
STOCK
PRICE
PREDICTION
USING MACHINE LEARNING AND LINEAR REGRESSION

Team Members Details

SL NO.	NAME	UNIVERSITY ROLL
1.	SUBRATA MUKHERJEE	16500119013
2.	SAGNIK DAS	16500119020
3.	SAGAR MAITI	16500119016
4.	DEBARGHA ADHIKARY	16500119037

Under the Supervision of

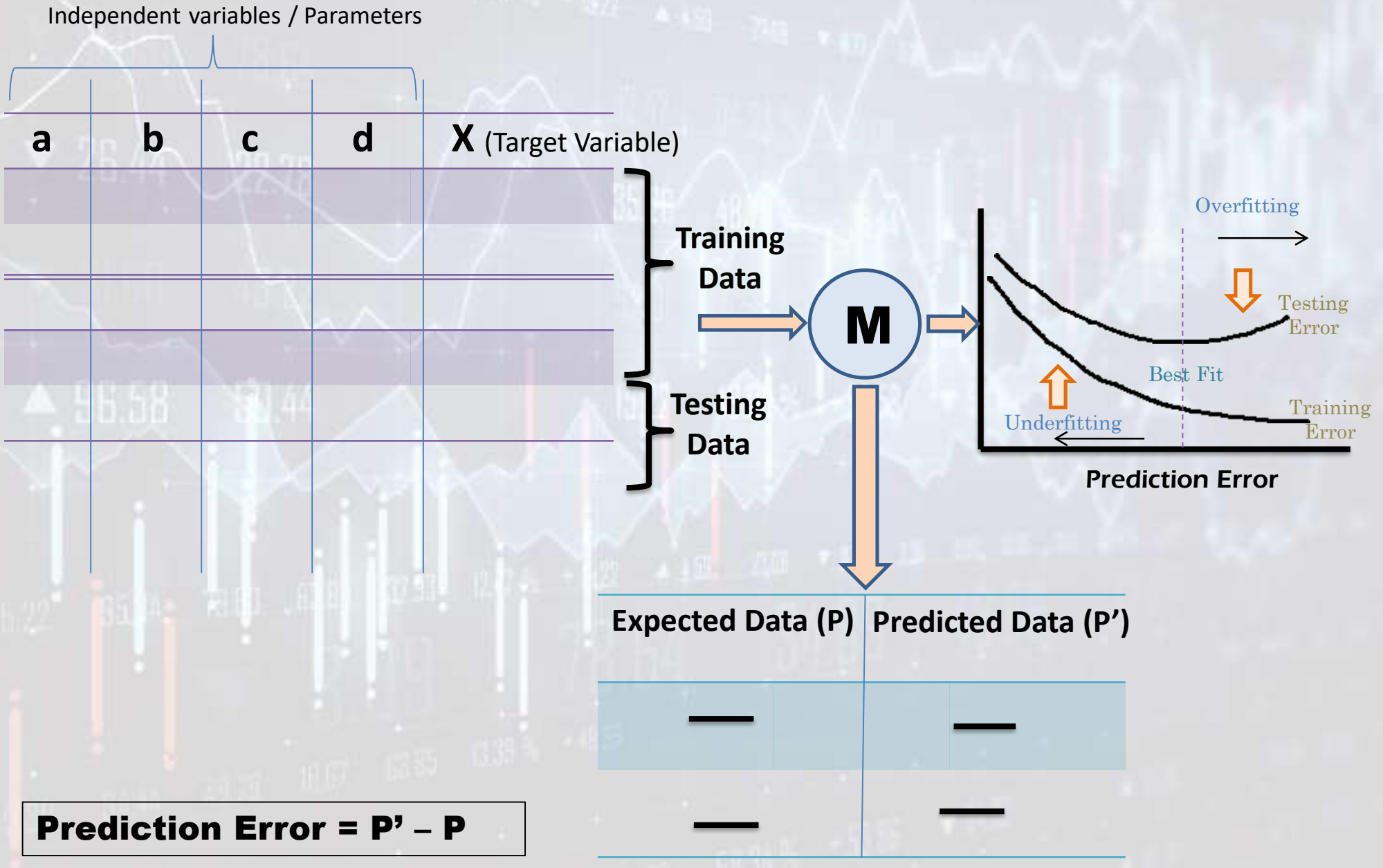
Mr. DEBAJYOTI BAGCHI
DEPARTMENT OF COMPUTER SCIENCE ENGINEERING
CALCUTTA INSTITUTE OF ENGINEERING AND MANAGEMENT
PRASANTA SUR CAMPUS
KOLKATA-700040, INDIA
MAY 2023

