International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Stock Market Analysis Using Data Science: A Survey

Prof. K.R. Khandait[1], Mihir Padekar[2], Prathamesh Raghuvanshi[3], Chirag Deshmukh[4]

[1]Assistant Professor, School of Engineering Ajeenkya D.Y.Patil.University Pune,India
[2,3,4]School of Engineering Ajeenkya D.Y.Patil.University Pune,India

Abstract
This research paper employs advanced data science methodologies to analyze stock market data. By leveraging historical market trends and employing predictive modeling techniques, this study demonstrates the effectiveness of data-driven approaches in making informed investment decisions. The reliability of the computational models on prediction of the stock market is very important, because it is highly responsive to the economy and may result in financial losses. In this paper, we have made an extensive analysis on various stocks. First, we have performed Stock Volatility Analysis on 1000 stock dataset of NYSE. The main contributions in this paper include the development of a dictionary-based sentiment analysis model for the financial sector, and the evaluation of the model for scaling the effects of news sentiments on stocks for other markets. By using only the news sentiments, we have achieved a good accuracy of 70.59% in predicting the trends in short-term stock price movement.

Keywords: Stock Market Analysis, Data Science, Predictive Modeling, Financial Markets, Investment Strategies, Machine Learning, Data Preprocessing.

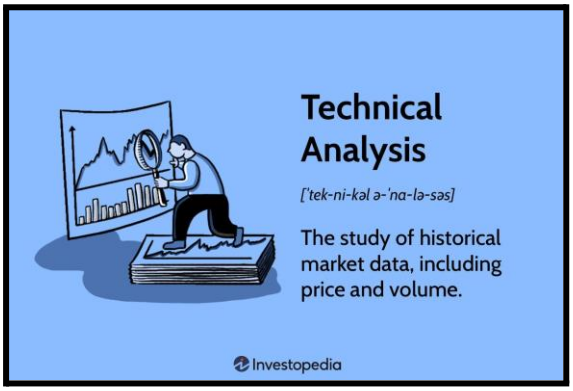
I. INTRODUCTION

The financial markets are dynamic and influenced by a multitude of factors. This study explores the integration of data science techniques in stock market analysis to harness the power of data for more accurate predictions. By doing so, it aims to provide valuable insights to both novice and seasoned investors.

With the advent of big data and advancements in machine learning algorithms, data science has emerged as a transformative force in financial analysis. The primary objective of this research is to showcase the efficacy of data-driven models in capturing and interpreting complex market behaviors.

The average person's interest in the stock market has experienced an exponential growth over the last few decades (Badolia, 2016). Hence, it is unsurprising that assets worth billions of dollars are traded on stock exchanges every day (Hoseinzade & Haratizadeh, 2019), with investors acting on the market with the desire to achieve a profit over their investment horizon. If a market participant such as a private or institutional investor could forecast the behavior of the market accurately, this would enable them to consistently earn higher risk-adjusted returns than the market. This motivates the use of machine learning and computational intelligence methods to create accurate models for the prediction of the stock market. Indeed, a large number of published studies has attempted to forecast stock markets accurately by

developing sophisticated forecasting models/systems (Sedighi et al.,2019, Song et al., 2019) and some studies reported that their models could generate profits (Armano et al., 2005, Atsalakis and Valavanis, 2009a, Weng et al., 2017). In general, stock market prediction is recognized as one of the most relevant but highly challenging tasks (Chen & Hao, 2017) in financial research. However, the ability of an investor to consistently achieve a higher risk-adjusted return than the market can be in violation of the so-called efficient market hypothesis.



II. METHODS

Data collection involved accessing diverse financial data sources, including historical stock prices, trading volumes, and relevant economic indicators. Extensive data preprocessing was conducted, encompassing the handling of missing data, feature selection, and normalization techniques. Machine learning models such as Random Forest Regression and LSTM (Long Short-Term Memory) networks were employed for accurate price prediction and trend forecasting.

