ISSN No:-2456-2165

Stock Price Trend Forecasting Using Machine Learning

Ujwal C, Varun U K, Vishwas Gowda U, Manoj K Y
Computer Science and Engineering
SJB INSTITUTE OF TECHNOLOGY
Bengaluru, India

Kala Chandrashekar
Assistant Professor
Computer Science and Engineering
SJB INSTITUTE OF TECHNOLOGY
Bengaluru, India

Abstract:- This paper here presents how different algorithms can be implemented to find the stock price. Algorithms such as Linear Regression, KNN, Random Forest Regression, Elastic Net and LSTM model are implemented. The main aim of this paper is to find the trend of the stock i.e predict whether the price of the stock is going to increase or decrease the next day. The average of the predicted values are taken and the value is predicted. Since predicting the future value of the stock is highly dependent on various factors such as current trend, social media engagements, public involvement etc. Hence the best way to get the exact value of the stock is by predicting one day into the future.

Keywords:- *Stock Prediction, Machine Learning (ML), Regression*

I. INTRODUCTION

The stock market fluctuates rapidly, and there are numerous complex financial indicators. However, the Machine learning advancements provide an opportunity to profit consistently from the stock market and can also assist specialists in identifying the most useful signs to make better predictions. The ability to estimate market value is critical to maximise profit.

Financial models have been used by investment businesses, hedge funds, and even individuals to better understand market behaviour and make profitable investments and trades. Historical stock prices and corporate performance data provide a wealth of information ideal for machine learning algorithm to process. Even machine learning programs find it difficult to predict future pricing. There are simpler questions to answer that may provide benefits as well. Will tomorrow's closing price, for example, be higher than today's closing price?

Machine learning is based on data. The method entails identifying a desirable trade and then asking the machine learning model-fitting process to look for patterns in the data.

In comparison, rule-based trading systems were developed previously. This method entails computing some indicators and then waiting to see what occurs. Almost many of the systems that result are single decision trees. More complicated machine learning models always outperform simple regression models in machine learning contests and in trading.

II. METHODOLOGY

a) Data discretization: Here, after obtaining the datasets we are converting continuous data attribute values into a finite set of intervals, and associating each interval with a particular data value.

b) Data transformation: during this method, the dataset which is loaded into the project is converted into object format for pre-processing.

c) Data Cleaning: the information which is transformed is checked for errors. Redundant data is removed (Normalisation) from the dataset. The values which aren't required are deleted.

d) Data Integration: The data and the information after preprocessing is split into training and testing data for prediction, after which output is obtained.

ISSN No:-2456-2165

The following algorithms were used for the stock prediction:

➔ Linear Regression.
➔ Random Forests.
➔ K Nearest Neighbor (KNN).
➔ Elastic Net.
➔ LSTM Model.

A. Linear Regression:

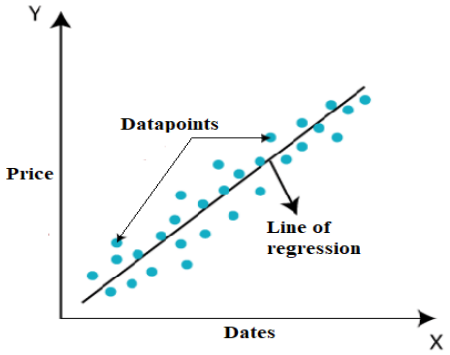

Fig 1. Representation of Linear Regression

Linear regression is a valuable tool for technical and quantitative analysis in financial markets because it examines two different variables to determine a single relationship.

Traders can detect when a stock is overbought or oversold by plotting stock values along a normal distribution (bell curve).

A trader can use linear regression to discover crucial price points such as entry, stop-loss, and exit prices. The system parameters for linear regression are determined by the price and time period of a stock, making the procedure generally applicable.

B. Random Forests:

Ensemble learning techniques are used to create random forests. Ensemble simply refers to a group or a collection, in this case a group of decision trees together referred to as a random forest. Ensemble models are more accurate than individual models since they combine the results of the many models to provide a final conclusion.

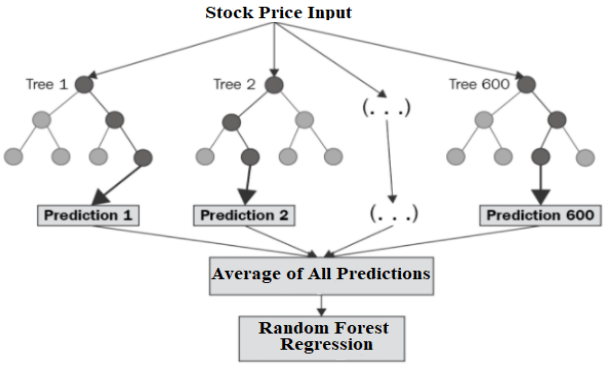

Fig 2. Representation of Random Forests Regression

A process called bootstrap aggregating or bagging is used to select features at random. A number of training subsets are formed from the dataset's collection of features by selecting random features with replacement. This means that a single feature may appear in many training subsets at the same time.

