Stock Price Prediction

Name:Ashanpreet Singh Sra

Roll No.-20103033

Department: Computer Science

NIT Jalandhar

Jalandhar,India

Email ID-ashanpreetss.cs.20@nitj.ac.in

Name:Ayush Kumar Jha

Roll No.-20103035

Department: Computer Science

NIT Jalandhar

Jalandhar,India

Email ID-ayushkj.cs.20@nitj.ac.in

Name:Bhart Bansal

Roll No.-20103039

Department: Computer Science

NIT Jalandhar

Jalandhar,India

Email ID-bhartb.cs.20@nitj.ac.in

Name:Cherry Kataria

Roll No.-20103045

Department: Computer Science

NIT Jalandhar

Jalandhar,India

Email ID-cherryk.cs.20@nitj.ac.in

***Abstract*—In the domain of Financial Management, Stock Market has its own importance and a role to play. Prediction of Stock Market has proven to be a difficult task due to its changing computational nature and complexity but now a few models have been developed to predict stock prices with the help of artificial intelligence and Random Forest Techniques. In this work we present a vigorous and precise model of stock price prediction using Machine Learning methods. We have taken dataset of WIPRO and try to predict its stock value, the next day. We have used LSTM (long short-term memory) and RNN architecture to build forecasting framework for stock prices. With these memory cells, networks can effectively associate memories and input remote in time. The financial data: Open, High, low and Close prices of stock are used for creating new variables which are used as input to the model. We have imported Keras libraries and used packages like Sequential, Dense, LSTM and Dropout. Extensive results have been presented on the performance of these models.**

Keywords—Stock Price Prediction, moving average, linear regression, KNN, LSTM, financial forecasting, NETFLIX, TATA

I. INTRODUCTION AND MOTIVATION

Machine Learning has become one of the most reliable tool in financial market. It is used to manage investment efficiently. Machine learning has been used widely in financial sector to provide better decisions in predicting stock market prices.

Price prediction based on a few parameters would be simple, but the outcome could be erroneous because other aspects that were left out could be crucial in understanding the price. The fluctuation of stock prices Various factors, such as the economy, can influence the prices of individual stocks. Economic development. It's a challenge.to manually analyse all factors , it would be preferable if were tools for assisting with the data analysis within a reasonable amount of time Making the best decision in a timely manner As such a large organisation, it has faced numerous challenges. For predicting the future, a large amount of data is required. The price of a stock fluctuates.

Investors can benefit from the examination of this huge data, as well as analysing the direction of stock market indices .With the huge success of machine learning in many domains, finance research has gotten a lot of attention and is being investigated constantly. Thus, in this work, a desktop study was undertaken to investigate the application of machine learning in finance: methods and approaches were used, with the exclusive focus on stock prediction.

We compared daily training outcomes for various companies stock prices, demonstrating that the LSTMs strategy can deliver high enough accuracy. This indicates that it can be used successfully in practise. We also show that RNNs must be trained over a large number of epochs to avoid overfitting to the dataset, with final weights carefully chosen and early termination. Our work is based on NETFLIX and TATA-GLOBAL stock prices prediction.

Therefore, predicting its stock prices has been of great interest.

II Previous Works

The literature seeking to establish or reject the efficient market hypothesis can be divided into three strands based on the variables used and estimating and forecasting methodologies used. The first strand consists of cross-sectional data studies utilising simple regression techniques [10-14]. The second line of research has used time series models and methodologies to forecast stock returns using economic tools such as the Autoregressive Integrated Moving Average (ARIMA), Granger Causality Test, Autoregressive Distributed Lag (ARDL), and Quantile Regression [15-18]. The third strand covers work on stock return prediction using machine learning technologies [19-23]. The incapacity of present stock price prediction solutions to predict stock price movement in a short-term interval is their biggest flaw. The current study seeks to remedy this problem by utilising the capacity of deep neural networks to model and forecast stock price behaviour.

