e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
Volume:03/Issue:06/June-2021 Impact Factor- 5.354 www.irjmets.com

STOCK PRICE PREDICTION USING LSTM, RNN

Rachit A Mishra*[1], Karan L Janani*[2], Nikhil S Pachpande*[3], Manthan A Patel*[4]

*[1,2,3,4]Computer Science And Engineering, SKNSITS (Lonavala), India.

ABSTRACT

Stock market trading is the one of the most important activity in the finance world. Normally in order to predict the market, investors used to analyze stock price and indicators using news related to these stocks. Previously work in industry was focused on classifying the news related to stocks price as positive, negative and neutral and demonstrating the effect on stock price. In this paper we propose automated trading system that integrates mathematical functions, machine learning and other external factors that will be trained from the available stocks data then uses the acquired knowledge for an accurate stock closing price prediction. The programming language is used to predict the stock market using machine learning is Python. To achieve this goal, this project uses Deep Learning models, Long-Short Term Memory (LSTM) Neural Network algorithm, to forecast stock prices. For data with timeframes recurrent neural networks (RNNs) and also Linear Regression and K-Nearest Neighbor are used to compare their efficiencies with each other.

I. INTRODUCTION

Since the development in the AI field, many categories of our existing technology environment has seen a drastic improvement because as it uses this technology to predict their future occurrences. One of which is Stock Market Closing Price prediction which is being discussed in this paper. The share market is an assortment of buyers and sellers of shares or stocks which represent the partial ownership of that business, it basically represents a company's current as well as future growth which in turn will help in the economy. There are various factors involved that drive the prices of stocks on daily basis, which includes the company's market value, its competition, net profit, stability in the sector of finance, the other aspects that may make difference is oil price, foreign exchange rate, political decision as well as government, these factors are beyond the control of that company [1][2]. Researchers all over the world have come up with every possible solution to predict the prices, many techniques are still being tested, applications like reinforcement learning is being applied in this area, thus technologies like these have given the stock market prediction models a concrete base to start for its improvement [3]. Hence in this paper few machine learning and deep learning techniques are studied, to find which has more efficacy.

II. LITERATURE REVIEW

There are number of researches done in the field of predicting stock market prices including LSTM. Pretty much every data mining and prediction techniques were applied to forecast of stock prices, various different features and attributes were used for this purpose. There are basically three main categories of analysis of stock market and price prediction (a) Fundamental analysis, (b) Technical analysis and (c) Time series analysis [1].

[4] In this the author proposed the CNN and LSTM determining environment with addition of some features like financial news and the stock market's historical data. It produced seven prediction models, as indicated by the group of learning strategy, the seven models were developed into one gathering model to get a final model. But, unfortunately all the models did not perform well thus they had low prediction accuracy.

[5] Lai, C.Y. proposed an LSTM model which utilized the average of the past five days' data of the stock market which were (open, high, low, volume, and close) as the input. The forecast was then utilized as part of the average of the stock value data for the following five days through the ARIMA technique. In addition, he used specialized analysis indicators whether to buy, or to keep on holding or to sell the stocks

[6]A model proposed by this author was based on LSTM and the sentiment data which was collected over the time in the stock market. The result was more than good as compared to CNN models.

[7] In the paper proposed by Rana, M., he proposed a LSTM model that beat the LR and SVR models. He also additionally contrasted the diverse actuation functions with distinctive analyzer and concluded that the tanh enactment with the Adam calculation performs best with an accuracy of 98.49% which is hugely better.

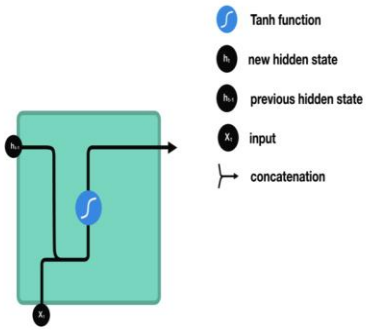
[8] Tiwari, S. and A. Gulati proposed model based on decision tree and neural network concluded that the stock market's subject are the investors that can significantly affect the market's value. Despite this the model can successfully give a good result based on market behaviors as an input.

[9] AL khatib, Khalid et al. paper's on KNN approach studied the Jordanian Stock Market, for this five-year major companies' data was used and studied it was concluded that the result of predicted prices was drastically close to the actual prices which ensured the efficacy of KNN approach.

Recurrent Neural Network

The Recurrent Neural Network is actually a part of the deep learning method that works with the help of similar ways to the convolutional neural network. The recurrent neural network has a problem of losing the previous information as it moves for in its hidden layers. The main reason for this issue is the sequence of data fed to the input layer. If the sequence is long then it is highly possible that the RNN might lose a significant amount of data before it actually reaches the output layer. The Recurrent Neural Network has to deal with a problem called as vanishing gradient problem which is important for allotting the weights or calculating the weights for a particular node or input layer while back propagation. If the vanishing gradient value becomes too small then it doesn't contribute too much learning.

