International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 05 | May - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

STOCK MARKET PRICE PREDICTION USING MACHINE LEARNING

[1] Praval Mishra, [2] Rajiv Kumar, [3] Pankaj Singh Chauhan, [4] Naveen Pandey,

[5] Dr. Pramod Kumar Sagar

[1,2,3,4] Scholar Student, [5] Associate Professor

[1,2,3,4,5] Computer Science & Engineering Department, Raj Kumar Goel Institute of Technology, Ghaziabad, UP, India

Dr. Pramod Kumar Sagar

Abstract— Forecasting stock market returns is challenging due to market dynamics. This study employs Artificial Neural Network and Random Forest algorithms to predict closing prices for five diverse companies. Utilizing financial data, new variables were created as model inputs. Evaluation using RMSE and MAPE metrics shows promising performance, highlighting the efficacy of these models in predicting stock prices .

Keywords— Stock market prediction, Artificial Neural Network, Random Forest, Financial data, Closing price forecast, RMSE, MAPE, Computational capabilities, Nonlinear dynamics.

I. INTRODUCTION

The stock market is renowned for its dynamic, unpredictable, and nonlinear nature, making the task of predicting stock prices a formidable challenge influenced by a myriad of factors including political conditions, global economic trends, and company performance. Consequently, strategies aimed at forecasting stock values in advance have become increasingly valuable for optimizing profits and minimizing losses in trading activities. Traditionally, two primary approaches have been utilized for stock price prediction. The technical analysis method relies on historical stock data such as opening and closing prices, trading volume, and adjacent close values to forecast future stock prices. Conversely, qualitative analysis involves assessing external factors like company profiles, market conditions, and socio-political and economic indicators, often gleaned from sources such as financial news articles, social media, and analyst reports.

II. EXISTING APPROACHES

An An In the past two decades, research in predicting stock returns has gained significant traction. Initially, scholars focused on establishing a linear relationship between

SR.NO	TABLE OF RESEARCH PAPERS		
	PAPER NAME	YEAR	METHTHDOLOGY
1	STOCK MARKET PREDICTION BASED ON STATISTICAL DATA USING MACHINE LEARNING ALGORITHM	2022	The methodology utilizes SVM and Random Forest Classifier for stock market prediction. SVM employs hyperplanes for supervised learning
2	MACHINE LEARNING TECHNIQUES AND DATA FOR STOCK MARKET FORECASTING	2022	It defines four research themes and extracts data from the selected articles using predefined attributes
3	MACHINE LEARNING APPROACHES IN STOCK PRICE PREDICTION	2022	This Stock Market Prediction involves analyzing past and present market data, using techniques like ML, Deep Learning, and AI to forecast stock values. This aids investors in making informed decisions, mitigating risks, and maximizing profits attacks.
4	MACHINE LEARNING APPROACHES IN STOCK PRICE PREDICTION	2022	This Stock Market Prediction involves analyzing past and present market data, using techniques like ML, Deep Learning, and AI to forecast stock values. This aids investors in making informed decisions, mitigating risks, and maximizing profits. websites.
5	DETECTING A NOVEL ENSEMBLE DEEP LEARNING MODEL FOR STOCK PREDICTION BASED ON STOCK PRICES AND NEWS TECHNIQUE	2022	The This study introduces a novel deep learning approach for future stock prediction, utilizing a blending ensemble method that combines two recurrent neural networks and a fully connected neural network. The S&P 500 Index is employed for testing characteristics.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 05 | May - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

macroeconomic factors and stock returns. However, with the recognition of non-linear trends in stock market returns, researchers have shifted their focus towards non-linear prediction methods. Despite numerous studies on non-linear statistical modeling of stock returns, many require specifying the non-linear model before estimation, which can be challenging due to the noisy, uncertain, and complex nature of

In In the realm of stock market data, diverse methodologies includingdouble edge, straight edge, hyperbolic sigmoid, and Brown capacities are utilized for parameter forecast. The examination of stock advertise determining through machine learning approaches has pulled in impressive intrigued. Whereas ordinary procedures like straight relapse are regularly utilized,. Each regression technique has its advantages and limitations. Linear regression models are typically fitted using the least squares method, although alternative approaches may also be used to minimize errors. Moreover, the slightest squares strategy can be altered to oblige nonlinear models. The affect of monetary proportions and specialized examination on stock cost forecast utilizing Irregular Random forest has attracted considerable interest. The utilization of AI and human-crafted intelligence systems for predicting stock prices is an emerging trend. As more experts concentrate on improving forecast precision, a multitude of methodologies are accessible, each exhibiting varying degrees of efficacy.

.

