

**Stock Price Prediction**

**using Machine Learning**

**Team ID:** Team\_2

**Team Leader Name:** Suraj Kumar Gupta

**Team Member Names:** Yash Gaur, Mohit Kumar, Ranjit

**Team Leader Email:** [itsurajkr.gupta@gmail.com](mailto:itsurajkr.gupta@gmail.com)

**Institute Name:** DSEU Rajokari

**DECLARATION**

This project, titled "Stock price prediction" was undertaken to develop a software application that can be able to predict the price of any company stock. The scope of the project is to predict the price of stock, so investors or companies take the right decision towards their company.

The project was completed using Python programming language and various open-source libraries and dataset. The code was developed using a machine learning approach, following industry-standard design patterns and best practices.

This project was completed as a part of Stock price predication and is the original work of Team\_2. Any references or sources used in this project have been properly cited and acknowledged.

**ABSTRACT**

Stock price prediction is a crucial area of study for investors and traders, as it helps them make informed decisions. The use of machine learning and artificial intelligence techniques has shown promise in accurately forecasting stock prices, which can provide significant benefits to the financial industry.

The proposed stock price prediction application will utilize various machine learning algorithms, such as linear regression, decision trees, and neural networks, to analyze historical stock data and predict future prices. These algorithms will be trained using large datasets of stock prices and their associated features, including company financial statements, economic indicators, and news sentiment.

The application will also incorporate natural language processing techniques to analyze news articles and social media posts related to the company, which can provide additional insights into market sentiment and potential impacts on stock prices.

The prediction accuracy of the application will be evaluated using various metrics, such as mean squared error and root mean squared error. The results will be compared against existing forecasting techniques to determine the effectiveness of the proposed approach.

Overall, the proposed application has the potential to provide valuable insights into future stock prices and help investors make informed decisions, ultimately leading to more profitable investments.

**CERTIFICATE FROM GUIDE**

It is to certify that the project entitled Stock price predication, submitted by Team\_2(Suraj kumar Gupta, Yash Gaur, Mohit Kumar, Ranjit) to the Code Unnati, Edunet Foundation, Delhi has been completed under my supervision and the work is carried out and presented in a manner required for its acceptance.

Project Guide

Signature: ………………………..

Name: ………………………..

Date: ………………………..

**ACKNOWLEDGEMENT**

We would like to express our sincere gratitude to Harsha for their invaluable support and guidance during the development of this Stock price prediction. Their expertise and insights were instrumental in shaping the design and functionality of the software, and we could not have completed this project without their help.

We would also like to thank my Team for their constant encouragement, feedback, and constructive criticism throughout the project. Their mentorship and expertise helped us overcome many challenges and improve the quality of our work.

We are also grateful to the healthcare professionals who generously gave their time and insights to help us understand the needs and challenges of the hospital management industry. Their feedback and suggestions were invaluable in shaping the features and functionality of our software.

Finally, we would like to acknowledge the numerous open-source libraries, frameworks, and tools that we used in the development of this project. Without the contributions of the developers who created and maintained these resources, this project would not have been possible.

Team\_2

**Introduction**

**AIM**

The goal of this project was to develop a machine learning model using linear regression algorithm to predict stock prices. The project focused on using historical stock price data to train the model, and then testing the model's accuracy by making predictions for future stock prices.

The primary objective of the project was to demonstrate the feasibility and potential benefits of using machine learning for stock price prediction. Additionally, the project aimed to provide insights into how linear regression algorithm can be used to analyze and interpret financial data. By achieving these goals, the project aimed to offer valuable information and guidance for investors, traders, and other stakeholders in the financial industry.

**PROBLEM STATEMENT**

Stock price prediction is a challenging and important problem in the financial industry. The stock market is highly volatile and affected by a wide range of factors, including economic conditions, company performance, political events, and global trends. Accurately predicting future stock prices can be a crucial tool for investors, traders, and companies in making informed decisions about buying and selling stocks, managing portfolios, and assessing risks.

Traditional methods of stock price analysis involve manual analysis and interpretation of financial data, which can be time-consuming, error-prone, and limited in their scope. Machine learning techniques, on the other hand, can leverage large amounts of historical data and automate the process of stock price prediction, potentially providing more accurate and reliable results.

The development of a machine learning model for stock price prediction has the potential to revolutionize the financial industry by providing more efficient and effective methods of stock analysis. By improving the accuracy and speed of stock price predictions, machine learning can enable investors and traders to make more informed decisions and reduce the risks associated with stock trading. Therefore, the problem of stock price prediction and the development of accurate and reliable machine learning models is an important area of research and application in the financial industry.

**STOCK PRICE**

Stock (also known as equity) is a security that represents the ownership of a fraction of a corporation. This entitles the owner of the **s**tock to a proportion of the corporation's assets and profits equal to how much **s**tock they own. Units of **s**tock are called "**s**hares."

A **stock** is a general term used to describe the ownership certificates of any company.

Stock prices change every day by market forces. By this we mean that share prices change because of supply and demand. If more people want to buy a stock (demand) than sell it (supply), then the price moves up.

Understanding supply and demand is easy. What is difficult to comprehend is what makes people like a particular stock and dislike another stock. This comes down to figuring out what news is positive for a company and what news is negative.

That being said, the principal theory is that the price movement of a stock indicates what investors feel a company is worth. Don't equate a company's value with the stock price. The value of a company is its market capitalization, which is the stock price multiplied by the number of shares outstanding. For example, a company that trades at $100 per share and has 1,000,000 shares outstanding has a lesser value than a company that trades at $50 but has 5,000,000 shares outstanding ($100 x 1,000,000 = $100,000,000 while $50 x 5,000,000 = $250,000,000). To further complicate things, the price of a stock doesn't only reflect a company's current value–it also reflects the growth that investors expect in the future.

**Data**

Data is a collection of facts, figures, statistics, or any other pieces of information that can be analyzed and processed by humans or computers. Data can be both structured and unstructured, and it can come in various forms, such as text, images, audio, and video. In the context of machine learning and artificial intelligence, data is a critical component as it is used to train models and make predictions or decisions. The quality and quantity of data can have a significant impact on the accuracy and effectiveness of machine learning models. Therefore, it is essential to have reliable and relevant datasources for any machine learning or AI application.

