Seminar Report

On

Stock market Analysis By training a Model using ML

By

Siddharth Varangaonkar

1032201708

Under the guidance of

Prof. Laxmi Bhagwat

MIT-World Peace University (MIT-WPU)
Faculty of Engineering
School of Computer Engineering & Technology

*** 2022-2023 ***

MIT-World Peace University (MIT-WPU)

Faculty of Engineering
School of Computer Engineering & Technology

CERTIFICATE

This is to certify that Mr. Siddharth Varangaonkar of B.Tech., School of Computer Engineering & Technology, Semester- VI, PRN. No. 1032201708, has successfully completed seminar on

Stock market Analysis By training a Model using ML

To my satisfaction and submitted the same during the academic year 2022 - 2023 towards the partial fulfillment of degree of Bachelor of Technology in School of Computer Engineering & Technology under Dr. Vishwanath Karad MIT- World Peace University, Pune.

Prof. Laxmi Bhagwat
Seminar Guide

Dr. Vrushali Kulkarni

Head
School of Computer Engineering & Technology

List Of Figures

List Of Tables

Abbreviations

Abbreviations	Meanings
LSTM	Long short-term memory
ANN	Artificial neural networks
SVM	Support vector machines

Acknowledgement

I **Siddharth Varangaonkar**, roll no. **PA45** would like to thank my guide and coordinator **Prof. Laxmi Bhagwat** for her support, timely guidance, openness, insightfulness, and tolerance throughout the completion of this report, her feedbacks and constructive criticism were very valuable to rightly direct me.

I am appreciative that MIT World Peace University's faculty and staff gave me the tools and facilities I needed to complete this research. I would like to express my gratitude to the library staff who provide me with credentials of digital library.

Finally, I would like to acknowledge the support and encouragement of my friends who willing cleared my doubts. Thank you all for your support, guidance, and encouragement.

INDEX

ABSTRACT

The stock market is a dynamic and complex financial arena where individuals and organizations can trade shares of publicly listed companies. It serves as a platform for companies to raise funds for expansion and innovation, while also providing opportunities for investors to grow their wealth. The Seminar focuses on using a type of neural network called Long Short-Term Memory (LSTM) to predict stock prices. The stock market is known for its complexity and volatility, which makes accurate predictions a challenging task. However, LSTM networks have gained popularity due to their ability to capture long-term patterns in sequential data, making them suitable for modeling time series data like stock prices.

In this Seminar, historical stock price data is collected and prepared for input into the LSTM model. The LSTM architecture is then built, consisting of multiple LSTM layers followed by a dense output layer. Dropout regularization is applied to prevent the model from overfitting the training data and to enhance its generalization capability. The constructed LSTM model is trained using the historical stock price data, and its performance is evaluated using separate test data that the model has not seen before. Evaluation metrics such as mean squared error and accuracy are used to measure the model's predictive ability. The experimental results demonstrate that the LSTM model is effective in predicting stock prices. The model achieves promising accuracy and demonstrates its ability to identify patterns and trends in the stock market data. These findings contribute to the existing research on using deep learning techniques, like LSTM, for forecasting financial time series.

Keywords: - LSTM, ANN, Stock Market, Technical indicators, Neural networks, Financial Data, Time series Analysis

1. INTRODUCTION

1.1 Overview

In Todays, world the people are trying to increase their assets and to became financially stable through searching means for second source of Income. Such type of income can be referred as indirect income. The latest trend in the area of Finance had suggested that people are keen in buying and selling stocks.

Now let us now have a look on what is stock and understand its basic structure. There are many definitions to describe the essence of stock. To keep it into simple words stock can be coined as a way the ownership of the company can be decided and is distributed and hold accordingly [1].

The person who holds about 51%of the stock of the company Became the person having the power to take the major financial decisions about the Organization. When the company faces some kind of loss in the sector then they release the stock in the market. Hence the price of the stock Reduces and when there is profit the major shareholders hold their stock and take the benefit from then and they helps the company to grow and prosper.

1.2 Stock Structure

The stock structure of a company is a critical aspect of its overall financial makeup. It refers to the way in which the ownership of the company is divided among its shareholders through the issuance of stocks. Each share of stock represents a portion of ownership in the company, and shareholders are entitled to certain rights, such as voting on major company decisions and receiving dividends [4].

The stock structure can vary depending on the company and the type of stocks issued. Common stock is the most common type and gives shareholders voting rights, but may have limited dividend payouts. Preferred stock, on the other hand, does not typically have voting rights but may have higher dividends and priority in the event of bankruptcy or liquidation [5].

The number of shares a company-issues can impact the power and influence of individual shareholders. If a company issues more shares, the percentage of ownership held by existing shareholders can be diluted, reducing their influence over company decisions [11].

Understanding the stock structure is essential for investors and stakeholders in evaluating a company's financial health and potential for growth. It can also impact the cost of capital, as investors may be more willing to invest in a company with a stable and attractive stock structure. Companies may adjust their stock structure to raise capital, reward shareholders, or maintain control of the company, making it important to monitor changes in the stock structure over time [8].

Figure. 1 Stock graphs

1.3 Stock Types

To have a brief overview about the stock market let us look into different type of stocks: -

Companies sell stocks to investors to raise funds. There are different types of stocks that a company can issue: -

- Common stock is the most basic type that represents ownership in a company. Common stockholders can vote on important company decisions and may receive dividends. Preferred stock doesn't allow voting rights but pays a fixed dividend and gives stockholders a higher claim on assets and earnings if the company goes bankrupt.
- Dual-class stock is when a company issues two classes of stock with different voting rights. Treasury stock refers to the company's own repurchased shares, which don't have voting rights or receive dividends.
- Penny stocks are low-priced stocks from small companies and are riskier due to their volatility and lack of regulation. It's important for investors to understand the features of each type of stock before investing their money.

