SRM Institute of Science and Technology
College of Engineering & Technology

18CSC305J Artificial Intelligence – Mini Project

Tesla Stock Price Prediction with Machine Learning (TSLA)

Team Members

RA2111030010062 PRITISH V

RA2111030010058 YUVAN DAYAKAR

SRM
INSTITUTE OF SCIENCE & TECHNOLOGY
(Deemed to be University u/s 3 of UGC Act, 1956)

Problem Statement

1. Every day, the stock market is covered in the news. Every time it achieves a new high or low, you hear about it. If a reliable algorithm could be developed to forecast the short-term price of a specific stock, the rate of investment and business opportunities on the stock market may rise. The PCC (Pearson Correlation Coefficient) and BLS (Board Learning System) have been used in the past to predict stock prices.
2. The goal of this research is to determine whether it is possible to develop a model that will accurately and efficiently predict stock price by providing future predictions based on data analysis using regression models.

Figure 1. Stock Market Background

SRM
INSTITUTE OF SCIENCE & TECHNOLOGY
(Deemed to be University u/s 3 of UGC Act, 1956)

ABSTRACT

(TSLA), one of the leading companies in the electric vehicle and renewable energy sectors. The aim is to develop a predictive model that accurately forecasts the closing price of Tesla stock, aiding investors in making informed trading decisions. We begin by collecting historical stock price data for Tesla spanning several years, along with relevant financial indicators and market data. Feature engineering techniques are employed to extract meaningful features, including past stock prices, trading volume, market indices, economic indicators, and sentiment analysis from news sources and social media. Several machine learning algorithms are explored for time series forecasting, including linear regression, decision trees, random forests, and deep learning models such as recurrent neural networks (RNNs) and long short-term memory networks (LSTMs). The performance of each model is evaluated using metrics such as mean squared error (MSE) and mean absolute error (MAE) on a validation dataset. Hyperparameter tuning is conducted to optimize the chosen models, and robustness is ensured through techniques like cross-validation and time-based splitting. The trained models are then deployed to a production environment for real-time prediction, with monitoring systems in place to track performance and facilitate periodic retraining. Experimental results demonstrate the efficacy of the proposed machine learning approach in predicting Tesla's stock price, with the best-performing models achieving high accuracy and outperforming baseline methods. Insights gained from the models provide valuable information on the factors influencing Tesla's stock price and contribute to a better understanding of market dynamics.

SRM
INSTITUTE OF SCIENCE & TECHNOLOGY
(Deemed to be University u/s 3 of UGC Act, 1956)

INTRODUCTION

The financial markets, with their inherent complexity and volatility, have long fascinated investors and researchers alike. Among the myriad challenges facing investors, predicting stock prices stands out as a particularly daunting task due to the multitude of factors influencing market dynamics. Despite this challenge, accurate stock price prediction holds immense value for investors, enabling them to make informed trading decisions and optimize their investment strategies.

