Volume: 28, Issue: 1
Page: 219-228
2023

International Journal of Science and Business

Journal homepage: ijsab.com/ijsb

The Impact of Big Data Analytics on Stock Price Prediction in the Bangladesh Stock Market: A Machine Learning Approach

Md Zahidul Islam, Md Maruful Hoque Chowdhury, & Mohammad Momen Sarker

Abstract

The stock market is a volatile and complex environment impacted by various unpredictable factors, making accurate stock price prediction challenging. This research paper explored the potential and capability of big data analytics and machine learning techniques in terms of enhancing stock price prediction accuracy in the setting of the Bangladesh stock market. The methodology adopted in the study entailed a data gathering process, which comprised collecting financial data from the Bangladesh stock market, such as news articles, financial statements, macroeconomic indicators, and historical stock prices. Based on a literature review, various fundamental and technical indicators are chosen as predictive features. The research paper employed a combined methodology that consolidates technical calculations and sentimental analysis to predict and forecast stock market patterns. By adopting machine learning and sentiment analysis techniques, this technique provides future predictions for the stock market while considering the impact of political events, economic factors, and dynamics in social media. The consolidation of big data analytics enables real-time predictions of stock market movements. The sentiment analysis algorithm facilitates prompt and extensive evaluations of tweets and news articles. As a result, the integration of technical and sentiment analyses greatly enhances the accuracy of stock market predictions.

IJSB
Accepted 26 October 2023
Published 29 October 2023
DOI: 10.58970/IJSB.2216

ISSN: 2520-4750 (Online) 2521-3040 (Print)

Keywords: *Big data, Machine Learning, BD Stock market, Stock prediction, Sentiment Analysis.*

About Author (s)

Md Zahidul Islam, Gannon University, PA, USA. BBA- Accounting & Information Systems (AIS), United International University, MBA- Accounting & Information Systems (AIS), University of Dhaka, Dhaka, Bangladesh.

Md Maruful Hoque Chowdhury (Corresponding Author), Business Development Manager, UNICAP Securities Limited, Dhaka, Bangladesh. MBA-Northern University Bangladesh. BBS-Accounting, Government College of Commerce, Chattogram. Under National University, Bangladesh.

Mohammad Momen Sarker, Executive Commercial, Bengal Group of Industries, Dhaka, Bangladesh. MBA. BSS-Accounting, University of Chittagong, Bangladesh.

1.0 Introduction

According to Islam et al (2023), predicting future stock prices precisely and accurately has always been problematic due to the inherent complexity and uncertainty involved in stock market movement which highly depends on many unpredictable political, financial, economic, and psychological factors. Traditionally, approaches like technical analysis comprising chart patterns and pattern predictions or fundamental evaluation using macroeconomic variables and organizational financials have been employed for stock price forecasting. Nevertheless, according to Muhammad (2022), the reliability and accuracy of such predictions are limited due to their simplistic assumptions and inability to capture all relevant factors impacting stock prices. This paper aims to examine the possibility of machine learning and big data techniques for enhancing stock price prediction accuracy in the Bangladesh stock market. The research paper applies distinct machine learning algorithms on big historical stock prices and financial datasets and compares and contrasts their performance for short-term price forecasting.

The prime objective of this study is to explore the implication of big data analytics on stock price prediction in the Bangladesh stock market. The research paper targets to achieve the following specific objectives: (1) To employ machine learning algorithms for stock price prediction in the Bangladesh Stock market. (2) To examine the implications of big data analytics on stock price prediction accuracy. (3) To identify the opportunities, challenges, and future directions of applying big data analytics in stock price prediction. In that respect this research paper targets to address the following research questions: (a)How can big data analytics be employed for stock price prediction in the Bangladesh stock market? (b)What are the most efficient machine learning methods for predicting stock prices? (c)What is the implication of big data analytics regarding stock price prediction accuracy? (d)What are the opportunities and challenges of applying big data analytics in stock price prediction?

