Analysing Tesla Stock Prices Using Machine Learning Algorithm

Dr. Chaitanya Kishore Reddy.M¹ ¹Professor & Dean-IT, NRI Institute of Technology, A.P, India-521212

V.Luke Sundar³ ³UG Scholar, Dept. Of IT, NRI Institute of Technology, A.P, India-521212

Abstract:- Long-term research has focused heavily on the prediction of stock values. Predicting stock prices has been one of the biggest concerns in recent years. Trading stocks is a significant economic activity that contributes to society and enables individuals to increase their income. Making a prediction about the stock market involves attempting to anticipate the future value of a company's stock or another financial instrument traded on a stock exchange. Gains for investors will be maximized if a stock's future price can be accurately predicted. The stock market's issue is that it occasionally displays an erratic pattern that could result in a crash. Stock price predictions frequently involve machine learning techniques. We employ a variety of machine learning techniques, both supervised and unsupervised, to inform investors of stock price increases and decreases. Data acquisition, dataset preprocessing, feature extraction, feature-based stock price prediction, and result presentation were the five stages of the process. Data is initially gathered from various social media platforms and historical company information. Pre-processing is the second phase, where incorrect, duplicate, and dirt data are removed. In the third phase, data sets are reduced and meaningful data are chosen. In the fourth phase, predictions are made utilizing various machine learning methods, including supervised and unsupervised learning techniques. Now, various methods are used in the final phase to determine correctness.

Keywords:- Tesla, Elon Musk, Stock Market, Linear Regression, Machine Learning

I. INTRODUCTION

Tesla Inc., an American automaker, was established in 2003. It is one of the businesses that produced electric vehicles at a level that allowed for their availability on the consumer market[Fig:1]. The stock price of Tesla began to decline in the beginning of 2019. This instilled dread in Tesla owners, and the majority of them began selling their cars. The used automobile market has seen a significant spike in demand for Tesla vehicles. However, the cost of a Tesla vehicle was heavily influenced by its stock price, which had an impact on the market for used Tesla cars. Tesla automobiles heavily rely on the post-sale services that

K.Pradeep² ²UG Scholar, Dept. Of IT, NRI Institute of Technology, A.P, India-521212

A.Devi Priya⁴ ⁴UG Scholar, Dept. Of IT, NRI Institute of Technology, A.P, India-521212

the firm offers. In order to keep the vehicles usable in the market and also decline of the stock prices the accurate prediction of used car prices was very important for the new and existing customers and the company itself. Since the commencement of the stock market, investors have sought to accurately predict stock prices.[2]

Few stocks have been as fascinating and contentious over the past two years as Tesla. By upending a crucial industry, the business hopes to profoundly alter the course of human history. Elon Musk became one of the most prominent CEOs in the world thanks to Tesla, which also made a lot of money for its investors. Market investors, however, are understandably pessimistic about how high the Tesla stock may rise. Tesla just released their Q3 2020 manufacturing and delivery figures. Despite having delivered 241,000 vehicles, the firm only produced 237,000. With these figures, it is very nearly producing 1 million units annually. While the international market leader Volkswagen group produces 9.3 million vehicles annually, domestic brands like Ford produce over 4 million.



II. TECHNOLOGY USED

Machine Learning:

Computers can learn intuitively thanks to the scientific field of machine learning. Machine learning has emerged as one of the most fascinating technologies ever. The moniker signifies that the computer becomes more human-like through its capacity to learn. There may be a lot more locations than one might assume where machine learning is currently being actively deployed.

Two of the most popular machine learning techniques are supervised learning and unsupervised learning, which train algorithms using example input and output data that is labelled by humans. Unsupervised learning involves giving the algorithm no labelled data in order to allow it to find structure within its input data, and supervised learning uses example input and output data that is unlabeled. Let's look into these strategies more closely.

Supervised Learning:

Examples of inputs and intended outputs are given to the model or algorithm, which then looks for patterns and relationships between the two. Learning a general rule that connects inputs and outputs is the objective. The training procedure is carried out repeatedly until the model's accuracy on the training set reaches the target level. Here are a few actual examples:

You practise picture classification using labels and images. Then you present a new image in the future with the expectation that the computer will identify the new object.

