WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Stock Price Prediction

[1]Mr Mohan Rajadurai, [2]Vishnu Ram M,[3] Sri Ragavendhira S,[4] Yogesh P,[5] Sanjaykumar S

[12345]Sri Shakthi Institute of Engineering and Technology, Coimbatore, 641062, India.

A B S T R A C T:

Stock price prediction is a crucial yet challenging task due to the complex, volatile, and non-linear nature of financial markets. This study explores the application of machine learning and time series analysis techniques—specifically ARIMA, Long Short-Term Memory (LSTM) networks, and Linear Regression models—for forecasting stock prices. Additionally, sentiment analysis on Twitter data is integrated to assess the impact of public opinion on stock movements. Historical stock price data and social media sentiment were collected and preprocessed to train and evaluate each model. The models were compared based on metrics such as RMSE and MAPE to determine their predictive accuracy. Results indicate that the LSTM model, when combined with sentiment analysis, outperforms traditional methods in capturing market trends and anomalies. This hybrid approach demonstrates the potential of combining deep learning with real-time sentiment for more accurate and timely stock price predictions.

Keywords: Stock Price Prediction, Machine Learning, Time Series Analysis, ARIMA, LSTM, Linear Regression, Sentiment Analysis, Twitter Data, Financial Forecasting, Deep Learning

Introduction

Stock markets play a critical role in the global economy, serving as a barometer for financial health and providing a platform for investment and capital growth. Accurately predicting stock prices has long been a subject of interest for investors, analysts, and researchers due to the potential financial gains and risk management opportunities it offers. However, stock prices are influenced by a multitude of factors, including historical trends, economic indicators, and public sentiment, making them highly volatile and difficult to forecast.

In recent years, advancements in machine learning and deep learning have opened new avenues for financial forecasting. Techniques such as AutoRegressive Integrated Moving Average (ARIMA), Linear Regression, and Long Short-Term Memory (LSTM) networks have shown promise in modeling complex, non-linear time series data. Furthermore, with the rise of social media platforms like Twitter, sentiment analysis has emerged as a powerful tool for capturing market sentiment and enhancing prediction models.

This study aims to explore and compare various predictive models for stock price forecasting, integrating historical market data with real-time sentiment extracted from Twitter. By evaluating the performance of traditional and deep learning approaches, this research seeks to identify effective strategies for improving forecast accuracy and supporting data-driven investment decisions.

Literature Review

Stock price prediction has been extensively studied using both statistical and machine learning methods. Traditional models like ARIMA are effective for linear time series data but struggle with non-linear patterns. Machine learning algorithms such as Linear Regression, SVM, and Random Forest have shown improved performance by capturing complex relationships in data. More recently, deep learning models like LSTM have outperformed traditional methods due to their ability to model sequential dependencies. Sentiment analysis from social media platforms like Twitter has also been integrated into prediction models to reflect public mood. Hybrid models combining historical data and sentiment analysis are emerging as the most effective approach. Despite progress, challenges like market volatility and model overfitting persist.

Methodology

Historical stock price data and Twitter sentiment were collected and preprocessed. Technical indicators were calculated, and sentiment scores were extracted using NLP techniques. ARIMA, Linear Regression, and LSTM models were trained using this combined data. Model performance was evaluated using RMSE and MAPE to compare prediction accuracy.

3.1 System Overview

The system follows a modular pipeline, consisting of the following major components:
Data Acquisition – Collection of historical stock price data from financial APIs and real-time sentiment data from Twitter feeds.
Preprocessing – Cleaning and normalization of stock data, including calculation of technical indicators and sentiment analysis of tweets.
Model Development– Application of ARIMA, Linear Regression, and LSTM models for stock price prediction based on the processed data.
Prediction Calculation – Model inference to generate predicted stock prices for future time points.
Visualization – Real-time visualization of stock price predictions along with historical data and performance metrics.
Backend Processing (Flask API) – Integration of model inference with a Flask-based API, providing HTTP endpoints for users to request predictions and view results.

3.2 Data Collection

Historical stock price data was collected from financial APIs, and relevant Twitter data was extracted based on keywords and stock symbols.

3.3 Data Preprocessing

Stock data was cleaned, missing values handled, and features normalized. Technical indicators (e.g., moving averages, RSI) were computed. Tweets were cleaned, tokenized, and processed for sentiment analysis.