1.1 Background

According to Vigila (2018), the stock market plays a pivotal role in any economy, enabling organizations to raise capital and investors to engage in wealth creation. As such, accurate stock price prediction is of significant interest to investors, policymakers, and financial analysts, as it assists in terms of managing risks, making informed investment decisions, and maximizing returns. Traditionally, stock price prediction has depended on fundamentals, market sentiment, and technical analyses. Nevertheless, with the dawn of big data analytics and machine learning, there is an escalating interest in examining these techniques' capabilities in predicting stock prices more accurately. In the recent past, the availability of a large volume of structured and unstructured stock market data in conjunction with advancement in machine learning algorithms has opened up new possibilities for enhanced stock price prediction. Big data analytics comprises collecting, processing, and evaluating large datasets constituting multiple variables ranging from historical stock prices, trading volumes, and market sentiments to organizational financials, economic indicators, and news articles for obtaining useful insights (Vigila, 2018). When consolidated with powerful machine learning techniques, big data holds promise for establishing more robust predictive models that can pinpoint complex patterns and associations in the data to forecast stock prices with higher accuracy. While the majority of studies have examined the effects of machine learning and big data analytics on stock markets in developed economies, uprising markets like Bangladesh remain relatively unexplored. The Bangladesh stock market, which is famously known as *the Dhaka Stock Exchange (DSE)*, was established in 1954 and has grown substantially over the years (Muhammad, 2022). Nonetheless, accurate stock price prediction still proves to be a challenge for investors in this market because of its limited availability and unique dynamics of high-quality historical data (Vigila, 2018).

This research is significant for a myriad of reasons. First, it undoubtedly adds value to the present literature by examining the application of big data analytics in stock price prediction in the setting of the Bangladesh stock market. Second, it offers insights into the efficiency of machine learning algorithms for stock price prediction, which can assist financial analysts and investors in making more informed decisions (Mahtab et al., 2022). Third, the study pinpoints the opportunities and challenges related to the use of big data analytics in stock price prediction, informing researchers and policymakers about the potential benefits and limitations of these techniques.

2.0 Literature Review

2.1 Understanding Big Data Analytics

As per Mahtab et al. (2022), big data comprises various types of data sources that can be classified differently based on their structure. Textual data may come in a structured format appropriate for databases, an unstructured format such as tweets, or a semi-structured format like XML or JSON files. Multimedia content contributes highly to big data volumes, images, audio, and video creating huge repositories of digitized content daily. According to recent approximation, over 2.5 quintillion digital content is developed internationally every day from sources as diverse as social media posts to sensor readings. Nevertheless, within these huge troves of accumulating data also lies opportunities for new insights. By utilizing big data analytics methodologies, useful data can be extracted and trends pinpointed from the spontaneous recordings of transactions and human activities now continuously streamed online and via omnipresent IoT devices. While irrelevant and noisy content ought to be filtered out, evaluating linguistic, conceptual, and patterns hidden in the flood of daily digital activities enables evidence-based predictions valuable across various applications (Mahtab et al., 2022). One promising domain is network optimization, as understanding utilization habits and surfing behaviors at scale can guide network infrastructure planning and reinforce the quality of experience.

2.2 Big Data Analytics in Finance

According to Hasan (2022), big data analytics has revolutionized various sectors, such as finance. In the financial sphere, big data analytics denotes the process of gathering, evaluating, and interpreting big volumes of unstructured and structured data to extract crucial insights and make data-driven decisions. The adoption of big data analytics in finance has facilitated financial institutions to enhance risk management, fraud detection, client segmentation, and investment decision-making. The availability of a large volume of financial data, including historical stock prices, social media sentiment, news articles, and macroeconomic indicators, has opened up new opportunities for accurately predicting stock prices (Hasan, 2022).

