STOCK MARKET PREDICTION USING MACHINE LEARNING

Shubhangi Kadam[1], Shresth Jain[2], Shrikant Kamineni[3], Shweta Narute[4], Prof. Mohan Pawar[5]
[1,2,3,4]*Student, MIT School of Engineering, Pune*
[5]*Assistant Professor, MIT School of Engineering, MIT-ADT University, Pune*

***Abstract*- The goal of Stock Market Prediction is to forecast the closing prices of the firm Bajaj-Finance on a given day using machine learning algorithms such as Linear Regression and Support Vector Machine. The company's historical and real-time data (close, high, low, volume, open, and Adj-Close) were used to make the prediction. The results of the investigation revealed that machine learning algorithms were capable of producing very accurate forecasting results.**

The use of machine learning to produce predictions that support the current stock closing price by training on historical values is a recent trend in available stock market prediction technologies. Machine Learning makes use of a variety of models to make accurate predictions. Stock trading is one of the most essential activities in the financial sector. The act of attempting to anticipate the long-term future value of a stock or other financial instrument traded on a financial exchange is known as the stock market prediction. The prediction of a stock using Machine Learning is demonstrated in this study. Most stockbrokers use technical and fundamental analysis, also known as static analysis when making stock predictions. Python is the recommended programming language for applying machine learning to anticipate the stock market. In this research, we present a Machine Learning approach that will be taught using publicly available stock data to build intelligence and then use that intelligence to make an accurate prediction. For shorter base period lengths and forecast horizons, all algorithms performed better.

***Index Terms*- Stock market prediction, Linear Regression, Support Vector Machine, and Support Vector Regression are some of the terms used in this paper.**

I. INTRODUCTION

Stock market forecasting is an important part of investment theory and practice, especially with the advancements in automated trading systems for capital markets. The stock market is known for being dynamic, unpredictable, and non-linear. Predicting stock prices is difficult because they are influenced by a variety of factors such as political events, global economic conditions, corporate financial reports and performance, and so on. Thus, in order to maximize profits and avoid losses, strategies for predicting stock values in advance by examining the pattern over the preceding few years could be quite valuable for generating stock exchange movements. For predicting a corporation's stock price, two basic methodologies have been offered in the past. For estimating the stock's longer-term price, technical analysts examine previous stock prices such as closing and opening prices, the volume traded, adjacent close values, and so on. Despite the fact that the trend during a stock market projection isn't a replacement item, this topic continues to be debated by numerous organizations. There are two forms of stock research that investors conduct before investing in a stock. The first is elemental analysis, in which investors examine the intrinsic worth of stocks as well as the performance of the industry, economy, and political atmosphere to determine whether to invest or not. Technical analysis, on the other hand, is the examination of statistics created by market activity, such as historical prices and volumes, to determine the evolution of stocks.

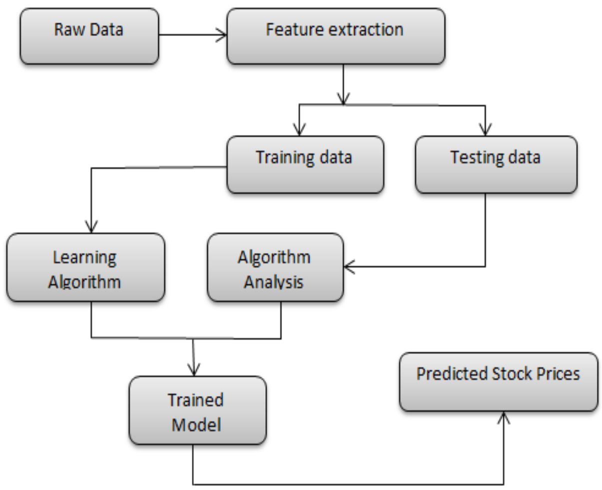
II. LITERATURE SURVEY

Many significant changes have occurred in the financial market environment over the last 20 years. The expansion of effective communication and trading services has broadened the range of options available to investors.