Market Prediction/Regression: You teach the computer using past market data and ask it to forecast a future price.

➤ Unsupervised Learning:

The learning system is not given any labels; instead, it is left to its own devices to identify structure in the data. It is employed to divide the population into various groups. Unsupervised learning may serve as a primary objective (discovering hidden patterns in data).

Clustering is a crucial step in scientific study where you ask the computer to group together comparable pieces of data into clusters.

High Dimension Visualization: We can see high dimension data with the aid of the computer.

Generated Models: A model can produce more data once it has captured the probability distribution of your input data.This can help to strengthen your classifier greatly.

Unsupervised learning may have the simple objective of finding hidden patterns in a dataset, but it may also have the objective of feature learning, which enables the computing device to automatically find the representations required to categorise the raw data.

> Approaches

Understanding and utilising machine learning algorithms requires a background in statistics because computational statistics and machine learning are closely related fields.

CSS separates the document content from document presentation. There are three types of CSS. They are Inline CSS, Internal CSS and external CSS.

Given that correlation and regression are frequently used methods for examining the relationship between quantitative variables, it may be helpful to define them first for those who have not studied statistics. A measure of association between two variables that are not categorised as dependent or independent is called correlation. At its most basic, regression is used to look at the relationship between two independent variables. Regression allows for prediction since statistics may be used to forecast the dependent variable when the independent variable is known.

Machine learning strategies are always being improved.For our purposes, we'll go through a couple of the well-liked machine learning techniques that are now in use.

Predictive Modelling

Predictive modelling, in summary, is a statistical technique that uses machine learning and data mining to predict and anticipate likely future outcomes using historical and existing data. The prediction model predicts the market as Positive or Negative based on a variety of inputs [1].It functions by analysing recent and old data and applying what it discovers to a model created to predict expected outcomes. Almost everything can be predicted with predictive modelling, including TV ratings, a customer's future purchase, credit risk, and company earnings

The majority of predictive models operate quickly and frequently finish their computations in real time. Because of this, businesses like banks and merchants may, for instance, forecast the risk associated with an online mortgage or credit card application and quickly approve or reject the request based on that assessment. Although the computation of some more complex predictive models, such as those used in computational biology and quantum computing, takes longer than that of a credit card application, it can now be completed much more quickly than in the past due to advancements in technology, including computing power.

III. ALGORITHM USED

Linear Regression:

The term "linear regression" refers to a machine learning algorithm based on supervised learning. Regression processing is carried out. To model a goal prediction value, regression uses independent variables. The main purpose of it is to establish the relationships between variables and forecasts. The amount of independent variables that a regression model uses and the sort of relationship that it considers between the dependent and independent variables are two factors that make a regression model different [Fig:2].

By employing an independent variable as a basis, linear regression is used to forecast the value of a dependent variable (y) (x). This regression method therefore discovers a linear relationship between x (the input) and y (the output) (output). Linear regression was thus first used.



IV. SOFTWARE REQUIREMENTS SPECIFICATION

Functional Requirements:

In the technical perspective, functional requirements address the functionality of the software. It improves and describes the component flow and structural flow of the same. The functional statement works with categorising and learning from the same dataset's raw datasets. Then the datasets are grouped into clusters, and any degradation of the groupings is examined for efficiency. After the dataset has been cleaned, the machine learns the data and identifies the pattern set for it. It then goes through a number of iterations and produces results.

Functional specifications specify what the software should be able to do (the functions). Consider the fundamental functions. Functional requirements should be written in the future tense because the "functions" are already established before development.In the technical perspective, functional requirements address the functionality of the software. It improves and describes the component flow and structural flow of the same. The functional statement works with categorising and learning from the same dataset's raw datasets. Then the datasets are grouped into clusters, and any degradation of the groupings is examined for efficiency. After the dataset has been cleaned, the machine learns the data and identifies the pattern set for it. It then goes through a number of iterations and produces results.