III Proposed Solution

In the domain of Financial Management, Stock Market has its own importance and a role to play. Prediction of Stock Market has proven to be a difficult task due to its changing computational nature and complexity but now a few models have been developed to predict stock prices with the help of artificial intelligence and Random Forest Techniques. In this work we present a vigorous and precise model of stock price prediction using Machine Learning methods. We have taken dataset of NSE-TATAGLOBAL11 and try to predict its stock value, the next day. We have used LSTM (long short-term memory), Moving Average, Linear Regression and K Nearest Neighbour to build forecasting framework for stock prices. With these memory cells, networks can effectively associate memories and input remote in time. The financial data: Open, High, low and Close prices of stock are used for creating new variables which are used as input to the model.

- The first price of any listed stock at the start of a trading day on an exchange is known as the opening price.
- The high and low prices are the stock's highest and lowest prices on that day. Traders typically use these statistics to determine the stock's volatility.
- Closing Price is the stock's price at the end of the trading day.
- Volume is the total number of stocks or contracts traded in all marketplaces for a securities over a specific time period.
- Adjusted Closed Prices is regarded as the stock's genuine price, and it depicts the stock's value after dividends have been distributed.

We have imported libraries and used packages like Sequential, Dense, LSTM and Dropout. Extensive results have been presented on the performance of these models.

Firstly, the libraries that we have imported are Numerical python(numpy), matplotlib for plotting of graphs to show trends, pandas for our dataframe to read files, datetime, rcParams for runtime configurations to default styling of sheets.

For normalizing our data we have to import efficient tools. The sklearn package provides us with efficient to handle our normalization. Here the tool imported from sklearn.preprocessing is MinMaxScaler. Minmaxscaler has default scaling range from 0,1. But we can enter our preferred scale using the feature_range argument and specify the tuple.

For manipulating date and time we are using tool called to_datetime. Specified date format is %Y-%m-%d. Plt.plot gives the graph after plotting our values.

In the above code, we are creating a new dataframe of data and target variable. The new_data created has table with column names Date and Close.
These are declared under for loop for the so that range should be till the length of the data provided. Next we are splitting our data into train set and validation set.

First model technique that we are using is known as Moving Average method. A'moving average,' in simple words, is a statistical tool for determining the direction of a trend. It does so by aggregating a subset of data points over a given time period and dividing the sum by the number of data points in the subset to produce an average. It's called a moving average since the computation is repeated as the number of data points increases over time.
Root Mean Square Error (RMSE) is a metric for determining how well a regression line fits the data points.

Based on the validation set we plotted our data according to the predictions and close.

Next model that we have applied is the Linear regression. It is a statistical method that is used for predictive analysis. Linear regression makes predictions for continuous/real or numeric variables.

For implementation in Linear Regression we are again going to split our dataset into train set and validation set or valid set. The train set will have values till 987 and the valid set will have values after 987. The corresponding values in valid set is 248. Now for train and valid we will be doing x_train and y_train. Here we will drop the close column.
Tool LinearRegression will be imported from the package sklearn.linear_model. Now the fitting will be done using the argument model.fit

Next model that we have applied is the K Nearest Neighbour. K-Nearest Neighbour is a Supervised Learning-based Machine Learning algorithm that is one of the most basic. The K-NN algorithm assumes that the new case/data and existing cases are similar and places the new case in the category that is most similar to the existing categories. The

K-NN method maintains all of the available data and classifies a new data point based on its resemblance to the existing cheval. This means that new data can be quickly sorted into a suitable category using the K-NN method.

GridSearchCV is a library function in the model selection package of sklearn. It assists in looping over specified hyperparameters and fitting your estimator (model) to your training data.

We are using our scaling tool i.e., MinMaxScaler to scale our data within certain range. Then we use the GridsearchCV tool to find the best parameters so that our KNN algorithm applies correspondingly. The train set and valid set have been created which contain the scaled values . Hence we start to to fit our model and then make prediction based on the valid set.

Next method used in the development of our model in the LSTM model which is based on the Recurrent Neural Networks. Long Short Term Memory Networks (LSTMs) are a type of RNN that can learn long-term dependencies. They are currently frequently utilised and function exceptionally effectively on a wide range of situations. LSTMs are specifically developed to prevent the problem of long-term dependency.

