International Journal of Engineering Applied Sciences and Technology, 2022
Vol. 7, Issue 12, ISSN No. 2455-2143, Pages 86-95
Published Online April 2023 in IJEAST (http://www.ijeast.com)

EXPLORING THE PREDICTIVE POWER OF FINANCIAL DATA

Sahil Tushamar, Samapan Kanojia, Bineet Kumar and Ashutosh Vashishth
Department of Computer Science and Engineering,
Bharati Vidyapeeth College of Engineering
Paschim Vihar East, New Delhi, India

Dr.Aarti
(Assistant Professor)
Bharati Vidyapeeth College of Engineering
Paschim Vihar East, New Delhi, India

Abstract: **Learning to forecast stock prices and trends is more common than ever in the era of big data. We created a machine learning-based model to forecast stock price changes using ten years' worth of stock market data from Google. The proposed comprehensive solution includes stock trade preprocessing datasets, using an unique deep learning-based stock trend forecasting model in conjunction with a variety of feature engineering methodologies. We perform in--depth evaluations of commonly used machine learning models. The technique is very accurate overall at predicting stock market trends. The thorough design and evaluation of prediction period durations, engineering functions, and data pretreatment techniques contribute to the finance and technology academic researchers. Keywords: forecasting, machine learning, stock market trends,**

I. INTRODUCTION

In the stock market, there is an environment that is both unexpected and turbulent where investors constantly seek ways to maximize their profits through informed decision-making. As a result, financial research has begun to focus more on stock market forecasting. Numerous methods and techniques are developed to help investors anticipate future trends and make informed investment decisions.

Machine learning, which uses mathematical and statistical models to evaluate massive amounts of data and generate predictions based on patterns and trends, is one of the most promising methods for stock market forecasting. In this research paper, we will explore the performance of five popular machine learning models for stock market prediction: Support Vector Machines (SVM), Neural Networks (NN)**,** Recurrent Neural Networks (RNN), ARMA-LSTM hybrid model, and Deep Reinforcement Learning (DRL) model.

SVM is a machine learning model which has gained significant popularity due to its ability to handle high-dimensional data and its suitability for classification and regression problems. In order to make predictions about fresh data points, SVM first determines the hyper plane that maximally separates the data points in a particular dataset.

Deep learning models NN and RNN have been very popular lately thanks to their ability to recognize intricate patterns in data. NN models are particularly well-suited for supervised learning problems, while RNN models are better suited for time-series prediction problems. Since they can find long-term links in the data, RNN models have been demonstrated to be effective at forecasting stock prices.

LSTM model from deep learning and the conventional Autoregressive Moving Average (ARMA) model are combined in the ARMA-LSTM hybrid model. The LSTM model is a deep learning model that can capture long-term interdependence in data, whereas the ARMA model is a statistical model that is frequently used for time-series analysis. Combining these two models can improve prediction accuracy compared to using either model alone.

Finally, the DRL model is a new and emerging model that combines reinforcement learning and deep learning. This model learns from experience by taking actions and observing the resulting rewards or penalties. The DRL model has shown promising results in stock market prediction, as it can adapt to changing market conditions and learn optimal strategies.

This research paper aims to provide an in-depth analysis of these five models for a stock market forecast. We will review the theoretical foundations of each model, as well as their strengths and weaknesses. We will also compare the performance of these models using

real-world stock market data and evaluate their prediction accuracy, efficiency, and robustness. This research can assist investors and financial experts in the stock market by offering insights into the efficacy of various models for stock market forecasting.

International Journal of Engineering Applied Sciences and Technology, 2022
Vol. 7, Issue 12, ISSN No. 2455-2143, Pages 86-95
Published Online April 2023 in IJEAST (http://www.ijeast.com)

II. LITERATURE SURVEY

2.1 **Stock Market Prediction Using Machine Learning**

V.K.S Reddy, a student at the S.N.I.S.T in Hyderabad, India, conducted the study.

To predict the fluctuations of stock indices, the study proposed combining data gathered from several international financial markets in machine learning algorithms. A sizable dataset of values gathered from numerous international financial marketplaces is used by the SVM algorithm.

