International Journal of Current Trends in Engineering & Technology
www.ijctet.org, ISSN: 2395-3152
Volume: 06, Issue: 06 (November-December, 2020)

Enhanced Stock Market Analysis with Supervised Machine Learning Algorithm

Raju Bilwal, Rajeev Gupta
Department of Computer Science and Engineering, SISTEC, RGPV, Bhopal, India

Abstract— Supervised learning involves training a model on labelled data from a data source, often utilising LSTM models, which have demonstrated their effectiveness in text classification. These models are proficient in understanding sequential data patterns. Machine Learning algorithms use previously recorded data as training samples to predict future stock price trends. Supervised learning is a crucial task within the field of machine learning. It revolves around classifying and processing data through the application of machine learning techniques. The stock market is frequently employed in investment schemes that promise high returns but also carry inherent risks. Stock returns exhibit significant volatility, influenced by historical stock prices, prevailing market trends, financial news, and social media sentiments. Various methodologies like technical analysis, fundamental analysis, time series analysis, and statistical analysis are employed to forecast stock values. However, none of these methods has been conclusively established as reliable prediction tools. Many stockbrokers rely on technical, fundamental, and time series analyses when making stock predictions. Python is the predominant programming language employed for stock market prediction using machine learning. There are two main types of predicted stock analyses: first, a fundamental analysis of stock information, and second, a daily return analysis involving four prominent companies—Google, Apple, Microsoft, and Amazon. In the fundamental analysis of stock information, actual metrics like stock closing price, stock sales volume, and adjusted closing price are considered to predict the overall stock behaviour of the four companies, as mentioned earlier. This analysis encompasses stock purchases, sales, and daily return predictions. Daily return analysis is performed for each company, and these results are enumerated. This process provides comprehensive insights into the predicted stock behaviour of Google, Apple, Microsoft, and Amazon. These predictions aim to provide reliable stock buy and sell signals and accurate overall analyses. A proposed machine learning (ML) approach leverages available stock data from the mentioned companies (Google, Apple, Microsoft, Amazon) for training. This approach aims to acquire intelligence from the data and utilise the acquired knowledge for precise stock price predictions.

KEYWORDS: Machine learning, Supervised machine learning, finance, stock price prediction, sentiment analysis. Prediction, Classification, Optimisation, Stock Market, Data Analysis, LSTM.

I. Introduction

Stock investment has been steadily gaining popularity in recent times. The rise in the number of stock investors can be attributed to the proliferation of various stock prediction techniques available today. Most of these techniques heavily rely on time series data. The primary objective behind using these techniques is to maximise profits while minimising losses.

A stock market, or equity market, is a public platform that facilitates trading company stocks (shares) and derivatives at agreed-upon prices. These securities can be listed on a stock exchange or traded privately. Stocks are traded on stock exchanges, which are entities formed by corporations or mutual organisations specialising in bringing together buyers and sellers of these securities. In India, the bulk of trading occurs on two major stock exchanges: the Bombay Stock Exchange (BSE) and the National Stock Exchange (NSE). The BSE has been operational since 1875, while the NSE was established in 1992 and began trading in 1994. Despite their differing founding dates, both exchanges follow similar trading mechanisms, trading hours, and settlement processes. Financial market forecasting relies on specific principles, theories, and models to analyse financial markets and predict their future trends. Changes in stock prices are significantly influenced by human opinions and expectations regarding the future performance of stocks or shares [2]. A systematic approach to forecasting the stock market involves several operational steps: selection of input variables, Data preprocessing, Feature selection and extraction, training using prediction/classification models, and Evaluation of the proposed model's performance. The initial step in stock market forecasting entails choosing input features modelled using Computational Intelligence (CI) methods and determining the output to be predicted. Various fundamental and technical input variables are available for stock market forecasting. Selecting the appropriate input variables is a critical challenge, as deciding which ones to use is not straightforward. Following the input selection, the

subsequent step involves data preprocessing, enhancing the predictive capabilities of models [3,4]. Data preprocessing aims to improve the quality and usability of the data before it's fed into forecasting models.