For LSTM we have to import various libraries such as Sequential, Dense, Dropout, LSTM. They are imported from keras package which are widely used in the imlementation of these kinds of models.

Dataframe have been created for data and the target variable. Considering the length of our data we have created a loop for which our new_data will contain the columns of Date and Close for the length provided by dataset.
Index have been defined which will have the starting point of Date.

Now to train our model we are making the train set and for testing we are making the test set. Train set will contain values from 0 to 987 and the valid set will contain values from 987 to last (having 248 values).
Next we will convert our dataset into x_train and y_train. The MinMaxScaler will define the range of values. Appending of values will be done from 0 to 59 then 1 to 60 likewise.

Compilation will be done using model.compile where argument loss means how well our data performs in predicting the outcome. Here we use the optimizer to optimize our algorithm. It has the optimizing instance.
Now our test set has been developed and the closing_price will do the prediction based on x_test provided.

After all the computation we have the graph plotted for train and valid.

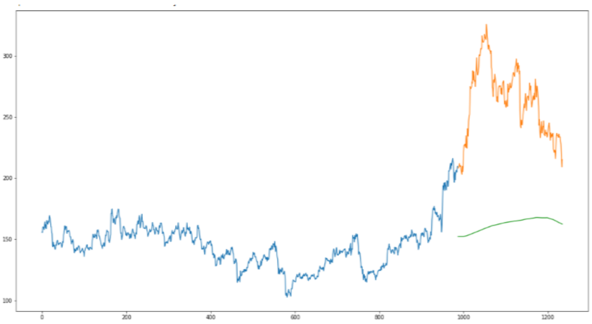
IV Experimental Results And Comparisons

I Moving Average

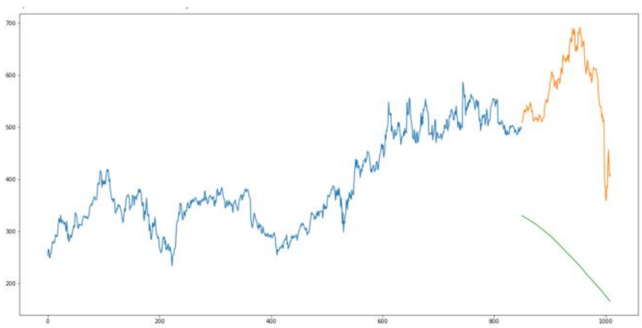
A moving average is a statistic that measures how much a data series has changed over time. Technical analysts in finance frequently use moving averages to examine price patterns for individual securities. A rising trend in a moving average could indicate an increase in the price or momentum of a security, whereas a falling trend could indicate a decline.
The average of a collection of previously observed values will be used to predict the closing price for each day. We'll utilise the moving average technique instead of the basic average, which considers the most recent set of numbers for each prediction.

In other words, the projected values are taken into account for each consecutive step while the oldest observed value is removed from the collection. Here's a basic diagram that can help you comprehend things better.

FOR TATA-GLOBAL



FOR NETFLIX



Although the RMSE is close to 105, the results are not encouraging (as you can gather from the plot). The anticipated values are within the same range as the train set's observed values (there is an increasing trend initially and then a slow decrease).

	NETFLIX	NSE - TATAGLOBAL
RMSE	328.2054	104.5145
MSE	107718.791	10923.208524

VARIANCE SCORE	-0.49313	-0.02969

ii Linear Regression

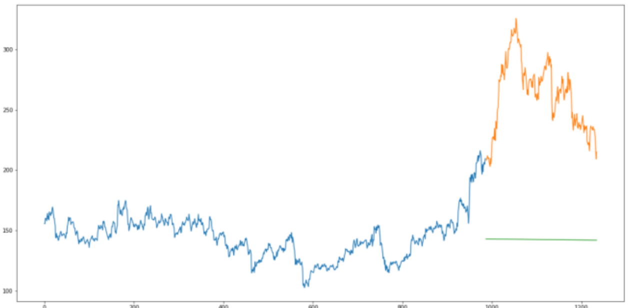
The linear regression algorithm reveals a linear relationship between a dependent (y) variable and one or more independent (y) variables, thus the name. Because linear regression reveals a linear relationship, it determines how the value of the dependent variable changes as the value of the independent variable changes.

