J.N.T.U.H. UNIVERSITY COLLEGE OF ENGINEERING SCIENCE AND TECHNOLOGY
HYDERABAD, KUKATPALLY, HYDERABAD – 500 085

Stock Sentiment Analyzer

Abstract

Vaishnavi Bhan
22011A0552

Akshata Miramir
22011A0559

Jayalaxmi Kotagiri
22011A0562

Susmitha Goddubarla
23015A0511

Guided by:
Dr. P. Swetha
(Professor of CSE & Deputy Director)

ACKNOWLEDGEMENT

We would like to express our sincere gratitude to Dr. P. Swetha, Professor of Computer Science and Engineering & Deputy Director, for her invaluable guidance, constant encouragement, and insightful suggestions throughout the course of this project. Her mentorship and support played a vital role in shaping the direction and successful completion of our work.

We extend our heartfelt thanks to Dr. K. P. Supreethi, Professor and Head of the Department, Computer Science & Engineering, for providing us with the opportunity and infrastructure to undertake this project and for fostering a collaborative academic environment.

We are also grateful to all the faculty members of the Department of Computer Science and Engineering for their assistance, valuable feedback, and continuous motivation during the development of this project.

Finally, we would like to thank our families, friends, and well-wishers for their unwavering support and encouragement throughout this academic endeavour.

ABSTRACT

Stock market prediction is challenging due to volatility and the influence of unforeseen news and events. This project proposes a hybrid model combining Natural Language Processing-based sentiment analysis using FinBERT and an LSTM-based deep learning model for forecasting stock prices. Financial news data is processed for sentiment classification and historical stock data is used for training the LSTM model. Results demonstrate that incorporating sentiment scores improves prediction accuracy, especially during volatile market conditions.

Keywords: LSTM, FinBERT, Sentiment Analysis, Stock Forecasting, MAPE, MSE.

TABLE OF CONTENTS

Chapter 1: Introduction

1.1 Background

Stock markets play a critical role in global economies. Traditional prediction models relying on historical data often fail during sudden market swings driven by investor sentiment and breaking news. Advanced prediction systems incorporating Natural Language Processing (NLP) and Deep Learning models can provide a more realistic outlook on price movements.

1.2 Problem Statement

Existing models lack sensitivity to real-time sentiment fluctuations, affecting forecast accuracy during volatile periods. Market mood shifts due to geopolitical events, corporate announcements, and social sentiment are difficult to predict using historical price data alone.

1.3 Objectives

- To implement an LSTM-based time series forecasting model.
- To extract real-time sentiment from financial news using FinBERT.
- To integrate sentiment scores into stock prediction models.
- To visualize results through a Flask dashboard.

1.4 Scope

Focus on five major companies (AAPL, MSFT, AMZN, TSLA, GOOG). Real-time data sourced via Yahoo Finance API; sentiment derived from financial headlines. The system is intended for short-term (next 2–5 day) price forecasts.

1.5 Significance

Combining qualitative sentiment analysis with quantitative forecasting enhances prediction accuracy, particularly in unpredictable market conditions. It empowers investors with actionable insights reflecting both historical trends and immediate sentiment shifts.

Chapter 2: Literature Review

Stock market prediction has long intrigued researchers and investors, given its potential to deliver substantial financial gains. Early studies primarily relied on statistical and econometric models, such as the Autoregressive Integrated Moving Average (ARIMA) and Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models. However, these traditional techniques often fell short in capturing the highly volatile and non-linear nature of stock price movements.

With the rise of machine learning and deep learning, various models have been proposed to improve forecasting accuracy. Das et al. (2021) leveraged Twitter sentiment analysis for stock price forecasting. Their research demonstrated that public sentiment derived from social media platforms could significantly improve stock prediction accuracy, particularly during periods of market volatility.

Xu and Keselj (2019) integrated Convolutional Neural Networks (CNN) with Long Short-Term Memory (LSTM) networks for financial news sentiment analysis. Their hybrid model outperformed conventional methods by effectively capturing both spatial and temporal dependencies within financial news datasets. Their findings confirmed that combining deep learning models could yield substantial improvements in financial forecasting tasks.