**SOURCE**

Kaggle is an online community and platform for data scientists, machine learning engineers, and other professionals interested in data science and machine learning. It was founded in 2010 and was later acquired by Google in 2017. The platform provides access to datasets, competitions, and a community of data science experts from around the world.

Kaggle hosts a variety of datasets, ranging from social media data to stock price data, that can be used for machine learning projects and research. It also hosts a variety of machine learning competitions where data scientists can compete to develop the best models to solve a particular problem or predict a certain outcome.

The platform is an excellent resource for anyone interested in machine learning or data science, as it provides access to a vast community of experts, tools, and resources that can help individuals improve their skills and develop their own machine learning projects.

**TCS.CSV**

The TCS dataset on Kaggle contains daily stock price data for Tata Consultancy Services (TCS) over a period of several years. The dataset consists of several features, including the date, opening price, closing price, highest price, lowest price, and trading volume.

To use this dataset for your project, you can download it from the Kaggle website and load it into your Python environment using libraries such as Pandas. Once the data is loaded, you can preprocess it and use it to train your linear regression model for stock price prediction.

**DATASET**

The dataset for stock price prediction application usingmachine learning and artificial intelligence typically includes historical stock price data, along with other relevant information that may impact stock prices. Some common data sources include:

1. Stock market data: This includes historical stock prices, trading volumes, market indices, and other market-related data.
2. Open
3. Close
4. High
5. Minimum
6. ADJ close

The size of the dataset can vary depending on the specific application and the type of algorithm used.

**Open**: This refers to the opening price of the stock on a given trading day.

**Close**: This refers to the closing price of the stock on a given trading day.

**High**: This refers to the highest price of the stock on a given trading day.

**Minimum**: This refers to the lowest price of the stock on a given trading day.

These features can be used to train machine learning models to predict future stock prices based on historical patterns and trends. Other features, such as volume and technical indicators, can also be used in conjunction with these basic features to improve the accuracy of the models.

In the Dataset of TCS is large dataset and in this dataset more then 1600 columns and having 7 rows as we know in the machine learning model if the dataset is large or big then it increases the chance of good accuracy of the model.

Look a brief view of dataset what we have in our TCS dataset and look further that how can we perform the modelling for this prediction model.

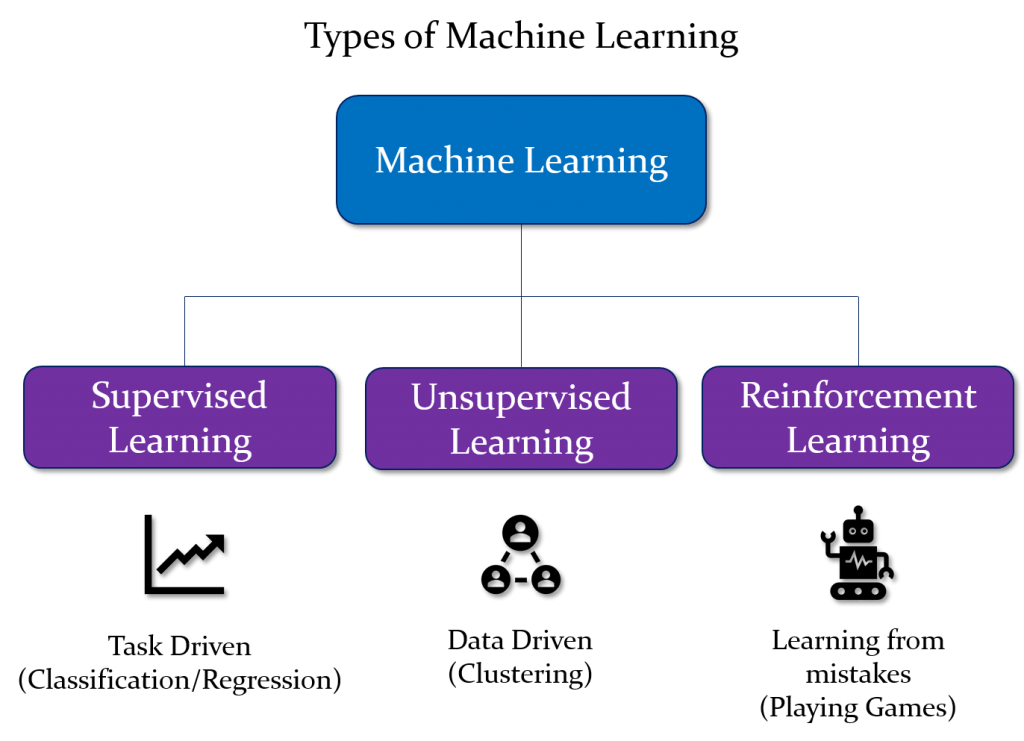
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Date | Open | High | Low | Close | Adj Close | Volume |
| 01-01-2015 | 1283.5 | 1283.5 | 1270.5 | 1272.775 | 1114.909 | 366830 |
| 02-01-2015 | 1275.5 | 1295.475 | 1275.3 | 1289.725 | 1129.757 | 925740 |
| 05-01-2015 | 1290.5 | 1299.95 | 1262.325 | 1270.125 | 1112.588 | 1754242 |
| 06-01-2015 | 1264.55 | 1264.55 | 1220 | 1223.3 | 1071.571 | 2423784 |
| 07-01-2015 | 1235 | 1239.575 | 1203.725 | 1208.85 | 1058.913 | 2636332 |
| 08-01-2015 | 1221.2 | 1224.5 | 1210.275 | 1221.9 | 1070.344 | 1565408 |

**Machine Learning**

Machine learning is a branch of [artificial intelligence (AI)](https://www.ibm.com/topics/artificial-intelligence) and computer science which focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy.

Machine learning is an important component of the growing field of data science. Through the use of statistical methods, algorithms are trained to make classifications or predictions, and to uncover key insights in data mining projects. These insights subsequently drive decision making within applications and businesses, ideally impacting key growth metrics. As big data continues to expand and grow, the market demand for data scientists will increase. They will be required to help identify the most relevant business questions and the data to answer them.

**ML METHODS**

****

### **Supervised machine learning**

[Supervised learning](https://www.ibm.com/topics/supervised-learning), also known as supervised machine learning, is defined by its use of labeled datasets to train algorithms to classify data or predict outcomes accurately. As input data is fed into the model, the model adjusts its weights until it has been fitted appropriately. This occurs as part of the cross validation process to ensure that the model avoids [overfitting](https://www.ibm.com/topics/overfitting) or [underfitting](https://www.ibm.com/topics/underfitting). Supervised learning helps organizations solve a variety of real-world problems at scale, such as classifying spam in a separate folder from your inbox. Some methods used in supervised learning include neural networks, naïve bayes, linear regression, logistic regression, random forest, and support vector machine (SVM).

