VFAST Transactions on Software Engineering
http://vfast.org/journals/index.php/VTSE@ 2023, ISSN(e): 2309-3978, ISSN(p): 2411-6246
Volume 11, Number 2, April-June 2023 pp:01-10

Financial Prices Prediction of Stock Market using Supervised Machine Learning Models

Muhammad Rehman*, Muhammad Fuzail, Muhammad Kamran Abid, Naeem Aslam

Department of Computer Science, of Engineering and Technology, Multan, Punjab, Pakistan
*Corresponding author email: 2k19mscs111@nfciet.edu.pk

ABSTRACT

The process of predicting stock market movements may initially appear to be non-statistical due to the multitude of factors involved. However, machine learning techniques can be utilized to establish connections between past and present data, enabling the training of machines to make accurate assumptions based on the information. By effectively linking historical data to current data using machine learning, it becomes possible to make precise predictions regarding stock performance. These predictions can lead to substantial profits for individuals and their brokers. Traditionally, stock market predictions have exhibited chaotic patterns rather than randomness, which is why a thorough analysis of a market's historical data allows for predictions to be made. Machine learning offers an efficient means of modeling such processes. By closely aligning market predictions with actual values, the analysis's accuracy can be raised greatly. The field of stock prediction has seen a growing interest in machine learning among researchers due to its effectiveness and precision. Regression-based models are commonly employed when the objective is to forecast continuous values based on independent variables. To predict stock market closing prices for the upcoming ten to fifteen days, we used SVR, RF, KNN, LSTM, GRU, and LSTM with GRU in our study. In regression modeling, the R-squared (R2) value represents the percentage of difference explained by the independent variable(s). A higher (R2) value near to 1 indicates better performance. Our experiments yielded R2 values of 0.832, 0.832, 0.574, 0.838, 0.825, and 0.815 for SVR, RF, KNN, LSTM, GRU, and LSTM with GRU, respectively. Consequently, the most effective model for correctly predicting stock market closing prices is the LSTM learning model, which had the greatest R2 value of 0.838.

KEYWORDS

Supervised machine learning, Stock market, ANN, CNN, RNN, LSTM, and SVM.

JOURNAL INFO

HISTORY: Received: April 10, 2023 Accepted: May 17, 2023 Published: May25, 2023

1. INTRODUCTION

Factors influencing the value of stocks and the capital market's influence on the economy since always enjoyed the interest of economists. Relationships between market valuations capital, and new information on individual values are the foundations of an efficient market, in which a better-informed individual has the chance to make better decisions on investment. It is therefore important to understand the fundamentals of capital market pricing it is also an issue for practitioners, and knowledge of the determinants of stock prices is an advantage information that can lead to above-average profits. Market participants use different approaches to determining the potential profitability of investments and the reduction of risk capital associated with the placement of capital [1]. Traditional methods include fundamental analysis, portfolio analysis, and technical analysis (and their combinations), but as digitization continues, the approach to investing is gradually changing. Algorithmic trading, which uses large amounts of power, is becoming a common phenomenon in computing for making instant transactions. Combining the development of statistics and probability applications and greater availability of high computing power contributed to the popularization of machine learning and networking that estimates results through iterative learning of relationships between data [2]. Thanks to easy access to the stock market and macroeconomic data, it becomes a popular application of these methods also for forecasting future values of stock exchange rates. Data science is a constantly evolving field of science, and discoveries in it give rise to a chance for greater forecasting accuracy, including the stock exchange's stock price [3].

The challenge of predicting financial prices in the stock market using supervised machine learning models is to forecast stock values based on past performance and other important variables. This is typically done by training a machine learning model on data from the past stock market, including previous stock prices and trading volume, and other financial indicators, and then applying the trained model to new data to produce predictions. The objective is to develop a model that can correctly forecast future stock prices, which traders as well as investors may use to decide intelligently whether to purchase and sell stocks [4]. This is a challenging problem because stock prices are affected by many factors, including economic conditions, company performance, and market sentiment, and these factors can change quickly and unpredictably. Machine learning, especially its most computationally complicated method, deep neural networks

This work is licensed under a Creative Commons Attribution 3.0 License.

are gaining popularity in multiple fields such as genetics [5], sarcasm detection [6], disease detection [7], image retrieval [8], and stock market analysis [9]. It is influenced by many factors, such as greater universal availability of computing power, thanks to more and more technologically advanced personal computers, and Digitization results in greater availability of data in the form of computer files and databases. The rapid development of commercial solutions in this field and its growing popularity contribute to the great interest of the scientific community [10]. This, in turn, is associated with significant progress and a rapid pace of changes in machine learning trends. The scale of these changes features includes MIT researcher Lex Friedman, who is at the beginning of each. The year discusses the prevailing trends and the changes implemented by scientists that are becoming become the new norm [11]. Among relatively new achievements in this field, giving promising results for time series, there are recursive networks and LSTM networks [6] that can remember dependencies between distant points in time [12]. Models based on them will be used in this work for modern tools based on which the transaction system will be built. Every investor strives to be as effective as possible in his trading system. Even though the achievement of 100% (or even close to this threshold) effectiveness seems to be impossible with the use of only historical data, it is assumed among practitioners that achieving effectiveness greater than 50% is a good result. Neural networks are a solution that is increasingly used to build transactional systems, which is why this study's objective is to examine how they are used to generate investment signals. The literature review shows that neural networks, and especially recursive networks, spend prove to be a promising tool to support investment decisions. Hence the hypothesis of this thesis is that networks built based on LSTM neurons using historical data such as stock prices, commodity prices, exchange rates, and data macroeconomic and technical analysis indicators can be an effective tool for generating investment signals [12].