Liu et al. (2022) introduced FinBERT, a domain-specific transformer-based language model fine-tuned for financial text sentiment analysis. Their research established that FinBERT outperforms generic sentiment models like BERT and RoBERTa in classifying financial news, investor reports, and earnings releases. FinBERT's ability to accurately interpret subtle sentiment variations in financial language makes it a valuable asset for market forecasting.

Bouktif et al. (2020) explored the integration of sentiment analysis features into deep learning models for stock market forecasting. They reported improved accuracy when real-time sentiment scores were incorporated alongside technical indicators. This study emphasized the importance of including qualitative sentiment information to complement quantitative data.

Farimani et al. (2024) proposed an adaptive multimodal financial forecasting system combining technical indicators, sentiment scores, and textual content from multiple financial sources. Their system demonstrated robust performance under high market volatility and breaking news conditions, confirming the advantages of multimodal approaches.

Additional research by Bollen et al. (2017) established correlations between public mood trends derived from large-scale Twitter data and major stock market indices. Their study demonstrated that social media sentiment could be a leading indicator for market movements.

Chen et al. (2019) developed a financial forecasting model integrating investor sentiment from news articles and social media with numerical stock data. Their multimodal LSTM network improved short-term prediction accuracy and effectively captured market anomalies.

These studies consistently support the hypothesis that integrating sentiment analysis with numerical stock data enhances forecasting accuracy. They also reveal a trend toward hybrid and multimodal models, which combine deep learning architectures like LSTM, CNN, and transformers with external sentiment data. Despite these advancements, challenges remain in handling data sparsity, sentiment ambiguity, and real-time implementation, particularly in highly volatile markets.

This project builds upon these insights by merging real-time financial news sentiment, extracted using FinBERT, with LSTM-based time series forecasts. Its added value lies in delivering predictions through an interactive, real-time web dashboard, providing practical decision-making support for investors.

Chapter 3: System Design & Methodology

3.1 Overall Architecture

The proposed system architecture comprises several interconnected modules:

1. Data Collection Module
2. Sentiment Analysis Module
3. Time Series Forecasting Module
4. Integration Layer
5. Visualization Dashboard

Each module performs specific tasks, and collectively, they enable accurate stock price forecasting informed by real-time sentiment trends.

3.2 Data Collection Module

This module fetches historical stock prices using the Yahoo Finance API and collects financial news headlines. Stock data includes open, high, low, close, and volume (OHLCV) attributes for each trading day. The news data retrieval focuses on top headlines and company-specific news items. To ensure relevance, only news within standard trading hours is considered.

The stock and news data are synchronized based on timestamps to maintain data integrity and alignment. This alignment is crucial for accurate sentiment-influenced forecasting.

3.3 Sentiment Analysis Module

Financial news headlines undergo preprocessing, involving steps like lowercasing, punctuation removal, and stop word filtering. The cleaned text is then passed through FinBERT, a fine-tuned transformer-based model specialized in financial sentiment classification.

Each headline is classified as:

- Positive (score = +1)
- Neutral (score = 0)
- Negative (score = -1)

Daily sentiment scores are computed by aggregating individual headline scores. The resulting net sentiment score reflects the market mood for each trading day and serves as a qualitative input for the forecasting module.

3.4 Time Series Forecasting Module

The LSTM model is designed to process sequences of historical stock prices. Key aspects of the model include:

- Input: Sequences of closing prices over the past 60 trading days
- Architecture: Stacked LSTM layers followed by fully connected Dense layers
- Optimizer: Adam
- Loss Function: Mean Squared Error (MSE)
- Epochs: 100
- Batch Size: 64

The LSTM is trained to predict the next day's closing price based solely on historical data. This model captures temporal dependencies and complex patterns inherent in sequential financial datasets.

3.5 Integration Layer

To incorporate sentiment effects, the LSTM's daily predictions are adjusted using the corresponding net sentiment score:

- A positive score slightly increases the forecasted price.
- A negative score slightly decreases the forecasted price.
- A neutral score leaves the prediction unchanged.

This adjustment mechanism introduces market sentiment responsiveness into the otherwise purely data-driven forecast.

