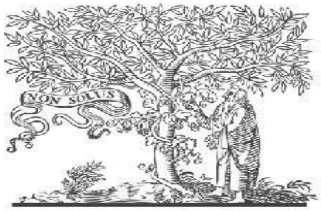
International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

COPY RIGHT



ELSEVIER SSRN

2023 IJIEMR. Personal use of this material is permitted. Permission from IJIEMR must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. No Reprint should be done to this paper, all copy right is authenticated to Paper Authors

IJIEMR Transactions, online available on 07th Sept 2023. Link :http://www.ijiemr.org/downloads.php?vol=Volume-12&issue=Issue 09

10.48047/IJIEMR/V12/ISSUE 09/20

Title **A study on Stock Market Price Analysis**

Volume 12, ISSUE 09, Pages: 179-185

Paper Authors **Dr. UdayaSri Kompalli, Sudhir K**

USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper As Per UGC Guidelines We Are Providing A Electronic Bar Code

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

A study on Stock Market Price Analysis

Dr. UdayaSri Kompalli
Dept. of Computer Science P.B. Siddhartha College of Arts & Science
Vijayawada, AP, India
kudayasri@pbsiddhartha.ac.in
Sudhir K
Dept. of Computer Science P.B. Siddhartha College of Arts & Science
Vijayawada, AP, India
ksudhir@pbsiddhartha.ac.in

***Abstract*—Machine Learning plays an important role in the society. It is increasing the dependency on the ML techniques to increase the efficiency and accurate decision making on investments. One of the places where the smart decision is to be applied is the Stocks. Stocks are the main instruments of investment in the market which are helpful to the society and also to the individual. There is a saying that "Right Decision at Right Time" will effect the financial position with in no time. If wrong decision is taken this may lead to problems. A simple Calculation of predicting the price fluctuations in the Stock Market is the main smart decision to follow. So, for this situation, in this study we are trying to apply the Machine Learning Techniques for correct guidance in decision Making. As this is related to Time series data, ie., recording the previous day to day transaction and making the machine learn the things and predict for the future dates. . In this paper we propose a Machine Learning (ML) approach that will be trained from the available stocks data and gain intelligence and then uses the acquired knowledge for an accurate prediction. In this context this study uses a machine learning technique called Support Vector Machine (SVM) to predict stock prices.**

Keywords—Time Series, Machine Learning, Stock Market, Prediction, Logistic Regression, Support Vector Machine, XG Boost.

I. INTRODUCTION

Stock Markets play an important role in the society to build the financial status. Prediction of the Stock Market is a challenging task. Due to the fluctuating nature of the stock, the stock market is too difficult to predict. Stock prices are fluctuating every day. Estimating of the stock market has a high demand for investors. The prediction of the stock market is a costliest game play where an individual can get profit or even too much of loss with wrong decision. In this paper we are trying to design a model which will support the investor to take right decision with Machine Learning Algorithms, Machine Learning is a branch of Artificial Intelligence which will make the machine to learn the things by its own on historic data and prepare a model which is able to predict the future basing on the past data. As the stock market plays vital role in building the financial position of the society.

Machine Learning is the field of study that gives computers the capability to learn without being explicitly programmed. ML is one of the most exciting technologies that one would have ever come across. As it is evident from the name, it gives the computer that makes it more similar to humans: The ability to learn.

II. LITERATURE SURVEY

[1] In this paper the author has discussed about the stock market analysis using linear regression and SVM with time series data. The data was taken from the Coca-Cola company and compared the results of the both linear regression and SVM. An overview of all the techniques is discussed in this paper by the author.
[3] In this paper the author have analysed the stock market price using Linear Regression on companies with S&G 500 index. The analysis was taken from real time data and applied VIF , Linear Regression and also R2 and RMSE [2] In this paper the author have implemented different algorithms for measuring the accuracy and found that the logistic regression algorithm is ideal for predicting the market price of a stock based on

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

various data points from the historical data. The algorithm will be a great asset for brokers and investors for investing money in the stock market since.