2.3 Stock Price Prediction

Stock price prediction is a perplexing task due to financial markets' sophisticated and dynamic nature. Traditional techniques for stock price prediction comprise fundamental analysis, which entails evaluating organizational financial statements, sector trends, and economic components to approximate its technical analysis and intrinsic value, which depend on historical price trends and trading volumes to pinpoint patterns and make trading decisions (Kadam, 2022). Nevertheless, these techniques have challenges in terms of capturing the intricate associations and patterns present in financial data. In this research paper, the researcher will harness the capability and power of big data to forecast and predict trends in the stock market. The stock market has gradually become a pillar of global commerce, attracting millions of participants globally, Participating in diverse trading and investment activities. The art and science of prediction hold pivotal significance within the stock market,

as precision in forecasting can lead to substantial profits. Moreover, the stock market serves as a barometer for a country's economic health, intricately associated with the global economic landscape. The dynamic nature of the stock market is unmistakable, with data changing by the second. The advent of social media platforms has introduced an additional layer of complexity. Numerous websites facilitate discussions and opinions about various companies, constituting a rich source of sentiment data. However, it's essential to acknowledge that these opinions are not always impartial; bias can seep in, potentially compromising the accuracy of predictions (Kadam, 2022).

2.4 Machine Learning Techniques in Stock Price Prediction

As per Hasan (2022), machine learning approaches have obtained popularity in stock price prediction because of their capability to analyze and evaluate huge volumes of data, pinpoint complex patterns, and make accurate predictions and forecasting. Supervised learning algorithms e.g. decision trees, random forests, linear regression, support vector machines (SVM), and neural networks, have been widely adopted in terms of predicting stock prices (Lyhyaoui, 2022). These algorithms learn from historical data to capture the association between various attributes and the target variable (stock prices) and then make predictions on new, unseen data.

2.5 Big Data Analytics and Stock Price Prediction: The Bangladesh Context

In the recent past, the Bangladesh stock market has encountered significant growth, capturing international and local investors. With the escalating availability of financial advancement and data in big data analytics and machine learning, there is an escalating interest in terms of applying these approaches to predict stock prices accurately and precisely in the Bangladesh context (Muhammad, 2022). Nonetheless, the employment of big data analytics in stock price prediction in Bangladesh is relatively promising, and there is a need for empirical research to examine its impact, challenges, and opportunities. According to Muhammad (2022), big data analytics can significantly influence stock price prediction and forecasting in the Bangladesh stock market. By using large volumes of financial data and employing machine learning algorithms, accurate predictions can be attained, leading to enhanced risk management and investment decisions. Nonetheless, challenges such as model interpretability, data quality, and overfitting need to be addressed. Examining opportunities such as alternative data sources, enhanced feature engineering, ensemble learning techniques, and ethical considerations can further reinforce the application of big data analytics in stock price prediction and forecasting in Bangladesh. These advancements can motivate investors, policymakers, and financial analysts, to make more informed decisions and lead to the growth and stability of the Bangladesh stock market.

3.0 Methodology

3.1 Data Collection

The first stage in the research methodology comprised gathering relevant financial data from the Bangladesh stock market. This comprised financial statements of organizations, news articles, macroeconomic indicators, historical stock prices, and social media data. The data can be obtained from various sources, such as stock exchanges, financial databases, and online platforms. For this study, daily stock market and organizational financial data were collected for 10 top organizations listed on the DSE from January 2022 to December 2022. The organizations were chosen based on their market capitalization ranking. Furthermore, historically adjusted closing stock prices, the following 11 fundamental and technical indicators were selected as predictive features based on the literature review:

- Trading volume

- Relative strength index (RSI)
- Moving average (5-day, 10-day, 20-day)
- Earnings per share (EPS)
- Moving average convergence divergence (MACD)
- Return on equity (ROE)
- Price to earnings (P/E) ratio
- Current Ration
- Net profit margin
- Dividend Yield
- Debt to equity ratio

3.2 Data Preprocessing

After gathering the data, preprocessing approaches were applied to transform and clean raw data into the appropriate format for analysis. This comprised handling missing data, outlier identification and treatment, feature scaling, and data normalization. Moreover, text data from social media and news articles was processed using natural language processing techniques to extract sentiment and relevant features. This resulted in a multivariate time series dataset comprising 12 variables entailing the stock price for each organization. The data was preprocessed to manage missing values, and outliers and normalize attribute scales before feeding into machine learning models. Then 70% of the data was adopted for training and the remaining 30% was held out as a test set for evaluation.

