Project Proposal: Event-Driven Stock Market Prediction Using Artificial Intelligence with Technical Indicators

Dhanush Vemulapalli
AP22110010017
dhanush_vemulapalli@srmap.edu.in

Doondi Venkata Ganesh Chettipalli
AP22110010058
doondivenkataganesh_c@srmap.edu.in

Sriya Palla
AP22110010705
sriya_palla@srmap.edu.in

Niharika Kancharlapalli
AP22110010027
niharika_kancharlapalli@srmap.edu.in

Abstract

The aim of this project is to design and develop a comprehensive stock market analysis and decision-making system that integrates real-time data retrieval, advanced forecasting methods and sentiment analysis to help traders and investors make informed decisions. The inherent complexity and volatility of the stock market requires a multi-dimensional approach that goes beyond traditional analytical methods. By using a combination of technical indicators, fundamental metrics and sentiment-based analysis, this system delivers actionable insights such as price predictions, technical signals and risk management recommendations. The core of this project lies in the synthesis of predictive modeling techniques such as Long Short-Term Memory (LSTM) networks and ARIMA with sentiment analysis from financial news and social media as well as traditional financial metrics such as EPS, P/E and ROE. This integration enables a holistic approach to understanding stock market dynamics. The system forecasts short and medium-term stock prices, assesses market dynamics and trends using technical indicators such as RSI and MACD, and assesses the company's health using fundamental analysis. Sentiment analysis of financial news articles will assess market sentiment and its potential impact on stock trends. A decision engine synthesizes these inputs and generates actionable recommendations such as "Buy," "Hold," or "Sell," accompanied by stop loss and target prices.

This project provides an innovative and adaptable framework for addressing the complexities of the stock market. By combining various methods into a unified system, it provides users with data-driven, risk-balanced decision support.

Keywords : Artificial intelligence, event driven, stock, stock market, sentiment analysis, decision making, technical analysis, indicators, financial news, natural language processing

Disclaimer :

This document is a project proposal designed to outline the preliminary ideas, objectives, and methodologies of the proposed study. The content presented, including the evaluation plan, expected contributions, and references, is subject to change and refinement as the project progresses.

The methods described are proposed approaches and may be adjusted based on new findings, emerging challenges, or additional data obtained during the project process. Similarly, the references cited represent an initial review of relevant literature and may evolve as further resources are identified and analyzed.

The author acknowledges that this proposal is not exhaustive and serves as a foundational blueprint to guide the project process. Stakeholder feedback, peer review, and ongoing learning may lead to modifications in the scope, methods, or focus of the study.

Readers are advised to consider this document as a preliminary framework rather than a finalized project plan.

---

1 Introduction

The stock market is inherently complex and volatile, making decision making a challenging task for traders and investors. Traditional methods rely heavily on technical or fundamental analysis and often neglect sentiment and predictive models. The rise of machine learning and natural language processing (NLP) offers the opportunity to bridge this gap through the use of data-driven techniques. This project proposes a unified system that combines

Short- and medium-term price forecasting : The system will forecast short- and medium-term stock prices.

Fundamental analysis : Evaluate company health through fundamental analysis.

Technical Analysis : The system will assess market momentum and trends using technical indicators like RSI and MACD.

Sentiment analysis : Sentiment analysis of financial news articles will gauge market sentiment and its potential impact on stock trends.

This project provides an innovative and adaptable framework for addressing the complexities of the stock market. By combining different methods into a unified system, it provides users with data-driven, risk-balanced decision support.

---

2 Objectives

1. **Develop a forecasting model** for short- (3 days) and medium-term (8 days) stock price prediction.
2. **Incorporate technical indicators** (e.g., RSI, MACD, OBV) to evaluate stock momentum and trend.
3. **Leverage fundamental analysis** by integrating financial ratios such as EPS, P/E ratio, and ROE.
4. **Perform sentiment analysis** on financial news articles to gauge market sentiment and its impact on stock trends.
5. **Design a decision-making module** that synthesizes all inputs into actionable recommendations (Buy, Hold, Sell).

---

3 Proposed Methodology

3.1 Data Collection :

- **Real-Time Stock Data**: Historical stock prices for the past 21 days are fetched using APIs like Alpha Vantage or Yahoo Finance. These time-series datasets are crucial for calculating technical indicators such as RSI, MACD, and Support/Resistance levels using libraries like TA-Lib or pandas-ta.
- **News Articles**: Relevant financial news and social media updates are scraped to extract textual data. These datasets often contain event headlines with sentiment labels (positive, neutral, or negative), aiding sentiment analysis.
- **Financial Metrics**: Company-specific financial data is collected from reliable sources to enhance contextual understanding.
- **Event-Based Data**: Stock price movements linked to specific events are included for event-driven market analysis.

