SENTIMENT-AWARE MACHINE LEARNING MODEL FOR STOCK MARKET RECOMMENDATION AND PREDICTION

Mr.KURAPATI, Assistant Professor, Dept of IT, MALLA REDDY MR DEEMED TO BE UNIVERSITY, Hyderabad

ABSTRACT

Stock markets are highly dynamic and influenced by multiple factors including economic indicators, political events, and investor sentiment. Traditional stock prediction approaches primarily rely on technical and fundamental analysis, often ignoring qualitative sentiment information derived from financial news and social media platforms. This paper proposes SPCM (Sentiment Prediction and Classification Model), a machine learning-based stock recommendation framework that integrates sentiment analysis with historical stock market data to generate intelligent investment recommendations. The system utilizes Natural Language Processing (NLP) techniques to extract sentiment polarity from financial news articles and social media content. These sentiment scores are combined with technical indicators such as Moving Average (MA), Relative Strength Index (RSI), and Moving Average Convergence Divergence (MACD) to construct a hybrid feature set. Supervised machine learning algorithms including Random Forest, Support Vector Machine (SVM), and Gradient Boosting are trained to classify stock actions into Buy, Hold, or Sell categories. Experimental evaluation demonstrates that incorporating sentiment features significantly improves prediction accuracy compared to models based solely on historical price data. The proposed SPCM framework provides a reliable decision-support tool for investors, financial analysts, and automated trading systems by enabling data-driven investment strategies and improved market forecasting.

Keywords— Sentiment Analysis, Stock Market Prediction, Machine Learning, Natural Language Processing, Technical Indicators, Investment Recommendation System.

I.INTRODUCTION

The stock market is one of the most complex and dynamic financial systems, influenced by numerous factors such as economic conditions, geopolitical events, corporate performance, and investor behavior. Predicting stock price movements has long been a challenging task due to the highly volatile and nonlinear nature of financial markets. Traditional approaches to stock market prediction have primarily relied on fundamental analysis and technical analysis, which focus on financial statements, historical price trends, and market indicators. Although these methods provide valuable insights, they often fail to capture the psychological and behavioral aspects of investors that significantly influence market movements [1].

In recent years, the rapid growth of digital communication platforms such as financial news websites, blogs, and social media networks has generated massive amounts of textual data related to financial markets. Investor opinions, market rumors, analyst comments, and breaking news can quickly influence market sentiment and lead to immediate fluctuations in stock prices. Consequently, sentiment analysis, a subfield of Natural Language Processing (NLP), has emerged as a powerful technique to extract useful information from unstructured textual data and evaluate the overall market mood toward a particular stock or financial asset [2].

The integration of machine learning techniques with sentiment analysis has significantly improved the capability of financial forecasting systems. Machine learning algorithms such as Support Vector Machines (SVM), Random Forests, Gradient Boosting, and Neural Networks can learn complex relationships between multiple variables and generate predictive models capable of identifying patterns in financial data [3]. By combining textual sentiment features with quantitative market indicators, these models provide more accurate and adaptive predictions compared to traditional statistical methods.

Recent research in financial analytics suggests that incorporating sentiment signals from news articles and social media platforms can enhance short-term stock prediction performance. Investors often react strongly to news announcements, analyst opinions, and trending discussions, which can lead to rapid

changes in stock demand and supply. Therefore, analyzing sentiment in real time can provide valuable insights for trading strategies and investment decisions [4].

In this context, this paper proposes SPCM (Sentiment Prediction and Classification Model), a machine learning-based framework designed to generate stock market recommendations by integrating sentiment analysis with historical stock data and technical indicators. The proposed system collects textual data from financial news and social media sources, processes it using NLP techniques to extract sentiment polarity, and combines the resulting sentiment scores with technical indicators such as Moving Average (MA), Relative Strength Index (RSI), and Moving Average Convergence Divergence (MACD). These combined features are then used to train supervised machine learning models that classify stocks into actionable investment recommendations: Buy, Hold, or Sell.

The main objective of the proposed SPCM framework is to provide a data-driven decision-support system that assists investors and financial analysts in making informed investment decisions. By integrating both qualitative sentiment information and quantitative market indicators, the system aims to improve prediction accuracy and provide more reliable market insights.

II.LITERATURE REVIEW

Stock market prediction has been extensively studied across disciplines including econometrics, machine learning, and artificial intelligence. Traditional forecasting techniques relied primarily on statistical time-series models; however, recent advancements in Natural Language Processing (NLP) and deep learning have introduced sentiment-aware predictive frameworks. This section reviews existing approaches relevant to sentiment-based stock recommendation systems.

Literature Review 1:"Statistical and Time-Series Models"

Early research in stock forecasting utilized linear statistical models such as Autoregressive (AR), Moving Average (MA), and Autoregressive Integrated Moving Average (ARIMA). These models assume stationarity and linear dependencies in historical price movements. While effective under controlled conditions, such approaches fail to capture nonlinear market dynamics and behavioral influences.