1.4 Prediction Aspect of Stocks

As we had understood the basic meaning of the stock and how the stock market works. So now let us discuss about the prediction aspect of the stock market: -

Stock prediction is a technique that uses statistical and analytical methods to estimate how well a particular stock or the entire stock market will do in the future. It's really important for investors to have this kind of information to decide whether to buy, sell or keep stocks [3].

To make stock predictions, investors analyze a variety of factors that could impact a company's financial performance such as financial reports, economic indicators, industry trends and news events. They use models, such as machine learning algorithms, to look at past data and make projections about future stock prices and trends [7].

But, predicting stock performance isn't an exact science and there are many factors that can affect stock prices that are difficult to predict. Stock markets can also be really volatile and can suddenly shift due to unexpected news or events [5].

However, despite these challenges, many investors and traders still rely on stock prediction to make informed investment decisions. By using a combination of tools and judgement, investors can get a better understanding of the stock market and make smart investment choices [1].

To make accurate prediction regarding the price of the sock and whether or not be buy the stock the Data Analyst that are been working in our domain had come up with some of the ML and Deep learning Algorithms [9].

The Algorithm are been discussed below: -

Regression analysis uses statistical techniques to analyze historical data and find patterns that can be used to predict future stock prices.

Artificial Neural Networks (ANN) mimic the behavior of the human brain and analyze large amounts of data to predict stock prices. Genetic algorithms use a genetic process to optimize trading strategies and identify profitable stocks.

Support Vector Machines (SVM) use a mathematical model to identify patterns in data and predict future stock prices. Decision trees use a tree-like structure to analyze data and predict whether a stock price will rise or fall based on specific criteria.

Random Forests combine multiple decision trees to make more accurate predictions about future stock prices and are particularly useful for analyzing complex data sets. The choice of algorithm used depends on the specific data and problem at hand, as each algorithm has its own strengths and weaknesses.

1.5 Latest trends

One of the latest trends is the use of deep learning techniques, such as Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks. These techniques can analyze vast amounts of data, including news articles, social media, and other unstructured data sources, to make more accurate predictions about future stock prices [4].

Another trend is the use of natural language processing (NLP) to analyze news articles, earnings calls, and other textual data to identify patterns and trends that can impact stock prices. This technology is particularly useful in identifying sentiment and tone, which can be strong indicators of future stock performance [8].

Additionally, there is a growing interest in the use of blockchain technology to improve the accuracy and transparency of stock market predictions. By creating a decentralized and secure database of financial data, blockchain technology could potentially provide more reliable and accurate predictions about future stock prices [3].

Finally, there is an increased focus on the use of real-time data sources, such as social media and satellite imagery, to identify early indicators of market trends and make faster and more informed investment decisions [2].

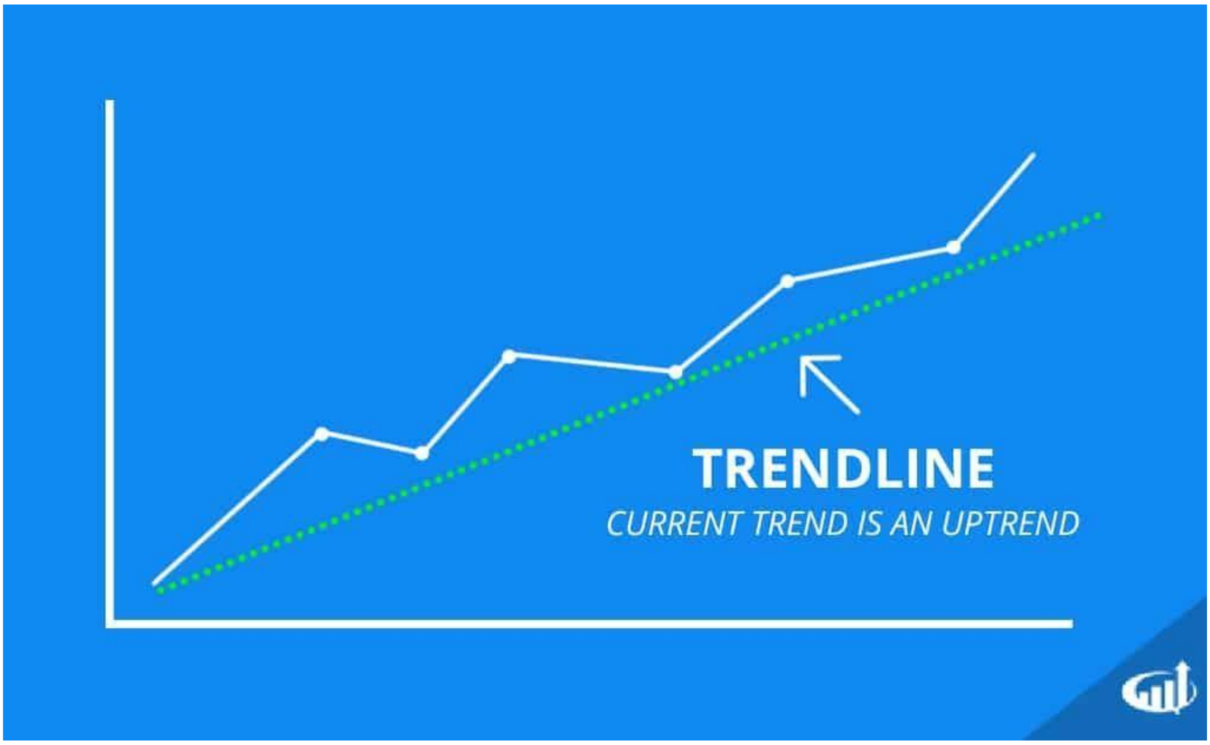

Figure. 2 Current trend Line

2. LITERARY SURVEYS

This study reveals that by combining Artificial Neural Network and Random Forest techniques with the enhanced computational power provided by artificial intelligence, the accuracy of predicting stock prices has greatly improved. The researchers evaluated the models using well-established measures like Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) to gauge their performance. The low values obtained for these indicators indicate that the models have demonstrated their effectiveness in accurately predicting the closing prices of stocks [3].