Also, SVMs do not pose over fitting problems. Numerical results show high efficiency. This model produces higher profits compared to our chosen benchmark.

Technique Used	• Support Vector Machine(SVM) with RBF kernel
Comments	• SVM algorithm not suitable for large data sets • When the data set has additional target classes that overlap, it does not function well. • The model relies on historical data for training and prediction • Lack of transparency • The model may be susceptible to overfitting

2.2 **Forecasting the Stock Market Index Using Artificial Intelligence Techniques**

Lufuno Ronald Marwala's research findings were submitted to the University of the Witwatersrand in Johannesburg. To forecast future stock market price indices, three artificial intelligence algorithms have been used. In this case, the JSE's aggregate price index was predicted using artificial NN, SVM, and a neuro-fuzzy system. This project's methodology involved estimating All Share Index future prices. Predicting the direction of future prices has been the topic of substantial studies utilising all three strategies.

The amount of the price increase or decrease cannot be predicted by direction alone. The outcomes demonstrate that the index's price can be predicted by all three approaches. Compared to forecasting direction, price prediction is far more difficult. If it can be correctly forecast, it can offer additional insight into how stock values will move in the future.

This study's other goal was to demonstrate the predictability of previous stock prices. They learned that data from historical stock prices can be analysed to predict future stock values.

Technique Used	• Artificial intelligence methods like SVM, NN, and neuro-fuzzy systems
Comments	• Random walk method outperformed all the other techniques. • Focus on a specific stock market index • Data availability: • The study may have limitations in selecting the most appropriate technique for the specific dataset and research question.

International Journal of Engineering Applied Sciences and Technology, 2022
Vol. 7, Issue 12, ISSN No. 2455-2143, Pages 86-95
Published Online April 2023 in IJEAST (http://www.ijeast.com)

2.3 **Automated Stock Price Prediction Using Machine Learning**

The study was conducted by Mariam Moukalled and Wassim El-Hajj Mohamad Jaber at the American University of Beirut.

Data for these models was gathered from two sources: (i) Reuters historical stock market data and (ii) news sentiment released for certain stocks. During the course of ten years, this information was gathered for four different strains. While calculating the technical characteristics and adding a sentiment, three options were taken into account. These three possibilities were also computed and utilised as raw data for the model. For prediction, the Al framework uses DNN, RNN, SVR, and SVM. Testing the suggested predictive model on the stocks of APPL, AMZN, GOOGL, and FB using data gathered from 01/01/2008 to 31/12/2017 produced an accuracy of 82.91%.

Technique Used	● Machine Learning The performance of (SVM), (SVR), (FFNN), and (RNN) in forecasting the direction of the day's closing price in relation to the day's close price was compared.
Comments	● Failed to create a risk management system to monitor the accuracy of predictions-based profitability. ● The data were not grouped according to various timeframes. ● Finally, it might aim to improve price prediction accuracy.

2.4 **Stock Price Correlation Coefficient Prediction with ARIMA LSTM Hybrid Model.**

Hyeong Kyu Choi, B, did the research work. A Student Dept. of Business Administration Korea University Seoul, Korea.

LSTM recurrent neural networks were used to forecast nonlinear trends after employing the ARIMA-LSTM hybrid model to eliminate linearity in the ARIMA modelling step. According to test results, the ARIMA-LSTM hybrid model performs significantly better than similar financial models. The effectiveness of the model is verified over various time periods and asset combinations using several measures, including MSE, RMSE, and MAE. The value was almost half that of the constant correlation model, which was the best of the four financial models in the experiment. We can assume there is. ARIMA LSTM models are, therefore, of great importance as correlation coefficient predictors for portfolio optimization.

Technique Used	● ARIMA-LSTM hybrid model
Comments	● The testing results demonstrated that this model outperforms other comparable financial models by a significant margin. Model performance was verified using a variety of metrics, including the MSE, RMSE, and MAE, for both various periods of time and different asset pairings. ● The Constant Correlation model, which in our experiment outperformed the other four financial models, saw values that were almost halved. ● Prior to the year 2008, the experiment was not conducted. The model could therefore be vulnerable to certain financial circumstances that did not exist between 2008 and 2017.