In RNN first the data needs to be transformed into meaningful vectors or machine-readable vectors to be precise enough. Now these vectors are fed to the neural network or the input layer of the network one at a time. While processing, the output of the first hidden state acts as one of the two inputs in the subsequent hidden state. These hidden states act as the memory of the neural network. This method helps to store the data from the previous hidden layers of the neural network.



TANh Activation

In This particular function is known as an activation function that helps to regulate the values flowing in the recurrent neural network explained above. This function helps to convert the values to always stay between -1 to 1. The vectors undergo various a lot of transformation because of the mathematics involved during the flow of data while the neural network is running. Regulating the output becomes slightly easy as the values in the hidden layers are always between -1 and 1. So that's an RNN. It has very few operations internally but works pretty well given the right circumstances (like short sequences). RNN's uses a lot less computational resources than its evolved variants, LSTM's and GRU's.

Long Short-Term Memory

The Long Short-Term Memory has a similar working method like that of the recurrent neural network explained above. It converts the data to machine readable vectors and let it process through the network with the help of forward propagation. The main difference between these two methods is the number of cells of different states through which these two methods operate. The LSTM has states such as the

- Forget Gate
- Input Gate
- Cell State
- Output Gate

These operations are the reason that helps the LSTM to keep or forget the data that it gets as on output through each hidden layer.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
Volume:03/Issue:06/June-2021 Impact Factor- 5.354 www.irjmets.com

The core concept here is the cell state and its hidden state. The cell state helps to transport the data all the way to the output state of the neural network.

III. SYSTEM REQUIREMENTS

Using Python, it is easy to represent real world entities due to its easy-to-read syntax. Python dominates other sets of programming languages due to its rich set of libraries which has played a vital role in AI, Big Data, Software testing, Automation and so on. Numerous already made packages are present to carry out computations based on requirements in this model. For this paper, we have imported NumPy, NLTK, Pandas, Matplotlib, Scikit-learn from python libraries. NumPy: This package is designed for efficient scientific computation. With data structures, executing multi-dimensional arrays and matrices, NumPy is already dominating python. NumPy usage is even aimed at large array processing.

Pandas: "Python Data Analysis Library" is written for Data handling and analysis. When it comes to analyzing data with python, pandas are the preferred tool. Pandas Data frame method data manipulation and analysis is easy when compared with other methods.

TensorFlow: TensorFlow is a library created by Google so as to build various deep learning models. It performs mathematical operations on large multi-dimensional data that can be further generalized as tensors.

IV. PROPOSED MODEL

The model consists of 3 most important stages namely the feature selection, dimensionality reduction and lastly, it's the model building. In the feature selection technique, we check the dataset and try to select the most important ones out of those that can help us the most in model building and prediction of the stock value. Further we need to use the dimensionality reduction techniques for us to visualize the data in 2d and 3d graphs. PCA and tSNE are the two reduction techniques that we will be using here in this case. And finally, the model building part where we will use the LSTM technique to predict the stock prices.

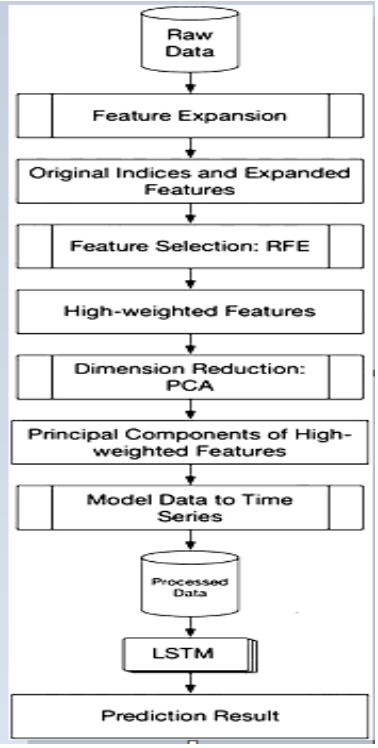


Fig. Proposed workflow

PCA: Principal Component Analysis is a dimensionality reduction technique that helps us convert a large number of features to desirable number of components. The basic idea here is to convert the feature so that the mean of the feature is equal to 0 and the standard deviation is equal to 1. The catch here is that in this process we can lose a significant amount of information while doing this. Here we need to be precise enough while choosing the number of components that we are converting while we are using the PCA algorithm. After this

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
Volume:03/Issue:06/June-2021 Impact Factor- 5.354 www.irjmets.com

step the system will get a modified matrix with a smaller number of features which are nothing but the principal components.