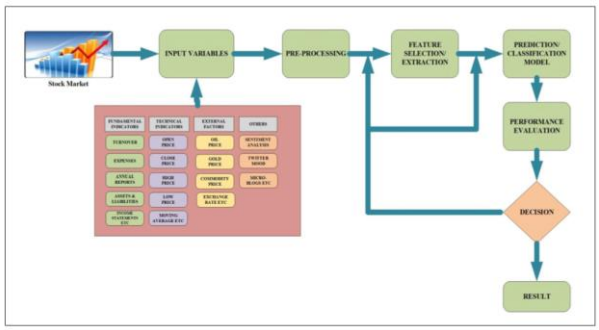

Figure 1 Generalised framework for stock market forecasting

The preprocessing mechanism is utilised for noise removal, outlier detection, handling missing values, and normalising data. In the third phase, various feature selection or extraction techniques are applied to identify the most representative variables from the input data. This helps to reduce the dimensions of the input data and subsequently lowers the computational complexity of the model. The subsequent phase involves selecting the appropriate Computational Intelligence (CI) methods for forecasting and then training the model using training data. The final phase of performance evaluation encompasses choosing suitable performance metrics, measuring the model's accuracy, and making informed decisions based on the results. In this article, each selected paper is discussed within a generalised framework for stock market forecasting [5].

II. Related Work

Many prediction models have been proposed to date for forecasting stock prices and stock market trends. Some machine-learning techniques have been discussed in this paper. Table 1 displays a summary of all the techniques proposed by various researchers. All the techniques have been classified into categories like classification techniques, regression techniques, ensemble algorithms, evolutionary techniques, deep learning, hybrid models, and others. In Shobana et al. [10], the complexity of analysing market trends depends on several external factors, some of which are not under one's control. This work analyses stock market trends using machine learning and nature-inspired techniques. Some of the techniques used in this paper are Decision Tree, PSO (Particle Swarm Optimisation), Black-Hole, and Naïve Bayes. After analysing the trends with standard techniques, we propose a new approach to analysing stock market indices. We calculate and compare accuracy across different techniques and algorithms. We outline the design of the proposed model along with its salient features and customisable parameters. Finally, we test our model on one year of the Nifty stock index dataset in real-time, analysing values based on data from the past three months.

In Preethi et al. [11], stock prices are important indicators for companies, and various factors can influence their values. Different events can impact public sentiments and emotions, affecting stock market trends. Due to the complexity of these factors, stock prices are dynamic, noisy, and nonlinear time series data. Machine learning has been applied to this area due to its capability to handle such problems. Learning-based methods for stock price prediction are popular, and strategies have been enhanced to improve their performance. However, successful stock market prediction remains a challenge. While news articles and social media data are useful, no effective method exists to incorporate social media for better financial market analysis. In Dang et al. [12], the web has become a carrier of vast information in modern society. Analysing and data mining this information allows emotional analysis to identify emotional tendencies and even evolutionary emotion rules. Like common search engines, information search technology relies on keywords and cannot support our needs, as emotions and opinions cannot be sorted by keywords. Emotion analysis involves altitude from the speaker, sentiment analysis, and classification of opinions as for or against. Text-based emotion analysis includes natural language processing, information retrieval, data mining, and artificial intelligence. In Sable et al. [13], accurate stock price prediction requires considering various factors influencing stock prices. Due to the superposition of multiple factors, nonlinear methods are commonly used for predicting fluctuating stock prices. Nonlinear regression analysis is frequently employed, but regression analysis is not always suitable due to the complexity of stock market factors. Stock price impact factors are complex, involving internal market and company-related actors. In Kunal Pahwa et al. [14], the stock market is complex, and this paper proposes using machine learning algorithms to predict future stock prices. The goal is to make this unpredictable business format more predictable by implementing machine learning algorithms and open-source libraries. The approach is based on assumptions and numerical outcomes. In Venkata Sasank et al. [15], sentiment analysis is applied to Twitter data to observe correlations between stock price changes and public opinions. Positive sentiments in social media can impact stock prices positively. The study shows a strong association between stock price fluctuations and public sentiments in tweets. In the finance world, stock trading plays a crucial role. Stock market prediction involves determining the future value of stocks using