2.5 **A Deep Reinforcement Learning Library for Automated Stock Trading in Quantitative Finance**

Researchers Bowen Xiao, Christina Dan Wang, Qian Chen, Runjia Zhang, Liuqing Yang, Xiao-Yang Liu, and Hongyang Yang carried out the study.

The FinRL library was presented in this article. For the aim of teaching and demonstrating automated stock trading, this DRL library was created especially for that purpose. Scalability, a complex market environment, and all-inclusive tools for performance measurement, especially for quant investors and strategy developers, are the defining features of FinRL. Market simulators, trading agent learning

International Journal of Engineering Applied Sciences and Technology, 2022
Vol. 7, Issue 12, ISSN No. 2455-2143, Pages 86-95
Published Online April 2023 in IJEAST (http://www.ijeast.com)

techniques, and successful methods can all be easily customised. FinRL follows a training-validation-testing flow for trading strategy design and provides automated back testing and benchmarking. Strategies that are repeatable and profitable in various situations utilising FinRL: single stock trading, multiple stock trading, and an internal method for stock information penetration. Implementing effective DRL-driven strategies is now simple, quick, and pleasant thanks to the FinRL library.

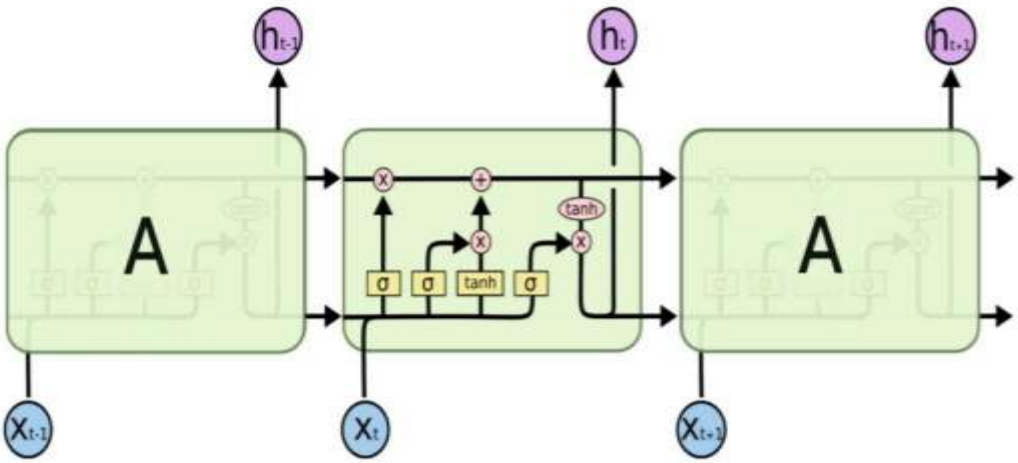
Technique Used	• Deep reinforcement learning (DRL)
Comments	• Unlike ML regression/classification models that forecast the likelihood of future outcomes, DRL uses a reward function to maximise future rewards. • Lack of interpretability of the DRL algorithms • Challenge of handling high-dimensional state variables due to the large volume of financial data • Difficulty of incorporating transaction costs and market impact into the reward function

III. LSTM

Recurrent neural networks (RNNs) with Long Short-Term Memory (LSTM) can deal with long-term dependencies and vanishing gradient issues that can occur with conventional RNNs. The first LSTMs were presented by Hochreiter and Schmidhuber in 1997 and have since become a well-liked tool in speech recognition, natural language processing, and time-series forecasting.

LSTM is composed of three gates: an input gate, a forget gate, and an output gate..Although the forget gate controls the flow of data that needs to be deleted from the previous time step, the input gate controls the flow of new data into the memory cell. The output gate controls the cell's output based on the input that is currently being received and the condition of the internal memory.

The LSTM solves the issue by adding extra memory cells and gates that enable the network to choose forget or remember data from earlier time steps, enhancing its capacity to detect persistent dependencies in data.



Four interconnected layers are found in the repeating module of an LSTM.