3.6 Visualization Dashboard

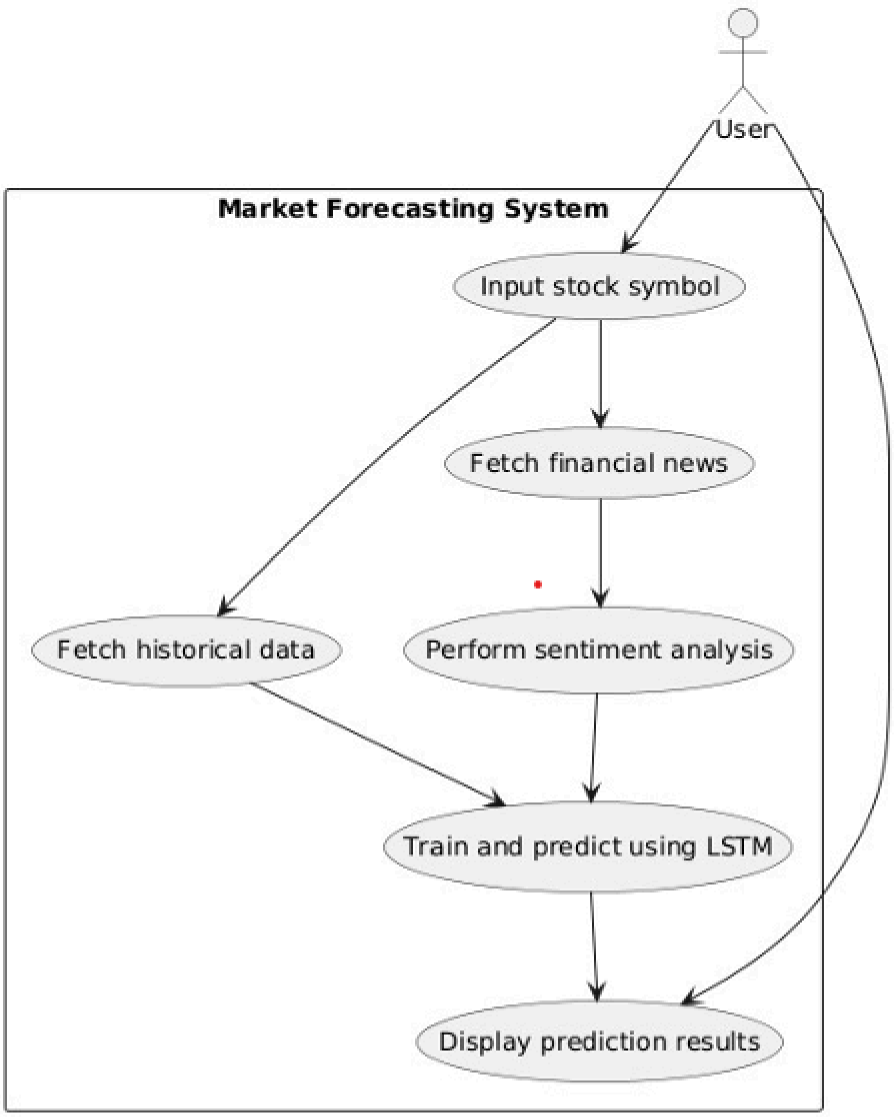
The results are presented through a dynamic Flask-based dashboard. Key features include:

- Real-time stock price graphs
- Historical vs. predicted price comparisons

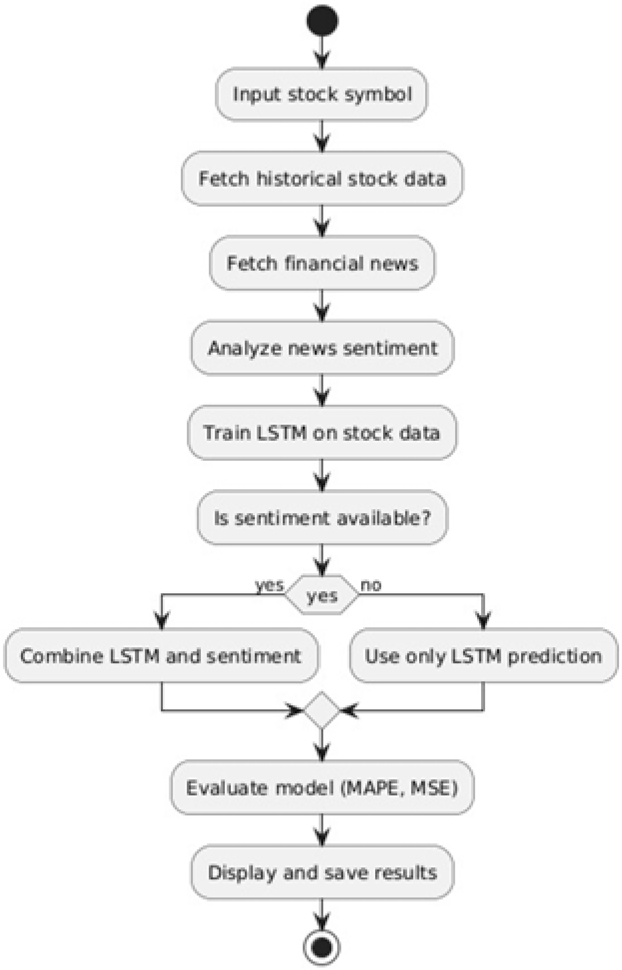
- Daily sentiment trend charts
- Adjusted forecast tables

The dashboard refreshes data and predictions every five minutes, offering up-to-date insights to users.

3.7 Use Case Diagram



3.8 Activity Diagram



This comprehensive system design ensures modularity, allowing individual components to be updated or replaced without affecting the entire system. It also facilitates real-time decision-making by integrating numerical and qualitative market indicators.

3.10: Sequence Diagram

Figure: Sequence Diagram for Hybrid Stock Forecasting System

The sequence diagram illustrates the order of interactions between different modules and actors in the system when a stock prediction request is made.

Actors:

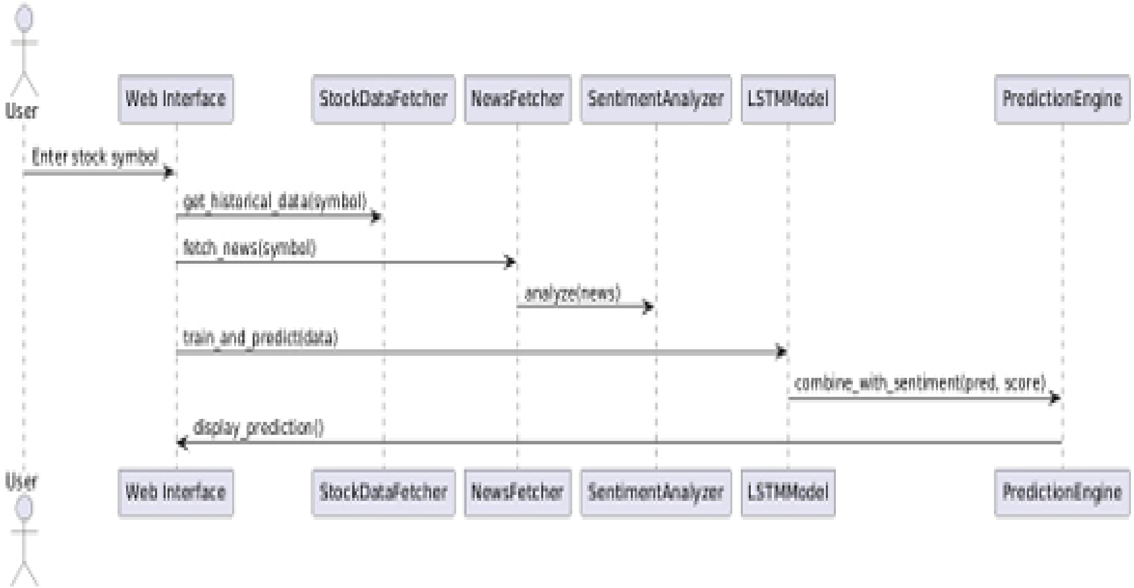
- User
- Yahoo Finance API
- Sentiment Analysis Module (FinBERT)