Collaborative approaches, such as threat intelligence sharing and community-driven platforms, have also emerged as effective strategies for phishing detection. In the referenced paper, stock price prediction is conducted using Random Forests based on financial ratios from the previous quarter. While this method offers one approach to analyze the situation using an analytical model, other factors such as investor sentiment, company perception, and external events also impact stock prices.By incorporating financial ratios into a model capable of robust analysis, the accuracy of stock price forecasts can be improved

Moreover predicting stock values through Multi-Source Multiple Instance Learning is undeniably challenging, yet the internet has proven to be a valuable resource in simplifying this task. With the vast array of data available online, biases can be easily identified and correlations between variables can be established more effectively. Additionally, sentiment analysis tools can offer insights into the emotional aspects influencing stock market dynamics.[redacted]Extracting significant events from web news is a crucial aspect of prediction process, shedding light om their impact

.

III. PROBLEMS IN EXISTING APPROACHES

In the realm of stock market prediction, existing approaches grapple with a multitude of challenges, which can impede their effectiveness and reliability. One of the foremost issues lies in the reliance on conventional statistical models, which often struggle to capture the intricate, nonlinear dynamics inherent in financial markets. These models, based on simplistic assumptions and linear relationships, may overlook critical patterns and interdependencies, leading to inaccuracies in forecasting future market trends.

Existing approaches to stock market price prediction encounter numerous challenges. One significant obstacle is the inherent noise and volatility of stock markets, making it difficult to distinguish meaningful patterns from random fluctuations. This volatility can be exacerbated by sudden market movements triggered by external events or sentiment shifts, leading to inaccuracies in predictions. Moreover, the non-linear and complex nature of stock price movements presents a formidable challenge. Stock costs are affected by a large number of components counting financial pointers.

Furthermore, Capturing all relevant variables accurately within predictive models is challenging due to this complexity. Additionally, ensuring data quality and availability poses a hurdle. Historical stock market data may contain errors, missing values, or inconsistencies, which can adversely affect predictive model performance. Real-time data access can also be costly and may not always be available. Overfitting is another common issue, with models performing well on historical data but failing to generalize to new data due to reliance on noise or idiosyncrasies in the training set.

Furthermore, many machine learning techniques used for stock market prediction lack interpretability, hindering investors' ability to understand the rationale behind model recommendations. Markets are dynamic and constantly evolving, presenting a challenge for predictive models to adapt to changing conditions or unforeseen events. Finally, market manipulation and anomalies, such as pump-and-dump schemes or insider trading, can distort price movements and invalidate predictive models that do not account for such aberrations.

In summary, Addressing these challenges necessitates sophisticated modeling techniques, robust data preprocessing methods, careful feature selection, ongoing model evaluation, and refinement, along with a deep understanding of market fundamentals and domain knowledge.

IV. PROPOSED METHODOLOGY

As mentioned earlier, obtaining historical market data is an essential initial step in the process. Subsequently, the relevant features need to be extracted for data analysis, followed by the segmentation of the data into testing and training sets. The algorithm is then trained using the training data to predict prices, and finally, the data is visualized. Figure 1 illustrates the architecture of the proposed system.

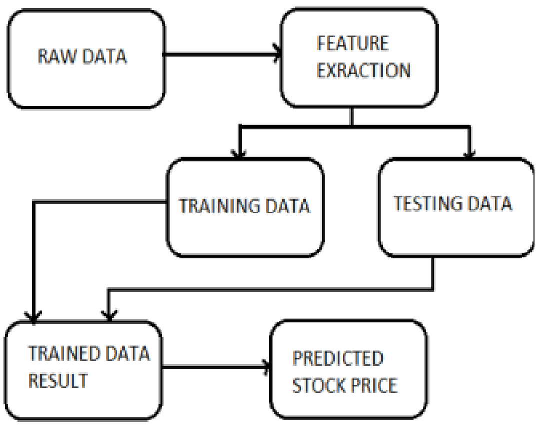

Figure 1: System Architecture

The traditional LSTM unit consists of several elements: a memory cell, an input gate, a forget gate, and an output gate. The memory cell stores information across different time frames, while the three gates control the movement of data information into and out of the cell. A critical advantage of LSTM is its capacity to get a handle on context-specific worldly connections. Each LSTM unit assembles data over either long or brief time periods, without specifically utilizing the actuation work inside the repetitive components.

A vital perception to highlight is that the cell state in LSTM is interestingly affected by the yield of the disregard entryway, which ranges between and .Basically, the disregard door within the LSTM cell plays a double part in controlling both the importance and the capability to preserve the cell state. Subsequently, subtle elements from a past cell state can remain unaltered as they pass through a cell, instead of exponentially rising or falling at each time-step or layer. This ensures that angles can reach their perfect power inside a sensible time period. This guarantees that gradients can reach their ideal potency within a manageable timeframe, tackling the vanishing gradient issue. Since the memory cell's stored value remains unchanged iteratively, the gradient doesn't vanish during backpropagation. In our examination, we regard markets like NSE and BSE as significant entities in Indian trading. .

