A PROJECT REPORT

On

STOCK PRICE PREDICTION

Submitted to

KIIT Deemed to be University

In Partial Fulfillment of the Requirement for the Award of

BACHELOR'S DEGREE IN

COMPUTER SCEINCE & ENGINEERING

BY

Chirag Mishra	**2005578**
Sajal Kumar Ujjwal	**2005603**
Sarthak Singh	**2005608**

UNDER THE GUIDANCE OF
Dr. Arup Abhinna Acharya

SCHOOL OF COMPUTER ENGINEERING
KALINGA INSTITUTE OF INDUSTRIAL TECHNOLOGY
BHUBANESWAR, ODISHA - 751024

May 2023

A PROJECT REPORT

On

STOCK PRICE PREDICTION

Submitted to

KIIT Deemed to be University

In Partial Fulfillment of the Requirement for the Award of

BACHELOR'S DEGREE IN

COMPUTER SCEINCE & ENGINEERING

BY

Chirag Mishra	2005578
Sajal Kumar Ujjwal	2005603
Sarthak Singh	2005608

UNDER THE GUIDANCE OF
Dr. Arup Abhinna Acharya

SCHOOL OF COMPUTER ENGINEERING
KALINGA INSTITUTE OF INDUSTRIAL TECHNOLOGY
BHUBANESWAR, ODISHA - 751024

May 2023

KIIT Deemed to be University

School of Computer Engineering
Bhubaneswar, ODISHA 751024

CERTIFICATE

This is certify that the project entitled

STOCK PRICE PREDICTION

submitted by

Chirag Mishra	2005578
Sajal Kumar Ujjwal	2005603
Sarthak Singh	2005608

is a record of bonafide work carried out by them, in the partial fulfilment of the requirement for the award of Degree of Bachelor of Engineering (Computer Science & Engineering OR Information Technology) at KIIT Deemed to be university, Bhubaneswar. This work is done during year 2022-2023, under our guidance.

Date: 05/05/2023

Dr. Arup Abhinna Acharya
Project Guide

Acknowledgement

We are profoundly grateful to **Dr. Arup Abhinna Acharya, Dean, CSE** of **KIIT University** for his expert guidance and continuous encouragement throughout to see that this project rights its target since its commencement to its completion.

2005578 Chirag Mishra
2005603 Sajal Kumar Ujjwal
2005608 Sarthak Singh

ABSTRACT

Accurately predicting future stock prices is difficult due to the stock market's great volatility and complexity. The ability to make informed decisions about the purchase, sale, or ownership of stocks, however, is advantageous for investors and can result in accurate stock price predictions. In this project, one of the top two-wheeler producers in India, Hero Motors, we try to forecast stock prices.

We gathered Hero Motors' historical stock price data and performed linear regression on it. In addition, we looked at a number of variables that affect Hero Motors stock prices, including macroeconomic statistics, company-specific news, market trends, and investor sentiment.

KEYWORDS:
1. API
2. Linear Regression
3. Stocks
4. Hero-motocorp
5. Machine Learning
6. Javascript

Contents

List of Figures

Chapter 1

Introduction

Investors, traders, and scholars have always found stock price forecasting to be a fascinating subject. Accurately predicting the future price of a company is difficult due to the unpredictability of the stock market. Investors may benefit from having the capacity to forecast stock prices since it enables them to make well-informed decisions about whether to buy, sell, or keep stocks.

In this project, one of India's top two-wheeler producers, Hero Motors, has its stock prices predicted. Hero Motors is a publicly traded company, and a number of variables, including macroeconomic statistics, company-specific news, market trends, and investor sentiment, have an impact on the price of its shares.

We will employ linear regression techniques to forecast Hero Motors stock values. We will examine Hero Motors' historical stock price data to look for patterns and trends. We will create predictive models based on these trends that can project Hero Motors stock prices in the future.The project report will include a detailed analysis of the Hero Motors stock price forecast.