3.3 Feature Selection and Engineering

As per Divyavalli (2023), feature selection is a pivotal stage in stock price prediction as it assists in detecting the majority of relevant features that lead to the prediction task. Diverse feature selection approaches, such as correlation analysis, and recursive feature elimination, were employed to choose the optimal subset of features. Feature engineering entails creating new features from the current ones to capture additional information that may improve prediction accuracy.

3.4 Machine Learning Techniques

Different machine learning algorithms were employed to predict and Forecast stock prices in the Bangladesh stock market. These comprised decision trees, SVM, linear regression, and random forests, as well as deep learning algorithms such as recurrent neural networks (RNN) and long short-term memory (LSTM) networks. The algorithms were trained on historical data and then used to make predictions on new, unseen data (Mahtab et al., 2022). Support Vector Regression (SVR) is a powerful nonlinear regression approach that maps input characteristics into a high dimensional space and determines the linear association to make predictions (Mahtab et al., 2022). Random Forest develops an ensemble of decision trees trained on randomly chosen subsets of features and data. This minimizes overfitting and enhances generalizability. LSTM is a form of recurrent neural network particularly appropriate for sequence prediction tasks like stock prices. It resolves the long-term dependency matter of standard RNNs. MLP is a feedforward neural network with several hidden layers that can learn highly sophisticated nonlinear associations in large datasets for prediction (Mahtab et al., 2022).

3.5 Proposed Architecture

The proposed architecture consolidates the power of big data and Recurrent Neural Networks (RNN), to attain precise stock value predictions. Within the domain of artificial neural networks, data analysis depends on historical data, facilitating automatic value predictions.

The neural network, having been programmed comprehensively, acts as an expert in its specific field, yielding highly accurate results (Rouf et al., 2021). Nonetheless, it's vital to recognize that although technical calculations are proficient, they may not always generate precise stock values. These inconsistencies emerge because of the multifaceted nature of the stock market, where values are impacted by factors such as quarterly or half-yearly results, international and national transactions, collaborations with other companies, anti-dumping duties on products, and broader economic conditions, among others.

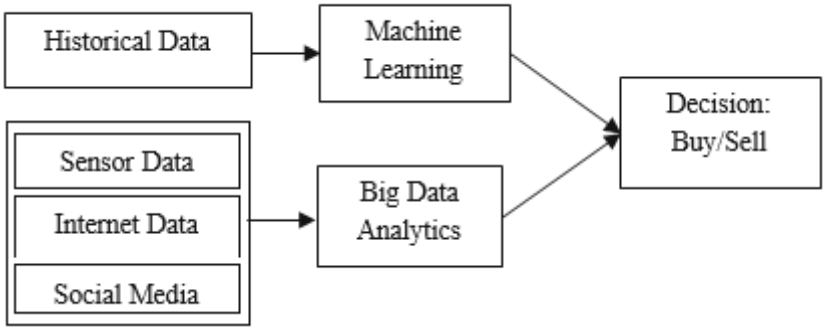

Figure 1: Showcases the architecture for predicting the stock market.

The prediction of future stock market values in this technique is achieved via the amalgamation of sentiment and technical analysis. Equation 1 was adopted to compute the technical value, which, in turn, acts as the basis for forecasting future stock market patterns.

Equation 1 can be represented as follows:

$$A = ((\sum_{i=1}^{n} Wi - W(i-1))/(n-1))^{1/2}$$

Where:

- A denotes the calculated technical value.
- Wi represents the closing value on the nth day.
- n represents the number of days considered for the prediction.

Wi denotes the closing value on the nth day, while n represents the number of days set for prediction. The calculation comprises determining the variation between closing values for two subsequent days, replicating this process for up to n days, and eventually deriving the average, which acts as the closing value for the concluding prediction as showcased in Equation 2.

$$B = \sum_{i=0}^{n} Vi/n$$

Where Vi represents the volume of the ith day the value is computed for n number of days and the average volume for n days is in Equation 3

$$C = (\sum_{i=0}^{n} Vi/n)/U$$

U represents the average volume of the month and the last prediction value is calculated by the Equation 4

$$D = ((\sum_{i=1}^{n} Wi - W(i-1))/(n-1))^{1/2} * ((\sum_{i=0}^{n} Vi/n)/U)$$

In this scenario, A denotes the average technical analytic value across the course of n trading days, while B represents the mean trading volume for n subsequent days. C, on the other hand, represents the ratio of the mean volume over n days to the average volume witnessed in a typical month. The parameter D acts as a vital indicator in this analysis. A positive value for D indicates prevailing buying pressure, which normally implies an expected increase in the share value shortly. By contrast, a negative D value represents a selling pressure scenario. When D approaches zero, it denotes that the investor may consider holding onto the shares.

