FinBERT–LSTM Based Smart Stock Prediction System

1 Overview

This report introduces a real-time stock prediction system based on a FinBERT–LSTM architecture. The system integrates sentiment extracted from financial news with historical monthly returns to forecast the next month's stock performance. Its overall design emphasizes modularity, data quality control, and deployment readiness, ensuring stable and reproducible inference in both local and server environments.

2 Model Architecture

The system consists of two main components. First, a domain-specific Transformer model (FinBERT) converts financial news headlines into sentiment scores for individual articles, which are then aggregated into a monthly sentiment indicator. Second, the LSTM network processes a rolling four-month sequence of feature vectors—each containing the monthly return and its corresponding sentiment value—and outputs a scalar representing the predicted return for the next month. The entire pipeline follows a modular design, allowing the NLP module and the time-series prediction module to be updated or replaced independently. The figure below illustrates the overall model architecture.

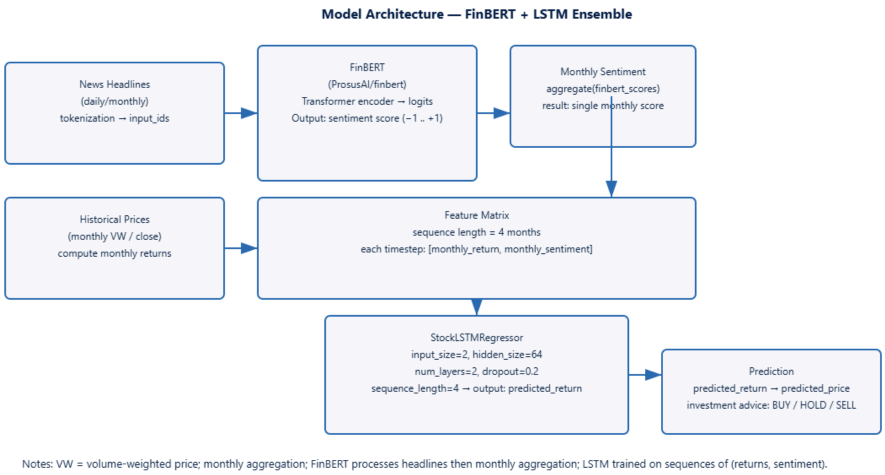

Figure: Model Architecture

3 FinBERT Sentiment Module

In the sentiment analysis module of this study, we employ the FinBERT pretrained model released by ProsusAI. FinBERT is based on the BERT architecture and has undergone domain-adaptive pretraining on large-scale financial corpora, enabling it to achieve higher accuracy in sentiment classification tasks involving financial news headlines, company announcements, and analytical reports. The model is loaded using

BertForSequenceClassification.from_pretrained("ProsusAI/finbert").

FinBERT outputs classification logits categorized into three classes: positive, negative, and neutral. This study further converts them into continuous sentiment scores to enhance the interpretability and usability of sentiment metrics, specifically by:

$$Sentiment\ Score = P(positive) - P(negative)$$

Subsequently, sentiment scores from multiple news articles within the same month are aggregated (using arithmetic mean by default) to construct robust monthly sentiment features.

4 LSTM Prediction Module

The LSTM predictor is specifically designed for stock return time series. Its input consists of a rolling sequence of the past four months, with each month represented by two features: the monthly return and the corresponding sentiment indicator. The model comprises two stacked LSTM layers, each with 64 hidden units, with a 20% dropout applied between layers to prevent overfitting. The representation of the final time step from the LSTM is then passed through a linear regression layer to produce the predicted return for the next month.

5 Training & Evaluation

The training dataset primarily consists of news data and historical stock price data. The news data is sourced from a Kaggle dataset, covering approximately 15,000 U.S. stocks with news headlines from 2009 to 2020. Stock price data is obtained via the Massive API, from which monthly price changes for the corresponding periods are calculated.

During data processing, the pre-trained FinBERT model is first used to compute sentiment scores for news headlines. These sentiment scores are then aligned with the corresponding monthly price changes to construct the training samples. Each training sample consists of a sequence of sentiment scores and price changes from the past four months, and the model's objective is to predict the price change for the following month.