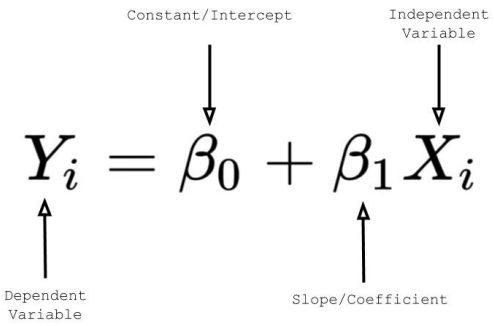
Chapter 2

Basic Concepts/ Literature Review

This section contains the basic concepts about the related tools and techniques used in this project. For research work, present the literature review in this section.

2.1 Linear Regression

Linear Regression: This statistical technique fits a linear equation to the data
in order to model the relationship between two variables. The relationship between the stock price and numerous influencing factors is modelled using linear regression in the context of stock price prediction.



2.1 Formula for Linear Regression

2.2 Variables

Dependent Variable: In a linear regression model, the variable we want to predict is the dependent variable. The stock price is the dependent variable when predicting stock prices. Here the dependent variable is Price. (Refer to formula of linear regression)

Independent Variables: In a linear regression model, the independent variables are those that are utilized to forecast the dependent variable. Independent variables in the context of stock price forecasting can include elements like macroeconomic statistics, company-specific news, market trends, and investor mood. Here the independent variable is time/date. (Refer to the Linear Regression Diagram)

2.3 Historical Data

Historical Data: Past data that is used to train a prediction model is referred to as historical data. Historical data can include previous stock prices, macroeconomic variables, company-specific news, market trends, and investor sentiment when used to forecast stock prices.

	A	B	C	D	E	F	G	H	I	J	K	L
1	Date	Symbol	Series	Prev Close	Open	High	Low	Last	Close	VWAP	Volume	Turnover
2	03-01-2000	HEROHONDA	EQ	1145.2	1200	1236.8	1171	1209.95	1207.6	1209.11	10631	1.285E+12
3	04-01-2000	HEROHONDA	EQ	1207.6	1200	1220	1176	1200	1199.85	1200.93	19849	2.384E+12
4	05-01-2000	HEROHONDA	EQ	1199.85	1155	1178	1140	1140	1143.8	1159.74	2525	2.928E+11
5	06-01-2000	HEROHONDA	EQ	1143.8	1183	1189.95	1150.25	1165	1161.4	1167.29	4375	5.107E+11
6	07-01-2000	HEROHONDA	EQ	1161.4	1140	1140	1069	1105	1106.55	1103.52	9093	1.003E+12
7	10-01-2000	HEROHONDA	EQ	1106.55	1105.05	1135	1100	1115.1	1126.35	1123.41	11113	1.248E+12
8	11-01-2000	HEROHONDA	EQ	1126.35	1100	1135	1095	1130	1125.1	1106.96	11333	1.255E+12
9	12-01-2000	HEROHONDA	EQ	1125.1	1176	1198	1150	1180	1174.95	1174.81	4353	5.114E+11
10	13-01-2000	HEROHONDA	EQ	1174.95	1198	1198	1140	1140	1146.4	1172.03	7221	8.463E+11
11	14-01-2000	HEROHONDA	EQ	1146.4	1194	1198	1157.5	1190	1190.3	1182.32	12898	1.525E+12
12	17-01-2000	HEROHONDA	EQ	1190.3	1216	1285.55	1216	1285.55	1285.55	1281.19	44550	5.708E+12
13	18-01-2000	HEROHONDA	EQ	1285.55	1339	1365	1270	1300	1294.65	1295.57	17246	2.234E+12
14	19-01-2000	HEROHONDA	EQ	1294.65	1300	1324.9	1270	1274.5	1272.55	1281.66	23057	2.955E+12
15	20-01-2000	HEROHONDA	EQ	1272.55	1275	1275	1236.3	1245	1252.9	1250.91	3987	4.987E+11
16	21-01-2000	HEROHONDA	EQ	1252.9	1240	1255.5	1235	1249	1242.55	1242.11	3125	3.882E+11
17	24-01-2000	HEROHONDA	EQ	1242.55	1201	1260.25	1201	1206.7	1212.1	1218.43	2624	3.197E+11
18	25-01-2000	HEROHONDA	EQ	1212.1	1201	1300	1170	1300	1218.6	1204	5056	6.087E+11
19	27-01-2000	HEROHONDA	EQ	1218.6	1221	1237	1183	1185	1187.3	1199.91	3095	3.714E+11
20	28-01-2000	HEROHONDA	EQ	1187.3	1180	1210	1180	1199	1197.55	1195.63	2427	2.902E+11
21	31-01-2000	HEROHONDA	EQ	1197.55	1187	1195	1171.1	1175	1175.3	1179.87	1491	1.759E+11
22	01-02-2000	HEROHONDA	EQ	1175.3	1160	1199	1125	1125	1146.25	1152.45	3516	4.052E+11