Steps in Big Data Analysis

Step 1: Data Collection: This stage will entail collecting data from different sources, including social media, and the internet. In the context of the stock market, numerous websites and channels provide valuable insights. This data is subsequently streamed into the Hadoop Distributed File System (HDFS). Consequently, the system retrieves organizational data to obtain company-associated data from online sources, concentrating on recent activities, and news articles, and the Mozenda web crawler is adopted for gathering news articles. The information is subsequently streamed into HDFS. The system will consequently perform sentiment analysis to determine the positive and negative sentiments associated with each company. Data is sourced from platforms such as NSE (National Stock Exchange), Financial Express, Economic Times, and Money Control websites. Additionally, expert opinions and insights are gathered from these sources.

Step 2: Data Analysis: This phase will apply Hadoop for Data Analysis for data analysis. This phase primarily focuses on the examination of news articles and the Stock Exchange department gathered from various sources. Before undertaking the analysis, preprocessing was imposed on the collected information. This entails tasks such as eliminating stop words, URLs, and duplicate entries from the dataset. In the setting of sentiment analysis, the processed data is evaluated by utilizing HDFS. For sentiment extraction from news and the internet, Hive is applied as a valuable tool. In this phase, data analysis is executed, affirming that the data is well-prepared and free from redundancy or noise before applying analytical techniques.

Algorithm: Sentiment analysis
Input: Data from distinct sources and avenues in the internet with Keyword.
Output: All the phrases with the keyword.
Start: Sentiment[R] = 0
For rows 1 to n
Contrast words in the dictionary for all rows R and adopt Sentiment Word.
Sentiment[R] = Sentiment[R] + 1

The news articles and stock exchange data were consolidated to offer an extensive view of sentiment regarding the organization. This sentiment analysis discerns whether the impact on the company is positive or negative. The resultant findings are visualized in graph form using a combination of R and Hadoop, offering a clear representation of the sentiment trends.

Step 3: Result – Present valuable and summarized outcomes to the user.

3.6 Performance Analysis

Performance metrics are calculated for the technical analysis, and subsequently integrated with sentiment analysis to pinpoint and rectify any false negatives and false positives. This holistic dimension reinforces stock forecasting accuracy, facilitating predictions that account for external factors such as international economics. In technical analysis, future values are

forecasted and predicted based on historical data, comprising volume and closing values. The prediction outcomes of the technical analysis are illustrated in Figure 2