In today's economy, there is a great deal of volatility in the stock market, and this complexity in the market is one of the key reasons why researchers and statisticians are striving to discover how you can predict it. Many people still claim that there is no way to forecast the stock market, despite the fact that there have been hundreds of research publications in the field over the years [13]. As a result, stock prices are affected by a wide range of factors that may or may not be known. Using past experiences to predict future price changes is a more common method of making stock market predictions. According to Lee Giles, Due to limited sample sizes, excessive noise, non-stationarity, and nonlinearity, the challenge of financial forecasting is difficult. Getting a large sample would require an extended period for stocks, since the information is often generated from actual time stock transaction values. Financial markets are not always stable over a longer period of time, so longer periods do not always yield the right results. If the elements that affect stock prices can be realized, stock price predictions may be possible without using previous prices. This thesis takes a closer look at each of the factors that contribute to such a problem. It is likely the most effective strategy for this type of application to rely on machine learning due to the complex nature of stocks. The use of some form of machine learning for Several research papers have reported on stock price forecasting, such as feed-forward neural networks [1][2], support vector machines [14], recurrent neural networks, linear SVM and MLPs, when we try to evaluate the short term price movements [15]. More recently [16][17] employed neural networks to forecast the next day's closing price. Comparative research by [16] showed that for various companies a simple artificial neural network outperformed a random forest based on their performance based on a simple neural network. The following are the main contribution of this research

- To examine how well various ensemble methods, including as random forest, gradient boosting, and k-nearest neighbors, predict the stock market.
- To examine the effects of different technical indicators and financial variables on stock market prediction using machine learning models.
- To assess the effectiveness of the models using several performance indicators, such as R-squared, mean absolute error, and Sharpe ratio.

Part 2 of the article, which includes the literature part, contains the remaining text. Part 3 provides explanations of the methodology, dataset and its processing, and models. The section discusses the outcomes of all employed models as well as a comparison of these outcomes.

2. LITERATURE REVIEW

A variety of elements, such as currencies and exchange rates, can have an impact on global enterprises. There is a rigorous analysis of the price movements of stock and exchange rates in [17]. In addition, raw material prices have also been shown to have an impact on stock prices, such as oil prices and aluminum prices.[18]. It is also important to consider the other stock markets as well [19]. Stocks traded on a daily basis [20] can be measured by the volume of trades. Stock prices can be dramatically impacted by changes in management as well as news on the company itself at the time of a change in management at a corporation [21]. Macroeconomic news can have a significant impact on stock values, such as information concerning changes in interest rates and inflation [22][23]. The volume of traded stock and its price may also be affected by speculations on Internet forums. In the short term, the psychological variables that influence stock pricing in a billion-dollar company can have a significant impact on its share price. Short-term, psychological factors can have a significant impact on how well stocks perform, since an Internet forum post is seemingly insignificant and peripheral, especially in the short run. Trends in the stock market are often the result of bubbles that burst and crashes that follow.

In bubbles, investors are driven by greed, while in crashes, investors are motivated by fear. This is referred to as herd behavior by some researchers. When traders rush in and out of the market, they join the herd of other traders [24]. Markets are governed by psychological forces such as greed, fear, and herd mentality. As well as these obvious factors, there are other, less obvious aspects that play a role as well. It is useful to note that the movement of stock prices on Mondays and Fridays is often different from one another for this reason. Despite the fact that sunny days can be good for the mood, they can also be useful for traders who are looking forward to a more optimistic market in order to fuel their optimism [25]. Does stock price movement follow a predictable pattern? Some researchers suggest that stock prices move according to the theory of the rand. It should be noted that Stock prices do not follow impossible difficult to the title 8 of a heavily cited paper. Stock returns are to some extent predictable, according to the authors of the paper. Thus, the past price behavior of a stock is an excellent indicator of future price movement. Many historians say that there are certain patterns in history that repeat themselves in the future. These are the result of past events being repeated over and over again. Considering that there are patterns in these patterns, it can be helpful to develop an understanding of them under the assumption that these patterns can be studied and analyzed. As a result, stock prices can also be predicted by these patterns [26]. However, economists are unable to agree on whether stock prices move randomly or not. There are supporters of both random walk theory and predictable movement's theory who believe that their theories have been empirically verified [27].

Considering the profits that good traders make on the stock market, we know that in terms of finances, it is possible to predict the stock market to a certain extent. There is still only a little over 50% accuracy in predicting whether the stock market is going up or down using machine learning [28]. While humans are capable of identifying patterns relating to stock price movements and factors that affect those prices, a computer is not as of yet able to do so to a satisfactory degree, even though it is capable of identifying patterns between them. In terms of definition, machine learning can be defined as the branch of computational intelligence that deals with the design and development of computer programs that are automatically improved with experience. "Improve with experience" means learning. Surely, what exactly is learning entails the question that arises here is the one which challenges us to answer. Learning is a function that allows animals to modify and reinforce or acquire new information, skills, patterns, and behaviors [29]. Unlike machines, animals have the ability to adapt to their environment, which makes them distinct from machines. Most of the modern machine learning research focuses on creating various algorithms that can be used on specific applications. With the help of machine learning, data is used to learn a model that will be used to create a prediction. It is possible to make predictions using this model. Codes generated by machine learning do not have hardcoded rules, but instead, establish their own rules through the use of models.

A large amount of data is one of the most important requirements for machine learning. An algorithm that uses machine learning is generally more effective when the data is better. By comparing the inputs with the expected outputs, the algorithm attempts to determine a correlation. In order to accomplish this, the training data and testing data are separated into two groups. Models are created using training data, and the performance of the models is tested using testing data, after which the models are re-generated. Getting the right data is one of the most important aspects of machine learning. In addition to having the correct size and relevance of the data, the data must also be relevant.

Throughout the rest of this thesis, there is more information about this topic to be discussed. In general, there are three different types of machine learning, which can be divided into the following categories: supervised learning, unsupervised learning, and semi-supervised learning.