III. PROPOSED METHODOLOGY

Machine learning is actively being used today, perhaps in many more places than one would expect. It is the
learning and building of algorithms that can learn from and make predictions on data sets that are provided to the model to learn There are two branches of machine learning.

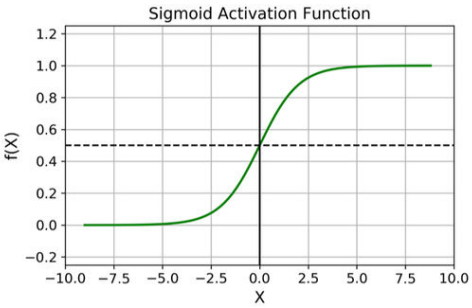
1. Supervised Learning
2. Unsupervised Learning

Supervised machine Learning on a predefined set of training samples, which then facilitate its ability to reach an accurate conclusion when given new data. When the data is divided or classified based on the class manually and let the model learn from the given data which is known. Unsupervised machine learning bunch of data and must find path there in. When the data is unclassified the model itself learns the path and pattern from the given data and predicts the results.

In this paper we are trying to apply three algorithms and evaluate the best model to predict with highest accuracy. We applied three algorithms as Logistic Regression, Support Vector Classifier and XGBClassifier.

Logistic Regression

Logistic Regression is a statistical procedure for predicting the value of a dependent variable from an independent variable when the relationship between the variables can be described with a discrete data. Logistic regression is a fundamental classification technique that belongs to the group of linear classifiers and is similar to polynomial and linear regression. Logistic regression is fast and relatively uncomplicated, to interpret the results. As it's a method for binary classification, it can also be applied to multiclass problems. To implement Logistic Regression one must have an idea about sigmoid curve. Where linear regression is a straight line passing through x and y axis Sigmoid curve shows the data as either 0 or 1. The logistic regression of some dependent variable x on the set of independent variables x = (x_1 , …, x_r), where r is the number of predictors (or inputs), start with the known values of the predictors x_i and the corresponding actual response (or output) y_i for each observation i = 1, …, n.

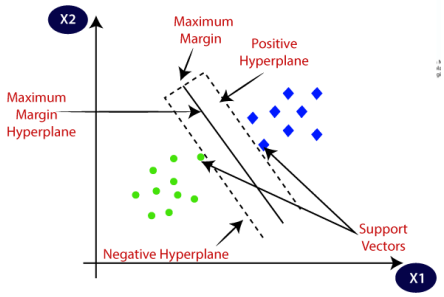


Logistic regression is a linear classifier, so we use linear function $f(x)=b_0+b_1x_1+\ldots b_nx_n$, also called the logit. The variables $b_0,b_1,\ldots b_n$ are the estimators of the regression coefficients, which are also called the predicted weights or just coefficients. The logistic regression function ($\mathbf{x}$) is the sigmoid function of ($\mathbf{x}$): ($\mathbf{x}$) = 1 / (1 + exp(−f($\mathbf{x}$)). As such, it's often close to either 0 or 1. The function ($\mathbf{x}$) is often interpreted as the predicted probability that the output for a given $\mathbf{x}$ is equal to 1. Therefore, 1 − (x) is the probability that the output is 0. Logistic regression determines the best predicted weights b_0 , b_1 , …, b_r such that the function p(x) is as close as possible to all actual responses y_i, i = 1, …, n, where n is the number of observations. The process of calculating the best weights using available observations is called model raining or fitting. There's one more important relationship between ($\mathbf{x}$) and ($\mathbf{x}$), which is that log(p($\mathbf{x}$) / (1 − p($\mathbf{x}$))) = f($\mathbf{x}$). This equality explains why ($\mathbf{x}$) is the logit. It implies that ($\mathbf{x}$) = 0.5 when ($\mathbf{x}$) = 0 and that the predicted output is 1 if ($\mathbf{x}$) > 0 and 0 otherwise.

