ISSN: 2278-4632
Vol-12 Issue-01 No.01: 2022

STOCK INDEX DATA ANALYSIS USINGLONG SHORT-TERM MEMORY MODEL IN MACHINE LEARNING

ORUGANTI. TARUN[1],SANGA. MEENA[2], BOYALAPALLI. ABHISHEK[3], BANDRAPALLI. PAVAN Student, Department of CSE, NRI INSTITUTE OF TECHNOLOGY, Vijayawada, A.P., India.

Dr.CH. SURYA KIRAN Professor, Department of CSE, NRI INSTITUTE OF TECHNOLOGY, Vijayawada, A.P., India

ABSTRACT

Analyzing stock index data is one of the most difficult tasks in the field of computation. There are many involved in the analyzing-physical factors vs. physiological, rational and irrational behavior, investor sentiment, market rumors, etc. All these aspects combine to make stock prices volatile and very difficult to analyze with a high degree of accuracy. We investigate data analysis as a game changer in this domain. As per efficient market theory when all information related to a company and stock index events are instantly available to all stakeholders/market investors, then the effects of those events already embed themselves in the stock index. So, it is said that only the historical spot price carries the impact of all other market events and can be employed to analyze its future movement. Hence, considering the past stock index as the final manifestation of all impacting factors we employ Machine Learning (ML) techniques on historical stock index data to infer future trend ML techniques have the potential to unearth patterns and insights we didn't see before, and these can be used to make unerringly accurate predictions. We propose a framework using LSTM (Long Short Term Memory) model and to analyze as well as prediction of future growth of a company.

INTRODUCTION

Data analysis has been used in all business for data-driven decision making. In share market, there are many factors that drive the share price, and the pattern of the change of price is not regular. This is why it is tough to take a robust decision on future price. Artificial Neural Network (ANN) has the capability to learn from the past data and make the decision over future. Deep learning networks such as Convolutional Neural Network (CNN), Recurrent Neural Network (RNN) etc. works great with multivariate time series data. We train our model from the past stock index data and calculate the future price and index of that stock. This future price, indexes used to

ISSN: 2278-4632
Vol-12 Issue-01 No.01: 2022

calculate the future growth of a company. Moreover, we found a future growth curve from different companies. Thus we can analyze and investigate the similarity of one company's future curve over another. Stock price of a listed company in a stock exchange varies every time an order is placed for sell or buy and a transaction completes.

The stock market is a vast array of investors and traders who stocks are governed by the principles of demand and supply, and the ultimate goal of buying shares is to make money by buying stocks in companies whose perceived value (i.e., share price) is expected to rise. Stock markets are closely linked with the world of economics -the rise and fall of share prices can be traced back to some Key Performance Indicators (KPI's). The five most commonly used KPI's are the opening stock price (`Open'), end-of-day price (`Close'), intraday low price (`Low'), intra-day peak price (`High'), and total volume of stocks tradedduring the day (`Volume').

This aspect of stock price movement can be used as an important tool to analyze the index of many stocks at once. Due to the sheer volume of money involved and number of transactions that take place every minute, there comes a trade-off between the accuracy and the volume of predictions made; as such, most stock analyzing systems are implemented in a distributed, parallelized fashion. These are some of the considerations and challenges faced in stock market analysis.

EXISTING SYSTEM

Traditional approaches to stock market analysis and stock price prediction include fundamental analysis, which looks at a stock's past performance and the general credibility of the company itself, and statistical analysis, which is solely concerned with number crunching and identifying patterns in stock index price variation. The latter is commonly achieved with the help of Artificial Neural Networks (ANN's), but these fail to capture correlation between stock prices in the form of long-term temporal dependencies. Another major issue with using simple ANNs for stock prediction is the phenomenon of exploding / vanishing gradient, where the weights of a large network either become too large or too small (respectively), drastically slowing their convergence to the optimal value. This is typically caused by two factors: weights are initialized randomly, and the weights closer to the end of the network also tend to change a lot more than those at the beginning. An alternative approach to stock market analysis is to reduce the dimensionality of the input data and apply feature selection algorithms to shortlist a core

ISSN: 2278-4632
Vol-12 Issue-01 No.01: 2022

set of features (such as GDP, oil price, inflation rate, etc.) that have the greatest impact on stock index prices or currency exchange rates across markets. However, this method does not consider long-term trading strategies as it fails to take the entire history of trends into account; furthermore, there is no provision for outlier detection.

PROPOSED SYSTEM

The prediction of stock value is a complex task which needs a robust algorithm background in order to compute the longer term stock index data. Stock index data are correlated within the nature of market; hence it will difficult to analyze the indices of stock data. The proposed algorithm using the stock market data to analyze the index using machine learning techniques like Recurrent Neural Network named as long short Term Memory (LSTM), in that process weights are corrected for each data points using stochastic gradient descent. This system will provide accurate outcomes in comparison to currently available stock index analysis algorithm. It is a recurrent network because of the feedback connections in this architecture. It has an advantage over traditional neural network due to its capability to process the entire sequence of data. Its architecture comprises the cell, input gate, output gate and forget gate.