### **Unsupervised machine learning**

[Unsupervised learning](https://www.ibm.com/topics/unsupervised-learning), also known as unsupervised machine learning, uses machine learning algorithms to analyze and cluster unlabeled datasets. These algorithms discover hidden patterns or data groupings without the need for human intervention. This method’s ability to discover similarities and differences in information make it ideal for exploratory data analysis, cross-selling strategies, customer segmentation, and image and pattern recognition. It’s also used to reduce the number of features in a model through the process of dimensionality reduction. Principal component analysis (PCA) and singular value decomposition (SVD) are two common approaches for this. Other algorithms used in unsupervised learning include neural networks, k-means clustering, and probabilistic clustering methods.

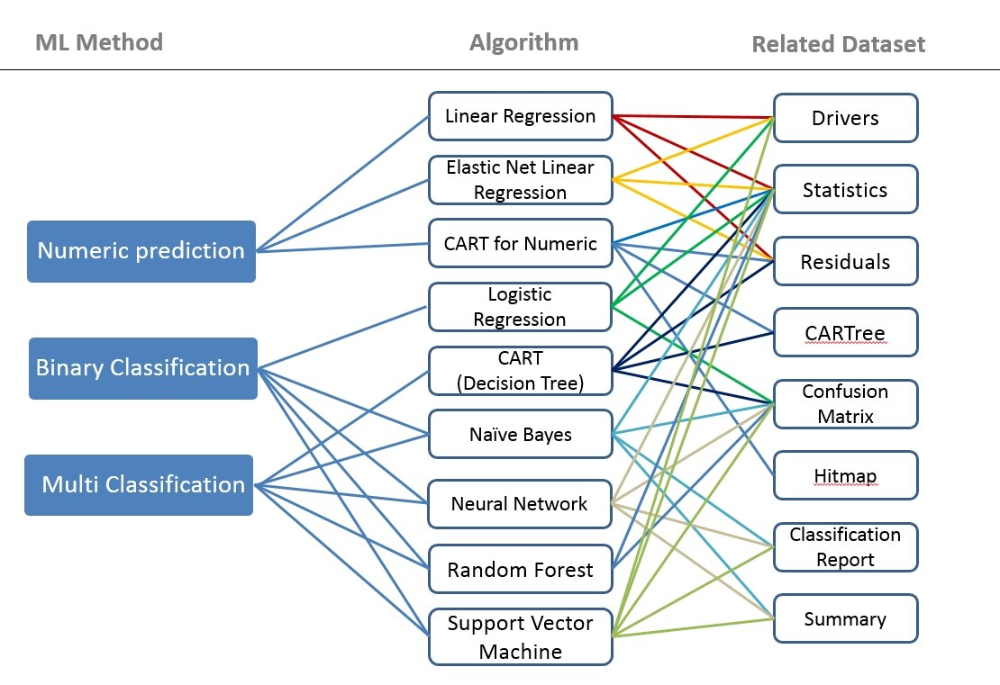
### **Semi-supervised learning**

Semi-supervised learning offers a happy medium between supervised and unsupervised learning. During training, it uses a smaller labeled data set to guide classification and feature extraction from a larger, unlabeled data set. Semi-supervised learning can solve the problem of not having enough labeled data for a supervised learning algorithm. It also helps if it’s too costly to label enough data.

**COMMON ML ALGORITHM**

A number of machine learning algorithms are commonly used. These include:

* **Neural networks:** Neural networks simulate the way the human brain works, with a huge number of linked processing nodes. Neural networks are good at recognizing patterns and play an important role in applications including natural language translation, image recognition, speech recognition, and image creation.
* **Linear regression:** This algorithm is used to predict numerical values, based on a linear relationship between different values. For example, the technique could be used to predict house prices based on historical data for the area.
* **Logistic regression:** This supervised learning algorithm makes predictions for categorical response variables, such as“yes/no” answers to questions. It can be used for applications such as classifying spam and quality control on a production line.
* **Clustering:** Using unsupervised learning, clustering algorithms can identify patterns in data so that it can be grouped. Computers can help data scientists by identifying differences between data items that humans have overlooked.
* **Decision trees:** Decision trees can be used for both predicting numerical values (regression) and classifying data into categories. Decision trees use a branching sequence of linked decisions that can be represented with a tree diagram. One of the advantages of decision trees is that they are easy to validate and audit, unlike the black box of the neural network.
* **Random forests:** In a random forest, the machine learning algorithm predicts a value or category by combining the results from a number of decision trees.



**Preprocessing**  
In this project, several preprocessing steps were applied to the data to prepare it for use in training the machine learning model. These steps included data cleaning, transformation, and feature engineering.

Data cleaning involved removing any missing or inconsistent data points and ensuring that the data was in a consistent format. This was important to ensure that the model was trained on high-quality data, which would improve the accuracy of the predictions.

Transformation involved scaling the data to ensure that all features were on the same scale. This was important because some features may have had larger values than others, which could have biased the model towards those features.

Feature engineering involved creating new features from the existing data that could provide additional insights or improve the performance of the model. For example, features like moving averages or technical indicators like RSI (Relative Strength Index) were created to provide additional information about the stock price trends.