> Non Functional Requirements:

Nonfunctional requirements address external aspects that are inherently nonfunctional. It serves as a tool for analysis. The same applies as it carries out the operations' performance evaluation. Since stock is practical and constantly changing, these additional effects and specifications enable it to receive the most recent updates and integrate in a single step where the technicians may work on and address any bugs or draughts that may exist. The efficiency and hit gain ratio are the non-functional parameters that are adhered to. The code's applicability for increased efficacy, implementation, and searching for the security console. With the help of integration and portability, the System is dependable and the performance is maintained.

- Product Charactersistics:
- Usability:

This refers to how easy it is for any type of stock trader and other stock market participants to understand the user interface of stock prediction software.

• *Efficiency*:

Keeping the closing stock prices as accurate as possible while using the least amount of time and data possible.

• Performance:

Performance is a feature of stock prediction software that refers to how responsive it is to different user interactions.

V. EXISTING SYSTEM

Tesla Stock Price Prediction Using Sentiment Analysis: \geq A tremendously volatile element of the financial world is the stock market. A very active area of research and analysis involves making precise predictions about numerous stocks. This research attempts to improve the accuracy of the prediction outcomes by extending the earlier ML prediction strategies that used artificial neural networks and fuzzy-based techniques. A blend of algorithmic trading is at the core of the research since many qualitative elements influence the decision to purchase or sell a stock. This study intends to investigate the particular connection between Elon Musk's Tweets and the price of Tesla stock. Our dataset, which had been pre-processed to eliminate any stop words, was used as the basis for exploratory data analysis as the main analysis technique. Combining these approaches and components produced a clear correlation that led to a certain conclusion: more tweets and engagements were correlated with higher closing prices for Tesla, and vice versa.

Sentiment Analysis:

Sentiment analysis, commonly referred to as opinion combines modern data science, artificial mining, intelligence (AI), and the humanities and languages. The fundamental premise of sentiment analysis is to identify the emotions and subjective feelings hidden behind a set of linguistic data through sentiment analysis. Many methods exist for conducting sentiment analysis. For instance, by studying big text datasets, a machine may learn to recognise human emotions. We used sentiment analysis because we thought that tweets' conveyed sentiments would be related to Tesla stock prices.Almost all facets of our lives need sentiment analysis. Pathos is a component of all advertisements, campaigns, and commercials, and sentiment analysis is employed in a variety of ways in these mediums. The subjective state behind Elon Musk's tweets and replies was applied to Tesla's stock as part of our study's broad use of sentiment analysis, which revealed unexpected links between the two.

Based on the predicted emotional language in each tweet, Twitter categorised each post into three categories: good, neutral, and negative. Using Python and the Scikitlearn tool, this was accomplished. The pre-processed tweets from the text column of the CSV file were used for this machine learning.

Disadvantages of Existing System:

- Main drawback is that a lot of training data is needed to be effective and that they fail to encode the position and orientation of objects.
- High computational cost is required in order to predict the future stocks.
- Without GPU data it is quite slow to train the model
- Mostly it depends on previous information for prediction

VI. PROPOSED SYSTEM

In this research, prediction modelling and regression techniques are combined in an effort to forecast the direction of the stock market. The data sets acquired are used to feed this model the most recent stock prices. The collected data is divided up into different subsets or data sets that are used to train and test the algorithm. The data is then modelled using a regression model in R or Python. This model applies a thorough search algorithm to the data sets and, depending on the results, generates a summary table. Regression and clustering approaches are used to determine if the stock's price has increased or decreased once the numbers are plotted on a chart. This model extrapolates the current stock prices based on calculations to produce a prediction after a certain amount of time. Algorithms for supervised machine learning are used to create and train the models. The output will take the form of a graph and update as the dataset does. This model uses supervised machine learning methods and predicts up to 69% in-sample accuracy and 37% out-ofsample accuracy.

A. Advantages of Proposed System:

Implementing linear regression is straightforward, and it is simpler to understand the output coefficients. This algorithm is the best to apply when you know the relationship between the independent and dependent variable has a linear relationship because it is less difficult than other algorithms. Over-fitting can occur with linear regression, however it can be prevented by employing crossvalidation, regularisation (L1 and L2) techniques, and some dimensionality reduction approaches. When the dataset can be separated into linear components, linear regression works effectively. It can be used to determine the type of relationship between the variables.