3.1.1 Dataset Details

- **Stock Market Data**: Time-series stock price data for multiple companies.
- **Sentiment Analysis for Financial News**: Headlines labeled with sentiment for extracting insights.
- **Event-Based Stock Market Dataset**: Mapping specific events to stock price changes.

These datasets collectively support predictive modeling by correlating stock prices, sentiment, and events, enabling more accurate insights into market behavior.

3.2 Short- and Medium-Term Forecasting

This stage focuses on predicting stock prices for short- and medium-term periods using advanced predictive models.

- **Model Selection**: Predictive models such as Long Short-Term Memory (LSTM) networks and AutoRegressive Integrated Moving Average (ARIMA) are employed for forecasting. LSTM is well-suited for capturing temporal dependencies in stock price data, while ARIMA is effective for analyzing linear trends and seasonality.
- **Evaluation Metrics**: Prediction accuracy is evaluated using metrics like Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE). These metrics provide insights into the models' performance, helping to refine them for better accuracy.
- **Outputs**: The models predict stock prices for the next 3 and 8 days. These forecasts are instrumental in determining crucial trading parameters such as stop-loss levels to minimize potential losses and target price levels to maximize gains.

By combining robust models with reliable evaluation metrics, this process enhances the accuracy of short- and medium-term forecasts, empowering traders and investors with actionable insights to make informed decisions in the dynamic stock market.

3.3 Technical Analysis

This phase involves applying technical analysis techniques to identify market trends and make informed trading decisions.

- **Indicators**: Key technical indicators such as Relative Strength Index (RSI), Moving Average Convergence Divergence (MACD), and On-Balance Volume (OBV) are calculated. Dynamic support and resistance levels are also identified to understand price movements and key turning points in the market.
- **Implementation**: Python libraries like TA-Lib and pandas are utilized to compute these indicators efficiently. TA-Lib provides a comprehensive suite of functions for technical analysis, while pandas facilitates data manipulation and integration of calculated indicators with historical stock price data.

By leveraging technical analysis, traders gain valuable insights into market momentum, trend reversals, and potential entry or exit points, which enhances decision-making in both short- and long-term investment strategies.

3.4 Fundamental Analysis

This step focuses on assessing a company's overall financial health and intrinsic value using key financial metrics:

- **Earnings Per Share (EPS)**: A profitability metric that indicates the company's earnings distributed per outstanding share, reflecting its ability to generate profit.
- **Price-to-Earnings (P/E) Ratio**: A valuation metric used to determine how the market values the company relative to its earnings, offering insights into whether the stock is overvalued or undervalued.

- **Return on Equity (ROE)**: A measure of financial efficiency that shows how effectively a company utilizes its equity to generate profits.
- **Dividend Yield**: An indicator of potential income generation, reflecting the ratio of annual dividends to the stock's price.

By combining these financial metrics, fundamental analysis provides a comprehensive understanding of a company's performance, profitability, and growth potential. This enables investors to make informed decisions regarding long-term investments.

3.5 Sentiment Analysis

This stage involves analyzing the sentiment of financial news articles to gauge market sentiment and its potential impact on stock prices.

- **Text Preprocessing**: News articles are tokenized and cleaned by removing stopwords, punctuation, and irrelevant text such as HTML tags and URLs. This step ensures that only meaningful information is retained for analysis.
- **Modeling**: Sentiment classification models are developed to analyze the processed text. Advanced transformer-based models like BERT are fine-tuned for this task to capture contextual nuances effectively. Alternatively, lightweight tools like VADER (Valence Aware Dictionary and sEntiment Reasoner) can be used for quicker sentiment analysis, especially for real-time applications. News data can be retrieved from reliable sources such as the News API for up-to-date inputs.
- **Output**: The analysis generates a sentiment score for each article or headline, categorizing it as positive, neutral, or negative. These scores are used to assess overall market sentiment and its influence on stock movements.

This approach integrates textual sentiment with market trends, providing valuable insights for making informed trading decisions.