[1] The authors Enke, D., and Thawornwong, S., described a system for forecasting stock exchange returns that included data processing approaches with neural networks. They used the variable relevance analysis technique in machine learning for data mining to explore the prediction capacity of financial and economic variables in this study. The authors looked at how well neural network models for level estimation and categorization worked.

[2] Tsong-Wuu Lin and Chan-Chien Yu conducted a study to see if using artificial neural networks was profitable (ANNs). They've turned forecasts into a basic trading technique, and its profitability is compared to a buy-and-hold method. They use a neural network to track the Taiwan Weighted Index and the S& P 500 in the United States. They discovered that a trading rule based on ANNs delivers superior returns than a buy-and-hold strategy in this study.
[3] 'Stock Closing Price Prediction Using Machine Learning Techniques,' according to the research article. Artificial Neural Networks and Random Forest algorithms were used to forecast the next-day closing price of five companies operating in various industries. The authors, Mehar Vijh, Deeksha Chandola, Vinay Anand Tikkiwal, and Arun Kumar, used financial data such as stock prices' Open, High, Low, and Close to create new variables that were used as inputs to the model. Standard strategic metrics such as RMSE and MAPE are used to assess the models. The low levels of those two indicators indicate that the models are effective at predicting the value of a company's shares.
[4] Sharma, Ashish, Bhuriya, Dinesh, and Singh, Upendra. "Analysis of stock market forecasting using machine learning." Electronics, communication, and aeronautical technology: an international conference in 2017 (ICECA). IEEE, 2017. Vol. 2. They examined a well-known efficient regression strategy for predicting stock market price using stock market data in this research. The findings of the multiple regression approach could be improved in the future by including more factors.
[5] Mehak Usmani et al., "Stock market forecast using machine learning approaches," 3rd international conference on computer and knowledge sciences, 2016. (ICCOINS). 2016 IEEE. The major goal of this study is to use several machine learning approaches to anticipate the market performance of the Karachi stock exchange (KSE) on a daily close. The model takes some features as input and forecasts whether the market will be positive or negative. Oil rates, gold and silver prices, interest rate, foreign exchange (FEX) rate, NEWS, and social media feed are among the attributes considered in the model. Simple Moving Average (SMA) and Autoregressive Integrated Moving Average (ARIMA) are two old statistical techniques that are utilized as input. Single Layer Perceptron (SLP), Multi-Layer Perceptron (MLP), Radial Basis Function (RBF), and Support Vector Machine (SVM) machine learning approaches are contrasted.
[6] Hegazy, Osman, Omar S. Soliman, and Mustafa Abdul Salam. "A stock market prediction model based on machine learning." arXiv preprint arXiv:1402.7351 (2014). To anticipate stock market prices, this study used a machine learning algorithm. Particle swarm optimization (PSO) and least square support vector machine performance are improved by the proposed technique (LS-SVM). To anticipate daily stock prices, the PSO algorithm is employed to optimize LS-SVM. They offer a proposed model based on historical data and technical indicators for stocks. To avoid overfitting and local minima concerns and enhance prediction accuracy, the PSO algorithm determines the optimal free parameter combination for LS-SVM. The suggested model was compared against an artificial neural network utilizing the Levenberg-Marquardt (LM) algorithm using thirteen benchmark financial datasets. The findings revealed that the suggested model has a higher prediction accuracy and that the PSO method has the ability to optimize LS-SVM.

III. PROPOSED SYSTEM



Above is the planned system for this project. First, we took raw data from a company called Bajaj Finance Limited. By extracting the data's characteristics, such as Open, Close, High, Low, Volume, and Adj Close. Our data has been divided into two datasets: train and test. Then, using Support Vector Regression, we trained our model with several kernels, analyzed the dataset using testing data, and plotted the graph. The model can then display the stock's anticipated closing values because it is trained with three kernels: RBF, polynomial, and linear. For linear regression, the same proposed system was employed.