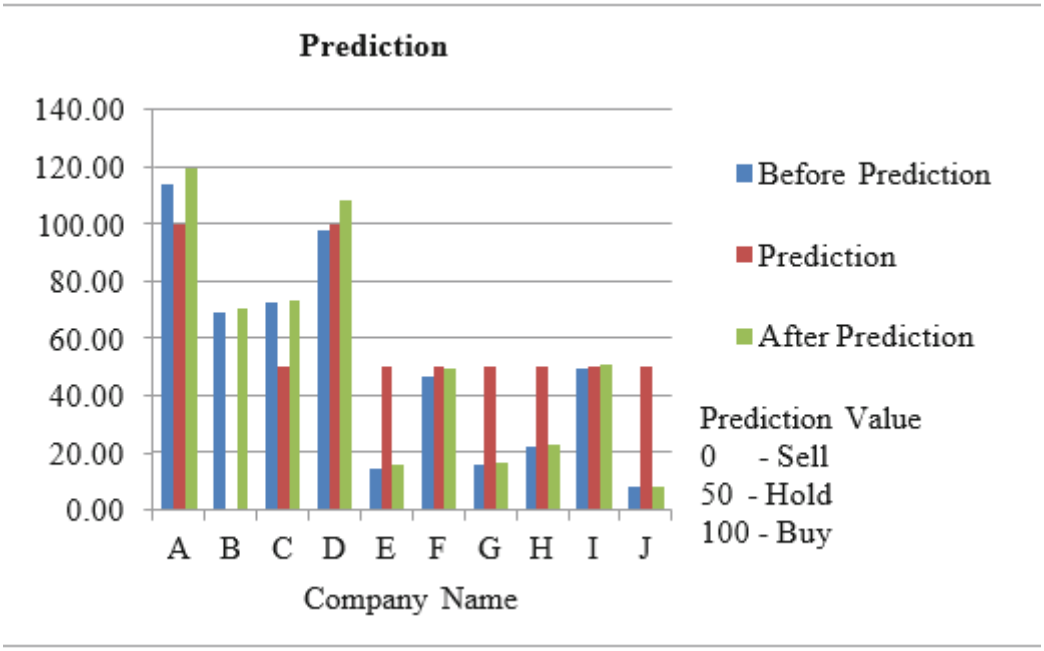

Figure 2 Showcases stock market prediction using technical analysis

The precision-recall metric is applied to evaluate the accuracy of the predicted results. Precision denotes the average probability value within the associated category, while recall represents the average probability value within the whole dataset. The precise definitions of recall and precision can be found in Equations 1 and 2 as showcased below.

$$\text{Precision} = \frac{tp}{tp + fp}$$

$$\text{Recall} = \frac{tp}{tp + fn}$$

Accuracy was computed by equation 3 as displayed below:

$$\text{Accuracy} = \frac{tp + tn}{tp + tn + fp + fn}$$

In this scenario, various parameters are employed to measure the accuracy of the predicted values, comprising true positives (tp), true negatives (TN), false positives (fp), and false negatives (fn). These parameters jointly contribute to evaluating the recall, precision, and overall accuracy of the technique. Particularly, this technique accomplishes a precision rate of 87%, a recall rate of 89%, and an accuracy rate of 89%. Maximizing the capabilities of big data analytics, sentiment analysis is undertaken, which entails evaluating stock predictions grounded on data from news channels, Stock market data, and internet sources. This analysis differentiates between negative and positive sentiments, ultimately facilitating more accurate stock predictions. The following figure showcases the prediction of the organizations.

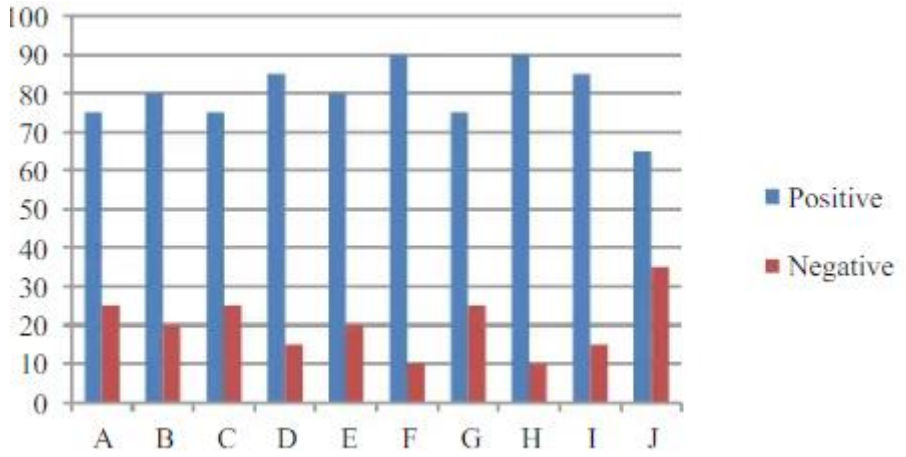

Figure 3: Displays stock market prediction using sentiment analysis.

By referring to the figure above, it is evident that by combining the sentiment analysis and technical the prediction result and outcome will be more accurate and assist the investor to make pro*fi*t in the stock market.