The link between the variables is represented by a slanted straight line in the linear regression model.

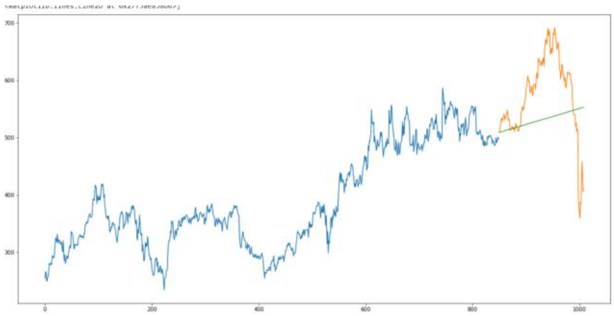
The independent variables are represented by x1, x2,....xn, and the weights are represented by the coefficients θ1, θ2, …. θn

$$Y = \theta_1 X_1 + \theta_2 X_2 + \dots \theta_n X_n$$

For TATA-GLOBAL



FOR NETFLIX



Linear regression is a straightforward technique that is easy to understand, but it has a few drawbacks. When employing regression techniques, one issue is that the model becomes overfit to the date and month columns. For cases like Big Mart sales, where the independent features are relevant for identifying the target value, a linear regression technique can perform well.

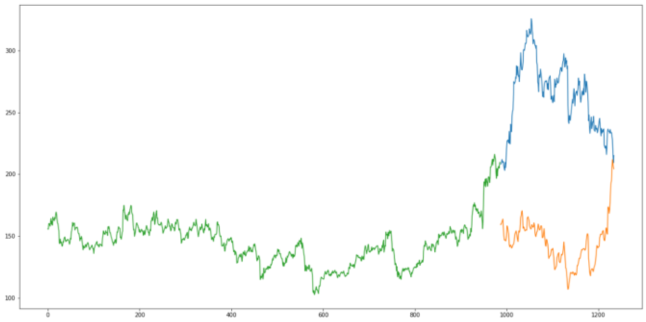
	NETFLIX	NSE - TATAGLOBAL
RMSE	83.3158	123.472
MSE	6941.527	15245.4637

VARIANCE SCORE	-0.0108	0.006

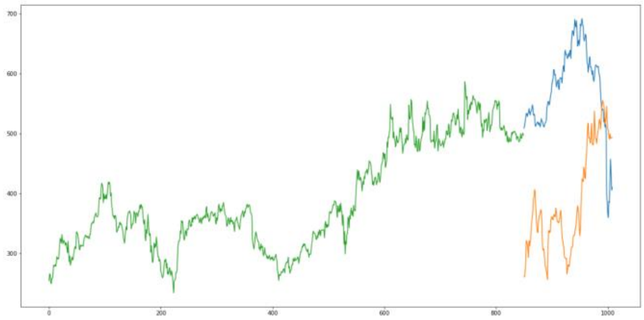
K Nearest Neighbors

The k-nearest neighbour algorithm is an example of instance-based learning used for classification in which the training data is maintained so that a classification for a new unclassified record may be obtained by comparing it to the most comparable records in the training set. The value of k specifies how many records from the training data set are used to categorise a record from the test data set.

For TATA-GLOBAL



For NETFLIX



The RMSE value does not differ significantly, but a plot of the projected and actual values should help to clarify things.