Data Cleaning: This is one of the most important steps while building a machine learning model. The data that we initially get is a raw dataset. This means that there might be a lot of anomalies in the data like null values, missing values, repeated values, etc. Here we need to use the libraries such as pandas and NumPy so that we can get a proper glance at our dataset. Next, we can use the inbuilt functions such as describe to get to know the standard deviation, mean of the dataset. Removing the null values, missing values is an important part of the data cleaning process. Apart from that we also need to check that the data should not be biased towards any one class label. If this is the case then the best approach is to add some random values to the dataset.

Further we need to divide the data set reasonably so that we do not train and test the model on the same data set because that gives a high probability of getting a high accuracy which might clearly not be the case. Best approach is to divide the data to 65% for the training dataset and 35% for the test dataset.

Exploratory data analysis is just a visualization part where we need to plot the graphs like the box plot and violin plot and the scatter plots with the help of the imported libraries such as matplotlib and seaborn. These libraries help us plot the different graphs that help us understand the behavior of the dataset and the variance and the biased nature of it.

V. IMPLEMENTATION

LSTM: As explained above LSTM consists of 4 stages.

After that the Principal Component Analysis part where I most important features are converted into j most important components that can be used to provide as an input to the LSTM input layers. Here we specify an LSTM model and add a conversion procedure for the stock price dataset to predict the proper stock values. The time series conversion function is used to help convert the principal component matrix to time series by shifting the input data frame according to the number of time steps.

In the first gate that the LSTM model has to offer the data, the data has to go through a sigmoid function which literally converts the data between 0 to 1. If the value is close to 0 then it means that the output needs to be forgotten while if it is close to 1 then it needs to be remembered. This here is the basic idea of the forget gate.

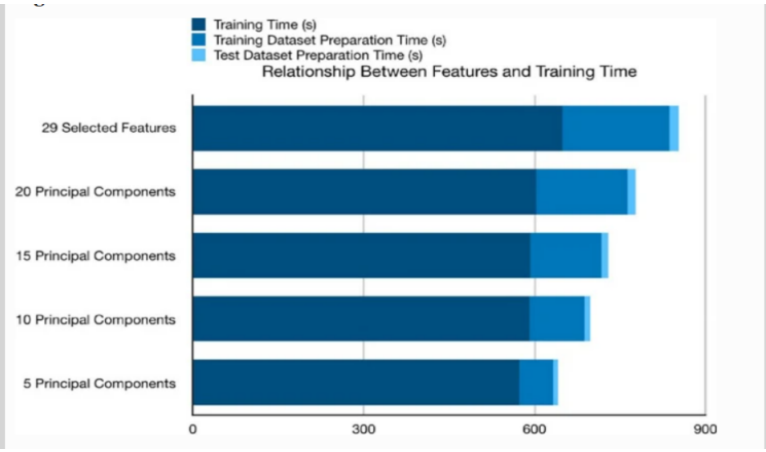
Next comes the input gate which has two different nodes to go through before combining and getting the output. One of them is a sigmoid function and the other is the tanh function. Together they are mutually added to get the output from the input gate.

Now the output from the input gate and the output from the forget gate together make a cell state. That cell state acts as the hidden layer or one input for the next activation layer.

Finally, the cell state needs to go through the tanh function and the input state needs to go through the sigmoid function and further they are multiplied to get an output of one hidden layer.

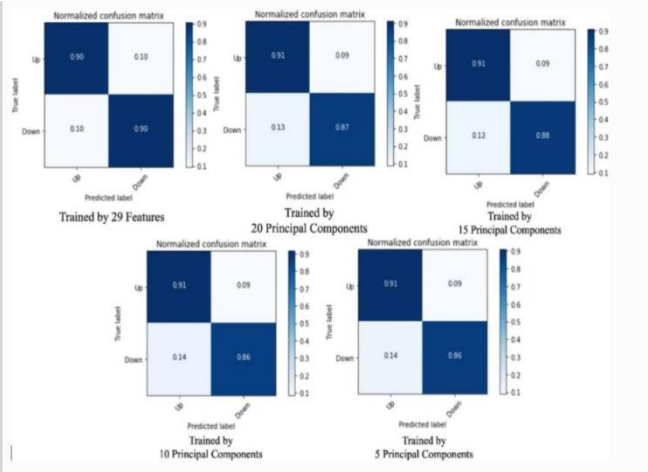
This helps the LSTM's gates learn what information is relevant to keep and what is irrelevant to forget during the training.