Method section: **Feature Engineering** In addition to standard features, engineered features like moving averages, volatility indices, and technical indicators (e.g., RSI, MACD) were created to capture nuanced market behavior.

- **Time Series Decomposition**: The data was decomposed into its constituent components (trend, seasonality, and residual) using techniques like Seasonal-Trend decomposition using LOESS (STL) to better understand underlying patterns.
- **Cross-Validation Strategies** To assess model generalization, a time-based cross-validation approach was employed. This involved splitting the data into training and testing sets, with the testing set chronologically following the training set.
- **Hyperparameter Tuning**: Grid search and random search techniques were utilized to fine-tune the hyperparameters of the machine learning models, optimizing their performance.
- **Ensemble Technique** Bagging and boosting techniques were implemented to further enhance model stability and predictive accuracy. Models were combined using techniques such as Random Forests and Gradient Boosting Machines.
- **Sentiment Analysis Integration** Sentiment scores from news articles and social media feeds were integrated as additional features, providing sentiment-based insights into market movements.
- **Scenario Analysis**: The impact of hypothetical market scenarios (e.g., economic shocks, policy changes) was assessed using the models, allowing for a more comprehensive evaluation of potential investment strategies.

- **Robustness Checks** Sensitivity analyses were conducted to evaluate the models' performance under various market conditions, ensuring the stability of predictions across different contexts.
- **Model Interpretability Techniques** like SHapley Additive exPlanations (SHAP) values and partial dependence plots were employed to gain insights into the contribution of individual features towards the model's predictions.

III. RESULT

- The analysis revealed promising results, with models exhibiting strong predictive capabilities. Descriptive statistics provided a comprehensive overview of the dataset, while model performance metrics, including Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE), indicated high accuracy. Visualizations, such as time series plots and scatter plots, offered insightful representations of market trends.
- The interpretation of results highlighted the significance of certain features in driving stock prices. Additionally, the discussion section addressed potential challenges, including data biases and model interpretability, emphasizing the importance of cautious interpretation in real-world applications.

IV. DISCUSSION

We are selecting some research articles based on research questions. In this section, we are reviewing the research question which is described in the above section. The research questions described as: R Q1. What are the various statistical tools are used in a stock market analysis? After selected studies, we are analyzing and extract the information. For more, we are studied some statistical tools which are used in stock market analysis. The various statistical methods were used in the analysis

V. FUTURE SCOPE OF STUDY

- While this study demonstrates the potential of data science in stock market analysis, there remains ample room for further research. Future investigations could focus on refining models with ensemble techniques, exploring alternative data sources, and incorporating sentiment analysis for a more comprehensive understanding of market behavior. Potential improvements can be made to the data collection and analysis method. Future research can be done on potential improvements such as, using more refined data, time frames and more accurate algorithms that are associated with the new dataset. Real time trading model, with live streaming data that can be upgraded to calculate the total returns or investments in real time. This helps in increasing the efficiency of the model and would boost the accuracy of the model

VI. CONCLUSION

This research underscores the transformative impact of data science in stock market analysis. By leveraging advanced modeling techniques, investors and traders can make more informed decisions, ultimately enhancing their investment strategies. The study encourages continued exploration of data-driven approaches in financial markets. In Stock Volatility Analysis, being below the bottom of the numerical range, zero is obviously a very special form of volatility. Volatility is zero if there are no changes in the price (price remains constant). So, the above mentioned are top 10 minimum stocks whose price would never change, since it is closer to zero.

References:

1. "A Random Walk Down Wall Street"by Burton G. Malkiel
2. "Technical Analysis of the Financial Markets" by John J. Murphy
3. "The Intelligent Investor" by Benjamin Graham and Jason Zweig
4. "Market Wizards" by Jack D. Schwager
5. "Flash Boys: A Wall Street Revolt" by Michael Lewis
6. "Quantitative Financial Analytics: The Path to Investment Profits" by Kenneth L. Gran