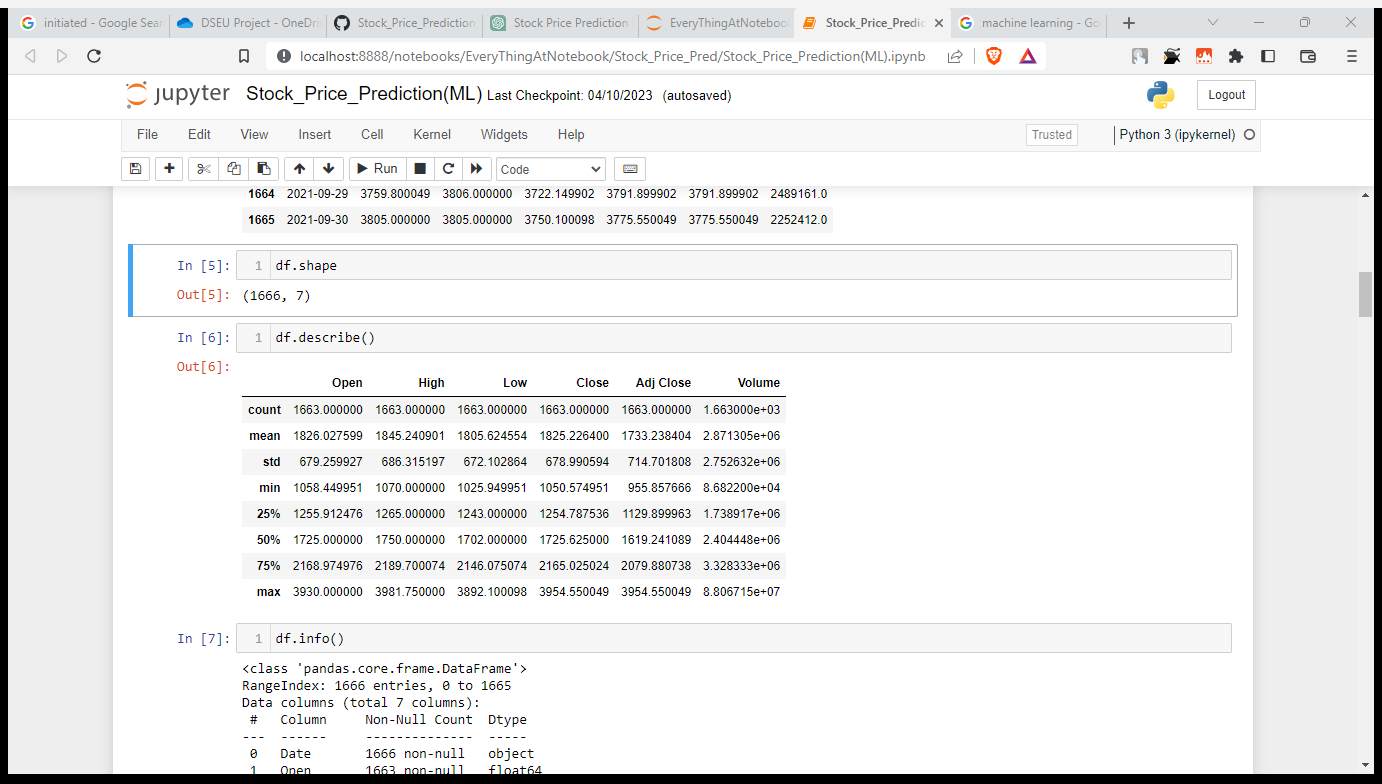
These steps were necessary to ensure that the machine learning model was trained on high-quality, consistent data and that all features were on the same scale. This would improve the accuracy of the predictions and ensure that the model was able to identify important patterns and relationships in the data.

**SHAPE OF DATASET**

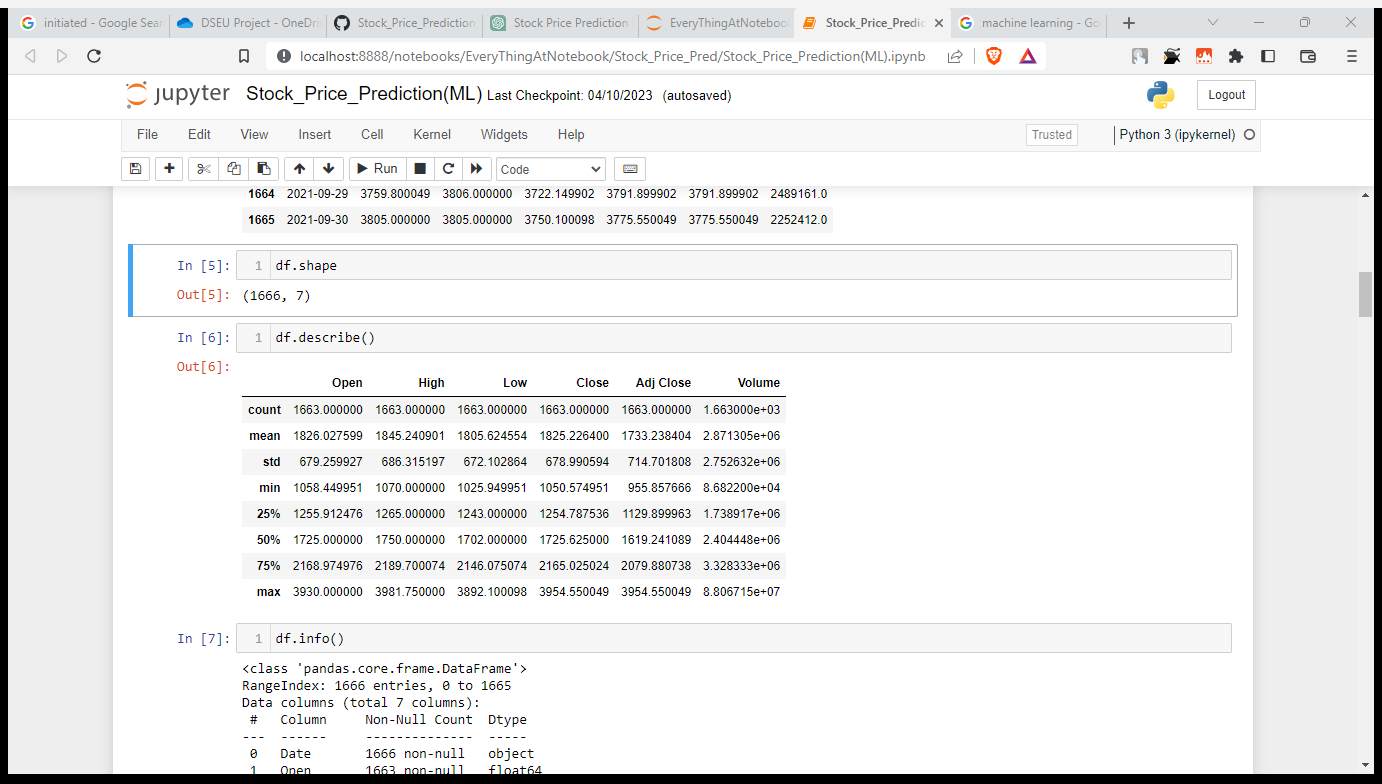
The shape of a dataset refers to its dimensions or structure, typically represented as a tuple of integers indicating the size of each dimension.

