IJRAR.ORG E-ISSN: 2348-1269, P-ISSN: 2349-5138

INTERNATIONAL JOURNAL OF RESEARCH AND ANALYTICAL REVIEWS (IJRAR) | IJRAR.ORG

An International Open Access, Peer-reviewed, Refereed Journal

Stock Price Analysis Through Machine Learning Models

Prasanta Kumar Sahoo, [2]Debabrath Sahoo

[1] Professor, School of Engineering, Dept. of CSE, Anurag University, Hyderabad, Telangana.

[2] B. Tech Student, Dept. of CSE-Data Science, Lovely Professional University, Punjab.

ABSTRACT: In this era of digital economy like India mostly young generation in the society are now interested in investing in share market and stock exchanges because of the lucrative financial gain. In Indian stock market retail participants holds a major portion of share other than foreign institutional investor and domestic institutional investor. Accurate stock price prediction has long been a critical area of research in the finance sector, as it can inform investment strategies and decision-making processes. Predicting how the stock market will perform in future is one of the most difficult things to do. Prediction of stock market returns is a very complex issue as it depends on so many factors such company financial status, global economy, national growth and government policy etc. These days' stock prices are affected due to many reasons like company related news, political, social economic conditions and natural disasters. There are a lot of research is taking place to predict the value of stock market index and also day to day movement of the index. Many models are available and were developed over a period of time to predict the value of a particular stock price over a period of time future but could not able to predict accurately as it fails to understand the market sentiment. This research paper implemented many models such as auto regressive model, KNN and LSTM for the prediction of stock prices. This research paper work also computed the accuracy of the prediction by comparing the predicted values with the actual values over a period of time.

Key Words: Stock Price, Machine Learning, ARIMA

I. INTRODUCTION

The stock price prediction is always a critical area of focus in finance, as it can suggest investment strategies. In this research paper use of machine learning techniques explored to forecast future stock prices by leveraging historical market data. Predicting stock prices is a highly complex and challenging in the field of finance. Stock prices are influenced by a wide range of factors, including company performance, market sentiment, economic indicators, and global events. Thus the prediction of stock price is a very complex task and has becoming more and more difficult today than earlier days. Now a day's the price of a particular stock depends on so many factors such as news related to company, political scenario and natural disasters etc. The prediction of stock price is one of the most important issues to be investigated in academic and financial researches [1]. Today stock price prediction is becoming very complex than before as stock prices are not only affected due to company's financial position but also due to socio economical condition of the country, national growth, political atmosphere, global economy and natural disasters etc. The return from the share market is always uncertain and ambiguity in nature hence traditional techniques will not give accurate prediction [2]. At the age of AI, due to data analytics and rapid data processing has made the stock prices to fluctuate at a very rapid rate. The investor does have an idea how the prices of a particular stocks are going to increase or decrease on the next couple of days before investing in any stock. One need to combine several computing techniques in order to predict the price of a stock and the nature of the stock market with some accuracy. Over the period of time the very traditional capital market concept has changed and various new methods of analytics have been emerged [3]. There are so many people are earning their bread and butter by investing in stock market in a day to day basis. So accurate prediction of stock market price is very important for their survival [4]. There is many research is going on over many years to forecast the stock prices or stock index. There is important information such as economic variables such as interest rates and exchange rates, industry specific information such as consumer price index and industrial production growth rate and company specific information such as profit loss statements, revenue income, margin and dividend yields which are very much necessary

for the prediction of stock price. The value of the share increases as many people buy it and decreases as many people sell it. A broker helps the customers by making analysis and suggest good stocks to buy. Various important technical, fundamental and statistical indicators have been used for analysis and used for prediction of results. However, not even one technique or combination of many techniques has been successful to beat the market. In general, technical analysis approach [5, 6, 7 and 8], that predicts stock prices based on historical prices and volume, the Dow Theory, basic trends, price patterns over a period of times, graphs and oscillators, is commonly used by stock analyst for prediction of stock prices. Advanced artificial intelligent techniques ranging from pure mathematical models and expert systems [9, 10] to neural networks [11, 12, 13, 14, 15, 16 and 17] have also been used by many analysts to predict stock price over a period of time. Ultimately, most of the researchers have developed variety of mathematical methods to predict future share price using artificial neural network. In this research work the future price of one stock is predicted based on its past day prices as well as other related present day prices. Artificial neural network is a subset of artificial intelligence where artificial neural network back propagation algorithm is used with the feed forward neural network to predict the price of a stock market. In [18] the researcher analyzes and forecast the stock market index with Markov properties. In [19] the author has proposed a new model to predict stock price. An HMM is a state machine for a system based on Markov process with unobserved states.

