Tesla Stock Closing Price Forecasting Using Reddit Sentiment Analysis and LSTM Model

Author - Rakchanok Wongsara
Advisor - Associate Professor Dr.Chumphol Bunkhumpornpat

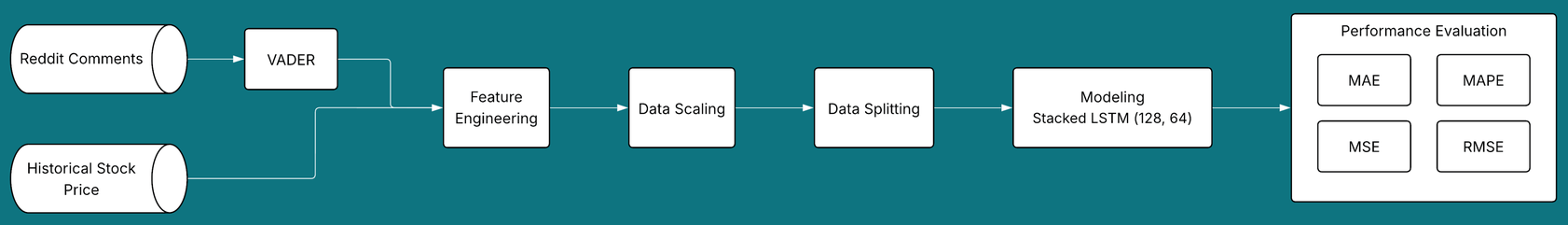
ABSTRACT

This study investigates the effectiveness of integrating social media sentiment with historical stock price data to forecast Tesla's closing prices using NLP and Deep Learning. Reddit data were analyzed with VADER to produce daily sentiment scores and combined with stock prices from Yahoo Finance. A stacked LSTM model with time-series splitting was employed and compared with a price-only baseline. Results indicate that incorporating sentiment improves short-term stock price forecasting performance.

INTRODUCTION

Social media platforms such as Reddit reflect retail investor sentiment that can influence stock price movements. Traditional forecasting approaches primarily rely on historical price data and often overlook behavioral signals. This study integrates Reddit-derived sentiment with Tesla historical price data and employs a stacked LSTM model to evaluate whether sentiment information improves short-term stock price forecasting performance compared with price-only models.

METHODOLOGY



Data Collection

- Reddit comments (PushshiftDumps)
- Tesla Stock prices (Yahoo Finance)
- Period: 2022–2024

VADER Sentiment Analysis

- Text cleaning
- VADER scoring -> daily sentiment

Feature Engineering

- Merge sentiment + price
- 45-day lookback window

Data Scaling

- Z-score standardization

$$X_{scaled} = \frac{X - \mu}{\sigma}$$

Mean Absolute Error (MAE)

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

Mean Squared Error (MSE)

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

Root Mean Squared Error (RMSE)

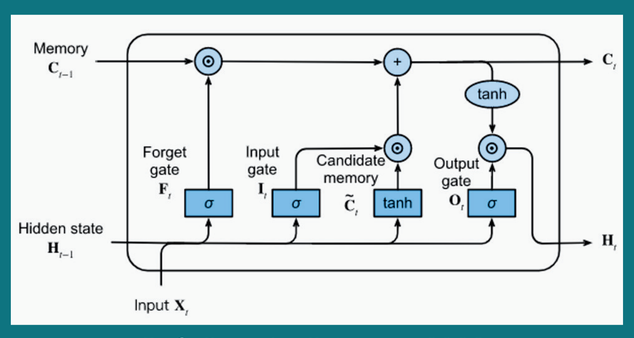
$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

Mean Absolute Percentage Error (MAPE)