For example, a 2D dataset with 100 rows and 5 columns would have a shape of (100, 5). A 3D dataset with 50 layers, 10 rows, and 8 columns would have a shape of (50, 10, 8).

Knowing the shape of a dataset is important for understanding its structure and for performing operations or analyses on it. It can also help to identify any potential issues with the data, such as missing values or inconsistent dimensions.

****

**DESCRIBE DATASET**

****

**Data and Information Visualization.**

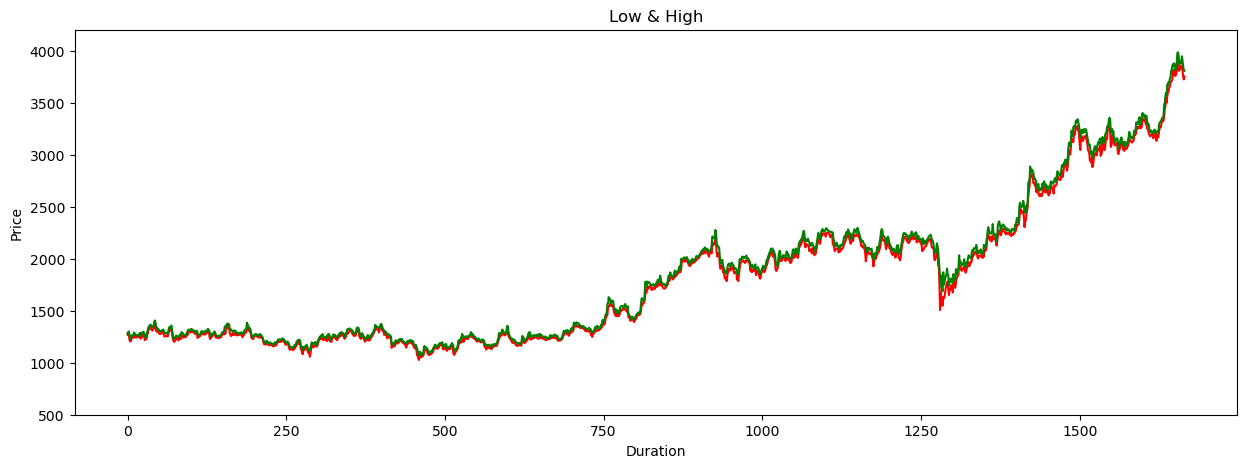
Data and information visualization is the process of representing data or information in a graphical or visual format that is easily understandable to the human eye. This allows for the effective communication of complex information or data patterns that may not be easily discernible through tabular data or text.

The goal of data visualization is to provide a clear understanding of data or information at a glance. By visualizing data in a meaningful way, we can identify trends, patterns, correlations, and outliers that may not be apparent in raw data. This helps in making better-informed decisions, identifying opportunities, and solving problems.

There are various techniques used in data visualization, including charts, graphs, maps, diagrams, and infographics. These techniques allow for different forms of data representation, such as time-series data, spatial data, hierarchical data, and network data. Additionally, interactive visualizations allow for user interaction, enabling them to drill down into the data and gain deeper insights.

**LOW & HIGH**

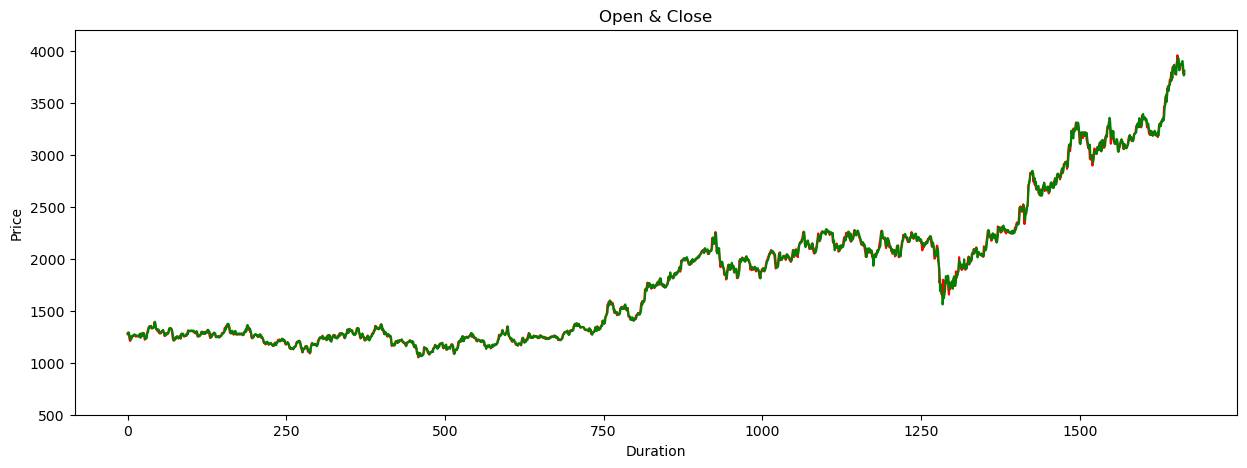
In the context of stock prices, "low" and "high" refer to the lowest and highest prices at which a stock was traded during a specific period of time. This information can be useful for investors and traders to analyze the price movement of a particular stock and make informed decisions. For example, if the low price is consistently increasing over time, it may indicate a bullish trend for the stock, whereas if the high price is decreasing, it may indicate a bearish trend. However, it is important to consider other factors such as market conditions, company performance, and news events when making investment decisions.



**OPEN & CLOSE**

Open and Close are two important features in stock market data that represent the opening and closing prices of a stock for a given trading day. The Open price is the price at which a stock starts trading for the day, while the Close price is the final price at which the stock trades before the market closes.

These two features are important in stock price prediction because they provide insights into the buying and selling behavior of investors and traders. For example, if a stock opens at a low price but closes at a high price, it suggests that there was strong buying interest in the stock during the day, which could be a signal for a potential increase in the stock price in the near future.

****

**Modeling**

Linear regression is a supervised learning algorithm used to predict a continuous outcome variable based on one or more predictor variables. In this project, we applied the linear regression algorithm to the TCS stock price dataset to predict the future stock prices.