II. LITERATURE REVIEW

Predicting stock prices is a highly complex and challenging in the field of finance. Stock prices are influenced by a wide range of factors, including company performance, market sentiment, economic indicators, and global events. Accurately forecasting these prices can provide traders and investors with a significant edge in the market, helping them make good decisions. However, due to the inherent volatility and non-linearity of financial markets, traditional statistical methods often fall short in capturing the complexities of stock price movements. In recent years, advancements in machine learning and data science have opened new avenues for improving stock price prediction accuracy. These techniques allow for the integration of vast amounts of historical data and the extraction of complex patterns that may not be apparent with conventional approaches. Machine learning is the method of exploring and modeling a huge amount of data. It helps to sorts out the data through large stock market data sets to find patterns and analyze the relationship between the attributes. Machine Learning is the study of collecting, cleaning, pre-processing, analyzing, gaining useful insight information from data [20]. Stock price prediction has long been a critical area of research due to the immense potential it holds for traders and investors. Financial markets are known for their volatility, driven by a wide range of factors including economic indicators, investor behavior, and global events. As a result, numerous models have been developed over the years to understand and predict stock price movements. Among these, time series models have emerged as one of the most effective methods for analyzing historical stock prices to forecast future values.

A) Traditional Time Series Models

Time series analysis is crucial for financial forecasting as it captures temporal patterns within data. One of the most widely used models in this domain is the Autoregressive Integrated Moving Average (ARIMA) model. ARIMA is a very simple model and able to handle univariate data with linear trends.

It integrates three components:

- Autoregression (AR)
- Differencing (I) to achieve stationarity
- Moving averages (MA) to address errors.

ARIMA models have been extensively applied in financial forecasting due to their robustness in capturing linear relationships in data. However, the model has limitations in handling non-linear dynamics and is often insufficient for complex, volatile stock market data. To address seasonal patterns in time series data, the Seasonal ARIMA (SARIMA) model was developed as an extension of ARIMA.

- SARIMA incorporates seasonal differencing along with seasonal autoregressive and moving average components. This makes it suitable for stocks that show cyclic patterns, such as those tied to quarterly or annual financial events. However, SARIMA shares similar limitations with ARIMA when it comes to non-linearity and abrupt market changes.

- Another traditional model used in stock price forecasting is the Exponential Smoothing State Space (ETS) model, which focuses on trend and seasonality by applying exponentially weighted averages to historical data.

- This model is particularly effective when there is a clear trend in stock prices, as it gives more importance to recent data. Despite this, ETS is often outperformed by more advanced models, especially in highly volatile or unpredictable markets.

B) Volatility Modeling

A significant characteristic of financial markets is their time-varying volatility. Traditional models such as ARIMA and SARIMA do not explicitly account for changes in volatility, which can affect prediction accuracy. To overcome this limitation, the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model was introduced. GARCH models the variance of the residual errors over time, capturing the volatility clustering commonly seen in stock prices. This is particularly useful in understanding market risks and can improve financial decision-making in areas such as option pricing. However, GARCH models focus on volatility prediction rather than forecasting actual price levels, making them more suitable for risk management than price prediction.

C) Machine Learning Approaches

In recent years, machine learning has gained prominence in financial forecasting, with models like Long Short-Term Memory (LSTM) networks demonstrating significant potential.

- LSTM, a type of recurrent neural network (RNN), excels at capturing long-term dependencies in sequential data, making it highly effective for stock price prediction. Unlike traditional models, LSTM can capture complex, non-linear relationships and can remember important patterns over long periods.

This gives it an advantage in forecasting highly dynamic financial markets where price movements are influenced by a range of external factors.

Despite their strengths, machine learning models like LSTM come with challenges. They require large datasets for training and are often criticized for their "black box" nature, which makes them difficult to interpret.