LSTM networks are popularly used on time-series data for classification, processing, and making prediction and analysis. The reason for its popularity in time-series application is that there can be several lags of unknown duration between important events in time series. In this task, the future stock index data are analyzed, predicted using the LSTM. Our task is to analyze stock indexes for a few days, which is a time series problem. The historical index data are collected automatically using the "nsepy" library of python. This data set contains observations with attributes. After processing, only dates and OHLC (Open, High, Low, Close) columns, a total of 5 columns are taken as these columns have main significance in the dataset. The LSTM model is trained on the entire dataset, and for the testing purpose, a new dataset is fetched for the duration of the stock index data to be analyzed. The stock index analysis for this new duration will be analyzed by the already trained LSTM model, and the analyzed index data will be plotted against the original index data to visualize the model accuracy. The network is trained and evaluated with various sizes of input data to urge the graphical outcomes.

IMPLEMENTATION

LSTM is very sensitive to the scale of the data, Here the scale of the close value is in a

ISSN: 2278-4632
Vol-12 Issue-01 No.01: 2022

kind of scale, and we should always try to transform the value. Here we will use min-max scalar to transform the values from 0 to 1. Firstly, we can download the raw stock index dataset of selected companies are collected from the NSE (National Stock Exchange) official website.

Data Pre-Processing: This step incorporates the following

a) Data discretization: Part of data reduction but with particular importance, especially for numerical data.
b) Data transformation: Normalization.
c) Data cleaning: Fill in missing values.
d) Data integration: Integration of data files. After the dataset is transformed into a clean dataset, the dataset is divided into training and testing sets so as to evaluate.

Feature Selection: In this step, data attributes are chosen that are going to be fed to the neural network. In this study Date & Close price are chosen as selected features.

Train the NN model: The NN model is trained by feeding the training dataset. The model is initiated using random weights and biases. Proposed LSTM model consists of a sequential input layer followed by 3 LSTM layers and then dense layer with activation. The output layer again consists of a dense layer with a linear activation function.

LSTM has three gates:

a) Input Gate: The input gate adds information to the cell state.
b) Forget Gate: It removes the information that is no longer required by the model.
c) Output Gate: Output gate at LSTM selects the information to be shown as output.

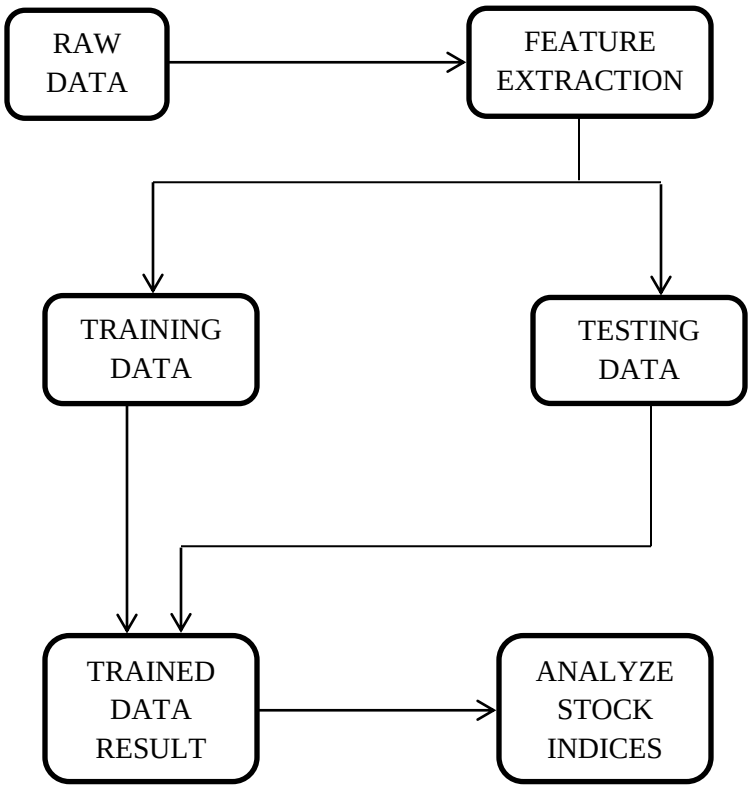
Output Generation: The RNN generated output is compared with the target values and error difference is calculated. The Back propagation algorithm is used to minimize the error difference by adjusting the biases and weights of the neural network. By calculating deviation we check the percentage of error of our prediction with respect of the neural network.

Visualization: After fitting the data with our model we use it for prediction. We must use inverse transformation to get back the original value with the transformed function. Now we can use this data to visualize the prediction.