LSTM Step by Step Walkthrough

(Long Short-Term Memory) is a recurrent neural network designed to handle the vanishing gradient problem and effectively capture long-term dependencies in sequence data. Here is a step-by-step walkthrough of how an LSTM works:

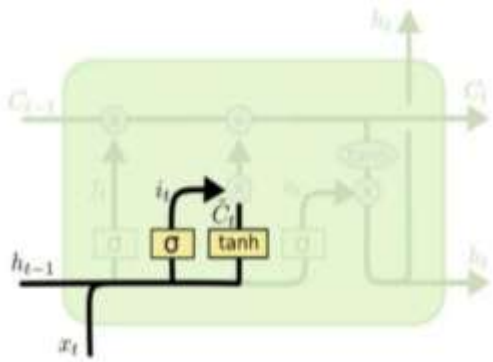
1. Input gate: The initial stage is deciding which input

International Journal of Engineering Applied Sciences and Technology, 2022
Vol. 7, Issue 12, ISSN No. 2455-2143, Pages 86-95
Published Online April 2023 in IJEAST (http://www.ijeast.com)

sequence data should be saved in the state of the LSTM cell. The associated input element is eliminated if the gate output is close to 0, and retained if it is close to 1.

The choice of fresh information to be stored in the cell state is made by the input gate. It starts with the conjunction of the previous hidden state, h(t-1), and the current input, x(t), and runs it through two different layers: a sigmoid layer to create an input gate activation vector, i(t), and a tanh layer to create a candidate cell state vector, c. (t).

The input gate and candidate cell state layers' respective weight matrices are Wi and Wc, whereas the bias vectors are bi and bc.



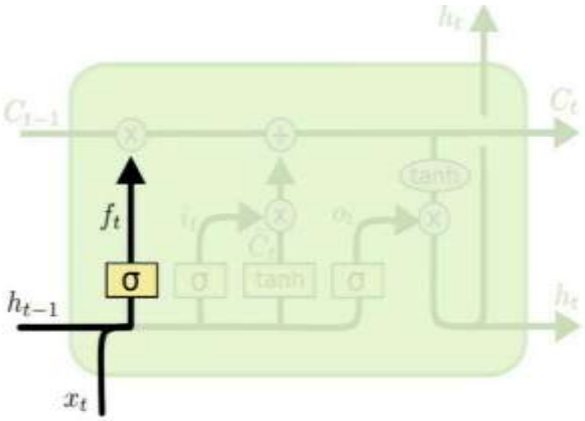
$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] \; + \; b_i\right)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] \; + \; b_C)$$

2. Forget gate: Next, the LSTM cell decides which information to forget from the previous state. A forget gate is used to do this; it accepts as inputs the previous hidden state and the current input vector and outputs a value between 0 and 1 for each component of the previous state. The matching state element is ignored if the gate output is close to 0, and maintained if it is close to 1.

The forget gate makes the choice of what data to remove from the cell state. It concatenates the current input, x(t), with the prior hidden state, h(t-1), and then applies a sigmoid function to create a forget gate activation vector, f(t).

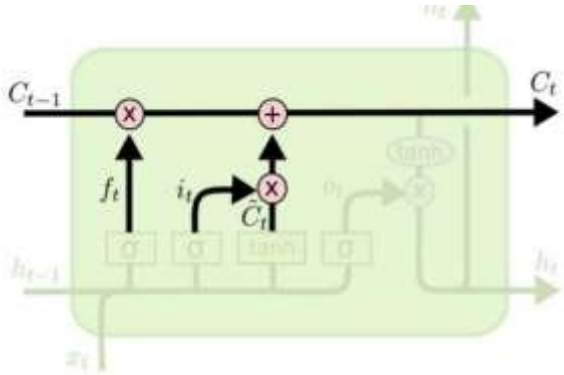
Wf is the forget gate's weight matrix, while bf is its bias vector. The forget gate activation vector values are kept within the range of 0 and 1 thanks to the sigmoid function.



$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] \; + \; b_f\right)$$

3. Update gate: The LSTM cell then determines how much new information from the input should be added to the current state.
4. An update gate is employed in this process. Its inputs are the current input vector and the previous concealed state, and its output is a new candidate state vector.The amount of the candidate state vector that should be eliminated or added to the existing state is decided. By merging the prior cell state, C(t-1), the forget gate activation vector, f(t), and the candidate cell state vector, c, the cell state, C(t), is updated (t).