Relationships between feature and training time:

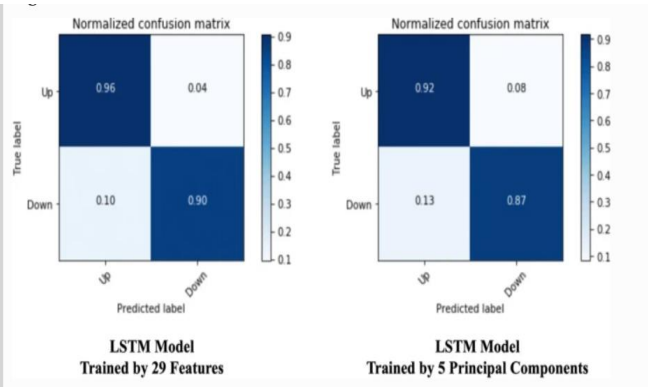


e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
Volume:03/Issue:06/June-2021 Impact Factor- 5.354 www.irjmets.com

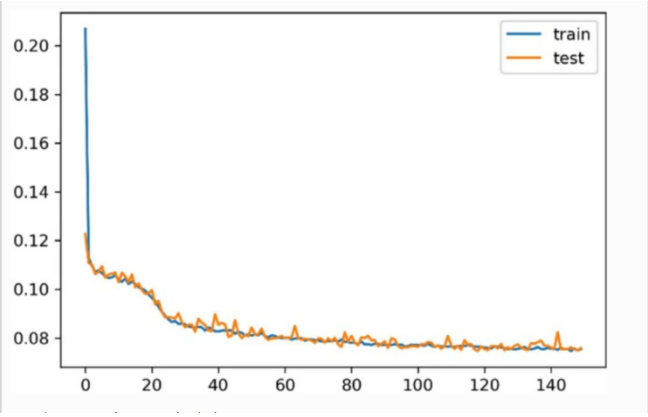
Confusion matrix for different number of components:



Confusion matrix for 5 PCA:



Learning curve for LSTM model:



VI. CONCLUSION

In this project, we found Long-Short term memory (LSTM) shows better prediction accuracy than the classification Algorithms like KNN and Linear Regression which were previously being studied. The stock market is hard to monitor and requires plenty of factors to determine when trying to interpret the movement and predict the price of a stock. Classification algorithms have some problems while predicting the price of the stocks. Linear regression model overfits to date and month instead of taking in the account the previous values from the point of prediction the consider the value from the same date a month ago or the same date a year ago and KNN is almost similar to the linear regression model. We can safely say that regression algorithms have not performed well. While LSTM can store past information that is important, and forget the information that is not important in its memory cell as such they can keep track of dependencies between stock price for a long period while performing predictions and when we increased the size of the dataset it gives better accuracy at its core

stock market is the stock market is dependent upon human emotion and sentiments of the market, numerical data and technical analysis have their limitations to overcome this problem in the future we also have to add sentimental analysis and news feed analysis for social media platforms. We can link this sentimental and newsfeed analysis with LSTM to better train weights of LSTM model and further improve accuracy.

VII. REFERENCES

[1] Ghosh, A., Bose, S., Maji, G., Debnath, N., & Sen, S. (2019). Stock Price Prediction Using LSTM on Indian Share Market.

[2] Hu, Zexin, Zhao, Yiqi and Khushi, Matloob, (2021), A Survey of Forex and Stock Price Prediction Using Deep Learning, Papers, arXiv.org, https://EconPapers.repec.org/RePEc:arx:papers:2103.09750.

[3] Chen, S.; He, H. Stock prediction using convolutional neural network. In 2018 2nd International Conference on Artificial Intelligence Applications and Technologies (AIAAT 2018); IOP Publishing: Shanghai, China, 2018.

[4] Cai, S.; Feng, X.; Deng, Z.; Ming, Z.; Shan, Z. Financial news quantization and stock market forecast research based on CNN and LSTM. In Proceedings of the International Conference on Smart Computing and Communication, Tokyo, Japan, 10–12 December 2018; Springer: Berlin/Heidelberg, Germany, 2018.

[5] Lai, C.Y.; Chen, R.-C.; Caraka, R.E. Prediction stock price based on different index factors using LSTM. In Proceedings of the 2019 International Conference on Machine Learning and Cybernetics (ICMLC), Kobe, Japan, 7–10 July 2019

[6] Chen, M.-Y.; Liao, C.H.; Hsieh, R.-P. Modeling public mood and emotion: Stock market trend prediction with anticipatory computing approach. Comput. Hum. Behav. 2019, 101, 402–408

[7] Rana, M.; Uddin, M.; Hoque, M. Effects of Activation Functions and Optimizers on Stock Price Prediction using LSTM Recurrent Networks. In Proceedings of the 2019 3rd International Conference on Computer Science and Artificial Intelligence, Beijing, China, 6–8 December 2019; Association for Computing Machinery: Normal, IL, USA, 2019; pp. 354–358

[8] Tiwari, S. and A. Gulati. “Prediction of Stock Market from Stream Data Time Series Pattern using Neural Network and Decision Tree.” (2011).

[9] Alkhatib, Khalid et al. “Stock Price Prediction Using K-Nearest Neighbor (kNN) Algorithm.” (2013).