3.4 Sentiment Analysis:

A pre-trained NLP model was used to classify tweet sentiment (positive, negative, neutral). Sentiment scores were aggregated daily and merged with stock data.

3.5 Model Development

Three models were developed:

- ARIMA for statistical time series forecasting
- Linear Regression as a baseline ML model
- LSTM for deep learning-based sequence prediction

3.6 Training and Testing

Models were trained using a rolling window method to simulate real-time prediction. Input features included price history, technical indicators, and sentiment scores.

Implementation Details

The implementation of the stock price prediction system involves several steps, starting from data collection and preprocessing to the deployment of prediction models via a Flask-based API. Below are the details of each component.

4.1 Data Acquisitiont

Stock price data is collected using APIs like Alpha Vantage or Yahoo Finance, which provide historical stock prices (open, close, high, low, and volume). Sentiment data is gathered in real-time from Twitter using the Tweepy library, filtering tweets based on stock symbols or company names.

4.2 Data Preprocessing

The collected data undergoes preprocessing, including:

- Stock Data: Handling missing values, normalization of numerical features, and generating technical indicators such as moving averages and Relative Strength Index (RSI).
- Sentiment Data: Tweets are cleaned by removing stop words, punctuations, and irrelevant content. Sentiment analysis is performed using Natural Language Processing (NLP) techniques, such as the VADER sentiment analysis model or pre-trained transformer models like BERT. The sentiment scores are aggregated daily.

4.3 Model Development

ARIMA (AutoRegressive Integrated Moving Average): A statistical model used for time series forecasting. ARIMA is tuned with parameters (p, d, q) based on the stock's historical data to model the temporal patterns.
Linear Regression: A simple machine learning model used as a baseline. It predicts stock prices based on linear relationships between historical prices and technical indicators.
LSTM (Long Short-Term Memory): A type of Recurrent Neural Network (RNN) specialized in learning long-term dependencies. LSTM is used to capture complex, non-linear patterns in the stock price data and improve prediction accuracy over time.

4.3 Prediction Calculation

After training, the models are used to predict future stock prices based on the most recent data. The predictions are generated for a specific time horizon, such as daily or weekly forecasts.

4.4 Visualization

The system provides a visual interface (e.g., via a web app built with Flask) where users can:

- View stock price predictions alongside historical data.
- See the effect of sentiment analysis on stock price movements.
- Compare the performance of different models in terms of forecast accuracy.

Result and Discussion

The results clearly indicate that **LSTM** is the most effective model for stock price prediction, especially when integrated with sentiment analysis. The LSTM's ability to model non-linear relationships and capture long-term dependencies made it more suited for forecasting stock prices compared to ARIMA and Linear Regression..

Conclusion

In conclusion, the LSTM model outperformed ARIMA and Linear Regression by effectively capturing non-linear patterns in stock price data. The integration of sentiment analysis further improved prediction accuracy, highlighting the importance of both historical data and market sentiment in stock price forecasting.

Acknowledgements

I would like to express my sincere gratitude to my advisor [Advisor's Name] for their valuable guidance and support throughout this project. Special thanks to [any collaborators, mentors, or peers] for their insightful feedback and assistance. I also appreciate the use of financial data APIs, sentiment analysis tools, and deep learning frameworks, which made this research possible. Lastly, I would like to thank my family and friends for their continuous encouragement.

REFERENCES:

1. J. Redmon Box, G. E. P., & Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control*. Holden-Day.
2. Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.
3. Brownlee, J. (2018). *Deep Learning with Python*. Machine Learning Mastery.
4. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735
5. Pang, B., & Lee, L. (2008). Opinion Mining and Sentiment Analysis. *Foundations and Trends in Information Retrieval*, 2(1–2), 1–135.
6. Sharma, D., & Sharma, S. (2019). Stock Price Prediction using Machine Learning Algorithms: A Review. *International Journal of Computer Applications*, 178(33), 1-5. https://doi.org/10.5120/ijca2019917972
7. Twitter Developer Documentation. (2020). *Tweepy: Python wrapper for the Twitter API*. Retrieved from https://www.tweepy.org
8. Zhang, Y., & Zheng, Y. (2019). Predicting Stock Price with LSTM Neural Networks. *International Journal of Computer Science and Information Technology*, 10(2), 1-9. https://doi.org/10.1016/j.jocs.2019.05.004