Data and Web Mining

Project Topic: TESLA Stock Prediction

Batch-2

Varish Sanghvi - AP19110010004
Gollapalli Sairam - AP19110010019
Kilaru Sravan - AP19110010030
Bavineni Nitin Krishna - AP19110010091
Varun Sai Karasala - AP19110010243

Abstract:
Stock Market Prediction is the method of determining future values of a company's stock. Due to the fluctuating nature of the stock, the stock market is too difficult to predict as stock prices are constantly changing every day. Estimating the stock market has a high demand for stock customers. Applying all extracted rules at any time is a major challenge to estimate the future stock price with high accuracy. In this project, a model is developed to predict the stock values of a company using different data mining techniques. Every day a huge number of companies will take part in the stock market by trading or investing. Thus analyzing such a huge market will prove beneficial to all investors of the system. An application which focuses on the patterns generated in this stock trade over the period of time, and extracting the knowledge from those patterns to predict future behavior of the stock market is essential. An application representing the information in visual form for user interpretation to keep the company stock or to sell the company's stock is a key requirement. In This Model, We proposed the application of Data mining Using Python to predict Stock Market prices of TESLA company and it could be used to guide an investor's decisions. The algorithm can be used for training a set of market data collected for the period of time.

Introduction:
Stock price prediction is a heated topic in prediction study of the financial area. The stock market is essentially a non-linear, nonparametric system that is extremely hard to model with any reasonable accuracy. Investors have been trying to find a way to predict stock prices and to find the right stocks and right timing to buy or sell. Most of the techniques used in technical analysis are highly subjective in nature and have been shown not to be statistically valid. Recently, data mining techniques such as linear regression, decision trees have been applied to this area. Data mining refers to extracting or mining knowledge from large data stores or sets. Some of its functionalities are the discovery of concept or class descriptions, associations and correlations, classification, prediction, clustering, trend analysis, outlier and deviation analysis, and similarity analysis.

Stock Market Prediction has always attracted people interested in investing in share market and stock of a company for large profits but it is very difficult to predict the stock values of a company as it depends on many factors and 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. However, with the introduction of Data Mining and its strong algorithms such as linear regression,decision trees, SVM and etc, the Stock Market Prediction advancements have begun as these are widely utilized by many organizations in Stock market prediction.

“The acceleration of the world's transition to sustainable energy” - with this mission statement Tesla has set the bars extraordinarily high in all its business sectors. The project goal is to predict the stock price of the Tesla company and we use different algorithms to predict it and in this project, we propose a model for forecasting the stock market trends based on the technical analysis using historical stock market data we are having.

Problem Survey:

- Linda Kaufman and Alexander J. Smola proposed a machine learning algorithm to determine the stock price using the concept of decision tree in their paper called “THE STOCK PREDICTION USING MACHINE LEARNING” which was published in International Research Journal of Engineering and Technology.
- Mohamma; Ashwini Kl; Tapas Kumar have worked on a similar project and published a paper titled “An Approach for Prediction of stock price using Machine Learning Algorithm” in which they collected a dataset of 900 cases and used few different attributes to predict the stock price using logistic regression algorithm. This paper is published in the 2019 International Conference on Electronics and Sustainable Communication Systems (ICESC).

- Jasic and Wood (2004) developed a machine learning model to predict daily stock market index returns using data from several global stock markets. The focus is on trying to support profitable trading.
- Chavan and Patil (2013) contribute to our understanding of stock market prediction by surveying different model input parameters found in nine published articles. They attempt to find the most important input parameters that produce better model prediction accuracy. Based on their survey, they find that most ML techniques make use of technical variables instead of fundamental variables for a particular stock price prediction.
- Behrouz Far, Pumsirirat and Garg Ishan have worked on a model using three different approaches namely Logistic Regression(LR), Decision Tree (DT) and Random Forest (RF) and concluded that Logistic Regression provided more accurate results compared to the other two approaches in their paper “Prediction of stock price using Machine Learning” published in International Journal of Advanced Science and Technology.

Proposed Model:

Our proposed model is to predict the TESLA stock price using the Linear Regression algorithm. Linear regression is the most fundamental and widely used kind of predictive analysis. The goal of regression is to look at two things: >>do a set of predictor variables do a good job of predicting an outcome (dependent) variable; and >>which variables in particular are significant predictors of the outcome variable, and how do they impact the outcome variable.

These regression estimations are used to illustrate how one dependent variable interacts with one or more independent variables. The simplest version of the regression equation with one dependent and one independent variable is y = c + b*x, where,

- y represents the estimated dependent variable score
- c represents the constant
- b represents the regression coefficient
- x represents the independent variable score

Assumptions to be considered while performing linear-regression analysis are:

- Data should be quantifiable for both dependent and independent variables.
- The dependent variable's distribution must be normal for each value of the independent variable. For all values of the independent variable, the variance of the dependent variable's distribution should be constant. The dependent variable should have a linear relationship with each independent variable, and all observations should be independent.