Table 1. Review of most Research-Related Articles

Authors	Dataset	Technique Used	Metrics Performance	Limitations
[30]	S&P500, CMSB, DJI	MLP, CNN, RNN, LSTM, CNN+RNN, CNN+ LSTM	MAE, RMSE, R-Square	It only use close price to predict the stock price and failed to integrate emotional factors like news and national policy
[31]	Five Stocks in SSE	CNN+ LSTM	MAPE	The predictability of this model partially relies on the fact that the strategy is hardly implementable given the trading restrictions in China
[32]	Google Trends	BPNN	Hit Ratio	More computation power and time required
[33]	China stock index price	SVM	-	The exchange rate used only SZ50
[34]	Chines A-share Market	SVM, RF, ANN	Sharp Ration	Only the Chinese dataset is used and the feature selection algorithm needs to be optimized

[35]	KOSPI200	XGB,ET, RF,DT,SVM_RBF,KNN	Accuracy, F1, Precision, Recall, AUC	Used only predefine No. of indicators
[36]	US stock market	MLP,RF, XGB(EGB), and LSTM using effective transfer entropy (ETE)	Accuracy and adjusted accuracy	More computation time required
[37]	CSI 300 stock index	SVR, LSTM, CNN, CNN-LSTM, TPM-NC, TPM	RMSE	More computation power and time required
[38]	Shanghai Composite Index stock	MLP, CNN, RNN, LSTM, BiLSTM, CNN-LSTM, CNN-BiLSTM, BiLSTM-AM	MAE, RMSE, R-Square	More computation power and time required
[39]	Chinese stock market	SVM,NB,MLP,RFE,FE+RFE+PCA+LSTM	Accuracy F1 score	This purpose model does not provide accurate results for more than a week

3. METHODOLOGY

Following the literature review, this research technique includes finding, selecting, processing, and analyzing information about a topic, enabling the topic. Our goal is to examine the terminology used to distinguish qualitative, quantitative, and mixed-methods strategies. ML classifiers have been proven to be effective and efficient in assembling models based on majority voting for any task. ML using a labeled dataset for the target task.

In Figure **1**, the proposed methodology shows 1. Data Collection: Collect historic stock market data such as stock prices, volume, etc. 2. Data Preprocessing: Clean and normalize the data to ensure that it is suitable to be used in the analysis. 3. Feature Selection: Select the features that are relevant to the prediction. 4. Model Building: Build the model using Algorithms for machine learning including Linear Regression, Random Forest, and Support Vector Machine, etc. Model Evaluation: Use several metrics, such as Mean Absolute Error and Root Mean Squared Error, to assess

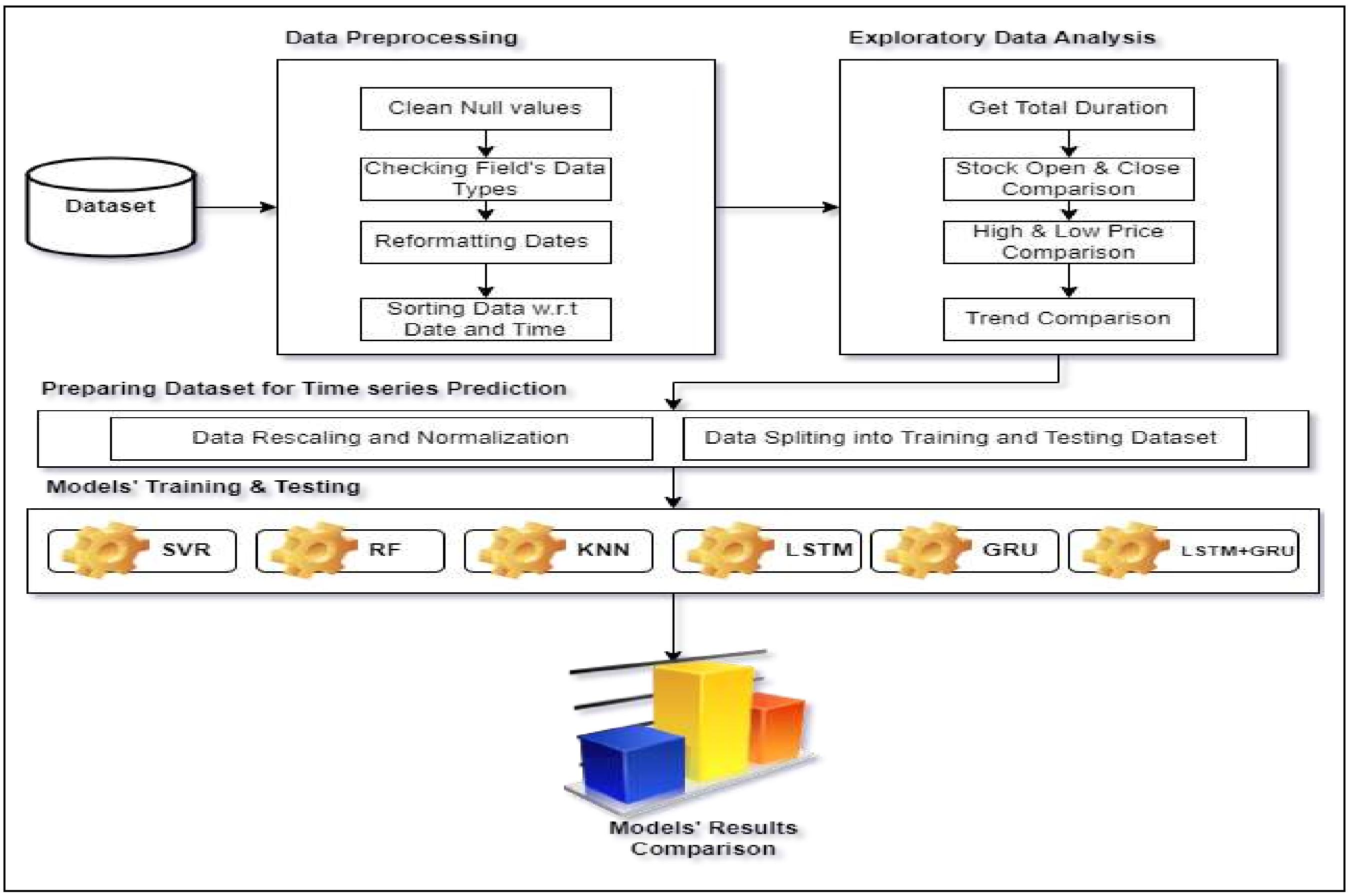

Figure 1. Proposed Methodology

the model's accuracy. 6. Model Deployment: Set up the model in a real-world setting to make predictions instantly.