Fig 3 Data Processing

As input to the models for training, attributes like price of open, high, low, close, and modified closing price are obtained from a large dataset. Data pre-processing techniques like normalisation and one hot encoding are then applied to the dataset. After that, the data is split into two sets, training and testing, with a ratio of 80:20 each. Then, a linear regression model is trained using this collection of data. Finally, root mean square error is used to analyse each of these modules.

> Collection:

The most important step in beginning with ML is to acquire accurate and high-quality data. Any verified source, such as the UCI dataset repository or data.gov.kaggle.in, can be used to acquire data. For instance, when studying for a competitive exam, students use the best study materials they have access to in order to learn the most effective content and produce the best outcomes. Similar to this, reliable and high-quality data will help the model learn faster and more effectively, and when tested, the model will produce cuttingedge findings. The process of gathering data uses up a tremendous amount of money, time, and resources.For instance, multiple photos of people in a range of human expressions are required when working on the Facial Expression Recognizer. Solid evidence guarantees that the outcomes of the model are valid and can be trusted upon.

> Preparation:

The information gathered can be in a raw form that cannot be immediately delivered to the machine. In this process, datasets are gathered from various sources, examined, and then a new dataset is created for further processing and research. This preparation can either be done manually or automatically. Data can also be prepared numerically, which would speed up the model's learning process. An image can be converted into a matrix of N x N dimensions, with each cell's value denoting a single picture pixel. [Fig:3]

> Input:

The data may not yet be machine-readable once it has been produced, thus certain conversion techniques are needed to turn the data into something that can be read by machines. To finish this exercise, a high level of computation and accuracy are necessary. Data can be acquired from a variety of sources, including Twitter comments, audio recordings, video clips, and MNIST Digit data (images).

> Processing:

At this point, processing the instructions given over a vast amount of data reliably and effectively requires the use of algorithms and machine learning techniques.

Output:

At this point, the machine generates conclusions that are clear to the user and meaningful. Reports, charts, movies, and other outputs are a few examples.

Storage:

This is the final step in which the obtained output and the data model data and all the useful information are saved for future use.[Fig:3].

C. Datasets:

Working with various dataset types is essential for success in the field of machine learning or to become a great data scientist. Finding a proper dataset for every type of machine learning project, however, is a challenging challenge. A dataset is a group of data that has been organised in some way. Any type of data, including series, arrays, and database tables, can be found in a dataset. A tabular dataset can be thought of as a database table or matrix where each row correlates to a dataset's fields and each column to a single variable. "Comma Separated File," or CSV, is the file type that is most frequently supported for tabular datasets. However, in order to store "tree-like data," we can use the JSON file more efficiently. Regression-based models are typically used to forecast continuous values using some provided independent data.[5]

➤ Kaggle Datasets:

One of the top places to find datasets for data scientists and machine learners is Kaggle. It makes it simple for users to locate, download, and publish datasets. Additionally, it offers the chance to collaborate with other machine learning professionals and complete challenging Data Science-related tasks.

We can quickly locate and download a high-quality dataset from Kaggle in a variety of formats.

D. Data Splitting:

The core of every ML problem is data. ML models are like bodies without a soul if they are not fed with the right data. But gathering data is no longer a significant issue in the 'big data' world of today. Every day, we consciously or unconsciously produce enormous datasets. Having an abundance of data available, however, does not make the issue go away. The quality data is useless unless it is used effectively, even though extracting meaning from raw data is an art in and of itself and necessitates solid feature engineering skills and domain expertise (in special circumstances) how to divide the data for training [Fig:4] and testing is the main issue that ML/DL practitioners encounter.Even though it initially appears to be a straightforward issue, only by delving deeply into it can its complexity be determined. Inaccurate training and testing sets may have unanticipated consequences on the model's output. It might cause the data to be overfitted or underfitted, and our model might end up producing biassed results.

➤ How should the Data be Divided?

The data should preferably be split into three sets: a train set, a test set, and a holdout cross-validation set (dev set). Let's first briefly discuss the meaning of these sets and the kinds of data they ought to contain.[Fig:4]

• Train Set:

The data that would be fed into the model would be in the train set. Simply put, our model would gain knowledge from this data.