We started with pre-processing by checking for missing values, scaling, and standardizing the dataset for linear regression. The model was then trained and analyzed by evaluating the learned data using an algorithm after partitioning the dataset into train and test sets. After that, we have the exact closing projected values.

IV. METHODOLOGY AND ALGORITHMS

The following algorithms are required to implement the Stock Prediction model:

1. Regression Linear

Linear regression is the simplest basic machine learning approach that may be used with this data. The linear regression model gives you an equation that shows how the independent variables are related.

The linear regression equation is written as follows:

a)

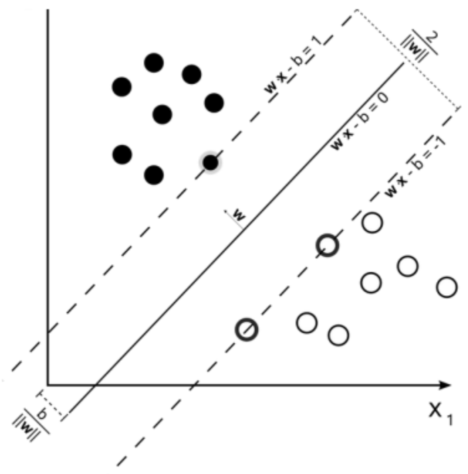
$$Y = \theta_1 X_1 + \theta_2 X_2 + \ldots \theta_n X_n$$

The independent variables are represented by x1, x2,....xn, and the coefficients are 1, 2,... The weights are represented by the number n.

Linear regression is a supervised learning machine learning algorithm. It carries out a regression task. Based on independent variables, regression models a goal prediction value. It is mostly utilized in forecasting and determining the link between variables. Different regression models differ in terms of the type of relationship they evaluate between dependent and independent variables and the number of independent variables they employ. Linear regression is used to predict the value of a dependent variable (y) based on the value of an independent variable (x). As a result of this regression technique, a linear relationship between x (input) and y (output) is discovered (output). As a result, the term Linear Regression was coined.

2.Support Vector Machine

The "Support Vector Machine" (SVM) is a supervised machine learning technique that can be used to solve classification and regression problems. It is, however, mostly employed to solve categorization difficulties. We represent each data item as an extended point in n-dimensional space (where n is the number of features you have), with the value of each feature being the value of a certain coordinate in the SVM algorithm. Then we accomplish classification by locating the hyper-plane that clearly distinguishes the two classes (look at the below snapshot).



SVM is useful since it can perform both classification and regression.

Because the output is a real number, it becomes extremely difficult to forecast the information at hand, which has an endless number of possibilities. A margin of tolerance (epsilon) is approximately in approximation to the SVM which may have already been sought from the matter in the event of regression.But, aside from this, there is a more difficult reason: the algorithm is more complicated, thus it must be taken into account. The general concept is to decrease mistakes by individualizing the hyper-plane that maximizes the margin while keeping in mind that some error is tolerable.

To execute the linear separation, the kernel functions translate the data information into a higher dimensional feature space.

b)

$$y = \sum_{i=1}^{N} (\alpha_i - \alpha_i^*) \cdot \langle \varphi(x_i), \varphi(x) \rangle + b$$

$$y = \sum_{i=1}^{N} (\alpha_i - \alpha_i^*) \cdot K(x_i, x) + b$$

We utilized the RBF kernel (Radial Basis Function) A radial basis function is a real-valued function whose value is solely determined by the distance from the origin, or alternatively, the distance from some other point referred to as a center. A radial function is any function that satisfies the property. RBF stands for "Response Function Local."

c) There are several versions of the kernel function K(xi, xj):

c.1) Linear kernel: K(xi, xj) = xTi xj, K(xi, xj) = xTi xj, K(xi, xj) = xTi

c.2) Degree d polynomial kernel: K(xi, xj) = (1 + xTi xj/c)d,

c.3) RBF kernel: K(xi, xj) = exp(kxi xjk2/2),

c.4) MLP kernel: K(xi, xj) = tanh(kxTi xj +),

d, c, k, and k are constants. The linear kernel, we note, corresponds to the

(x) = x is a linear function.