A wide range of machine learning approaches can be employed for predicting stock market trends. Nayak et al. compared different supervised machine learning techniques in their study. Bailings et al. utilized random forest (RF), AdaBoost, kernel factory, neural networks (NN), support vector machines (SVM), and k-nearest neighbors (KNN) to predict the direction of the stock market over a one-year period. Patel et al. discussed various machine learning models including Artificial Neural Networks (ANN), SVM, RF, and Naïve Bayes, and also made predictions on stock market indices using ANN, SVM, and RF. ANN is recognized as one of the most widely used models and offers different approaches, as highlighted by Vui et al. Several researchers have explored these ANN models, such as Bing et al. who employed Backpropagation Neural Networks (BPNN) to predict the Shanghai Stock Exchange Composite Index [2]. Wensheng et al. compared Nonlinear Independent Component Analysis (NLICA) and BPNN for the Asian stock market. Selvin et al. discussed approaches using Long Short-Term Memory (LSTM), Recurrent Neural Networks (RNN), and Convolutional Neural Network-Sliding Window (CNN-sliding window). Chen et al. focused on the use of the LSTM model. Nelson et al. compared LSTM with RF and Multilayer Perceptron (MLP), while G. et al. explored MLP, RNN, LSTM, and CNN approaches [7].

In their study, Bailings et al. utilized a range of machine learning algorithms, including Random Forest (RF), AdaBoost, kernel factory, neural networks (NN), support vector machines (SVM), and k-nearest neighbors (KNN), to predict the direction of the stock market over a specific time frame. Their research emphasized the importance of employing multiple models to capture the complexity of market dynamics accurately [11].

The existing body of literature on stock market prediction encompasses a wide range of machine learning techniques. Researchers have investigated and compared various models including Artificial Neural Networks (ANN), Random Forest (RF), Support Vector Machines (SVM), Naïve Bayes, and k-nearest neighbors (KNN). By combining Artificial Neural Network and Random Forest techniques, along with the increased computational power of artificial intelligence, notable improvements in accuracy have been achieved in predicting stock prices. The effectiveness of these models has been validated through the use of evaluation metrics like Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE). Moving forward, future research in this field is expected to continue exploring new machine learning approaches and refining existing models to further enhance the ability to forecast stock market trends in the ever-changing financial landscape.

Predicting the future price of stocks accurately can lead to significant profits. Over the years, various approaches have been explored to forecast stock trends. In this research, a novel framework for predicting stock prices is introduced, which incorporates two popular models: the Recurrent Neural Network (RNN) model, specifically the Long Short Term Memory (LSTM) model, and the Bi-Directional Long Short Term Memory (BI-LSTM) model. The simulation results indicate that by using these RNN models, specifically LSTM and BI-LSTM, with appropriate tuning of hyperparameters, the proposed framework achieves a high level of accuracy in forecasting future stock trends [5].

In our Seminar, we searched through articles published between 2011 and 2022 in the Scopus and Web of Science databases. Our goal was to examine various feature selection and extraction methods that have been successfully used in stock market analyses presented in these articles. We also investigated how combining feature analysis techniques with machine learning (ML) methods can improve performance and evaluated their effectiveness [1].

Additionally, we reviewed other surveys, stock market input and output data, and analyzed different factors related to stock market prediction. Our research found that certain techniques, such as correlation criteria, random forest, principal component analysis (PCA), and autoencoder, have been widely used for feature selection and extraction in stock market applications. These methods have shown promise in improving prediction accuracy across various studies [2].

Overall, our study suggests that the selected feature selection and extraction methods are effective in improving stock market prediction accuracy. These findings provide valuable insights into the current literature and can guide future research in this field [1].

Support Vector Machines (SVM) have gained popularity as a Machine Learning (ML) algorithm for predicting stock prices. However, many studies in algorithmic investments utilizing SVM have overlooked the issue of overfitting when dealing with high-noise and high-dimensional input datasets. To address this, our study introduces a novel ensemble classifier called GASVM, which combines Support Vector Machine with Genetic Algorithm (GA) for feature selection and SVM kernel parameter optimization in stock market prediction [8].

The Genetic Algorithm (GA) is employed in our study to simultaneously optimize various design factors of SVM. We conducted experiments using more than eleven years of stock data from the Ghana Stock Exchange (GSE), and the results were compelling. The outcomes demonstrated that our proposed model, GASVM, outperformed other traditional ML algorithms such as Decision Tree (DT), Random Forest (RF), and Neural Network (NN) in predicting stock price movements ten days in advance. GASVM achieved a superior prediction accuracy of 93.7%, surpassing RF (82.3%), DT (75.3%), and NN (80.1%) [9].

In summary, our research highlights the limitations of SVM in high-noise and high-dimensional datasets and proposes GASVM as a solution. The GASVM model, incorporating GA-based feature selection and SVM parameter optimization, significantly outperforms other classical ML algorithms in predicting stock price movements [12].

Sr No	**Publication Title**	**Publication Year**	**Methods**	**Gaps in publication work**
1	Stock Closing Price Prediction using Machine Learning Techniques	2019	ANN, Random Forest	Deep learning models could be developed.
2	Time series data analysis of stock price movement using machine learning techniques	2020	Time series Modeling	Adding more historical data And technical indicators.
3	An improved deep learning model for predicting stock market price time series	2020	Empirical wavelet transform decomposition , Weighted regularized extreme learning machine	To apply the model in other financial time prediction model
4	Deep Learning-Based Stock Price Prediction Using LSTM and Bi-Directional LSTM Model	2020	LSTM and BI Directional LSTM	Deep learning models could be developed.