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

SVM – Support Vector Machine (Classifier)

An SVM maps input (real valued) feature vectors into a higher dimensional feature space through some nonlinear mapping. SVMs are developed on the principle of structural risk minimization [16]. -Structural risk minimization seeks to find a hypothesis (h) for which one can find lowest probability of error whereas the traditional learning techniques for pattern recognition are based on the minimization of the empirical risk, which attempt to optimize the performance of the learning set.



XGBClassifier

The XGBoost or Extreme Gradient Boosting algorithm is a decision tree based machine learning algorithm which uses a process called boosting to help improve performance. regularly produces results that outperform most other algorithms, such as logistic regression, the random forest model and regular decision trees. This is an boosting algorithm which is used for Hyper Parameter Tuning and improve the accuracy.

IV. IMPLEMENTATION

In this paper we used the dataset which is realtime data of Quote-Equity-TATAGLOBAL-EQ-01-08-2021-to-21-08-2023. from https://www.nseindia.com/get-quotes/equity?symbol=TATAGLOBAL. The dataset is having the following structure

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 495 entries, 0 to 494
Data columns (total 14 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   date      495 non-null    object
 1   series    495 non-null    object
 2   open      495 non-null    float64
 3   high      495 non-null    float64
 4   low       495 non-null    float64
 5   prevc     495 non-null    float64
 6   ltp       495 non-null    float64
 7   close     495 non-null    float64
 8   vwap      495 non-null    float64
 9   wh        495 non-null    float64
 10  wl        495 non-null    float64
 11  volume    495 non-null    float64
 12  value     495 non-null    float64
 13  ntraders  495 non-null    float64
dtypes: float64(12), object(2)
memory usage: 54.3+ KB
```

The dataset contains total of 14 features and 495 rows.

The sample data of first 10 records is

```
df = pd.read_csv('stock.csv')
df.head()
```

	date	series	open	high	low	prevc	ltp	close	vwap	wh	wl	volume	value	ntraders
0	2023-08-18	EQ	845.00	848.95	835.50	844.25	843.4	842.55	843.26	883.95	686.6	1145798.0	9.662015e+08	38556.0
1	2023-08-17	EQ	835.00	845.95	835.00	840.70	844.9	844.25	841.71	883.95	686.6	1105325.0	9.303646e+08	51743.0
2	2023-08-16	EQ	840.90	843.45	836.80	844.70	838.0	840.70	840.15	883.95	686.6	360863.0	3.031787e+08	14680.0
3	2023-08-14	EQ	838.85	845.90	829.65	836.90	845.0	844.70	840.85	883.95	686.6	754694.0	6.345808e+08	32827.0
4	2023-08-11	EQ	849.70	853.85	835.00	850.55	835.8	836.90	840.05	883.95	686.6	909080.0	7.636749e+08	31864.0

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

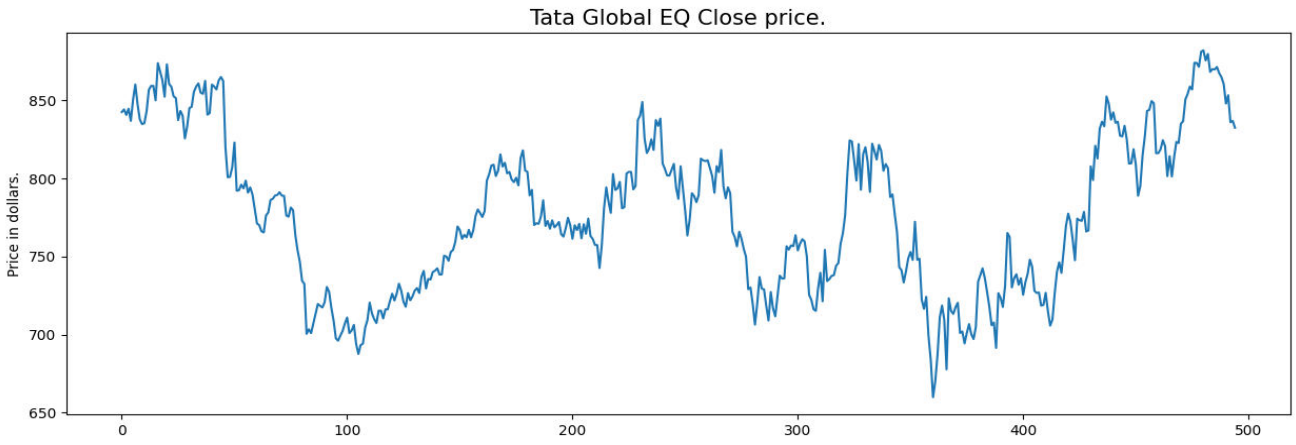
V. RESULTS

After collecting the real time data of TATA Global EQ stock market price. The structure is observed and after the data cleaning process is done on the data set. As the data is realtime data there are no null values. Calculating the summary of the numeric data by using the describe() in python.