KRBF (x, x') = exp [-y ||x -x'||2]

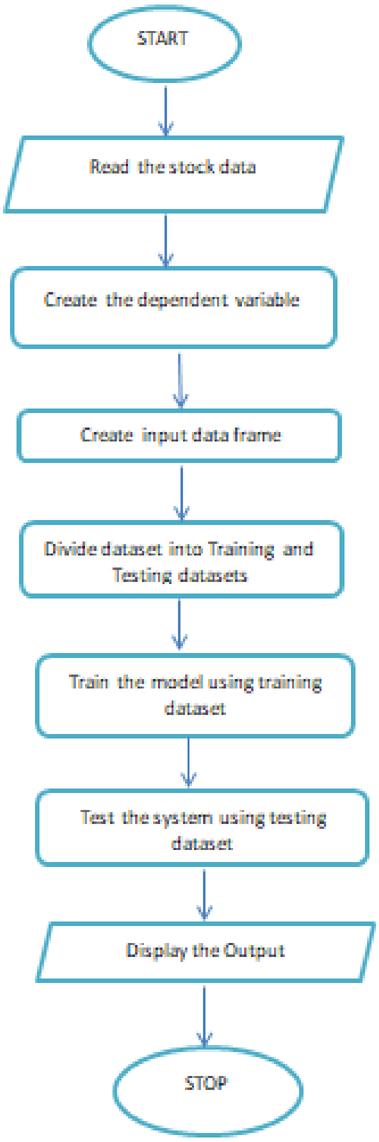
The RBF Kernel is just a low-band pass filter, which is commonly used in Signal Processing to smooth pictures. The RBF Kernel operates as a prior, weeding out non-smooth solutions.

d)

Gaussian Radial Basis function

$$k(\mathrm{x}_i, \mathrm{x}_j) = \exp\left(-\frac{\left\| \mathrm{x}_i - \mathrm{x}_j \right\|^2}{2\sigma^2} \right)$$

V. FLOWCHART



VI. OUTPUT

We have a dataset for the company “BajFinance.NS” here. Yahoo Finance provided the most up-to-date information. The data spans a ten-year period, from January 1, 2010, to November 6, 2021, when BajajFinance was founded. The data includes stock information such as the high, low, open, volume, close, and adjacent close. Only the stock's day-by-day closing price has been extracted.

1) Describing dataset:

1.1) The head data of our dataset

Date	High	Low	Open	Close	Volume	Adj Close
2010-01-04	33.509838	32.324509	32.324509	33.334953	466064.0	30.912809
2010-01-05	34.073353	32.645130	34.010201	33.111488	318779.0	30.705582
2010-01-06	34.102501	32.567402	33.130920	33.587563	192881.0	31.147062
2010-01-07	33.995628	32.936604	33.995628	33.383530	138876.0	30.957851
2010-01-08	34.879765	33.271797	33.432110	33.990768	312696.0	31.520975

1.2) The tail of our dataset

Date	High	Low	Open	Close	Volume	Adj Close
2021-06-07	5843.850098	5682.450195	5800.000000	5729.750000	2835805.0	5729.750000
2021-06-08	5817.000000	5720.149902	5751.000000	5755.399902	1789007.0	5755.399902
2021-06-09	5795.000000	5627.299805	5795.000000	5674.500000	1538112.0	5674.500000
2021-06-10	6115.000000	5677.000000	5715.000000	6086.399902	5296120.0	6086.399902
2021-06-11	6230.000000	6045.200195	6105.350098	6120.000000	3702721.0	6120.000000

Based on the current Adjusted Close price, we will forecast the price of a stock for the next 60 days. Because we only require the Adjusted Close (Adj. Close) price, I'm only pulling data from the column 'Adj Close' and storing it in the variable 'df.' The first primary 5 rows of the new data set are then printed.