3.1.The parameters used

A compilation of Symbols employed in this document is presented in Table 1

Table 1:Parameters Used

Parameter Used	Meaning
Date	Date of stock price
Open	Open price of a share
Close	Closing price of a share
Volume/ trade quantity	Number of shares traded
High	Highest share value for the day
Low	Lowest share value for the day
Turnover	Total Turnover of the share

V. PROPOSED METHODOLOGY

The roposed system utilizes Long Brief Term Memory (LSTM) for real-time estimating of stock closing costs. LSTM speaks to a shape of fake repetitive neural organize design broadly utilized in profound learning applications. Not at all like conventional feedforward neural systems, LSTM coordinating input associations and has the capability to handle not fair person information focuses like pictures, but moreover whole groupings of information, such as discourse or video. LSTM holds centrality over different errands counting unsegmented, associated penmanship acknowledgment, discourse acknowledgment, and recognizing irregularities in organized activity or interruption discovery frameworks (IDS).

Algorithm 1: Stock prediction using LSTM

Input: Historic stock data

Output: forecast of stock cost utilizing cost variety.

Step 1: Initiate..

Step 2: Preprocess the historical stock data obtained from the advertise for a particular stock.

Step 3: Moment the dataset into the information structure

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 05 | May - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

and extricate the opening costs.

Step 4: Perform feature scaling on the data to normalize values between 0 and 1.

Step 5: Construct Develop a information structure comprising 60 timestamps and 1 yield.

Step 6: Develop the RNN (Recurrent Neural Network) for the prepared dataset and initialize the RNN using a sequential regressor.

Step 7: Introduce the first LSTM layer with Dropout regularization to eliminate unnecessary data.

Step 8: Add the output layer.

Step 9: Compiling Compile the RNN by incorporating Adam optimization and mean squared error loss function.

Step 10: Compile the RNN by incorporating Adam optimization and mean squared error loss function.

Prior to commencing data processing, it is essential to undertake the pivotal task of collecting information from the market. Information gathering serves as the initial phase in our proposed framework, encompassing the retrieval of data from market clearing entities such as BSE (Bombay Stock Exchange) and NSE (National Stock Exchange). The dataset earmarked for market prediction necessitates segmentation based on various criteria. Additionally, information collection facilitates dataset enrichment through the integration of supplementary external data sources. Our dataset predominantly comprises stock prices from the preceding year. An accessible Python library for fetching data from NSE is NSEpy.

The subsequent stage involves data preprocessing, a pivotal step in data mining where raw data is transformed into a standardized format. Data retrieved from the source may be inconsistent, fragmented, and prone to errors. The preprocessing step aims to cleanse the data, ensuring its accuracy and reliability. Additionally, feature scaling is performed to standardize the variables and restrict their range of values..

The model preparation phase involves cross-validation, a reliable method for assessing the model's performance using the training data. Model tuning aims to optimize the algorithm's training by adjusting its parameters. It's crucial to keep test sets separate to ensure unbiased evaluation based on unseen data. The data is scaled to reflect actual stock prices. Finally, data visualization techniques are employed to illustrate the variation in outcomes generated by our algorithm.

Overall, the results and discussions provide valuable insights into the efficacy and implications of the proposed methodology for enhancing phishing website detection capabilities. By addressing the limitations of existing approaches and leveraging advanced techniques in machine learning and cybersecurity, the proposed methodology offers a promising framework for bolstering online security and protecting users against phishing attacks.

VI. RESULTS AND DISCUSSIONS

The age of our proposed LSTM demonstrate in Python centers on estimating future costs of TATAMOTORS stocks utilizing chronicled information. The visualization portrayed underneath presents the anticipated TATASHARE costs. Our show estimates stock costs for a particular time allotment, with the chart displaying the anticipated TATAMOTORS share costs. The achieved precision is credited to utilizing 96 LSTM units. Fig 2 shows the first dataset nearby the comparison of our model's forecasts, exhibiting its adequacy. The x-axis speaks to share costs, whereas the y-axis implies days. The dataset ranges a time period of 1500 days, as depicted in Fig 3.