2.2. Data-set Used

2.4 Evaluation Metrics

Evaluation metrics are used to gauge a prediction model's degree of accuracy. Commonly used evaluation measures for stock price prediction include mean squared error, mean absolute error, and root mean squared error.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \tilde{y}_i)^2$$

2.3. Formula for MSE

Chapter 3

Problem Statement / Requirement Specifications

A variety of economic, political, and social issues have an impact on the stock market, which is a dynamic and complicated system. It is difficult to predict stock prices accurately, but it is necessary for investors, traders, and other stakeholders to do so in order to make wise decisions about whether to buy, sell, or hold stocks. The goal of this project is to anticipate Hero Motors' stock prices using linear regression and assess how accurate the predictions are. Hero Motors is a publicly traded firm. On the basis of an examination of the stock price prediction using linear regression, the project will also seek to pinpoint the variables that affect Hero Motors stock prices and offer investor advice.

3.1 Project Planning

Phase 1: Planning and Requirements Gathering
Phase 2: Data Collection and Pre-processing
Phase 3: Model Development
Phase 4: Model Evaluation
Phase 5: Feature Importance
Phase 6: Recommendation System and Final Report

3.2 Project Analysis

Review the project scope and objectives:Examining the project scope and objectives will help the project team make sure that they are in line with the needs of the stakeholders.
Review the project timeline: The project team will examine the project timeframe to make sure it is practical and realistic.
Review of the model assessment plan: The model evaluation plan will be examined by the project team to make sure it is complete and accurate.
Review of the feature importance analysis plan: The feature importance analysis plan will be reviewed by the project team to make sure it is sound and founded on best practises.

3.3 System Design

System Architecture: System architecture entails specifying the elements of the system and their relationships to one another. The system has a user interface on the web for showing the predictions, a linear regression model for forecasting future stock prices and a data pre-processing module where data is imported, cleaned and checked for null values.

Data sources and data flow: This entails determining the data sources used by the system which have been taken from websites like Kaggle where there are accurate datasets.. Determining how the data flows through the system, including how it is ingested, cleansed, and converted before being utilized for prediction is also done.

3.3.1 Design Constraints

Hardware Constraints: The project team should consider the limitations of the hardware components, such as memory, processing power, and storage capacity, while designing the system.

Software Constraints: The project team should consider the limitations of the software components, such as compatibility, licensing, and interoperability, while designing the system.

Data Constraints: The project team should consider the limitations of the data, such as availability, quality, and security, while designing the system.

Time Constraints: The project team should consider the time constraints and develop a realistic timeline for designing and developing the system.

Performance Constraints: The project team should consider the performance constraints, such as response time, throughput, and scalability, while designing the system.

Chapter 4

Implementation

In this section, present the implementation done by you during the project development.

4.1 Methodology OR Proposal

Data Gathering: The necessary data was gathered in the first step in the implementation process.

This was in the form of CSV files which contained historical data of Hero Motors' stocks, which was taken from various sources from the internet.
Data pre-processing: The accuracy of the model may be impacted by missing values, outliers, or noise in the collected data. To eliminate the abnormalities and enhance the data's quality, the data was cleaned, converted, and normalized.

Feature Selection: The next step was to identify the relevant features for the prediction model. Here the dependant variable is price while the independent variable is time/date.

Model Selection: Linear regression model was used as it gives accurate prediction for continuous data.

