

REAL TIME STOCK TREND ANALYSIS USING NEWS SENTIMENT

Project Reference No.: 48S_BE_5869

College : Shridevi Institute of Engineering and Technology, Tumakuru

Branch : Artificial Intelligence and Data Science

Guide(s) : Dr. Girish L

Student(s): Mr. Niranjan K V

Mr. T M Yuva Prasad

Mr. Ullas P M

Mr. Srinivasulu V

Keywords:

Stock Forecasting, SARIMA, Streamlit Web App, Machine Learning, Financial Analytics, Time Series Analysis, NewsAPI, Data Visualization, ARIMA, Regression Metrics, Real-Time Data Integration

Introduction:

The stock market is inherently volatile and influenced by multiple factors, including investor sentiment driven by real-time news. In this project, we aim to enhance the accuracy of stock price predictions by integrating machine learning models with news sentiment analysis and historical prices. By using the SARIMA model for time-series forecasting and incorporating real-time news sentiment derived through natural language processing, we create a hybrid predictive framework. This project also involves the development of an interactive Streamlit-based web application that delivers forecasts and sentiment insights to end users in a user-friendly format. learning models, the project aims to bridge the gap between real-time sentiment and stock performance for smarter, sentiment-informed decision-making.

Objectives:

The report aims to develop a user-friendly application leveraging machine learning, specifically the Seasonal Auto Regressive Integrated Moving Average (SARIMA) model, for stock price prediction. It integrates real-time news sentiment analysis to enhance predictions.

- Develop a SARIMA-based model integrated with news sentiment analysis.
- Build an interactive Streamlit web app for stock price forecasting.
- Evaluate model performance using statistical metrics.
- Analyse the impact of news sentiment on prediction accuracy.
- Identify limitations and propose areas for future research.

Methodology:

Data Collection: Historical stock price data will be retrieved using the yfinance library. Real-time news data will be collected via the NewsAPI for sentiment analysis.

2. Data Preprocessing: Raw data will be cleaned, missing values handled, and time series indexed. Sentiment scores from news articles will be derived using natural language processing techniques.

3. Model Selection: SARIMA (Seasonal AutoRegressive Integrated Moving Average) will be chosen for its strength in capturing trend and seasonality in time series data.

4. Model Parameter Optimization: Optimal SARIMA parameters will be determined using ACF, PACF plots, and criteria like AIC/BIC.

5. Model Training and Validation: The SARIMA model will be trained on historical data, validated using statistical metrics (e.g., RMSE, MAE, R^2), and tested on unseen data.

6. Sentiment Integration: Sentiment scores from news analysis will be added as a feature to enhance prediction accuracy.

Model Evaluation: Prediction performance will be evaluated with regression and classification metrics, including MAPE and precision-recall measures.

8. Web Application Development: A Streamlit-based interactive web app will be created to display historical data, predictions, and sentiment trends.

9. Data Visualization: Visualizations, including prediction trends and news sentiment impacts, will be implemented using Plotly for an interactive user experience.

10. Comparative Analysis: SARIMA predictions will be compared against other models like Long Short-Term Memory LSTM and baseline approaches (e.g., moving averages).

11. Iterative Testing: Regular iterations will ensure alignment with real-world market conditions and address identified challenges.

12. Performance Documentation: Comprehensive documentation of evaluation metrics and system performance will be maintained for insights.

13. User Feedback: Feedback will be collected to refine the application's usability and accuracy.

14. Scalability Design: Modular system architecture will enable integration of additional features, such as alternative models or expanded data sources.

15. Deployment and Maintenance: The web app will be deployed with mechanisms for real-time updates and regular maintenance for sustained performance.

Result:

Comprehensive Stock Analysis Tool:

- Integrates stock data, predictions, and news in one platform

Advanced Forecasting:

- Uses SARIMA model for historical and future price predictions

Performance Validation:

- Includes multiple metrics to assess model accuracy

Visual Data Representation:

- Features interactive charts for clear trend visualization

Project Outcome & Industry Relevance:

The project successfully delivered a functional and scalable web application capable of real-time stock forecasting integrated with sentiment analytics. While not directly connected with an industry partner during development, the system holds high potential for adoption in fintech, investment advisory, and trading platforms. Its modular architecture supports the inclusion of more data sources or forecasting models, making it a valuable tool for professionals looking to combine quantitative and qualitative data for stock analysis.

Working Model vs. Simulation/Study:

This project extends beyond a theoretical study and includes a fully working model. The SARIMA forecasting model was trained and tested using real historical data, and

sentiment scores were computed using actual news articles fetched via NewsAPI. The predictions and insights are delivered through a deployed Streamlit application, making it a real-time interactive tool rather than a mere simulation.

Project Outcomes and Learnings:

1. Development of a robust SARIMA-based stock price prediction model capable of capturing trends and seasonality.
2. Integration of real-time news sentiment analysis to enhance prediction accuracy and contextual relevance.
3. Creation of an interactive, user-friendly Streamlit web application for stock analysis and forecasting.
4. Accurate short-term stock price predictions with improved performance metrics such as RMSE, MAE, and R^2 .
5. Insightful visualizations of historical data, forecasts, and sentiment trends for better decision-making.
6. Quantification of the impact of news sentiment on stock price movements.
7. Comparative evaluation demonstrating the superiority of SARIMA with sentiment analysis over baseline models.
8. Scalability and modularity of the system, allowing future integration of additional data sources or models.
9. Enhanced understanding of stock market dynamics through combined quantitative and qualitative analyses.
10. A valuable educational tool that bridges the gap between time series forecasting and real-world financial applications.

Future Scope:

The future scope of this project includes:

1. Extend to multilingual sentiment analysis across global markets.
2. Integrate news media data and financial reports for comprehensive prediction.
3. Improve model robustness by incorporating reinforcement learning.
4. Provide mobile application support and alerts for investors.