• Dev Set:

The trained model is validated using the development set. As the foundation for our model evaluation, this setting is the most crucial.

• Test Set:

The data used to test the trained and approved model are in the test set. It reveals the effectiveness of our entire model and the likelihood that it will forecast an illogical event.



E. Training Dataset:

The training dataset, which is used to train or fit the machine learning model, is the largest (in terms of size) subset of the original dataset. In order for the ML algorithms to learn how to make predictions for the given task, training data is first fed into them.

Whether we are using supervised learning or unsupervised learning algorithms, the training data changes.

In unsupervised learning, inputs are not tagged with the appropriate outputs, hence the training data contains unlabeled data points. In order to produce predictions, models must extract patterns from the provided training datasets.

In contrast, labels are included in the training data for supervised learning in order to help the model be trained and predictions made.

The model's accuracy and predictive power are heavily influenced by the kind of training data we give it. It implies that the model will perform better the higher the quality of the training data. A typical ML project's training data makes up more than or equal to 60% of the total data.

> Test Dataset:

Once the model has been trained using the training dataset, it is time to test it using the test dataset. This dataset assesses the model's performance and guarantees that it can generalise well to new or unexplored datasets. The test dataset is a different subset of the original data from the training dataset. When the model training is finished, it utilises it as a benchmark because it has some similar features and a similar class probability distribution. A well-organized dataset called test data provides information for each type of scenario the model might encounter in the actual world. The test dataset typically represents 20 to 25 percent of the total original data for an ML project.

At this point, we may also examine and contrast the testing accuracy with the training accuracy, or, more specifically, the accuracy of our model when applied to the test dataset in comparison to the training dataset. The model is considered to have overfitted if its accuracy on training data is higher than its accuracy on testing data.

VII. TOOLS USED

> *Python* :

Python was the language we decided to use for our project. For many reasons, this was an easy call.

- There is a sizable community that supports Python as a language. A simple visit to Stack Overflow can fix any issues that may arise. The most often used language on the site, Python, provides the most straightforward answers to all queries.
- Python has several strong tools available for scientific computing packages. Packages like NumPy, Pandas, and SciPy are thoroughly documented and completely free. These packages will drastically reduce and vary the amount of code required to create a certain programme. This speeds up repetition.
- Python is a language that is tolerant and allows programmes that appear to be pseudocode.When the pseudocode provided in tutorial papers is required and checked, this can be useful. This step can occasionally

be completed quickly using Python. Python, though, is not without flaws. Python is a dynamically written language, and its modules are well-known for their duck writing. This could be annoying if a package approach returns something that, for example, resembles an array but isn't one. Additionally, the return type of a method was not explicitly stated in the standard Python documentation, which would have required extensive trial-and-error testing in a less capable language. This issue makes it more challenging than necessary to learn how to utilise a replacement Python package or library.

> Numpy:

Python's Numpy package offers higher-level scientific and mathematical abstractions wrapped in python. It serves as the foundational library for scientific computing and includes facilities for integrating C, powerful n-dimensional array objects, C++, etc. Additionally, it helps in linear algebra and random number generation.

The array type in Numpy adds an effective data structure for numerical computations, such as manipulating matrices, to the Python language. Additionally, Numpy offers fundamental numerical operations, such as tools for identifying Eigenvectors.

Scikit Learn:

A free machine learning library for Python might be Scikit-learn. Numerous techniques for classification, clustering, and regression are included, including random forests, k-neighbors, and support vector machines. It also supports SciPy and NumPy, two Python scientific and mathematical libraries. Scikit-learn is a programme that is specifically written in Python, with the main algorithms implemented in Cython for performance. A wrapper in Cython for LIBSVM enforces support vector machines. Specifically, logistic regression and linear support vector machines using a wrapper similar to LIBLINEAR.

> Matplotlib:

Matplotlib is an excellent Python visualisation package for 2D array displays. To handle the larger SciPy stack, a multi-platform data visualisation toolkit called Matplotlib was developed and is based on NumPy arrays. John Hunter introduced it for the first time in 2002.

One of visualization's main benefits is that it allows us visual access to enormous amounts of data in easily understandable forms. In Matplotlib, there are many different types of plots, such as line, bar, scatter, and histograms.