5	Predicting stock market trends using machine learning and deep learning algorithms via continuous and binary data; a comparative analysis on the Tehran stock exchange	2020	Decision Tree, Random Forest	Using an ensembled Learning Model

Table. 1 Literary Survey

3. DETAILS OF DESIGN/ TECHNOLOGY

3.1 Block Diagram

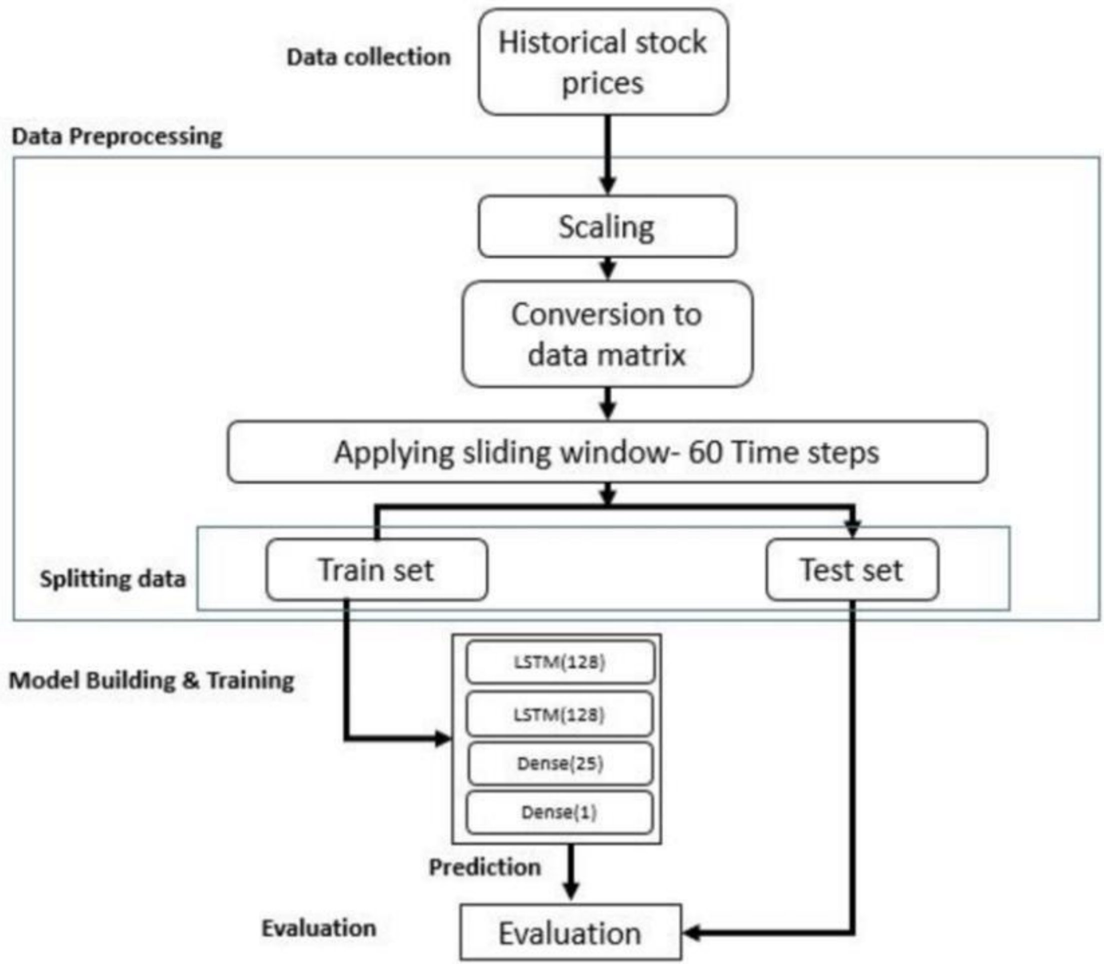

Figure. 3 Block diagram for LSTM model

3.2 Technologies used

3.2.1 LSTM

LSTM, or Long Short-Term Memory, is a type of neural network that specializes in analyzing sequences of data, such as text or time series. It addresses a challenge faced by traditional recurrent neural networks known as the vanishing gradient problem, enabling it to capture long-term relationships in the data.

At the core of an LSTM network is the LSTM cell, which consists of different gates that control the flow of information. These gates determine what information is retained, forgotten, or outputted at each step. The cell state acts as a memory, allowing important information to be preserved over longer sequences.

LSTMs have proven to be highly effective in various domains, including language translation, speech recognition, and even predicting stock market trends. Their ability to understand complex patterns and dependencies makes them well-suited for these tasks. By adjusting their parameters through a process called backpropagation, LSTMs continuously learn from the input data and improve their predictions or generate meaningful outputs.

The key strengths of LSTMs lie in their proficiency in handling sequential data, capturing long-term connections, and modeling intricate patterns. These qualities have contributed to the widespread use of LSTMs in deep learning applications, particularly when dealing with data that evolves over time.