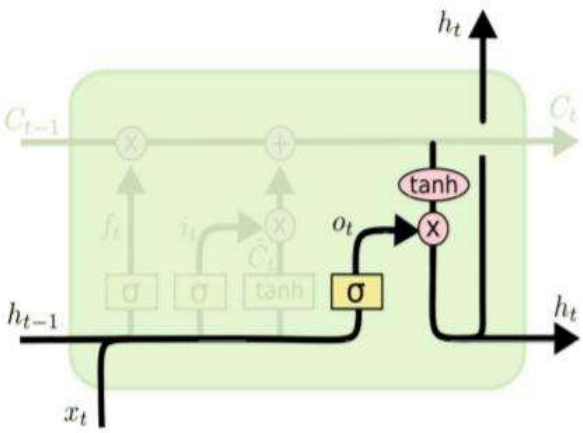
International Journal of Engineering Applied Sciences and Technology, 2022
Vol. 7, Issue 12, ISSN No. 2455-2143, Pages 86-95
Published Online April 2023 in IJEAST (http://www.ijeast.com)



$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

5. Output gate: Finally, the LSTM cell decides what output to produce based on the updated state. This is done using an output gate, which takes the updated state and the current input vector as inputs and outputs a value between O and 1 for each element of the updated state.

The output gate makes the choice of what data to output from the cell state. It concatenates the updated cell state, C(t), the current input, x(t), and the prior hidden state, h(t-1), as input and runs it through a sigmoid layer to create an output gate activation vector, o(t). It then creates a new hidden state, h(t), by passing the modified cell state through a tanh layer .
Here, the output gate layer's weight matrix (Wo) and bias vector (bo) are both present.



$$o_t = \sigma\left(W_o\left[h_{t-1}, x_t\right] + b_o\right)$$
$$h_t = o_t * \tanh\left(C_t\right)$$

6. Hidden state: The final output of the LSTM cell, which is transferred on to the following time step or network layer, is its hidden state.
7. Memory cell: The forget gate, input gate, and update gate are combined to update it, which is in charge of storing and transmitting information between timesteps.

Overall, the LSTM architecture allows for the selective retention and updating of information in the memory cell, which makes it effective for modeling long-term dependencies in sequential data.

2.6 Algorithm
This section details the algorithms that are used:

International Journal of Engineering Applied Sciences and Technology, 2022
Vol. 7, Issue 12, ISSN No. 2455-2143, Pages 86-95
Published Online April 2023 in IJEAST (http://www.ijeast.com)

Algorithm 1: LMS

Input:
x : input vector
d: desired vector
μ: learning rate
N: filter order
Output:
y: filter response
e: filter error
begin
  $M = size(x)$;
  $x_n(0) = w_n(0) = [0\ 0\ \ldots\ 0]^T$;
  while $n < M$ **do**
    $x_{n+1} = [x(n);\ x_n(1:N)]$;
    $y(n) = w_n^H * x_n$;
    $e(n) = d(n) - y(n)$;
    $w_{n+1} = w_n + 2\mu e(n) x_n$;
  end
end

Algorithm 2: LMSPred

Input:
x: input vector
l: quantity of days to predict
μ: learning rate
N: filter order
Output:
y: filter response
begin
  $M = size(x_d)$;
  $x_n(0) = w_n(0) = [0\ 0\ \ldots\ 0]$;
  $x_d = [0\ 0\ ..\ 0\ x]$;
  while $n < M$ **do**
    $x_{n+1} = [x_d(n);\ x_n(1:N)]$;
    $y(n) = w_n^H * x_n$;
    if $n > M - l$ **then**
      $e = 0$;
    else
      $e(n) = d(n) - y(n)$;
    end
    $w_{n+1} = w_n + 2\mu e(n) x_n$;
  end
end

International Journal of Engineering Applied Sciences and Technology, 2022
Vol. 7, Issue 12, ISSN No. 2455-2143, Pages 86-95
Published Online April 2023 in IJEAST (http://www.ijeast.com)

IV. RESULTS

4.1 The dataset

This section describes the data that was obtained from open sources and the most recent data file created. Since there are many different types of stock market data, we first analysed comparable publications from the survey of economic research papers to identify the best methods for gathering data. We recommended using a dataset data structure after data collection. Following, we provide a detailed description of the dataset, including information on its data structure and the tables that each data category contains.