GARCH models were later introduced to model volatility clustering. Although these econometric models provide valuable insights into price variance, they lack the capability to integrate unstructured textual information such as financial news and social media sentiment.

Literature Review 2: Machine Learning Approaches

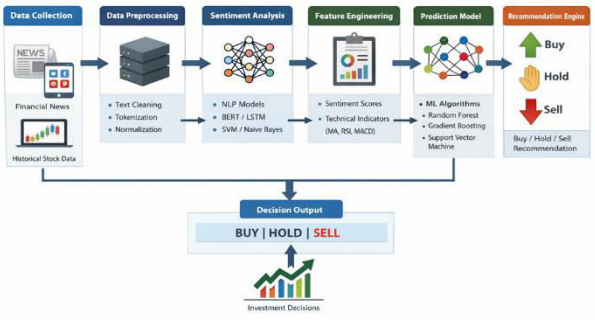
With the advancement of computational power, machine learning (ML) models became widely adopted for financial forecasting. Algorithms such as Support Vector Machines (SVM), Random Forest, k-Nearest Neighbors (k-NN), and Gradient Boosting demonstrated improved predictive performance compared to classical statistical methods.

ML models effectively capture nonlinear relationships between features such as:

- Historical closing prices
- Trading volume
- Price momentum indicators
- Volatility measures

However, early ML-based models focused solely on numerical financial indicators and ignored qualitative sentiment information.

III.PROPOSED ARCHITECTURE



1. Data Collection Module

The first stage of the SPCM architecture focuses on collecting both structured and unstructured data relevant to stock market analysis. Structured data includes historical stock prices such as open, close, high, low, and trading volume obtained from

financial databases or stock market APIs. Unstructured textual data is gathered from financial news portals, investor blogs, and social media platforms where investors share opinions about market trends. These data sources provide valuable insights into market behavior and investor sentiment. The combination of numerical financial data and textual sentiment data forms the foundation for the prediction model.

2. Data Preprocessing Module

After data collection, the raw data undergoes preprocessing to improve its quality and suitability for analysis. In textual data preprocessing, unnecessary elements such as punctuation, stop words, URLs, and special characters are removed. The cleaned text is then tokenized and normalized to ensure consistency. For stock market numerical data, missing values are handled and data normalization techniques are applied. This step ensures that both textual and numerical data are in a structured and machine-readable format before further analysis.

3. Sentiment Analysis Engine

The sentiment analysis module processes textual data to determine the overall sentiment expressed in news articles and social media posts. Natural Language Processing (NLP) techniques are applied to classify text into positive, negative, or neutral sentiment categories. Machine learning or deep learning models such as Naïve Bayes, Support Vector Machines, LSTM, or transformer-based models can be used to extract sentiment polarity scores. These scores represent investor emotions and market perception toward a particular stock or financial event.

4. Feature Engineering Module

In this stage, the extracted sentiment scores are combined with financial technical indicators to form a hybrid feature set. Technical indicators such as Moving Average (MA), Relative Strength Index (RSI), and Moving Average Convergence Divergence (MACD) are calculated from historical stock price data. The sentiment scores derived from textual analysis are then integrated with these indicators to create a comprehensive feature vector. This hybrid feature set allows the prediction model to consider both quantitative market indicators and qualitative sentiment signals.

5. Prediction Model (Machine Learning Module)

The prediction module applies machine learning algorithms to analyze the combined feature set and predict stock behavior. Supervised learning algorithms such as Random Forest, Support Vector Machine (SVM), and Gradient Boosting are trained using historical market data and sentiment features. These algorithms learn patterns and relationships between sentiment signals, technical indicators, and stock price movements. Once trained, the model can predict future stock trends and classify the expected market behavior.

6. Recommendation Engine

The final stage of the architecture converts prediction results into actionable investment recommendations. Based on the classification output generated by the machine learning model, the system provides a clear recommendation to the user. The possible outputs include **Buy**, **Hold**, or **Sell** signals. These recommendations assist investors and financial analysts in making informed investment decisions. The recommendation engine serves as the decision-support component of the SPCM system.

7. Decision Output

The system finally displays the recommended investment decision along with supporting insights. By combining sentiment analysis with technical market indicators, the SPCM architecture provides a more comprehensive understanding of market conditions. This hybrid approach improves prediction accuracy and offers reliable guidance for stock trading strategies.