Model Training: The selected model was trained on the preprocessed data.

Model Evaluation: The performance of the model was evaluated based on various metrics such as Mean Squared Error (MSE) and R-squared value.

4.2 Testing OR Verification Plan

Performance Testing: Performance testing was performed to evaluate the system's performance under different load conditions.

Accuracy Testing: Accuracy testing should be performed to evaluate the model's accuracy in predicting the stock prices.

Test ID	Test Case Title	Test Condition	System Behavior	Expected Result
T01	Data Processing	Data has null values	System handles null values correctly	Null values are removed.
T02	Data Visualization	Plotting Chart	System successfully creates a plot of the data	Plot is displayed correctly.
T03	Model Evaluation	Training set performance	The system calculates the r2_score and MSE for the training set.	The scores are reasonable and accurate which is near 80 percent.

Training and testing accuracy (as shown on Jupyter terminal):

```
In [36]:
scores = f'''
{'Metric'.ljust(10)}{'Train'.center(20)}{'Test'.center(20)}
{'r2_score'.ljust(10)}{r2_score(Y_train, lm.predict(X_train))}\t{r2_score(Y_test, lm.pre
{'MSE'.ljust(10)}{mse(Y_train, lm.predict(X_train))}\t{mse(Y_test, lm.predict(X_test))}
'''
print(scores)
```

```
Metric          Train               Test
r2_score  0.8246772876773119    0.8177030866648783
MSE       207474.37029707094    209775.29679981153
```

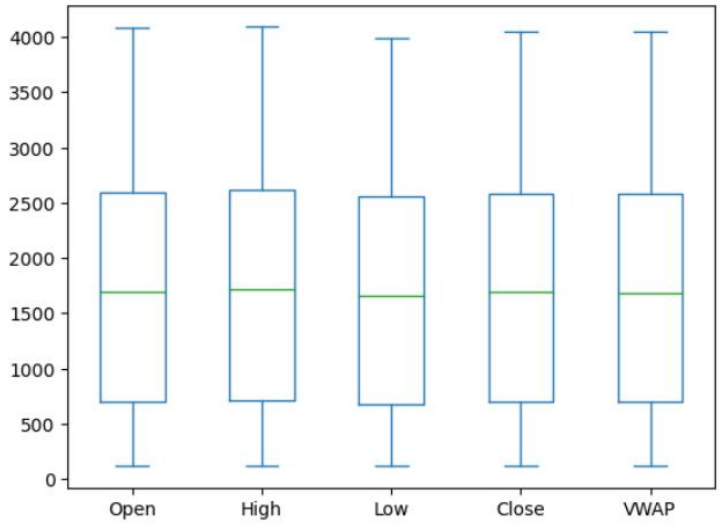
4.1. Training & Testing Accuracy

4.3 Result Analysis and Screenshots

For fresh data, the system correctly forecasts stock prices. Values predicted are fairly accurate. The date was analyzed to predict the stocks on various socio-cultural, economic and environmental factors. For example: Upon analysis, it was found that the sale of vehicles increased in the September-October time as it was time for Dhanteras and Deepavali.