INTRODUCTION

The objective of a stock price prediction project using machine learning is to build a model that can accurately forecast the future stock prices of a company based on historical stock market data. The goal is to create a model that can identify patterns and relationships in the data to make predictions on future prices. The model should be able to analyze large amounts of data, identify trends, and make predictions that are as accurate as possible.

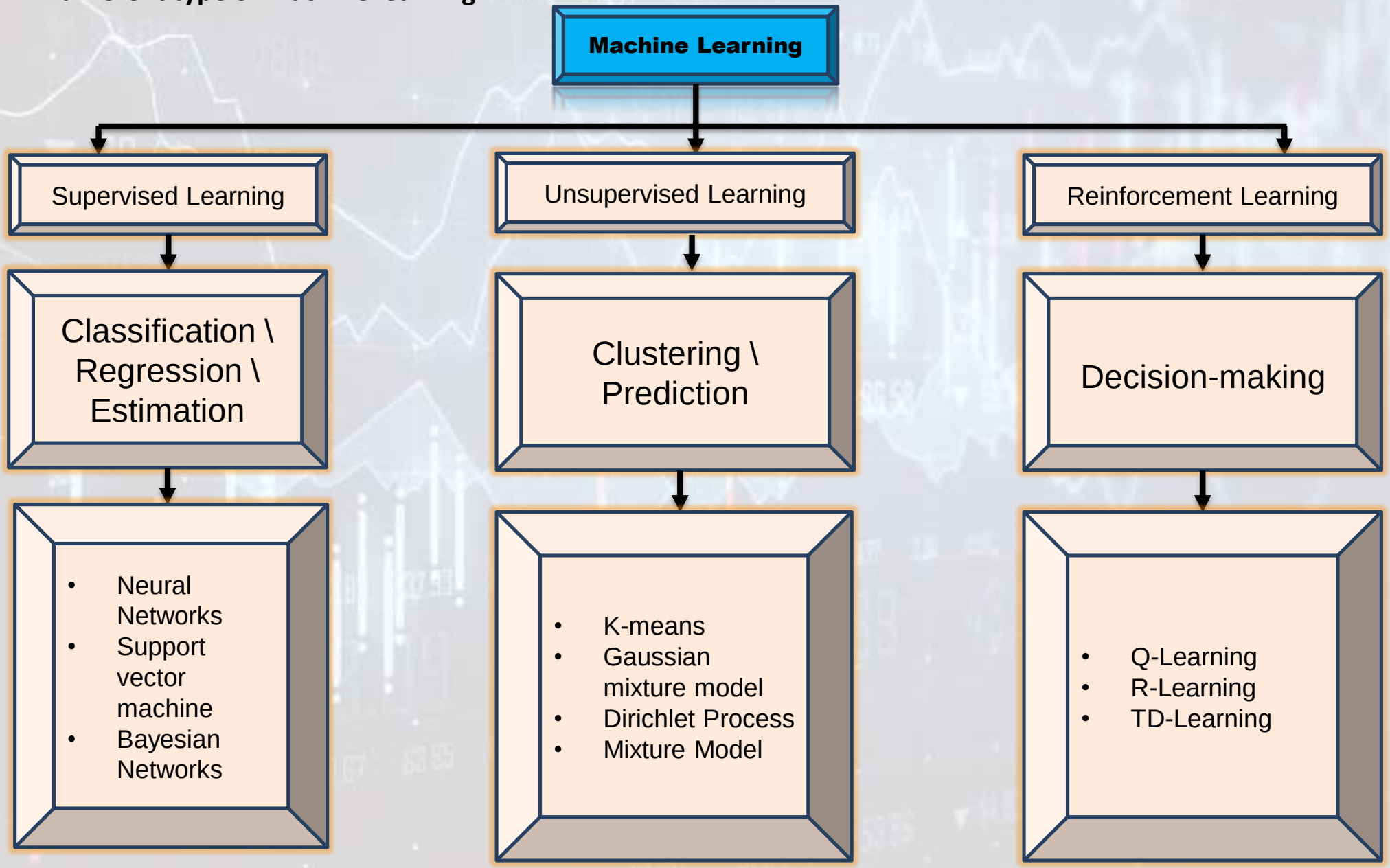
Stock market prediction and analysis are some of the most difficult jobs to complete. There are numerous causes for this, including market volatility and a variety of other dependent and independent variables that influence the value of a certain stock in the market. These variables make it extremely difficult for any stock market expert to anticipate the rise and fall of the market with great precision. A correct prediction of stocks can lead to huge profits for the seller and the broker. Frequently, it is brought out that prediction is chaotic rather than random, which means it can be predicted by carefully analyzing the history of respective stock market. Machine learning is an efficient way to represent such processes.