D) Comparative Studies

Several comparative studies have been conducted to evaluate the effectiveness of different time series models for stock price forecasting. For example, Zhang (2017) demonstrated that traditional models such as ARIMA work well for short-term forecasting, particularly when there is minimal volatility [21]. However, the study found that machine learning models, especially LSTM, perform better over the long term due to their ability to capture complex patterns and non-linear relationships. Similarly, Patel (2015) suggested that hybrid approaches, which combine traditional models like ARIMA with machine learning techniques, often yield superior results by leveraging the strengths of both methodologies [22].

E) Hybrid Models

Recent research has explored the potential of hybrid models that combine the strengths of traditional statistical techniques and modern machine learning algorithms. For example, integrating ARIMA with LSTM can yield better forecasting results by utilizing ARIMA's interpretability for linear trends and LSTM's ability to capture complex non-linearities.

III. DATA SET DESCRIPTION

The stock data is collected from New York Stock Exchange for our investigation. In The study utilizes data from major companies, including TSLA(Tesla), AAPL(Apple), and EW-MAX, which contain key features such as Open, High, Low, Close, Adjusted Close, and Volume. In addition to these core features, technical indicators such as Moving Averages, Bollinger Bands, and the Relative Strength Index (RSI) are calculated to provide deeper insights into market trends and momentum. The project implements various time series forecasting methods, including ARIMA, and other machine learning models, to compare their effectiveness in predicting short-term price movements. The primary objective is to develop a robust model that can accurately predict future stock prices with minimal error. The model is evaluated based on metrics such as Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) to ensure its reliability. Additionally, we explore the impact of external factors such as trading volume and broader economic indicators on stock price movements. This project also emphasizes the importance of proper data preprocessing, including handling missing data, scaling features, and managing time series data complexities. Through this approach, we aim to provide a valuable tool for traders and investors to anticipate market fluctuations and enhance their decision-making processes. By utilizing historical price data and incorporating additional technical indicators, machine learning models can be trained to recognize trends and predict future price movements. This research aims to develop a predictive model for stock prices using historical data from prominent companies such as Tesla (TSLA), Apple (AAPL), and EW-MAX. The model will be built using time series forecasting techniques, complemented by feature engineering to enhance predictive capabilities. The dataset used in this project comprises historical stock price data for three major companies:

- Tesla (TSLA)
- Apple (AAPL)
- EW-MAX

A. Key Features

The primary features in the dataset for each stock are as follows:

- Date: The date corresponding to the stock price data.

- Open: The stock price at the beginning of the trading day.

- High: The highest price of the stock during the trading day.

- Low: The lowest price of the stock during the trading day.

- Close: The stock price at the end of the trading day.

- Adjusted Close: The closing price adjusted for stock splits, dividends, and other corporate actions. This feature reflects the true value of the stock for long-term analysis.

- Volume: The total number of shares traded during the trading day.

Each of these features provides valuable insights into stock performance over time, which is crucial for accurate time series forecasting and model training.

B. Data Sources

The stock price data for Tesla, Apple, and EW-MAX was collected from publicly available financial platforms. This data spans multiple years, providing a rich historical dataset for time series analysis.

- TSLA.csv: Contains stock data for Tesla, a leading electric vehicle manufacturer.
- AAPL.csv: Contains stock data for Apple, a global technology company known for its consumer electronics and software services.
- EW-MAX.csv: Contains stock data for EW-MAX, which is presumed to represent another entity in this project (details can be clarified based on the context of the project).

C. Technical Indicators

In addition to the core stock price features, technical indicators are derived to gain deeper insights into market trends and momentum. These indicators are calculated using the raw stock price data and include:

- Moving Averages (MA): A smoothing technique that helps identify trends by averaging the stock's closing prices over a specific number of days (e.g., 5-day, 10-day, or 50-day moving averages).
- Relative Strength Index (RSI): A momentum oscillator that measures the speed and change of price movements. It oscillates between 0 and 100 and is used to identify overbought or oversold conditions in a stock.

These technical indicators are integrated into the dataset to enhance the predictive capabilities of the models, as they can provide additional insights into stock price trends and reversals.