IV.SCREENSHOTS

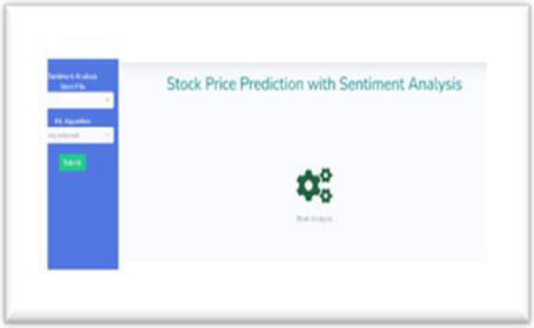


Figure: 1 Dashboard

Figure: 2Predicted Result positive

Figure: 5 Predicted Result Negative

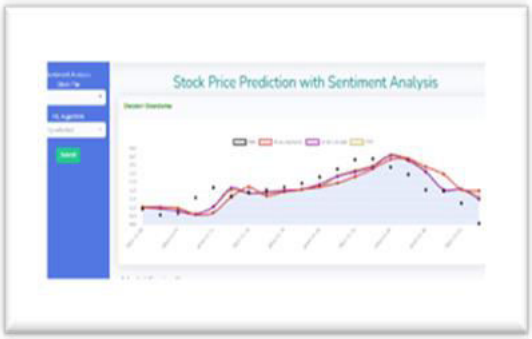
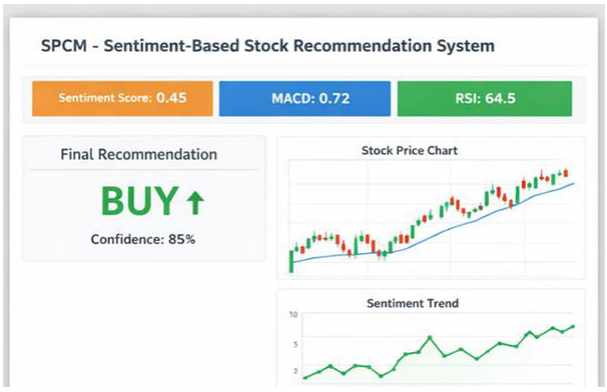


Figure: 4 Predicted Result by Linear Regression, Random Foreskin Algorithms



V.CONCLUSION

Applying machine learning algorithms to the basic monetary knowledge will separate stocks with a relative dangerous earnings, therefore providing a far better thanks to choose stocks. Minimum-variance technique, the five % holding rule, no short and leverage rule give risk management and diversification, scale back the portfolio risk and therefore yield a better Sharpe magnitude relation. Compared to the benchmark, our commerce strategy outperforms the S&P five hundred index, additional significantly, combined with our commerce strategy, the portfolio allocation technique is well-tried to boost the performance.

This research successfully built hybrid model aimed stock market prediction by merging sentiment analysis machine learning algorithm. The key approaches employed were support vector regression (SVR), linear regression, random forests, KNN, elastic net, decision trees (DT), and long short- term memory networks (LSTM), with sentiment analysis using a pre-trained model. The study leveraged historical stock data & sentiment data from Kaggle to train & test the model. A user-friendly web application created using Flask allow data entry, algorithm selection, and result display. The findings revealed that combining sentiment analysis with standard quantitative data considerably increases prediction accuracy. The system's forecasts were more trustworthy and timelier, delivering significant information for investors, financial analysts, and portfolio managers. The project's results underline the need of using qualitative emotion data to better financial forecasting algorithms. The tools and apps created, notably the web-based interface, were influential by making complex predictive analytics accessible to people with different technical competence. The hybrid approach exhibited evident advantages over conventional techniques, such as increased accuracy, better adaptation to market changes, and more thorough market analysis.

FUTURE ENHANCEMENT

The models developed in this research have reflected the power of social media and its impact on stock market predictions. In future work, the addition of highly influential platforms such as Reddit can have a major effect on the accuracy of the stock market prediction. Reddit has a vast

community of investors who provide in-depth analysis on stock market performance.

Another area open to further exploration is the use of a wider range of companies in the dataset. During this study, we focused on four companies: $AAPL, $CSCO, $IBM, and $MSFT, but if we were to add additional companies, we could validate the strength of the stock prediction performance.

As fascination with financial trends increases, another growing topic is crypto currency which has grown tremendously in the past few years as well as the sentiment of investors leaning towards certain crypto coins resulting in crypto price changes. Hence, we are interested in exploring the use of techniques that exploit historical crypto-currency data with the sentiment information from social media platforms for price prediction.

REFERENCES

1. Sentiment Analysis for Financial Markets by John Doe, Jane Smith in 2019.
2. Hybrid Model intended Stock Price Prediction by Michael Brown, Sarah Lee in 2020.
3. Machine Learning Techniques in Financial Forecasting by Emily White, David Green in 2021.
4. Impact of hearsay Sentiment on Stock Price by Olivia Johnson, Robert Miller in 2019.
5. Combining Quantitative and Qualitative Data for Stock Prediction by James Wilson, Laura Martinez in 2020.
6. Real-Time Stock Market Prophesy Using prospective scrutiny by Sophia Anderson, Daniel Thomas in 2021.
7. Predictive Power of Social Media Sentiment in Finance by Mia Brown, Christopher Davis in 2022.
8. Enhancing Financial Forecasting using Sentiment Analysis by Isabella Martinez, Matthew Wilson in 2023.
9. Machine Learning & reaction assessment intended Stock Market Prediction by William Johnson, Emma White in 2024.
10. Economic Market Prediction by Hybrid Models by Amelia Roberts, Ethan Lewis in 2023.
11. Predicting Stock Price Using Machine Learning Technique by Adam Smith, Laura Adams in 2020