Following diagram shows the fundamental of machine learning :



Prediction Error = P' – P

Different type of machine learning used in different purposes. **The following flow chart shows the different type of machine learning :**



It predicts a market value close to the tangible value, thereby increasing the accuracy. Introduction of machine learning to the area of stock prediction has appealed to many researches because of its efficient and accurate measurements. **Lots of researches done on this field, some of those research studies done below :**

In Ashfaq Shaikh, Ajay Panuganti, Maaz Husain &Prateek Sing's paper[1] taken approach, the machine learning model uses historical prices and human sentiments as two different inputs, and the output is distinguished as a graph showing the future prediction and a label (positive neutral and negative), respectively.

In Amit Gupta, T.J.Nagalakshmi's paper[2] taken approach, stock costs of an organization relies upon different factors Thus, the financial exchange forecast turns into a great deal troublesome and testing. Under regression methodology linear regression and support vector regression algorithm is used here and it should great results. The model is incredibly trendy and that we examine in foreseeing the stock costs in all respects precisely.

In IshitaParmar; Navanshu Agarwal; SheirshSaxena; Ridam Arora; Shikhin Gupta; HimanshuDhiman; LokeshChouhan's paper[3] taken approach, Machine learning itself employs different models to make prediction easier and authentic. The recent trend in stock market prediction technologies is the use of machine learning which makes predictions based on the values of current stock market indices by training on their previous values. Here Regression technique is used to predict stock values based on factors(open, close, low, high, volume) and it showed great results.

Above papers can predict the market scenario with advantages like

- Simplicity
- Flexibility
- Predictive power
- Robustness

But with some difficulty there are some disadvantages also with this such as

- Linearity assumption
- Sensitivity to outliers
- Over fitting
- Multicollinearity

To resolve the issue we use linear regression in our project ;

Linear regression is a simple yet powerful technique that provides insights into the relationship between variables and allows for prediction and inference. However, its effectiveness depends on the underlying assumptions being met and the presence of a linear relationship between the dependent and independent variables.
It handles over fitting pretty well using dimensionally reduction techniques, regularization, and cross-validation.