3.6 Decision-Making Module

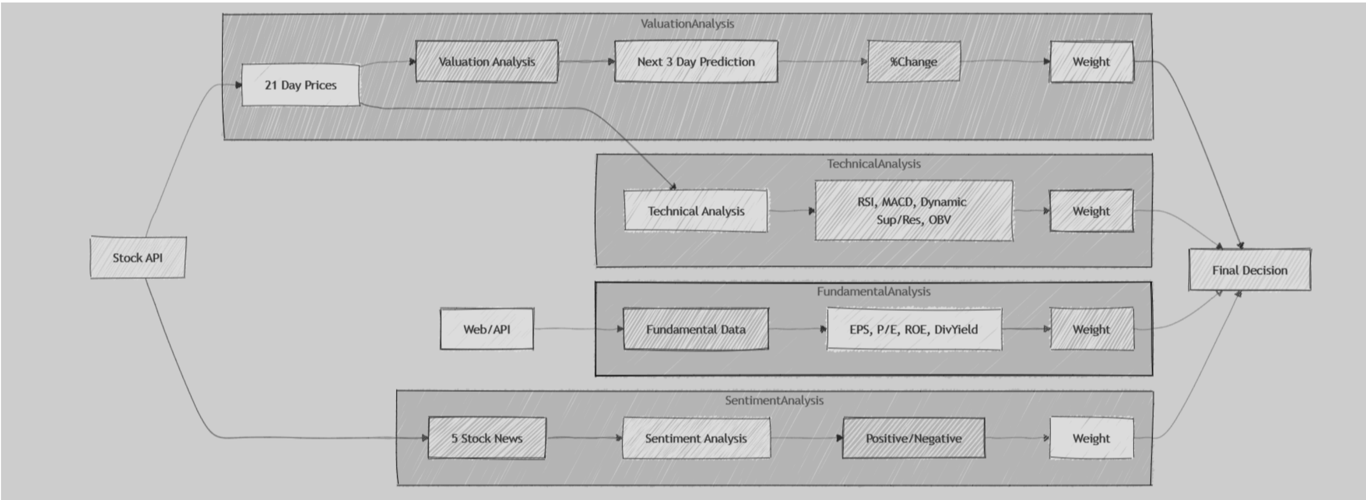
This module synthesizes insights from multiple analyses to provide actionable investment recommendations.

- **Weighted Scoring System**: A scoring framework is developed by assigning weights to various inputs, including:
  - Forecasting outputs (e.g., short- and medium-term price predictions).
  - Technical indicators (e.g., RSI, MACD, and support/resistance levels).
  - Fundamental metrics (e.g., EPS, P/E ratio, ROE, and dividend yield).
  - Sentiment analysis scores (e.g., positive, neutral, or negative sentiments from news articles).
- The weights reflect the relative importance of each input based on historical performance and market conditions.
- **Output Recommendations**: The module integrates the weighted scores to classify stocks into one of three categories:
  - **Buy**: Favorable metrics indicate strong growth potential.
  - **Hold**: Neutral signals suggest monitoring without immediate action.
  - **Sell**: Unfavorable metrics recommend reducing exposure to the stock.

- Additionally, the module provides suggested stop-loss levels to mitigate potential losses and target prices to guide profit-taking decisions.

By combining multiple analysis techniques, this decision-making module offers a comprehensive and balanced approach to stock market investment strategies.

3.7 Design Architecture



---

4 Evaluation Plan

1. **Backtesting**: The forecasting and decision-making models are rigorously tested using historical stock data to simulate real-world trading scenarios. This process helps validate the accuracy and reliability of the models under varying market conditions.
2. **Performance Metrics**:
   - **Prediction Accuracy**: Metrics such as Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) are used to evaluate the precision of short- and medium-term stock price forecasts.
   - **Decision-Making Effectiveness**: The effectiveness of recommendations (Buy, Hold, or Sell) is assessed by calculating key financial outcomes, such as the percentage of profit or loss generated over a given period.
3. **Comparison with Benchmarks**: The system's performance is compared to traditional trading strategies (e.g., buy-and-hold or moving average-based approaches) and existing automated systems. This benchmarking helps highlight the system's competitive advantages and identify areas for improvement.

---

5 Expected Outcomes and Contributions

1. **A Robust, Real-Time Stock Market Analysis Framework**: Development of a comprehensive system capable of analyzing real-time stock market data. The framework seamlessly integrates forecasting, technical analysis, fundamental analysis, and sentiment analysis to deliver actionable insights for investors and traders.
2. **Insights into the Interplay Between Sentiment and Stock Price Movements**: Exploration of the relationship between market sentiment—derived from news articles, social media, and other textual data—and stock price fluctuations. This contributes to a deeper understanding of how sentiment impacts market behavior and enhances predictive modeling.
3. **A Unified Methodology for Combining Diverse Analysis Techniques**: Establishment of a methodology that effectively combines diverse approaches, including machine learning models, technical indicators, and financial metrics, into a cohesive decision-making system. This approach sets a precedent for future systems integrating multiple perspectives for improved accuracy and reliability.

These contributions aim to advance stock market analysis by offering a novel, holistic, and data-driven solution tailored to the needs of modern investors.

---

6 Conclusion

This project focuses on developing a unified system that integrates predictive modeling, sentiment analysis, and financial metrics to facilitate informed stock market decision-making. By combining diverse methodologies into a cohesive framework, the system addresses the complexities of financial markets and provides actionable insights for investors.

The proposed solution offers a novel, data-driven approach to analyzing market trends, incorporating real-time adaptability and a weighted decision-making module. Its potential applications extend beyond trading, contributing to advancements in financial analytics. This work holds promise for publication in esteemed finance and data science conferences or journals, showcasing its innovation and practical relevance.