1.3)

Date	Adj Close
2010-01-04	30.912809
2010-01-05	30.705582
2010-01-06	31.147062
2010-01-07	30.957851
2010-01-08	31.520975

Here we have created a variable to store the number of days into the future we want to predict got the new data like

1.4)

Date	Adj Close	Prediction
2021-06-07	5729.750000	NaN
2021-06-08	5755.399902	NaN
2021-06-09	5674.500000	NaN
2021-06-10	6086.399902	NaN
2021-06-11	6120.000000	NaN

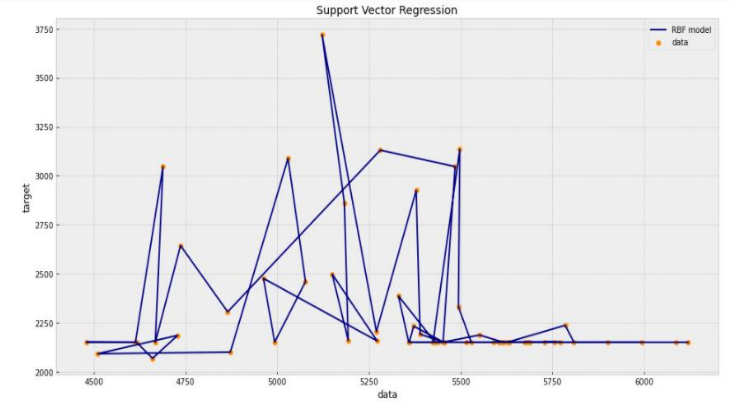
2) Using Support Vector Regression

2.1) Accuracy got by Support Vector Machine

```
svm rbf confidence:   0.7292161839316809
svm linear confidence:   0.530656651366114
svm polynomial confidence:   0.633596778618
```

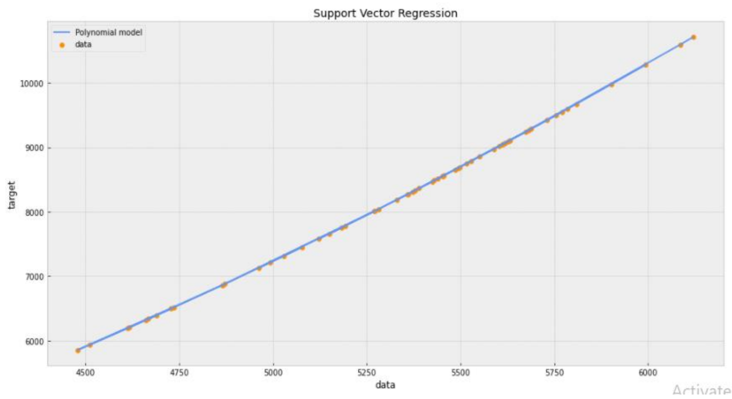
2.2) Predicted values and plot graph got by using RBF kernel

```
[2150.38686775 2150.38686775 2150.38686775 2232.47266332 2150
 2191.47711163 2924.74804123 2204.60356721 3718.97612392 2860
 2158.93230961 2498.14432643 2157.26324935 2475.9531435  2150
 2459.53364753 3088.42151332 2099.9601052  2091.79066942 2185
 2068.34910823 2150.38686775 2151.2492293  2150.38686775 3046
 2150.38654776 2643.44504414 2305.39331763 3130.27448941 3046
 2150.38686775 2186.92962346 2150.38686775 2150.41157046 2150
 2330.94583344 3134.83050817 2150.4771108  2150.38686775 2386
 2150.38686789 2150.38686775 2150.38686775 2150.3897167  2150
 2150.38686775 2150.38686775 2150.90290074 2150.38686775 2150
 2150.38686775 2236.68390869 2150.38686775 2150.38686775 2150
 2150.38686775 2150.38686775 2150.38686775 2150.38686775 2150
```