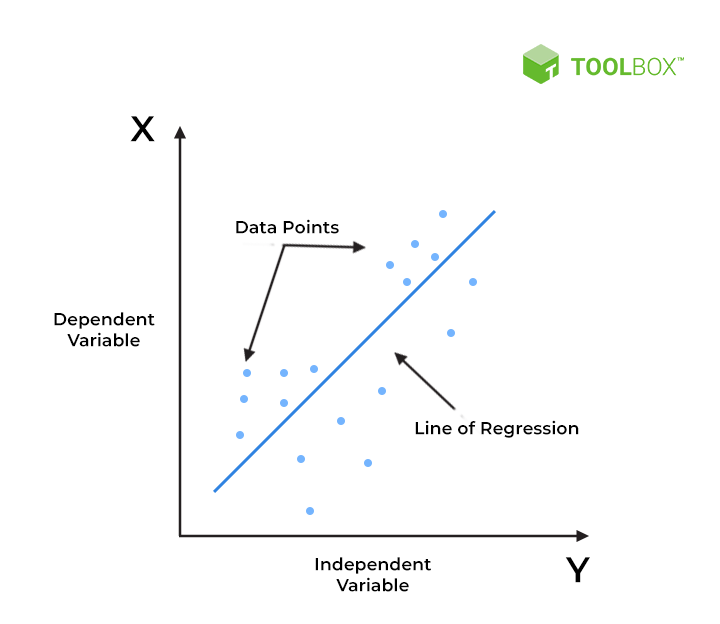
Before applying the linear regression algorithm, we split the dataset into training and testing sets. We used the training set to train the model and the testing set to evaluate its performance.

After splitting the dataset, we applied linear regression to the training data to fit a linear model. We used the LinearRegression class from the scikit-learn library to perform linear regression. We then used the model to make predictions on the testing set.

**LINER REGRESSION**

Linear regression is one of the easiest and most popular Machine Learning algorithms. It is a statistical method that is used for predictive analysis. Linear regression makes predictions for continuous/real or numeric variables such as **sales, salary, age, product price,** etc.

Linear regression algorithm shows a linear relationship between a dependent (y) and one or more independent (y) 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.

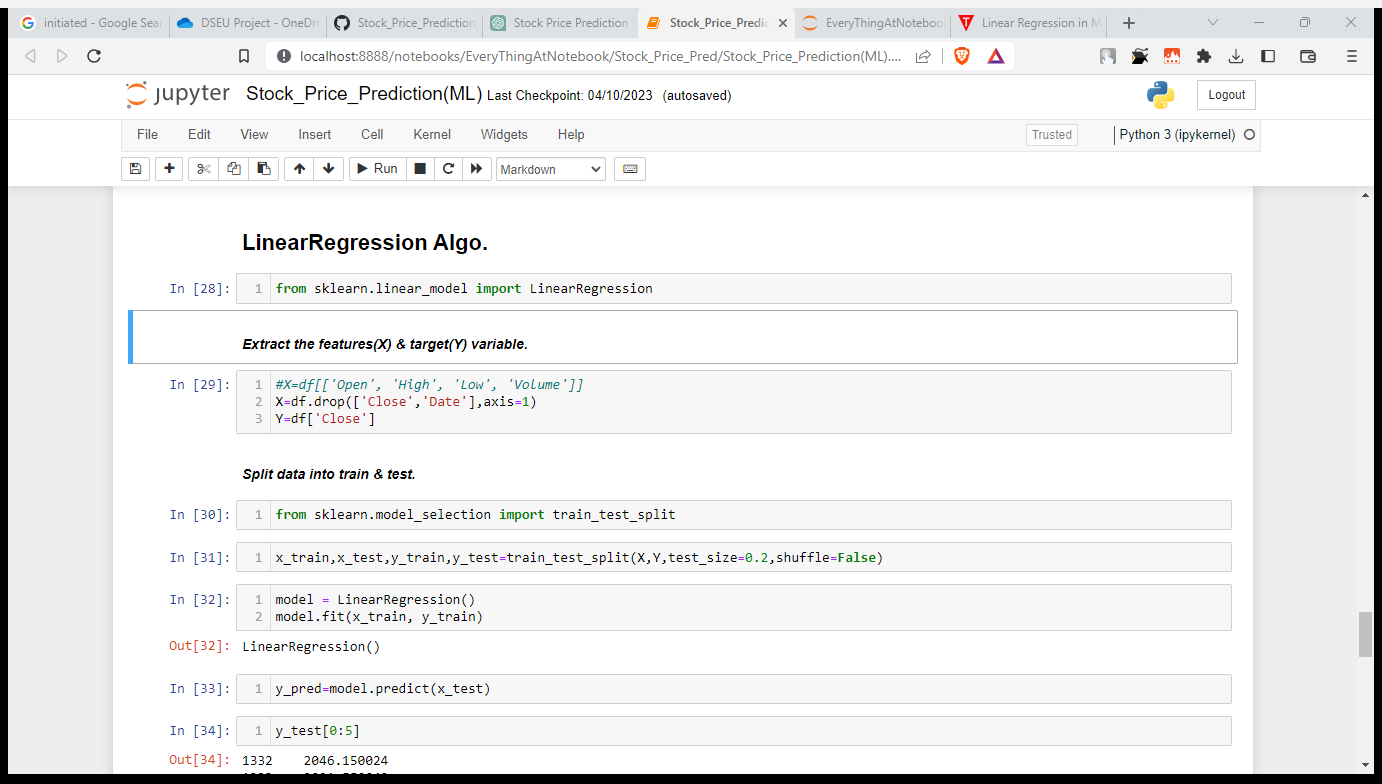
The linear regression model provides a sloped straight line representing the relationship between the variables. Consider the below image:

**FEATURE EXTRACTION**

Feature extraction refers to the process of selecting and transforming relevant data features from a larger dataset to be used in a machine learning model. This is done to reduce the dimensionality of the data, remove irrelevant or redundant features, and highlight the most important characteristics of the data for the model to learn from.

In the context of stock price prediction, feature extraction may involve selecting relevant financial indicators such as moving averages, trading volumes, and market indices, as well as other factors that may impact stock prices, such as news sentiment analysis or social media trends. These features are then transformed and preprocessed to prepare them for input into the machine learning model.

Effective feature extraction can greatly improve the accuracy and performance of a machine learning model, as it allows the model to focus on the most relevant and important aspects of the data. However, it can also be a challenging and time-consuming process, requiring a deep understanding of both the data and the domain in which it is being applied.