Preliminary References

1. 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

2. Box, G. E., Jenkins, G. M., & Reinsel, G. C. (2008). *Time Series Analysis: Forecasting and Control*. Wiley.

3. Livieris, I. E., Kiriakidou, N., Pintelas, E., & Pintelas, P. (2020). A CNN-LSTM model for gold price time-series forecasting. *Neural Computing and Applications, 32*(23), 17351–17360.

4. Wilder, J. W. (1978). *New Concepts in Technical Trading Systems*. Trend Research.

5. Murphy, J. J. (1999). *Technical Analysis of the Financial Markets: A Comprehensive Guide to Trading Methods and Applications*. New York Institute of Finance.

6. Achelis, S. B. (2000). *Technical Analysis from A to Z*. McGraw Hill Professional.

7. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of the 2019 Conference of the NAACL-HLT*, 4171–4186.

8. Hutto, C., & Gilbert, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the International AAAI Conference on Web and Social Media, 8*(1), 216–225.

9. Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science, 2*(1), 1–8.

10. Penman, S. H. (2012). *Financial Statement Analysis and Security Valuation* (5th ed.). McGraw-Hill Education.

11. Damodaran, A. (2012). *Investment Valuation: Tools and Techniques for Determining the Value of Any Asset* (3rd ed.). Wiley.

12. Kumar, M., & Thenmozhi, M. (2012). Forecasting stock index movement: A comparison of support vector machines and random forest. *SSRN Electronic Journal.*

13. Cavalcante, R. C., Brasileiro, R. C., Souza, V. L. F., Nobrega, J. P., & Oliveira, A. L. I. (2016). Computational intelligence and financial markets: A survey and future directions. *Expert Systems with Applications, 55*, 194–211.

14. MacKinlay, A. C. (1997). Event studies in economics and finance. *Journal of Economic Literature, 35*(1), 13–39.

15. Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance, 62*(3), 1139–1168.

16. McKinney, W. (2010). Data structures for statistical computing in Python. *Proceedings of the 9th Python in Science Conference*, 51–56.

17. The TA-Lib Library. TA-Lib: Technical Analysis Library Documentation. Retrieved from https://ta-lib.org

18. Alpha Vantage. Alpha Vantage API Documentation. Retrieved from https://www.alphavantage.co

19. News API. Documentation. Retrieved from https://newsapi.org

20. Yahoo Finance. Yahoo Finance API. Retrieved from https://finance.yahoo.com

---

Glossary

1. ARIMA (AutoRegressive Integrated Moving Average): *A statistical model used for analyzing and forecasting time-series data. It captures linear trends and seasonality in data.*

2. BERT (Bidirectional Encoder Representations from Transformers): *A state-of-the-art natural language processing model developed by Google that understands the context of words in a sentence for sentiment analysis and other text-related tasks.*

3. LSTM (Long Short-Term Memory): *A type of neural network architecture designed for sequential data. It is particularly useful for capturing temporal dependencies in time-series data, such as stock prices.*

4. MACD (Moving Average Convergence Divergence): *A technical analysis indicator that shows the relationship between two moving averages of a stock's price. It helps identify trends and potential reversals.*

5. MAPE (Mean Absolute Percentage Error): *A metric used to measure the accuracy of a forecasting model. It calculates the percentage difference between predicted and actual values*.

6. OBV (On-Balance Volume): *A technical indicator that uses volume flow to predict changes in stock price. It reflects the cumulative buying and selling pressure.*

7. RSI (Relative Strength Index): *A technical indicator used to evaluate whether a stock is overbought or oversold. It ranges from 0 to 100.*

8. Sentiment Analysis: *A process that uses natural language processing to determine the sentiment (positive, neutral, or negative) of a given text, such as news articles or social media posts.*

9. Stop-Loss Level: *A predefined price at which a trader exits a trade to prevent further losses. It is a risk management strategy.*

10. Support/Resistance Levels: *Price points on a stock chart where the stock consistently stops falling (support) or rising (resistance). These levels are used to identify potential entry or exit points for trades.*

11. TA-Lib (Technical Analysis Library): *A Python library used for implementing technical analysis indicators like RSI and MACD. It provides functions for analyzing historical stock data.*

12. Technical Indicators: *Statistical measures used in technical analysis to evaluate market trends, momentum, and potential price reversals based on historical stock data.*

13. VADER (Valence Aware Dictionary and sEntiment Reasoner): *A lightweight rule-based sentiment analysis tool that is effective for analyzing the sentiment of social media text and other short-form content.*

14. Weighted Scoring System: *A method for decision-making that assigns different weights to various factors (e.g., technical indicators, sentiment scores) based on their importance to generate actionable recommendations.*