For the training strategy, 20% of the dataset is used as a validation set. The model is trained using the Adam optimizer and mean squared error (MSE) loss, combined with a dynamic learning rate scheduler and early stopping mechanism, for 50 epochs.The figure below shows the loss curves during training.

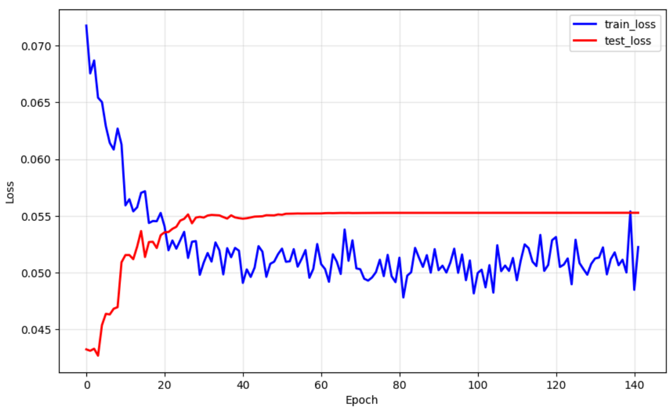

Figure: Training Loss

6 API Design & Real-time System

In addition, a web demo is provided to support real-time data retrieval, feature assembly, sentiment scoring, model inference, and output prediction. The request flow is as follows:

- Fetch the latest four months of monthly stock prices using the [Massive price API];
- Retrieve relevant news headlines for the same four months via the [Massive news API];
- Compute monthly sentiment scores using FinBERT;
- Assemble the input matrix for the past four months;
- Run LSTM inference and return the predicted return, corresponding stock price, and recommended trading actions.

The figure below shows the interface of the real-time web demo, using Apple Inc. (AAPL) as an example, and fully illustrates the entire workflow from data retrieval, feature assembly, and sentiment scoring to model prediction output.

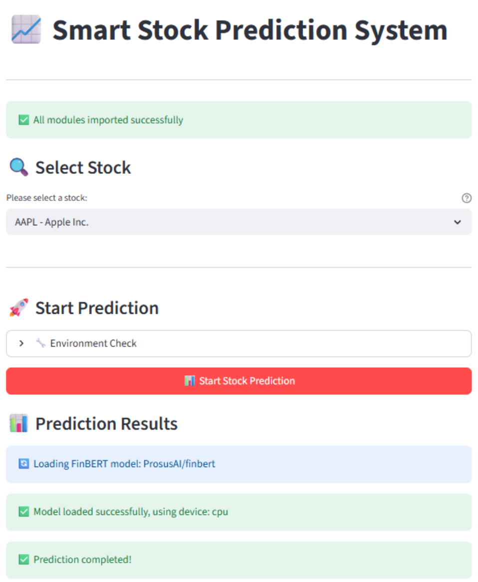

Figure: Demo Screenshot 1

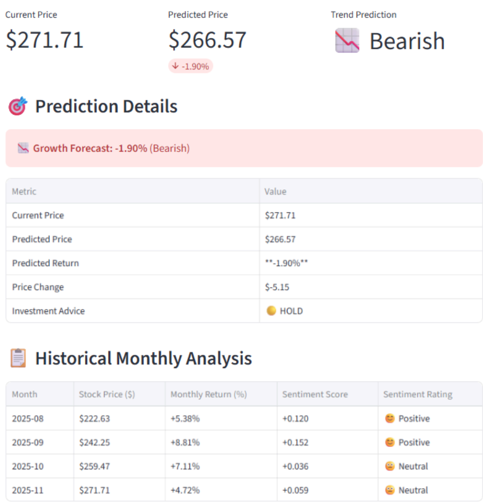

Figure: Demo Screenshot 2

Figure: Demo Screenshot 3

7 Conclusion

Combining FinBERT-derived sentiment with a compact LSTM time-series model produces a pragmatic and interpretable forecasting pipeline that improves responsiveness to new textual information while retaining temporal modeling of price dynamics. The modular design, built-in robustness measures, and API-oriented interfacing make this approach suitable for both experimental research and prototype production deployments. The code can be found at GitHub link, and the real-time prediction system is available at this website.