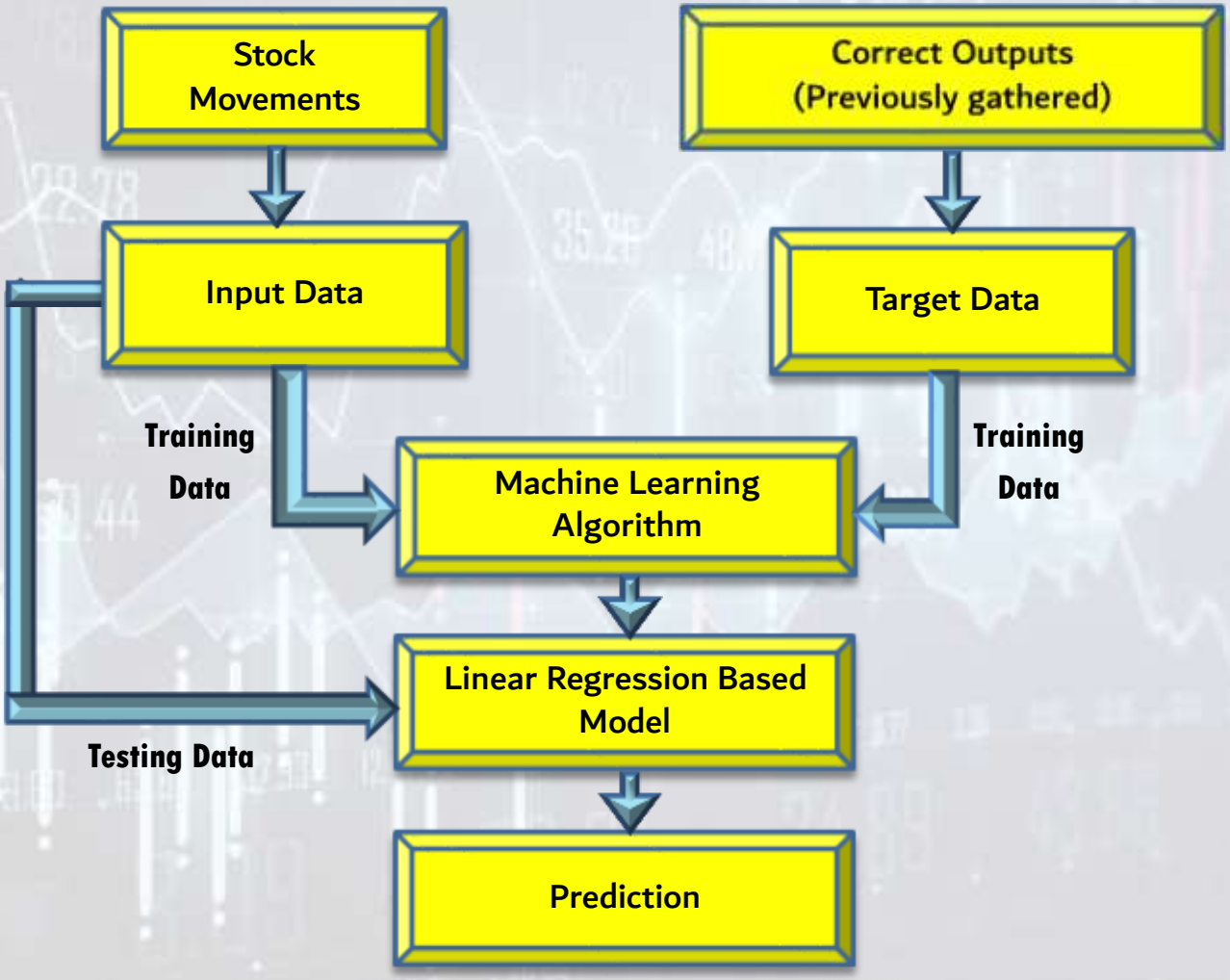
METHODOLOGY

Machine learning is a subfield of artificial intelligence (AI) that focuses on the development of algorithms and models that enable computers to learn from data and make predictions or decisions without being explicitly programmed. The working principle of machine learning involves several key components:

- ❑ Data collection
- ❑ Data preprocessing
- ❑ Feature extraction and selection
- ❑ Model selection and training
- ❑ Evaluation and validation
- ❑ Prediction or decision-making
- ❑ Iterative refinement

It's Important to note that machine learning models are only as good as the data they are trained on. High-quality, diverse, and representative data is crucial for the success of machine learning algorithms.