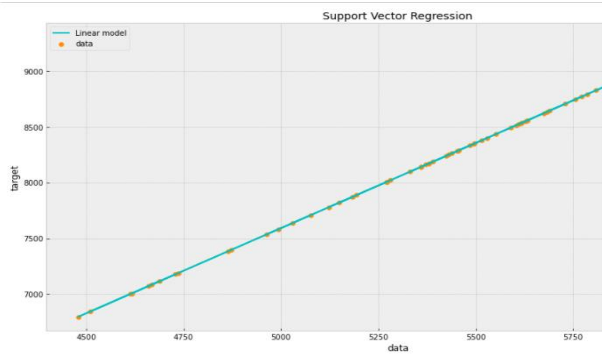
2.3) Predicted values and plot graph got by using Polynomial kernel

```
[ 8515.08311897  8566.55068239  8275.86922811  8312.0163508
  8558.31693154  8364.65544296  8332.90021436  8009.9375950
  7579.72454778  7755.00196202  7785.45663845  7659.1135548
  8015.38946922  7129.55297167  7214.70595869  7451.1492638
  7316.68014578  6881.5923977   5933.97516361  6496.3527654
  6316.71155798  6203.39561242  5855.06615099  6195.7840491
  6390.92782469  6336.26683937  6516.98887067  6861.8596119
  8041.19992489  8654.93090622  8554.20158459  8859.4498644
  9061.08927951  8748.09362311  8790.2306215   8682.3951336
  8693.76311447  8471.5089139   8271.67522048  8187.8709032
  8487.89281714  9291.75542018  9109.66042847  8976.5291039
  9111.39132152  9020.79376971  9070.19559899  9555.7026767
  9270.93762467  9045.39924076  9099.43224259  9599.7931558
  9677.08122814  9980.99784807 10286.71859723  9422.6141028
  9504.95811333  9246.4941167  10600.56647867 10715.2068977
```

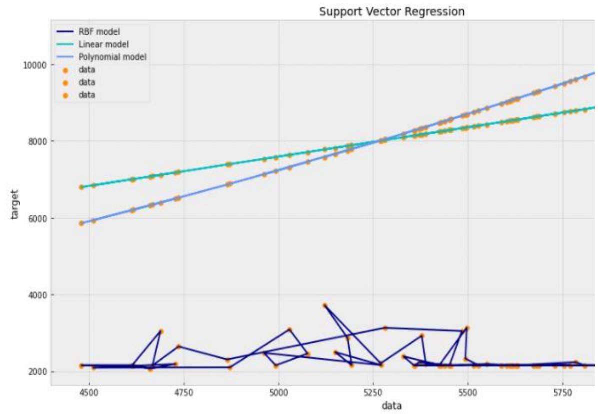


2.4) Predicted values and plot graph got by using Linear kernel

```
[8262.03015726 8287.85037224 8140.95131879 8159.36162596 8283.7246
 8186.09774809 8169.97960859 8004.21216016 7777.9436913  7870.9107
 7886.95266538 7820.18817167 8007.03846531 7533.95239549 7580.7027
 7709.03873091 7636.3154326  7396.14327436 6844.14593797 7176.8265
 7072.2473393  7005.48241835 6796.17220717 7000.97596531 7115.6386
 7083.70604139 7188.74274259 7385.06674009 8020.4072588  8332.0042
 8281.66284352 8433.29762058 8531.99423127 8378.29630833 8399.1515
 8345.67812555 8351.33082109 8240.10593026 8138.81253987 8095.9572
 8248.3560832  8643.52421479 8555.59836953 8490.74345308 8556.4390
 8512.36218434 8536.42505436 8769.41518955 8633.51679604 8524.3555
 8550.63366692 8790.27086942 8826.70895939 8968.56568437 9109.0478
 8706.16439729 8745.35297566 8621.75316911 9251.0583437  9302.3924
```