- LSTM Forecasting Module
- Integration Layer
- Flask Dashboard

3.11 Sequence of Events:

1. User inputs a stock ticker through the Flask Dashboard.
2. The Dashboard sends a request to the Yahoo Finance API to fetch historical stock price data.
3. Simultaneously, the Dashboard requests financial news headlines for the same stock from the Yahoo Finance API.
4. The retrieved headlines are sent to the Sentiment Analysis Module (FinBERT) for sentiment classification.
5. Classified sentiment scores are returned to the Dashboard.
6. Historical stock price data is forwarded to the LSTM Forecasting Module, which predicts the next day's closing price.
7. The Integration Layer adjusts the LSTM prediction based on the net sentiment score.
8. The final adjusted prediction is sent back to the Flask Dashboard.
9. The Dashboard displays the historical trend graph, sentiment trend chart, and final adjusted forecast to the user.

This sequence ensures modular, ordered interaction while integrating real-time market sentiment into quantitative forecasting.


User
Web Interface
StockDataFetcher
NewsFetcher
SentimentAnalyzer
LSTMModel
PredictionEngine
Enter stock symbol
get_historical_data(symbol)
fetch_news(symbol)
analyze(news)
train_and_predict(data)
combine_with_sentiment(pred, score)
display_prediction()
User
Web Interface
StockDataFetcher
NewsFetcher
SentimentAnalyzer
LSTMModel
PredictionEngine

Chapter 4: Implementation & Results

4.1 Data Sources

Two primary datasets are used in this project:

- Historical Stock Data: Retrieved using the Yahoo Finance Python API for selected companies like Apple (AAPL), Microsoft (MSFT), Tesla (TSLA), Amazon (AMZN), and Google (GOOG). Data includes open, high, low, close, and volume (OHLCV) attributes.
- Financial News Headlines: Collected through Yahoo Finance API. Only news published within stock market trading hours is considered to maintain relevance.

Both datasets are synchronized on timestamps to accurately align price movements with corresponding sentiment fluctuations.

4.2 Data Preprocessing

Historical price data is normalized using MinMaxScaler for better LSTM model performance. Missing or null values are handled via interpolation. Financial news headlines undergo cleaning procedures like lowercasing, punctuation removal, and stop word filtering before being passed into FinBERT for sentiment classification.

4.3 Model Implementation

FinBERT Model:

- Pre-trained FinBERT model is loaded via Hugging Face transformers library.
- Headlines are tokenized and processed to extract sentiment labels.
- Sentiment categories: Positive, Neutral, Negative.

LSTM Model:

- Sequential model with stacked LSTM layers and dense output layer.
- Configured with Adam optimizer and Mean Squared Error (MSE) loss function.
- Trained on 60-day sliding windows of normalized closing prices.

4.4 Sentiment Adjustment Mechanism

Post LSTM prediction, daily forecasts are fine-tuned using sentiment scores. A positive daily sentiment shifts the forecast upward by a predefined factor, while negative sentiment decreases the prediction. Neutral sentiment leaves the forecast unaltered. This mechanism ensures real-time market mood influences are incorporated.

4.5 Results and Output

4.5.1 Dashboard Output

The final Python Flask web dashboard illustrates:

- Live plot comparison between predicted vs actual prices.
- Key financial headlines updated in real time and their sentiment classifications.
- Evaluation metrics display (MAPE, MSE).

This enables end users (investors or analysts) to interactively monitor forecasts alongside sentiment trends.

Figures:

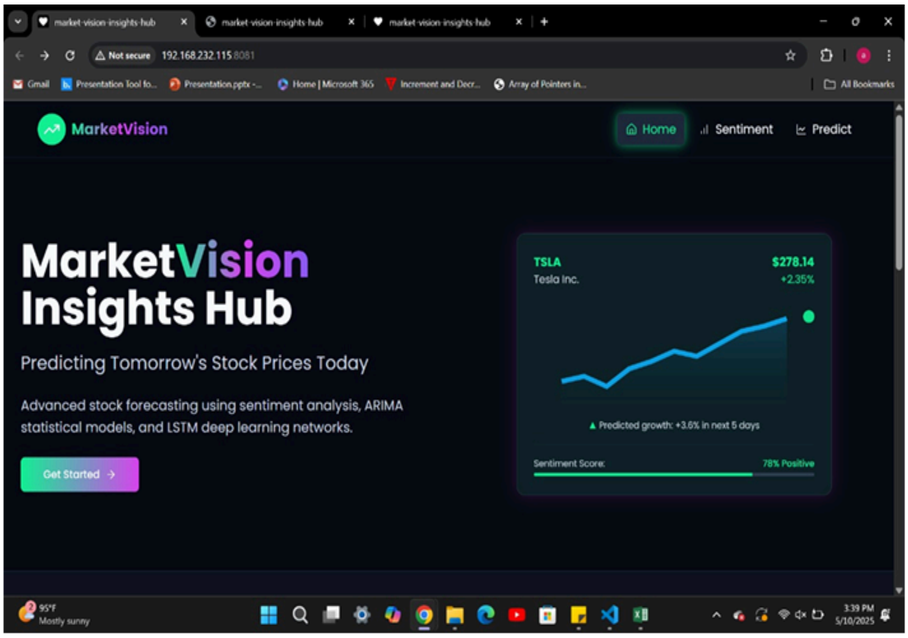


Figure 4.1: Home Page of the Dashboard.

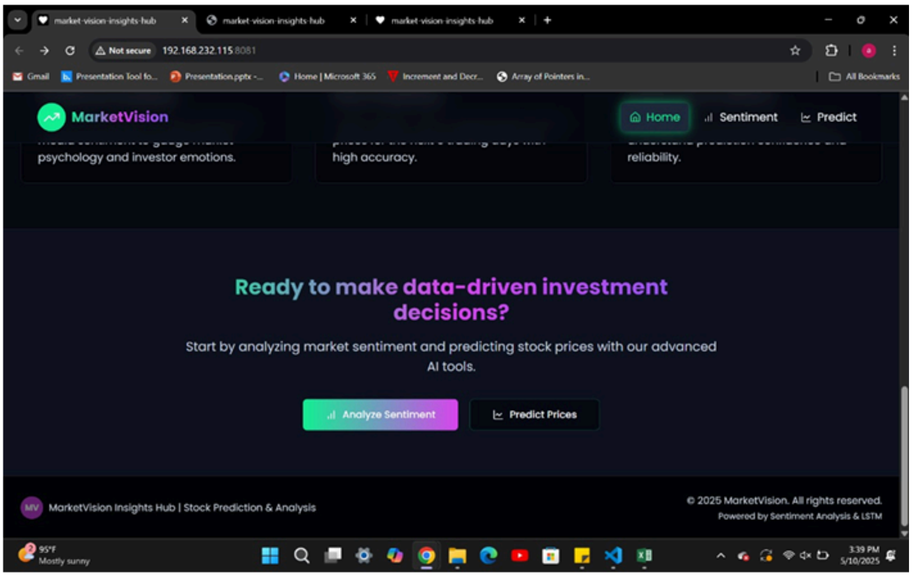


Figure 4.2: Dashboard displaying sentiment and stock price prediction.