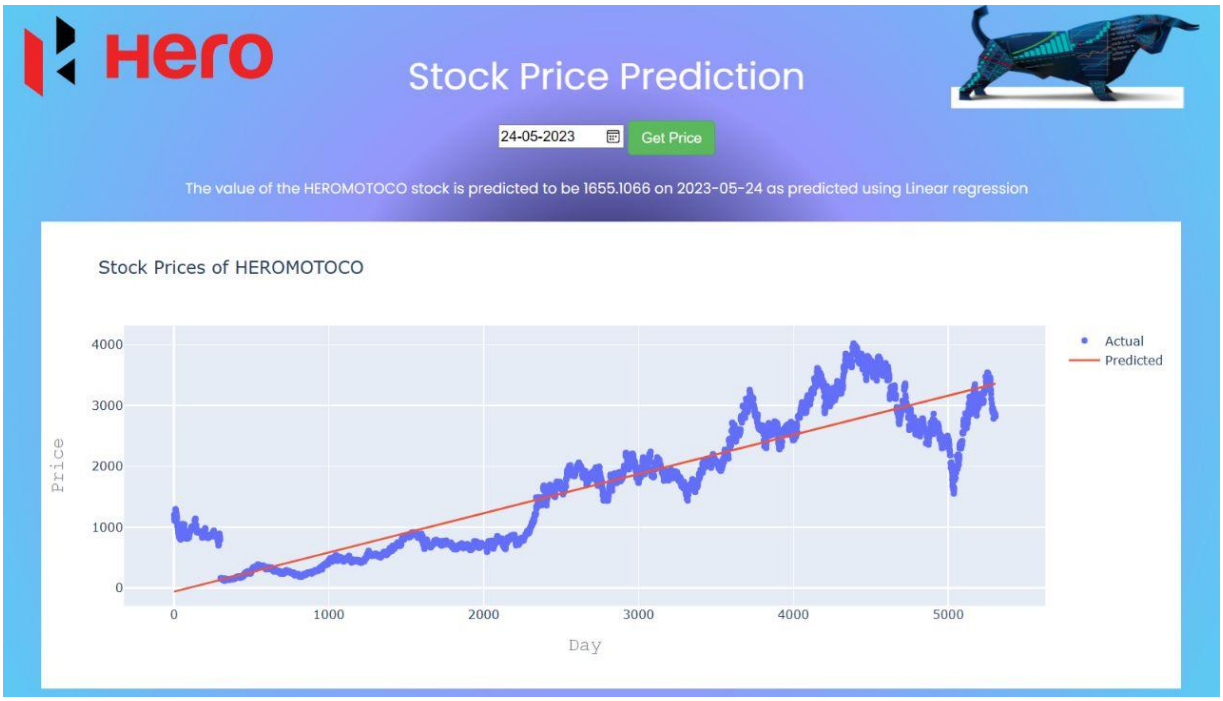
Screenshot of output (Box plot on Jupyter terminal):

```
In [24]:
heromotoco[['Open','High','Low','Close','VWAP']].plot(kind='box')
Out[24]:
<AxesSubplot:>
```



4.2. Box Plot on Jupyter Terminal

Screenshot of output (on web-page):



4.3. Screenshot of Web-page

Chapter 5

Standards Adopted

5.1 Design Standards

Modular design: The project should be organized into manageable modules with clearly defined interfaces to make it simpler to comprehend, test, and maintain.

Documentation: Project documentation should be thorough and include the requirements, design, code, and test cases. The project will be simple for other team members to understand and maintain if it is properly documented.

Scalability: To handle larger datasets and be readily expanded with new features, the project should be planned with scalability in mind.

Readability of the code: The code ought to be well-organized and simple to understand. This facilitates easier maintenance, easier debugging, and better comprehension of the code.

5.2 Coding Standards

Few of the coding standards are:

1. The code is well-documented, and comments have been added to each part to describe what it does.
2. Because of the code's usage of descriptive variable names, it is clear what each variable is used for.
3. The code is divided into several sections for data preparation and cleaning, model development and testing, and visualization, making it simpler to read and comprehend.
4. To make the code simpler and avoid creating anything entirely new, it makes use of third-party libraries like Pandas, NumPy, Matplotlib, and Plotly.
5. To guarantee the model's accuracy, the algorithm makes use of well-known machine learning methods including linear regression and train/test splitting.

5.3 Testing Standards

1. This worldwide standard for software testing, ISO/IEC/IEEE 29119, offers a thorough set of test techniques and procedures. It includes planning, designing, carrying out, and reporting on every facet of software testing.
2. ISO/IEC/IEEE 12207: This standard provides a framework for software life cycle processes, including testing. It defines the activities that are involved in software testing, including planning, specification, execution, and evaluation.

Chapter 6

Conclusion and Future Scope

6.1 Conclusion

In conclusion, the study shows that linear regression may be used to accurately forecast Hero Motors stock prices. A reasonable degree of accuracy in predicting the stock prices was achieved through the examination of past stock prices and the application of the linear regression model. According to the investigation, Hero Motors' closing price has a linear connection with time, and a straightforward linear regression model can accurately forecast future stock values. By adding more pertinent factors and utilizing more sophisticated machine learning techniques, the model's accuracy can be increased even more. Overall, the results of this study point to the potential value of using linear regression to anticipate stock prices for stakeholders and investors.

