Stock Price Movement Prediction Using Social Media Sentiment

Group ID: 25RC33

Group Members:

Name	Roll No.	Role & Responsibilities
Ankesh Kumar	2301CS06	Defined project scope, assigned tasks, integrated all modules, and oversaw report writing and submission.
Eshan Bhaskar	2301CS16	Drafted methodology and results sections, visualized findings, and compiled final report and presentation.
Krishan Kumawat	2301CS25	Collected financial and sentiment datasets, cleaned and preprocessed data, and prepared train-validation splits.
Neetish Kumar	2301AI15	Implemented text cleaning, sentiment scoring (BERT), and generated sentiment-based features for modeling.
Raushan Raj	2301CS82	Built and tuned ML models, evaluated performance metrics, and compared sentiment vs price baselines.

Introduction

This project aims to predict Tesla (TSLA) stock price movements by integrating social media sentiment with financial time-series data. Social media platforms like Twitter reflect investor mood and influence stock volatility. Tesla, being highly sensitive to news and public opinion, is a fitting candidate for such a study.

We extract sentiment scores from tweets using a pre-trained BERT-based model (FinBERT) and combine these with technical indicators (past returns, momentum) as inputs to a Long Short-Term Memory (LSTM) network. The LSTM, known for modeling temporal dependencies, is trained to forecast future returns (e.g., 7-day ahead). Our final model is evaluated on Root Mean Squared Error (RMSE) and directional accuracy (correct up/down predictions). We also simulate a trading strategy using stop-loss and take-profit rules to estimate profit. This approach builds on prior work showing that BERT-based sentiment indices and LSTM models improve forecasting accuracy.

Related Work

Previous studies have shown that combining textual sentiment with price data improves stock forecasts. For example, Gu (2020) incorporated a finance-tuned BERT ("Fin-

BERT") into an LSTM, boosting prediction accuracy by around 3%. Similarly, Hiew (2019) found significant enhancement using a BERT-based financial sentiment index together with LSTM.

Ko & Chang (2021) applied a BERT + LSTM model on news/forum data and reported 12% RMSE improvement over baselines. Liu (2025) used FinBERT and VADER sentiment inputs for multiple LSTM variants, finding that incorporating all news sentiment yielded the best TSLA forecasts. Asgarov (2023) showed that LSTM models trained on historical prices plus Twitter sentiment for Tesla and Apple achieved low errors and captured overall price trends.

Methodology

Data Collection & Preprocessing

TSLA historical prices (daily close, volume) were fetched from stock_yfinance_data from Hanna Yukhymenko, and tweets containing "Tesla" or "TSLA" were filtered out. Tweets were cleaned (removing URLs, mentions) and grouped by date.

Sentiment Scoring (BERT)

Each day's tweets were processed through a BERT model to obtain sentiment scores. We converted outputs into numeric sentiment indices (range: -1 to +1) and aligned them with stock price timelines. The "net sentiment" on day t was the mean of all tweet scores that day.

Feature Engineering

All numeric features were standardized (z-score). Engineered features included:

- Aggregated daily sentiment polarity
- Sentiment merged with stock data based on date alignment
- Normalized and time-aligned series using statistical scaling

Model Architecture

Our model used a sliding window of past 20 days of features as input to predict the next 7-day return. It consisted of two stacked LSTM layers (64 and 32 units) followed by a dense output neuron. Compiled using Adam optimizer with MSE loss, similar to architectures in Chaudhary (2025).

Training & Evaluation

The dataset was split chronologically (80% train, 20% test). Early stopping prevented overfitting. Evaluation used RMSE and directional accuracy (fraction of days where predicted and actual directions matched). Random baseline = 50% accuracy.

Trading Simulation

We tested a stop-loss/take-profit strategy based on model predictions. A long position was entered when positive return was forecasted; otherwise, no trade. Stop-loss triggered at 3% drop, take-profit at 2% gain. Portfolio performance was measured by cumulative returns, inspired by Polec (2020) and Arnold (2016).

Key Results

The LSTM model's predicted TSLA prices closely tracked actual prices with minimal lag. RMSE 0.1561 and directional accuracy 65.77%, outperforming traditional baselines. The stop-loss/take-profit strategy achieved a simulated profit of around 26%, validating the model's predictive power. Asgarov et al. (2023) reported similar tracking performance, and our results align with these findings. The LSTM effectively learned sentiment-driven trends reflected in social media data.

Conclusion

We developed a hybrid model combining BERT-based tweet sentiment with financial indicators in an LSTM framework for TSLA forecasting. Our key contributions:

1. **Sentiment Integration:** Using BERT for tweet-level financial sentiment.
2. **Modeling:** Multi-layer LSTM architecture for temporal dependency modeling.
3. **Evaluation:** Improved RMSE and directional accuracy over traditional models.

Limitations include possible misclassification by BERT on sarcastic tweets and variations in tweet volume.

References

1. Liu, X. (2025). *Tesla Stock Prediction Based on LSTM and News Sentiment Analysis.* Atlantis Press.
2. Asgarov, A. (2023). *Predicting Financial Market Trends using Time Series and NLP.* arXiv:2309.00136.
3. Ko, C.-R., & Chang, H.-T. (2021). *LSTM-based Sentiment Analysis for Stock Forecast.* PeerJ Comput. Sci.
4. Hiew, J. Z. G. et al. (2019). *BERT-based Financial Sentiment Index and LSTM-based Stock Return Predictability.* arXiv:1906.09024.
5. Chaudhary, R. (2025). *Advanced Stock Market Prediction Using LSTM.* arXiv:2505.05325.
6. Polec, J. (2020). *Stop-Loss, Take-Profit, Triple-Barrier & Time-Exit.* Medium.
7. Arnold, C. (2016). *Stop-Loss in Trading Strategy.* PythonProgramming.net.