** The embeddings were input into an **XGBoost classifier**, which learned to distinguish sentiment patterns. The model used an 80–20 split, training on 400 samples and testing on 100. XGBoost's decision-tree boosting helped achieve high accuracy and generalization on limited data.

The training process outputs accuracy metrics, precision-recall scores, and a confusion matrix to analyze class-wise performance. The model achieved balanced performance across positive, neutral, and negative classes.

Results and Discussion

The model achieved an accuracy of approximately **0.76**. It effectively distinguished neutral and positive sentiments, though negative samples were slightly underrepresented in the dataset. The confusion matrix showed that most neutral and positive predictions aligned with ground truth labels, confirming the model's generalization on unseen data.

Example Prediction:

"Markets decline as inflation rises" → Predicted: **Negative** (Confidence: 0.92)

"Sensex surges on strong quarterly earnings" → Predicted: **Positive** (Confidence: 0.89)

"Company reports steady Q2 performance" → Predicted: **Neutral** (Confidence: 0.85)

These results demonstrate that the FinBERT + XGBoost pipeline effectively interprets tone and emotion in short financial texts.

Future Work

To enhance this system, several improvements are planned:

- **Fine-tuning FinBERT** on extended datasets to adapt to evolving financial vocabulary and emerging market narratives.
- **Incorporating sequential deep learning** models such as **LSTM** or **BiLSTM** layers on top of embeddings

to capture temporal sentiment trends.

- **Integrating real-time data streams** from APIs (like Bloomberg or Yahoo Finance) to enable dynamic, continuous sentiment monitoring.
- **Developing a visualization dashboard** to display time-series sentiment evolution and correlation with stock indices.
- **Expanding multi-modal analysis** by combining textual sentiment with numeric market indicators for richer prediction insights.

These upgrades aim to transition the model from a static analyzer to a live, explainable financial sentiment intelligence system.

Conclusion

The News Sentiment Analyzer provides a powerful starting point for understanding how textual tone influences financial market perception. By leveraging FinBERT embeddings and XGBoost classification, it delivers interpretable, accurate predictions even with limited data. Future enhancements will allow the model to operate in real-time financial ecosystems, aiding investors, analysts, and decision-makers in sentiment-driven forecasting.