If a dataset comprises 20 features and subsets of 5 features are to be chosen at random to construct distinct decision trees, these 5 features will be chosen at random, and every feature can be part of more than one subset. This assures unpredictability, reducing the correlation between the trees and hence preventing overfitting.

The trees are built based on the best split after the features have been chosen. Each tree produces an output, which is regarded as a "vote" from that tree in favour of that output. The random forest chooses the final output/result that receives the most "votes," or in the case of continuous variables, the average of all the outputs is considered the final output.

C. K Nearest Neighbor (KNN):

The KNN algorithm is a simple supervised machine learning technique that may be used to handle classification and regression problems. k-NN is a sort of instance-based learning, often known as lazy learning, in which the function is only approximated locally and all computation is postponed until after the function has been evaluated.

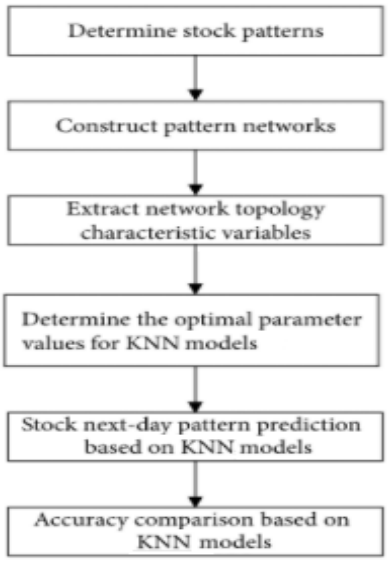

Fig 3. Representation of KNN Regression

The following is how kNN is used to estimate the closing price of a stock market:
a) Calculate the distance between the training samples and the query record, k.
b) Calculate k, the number of nearest neighbours.
c) Sort all of the training records by distance values.
d) For the class labels of k nearest neighbours, use a majority vote and assign it as the query record's prediction value.

D. Elastic Net:

Elastic net is a sort of regularised linear regression that includes two well-known penalties, the L1 and L2 penalty functions. This connection is a line with a single input variable, and it can be thought of as a hyperplane with greater dimensions that connects the input variables to the destination variable. The model's coefficients are determined by an optimization procedure that aims to reduce the total squared error between the predicted and expected goal values.

The advantage is that elastic net permits a balance of both penalties, which might result in greater performance on particular tasks than a model with only one penalty. The weighting of the sum of both penalties to the loss function is controlled by another hyperparameter called "lambda."

The fully weighted penalty is applied by default with a value of 1.0; the penalty is not applied with a value of 0. Lambada values as low as 1e-3 or even below are rather typical.

E. LSTM Model:

Long short-term memory (LSTM) is a deep learning architecture that uses an artificial recurrent neural network (RNN). LSTM has feedback connections, unlike normal feedforward neural networks. It can process not only single data points, but also complete data sequences.

Because they can store past information, LSTMs are extremely useful in sequence prediction issues. This is significant in our situation since a stock's historical price is critical in determining its future price.

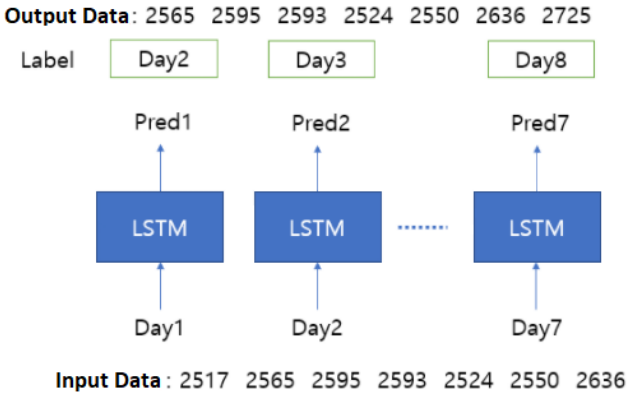

Fig 4. Representation of LSTM Regression

The LSTM model is trained on the full dataset, and a new dataset is obtained for testing purposes for the next month. The already trained LSTM model will estimate stock prices for this additional duration, and the predicted prices will be plotted against the original values to show the model's accuracy.