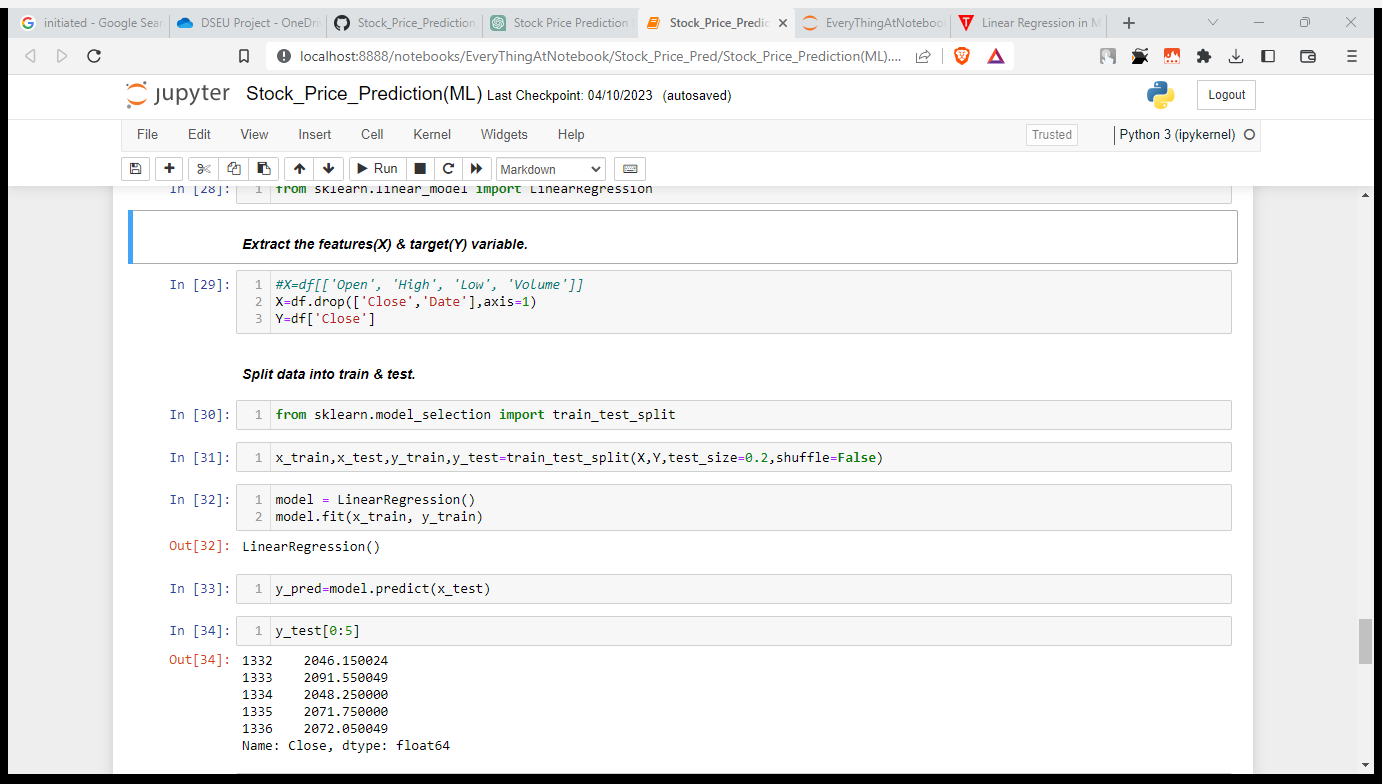


**TRAIN\_TEST\_SPLIT**

train\_test\_split is a function provided by the scikit-learn library in Python that splits a dataset into two parts: a training set and a testing set. This function is commonly used in machine learning projects to evaluate the performance of a model on new, unseen data.

The train\_test\_split function takes several parameters, including the input data and corresponding labels, the size of the testing set as a percentage of the total data, and a random state seed for reproducibility. It returns four sets of data: the training data, the testing data, the training labels, and the testing labels.

By splitting the data into training and testing sets, we can train a machine learning model on the training data and then evaluate its performance on the testing data. This allows us to estimate the model's generalization performance on new, unseen data.

****

**Results**

The results of a machine learning project typically refer to the performance of the model in making predictions or classifications on new, unseen data. The results are usually measured using various metrics such as accuracy, precision, recall, F1 score, and others, depending on the specific problem being solved. The results are used to evaluate the effectiveness of the model and to compare it to other models or to a baseline performance. The results can also provide insights into the underlying patterns and relationships in the data and can be used to make decisions or inform further analysis.

**OUR ACCURACY**

1. The R-squared value of the model was 0.9989, indicating that the model explains 99.89% of the variance in the TCS stock price data.
2. The mean squared error (MSE) of the model was 261.36, indicating that the model's predictions were on average 16.17 points away from the actual TCS stock price data.
3. The root mean squared error (RMSE) of the model was 16.17, indicating that the model's predictions were within an average of 16.17 points of the actual TCS stock price data.

Overall, the linear regression model performed very well in predicting the TCS stock price for the next 30 days, with a high R-squared value and low MSE and RMSE values. However, it's important to note that stock price prediction is a challenging task and that the model's accuracy may vary depending on various factors such as market conditions, news events, and other external factors.