```
df.describe()
```

	open	high	low	prevc	ltp	close	vwap	wh	wl	volume	value
count	495.000000	495.000000	495.000000	495.000000	495.000000	495.00000	495.000000	495.000000	495.000000	4.950000e+02	4.950000e+02
mean	778.317273	786.057475	769.150505	777.339596	777.276667	777.33000	777.539616	877.498384	617.134040	1.683535e+06	1.311442e+09
std	50.296256	49.847300	49.742622	49.765291	49.584685	49.75223	49.732458	13.025648	70.426311	1.299805e+06	1.022388e+09
min	659.000000	673.500000	650.200000	659.850000	662.000000	659.85000	661.970000	852.400000	458.950000	1.758370e+05	1.457256e+08
25%	736.000000	743.400000	725.850000	734.850000	734.975000	734.85000	734.695000	861.150000	577.050000	9.811900e+05	7.580008e+08
50%	774.300000	781.950000	765.850000	774.350000	774.250000	774.35000	774.530000	889.000000	650.200000	1.351519e+06	1.074242e+09
75%	815.000000	823.250000	804.125000	814.550000	813.025000	814.55000	814.635000	889.000000	650.200000	1.936892e+06	1.497975e+09
max	885.000000	889.000000	877.600000	882.100000	880.200000	882.10000	881.850000	889.000000	691.750000	1.535813e+07	1.213913e+10

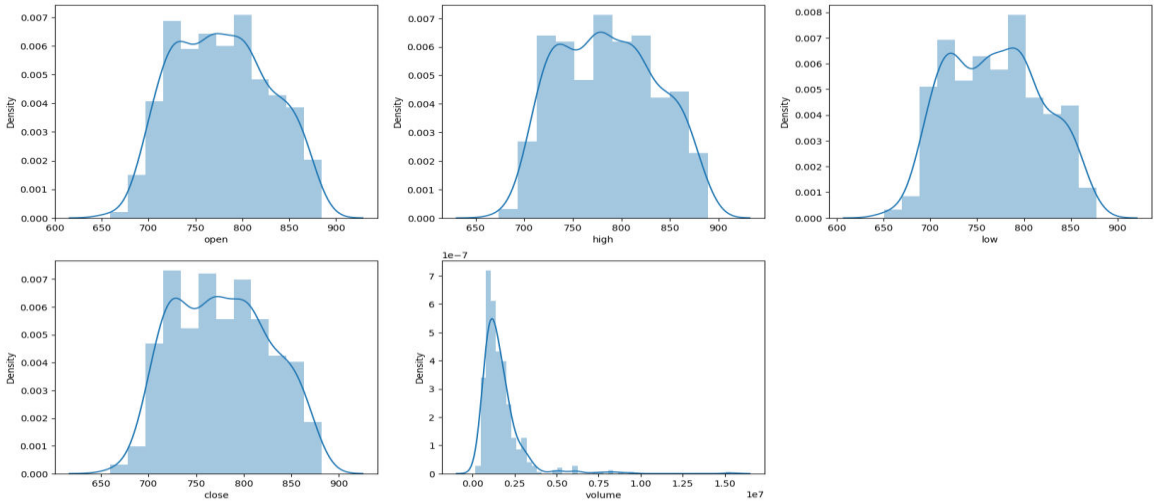
After describing it is observed the data as there are total 495 records. The mean of open is 778.31 and mean of high is 786.05 and so on. Here we need to observe the most important column is the closing price of each day by which we can predict the next opening price and take the decision to where to buy or sell the stocks. This helps in taking right decision. For this we observe graphically what is the impact of closing price on each day.