Data Structure:

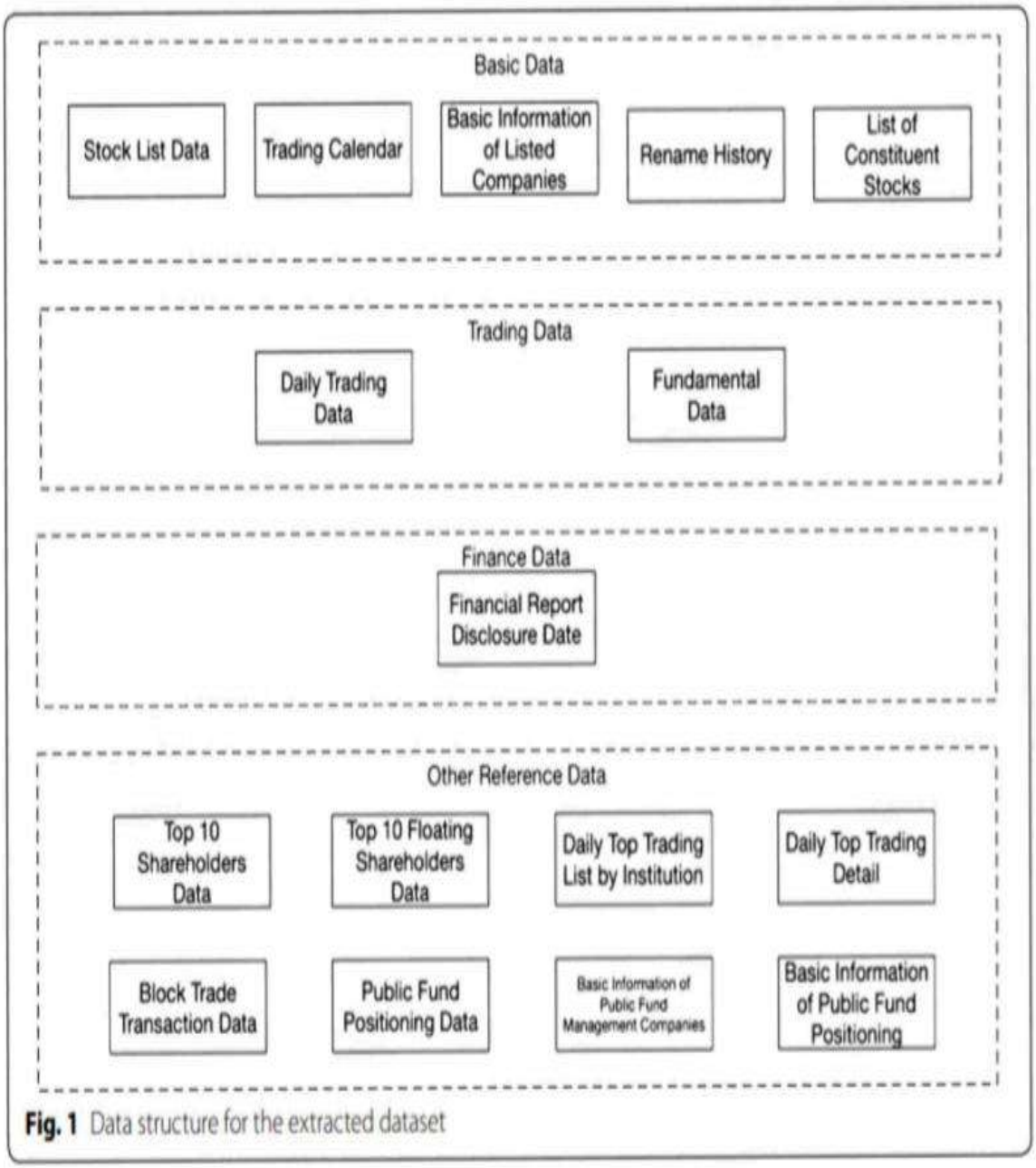
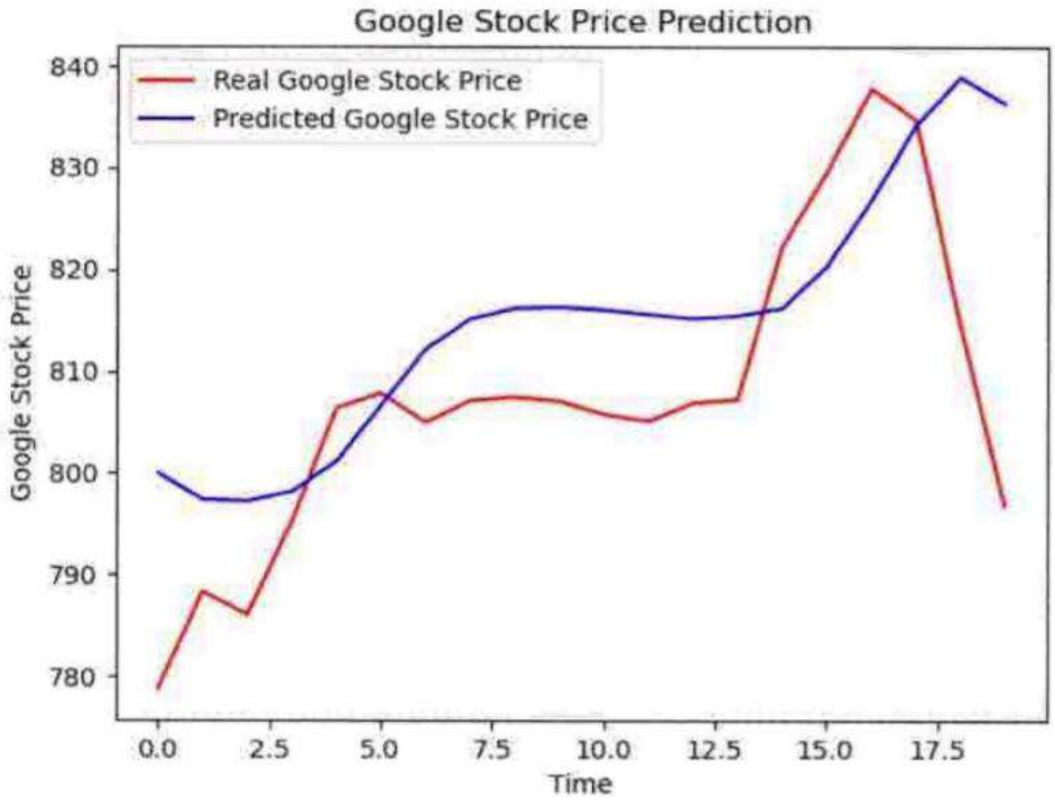