D. Data Preprocessing

Before the data is used for modeling, several preprocessing steps are applied:

- Handling Missing Data: Any missing values in the dataset are either imputed or removed to ensure data completeness.
- Scaling: Features like stock prices and volume are often on different scales, which can affect the performance of machine learning models. To address this, data is normalized or scaled using methods such as Min-Max scaling or Standardization.
- Time Series Indexing: Since stock prices follow a temporal order, the `Date` column is converted to a proper time index to maintain the sequence of the data for time series forecasting.

E. Dataset Summary

- Time Period: The dataset covers a range of dates that reflect the historical performance of each stock over several years, ensuring enough data points for accurate forecasting.
- Data Granularity: The data is collected at the daily level, which is suitable for short-term forecasting and understanding daily stock price movements.
- Number of Records: Each dataset contains thousands of records corresponding to each trading day over the chosen time period. This ensures the models have sufficient data for training and evaluation.

IV. PROPOSED WORK

The primary objective of this project is to build a robust and accurate stock price prediction model that can be utilized by traders and investors to make more informed decisions. The objective is to develop a simple model that can accurately predict future stock prices with minimal error. The model is evaluated based on metrics such as Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) to ensure its reliability. Additionally, the impact of external factors such as trading volume and broader economic indicators on stock price movements is also considered. The research work implements various time series forecasting methods, including ARIMA, and other machine learning models, to compare their effectiveness in predicting short-term price movements. This project also emphasizes the importance of proper data preprocessing, including handling missing data, scaling features, and managing time series data complexities. Through this approach, we aim to provide a valuable tool for traders and investors to anticipate market fluctuations and enhance their decision-making processes. By utilizing historical price data and incorporating additional technical indicators, machine learning models can be trained to recognize trends and predict future price movements. This research aims to develop a predictive model for stock prices using historical data from prominent companies such as Tesla (TSLA), Apple (AAPL), and EW-MAX. The model will be built using time series forecasting techniques, complemented by feature engineering to enhance predictive capabilities. In order to enhance the prediction capabilities, technical indicators such as Moving Averages, Relative Strength Index (RSI), Bollinger Bands, and others are included to capture market trends and momentum. Evaluation metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared are used to assess the accuracy and reliability of each model, ensuring that the final model is robust and effective. Various models are compared including traditional statistical methods (e.g., ARIMA) and advanced machine learning techniques (e.g. Random Forest, Gradient Boosting) to evaluate best performance.

V. CONCLUSION

Most of the people have taken trading in stock market as their live hood and others invest their hard earned money in stock market for investment purpose. So it is very important to build a robust and accurate stock price prediction model. The existing literature on stock price forecasting highlights the strengths and limitations of both traditional time series models and modern machine learning approaches. While models like ARIMA and SARIMA are effective in capturing linear patterns and seasonality, they often struggle with non-linearity and market volatility. Machine learning models, particularly LSTM networks, offer more sophisticated techniques for capturing these complexities but come with challenges in terms of data requirements and interpretability. As the field continues to evolve, hybrid models that combine the strengths of both traditional and machine learning approaches may offer the best solutions for stock price prediction in volatile financial markets. By analyzing past trends and leveraging the power of machine learning, the project aims to contribute to the growing field of quantitative finance and demonstrate the potential of data-driven approaches in financial forecasting

VI. REFERENCES

[1]. Kaya. M.I.Y., Karsligil. M.E. in 2ND IEEE International Conference on Information and Financial Engineering (ICIFE), pages 478-482, 17th to 19th September 2010.

[2]. Dr. P. K. Sahoo, Mr. Krishna charlapally, “Stock Price Prediction Using Regression Analysis”, International Journal of Scientific & Engineering Research, Volume 6, Issue 3, March-2015.

[3]. Podding, T and Rehkugler, H, “A world of integrated financial markets using artificial neural networks”, Nerocomputing, pages 251-273, 1996.

[4]. Alle Bhavana Kumari , Prasanta Kumar Sahoo, “A NOVEL MODEL TO PREDICT STOCK MARKET PRICE”, High Technology Letters ISSN NO : 1006-6748, Volume 26, Issue 12, 2020.

[5]. Murphy J.J., “Technical Analysis of Financial Markets – A Comprehensive Guide to Trading Methods and Applications”, New York Institute of Finance, 1999.