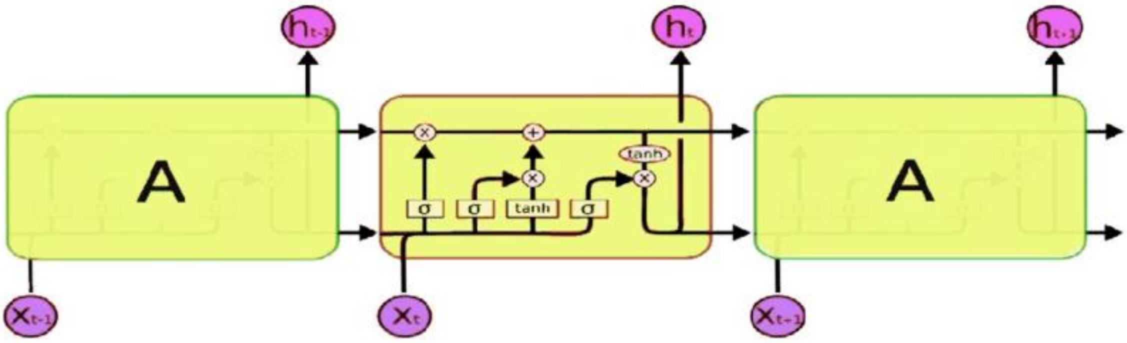

Figure. 4 Image of LSTM CELL

Advantages of LSTM

- One of the main strengths of LSTMs is their ability to capture long-term dependencies within sequential data. This means they can effectively retain and utilize important information over extended periods, allowing for accurate predictions in tasks that require understanding context over a lengthy sequence.

- LSTMs are also versatile when it comes to handling sequences of varying lengths. Regardless of the length of the input data, LSTMs can process and analyze it effectively, making them well-suited for tasks where the sequence length is inconsistent.

- Another advantage of LSTMs is their capability to retain crucial information within their memory cells over multiple time steps. This ensures that important contextual details are preserved, even in scenarios where there are gaps in the data. As a result, LSTMs can make informed decisions based on past information.

- Furthermore, LSTMs demonstrate robustness when dealing with noisy or incomplete data. They can filter out irrelevant information and focus on the most essential features, making them reliable in real-world scenarios where data may contain errors or inconsistencies.

- Useful in the field of natural language processing, speech recognition, time series forecasting, and many other applications. This adaptability allows LSTMs to tackle a wide range of sequence modeling tasks effectively.

- Lastly, LSTMs can be trained efficiently using a technique called backpropagation through time. This enables them to learn from the data and adjust their parameters effectively during the training process.

Disadvantages of LSTM

The following are the disadvantages of LSTM

- Computational Complexity: LSTMs are more computationally intensive compared to simpler neural network architectures. The presence of multiple gates and memory cells increases the computational requirements during training and inference, making LSTMs slower to train and run.

- Memory Requirements: LSTMs require more memory to store the parameters and state information of the network. This increased memory footprint can be a challenge when working with large-scale datasets or deploying models on devices with limited memory capacity.

- Overfitting: LSTMs can be susceptible to overfitting, especially when the available dataset is small. The complex nature of LSTMs and their ability to capture long-term dependencies can result in models that over-learn specific patterns in the training data, leading to poor generalization on unseen data.

- Difficulty in Interpretability: LSTMs are considered black-box models, making it difficult to interpret and understand how the model makes its predictions. This lack of interpretability can be problematic in domains where transparency and explain ability are crucial.

- Tuning Hyperparameters: LSTMs have several hyperparameters that need to be carefully tuned to achieve optimal performance. Finding the right combination of hyperparameters requires expertise, experimentation, and computational resources, which can be time-consuming and challenging.

- Limited Short-Term Memory: While LSTMs excel at capturing long-term dependencies, they may struggle with capturing short-term patterns. The network's focus on retaining long- term memory can sometimes hinder its ability to respond quickly to recent inputs or events.

Applications of LSTM are as follows: -

- Natural Language Processing: LSTMs are used in language-related tasks such as sentiment analysis, text generation, machine translation, and speech recognition. They excel in understanding the context and dependencies within text data, leading to improved performance in language processing applications.

- Time Series Forecasting: LSTMs are effective in modeling and predicting time-dependent data, including stock prices, weather patterns, energy consumption, and sales forecasting. Their ability to capture long-term dependencies and handle varying sequence lengths makes them well-suited for accurate time series analysis.

- Speech Recognition: LSTMs play a vital role in accurate transcription and voice command processing. They learn intricate patterns in spoken language and effectively model the temporal characteristics of audio data, enabling robust speech recognition systems.

- Gesture Recognition: LSTMs are applied to recognize and interpret hand gestures, body movements, and sign language. By capturing the sequential nature of gesture data, LSTMs improve the accuracy and understanding of human motions, facilitating various applications in human-computer interaction and robotics.

- Music Generation: LSTMs have been utilized to generate new musical compositions based on learned patterns from existing music. They capture the underlying structures and dependencies in music, allowing for the creation of compositions that resemble the style of the training data.

Figure. 5 Stock Analysis

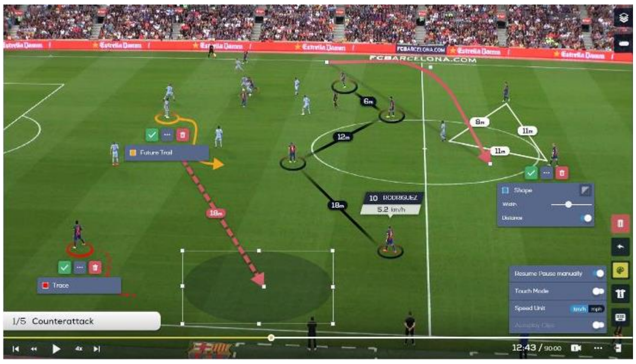

Figure. 6 Video Analysis

3.2.2 Numpy

Numpy is a powerful library designed specifically for scientific computing within the Python programming language. It provides extensive support for working with large arrays and matrices, and it offers a wide range of mathematical functions to perform efficient operations on these arrays.

The main building block of NumPy is its ndarray, or n-dimensional array, which serves as a versatile and efficient container for storing and manipulating large datasets. This data structure allows you to perform various operations on the elements of the array, such as performing calculations on individual values, extracting specific subsets of data using slicing techniques, reshaping the dimensions of the array, and even broadcasting operations across arrays of different sizes and shapes.

3.2.3Sklearn

Scikit-learn, also known as sklearn, is a popular and widely used machine learning library designed for the Python programming language. It offers a comprehensive suite of tools that cover various aspects of machine learning, including tasks like classification, regression, clustering, dimensionality reduction, and model selection.