Investigate different time interval: We repeated this process to predict the price at different time intervals. In this different time span, we calculate the percentage of error in the future

ISSN: 2278-4632
Vol-12 Issue-01 No.01: 2022

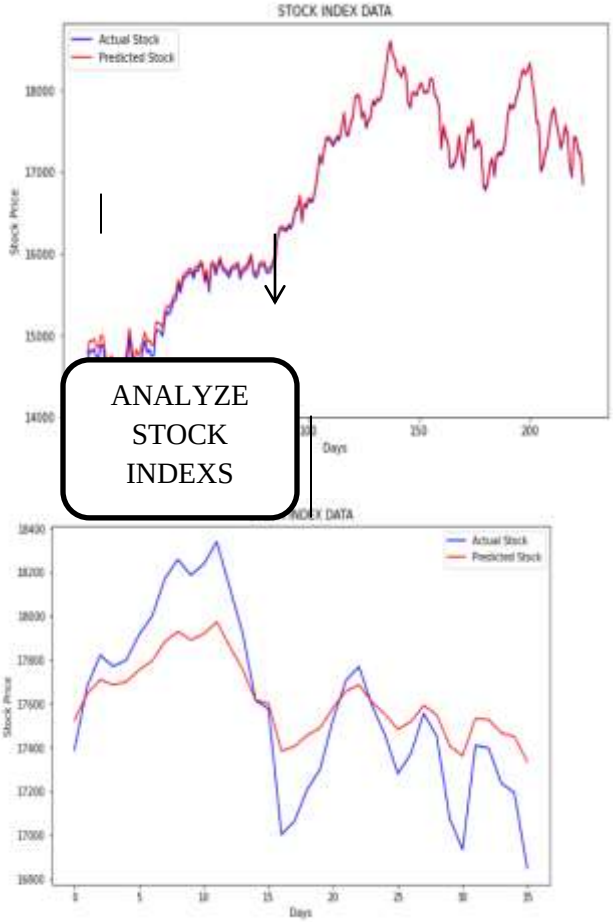
prediction. This would be different for different sectors. So, this will help to find a frame for the particular sector to predict future companies.



SAMPLE SCREENS

	Date	Open	High	Low	Close	Shares Traded	Turnover (Rs. Cr)
0	30-Mar-2021	14628.50	14876.30	14617.60	14845.10	594520567	35117.93
1	31-Mar-2021	14811.85	14813.75	14670.25	14690.70	436874691	29741.83
2	01-Apr-2021	14798.40	14883.20	14692.45	14867.35	430028478	29585.23
3	05-Apr-2021	14837.70	14849.85	14459.50	14637.80	500476690	33885.39
4	06-Apr-2021	14737.00	14779.10	14573.90	14683.50	474211156	31885.97
...	...	...	...	...	...	...	...
220	16-Feb-2022	17408.45	17490.60	17257.70	17322.20	244549223	21128.38
221	17-Feb-2022	17396.55	17442.90	17235.85	17304.60	232136131	19381.26
222	18-Feb-2022	17236.05	17380.80	17219.20	17276.30	189620888	16126.96
223	21-Feb-2022	17192.25	17361.05	17070.70	17206.65	215183301	18725.57
224	22-Feb-2022	16847.95	17148.55	16843.80	17082.20	300131995	24040.84

225 rows × 7 columns



ISSN: 2278-4632
Vol-12 Issue-01 No.01: 2022

CONCLUSION

In this, we have prepared several approaches to analyze the stock index values its movement patterns on a weekly forecast horizon using machine learning and deep learning regression models. Using the daily historical data of NSE index values during the period. We constructed, optimized and then tested the predictive models. Data pre-processing and data wrangling operations were carried on the raw data, and a set of derived variables are created for building the models. Among all the machine learning and deep learning based regression models, the performance of the LSTM based deep learning regression models were found to be far too superior to that of the machine learning based predictive models. The study has conclusively proved our conjecture that deep learning based models have much higher capability in extracting and learning the features of a time series data than their corresponding machine leaning. It also reveals that multi-variant is not a good idea in LSTM based regression, as univariant models are more accurate and faster in their execution.

FUTURE SCOPE FOR FURTHER DEVELOPMENT

The future enhancement we try to add new features to the existing one that features is news and sentiment of the country and company. These features increase the model to find the accurate value at add times. As a future scope of work, we will investigate the possibility of using generative adversarial networks (GAN) in time series analysis and forecasting of stock indexes and price.

REFERENCES:

[1].https://www.hindawi.com/journals/complexity/2020/6431712/

[2].https://www.ijert.org/stock-market-analysis-using-lstm-in-deep-learning

[3].https://www.analyticsvidhya.com/blog/2021/12/stock-price-prediction-using-lstm/

[4].https://www.researchgate.net/publication/341870475_Stock_Market_Analysis_using_LSTM_in_Deep_Learning

[5].https://analyticsindiamag.com/hands-on-guide-to-lstm-recurrent-neural-network-for-stock-market-prediction/

[6].https://www.researchgate.net/publication/356093237_Stock_Market_Prediction_Using_Machine_Learning_Techniques

[7].https://www.ijeat.org/wp-content/uploads/papers/v8i4/D6321048419.pdf.

[8].https://medium.com/analytics-vidhya/analysis-of-stock-price-predictions-using-lstm-models-f993faa524c4

[9].https://www.kdnuggets.com/2018/11/keras-long-short-term-memory-lstm-model-predict-stock-prices.html

[10].https://www.datacamp.com/community/tutorials/lstm-python-stock-market