various methods such as technical analysis, fundamental analysis, and time series analysis. This paper proposes a machine learning approach, specifically Support Vector Machine (SVM), for predicting stock prices across different market segments and frequencies [16]. In Shetty et al. [17], the relationship between public mood and stock market prices is explored. Chinese local microblogs and SVM are used to predict stock market trends, and the public mood is valuable for improving prediction accuracy. In Nirbhey et al. [18], an improved SVM version predicts US NASDAQ market trends. SVM performs better with fewer features, and regression experiments show its superiority over other algorithms. Hegazy et al. [19] present a unique use of SVM in sentiment analysis of Twitter data. The sentiment data is combined with historical data for stock price prediction, achieving an accuracy of around 76%. In Chouhan et al. [20] focus e.

III. Software Introduction (Python)

Python is a well-designed programming language that finds extensive use in real-world applications. As a highly dynamic, object-oriented, and general-purpose language, Python operates through an interpreter, rendering it suitable for a wide array of domains. The language's primary aim was to offer simplicity and comprehensibility, making it hailed as beginner-friendly and user-friendly in recent times. Python's popularity has surged as a language suitable for newcomers, replacing Java as the preferred choice for introductory programming. Python's flexibility is a hallmark attribute due to its dynamic typing. This attribute grants considerable adaptability and resilience in handling errors. A Python program can still be compiled and executed even in problematic sections. Python not only supports diverse programming styles, such as structural and object-oriented, but it also permits the integration of modular components from other programming languages. For instance, you can create a module in a language like C++ and seamlessly import and enhance it within a Python program – even designing graphical user interfaces for added functionality. Project Jupyter, initiated in 2014 by Fernando Pérez and Brian Granger, is dedicated to advancing open-source software, open standards, and interactive computing services across multiple programming languages. Derived from IPython, Project Jupyter embraces its name as a homage to Galileo's discovery of Jupiter's moons. This reference aligns with its support for the three core programming languages: Julia, Python, and R. The name and logo also pay tribute to Galileo's documented findings in notebooks. Project Jupyter oversees the development and maintenance of interactive computing tools like Jupyter Notebook, JupyterHub, and JupyterLab. The initiative's financial support is provided by NumFOCUS, further solidifying its presence in the programming community.

IV. Results and Discussion

Google, Apple, Microsoft, and Amazon Stock Market Price Prediction: Normalised data matches the input into the predictive model. Understanding the necessity of predicting stock price movements is crucial. Based on current data, the goal is to predict and visualise future stock market trends. This involves preprocessing, training, and eventually predicting stock prices. The analysis primarily focuses on two predictions: firstly, a basic analysis of stock information, and secondly, a daily return analysis involving the stocks of four major companies—Google, Apple, Microsoft, and Amazon.

(a)Basic Analysis of stock information

(i) Actual stock closing price

Basic Analysis of Stock Information and Comparative Analysis of Actual Stock Closing Prices: This section focuses on market data analysis, particularly on four major company stocks: Google, Apple, Microsoft, and Amazon. The analysis is based on preprocessing and training the data. In this section, I will explain how to retrieve stock information using pandas and how to analyse the fundamental attributes of a stock. The term' closing price' pertains to the share cost at the end of the trading day, as depicted in Figure 2.