Fig. 1 Data structure for the extracted dataset

International Journal of Engineering Applied Sciences and Technology, 2022
Vol. 7, Issue 12, ISSN No. 2455-2143, Pages 86-95
Published Online April 2023 in IJEAST (http://www.ijeast.com)

Date	Open	High	Low	Close	Volume
1/3/2017	778.81	789.63	775.8	786.14	1
1/4/2017	788.36	791.34	783.16	786.9	1
1/5/2017	786.08	794.48	785.02	794.02	1
1/6/2017	795.26	807.9	792.2	806.15	1
1/9/2017	806.4	809.97	802.83	806.65	1
1/10/2017	807.86	809.13	803.51	804.79	1
1/11/2017	805	808.15	801.37	807.91	1
1/12/2017	807.14	807.39	799.17	806.36	1
1/13/2017	807.48	811.22	806.69	807.88	1
1/17/2017	807.08	807.14	800.37	804.61	1
1/18/2017	805.81	806.21	800.99	806.07	1
1/19/2017	805.12	809.48	801.8	802.17	919
1/20/2017	806.91	806.91	801.69	805.02	1
1/23/2017	807.25	820.87	803.74	819.31	1
1/24/2017	822.3	825.9	817.82	823.87	1
1/25/2017	829.62	835.77	825.06	835.67	1
1/26/2017	837.81	838	827.01	832.15	2
1/27/2017	834.71	841.95	820.44	823.31	2
1/30/2017	814.66	815.84	799.8	802.32	3
1/31/2017	796.86	801.25	790.52	796.79	2