> Jupyter Notebook:

An open-source web programme called Jupyter Notebook allows users to create and share documents with equations, live code, and visualisations. Data can be used for a variety of purposes, such as data cleansing, data visualisation, data transformation, statistical modelling, machine learning, and many more.

VIII. SYSTEM ARCHITECTURE

In organisational theory and software engineering, a structure chart is a diagram that depicts the breakdown of a system to its most controllable components. The way a system is utilised, how it communicates with other systems, and how it interacts with the outside world are all reflected in its design. It specifies how one component of the system connects to the others and how data is transferred between them. Each module is symbolised by a box that carries its name. A system's architecture reveals how it is conceptualised in terms of its structure, operations, and connections. The term "system" in architecture typically refers to the software's architecture rather than the actual physical structure of the machines or structures. A system's architecture reflects how it is utilised, and as a result, it evolves through time.



Fig 5 System Architecture

IX. CONCLUSION

It appears that Tesla's stock prices will decline soon if they don't come up with a new strategy for communicating their goal. Given that other companies have begun making electric cars at significantly lower prices than Tesla, this may be viable.

A successful future price prediction for a stock will maximise profits for investors.

REFERENCES

- M. Usmani, S. H. Adil, K. Raza and S. S. A. Ali, "Stock market prediction using machine learning techniques," 2016 3rd International Conference on Computer and Information Sciences (ICCOINS), Kuala Lumpur, 2016, pp. 322-327.
- [2]. 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.
- [3]. H. Gunduz, Z. Cataltepe and Y. Yaslan, "Stock market direction prediction using deep neural networks," 2017 25th Signal Processing and Communications Applications Conference (SIU), Antalya, 2017, pp. 1-4.
- [4]. M. Billah, S. Waheed and A. Hanifa, "Stock market prediction using an improved training algorithm of neural network," 2016 2nd International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE), Rajshahi, 2016, pp. 1-4.
- [5]. H. L. Siew and M. J. Nordin, "Regression techniques for the prediction of stock price trend," 2012 International Conference on Statistics in Science, Business and Engineering (ICSSBE), Langkawi, 2012, pp. 1-5.
- [6]. K. V. Sujatha and S. M. Sundaram, "Stock index prediction using regression and neural network models under non normal conditions," INTERACT-2010, Chennai, 2010, pp. 59-63.
- [7]. S. Liu, G. Liao and Y. Ding, "Stock transaction prediction modeling and analysis based on LSTM," 2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA), Wuhan, 2018, pp. 2787-2790.
- [8]. T. Gao, Y. Chai and Y. Liu, "Applying long short term momory neural networks for predicting stock closing price," 2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS), Beijing, 2017, pp. 575-578.
- [9]. K. A. Althelaya, E. M. El-Alfy and S. Mohammed, "Evaluation of bidirectional LSTM for short-and longterm stock market prediction," 2018 9th International Conference on Information and Communication Systems (ICICS), Irbid, 2018, pp. 151-156

BIOGRAPHIES



Dr. Chaitanya Kishore Reddy. M is currently working as a Professor and Dean in the Department of Information technology NRI Institute Of Technology. at Pothavarappadu, Agiripalli, Krishna(dist), India. He received Ph.D. in Computer Science and Engineering and M. Tech in Computer Science and Engineering at Jawaharlal Nehru Technological university, Kakinada. He has Published 40 research papers in various National and International Journals and International Conferences. He is a member in ISTE, CSI, and IAENG. His research areas are Mobile Ad-hoc Networks, IoT, and Cloud Computing.



K. Pradeep is currently studying B.Tech with specification of Information Technology in NRI Institute of Technology. He had done a mini project on Tesla Stock Price Prediction.He had completed 2 NPTEL certification courses.



V.Luke Sundar is currently studying B.Tech with specification of Information Technology in NRI Institute of Technology. He had done a mini project on Tesla Stock Price Prediction. He had completed 2 NPTEL certification courses.



A.Devi Priya is currently studying B.Tech with specification of Information Technology in NRI Institute of Technology. She had done a mini project on Tesla Stock Price Prediction. She had completed 2 NPTEL certification courses.