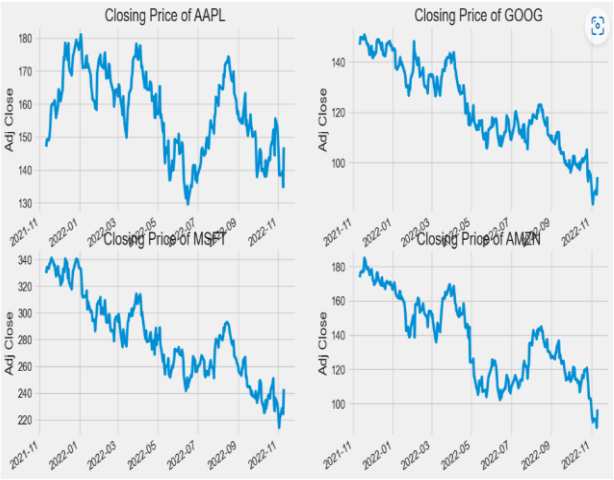

Figure 2: Comparative Analysis of Actual Stock Closing Prices

(ii) Actual stock sales volume

Basic Analysis of Stock Information and Comparative Analysis of Actual Stock Sales Volume: This section focuses on market data analysis, specifically involving four major company stocks: Google, Apple, Microsoft, and Amazon. The analysis is based on preprocessing and training the data. In this section, we explained how to retrieve stock information using pandas and how to

analyse the fundamental attributes of a stock. Figure 3 below illustrates this process.

(iii) Actual stock-adjusted closing price

While the closing price reflects the cost of shares at the end of the trading day, the adjusted closing price considers dividends, stock splits, and new stock offerings. This adjusted closing price is a more accurate indicator of stock value as it factors in these crucial elements, providing a starting point beyond where the closing price ends. This section focuses on fundamental stock analysis and a comparative assessment of adjusted closing prices for four prominent companies: Google, Apple, Microsoft, and Amazon. The analytical process encompasses initial market data preprocessing followed by subsequent training. Within this section, we delve into retrieving stock information using pandas. Additionally, we demonstrate fundamental stock attribute analysis, as depicted in Figure 4.

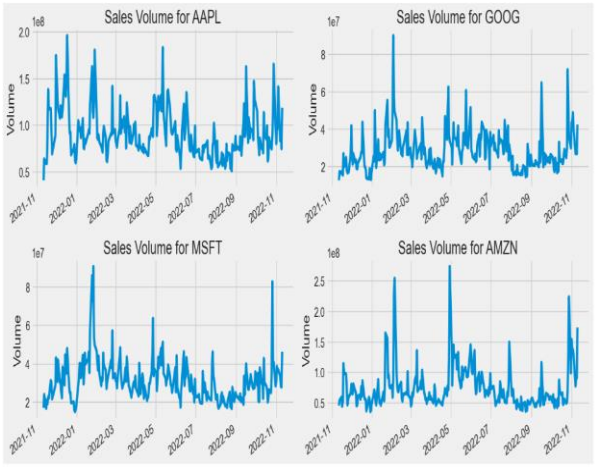

Figure 3: Comparative Analysis of Actual Stock Sales Volume

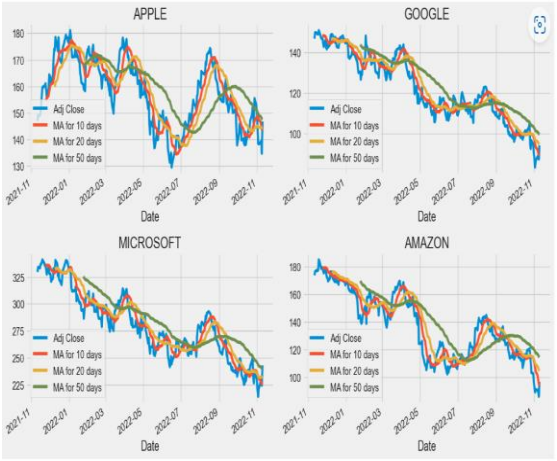

Figure 4: Comparative Analysis of Actual Adjusted Closing Prices

b) Daily Return Analysis

With our baseline analysis complete, let's delve deeper into the data. Our focus now shifts to analysing the stock's risk. A closer examination of the stock's daily fluctuations is necessary to achieve this, moving beyond its absolute value. Utilising pandas, we'll extract the daily returns for analysis across four prominent companies: Google, Apple, Microsoft, and Amazon.

(i) Company Daily Return Analysis

Excellent. With the groundwork laid, it's time for an overarching view of the average daily returns for the companies mentioned above—Google, Apple, Microsoft, and Amazon. Refer to Figure 5 below for visualisation.