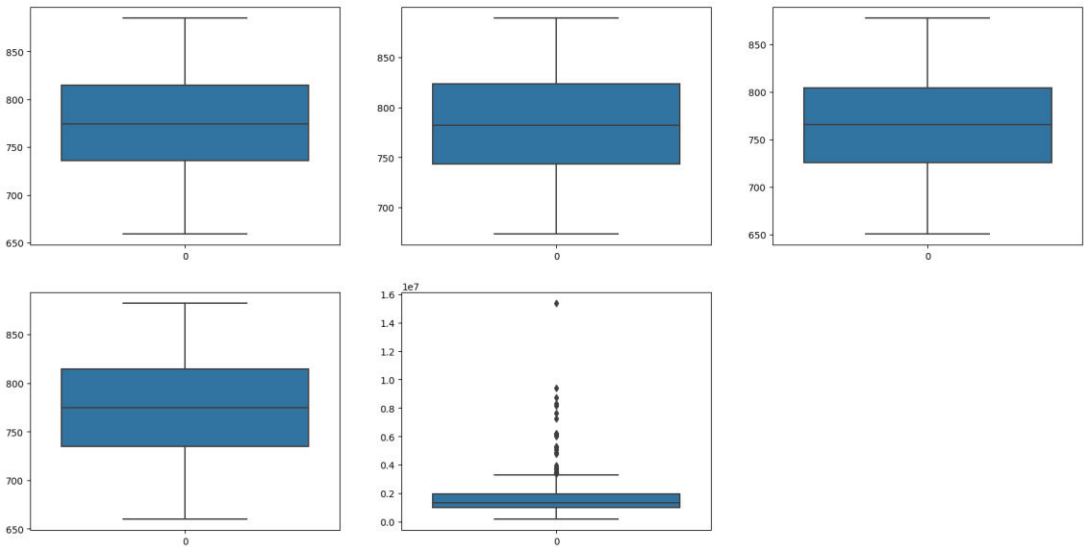
Next step is to observe the histograms of each column and identify how the market is going on. Open is the opening price , High is the highest price it reached on the day, Low is the lowest price reached on the day,

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

Close is the closing price of the day and volume is the number of investors who invested in the stocks.

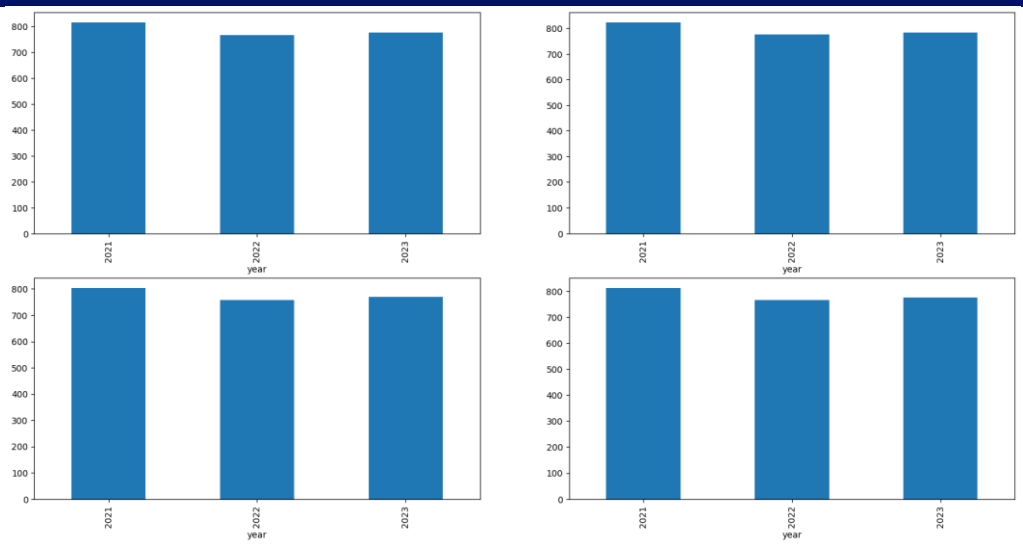


Next step is to check the outliers in the data. For this we draw boxplot for each and identify the outliers in the given data

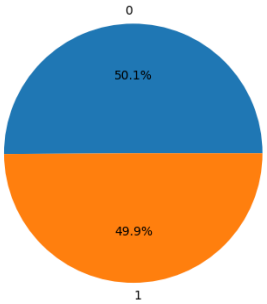


It is observed that only volume data have outliers. The next step is to observe the year wise average fluctuation in the prices.

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org



Now basing on this analysis we can now see whether an investor is buying or selling the stocks. How is taking the decision basing on these data



It is observed that 49% chances are to buy and 50% chances are to retain the stocks. Now after applying the training and testing on the collected data and the train_test_split is taken as 70% and 30%. Now apply different Machine Learning algorithms as Logistic Regression, SVClassifier, and XGBoost and test the accuracy of each model.