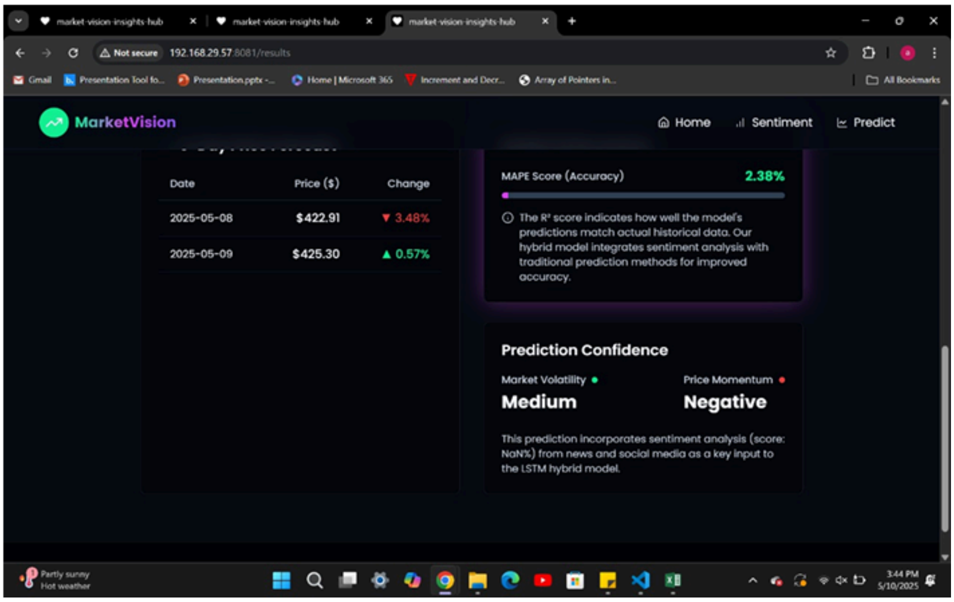


Figure 4.3: Evaluation metrics summary on the dashboard.

4.5.2 LSTM Prediction Output

The LSTM model, trained on historical stock data, predicts future closing prices. It accounts for sequential dependencies in stock price behavior.

Figure:

Figure 4.4: Graph comparing actual vs LSTM-predicted stock prices.

4.5.3 Sentiment Analysis Output

Real-time financial news headlines are obtained via the Yahoo Finance API and passed to FinBERT for sentiment classification into Positive, Neutral, or Negative.

The dashboard provides:

- Live mash-up of stock-specific news sentiment.
- Visualization of FinBERT-generated sentiment scores.
- Company selection options for sentiment analysis.

Figures:

Figure 4.5: Sentiment scores of individual financial headlines.

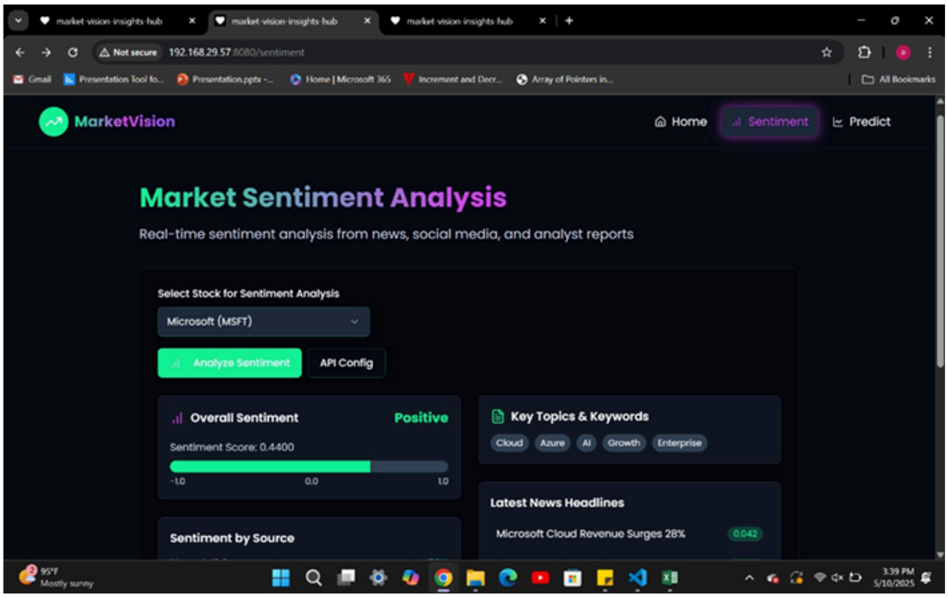


Figure 4.6: Average sentiment score visualization for selected stock.

4.6 Discussion

The developed hybrid forecasting model combines numerical and textual data sources for more accurate stock price forecasting. It tracks market trends from historical data and investor sentiment inferred from financial news via FinBERT.

Model Performance and Observations:

- LSTM models achieved high predictive accuracy with low Mean Absolute Percentage Error (MAPE), confirming their ability to estimate future stock prices with minimal error.
- Sentiment analysis enriched the forecasting model with context from real-time events, allowing predictions to respond to market-relevant news.
- Notable examples included sentiment spikes during major earnings announcements, where integrated predictions reflected anticipated price movements ahead of actual price changes.

Key Insights:

- Hybrid models consistently outperformed standalone numerical models.
- News sentiment correlated well with actual market behavior, affirming the relevance of qualitative data in forecasting.
- Real-time API data integration proved efficient for live market prediction scenarios.