6.2 Future Scope

Integration with other predictive models: The linear regression model can be merged with other predictive models, like neural networks or time series analysis, to increase the predictability of the results.

Addition of new features: At the moment, the model is only based on past stock prices. The model can be improved, though, by incorporating more elements, such as news sentiment analysis, macroeconomic indicators, or other financial data that could impact stock prices.

Real-time forecasting: The model has the ability to forecast stock values in real-time, which can be helpful for investors who must take quick decisions based on shifting market conditions.

References

For back-end:

- Node.js

For frontend:

- HTML
- CSS

For Data-set:

- Data-set refered from Kaggle

Others:

- Stock Closing Price Prediction using Machine Learning Techniques
- Stock price prediction using linear regression
- Stock Market Analysis Using Linear Regression

INDIVIDUAL CONTRIBUTION REPORT

STOCK PRICES PREDICTION

Chirag Mishra
2005578

Abstract: This project report includes a study on the use of linear regression to forecast Hero Motors stock values. The study explains the feature engineering data pre-processing methods, regression model application, and model performance assessment methods. The study offers insights into potential future research on the subject and suggests that linear regression can be a useful method for predicting stock values.

Individual contribution and learnings: The use of Machine Learning techniques to plot to represent the pattern in the graphical mode was my individual contribution to the stock price prediction project. Further this representation was used to derive the analysis of the stock prices for Hero Motors on a specific time frame and thus be helpful to the client. Further on the front end, I decided the color base and combination for making the webpage and its representation using CSS.

Individual contribution to project report preparation: My individual contribution was the SRS documentation along with the project's implementation while adhering to the standards.

Individual contribution for project presentation and demonstration: My role in preparing the presentation was the inference, limitations and conclusion and presenting and demonstrating the same.

Full Signature of Supervisor: Full signature of the student:

……. …….

INDIVIDUAL CONTRIBUTION REPORT

STOCK PRICES PREDICTION

Sajal Kumar Ujjwal
2005603

Abstract: This project report includes a study on the use of linear regression to forecast Hero Motors stock values. The study explains the feature engineering data pre-processing methods, regression model application, and model performance assessment methods. The study offers insights into potential future research on the subject and suggests that linear regression can be a useful method for predicting stock values.

Individual contribution and learnings: The student performed the task of linking the graph from ML to the webpage. In the webpage itself, he carreid out the entirety of the CSS work. He helped in the pre-processing & the training of the model using the proposed algorithm.

Individual contribution to project report preparation: The student compiled the final project report gathering materials provided by other members of the team along with providing the material for back-end themselves too.

Individual contribution for project presentation and demonstration: The student helped in the final culmination & presentation of the PPT based on searching of the facts & figures used in making the entire project's PPT.

Full Signature of Supervisor: Full signature of the student:

……. …….

INDIVIDUAL CONTRIBUTION REPORT

STOCK PRICES PREDICTION

Sarthak Singh
2005608

Abstract: This project report includes a study on the use of linear regression to forecast Hero Motors stock values. The study explains the feature engineering data pre-processing methods, regression model application, and model performance assessment methods. The study offers insights into potential future research on the subject and suggests that linear regression can be a useful method for predicting stock values.

Individual contribution and learnings: The student designed the front-end of the webpage based on the project using HTML & CSS as their source.

Individual contribution to project report preparation: The student helped in providing information regarding the standards adopted for making the entire project be it testing, coding or even design standards.

Individual contribution for project presentation and demonstration: The student provided research material to be incorporated in the PPT based on the topics he had to work on.

Full Signature of Supervisor: Full signature of the student:

……. …….

Plagiarism Checker X - Report

Originality Assessment

6%

Overall Similarity

Date: May 5, 2023
Matches: 188 / 3022 words
Sources: 19

Remarks: Low similarity detected, check with your supervisor if changes are required.

Verify Report:
Scan this QR Code