	NETFLIX	NSE - TATAGLOBAL
RMSE	221.4046	122.669
MSE	49020.0363	15047.753
VARIANCE SCORE	-1.9468	-0.4634

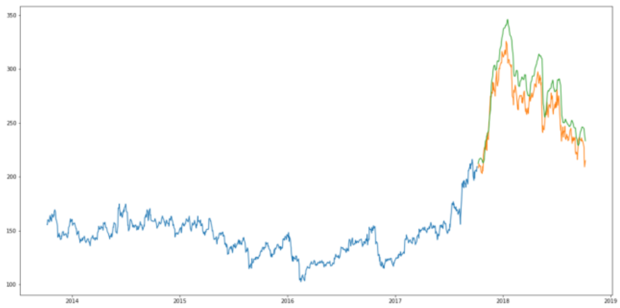
iv) Long Short Term Memory (LSTM)

Problems with sequence prediction have existed for a long time. They are regarded as one of the most difficult problems in the data science business to address. These difficulties span from estimating sales to identifying trends in stock market data, from comprehending movie plots to recognising your voice, from language translations to anticipating your next word on your iPhone's keypad.

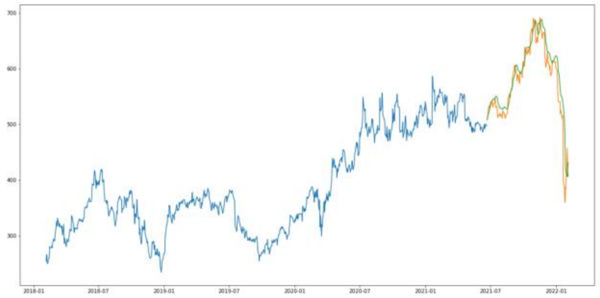
With the recent advances in data science, it has been discovered that Long Short Term Memory networks, often known as LSTMs, are the most effective answer for practically all of these sequence prediction problems.

In many areas, LSTMs outperform traditional feed-forward neural networks and RNNs. This is due to their ability to memorise patterns selectively over long periods of time.

For TATA-GLOBAL



For NETFLIX



	NETFLIX	NSE - TATAGLOBAL
RMSE	27.1809	18.9379
MSE	49020.0363	15047.7
VARIANCE SCORE	-1.9468	-0.47

V Conclusions and Future Works

Many investors all over the world are interested in stock investments. However, due to the multiple elements involved, choosing a selection is a difficult and complex undertaking. Investors are eager to foresee the stock market's future position in order to make successful investments. Small gains in predictive efficiency can be quite profitable. By giving supportive information such as the future direction of stock prices, a successful prediction system will assist investors in making more accurate and profitable investments. As a result, stock price forecasting is a critical activity that can be advantageous to investors. This research examined and contrasted the state-of-the-art of machine learning algorithms and methodologies used in finance, particularly stock price prediction. Various machine learning algorithms and methodologies have been examined in terms of input kinds, purposes, benefits, and drawbacks. Some machine learning algorithms and methodologies have been frequently chosen for stock price prediction due to their attributes, accuracy, and error acquired. Other information, such as politics, economic growth, financial news, and social media, may have an impact on the stock in addition to past pricing. Many research have shown that sentiment analysis has a significant influence on future prices. As a result, combining technical and fundamental studies could improve prediction efficiency, and it would be fascinating to include this into current ML research in the future.

VI References

[1] Luca Di Persio and Oleksandr Honchar, Recurrent neural networks approach to the financial forecast of Google assets, Volume 11, 2017

[2] Jaydip Sen, Stock Price Prediction Using Machine Learning and Deep Learning Frameworks, Proceedings of the Sixth International Conference on Business Analytics and Intelligence (ICBAI'2016)

[3] Mehar Vijh, Deekha Chandola, Vinay Anand Tikkiwal, Arun Kumar, Stock Closing Price Prediction using Machine Learning Techniques, International Conference on Computational Intelligence and Data Science (ICCIDS 2019)

[4] T. Kimoto; K. Asakawa; M. Yoda; M. Takeoka, Stock market prediction system with modular neural networks, Published: 28 November 2006 Stock market prediction with multiple classifiers Bo Qian & Khaled Rasheed

[5] Mehtabhorn Obthong1 a, Nongnuch Tantisantiwong2 b, Watthanasak Jeamwatthanachai1 c and Gary Wills, A Survey on Machine Learning for Stock Price Prediction: Algorithms and Techniques,

[6] Xiongwen Pang, Yanqiang Zhou, Pan Wang, Weiwei Lin & Victor Chang, An innovative neural network approach for stock market prediction, Published: 12 January 2018