One of the key advantages of sklearn is its user-friendly and consistent interface. It provides a well-defined and easy-to-use API that simplifies the process of implementing machine learning algorithms. Sklearn offers a wide range of algorithms and techniques that are implemented in a standardized way, making it convenient to experiment with different models and compare their performance.

3.2.4 MatplotLib

Matplotlib is widely used in the scientific and data analysis communities due to its versatility and extensive capabilities. It integrates well with other popular Python libraries, such as NumPy and pandas, which further enhances its functionality for data manipulation and analysis.

Overall, Matplotlib empowers users to effectively visualize their data by providing a comprehensive set of tools and customization options. Whether it's creating simple line plots or complex multi-panel figures, Matplotlib offers the flexibility and features needed to present data in a visually appealing and informative manner.

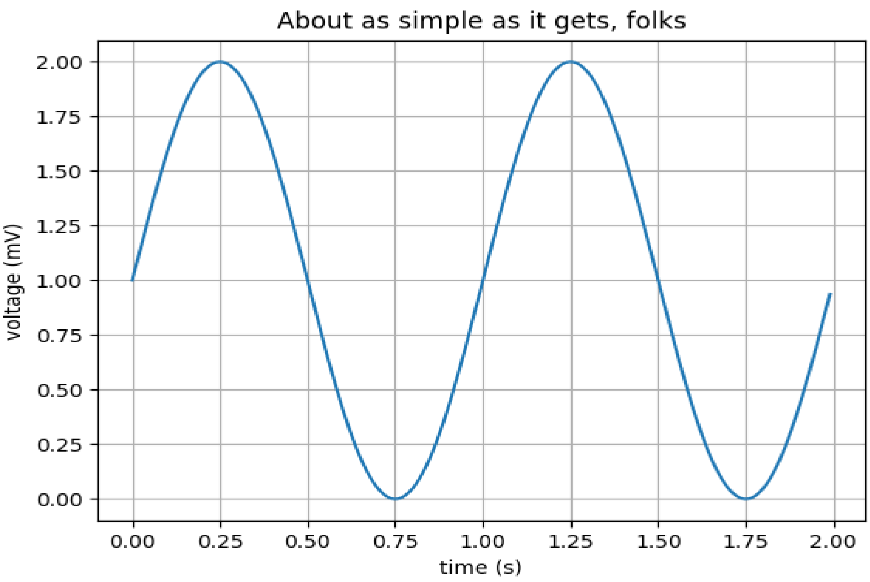

Figure. 7 Sample graph using Matplotlib

4. ANALYSIS AND EXPERIMENTAL WORK

4.1 Steps followed for designing the model are as follows: -

4.1.1 Data Collection

Data collection is the organized and methodical process of gathering and documenting information or data for research or analysis. It involves identifying appropriate sources and using various techniques like surveys, interviews, observations, or data mining to collect data. The collected data can be either numerical or descriptive, depending on the purpose of the research. Data collection is essential for gaining valuable insights, making informed decisions, and drawing meaningful conclusions. It requires careful planning, ethical considerations, and ensuring the accuracy and quality of the collected data.

For our experiment we had collected the data from Kaggle For the stock price prediction of the Tata Shares. For that purpose, we had used two data set one for the training purpose and the other dataset for the testing purpose.

The training data set contains 2035 rows and 8 columns.

	Date	Open	High	Low	Last	Close	Total Trade Quantity	Turnover (Lacs)
0	2018-09-28	234.05	235.95	230.20	233.50	233.75	3069914	7162.35
1	2018-09-27	234.55	236.80	231.10	233.80	233.25	5082859	11859.95
2	2018-09-26	240.00	240.00	232.50	235.00	234.25	2240909	5248.60
3	2018-09-25	233.30	236.75	232.00	236.25	236.10	2349368	5503.90
4	2018-09-24	233.55	239.20	230.75	234.00	233.30	3423509	7999.55
5	2018-09-21	235.00	237.00	227.95	233.75	234.60	5395319	12589.59
6	2018-09-19	235.95	237.20	233.45	234.60	234.90	1362058	3202.78
7	2018-09-18	237.90	239.25	233.50	235.50	235.05	2614794	6163.70
8	2018-09-17	233.15	238.00	230.25	236.40	236.60	3170894	7445.41
9	2018-09-14	223.45	236.70	223.30	234.00	233.95	6377909	14784.50

Figure. 8 Sample entries of training Dataset

The testing_dataset contains 200 rows and 8 columns

	Date	Open	High	Low	Last	Close	Total Trade Quantity	Turnover (Lacs)
0	2018-10-24	220.10	221.25	217.05	219.55	219.80	2171956	4771.34
1	2018-10-23	221.10	222.20	214.75	219.55	218.30	1416279	3092.15
2	2018-10-22	229.45	231.60	222.00	223.05	223.25	3529711	8028.37
3	2018-10-19	230.30	232.70	225.50	227.75	227.20	1527904	3490.78
4	2018-10-17	237.70	240.80	229.45	231.30	231.10	2945914	6961.65
5	2018-10-16	237.10	237.70	233.05	234.40	235.45	1723113	4052.25
6	2018-10-15	229.70	237.00	226.80	234.80	234.90	1224339	2845.68
7	2018-10-12	226.25	232.35	225.50	228.70	229.10	1165527	2675.91
8	2018-10-11	215.00	229.70	215.00	225.60	224.60	1293881	2890.85
9	2018-10-10	215.00	229.65	215.00	228.25	228.40	2919278	6557.95
10	2018-10-09	215.50	219.15	209.60	215.00	216.50	1844462	3940.70

Figure. 9 Sample entries of Testing Dataset

4.1.2 Data preprocessing

Data preprocessing is the initial step in analyzing data, where raw data is prepared and modified to be suitable for further analysis and modeling. This process involves various important stages that aim to ensure the data's quality and usability. These stages include cleaning the data by addressing missing values, outliers, and inconsistencies, integrating data from different sources into a consistent format, transforming the data by normalizing or scaling it to a standardized range, and reducing the complexity and size of the dataset through techniques like dimensionality reduction. The ultimate objective of data preprocessing is to enhance data quality, rectify errors, and make the data well-prepared for meaningful analysis and modeling purposes.