A. Dataset

A major component of machine learning is data. Improvements in learning algorithms (like deep learning) Data is a key element in machine learning. Due to the time required to label the data, it is typically challenging and expensive to develop high-quality labelled training datasets for supervised and semi-supervised machine learning algorithms. We used the Reliance stock price dataset to forecast the stock market's closing prices. It is a CSV file that was retrieved and covers the exact same period of time, from 8/19/2020 to 8/18/2021. 251 samples from the dataset are displayed in Figure **3**.

	A	B	C	D	E	F	G
1	Date	Open	High	Low	Close	Adj Close	Volume
2	8/19/2020	2141	2154	2121.35	2131.55	2124.715	15731396
3	8/20/2020	2120	2123.9	2088	2097.05	2090.326	10401212
4	8/21/2020	2118	2122	2077	2081.85	2075.174	11667129
5	8/24/2020	2091.4	2104.5	2070.5	2095.75	2089.03	15098991
6	8/25/2020	2106	2111.3	2078	2082.1	2075.424	8947563

Figure 3. Sample Dataset

B. Data preprocessing

During the data mining process, preprocessing refers to manipulating or dropping data in order to optimize performance. Preprocessing is particularly relevant in data mining and machine learning. The majority of data collection methods do not provide control, which results in out-of-range values, uncontrollable data combinations, and missing data when the data is collected. Before conducting any analyses, it is important to carefully screen data for these problems in order to avoid misleading results.

Data preparation is important in machine learning projects, particularly in computational biology. When there is a lot of duplicate and irrelevant information or noisy and untrustworthy data, learning might be more challenging.. Preparing and filtering data may take a long time. In the final training set, data can be cleaned, selected, normalized, one-hot encoded, transformed, and features extracted and selected.

C. Support Vector Regression

This supervised learning technique use the Support Vector Regression algorithm to forecast isolated values. Using the SVR algorithm, the best-fit line is determined by finding the hyperplane with the most points. In much the same way as SVMs, it finds a line that is best suited. As an alternative, an SVR is a regression model that is made to identify a line that is inside a specific threshold value in order to reduce the error between the real and predicted values. Given that the threshold value is defined by the separation between the hyperplane and boundary line in the dataset, SVR has a significant drawback in that it is challenging to scale up to datasets with more than a few tens of thousands of samples.

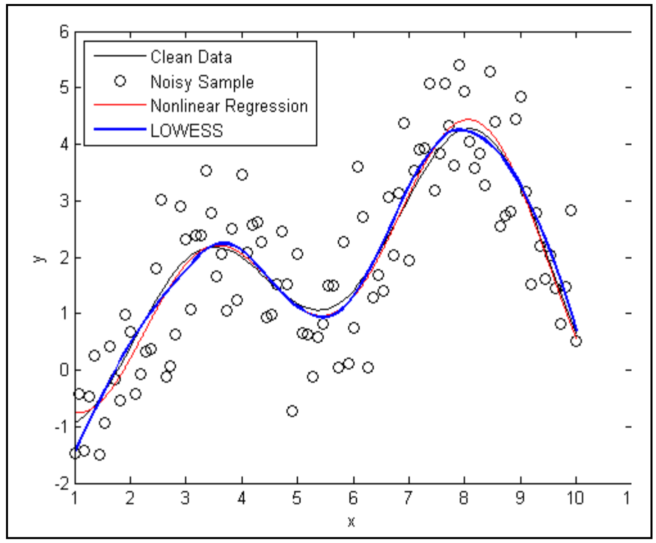

Figure 2. Support Vector Regression

In Figure **2**, the hyperplane is a line in the input space that divides the data points into two categories, one on either side of the line. The support vectors are the data points nearest to the hyperplane, and they determine the location and direction of the hyperplane. The SVR model then uses the support vectors to generate a regression model that best fits the data points. This model can then be used to predict future values.

Large datasets are generally recommended for linear SVRs or SGD Repressors. The linear SVRs are faster than SVRs, but they only take into account the linear kernel, and a linear kernel is the only input. Since the cost function disregards samples that are very close to the target and only considers a limited amount of training data, Support Vector Regression only relies on a limited amount of training data.

D. Random Forest

Machine learning encompasses several algorithms, including Random Forest, which aims to enhance model performance for tackling intricate problems. Random Forest is a commonly employed algorithm in both classification and regression tasks. To improve model performance, an approach called ensemble learning is utilized, which involves combining multiple classifiers. By merging these classifiers, the overall performance of the model is enhanced, emphasizing the objective of achieving superior results.

A number of decision trees are employed in the Random Forest classification model to create predictions using various subsets of the dataset, which are then averaged to increase accuracy. Instead of depending solely on one decision tree, the random forest model incorporates the forecasts from each tree plus a majority vote to anticipate the outcome. It consequently offers improved precision and guards against overfitting in more densely forested areas. The Random Forest algorithm is demonstrated in the following diagram.

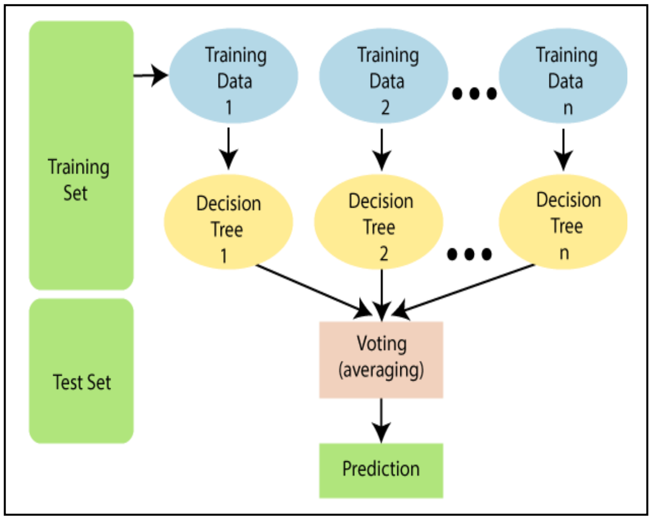

Figure 4. Random Forest

In Figure **4**, RF shows how to make predictions, the algorithm uses a voting system, where the final output is the one that receives the most votes from the individual decision trees. The Random Forest algorithm is robust to overfitting and provides accurate results even when the data is highly unbalanced. It also has a few hyperparameters that can be tuned to improve performance.