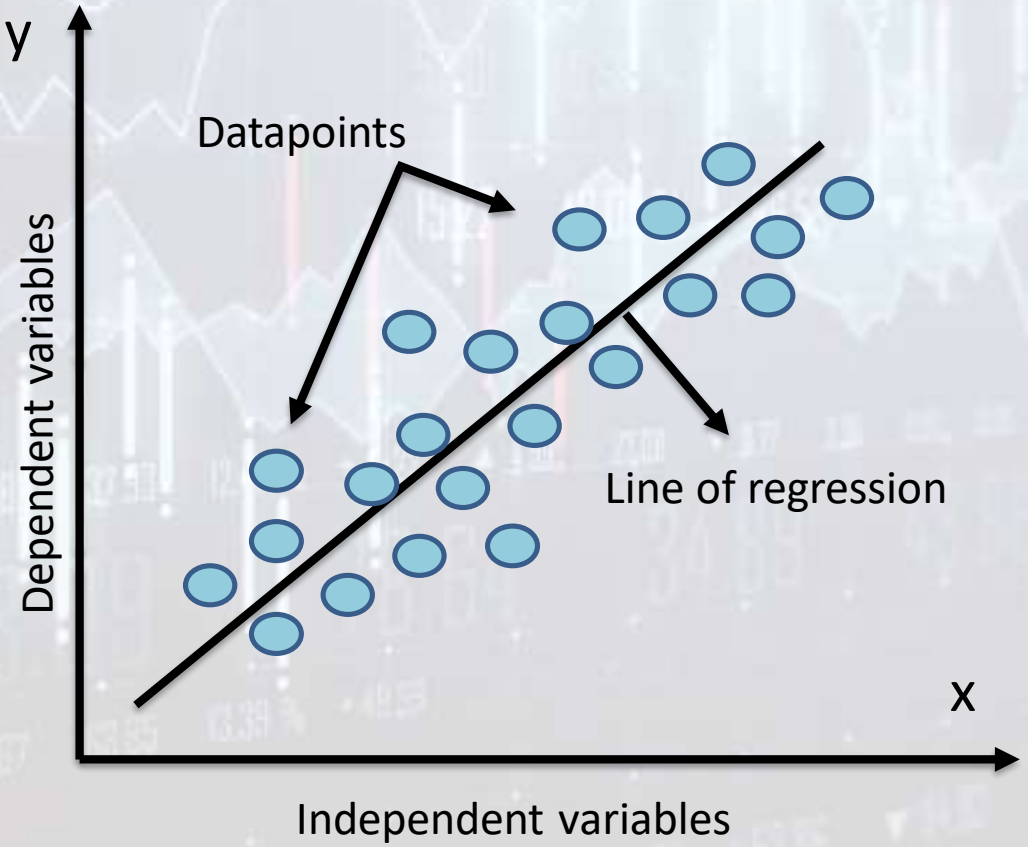
DATA FLOW DIAGRAM



WORKING OF LINEAR REGRESSION BASED MODEL

Linear regression algorithm shows a linear relationship between a dependent (y) and one or more independent (x) variables, hence called as linear regression. Since linear regression shows the linear relationship, which means it finds how the value of the dependent variable is changing according to the value of the independent variable.

The linear regression model provides a sloped straight line representing the relationship between the variables. Consider the below image:



Mathematically we can represent a linear regression as : $y = a0 + a1x + \varepsilon$

Here,

y = Dependent variable (Target Variable)

x = Independent variable (Predictor variable)

a0 = Intercept of the line (Gives an additional degree of freedom)

a1 = Linear regression coefficient (scale factor to each input value)

ε = Random error

The values for x and y variables are training datasets for Linear Regression model representation.

Regression is used for predicting continuous values through some given independent values. Work was carried out on csv format of data through Panda library and calculated the parameter which is to be predicted, the price of the stocks with respect to time. The data is divided into different train sets for cross validation to avoid over fitting. The test set is generally Kept 20% of the whole dataset. Linear regression as given by the above equation is performed on the data and then predictions are made, which are plotted to show the results of the stock market prices vs time.

RESULTS

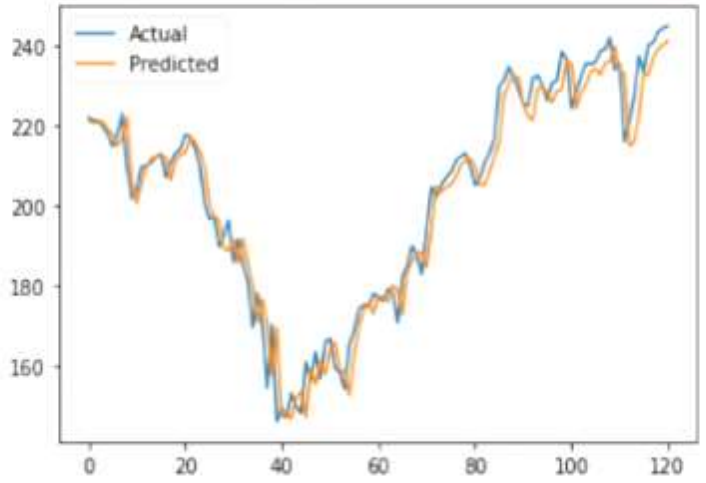
1. FACEBOOK :