To Normalize the data, we had used MinMaxscalar And After that we had used fit transform function to transform the data.

```
array([[0.6202352 ],
       [0.62226277],
       [0.64436334],
       ...,
       [0.16504461],
       [0.15896188],
       [0.16626115]])
```

Figure. 10 Normalized training set that contains value between 0 and 1

4.1.3 Incorporating TimeStamp

As to provide the input for the LSTM model we need to provide a 3d array. But first it requires timestamp. Therefore, 60-time stamps of the data are been created.

4.1.4 Model is created and trained for the price prediction

In the configuration of the LSTM layer, we set the number of units to 50, which determines the dimensionality of the output space. By setting return_sequences=True, we allow for the stacking of multiple LSTM layers, enabling the subsequent LSTM layer to receive a three-dimensional sequence input. The input_shape parameter specifies the shape of the training dataset, providing information about the input dimensions.

The Dropout layer is introduced with a dropout rate of 0.2. This means that during training, 20% of the neural network's layers will be randomly dropped or deactivated, helping to prevent overfitting and enhancing the model's ability to generalize well to new data.

Following the LSTM and Dropout layers, we add a Dense layer with a single output unit. This layer produces the final output of the model, providing the desired predictions or classifications.

For the compilation of the model, we employ the Adam optimizer, which is an efficient algorithm for optimizing neural networks. The mean_squared_error loss function is selected, which calculates the average squared difference between the predicted and actual values, serving as a measure of how well the model performs.

To train the model, we fit it to the training data for 100 epochs. Each epoch represents a complete iteration through the entire training dataset, allowing the model to learn from the data gradually. A batch size of 32 is used, meaning that the model's parameters are updated after processing 32 samples at a time. This batch-wise processing helps optimize memory usage and enhances the overall efficiency of the training process.

4.1.5 Testing the model

The model is tested against the test_dataset which is unknown to the model and it uses the relations formed during the testing Phase and it predicts The price of cost and it will be stored in the variable predicted variable cost.

4.1.6 Analyzing the result

The analysis of the result can be done through the help of various graph available in the matplotlib library. We here uses a line graph to plot the time vs Price of stock for the tata share. The black color lines indicated the actual cost while the green color lines indicated the predicted Stock price. It has been observed that our model at the start had show a bigger variation in estimation cost but as the time increases it almost start to overlap with the actual price line. And we were able to easily compare the variation.

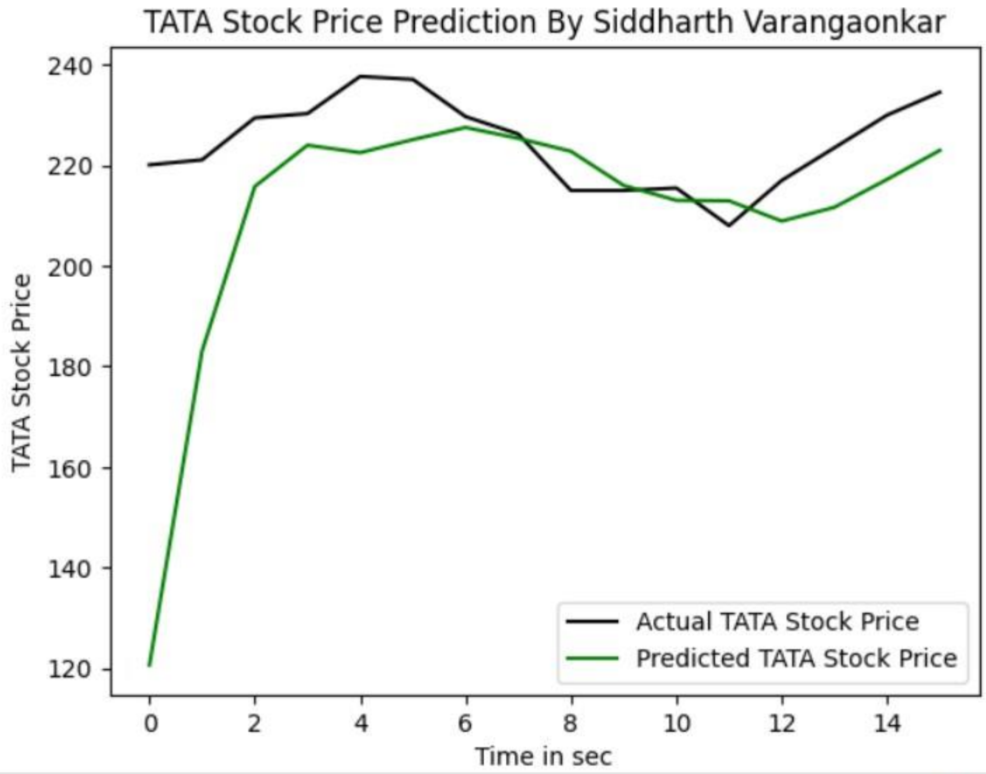

Figure. 11 **Comparison of Actual and estimated Cost**

4.2 Future Scope

It can be to Adding more historical data and technical indicators. To build a model which can be created as a generalized model for other time series data models like for predicting the score in the Football, Cricket matches. To use an ensembled learning model that may use AAN, Random Forest, LSTM in such a way that for each model 33.3% data is allocated and at the end all the data Are Combined And as a result a more precise and accurate model is being prepared.