E. K Nearest Neighbors

This algorithm provides supervised machine learning for the classification and regression of unknown variables. KNN is the acronym for K-Nearest Neighbor in machine learning. Identifying a new unknown variable's nearest neighbors is the key to predicting or classifying it. In order to gain a better understanding of this awesome algorithm, let's take a closer look at a real-world scenario that will help us understand how it works. Our closest peers often tell us that we have a lot in common, whether it is about our philosophies, working habits, or other characteristics.

Therefore, we can establish relationships with individuals who share similar characteristics to us. By using a distance-based approach, the KNN algorithm determines which class a new unknown data point belongs to based on the distance-based approach. In order to locate all of the nearby neighbors of the new unknown data point based on the distance-based approach, the algorithm uses a distance-based approach.

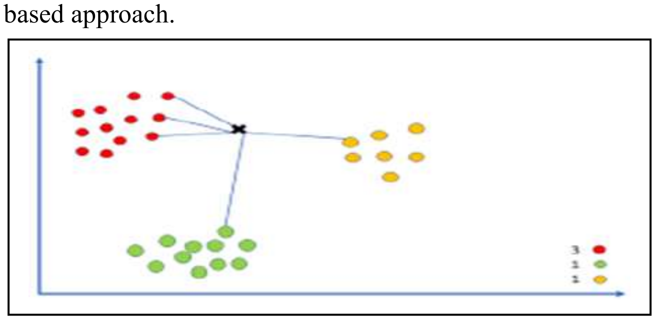

Figure 5. K-Nearest Neighbors

In Figure **5**, KNN shows how to make a prediction, KNN looks at the k (a user-defined parameter) closest data points to the query point and compares their classes to determine the class of the query point. The algorithm makes its prediction by taking the majority vote from the KNN.

F. Long Short-Term Memory

A Long Short-Term Memory (LSTM) is a type of recurrent neural network (RNN) specifically designed to tackle sequence prediction problems that involve order dependence and long-term memory. LSTMs have become a crucial component in various fields of deep learning, including machine translation, speech recognition, and more.

Unlike traditional RNNs, LSTMs have the ability to capture long-range dependencies and effectively model sequences with temporal dynamics. This is accomplished through a sophisticated architecture that incorporates memory cells, input and output gates, and forget gates. The memory cells allow LSTMs to remember information over extended periods, while the gates regulate the flow of information within the network.

LSTMs excel in tasks where context and sequential information play a vital role. Machine translation, for instance, requires the understanding of the context of a sentence to produce accurate translations. LSTMs can process sequences of words while maintaining an understanding of the entire context, resulting in improved translation quality.

Another application of LSTMs is in speech recognition. Speech is inherently sequential, and LSTMs can effectively model the temporal dependencies in speech data. By capturing long-term dependencies, LSTMs can handle variable-length input sequences and accurately recognize spoken words, leading to advancements in automatic speech recognition systems.

To fully comprehend LSTMs, it's essential to understand related concepts such as bidirectional modeling and sequence-to-sequence architectures. Bidirectional LSTMs leverage information from both past and future inputs, enhancing the model's understanding of the current context. This bidirectional approach is valuable in tasks like sentiment analysis, where the sentiment of a sentence may depend on the words that follow.

Sequence-to-sequence architectures utilize LSTMs to map input sequences to output sequences. This is particularly useful in tasks like machine translation, where an input sequence (source language) needs to be translated into an output sequence (target language). LSTMs enable the model to encode the input sequence into a fixed-size representation and then decode it into the desired output sequence.

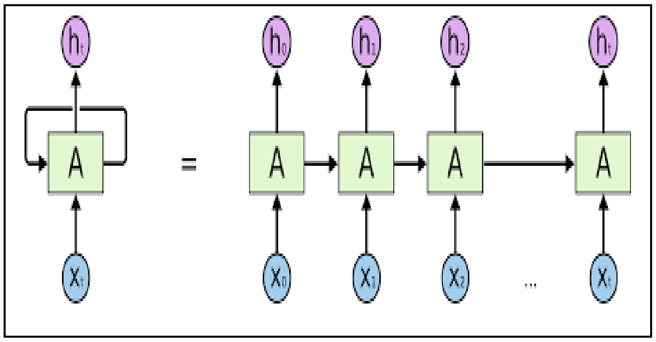

Figure 6. Long-Term memory

Figure **6**, LSTM shows that A Long Short-Term Memory (LSTM) network is made up of a series of repeating modules, each of which has four interconnected layers. These are the layers: The input layer is where the input data, which is often a series of values, are sent for the LSTM network to process. The forgetting gate layer decides whether information from the previous cell state should be maintained and which information should be forgotten.

The input gate layer plays a crucial role in identifying the relevant new information from the input that should be retained in the current cell state. The cell state layer is responsible for storing the combined information from both the forget gate layer and the input gate layer. On the other hand, the output gate layer determines which information from the cell state should be output. Finally, the output generated by the LSTM network is utilized for making predictions or decisions.

G. Gated Recurrent Unit

The Gated Recurrent Unit (GRU) is a type of recurrent neural network (RNN) architecture that is specifically designed to handle sequence data. Similar to LSTMs, GRUs are capable of capturing long-term dependencies and modeling sequences with temporal dynamics.

GRUs were introduced as a simplified version of LSTMs, aiming to reduce the complexity and computational requirements while maintaining comparable performance. They achieve this by combining the input and forget gates into a single update gate and modifying the memory cell.

The key components of a GRU include:

1. Update Gate: The update gate determines how much of the previous memory should be retained and how much of the new input should be incorporated into the current memory.
2. Reset Gate: The reset gate controls which parts of the previous memory should be forgotten, allowing the GRU to adapt to new input and update its internal representation.
3. Current Memory: The current memory represents the updated internal state of the GRU, which is a combination of the previous memory modified by the update and reset gates and the current input. GRUs excel in scenarios where memory and computational efficiency are important factors. They have demonstrated good performance in tasks such as machine translation, language modeling, speech recognition, and sentiment analysis. GRUs can effectively process sequential data, capture dependencies over long sequences, and generate accurate predictions or classifications.