Out[11]:

	Date	Actual	Predicted
0	2020-01-16	222.14	221.393213
1	2020-01-17	221.44	220.696224
2	2020-01-21	221.32	221.098561
3	2020-01-22	219.76	221.090701
4	2020-01-23	217.94	219.152618
...	...	...	...
116	2020-07-02	240.28	232.703244
117	2020-07-06	240.86	237.025522
118	2020-07-07	243.58	239.245923
119	2020-07-08	244.50	240.176012
120	2020-07-09	245.07	241.359188

121 rows × 3 columns

```
In [12]: df_pred[['Actual', 'Predicted']].plot()
Out[12]: <AxesSubplot:>
```



2. TESLA :

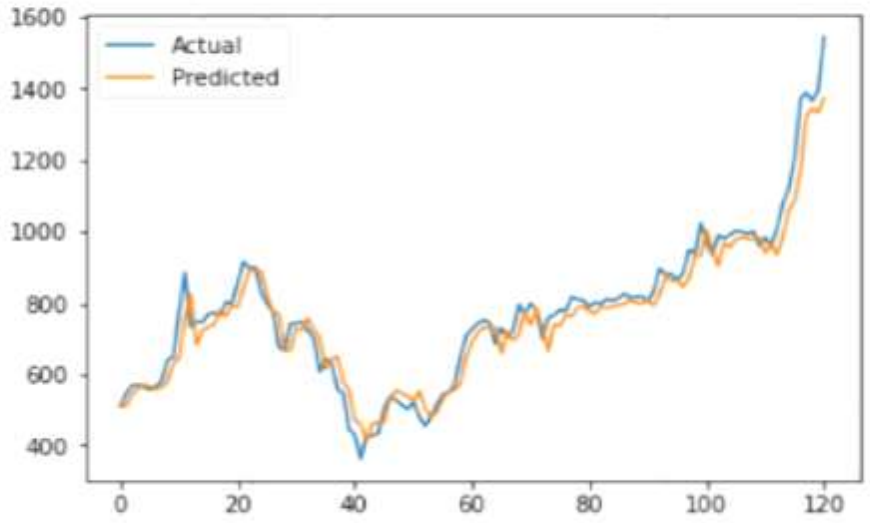
Out[14]:

	Date	Actual	Predicted
0	2020-01-16	510.50	518.889014
1	2020-01-17	547.20	510.685280
2	2020-01-21	569.56	544.412573
3	2020-01-22	572.20	561.497970
4	2020-01-23	564.82	572.873537
...	...	...	...
116	2020-07-02	1371.58	1169.005684
117	2020-07-06	1389.86	1323.728610
118	2020-07-07	1365.88	1344.902981
119	2020-07-08	1394.28	1332.692107
120	2020-07-09	1544.65	1373.506908

121 rows × 3 columns

```
In [15]: df_pred[['Actual', 'Predicted']].plot()
```

Out[15]: <AxesSubplot:>



3. IBM :

```
In [14]: df_pred = pd.DataFrame(Y_val.values, columns=['Actual'], index=Y_val.index)
         df_pred['Predicted'] = Y_val_pred
         df_pred = df_pred.reset_index()
         df_pred.loc[:, 'Date'] = pd.to_datetime(df_pred['Date'],format='%Y-%m-%d')
         df_pred
```