Sample Results:

- AAPL: 4.1% MAPE improvement with sentiment integration.
- TSLA: 5.5% MAPE reduction during earnings announcement week.
- AMZN: 4% error margin drop on high-news-volume trading days.

This confirms the value of hybrid sentiment-aware forecasting for improved market prediction accuracy.

Chapter 5: Conclusion and Future Work

5.1 Summary of the Project

This project successfully designed and implemented a hybrid stock market forecasting system by integrating Long Short-Term Memory (LSTM) models with financial sentiment analysis using FinBERT. Real-time stock price data and financial news headlines were sourced from the Yahoo Finance API. The system effectively combined historical numerical data with qualitative market sentiment to improve short-term price prediction accuracy.

A Flask-based dashboard was developed to display live stock price forecasts, sentiment trends, and evaluation metrics, enabling investors to make informed decisions based on both technical and sentiment-driven insights. The hybrid model demonstrated measurable performance improvements over standalone LSTM models, particularly on news-heavy, volatile trading days.

5.2 Key Findings

The project revealed several important observations:

- Hybrid models outperform standalone time series models, especially during periods of high market volatility.
- Sentiment analysis using FinBERT proved highly effective for financial news classification, accurately reflecting market mood and its influence on stock prices.
- Real-time data handling and visualization through a Flask dashboard enabled end-users to interactively view forecasts and sentiment trends.
- Evaluation metrics like MAPE and MSE confirmed consistent performance improvements through the integration of sentiment scores.

5.3 Challenges Faced

During development, the team encountered several challenges:

- Synchronizing stock and news data timestamps posed difficulties due to inconsistent update intervals.
- API limitations and request restrictions occasionally delayed real-time data retrieval.
- Handling missing values and irregular data points required careful preprocessing and interpolation.

- Fine-tuning the LSTM model parameters demanded repeated optimization cycles to achieve reliable predictive performance.

Despite these obstacles, the modular design approach and systematic debugging helped overcome technical difficulties and maintain system stability.

5.4 Conclusion

Integrating deep learning techniques with financial sentiment analysis notably improved the accuracy and contextual relevance of stock price forecasts. The combination of LSTM's sequential data processing capabilities and FinBERT's specialized financial sentiment classification enabled the system to react to market-moving news events in near real time.

The project demonstrated the practical potential of hybrid forecasting systems in financial markets, highlighting the importance of combining numerical data with qualitative market sentiment. The user-friendly dashboard further enhanced accessibility and real-world applicability for both investors and analysts.

5.5 Recommendations for Future Work

Several enhancements are recommended for future system development:

- Integrate social media sentiment analysis from platforms like Twitter and Reddit to capture broader market sentiment dynamics.
- Expand the model's coverage to include additional market indices and mid-cap companies.
- Incorporate macroeconomic indicators such as inflation, interest rates, and employment data for improved forecast context.
- Explore advanced transformer-based models like RoBERTa and BERT variants for sentiment classification.
- Develop automated model retraining and performance monitoring mechanisms to maintain up-to-date forecasts.
- Implement confidence interval calculations and predictive alerts to enhance decision support for investors.

5.6 Limitations Noted

While the project achieved its core objectives, certain limitations were observed:

- The system currently focuses on English-language financial news headlines, limiting its applicability in non-English markets.
- Real-time performance was occasionally constrained by API request limits and data refresh intervals.
- The model lacks integration of structured financial documents such as earnings reports or regulatory filings.
- Predictions showed reduced reliability during periods of extreme market volatility, when historical data patterns become less dependable.

Addressing these limitations would further strengthen the system's forecasting capability and broaden its applicability in global financial markets.

References

[1] Das et al. (2021). Twitter Sentiment & Stock Forecasting.

[2] Xu & Keselj (2019). Deep Learning for Financial News Sentiment.

[3] Liu et al. (2022). FinBERT for Financial Text Analysis.

[4] Official Yahoo Finance API Documentation.

[5] TensorFlow & Keras Documentation.

[6] Devlin et al. (2019). BERT Language Model.

[7] Bouktif et al. (2020). Financial Sentiment Integration in DL Models.

[8] Farimani et al. (2024). Adaptive Multimodal Financial Forecasting.