Figure 5: Company Daily Return Analysis

(ii) Company in enumerate daily return analysis

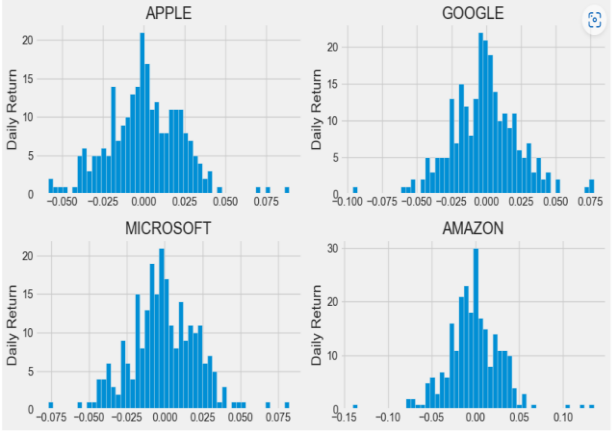

Figure 6: Company-wise Enumerated Daily Return Analysis

As anticipated from our pair plot analysis, both numerically and visually, we observe strong correlations among the daily stock returns of four prominent companies: Google, Apple, Microsoft, and Amazon. Notably, the enumerated daily stock return analysis reveals the most robust correlation. It's worth noting the intriguing observation that all the

technology companies exhibit positive correlations. Risk quantification takes various forms, but a fundamental approach is leveraging the daily percentage returns data we've collected. This involves comparing the anticipated return with the standard deviation of the daily returns, which represents risk. The histogram shows that these stocks exhibit lower risk and positive expected returns. Refer to Figure 6 above for visualisation. Our analysis encompasses two key predictive stock evaluation aspects. Firstly, the basic analysis of stock information, encompassing actual stock closing price, actual stock sales volume, and actual stock adjusted closing price, contributes to the overall prediction for stock activities in the four companies: Google, Apple, Microsoft, and Amazon. Secondly, based on the same set of companies, the daily return analysis augments our predictive capabilities. These analyses collectively provide reliable insights into stock purchasing, sales trends, and daily returns. In particular, the company-wise daily return analysis and the enumerated daily return analysis bolster the accuracy of our predictions. This accuracy is validated by an optimisation rate of approximately 98% in our training model for predicting the stock market prices of Google, Apple, Microsoft, and Amazon.

V. Conclusion

In conclusion, this paper highlights the significance of supervised learning in machine learning by emphasising the process of training datasets and aligning them with accurate classifications. The utilisation of LSTM models stands out as a powerful technique for text classification. The paper underscores the potency of machine learning across diverse applications, underscoring the pivotal role of data in this domain. The complexity of data analysis contrasts with the fundamental truth of data's immense value.

The study provides insights into fundamental algorithms employed in supervised machine learning, such as Random Forest, k-Nearest Neighbors, Support Vector Machines, and Multiple Linear Regression. The conclusion drawn is that these algorithms yield optimised outcomes in various scenarios.

Additionally, the paper conducts a comparative analysis of contemporary supervised machine learning techniques and envisions future trends. It focuses on stock investments, acknowledging the complexity of decision-making in this field due to multifaceted variables. The potential for enhanced predictive accuracy in stock market trajectory prediction is emphasised.

The paper's core involves predictive analyses based on fundamental stock information and daily return trends for prominent companies. Both analyses contribute to reliable predictions of stock activities, sales trends, and daily returns.

Overall, this paper offers a comprehensive understanding of supervised machine learning, shedding light on algorithmic intricacies and fostering optimised outcomes. It encapsulates present methodologies while anticipating future trends in the evolving landscape of supervised machine learning.

REFERENCES

[1]. Soni, D., Agarwal, S., Agarwal, T., Arora, P., & Gupta, K. (2018). Optimised Prediction Model for Stock Market Trend Analysis. 2018 Eleventh International Conference on Contemporary Computing (IC3).