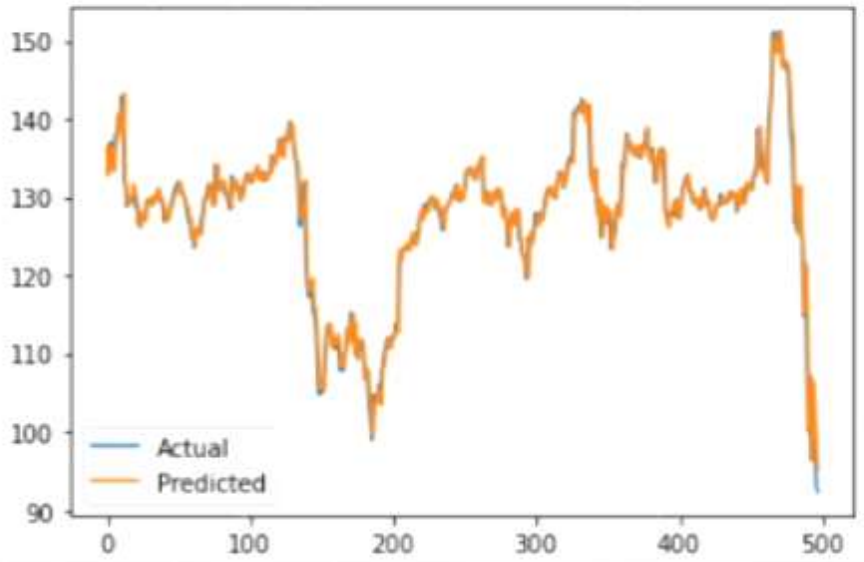
Out[14]:

	Date	Actual	Predicted
0	2018-03-29	133.37	136.382752
1	2018-04-02	133.17	132.761893
2	2018-04-03	136.96	133.619627
3	2018-04-04	136.88	136.390853
4	2018-04-05	133.81	136.385770
...	...	...	...
493	2020-03-16	103.88	96.399760
494	2020-03-17	100.86	106.404985
495	2020-03-18	97.73	102.631799
496	2020-03-19	92.91	98.834408
497	2020-03-20	92.31	95.070682

498 rows × 3 columns

```
In [15]: df_pred[['Actual', 'Predicted']].plot()
```

Out[15]: <AxesSubplot:>



4. GOOGLE :

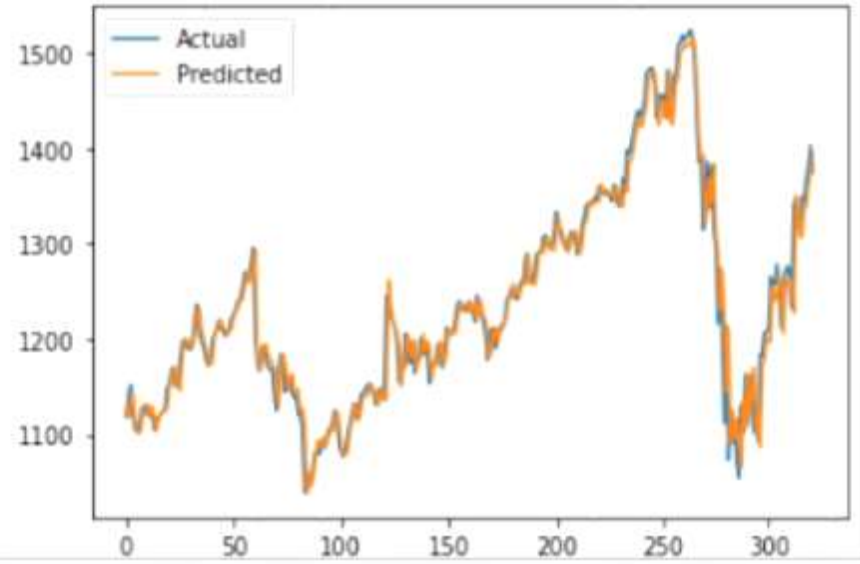
Out[16]:

	Date	Actual	Predicted
0	2019-01-31	1118.62	1126.264873
1	2019-02-01	1141.42	1118.131044
2	2019-02-04	1151.87	1139.733372
3	2019-02-05	1122.89	1143.017769
4	2019-02-06	1105.91	1118.366874
...	...	...	...
317	2020-05-05	1345.43	1339.532841
318	2020-05-06	1369.28	1339.216415
319	2020-05-07	1384.34	1361.989490
320	2020-05-08	1403.59	1372.787714
321	2020-05-11	1375.18	1394.198162

322 rows × 3 columns

In [17]:
```
df_pred[['Actual', 'Predicted']].plot()
```

Out[17]: <AxesSubplot:>



5. APPLE :

Out[29]:

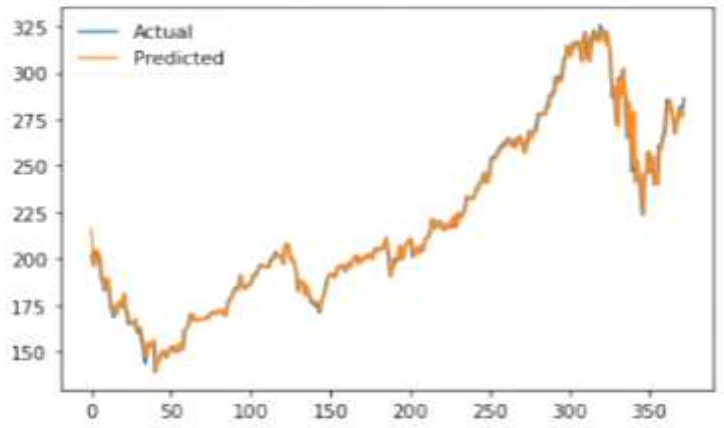
	Date	Actual	Predicted
0	2018-11-01	202.30	216.289777
1	2018-11-02	196.56	201.470183
2	2018-11-05	198.68	195.948932
3	2018-11-06	204.71	199.043601
4	2018-11-07	204.00	204.193666
...	...	...	...
368	2020-04-22	273.79	275.751755
369	2020-04-23	281.70	274.020244
370	2020-04-24	281.90	280.710044
371	2020-04-27	277.33	280.543603
372	2020-04-28	286.44	276.562125

373 rows × 3 columns

In [30]:

```
df_pred[['Actual', 'Predicted']].plot()
```

Out[30]: <AxesSubplot:>



CONCLUSION

Linear Regression technique have been utilized in this project on our datasets. This technique have shown an improvement in the accuracy of predictions, thereby yielding positive results. Use of this machine learning Techniques in the prediction of stocks have yielded Promising results. It has led to the Conclusion that it is possible to predict stock market With more accuracy and efficiency using machine Learning techniques.

In the future, the stock market prediction system can be further improved by utilizing a much bigger dataset than the one being utilized currently. This would help to increase the accuracy of our prediction models. Furthermore, other models of Machine Learning could also be studied to check for the accuracy rate resulted by them.

FUTURE SCOPE

- **Enhanced Prediction Models:** These models can capture complex patterns and dependencies in data, leading to more accurate predictions.
- **Real-Time Predictions:** Faster computing and streaming data will enable traders and investors to make more informed decisions with real time predictions.
- **Automated Trading Systems:** Automated trading systems that use stock price prediction models are a potential future development.
- **Risk Assessment:** Machine learning models can be used to assess risk factors, such as volatility, market risk, and systemic risk.

REFERENCES

[1] Ashfaq Shaikh, Ajay Panuganti, Maaz Husain &Prateek Sing, " Stock Market Prediction Using Machine Learning," First Online: 25 January 2021 , Springer, Singapore, Online ISBN : 978-981-15-8443-5

[2] Amit Gupta, T.J.Nagalakshmi, "Stock Price Prediction using Linear Regression in Machine Learning," International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075 (Online), Volume-8 Issue-12, October 2019 ,DOI:10.35940/ijitee.L3932.1081219

[3] IshitaParmar; Navanshu Agarwal; SheirshSaxena; Ridam Arora; Shikhin Gupta; HimanshuDhiman; LokeshChouhan, " Stock Market Prediction Using Machine Learning," 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC), 15-17 December 2018, DOI: 10.1109/ICSCCC.2018.8703332

[4] Zhiyong Zhao; Ruonan Rao; ShaoxiongTu, "Time-Weighted LSTM Model with Redefined Labeling for Stock Trend Prediction," 2017 IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI), 06-08 November 2017, DOI: 10.1109/ICTAI.2017.00184

[5] K. Raza, "Prediction of Stock Market performance by using machine learning techniques," 2017 International Conference on Innovations in Electrical Engineering and Computational Technologies (ICIEECT), Karachi, 2017, pp. 1-1.


THANK YOU ...