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right|$$

LSTM Model

- Stacked LSTM (128, 64)
- Time-series split (60/20/20)
- Adam + Early Stopping



Ref. Ottavio Calzone, 2022

Evaluation

- MAE, MAPE, RMSE, MSE
- Compared with price-only model

Figure 1: Comparison of training loss and validation loss for LSTM model

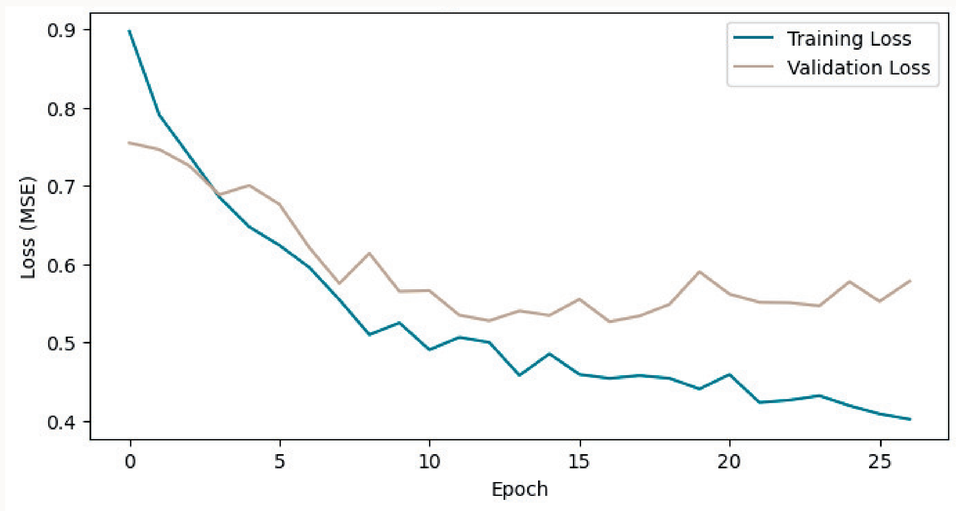
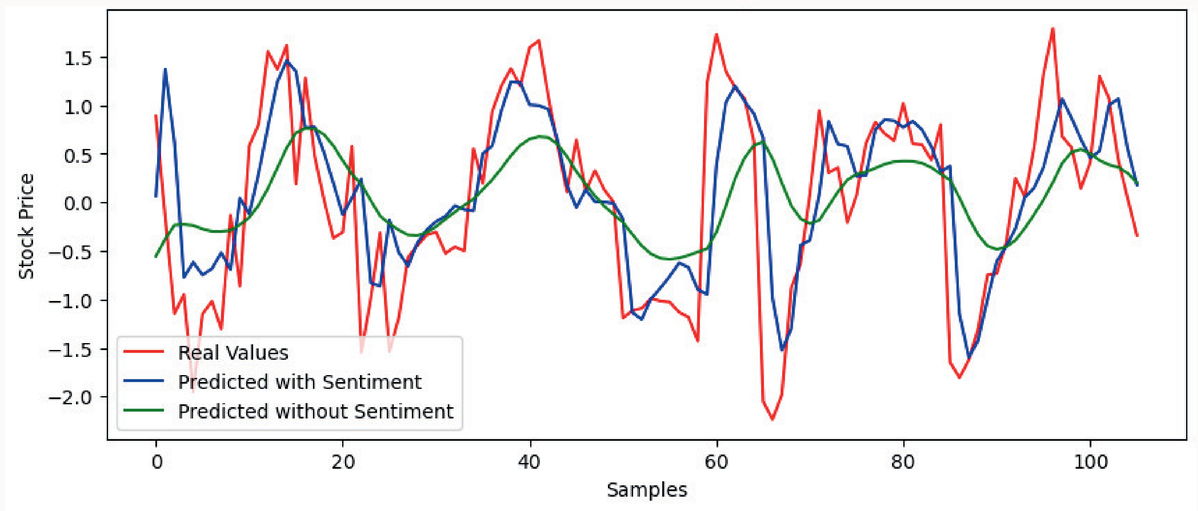


Figure 2: Stock price prediction by LSTM



RESULTS & DISCUSSION

The **sentiment-enhanced LSTM consistently outperforms the price-only baseline** across all evaluation metrics. RMSE decreases from 0.8896 to 0.7006, accompanied by clear reductions in MAE and MSE, indicating improved predictive precision. The training and validation loss curves suggest **stable convergence without significant overfitting.** Notably, the sentiment-integrated model **captures short-term fluctuations and market turning points more effectively**, whereas the baseline exhibits smoother but more delayed responses. Although MAPE remains relatively high—likely influenced by scaling effects and near-zero values—the overall **relative error is substantially reduced when sentiment information is incorporated.**

Table 1: Stock price prediction performance

Feature	Model Name	Evaluation Metric			
		MAE	MSE	RMSE	MAPE
With Sentiment	LSTM	0.494156	0.490829	0.700592	134.851089
Without Sentiment		0.688597	0.791335	0.889570	325.594269

Key Findings:

- Social media sentiment improves forecasting accuracy across all metrics.
- Sentiment-enhanced LSTM better captures short-term market movements.
- Training results indicate stable learning without clear overfitting.