[2]. Tang, X., Yang, C., & Zhou, J. (2009). Stock Price Forecasting by Combining News Mining and Time Series Analysis. 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology.

[3]. Lauren, S., & Harlili, S. D. (2014). Stock Trend Prediction using Simple Moving Averages Supported by News Classification. 2014 International Conference of Advanced Informatics: Concept, Theory and Application (ICAICTA).

[4]. Soni, D., Agarwal, S., Agarwal, T., Arora, P., & Gupta, K. (2018). Optimised Prediction Model for Stock Market Trend Analysis. 2018 Eleventh International Conference on Contemporary Computing (IC3). doi:10.1109/ic3.2018.8530457.

[5]. Wang, Z., Ho, S.-B., & Lin, Z. (2018). Stock Market Prediction Analysis by Incorporating Social and News Opinion and Sentiment. 2018 IEEE International Conference on Data Mining Workshops (ICDMW). doi:10.1109/icdmw.2018.00195.

[6]. Kaihui Zhang, Lei Li, Peng Li, & Wenda Teng. (2011). Stock Trend Forecasting Method based on Sentiment Analysis and System Similarity Model. Proceedings of 2011 6th International Forum on Strategic Technology. doi:10.1109/ifost.2011.6021163.

[7]. Xing, T., Sun, Y., Wang, Q., & Yu, G. (2013). The Analysis and Prediction of Stock Price. 2013 IEEE International Conference on Granular Computing (GrC). doi:10.1109/grc.2013.6740438.

[8]. Parmar, I., Agarwal, N., Saxena, S., Arora, R., Gupta, S., Dhiman, H., & Chouhan, L. (2018). Stock Market Prediction Using Machine Learning. 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC). doi:10.1109/icsccc.2018.8703332.

[9]. Cheng, C.-H., Su, C.-H., Chen, T.-L., & Chiang, H.-H. (2010). Forecasting Stock Market Based on

Price Trend and Variation Pattern. Lecture Notes in Computer Science, 455–464. doi:10.1007/978-3-642-12145-6_47.

[10]. Shobana, T., & Umamakeswari, Arumugam. (2016). A Review on Prediction of Stock Market using Various Methods in the Field of Data Mining. Indian Journal of Science and Technology, 9.

[11]. Preethi, G., & Santhi, B.. (2012). Stock Market Forecasting Techniques: A Survey. Journal of Theoretical and Applied Information Technology, 46.

[12]. Dang, L. Minh, & Duong, Duc. (2016). Improvement methods for Stock Market Prediction using Financial News Articles.

[13]. Sable, Rachna, Goel, Shivani, & Chatterjee, Pradeep. (2019). Empirical Study on Stock Market Prediction Using Machine Learning.

[14]. Kunal Pahwa, Neha Agarwal. (2019). Stock Market Analysis using Supervised Machine Learning. 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (Com-IT-Con), India.

[15]. Venkata Sasank Pagolu, Kamal Nayan Reddy Challa, Ganapati Panda. (2016). Sentiment Analysis of Twitter Data for Predicting Stock Market Movements.

[16]. Reddy, V. K. S. (2018). Stock Market Prediction using Machine Learning. International Research Journal of Engineering and Technology (IRJET), 5(10), 1033-1035.

[17]. Shetty, Nisha, & Pathak, Ashish. (2017). Indian Stock Market Prediction Using Machine Learning and Sentiment Analysis.

[18]. Nirbhey, Khalfay, Neeha, Soni, Vidhi, Vora, Deepali. (2017). Stock Prediction using Machine Learning: A Review Paper.

[19]. Hegazy, Osman, Soliman, Omar S., Abdul Salam, Mustafa. (2013). A Machine Learning Model for Stock Market Prediction. International Journal of Computer Science and Telecommunications, 4.

[20]. Chouhan, Lokesh, Agarwal, Navanshu, Parmar, Ishita, Saxena, Sheirsh, Arora, Ridam, Gupta, Shikhin, Dhiman, Himanshu. (2018). Stock Market Prediction Using Machine Learning.