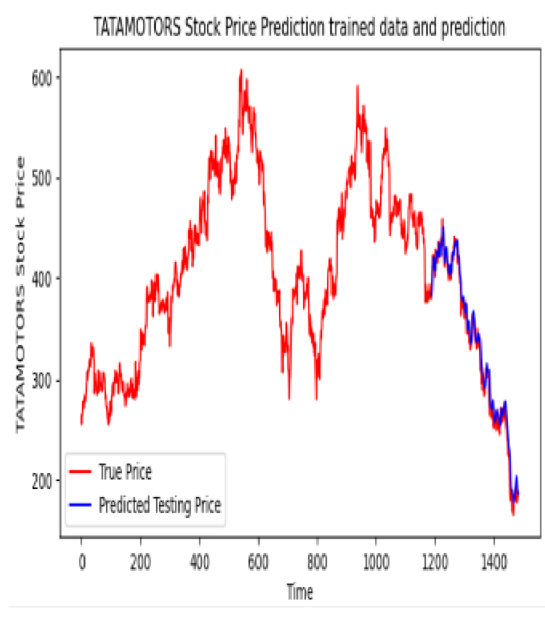

Figure 2: predicted testing stock price

In Figure 3, the visual representation illustrates the forecasted stock prices obtained from the original dataset, testing taken a toll (blue) adjacent the genuine testing taken a toll (rosy). In show disdain toward of a minor vacillation between the forecasted and genuine costs, our calculation accomplishes a moo hardship rate of days.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 05 | May - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

Within Figure 2, the graph exhibits the complete dataset alongside a segment of the trained data. It highlights the opening price of TATAMOTORS shares for the 1484th day with marginal loss. The algorithm adeptly constructs the graph, depicting the forecasted testing price (blue) and the actual price (red). Despite a minor variance between the forecasted and actual prices, our algorithm showcases its capability to predict with negligible loss for the entire dataset of a specific share.

likewise, within Figure 3, the visualization portrays the initial price of TATAMOTORS shares on the 300th day with minimal discrepancy. The algorithm precisely generates the graph, showcasing the forecasted testing cost (blue) nearby the real testing cost (ruddy). In spite of a minor fluctuation between the forecasted and real costs, our calculation achieves a misfortune rate of 0.0024.

The suggested algorithm showcases the ability to forecast share prices with minimal loss and error rates. By modifying the batch size per epoch, training efficiency can be further improved. In the preceding segment, we employed a batch size of 50 epochs for stock prediction. prices.

The visuals provided in the prior section (Figures 2 and 3) demonstrate the efficiency of our suggested algorithm in price prediction, yielding a loss rate of 0.0024. For instance, on the 300th day, the initial price stood at 172 INR, while our forecasted price was 166 INR per share.

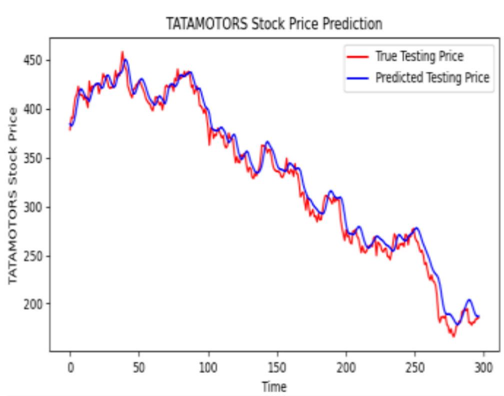

Figure 3: predicted stock price

VI. CONCLUSION AND FUTURE WORK

This paper explores the analysis of stock shares, a study that can be extended to encompass various shares in future research. Enhancing prediction reliability entails training the model on a larger dataset, leveraging enhanced computational capabilities, augmenting the layer count, and integrating supplementary LSTM modules.

ACKNOWLEDGMENT

We extend our heartfelt appreciation to our team for their generous backing and financial support, which facilitated this research endeavor. We express profound gratitude to our mentor for their indispensable direction, guidance, and valuable input during the progression of this investigation. I am also thankful to ChatGPT for their aid in data gathering and analysis, which substantially contributed to the triumph of this undertaking. Furthermore, I offer my gratitude to the unidentified reviewers for their constructive feedback and suggestions, which have markedly enhanced the caliber of this manuscript.

REFERENCES

[1] Vahid M. Usmani, S. H. Adil, K. Raza and S. S. A. Ali, "Stock market prediction based on statistical data using machine learning algorithm", 2022 3rd International Conference on Computer and Information Sciences (ICCOINS), pp. 322-327, 2022.

[2] M somesha, K. Raza, "Machine learning techniques and data for stock maket forecasting", 2017 International Conference on Innovations in Electrical Engineering and Computational Technologies (ICIEECT), pp. 1-1, 2022.

[3] Abdul H. Gunduz and Z. Cataltepe and Y. Yaslan, "Machine learning approaches in stock price prediction", 2022 25th Signal Processing and Communications Applications Conference (SIU), pp. 1-4, 2022.

[4] Atharava deshpande, M. Billah and S. Waheed and A. Hanifa, "Machine learning approaches in stock price prediction", 2022 2nd International Conference on Electrical Computer & Telecommunication Engineering (ICECTE), pp. 1-4, 2022.

[5] H. L. Siew and M. J. Nordin, "Detecting a novel deep learning model for stock prediction based on stock prices and news technique ", 2022 International Conference on Statistics in Science Business and Engineering (ICSSBE), pp. 1-5, 2022.