III. FLOWCHART

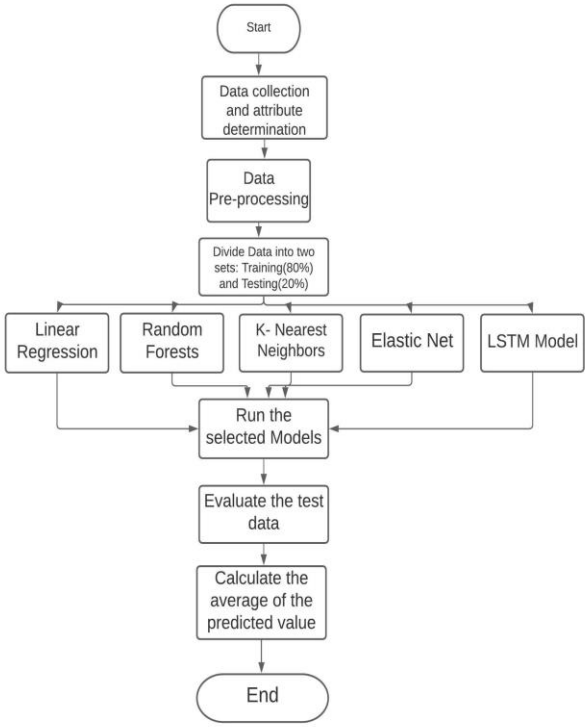

Fig 5. Flow Chart of the System

IV. RESULTS

This model was successfully built using different regression algorithms to predict the price of the stock for the next day.

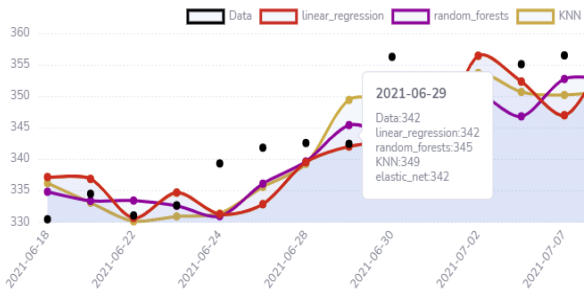

Fig 6. Snapshot of the project using different algorithms

The main advantage of this method is that the average of all the models is taken, as each and every model has its own advantages and disadvantages. The best way to minimise the error is to take the average.

The accuracy of the model build is determined using the mean error function, wherein the average of the difference between the actual value and the predicted value is taken.

Model	Mean Error
linear_regression	3.1125724
random_forests	1.3921287632353148
KNN	2.2208025
elastic_net	3.1096227
LSTM_model	3.7951625392675723

Fig 7. Accuracy of the models

V. CONCLUSION

Finally, stock price trend forecasting is used to forecast the direction of financial movement. For financial forecasting, regression is a promising method. Each approach, though, has its own set of advantages and disadvantages. The indicator function used can have a significant impact on the prediction system's accuracy.

Also, a certain Machine Learning Algorithm may be better suited to a specific type of stock, but the same algorithm may forecast other sorts of stocks with lesser accuracy.

REFERENCES

[1]. J. Jagwani, M. Gupta, H. Sachdeva and A. Singhal, "Stock Price Forecasting Using Data from Yahoo Finance and Analysing Seasonal and Nonseasonal Trend," 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS), 2018, pp. 462-467, doi: 10.1109/ICCONS.2018.8663035.

[2]. J. S. Park, H. Sung Cho, J. Sung Lee, K. I. Chung, J. M. Kim and D. J. Kim, "Forecasting Daily Stock Trends Using Random Forest Optimization," 2019 International Conference on Information and Communication Technology Convergence (ICTC), 2019, pp. 1152-1155, doi: 10.1109/ICTC46691.2019.8939729.

[3]. N. Powell, S. Y. Foo and M. Weatherspoon, "Supervised and Unsupervised Methods for Stock Trend Forecasting," 2008 40th Southeastern Symposium on System Theory (SSST), 2008, pp. 203-205, doi: 10.1109/SSST.2008.4480220.

[4]. S. Yao, L. Luo and H. Peng, "High-Frequency Stock Trend Forecast Using LSTM Model," 2018 13th International Conference on Computer Science & Education (ICCSE), 2018, pp. 1-4, doi: 10.1109/ICCSE.2018.8468703.
[5]. S. Dinesh, N. R. R, S. P. Anusha and S. R, "Prediction of Trends in Stock Market using Moving Averages and Machine Learning," 2021 6th International Conference for Convergence in Technology (I2CT), 2021, pp. 1-5, doi: 10.1109/I2CT51068.2021.9418097.
[6]. K. C, S. A. N. A, A. P, P. R, M. D and V. B. D, "Predicting Stock Prices Using Machine Learning Techniques," 2021 6th International Conference on Inventive Computation Technologies (ICICT), 2021, pp. 1-5, doi: 10.1109/ICICT50816.2021.9358537.
[7]. S. Vazirani, A. Sharma and P. Sharma, "Analysis of various machine learning algorithm and hybrid model for stock market prediction using python," 2020 International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE), 2020, pp. 203-207, doi: 10.1109/ICSTCEE49637.2020.9276859.
[8]. I. Parmar et al., "Stock Market Prediction Using Machine Learning," 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC), 2018, pp. 574-576, doi: 10.1109/ICSCCC.2018.8703332.