Steps and Assumptions to be considered:

1. Before proceeding, double-check that your data can be analyzed using linear regression. Before your data may be used, some assumptions must be satisfied.
2. On a continuous scale, the variables must be measured. Continuous variables include things like time, sales, weight, and test results.
3. Use a scatter plot to quickly see if the two variables are connected linearly.
4. The observations should not be connected in any way (that is, there should be no dependency).
5. In your data, there should be no obvious outliers.

Detailed Explanation of proposed model

- First we had taken a TESLA stock prediction dataset from the kaggle website and then loaded the dataset for preparing the model.
- Once data is loaded then we perform a few data pre-processing techniques such as (data cleaning, data reduction, data transformation) on the dataset.
- After completion of pre-processing we had visualized the pre-processed data to make some assumptions out of them.
- Once we are obtained with the pre-processed data then we divide the data into test data and train data to prepare the model.
- Now we need to train the model using Linear regression and then we are obtained with the results of our proposed model.

Results:

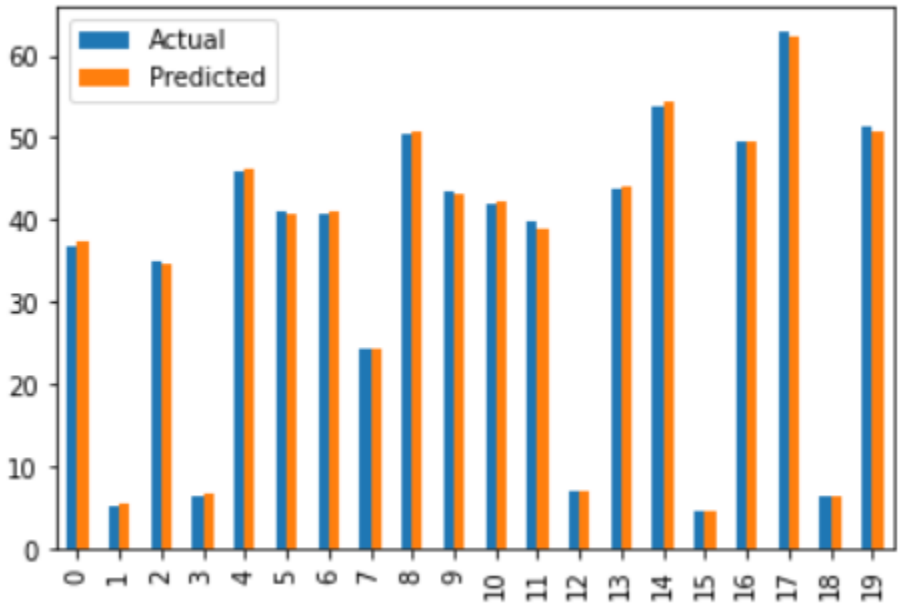
After training the model using the Linear regression algorithm we got an accuracy of **99.7%**.

	Actual	Predicted
0	36.903999	37.478809
1	5.278000	5.388071
2	35.066002	34.508804
3	6.454000	6.635981
4	45.816002	46.114189
5	41.080002	40.762093
6	40.807999	40.973945

The above image clearly shows the actual values vs the predicted values.
Where,

- Actual values are given in the dataset.
- Predicted values are obtained after training the model.

Here below we can see the bar graph of actual values vs predicted values for 20 rows.



From the Bar graph we can see that there is very little difference between the actual values and predicted values.

Conclusion:

Regression analysis is a powerful statistical process to find the relations within a dataset, with the key focus being on relationships between the independent variables (predictors) and a dependent variable (outcome). It can be used to build models for inference or prediction. Among several methods of regression analysis, linear regression sets the basis and is quite widely used for several real-world applications.

The model we have built can be used for inference of how the different predictors influence the outcome. It is far from perfect. There is still the presence of non-linearity and non-constant variance of errors. Moreover, the outliers and leverage points should be analyzed to find a better model. It may not (and most probably won't) give similar results when used to predict the outcome for new, unseen data. In order to use it for prediction, more concrete measures should be taken for ensuring the accuracy of the model. It still helps by providing good estimations of the significant relations between the predictors and the outcome. These estimations can be used to summarize the data in a more useful and presentable way and companies or any stakeholders or any stock market interested people can use this model to invest their money to buy or sell the stock of TESLA and manage their decisions accordingly.

References:

- Agresti, A. (1996). An Introduction to Categorical Data Analysis. New York: Wiley.
- Allen, D. M. (1974). The relationship between variable selection and prediction. Technometrics, 16, 125–127.
- Bowman, A. and Azzalini, A. (1997). Applied Smoothing Techniques for Data Analysis. Oxford: Oxford University Press.
- https://www.kaggle.com/
- https://www.kaggle.com/datasets/timoboz/tesla-stock-data-from-2010-to-2020