Compared to LSTMs, GRUs have a simpler architecture with fewer gates, making them computationally efficient and easier to train with limited data. However, their performance may vary depending on the specific task and

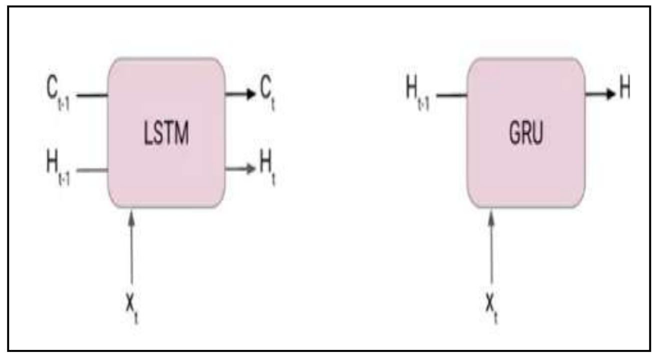

Figure 7. Gated Recurrent Unit

dataset, and it is often recommended to experiment with both LSTM and GRU architectures to find the best fit.

In **Error! Reference source not found.**, shows that Recurrent neural network (RNN) architectures include GRUs as a subtype. Update gate and reset gate cells make up its two separate cell kinds. Which information from the previous hidden state should be retained and which should be forgotten is determined by the update gate. This is done by calculating the dot product of the current input vector and the previously hidden state vector and then applying a sigmoid function to this result. This determines which information should be kept and which should be forgotten.

The reset gate determines which data from the current input should be retained and which should be erased. This is done by calculating the dot product of the current input vector and the previously hidden state vector and then applying a sigmoid function to this result. This determines which information should be kept and which should be forgotten. Finally, the output of the update gate and the output of the reset gate combine to form the output of the GRU. This combination is then used to produce the next hidden state. Another intriguing feature of GRU is that, in contrast to LSTM, it lacks a distinct cell state (Ct). Only the hidden state (Ht) is present. GRUs are easier to teach because of their simpler architecture.

H. LSTM with GRU

The Gates serve as an internal mechanism in LSTM

(Long-Short-Term Memory) and GRU (Gated Recurrent Unit), controlling which information is retained and which is deleted. GRU networks resolve the exploding and vanishing gradient problem by using this LSTM.

Results and Discussion

In this chapter, we present the model's suggested results. To generate scores related to regression, we applied all of the aforementioned learning models, including SVR, RF, KNN, LSTM, GRU, and LSTM with GRU.

- A comparison of the stock's original closing price and the forecasted close price.
- An estimation of the close price of stocks based on the last 15 days for the coming 10 days.

I. Evaluation metrics RMSE, MSE, and MAE

The main methods for calculating a model's error in forecasting quantitative data are the Root Mean Square Error (RMSE), Mean Square Error (MSE), and Mean Absolute Error (MAE).

L. Models Performance Comparison

In this research, I have applied six models to predict stock market trends. LSTM and RF give the highest VR and R2 among all applied models. The performance comparison of applied models is given in the **Error! Not a valid bookmark self-reference.** and graph Figure **8**. We had shown the prediction of all applied models in bellow diagram.

Table 2. Performance Comparison

Model	VR	R2
SVR	0.836	0.832
RF	0.844	0.832
KNN	0.614	0.574
LSTM	0.841	0.838
GRU	0.834	0.825
LSTM+GRU	0.838	0.815

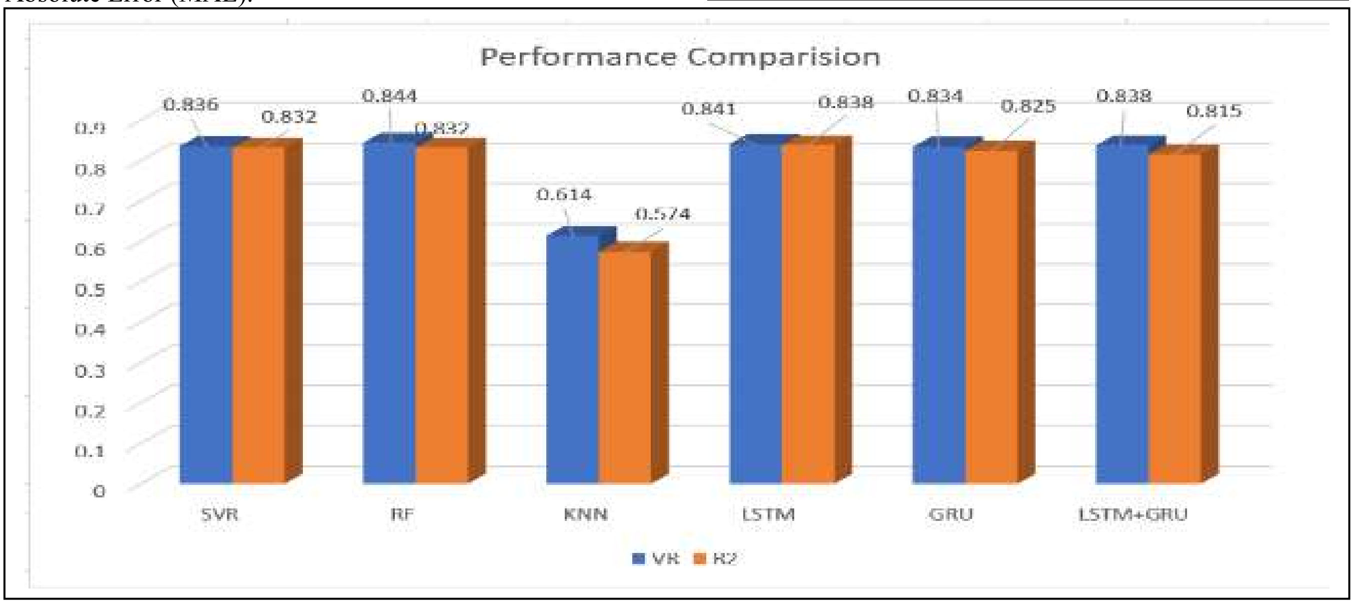

Figure 8. Performance Comparison

J. Explained variance regression score

The formula for the explained variance score, which explains the dispersion of mistakes in a given dataset, is as follows: Here, Var(y) represents the variance of actual values and forecast errors, respectively. Better squares of standard deviations of errors are indicated by scores near to 1.0, which are greatly desirable.