5 CONCLUSIONS

Thus, while studying various algorithm for predicting the price of stock, we came to know that we require algorithm which are efficient as well as self-learning models that gives us the most accurate prediction price by taking into consideration factors such as time series, start price, close price etc. We need to create models which decreases the risk factor replated to field of stock price prediction and should increase the assurance, reliability of the model and improving the quality of the result with each new set of data.

Hence, Stock market analysis using ML is a very interesting field of study and we should try to explore it and find the hidden patterns in the Algorithm and models Designed.

References

[1] D. C. V. A. T. A. K. M. Vijh, "Stock Closing Price Prediction using Machine Learning Techniques," *Procedia Computer Science,* vol. 167, pp. 599-606, 2020.

[2] S. K. K. Y. a. J. L. Mokhtari, "Effectiveness of artificial intelligence in stock market prediction based on machine learning," *arXiv preprint,* 2021.

[3] S. S. K. M. K. &. A. A. A. I. R. Parray, "Time series data analysis of stock price movement using machine learning techniques," *Soft Computing,* vol. 24, p. 16509–16517, 2020.

[4] M. a. N. P. a. J. H. a. S. S. a. M. A. Nabipour, "Predicting Stock Market Trends Using MachineLearning and Deep Learning Algorithms Via Continuous and Binary Data; a Comparative Analysis," *IEEE Access,* vol. 8, pp. 150199-150212, 2020.

[5] M. A. a. M. M. M. S. a. A. A. G. I. Sunny, "Deep Learning-Based Stock Price Prediction Using LSTM and Bi-Directional LSTM Model," *ntelligent and Leading Emerging Sciences Conference,* pp. 87-92, 2020.

[6] O. M. T. A. S. V. S. R. B. J. T. S. Vachhani H, "Machine learning based stock market analysis," *InInnovative Data Communication Technologies and Application: ICIDCA,* pp. 12-26, 2020.

[7] K. M. R. M. Hanauer MX, " Boosting agnostic fundamental analysis: Using machine learning to identify mispricing in European stock markets.," *Finance Research Letters,* p. 1;48:102856, 2022.

[8] M. Javed Awan, "Social media and stock market prediction: a big data approach." MJ Awan, M. Shafry, H. Nobanee, A. Munawar, A. Yasin et al.," Social media and stock market prediction: a big data approach," *Computers, Materials & Continua,* vol. 67.2, pp. 2569-2583, 2021.

[9] M. H. E. E.-S. A. a. E.-B. Sharaf, "StockPred: a framework for stock Price prediction," *Multimedia Tools and Applications,* vol. 80, pp. 17923-17954, 2021.

[10] S. S. K. P. a. M. Rahul, "Analysis of various approaches for stock market prediction," *Journal of Statistics and Management Systems,* vol. 23.2, pp. 285-293, 2020.

[11] H. B. M. a. P. Htun, "Survey of feature selection and extraction techniques for stock market prediction," *Financial Innovation,* vol. 9.1, p. 26, 2023.

[12] B. H. S. S. A. M. C. K. P. M. S. a. D. Singh, "ML-Based Interconnected Affecting Factors with Supporting Matrices for Assessment of Risk in Stock Market," *Wireless Communications and Mobile Computing,* 2022.

[13] A. S. Y. a. K. Jain, "Financial supervision and management system using Ml algorithm. Solid State Technology," vol. 63.6, pp. 18974-18982, 2020.

[14] S. G. J. a. G. Saha, " A survey of the application of graph-based approaches in stock market analysis and prediction," *International Journal of Data Science and Analytics,* vol. 14.1, pp. 1-15, 2022.

[15] O. a. P.-Q. Bustos, "Stock market movement forecast: A systematic review," *Expert Systems with Applications,* vol. 156, p. 113464, 2020.

7. BASE PAPER 1st PAGE

Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 167 (2020) 2091–2100

Procedia Computer Science

www.elsevier.com/locate/procedia

International Conference on Computational Intelligence and Data Science (ICCIDS 2019)

Optimizing LSTM for time series prediction in Indian stock market

Anita Yadav[a,*], C K Jha[a], Aditi Sharan[b]

[a]Department of Computer Science, AIM & ACT, Banasthali Vidyapith, Rajasthan- 304022, India
[b]School of Computer & Systems Sciences, Jawaharlal Nehru University, New Delhi - 110067, India

Abstract

Long Short Term Memory (LSTM) is among the most popular deep learning models used today. It is also being applied to time series prediction which is a particularly hard problem to solve due to the presence of long term trend, seasonal and cyclical fluctuations and random noise. The performance of LSTM is highly dependent on choice of several hyper-parameters which need to be chosen very carefully, in order to get good results. Being a relatively new model, there are no established guidelines for configuring LSTM. In this paper this research gap was addressed. A dataset was created from the Indian stock market and an LSTM model was developed for it. It was then optimized by comparing stateless and stateful models and by tuning for the number of hidden layers.

© 2020 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)
Peer-review under responsibility of the scientific committee of the International Conference on Computational Intelligence and Data Science (ICCIDS 2019).

Keywords: LSTM; Hyperparameters; Stateful Stateless; Hidden layers; Time series prediction

1. Introduction

Stock price prediction is the process by which future stock prices are forecast on the basis of past prices. Stock price prediction is useful for investors to increase the profits from stock trading. There are two traditional approaches to prediction: technical and fundamental analysis. Technical analysis tries to identify some patterns from historical data while fundamental analysis focuses on overall economy, the company's financial condition and its management [1]. Stock price prediction is a very challenging task as it is highly volatile, non-linear and dynamic and is impacted by various factors like political conditions, economic situation, trend, seasonality, investor psychology etc.

* Corresponding author. Tel.: +91-9899782208.
E-mail address: anitayadav_07@hotmail.com