2.5) Merging the graphs of all the kernels we can see the plot graph as follow



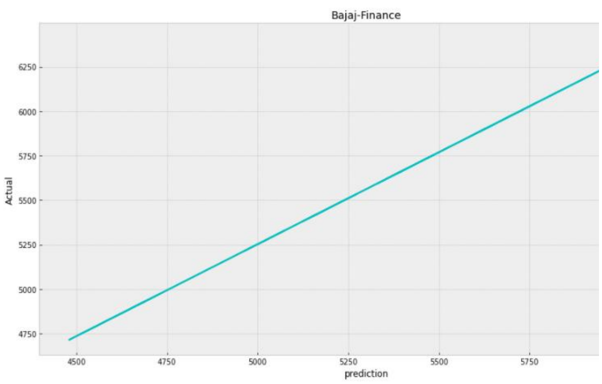
3) Linear Regression

Linear Regression confidence for 60 days is:

lr confidence: 0.8911011569583701

3.1) Linear Regression prediction values and graph for 60 days

```
[5706.3566848  5723.84502737 5624.34963661 5636.8
 5654.92796216 5644.01063239 5531.73599859 5378.4
 5452.31592353 5407.09593639 5533.65045976 5213.2
 5331.81501834 5282.55894891 5119.88987645 4746.0
 4900.5145551  4855.29456797 4713.5284405  4852.2
 4908.27547449 4979.41723554 5112.38765437 5542.7
 5719.65382774 5822.35666096 5889.20414767 5785.1
 5763.01139633 5766.84031866 5691.50786325 5622.9
 5697.09562416 5964.74376304 5905.19184371 5861.2
 5875.90700473 5892.20493544 6050.01029631 5957.9
 5901.82877864 6064.13506855 6088.81489325 6184.8
 6007.17012235 6033.7123656  5949.9980162  6376.2
```



4) Linear Regression prediction for the dataset of 3 years i.e from 2018 to 14-06-2021 have the accuracy as

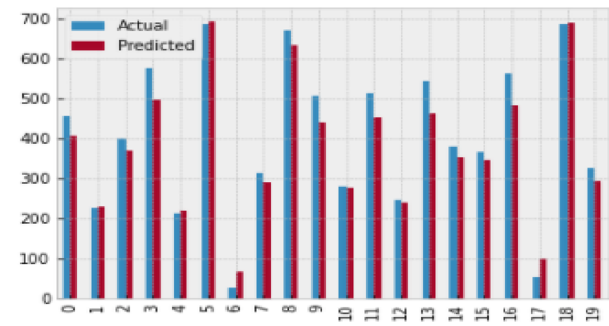
lr confidence: 0.9564746425858983

4.1) We can see here the actual and predicted values difference:

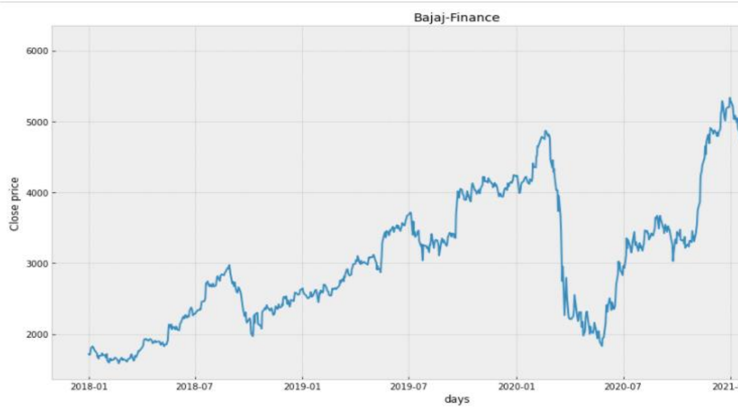
	Actual	Predicted
0	456	405.656215
1	225	229.575464
2	400	371.428570
3	577	498.391810
4	212	218.998840
5	687	692.861727
6	26	65.994621
7	312	289.490576
8	672	632.801804
9	507	438.939840
10	281	274.766540
11	514	452.926363
12	246	238.871702
13	545	462.384752
14	381	354.588252
15	365	347.423768
16	563	483.314580
17	53	97.301300
18	688	690.670706
19	327	294.343804