K. R^2 score for regression

R-squared (R2) is an indicator of statistical significance that depicts the percentage of variation for a dependent variable in a regression model that is explained by one or more independent variables.

1=Best
0 or < 0 = worse

In Figure **8**, a Comparison of performance reveals that the value of R-squared is thought to be at its best if it is close to 1. Based on our trials, we were able to determine the R2 values for SVR, RF, KNN, LSTM, GRU, and LSTM with

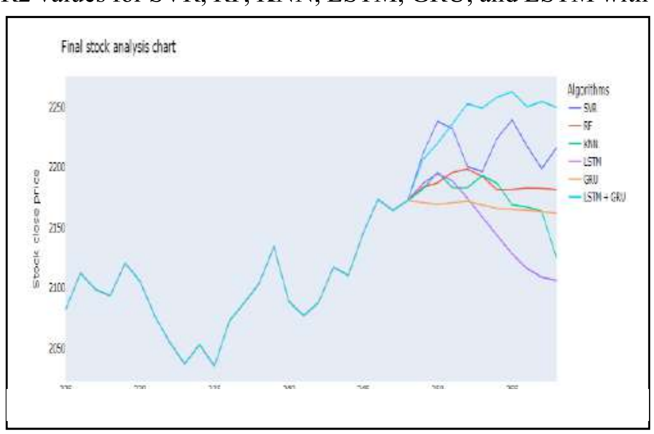

Figure 9. Final Stock Analysis

GRU to be 0.832, 0.832, 0.574, 0.838, 0.825, and 0.815. Due to its high R2 of 0.838, the LSTM learning model is the best one for overcasting stock market closing prices.

We performed the final stock analysis of algorithms. This diagram shows different algorithms predicting stock market close price.

4. CONCLUSION AND FUTURE WORK

There seem to be a lot of factors to be considered when predicting the stock market, so it does not seem statistical at first glance. In addition, it is possible to use machine learning techniques to link previous data to current data and train the machine to make appropriate assumptions based on the data. However, it is entirely feasible to utilize machine learning algorithms to connect historical data to recent data. If you can able to make a correct prediction of a stock, you can earn huge profits for yourself and your broker. As is often emphasized, predictions are chaotic rather than random in nature, which means they can be predicted by carefully examining the history of a particular stock market in order to predict them. Using machine learning, we can able to model such processes efficiently. The analysis' accuracy can be improved by forecasting the market value so closely to the tangible value. Due to its effectiveness and accuracy, machine learning has attracted a lot of researchers working in the field of stock prediction.

In this study, we utilized several Regression-Based Models, namely SVR, RF, KNN, LSTM, GRU, and LSTM with GRU, to make predictions on stock market closing prices for the next ten to fifteen days. The primary goal was to identify a model with a high R-squared, which is a statistical measure indicating the proportion of the dependent variable's variance explained by the independent variables in a regression model. A value close to 1 signifies a stronger fit.

After conducting our experiments, we obtained R-squared (R2) values of 0.832, 0.832, 0.574, 0.838, 0.825, and 0.815 for SVR, RF, KNN, LSTM, GRU, and LSTM with GRU, respectively. These values represent the goodness of fit for each model in explaining the variability in the dependent variable. Among them, the LSTM learning model exhibited the highest R2 value of 0.838, indicating its superior performance in accurately predicting stock market closing prices.

To summarize, when addressing this regression problem, our study concluded that the LSTM model outperformed the other models with an R2 value of 0.838, making it the most suitable choice for forecasting stock market closing prices.

We notice that the LSTM-supervised machine learning model required more CPU power rather than other models which would cause running processes for hardware resources. If we use a large data set then required more CPU power to evaluate the results. That is why we used exactly one year, if we used large data they may produce more accurate results.

The future accuracy of the stock market forecast system is probably going to be improved by a much larger dataset. To determine if other new machine learning models are accurate in the future, they can also be looked at. In terms of predicting stock prices, machine learning is a promising area of research. Deep learning-based models can also be used to analyze sentiment on how news affects a company's stock price.

REFERENCES

[1] K. Borowski, Fundamental analysis. Enterprise Valuation Methods. Difin SA, 2014.

[2] K. Borowski, Fundamental analysis. Enterprise Valuation Methods. Difin SA, 2014.

[3] M. Adya and F. Collopy, "How effective are neural networks at forecasting and prediction? A review and evaluation," J. Forecast., vol. 17, no. 5 - 6, pp. 481−495, 1998.

[4] M. H. Lang, K. V Lins, and D. P. Miller, "ADRs, analysts, and accuracy: Does cross listing in the United States improve a firm's information environment and increase market value?," J. Account. Res., vol. 41, no. 2, pp. 317–345, 2003.

[5] K. Vasista, "Role Of a Stock Exchange In Buying And Selling Shares," Int. J. Curr. Sci., vol. 12, no. 1, pp. 1770–2250, 2022.

[6] A. Naeem, A. H. Khan, S. u din Ayubi, and H. Malik, "Predicting the Metastasis Ability of Prostate Cancer using Machine Learning Classifiers," J. Comput. Biomed. Informatics, vol. 4, no. 02, 2023.

[7] H. Saleem, A. Naeem, K. Abid, and N. Aslam, "Sarcasm Detection on Twitter using Deep Handcrafted Features," J. Comput. Biomed. Informatics, vol. 4, no. 02, pp. 117–127, 2023.

[8] A. Naeem, M. S. Farooq, A. Khelifi, and A. Abid, "Malignant melanoma classification using deep learning: datasets, performance measurements, challenges and opportunities," IEEE Access, vol. 8, pp. 110575–110597, 2020.

[9] A. Naeem, T. Anees, K. T. Ahmed, R. A. Naqvi, S. Ahmad, and T. Whangbo, "Deep learned vectors' formation using auto-correlation, scaling, and derivations with CNN for complex and huge image retrieval," Complex Intell. Syst., pp. 1–23, 2022.