```
LogisticRegression() :
Training Accuracy :  0.9454387524240466
Validation Accuracy :  0.8782051282051282

SVC(kernel='poly', probability=True) :
Training Accuracy :  0.9432167097608274
Validation Accuracy :  0.8541666666666667

XGBClassifier(base_score=None, booster=None, callbacks=None,
              colsample_bylevel=None, colsample_bynode=None,
              colsample_bytree=None, early_stopping_rounds=None,
              enable_categorical=False, eval_metric=None, feature_types=None,
              gamma=None, gpu_id=None, grow_policy=None, importance_type=None,
              interaction_constraints=None, learning_rate=None, max_bin=None,
              max_cat_threshold=None, max_cat_to_onehot=None,
              max_delta_step=None, max_depth=None, max_leaves=None,
              min_child_weight=None, missing=nan, monotone_constraints=None,
              n_estimators=100, n_jobs=None, num_parallel_tree=None,
              predictor=None, random_state=None, ...) :
Training Accuracy :  0.9999595992243051
Validation Accuracy :  0.8429487179487178
```

After implementing the models the

Model	Training Accuracy	Validation Accuracy
Logistic Regression	0.94	0.87
SVClassifier	0.94	0.85
XGBoost	0.99	0.84

VI. CONCLUSION

In this paper we have implemented only the fundamental idea to predict the Stock market price on real time data. It is observed that the accuracy is being increased when the boosting algorithms are applied. By identifying and tuning the Hyper Parameters it is observed that the accuracy may be increased which helps in right Decision at Right Time. This is the most useful to take smart decision while investing in stocks. As the study of the market is changing from time to time with in seconds a keen observation is required on the process, This analysis may help the investor to take decision.

REFERENCES

[1] Stock Market Prediction using Linear Regression and Support Vector Machines Vaishnavi Gururaj#1, Shriya V R#2 and Dr. Ashwini K#3 #123 CSE Department, Global Academy of Technology, Bengaluru, India. International Journal of Applied Engineering Research ISSN 0973-4562 Volume 14, Number 8 (2019) pp. 1931-1934.
[2] Stock Market Prediction using Logistic Regression Analysis - A Pilot Study Shreenidhi Sriram1 , Sanjana Rangarajan2 1 Student, Department of Computer Science & Engineering, College of Engineering, Anna University, Guindy, Chennai, India. 2 Student, Department of Information Technology, College of Engineering, Anna University, Guindy, Chennai, India. International Journal for Research in Applied Science & Engineering Technology (IJRASET). 2020
[3] Stock Market Analysis Using Linear Regression,Taran Rishi, Department of Economics University of Northern Iowa. UnI Scholar works , 2022.
[4] “Stock Market Predication Using A Linear Regression” Dinesh Bhuriya et al.
[5] J. A. K. Suykens, "Nonlinear modelling and support vector machines." pp. 287-294.
[6] Dataset : https://www.nseindia.com/get-quotes/equity?symbol=TATAGLOBAL.