4.2) Bar Graph for Linear Regression

<AxesSubplot:>



4.3) The plot graph for our model is given below:



VII. CONCLUSION

Predicting stock market closing price returns is a difficult undertaking since stock values are always changing and are based on various parameters that follow complex patterns. We discovered that we can apply machine learning to anticipate and compare stock market prices in this article. The outcome demonstrates how historical data can be used to anticipate stock movement with reasonable accuracy.

VIII. ACKNOWLEDGEMENT

We would like to express our sincere gratitude to our Mentor Mr MOHAN PAWAR, Assistant Professor in Department of Computer Science Engineering, MIT ADT UNIVERSITY, Pune for his guidance, Encouragement and valuable suggestion. We would Also like to thank faculty of MIT ADT, Pune who helped Us to complete our project successfully.

REFERENCES

[1] Aparna Nayak, M. M. Manohara Pai, Radhika M.Pai “Prediction Models for Indian Stock Market” (2016) https://doi.org/10.1016/j.procs.2016.06.096

[2] S. D. Patel, D. Quadros, V. Patil, M. Pawale and Harsha Saxena, "Stock prediction using neural networks", *Int. J. Eng. Manag. Res.*, vol. 7, no. 2, pp. 490-493, 2017.

[3] Nisha Shetty, Ashish Pathak “Indian Stock Market Prediction Using Machine Learning and Sentiment Analysis” , December 2017 DOI:10.1007/978-981-10-8055-5_53

[4] Yasen Rajhy, Akeel Abdulkareem Alsakaa “ Back-propagation artificial neural networks in stock market forecasting application to the Warsaw Stock Exchange WIG20, January 2017 DOI: 10.5605/IEB.15.5

[5] Strader, Troy J.; Rozycki, John J.; ROOT, THOMAS H.; and Huang, Yu-Hsiang (John) (2020) "Machine Learning Stock Market Prediction Studies: Review and Research Directions," Journal of International Technology and Information Management: Vol. 28 : Iss. 4 , Article 3.

[6] Liheng Zhang, Charu Aggarwal, and Guo-Jun Qi. 2017. Stock Price Prediction via Discovering Multi-Frequency Trading Patterns. In Proceedings of KDD ’17, Halifax, NS, Canada, August 13-17, 2017, 9 pages. https://doi.org/10.1145/3097983.3098117

[7] Faten Subhi Alzazah and Xiaochun Cheng “Recent Advances in Stock Market Prediction Using Text Mining: A Survey” June 1st 2020, DOI: 10.5772/intechopen.92253

[8] Sidra Mehtab, Jaydip Sen, Abhishek Dutta “Stock Price Prediction Using Machine Learning and LSTM- Based Deep Learning Models” September 2020 cited as:arXiv:2009.10819v1

[9] K. Hiba Sadia, Aditya Sharma, Adarsh Paul, SarmisthaPadhi, Saurav Sanyal “Stock Market Prediction Using Machine Learning Algorithms” (IJEAT) ISSN: 2249 – 8958, Volume-8 Issue-4, April 2019.

[10] Kranthi Sai Reddy Vanukuru “Stock Market Prediction Using Machine Learning” November 2018 DOI:10.13140/RG.2.2.12300.77448

[11] Weng, B., Ahmed, M. A., & Megahed, F. M. (2017).” Stock market one-day ahead movement prediction using disparate data sources. Expert Systems with Applications”,79, 153-163. doi:10.1016/j.eswa.2017.02.041

[12] Li, X., Huang, X., Deng, X., & Zhu, S. (2014). “Enhancing quantitative intra-day stock return prediction by integrating both market news and stock prices information. Neurocomputing”, 142, 228-238. doi:10.1016/j.neucom.2014.04.043