[10] K. Pahwa and N. Agarwal, "Stock market analysis using supervised machine learning," in 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), 2019, pp. 197–200.

[11] G. P. Danezis, A. S. Tsagkaris, F. Camin, V. Brusic, and C. A. Georgiou, "Food authentication: Techniques, trends & emerging approaches," TrAC Trends Anal. Chem., vol. 85, pp. 123–132, 2016.

[12] D. Simberloff et al., "Impacts of biological invasions: what's what and the way forward," Trends Ecol. Evol., vol. 28, no. 1, pp. 58–66, 2013.

[13] K. Gemra, "The cost of financing operating activities through public bond issues," Ekon. i Organ. Przedsiębiorstwa, no. 3, pp. 113–125, 2018.

[14] M. McCombs, "A look at agenda-setting: Past, present and future," Journal. Stud., vol. 6, no. 4, pp. 543–557, 2005.

[15] M. J. Flannery and A. A. Protopapadakis, "Macroeconomic factors do influence aggregate stock returns," Rev. Financ. Stud., vol. 15, no. 3, pp. 751–782, 2002.

[16] A. Tsantekidis, N. Passalis, A. Tefas, J. Kanniainen, M. Gabbouj, and A. Iosifidis, "Using deep learning to detect price change indications in financial markets," 25th Eur. Signal Process. Conf. EUSIPCO 2017, vol. 2017-Janua, pp. 2511–2515, 2017, doi: 10.23919/EUSIPCO.2017.8081663.

[17] D. M. Cutler, J. M. Poterba, and L. H. Summers, "What moves stock prices?," National Bureau of Economic Research Cambridge, Mass., USA, 1988.

[18] R. A. Ajayi and M. Mougouė, “On the dynamic relation between stock prices and exchange rates,” J. Financ. Res., vol. 19, no. 2, pp. 193–207, 1996.

[19] N. APARGIS and S. M. MILLER, “Do structural oil-market shocks affect stock price,” Energy Econ., vol. 31, no. 1, pp. 569–575, 2009.

[20] Y. Hamao, R. W. Masulis, and V. Ng, “Correlations in price changes and volatility across international stock markets,” Rev. Financ. Stud., vol. 3, no. 2, pp. 281–307, 1990.

[21] A. R. Gallant, P. E. Rossi, and G. Tauchen, “Stock prices and volume,” Rev. Financ. Stud., vol. 5, no. 2, pp. 199–242, 1992.

[22] W. F. M. De Bondt and R. Thaler, “Does the stock market overreact?,” J. Finance, vol. 40, no. 3, pp. 793–805, 1985.

[23] J. B. Warner, R. L. Watts, and K. H. Wruck, “Stock prices and top management changes,” J. financ. econ., vol. 20, pp. 461–492, 1988.

[24] G. McQueen and V. V. Roley, “Stock prices, news, and business conditions,” Rev. Financ. Stud., vol. 6, no. 3, pp. 683–707, 1993.

[25] M. K. Brunnermeier and M. K. Brunnermeier, Asset pricing under asymmetric information: Bubbles, crashes, technical analysis, and herding. Oxford University Press on Demand, 2001.

[26] D. Hirshleifer and T. Shumway, “Good day sunshine: Stock returns and the weather,” J. Finance, vol. 58, no. 3, pp. 1009–1032, 2003.

[27] E. F. Fama, “The behavior of stock-market prices,” J. Bus., vol. 38, no. 1, pp. 34–105, 1965.

[28] N.-M. Jula and N. Jula, “Random Walk Hypothesis in Financial Markets,” Challenges Knowl. Soc., pp. 878–884, 2017.

[29] C. K.-S. Leung, R. K. MacKinnon, and Y. Wang, “A machine learning approach for stock price prediction,” in Proceedings of the 18th International Database Engineering & Applications Symposium, 2014, pp. 274–277.

[30] T. M. Mitchell and T. M. Mitchell, Machine learning, vol. 1, no. 9. McGraw-hill New York, 1997.

[31] W. Lu, J. Li, Y. Li, A. Sun, and J. Wang, “A CNN-LSTM-based model to forecast stock prices,” Complexity, vol. 2020, pp. 1–10, 2020.

[32] N. Jing, Z. Wu, and H. Wang, “A hybrid model integrating deep learning with investor sentiment analysis for stock price prediction,” Expert Syst. Appl., vol. 178, p. 115019, 2021.

[33] H. Hu, L. Tang, S. Zhang, and H. Wang, “Predicting the direction of stock markets using optimized neural networks with Google Trends,” Neurocomputing, vol. 285, pp. 188–195, 2018.

[34] J. Zhang et al., “Can the exchange rate be used to predict the shanghai composite index?,” IEEE Access, vol. 8, pp. 2188–2199, 2019.

[35] X. Yuan, J. Yuan, T. Jiang, and Q. U. Ain, “Integrated long-term stock selection models based on feature selection and machine learning algorithms for China stock market,” IEEE Access, vol. 8, pp. 22672–22685, 2020.

[36] K. K. Yun, S. W. Yoon, and D. Won, “Prediction of stock price direction using a hybrid GA-XGBoost algorithm with a three-stage feature engineering process,” Expert Syst. Appl., vol. 186, p. 115716, 2021.

[37] S. Kim, S. Ku, W. Chang, and J. W. Song, “Predicting the direction of US stock prices using effective transfer entropy and machine learning techniques,” IEEE Access, vol. 8, pp. 111660–111682, 2020.

[38] Y. Chen, W. Lin, and J. Z. Wang, “A dual-attention-based stock price trend prediction model with dual features,” IEEE Access, vol. 7, pp. 148047–148058, 2019.

[39] W. Lu, J. Li, J. Wang, and L. Qin, “A CNN-BiLSTM-AM method for stock price prediction,” Neural Comput. Appl., vol. 33, pp. 4741–4753, 2021.

[40] J. Shen and M. O. Shafiq, “Short-term stock market price trend prediction using a comprehensive deep learning system,” J. big Data, vol. 7, no. 1, pp. 1–33, 2020.