4.0 Conclusion

In summation, this study aimed to examine the possibility of machine learning and big data techniques for enhancing stock price prediction accuracy in the Bangladesh stock market. The research paper applies distinct machine learning algorithms on big historical stock prices and financial datasets and compares their performance for short-term price forecasting. The research paper employed a dual approach, consolidating sentiment analysis and technical calculations, to forecast stock market patterns. Employing machine learning and sentiment analysis, this approach offers future predictions for the stock market, considering the impact of economic factors, political events, and social media dynamics. The consolidation of big data analytics enables real-time stock market predictions. The sentiment analysis algorithm presents immediate, comprehensive evaluations of tweets and news articles. Therefore, the amalgamation of technical and sentiment analyses significantly enhances the accuracy of stock market predictions.

References

Divyavalli, K. (2023). *A Review on Long Short Term Memory based Stock Price Prediction Techniques*. IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/abstract/document/10128285

Hasan, M. R. (2022). Cybercrime Techniques in Online Banking. Int. J. of Aquatic Science, 13(1), 524-541. Retrieved from: https://www.journal-aquaticscience.com/article_158883.html (January 2022)

Kadam, Y. (2022). A Survey on Stock Market Price Prediction System using Machine Learning Techniques. Retrieved from: https://www.academia.edu/75020017/A_Survey_on_Stock_Market_Price_Prediction_System_using_Machine_Learning_Techniques

Kurani, A., Doshi, P., Vakharia, A., & Shah, M. (2023). A comprehensive comparative study of Artificial Neural Network (ANN) and Support Vector Machines (SVM) on stock forecasting. *Annals of Data Science*, *10*(1), 183–208. https://doi.org/10.1007/s40745-021-00344-x

Lyhyaoui, A. (2022). Stock price prediction based on SVM : The impact of the stock market indices on the model performance. Retrieved from:

https://www.academia.edu/78956809/Stock_price_prediction_based_on_SVM_The_impact_of_the_stock_market_indices_on_the_model_performance

Muhammad, T. (2022). Transformer-Based Deep Learning Model for Stock Price Prediction: A Case Study on Bangladesh Stock Market. Retrieved from: https://www.academia.edu/93246656/Transformer_Based_Deep_Learning_Model_for_Stock_Price_Prediction_A_Case_Study_on_Bangladesh_Stock_Market

Mahtab, Md. T., Zaman, A. G. M., Mahin, M. R., Mia, M. N., & Islam, Md. T. (n.d.). Stock Price Prediction: An Incremental Learning Approach Model of Multiple Linear Regression. Retrieved from: https://ajse.aiub.edu/index.php/ajse/article/view/490

NAHIL, A., & Academy, I. (2018). Stock price prediction based on SVM : The impact of the stock market indices on the model performance. Retrieved from: https://www.academia.edu/37201679/Stock_price_prediction_based_on_SVM_The_impact_of_the_stock_market_indices_on_the_model_performance

Rouf, N., Malik, M. B., Arif, T., Sharma, S., Singh, S., Aich, S., & Kim, H.-C. (2021). Stock Market Prediction Using Machine Learning Techniques: A Decade Survey on Methodologies, Recent Developments, and Future Directions. Retrieved from: https://www.mdpi.com/2079-9292/10/21/2717

Sheth, D., & Shah, M. (2023). Predicting stock market using machine learning: best and accurate way to know future stock prices. International Journal of Systems Assurance Engineering and Management, 14(1), 1–18. https://doi.org/10.1007/s13198-022-01811-1

Vigila, M., (2018). Forecasting of Stock Market by Combining Machine Learning and Big Data Analytics. Retrieved from: https://www.academia.edu/81523640/Forecasting_of_Stock_Market_by_Combining_Machine_Learning_and_Big_Data_Analytics

Cite this article:

Md Zahidul Islam, Md Maruful Hoque Chowdhury, Mohammad Momen Sarker (2023). The Impact of Big Data Analytics on Stock Price Prediction in the Bangladesh Stock Market: A Machine Learning Approach. *International Journal of Science and Business, 28*(1), 219-228. DOI: https://doi.org/10.58970/IJSB.2216

Retrieved from http://ijsab.com/wp-content/uploads/2216.pdf

Published by