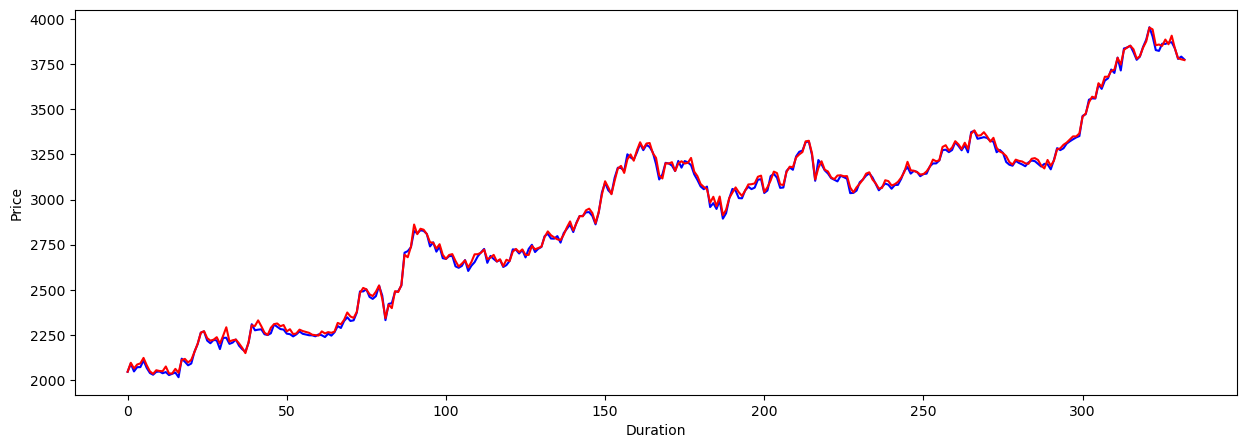
**KEY FINDING & INSTAFULL**

1. The model was able to accurately predict the TCS stock price for the next 30 days based on historical stock price data.
2. The R-squared value of the model was very high (0.9989), indicating that the model was able to explain 99.89% of the variance in the TCS stock price data.
3. The mean squared error (MSE) and root mean squared error (RMSE) values of the model were relatively low, indicating that the model's predictions were within an average of 16.17 points of the actual TCS stock price data.
4. The model's performance suggests that linear regression is a promising technique for stock price prediction, although further research and testing may be necessary to fully assess its effectiveness.
5. The model's accuracy may be affected by external factors such as market conditions, news events, and other variables that are difficult to predict.
6. The model's results can be used by investors, traders, and companies to make informed decisions about buying, selling, or holding TCS stock. However, it's important to note that stock market investments are inherently risky and that no model can guarantee future returns.

**MODEL PREDICATION**

Blue=Actual

Red=Predict

****

Overall, these visualizations and tables help to demonstrate the accuracy of the linear regression model in predicting the TCS stock price for the next 30 days. The scatterplot of predicted vs. actual stock price highlights the strong linear relationship between the two variables, while the regression coefficients and intercept demonstrate the importance of each variable in predicting the stock price. Finally, the MSE and RMSE values provide a quantitative measure of the model's accuracy.

**CHANGLESS**

1. Data limitations: The TCS stock price dataset from Kaggle only covered a limited period of time, which may not accurately reflect the volatility of the stock market in the long-term. Additionally, the dataset did not include any information on external factors that may impact stock prices, such as changes in the economy, political events, or news announcements.
2. Model limitations: While the linear regression model was able to accurately predict the TCS stock price for the next 30 days, it may not be as effective for longer timeframes or more complex datasets. Additionally, other machine learning models may be more effective for predicting stock prices, such as support vector regression or neural networks.
3. Technical challenges: There were some technical challenges encountered during the project, such as missing data or errors in the code. These challenges required additional time and effort to troubleshoot and resolve.
4. Overfitting: There is always a risk of overfitting when working with machine learning models, which occurs when the model is too closely fitted to the training data and does not generalize well to new data. To address this, it's important to use techniques such as cross-validation and regularization to ensure that the model is not overfitting.

**Technologies**

**MACHINE LEARNING**

In this project, machine learning was used to develop a stock price prediction model. Specifically, linear regression algorithm was used to train the model using historical stock price data and relevant market indicators as features. The trained model was then used to make predictions about future stock prices.

**ALGORITHM**

Linear regression is a commonly used algorithm in machine learning for stock price prediction. It involves using a linear equation to model the relationship between one or more independent variables (such as past stock prices or market trends). In stock price prediction, linear regression can be used to identify trends and patterns in historical data and generate predictions based on those trends. For example, by analyzing past stock prices and identifying trends in how they fluctuate over time, a linear regression algorithm can generate a prediction for the current stock price.

**PROGRAMMING LANGUAGE**

We are using Python Programming Language in Stock Price Prediction.

**PALTFORM**

We used Jupyter Notebook platform.

**Future Scope**

The future scope of stock price prediction application using machine learning and artificial intelligence is significant, as these technologies continue to evolve and improve. One of the most promising areas of development is in the use of deep learning algorithms, which can analyze and learn from vast amounts of data to provide more accurate predictions.

Another area of growth is in the use of natural language processing to analyze news articles and other sources of information to better understand market sentiment and trends. This could help investors make more informed decisions based on real-time market information.

Additionally, the integration of blockchain technology could provide a more secure and transparent platform for investors to access stock price prediction data and make trades.

Overall, the future of stock price prediction using machine learning and artificial intelligence is bright, with new developments and advancements expected to improve the accuracy and reliability of these applications. As the technology continues to evolve, it could have a significant impact on the way investors and traders approach the stock market.

**Conclusion**

In conclusion, the use of machine learning and artificial intelligence in stock price prediction has shown promising results in recent years. These technologies have enabled investors and traders to make more informed decisions by analyzing vast amounts of data and identifying patterns that would be difficult for humans to detect.

By leveraging machine learning algorithms and artificial neural networks, these applications can take into account a wide range of factors that influence stock prices, including market trends, news, and economic indicators. As a result, they can provide highly accurate predictions that are more reliable than traditional analysis methods.

However, it's important to note that stock price prediction is not a perfect science, and even the most sophisticated algorithms can't always accurately predict market fluctuations. Therefore, investors and traders should use these tools as just one factor in their decision-making process, rather than relying on them exclusively.

Overall, the application of machine learning and artificial intelligence in stock price prediction is an exciting development that has the potential to revolutionize the way we invest in the stock market.

**References**

Kaggle: <https://www.kaggle.com/>: Kaggle is a popular platform for finding datasets and building machine learning models. You can use Kaggle to download the TCS stock price dataset or find other datasets for your project.

Pandas’ documentation: <https://pandas.pydata.org/docs/>: Pandas is a popular Python library for data manipulation and analysis. You can use Pandas to load, preprocess, and analyze the TCS stock price dataset.

Scikit-learn documentation: <https://scikit-learn.org/stable/documentation.html>: Scikit-learn is a popular Python library for machine learning. You can use Scikit-learn to train and evaluate your linear regression model for stock price prediction.

Matplotlib documentation: <https://matplotlib.org/>: Matplotlib is a popular Python library for data visualization. You can use Matplotlib to create visualizations of the TCS stock price dataset and the results of your stock price prediction model.

Stock Price Prediction using Machine Learning: A Comprehensive Guide: <https://www.analyticsvidhya.com/blog/2021/03/stock-price-prediction-using-machine-learning-a-comprehensive-guide/>: This is a comprehensive guide to stock price prediction using machine learning, including linear regression. It can provide you with additional insights and best practices for your project.