V. CONCLUSIONS

Data extraction and Google preprocessing stock market dataset, engineering tasks, and a model for forecasting stock price movements based on long-short-term memory are all included in this study (LSTM). We gathered organised, clean data on the Google stock market for ten years. The LSTM prediction model acquired a high prediction accuracy that beats leading models in most similar works. The system adapts by developing a feature engineering technique. Also, we thoroughly evaluated this job. By evaluating the most extensively used machine learning models with our recommended LSTM model inside the design part of our proposed system, we derive numerous heuristic findings that might be future research challenges in technological and financial research. In contrast to previous attempts, the approach we propose is a fresh adaptation in that, rather of just developing another cutting-edge LSTM model, we constructed a tailored and personalised deep learning prediction system by combining sophisticated feature engineering with LSTM. By analysing the findings from the earlier sections, we get to the conclusion that the performance of the model can be significantly enhanced by developing a feature expansion method prior to recursive feature elimination.

Obtaining more thorough study on how technical indices affect various period durations is one potential future research path. When the most modern sentiment analysis methodologies are integrated with feature engineering and deep learning models, a more intricate prediction system that is trained on a variety of information, such as tweets, news articles, and other text data, has enormous potential for advancement.

ACKNOWLEDGMENT

We would like to express our gratitude to Dr. Aarti Sehwag for her insightful criticism and recommendations that helped the paper's quality and for encouraging us to continuously assess our work. We also want to express our gratitude to the Computer Science and Engineering Department.

VI. REFERENCES

[1]. Reddy, V. K. S. (2018). Stock Market Prediction Using Machine Learning. International Journal of Engineering Research and Technology, 11(5), 630-634.

[2]. Marwala, L. R. (2003). Forecasting the Stock Market Index Using Artificial Intelligence Techniques. South African Journal of Industrial Engineering, 14(1), 43-53.

[3]. Wassim El-Hajj, M. M., & Jaber, M. (2019). Automated Stock Price Prediction Using Machine Learning. Journal of Intelligent Learning Systems and Applications, 11(1), 18-30.

[4]. Choi, H. K. (2019). Stock Price Correlation Coefficient Prediction with ARIMA LSTM Hybrid Model. Journal of Computational Science, 34, 42-51.

[5]. Liu, X. Y., Yang, H., Chen, Q., Zhang, R., Yang, L., Xiao, B., & Wang, C. D. (2021). A Deep Reinforcement Learning Library for Automated Stock Trading in Quantitative Finance. arXiv preprint arXiv:2102.05431.

[6]. Zhai, X., Shen, L., Zhang, Y., Cao, B., & Lu, H. (2020). Stock price prediction using multi-head attention mechanism and improved CNN-LSTM model. IEEE Access, 8, 155626-155638.

[7]. Sun, S., Wang, Y., & Ye, Y. (2020). A deep learning model for stock price prediction using daily news. Knowledge-Based Systems, 197, 105844.

[8]. Zhou, Y., Zhang, L., Zhao, Y., & Xie, X. (2020). Stock price prediction based on machine learning and news sentiment analysis. Expert Systems with Applications, 146, 113158.

[9]. Zhang, J., Wang, T., Li, S., Liu, Y., & Li, J. (2021). An intelligent stock price prediction system based on sentiment analysis and deep learning. IEEE Access, 9, 17664-17673.

[10]. Ding, X., Zhang, S., Zhang, Y., & Gao, X. (2021). Research on stock price prediction based on BERT and LSTM. IEEE Access, 9, 64018-64027.

[11]. Wang, Y., Zhu, D., Li, B., & Yang, L. (2021). Stock price prediction using an attention-based neural network. Complexity, 2021, 1-11.

[12]. Li, Y., Li, Y., Zhao, T., Li, B., & Sun, Y. (2022). Improved stock price prediction based on machine learning and technical analysis. Journal of Ambient Intelligence and Humanized Computing, 13(4), 4067-4077.

[13]. Singh, A. (2021, March 1). Introduction to Long Short-Term Memory (LSTM). Analytics Vidhya.https://www.analyticsvidhya.com/blog/2021/03/introduction-to-long-short-term-memory-lstm/