{6]. Trippi, R.R., DeSieno, D., “Trading Equity Index Futures with a Neural Network”, the Journal of Portfolio Management, pp. 27-33, fall 1992.

[7]. Kryzanowski, L., Galler, M., Wright, D.W., “Using Artificial Networks to Pick Stocks”, Financial Analyst s Journal, pages 21-27, August 1993.

[8]. Li, E.Y., “Artificial Neural Networks and Their Business Applications”, Information & Management, vol. 27, pages 303-313, 1994.

[9]. Han J., Lu H., Feng L. "Stock Movement Prediction and N Dimensional Inter-Transaction Association Rules", Proceedings of the SIGMOD'96 Workshop on Research Issues on Data Mining and Knowledge Discovery (DMKD'98), 1998.

[10]. John G.H., Miller P. "Stock Selection Using Rule Induction", IEEE Expert, pages 52-58, October 1996.

[11]. A.N., Zapranis, A., Francis, G., "Stock Performance Modeling Using Neural Networks: A Comparative Study with Regression Models", Neural Networks, vol. 7, No. 2, pages 375-388, 1994.

[12]. Schoeneburg, E., "Stock Price Prediction Using Neural Networks: A Project Report", Neurocomputing, vol. 2, pages 17-27, 1990.

[13]. Yoon, Y., Guimaraes, T., Swales, G., "Integrating Artificial Neural Networks With Rule- Based Expert Systems", Decision Support Systems, vol. 11, 1994, pages 497-507, 1994.

[14]. Yoon Y., Swales G., "Predicting Stock Price Performance: A Neural Network Approach", Proceedings of the IEEE 24th Annual Conference on System Science, pages 156-162, 1991.

[15]. Choi J., Lee M., Rhee M., "Trading S&P 500 Stock Index Futures using a Neural Network", in the Proceedings of the 3rd Annual Conference on AI Applications on WS, pages 63-71, 1995.

[16]. Gately E. Neural Networks for Financial Forecasting, New York: Wiley, 1996.

[17]. Drossu R., Obradovic Z, "Rapid Design of Neural Networks for Time Series Prediction", IEEE Computational Science & Engineering, pages 78- 89, 1996.

[18]. Deju Zhang, Xiaomin Zhang, "Study on Forecasting the Stock Market Trend Based on Stochastic Analysis Method" in International Journal of Business and Management, Vol 4, No. 6, 2009.

[19]. M. R. Hassan and B. Nath, "Stock market Forecasting using Hidden Markov Models: a new Approach", Proceedings in 5th IEEE conference on Intelligent Systems Design and Applications, ISDA 2005.

[20]. Prasanta Kumar Sahoo, Pravalika Jeripothula, "Heart failure prediction using machine learning technique, Published in ELESVIER (SSRN) https://ssrn.com/abstract=3759562 ISSN: 1556-5068, January 2021.

[21]. Zhang, G.P.: Time Series Forecasting Using a Hybrid ARIMA and Neural Network Model. Neurocomputing 50, 159-175, January 2003.

[22]. A hybrid ARIMA and support vector machines model in stock price forecasting, omega 33(6): 497-505, December 2005.

VI. AUTHOR'S PROFILE

Dr. Prasanta Kumar Sahoo currently working as a Professor, Department of Computer Science and Engineering, School of Engineering, Anurag University, Hyderabad. His ORCID 0000-0002-5164 1010. He has completed his Ph.D. from Fakir Mohan University, Odisha in Computer Science Engineering. He has 22 years of teaching, research, and administrative experience. His Research Interest includes Cyber Security, Information Security, Machine Learning. He has published around 60 research papers in various reputed journals both at national and international levels. His research papers were cited both at the national and international levels, so far he had received 170 citations and his H index is 7. Many times Dr. Prasanta Kumar Sahoo won the best teacher award in various institutions for his contribution towards teaching and learning process. He is Certified Professional from BalaBit, completed Electronic Contextual Security Intelligence Examination Intermediate Level (ECSI). He has guided more than 50 projects both at UG and PG levels. He has delivered more than 15 guest lecturers. He has organized three national conferences and twelve faculty development program with immense success.

Debabrath Sahoo is pursuing Bachelor's degree in computer science and engineering at Lovely Professional university, Punjab.