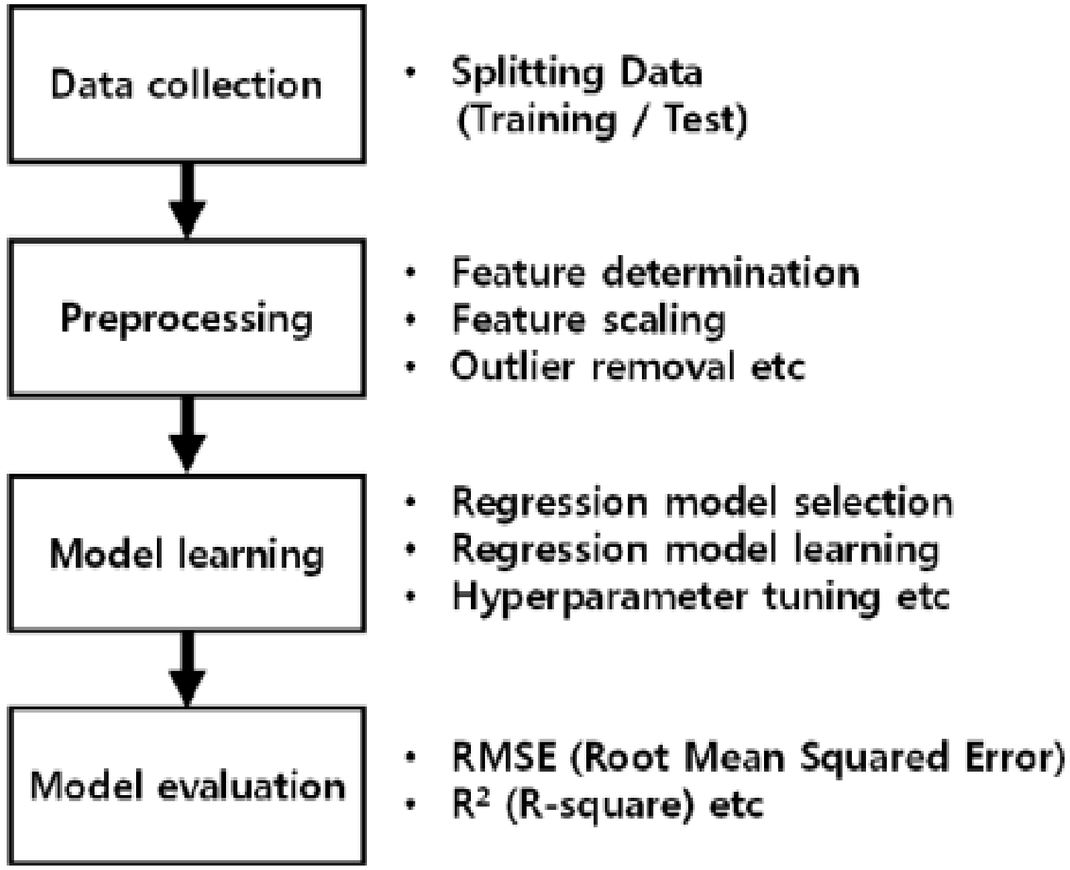
In recent years, machine learning techniques have emerged as powerful tools for addressing the stock price prediction problem. By leveraging historical market data and relevant features, machine learning models can learn complex patterns and relationships, thereby offering valuable insights into future price movements. In this context, we turn our attention to Tesla, Inc. (TSLA), a pioneering company in the electric vehicle and renewable energy sectors, known for its innovative products and disruptive impact on the automotive industry. The objective of this study is to develop a machine learning-based approach for predicting the future stock price of Tesla. By harnessing historical stock price data, along with a rich set of features encompassing financial indicators, market data, and sentiment analysis, we aim to build a predictive model that accurately forecasts the closing price of Tesla stock over a given time horizon. Such a model holds significant implications for investors, providing them with valuable insights into Tesla's stock performance and facilitating more informed investment decisions. In this introductory section, we provide an overview of the motivation behind this study, highlighting the importance of stock price prediction in the financial markets and the potential benefits it offers to investors. We also outlinehe structure of the paper, delineating the key components of our approach, including data collection, feature engineering, model selection, training, and evaluation. Finally, we set the stage for the subsequent sections by presenting the research objectives and outlining the contributions of this study to the field of machine learning in finance.

METHODOLOGY

Algorithms for machine learning (ML) learn from data. You must be sure to give them the pertinent information for the issue you're trying to resolve. Even if you have decent data, you still need to make sure that it is scaled, formatted, and even contains relevant landscapes. There is data everywhere. Terabytes of new data being created every second. Useful data for machine learning can take many different forms. Effective model training can be achieved with structured data. Extracting samples In this dataset we have total of 4263 rows of daily composed data of TESLA stock. Here we collected open price, closing price, high price of the day, volume and date. Tasks The main point of a task is to decide what to do with all the data we accrued. When different types of data are thrown at you, what do you do with them or how do you go about it? Some categories of tasks include classification and regression (under supervised learning) and clustering and generation (under unsupervised learning). Models Finding a mathematical relationship that links the input data to the outputs is what we now need to do. And this function should accomplish this correctly for all situations, not just those in which we already know the outcome. In other words, once you train the function and identify its precise restrictions, it should perform flawlessly even if it had never encountered test data during any of its training cases. Regression Model: Using the statistical method known as regression, an effort is made to control the form and strength of the relationship between one dependent variable (commonly denoted by Y) and a number of other factors (referred to collectively as independent variables). When it comes to valuing assets and understanding the relationships between a variety of elements, such as the price of a commodity and the shares of a firm that deals in that commodity, regression is a useful tool for investment and financial managers. Regression helps these managers understand the connections between numerous aspects. Methodology Summarization:

SRM
INSTITUTE OF SCIENCE & TECHNOLOGY
(Deemed to be University u/s 3 of UGC Act, 1956)

DATA FLOW



SRM
INSTITUTE OF SCIENCE & TECHNOLOGY
(Deemed to be University u/s 3 of UGC Act, 1956)

TYPES OF ML MODELS

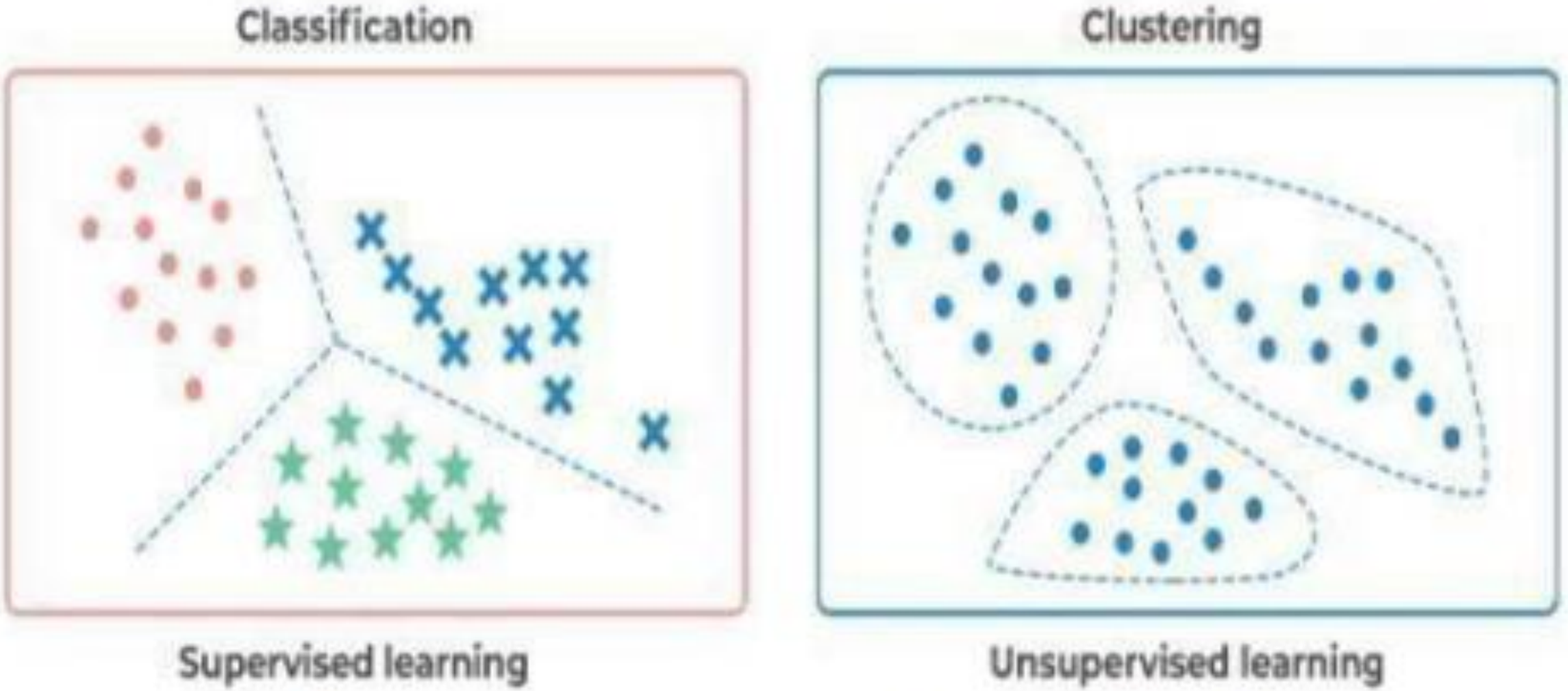


Figure 2.a) Supervised learningb) Unsupervised learning

REGRESSION MODEL

1. Regression analysis is a statistical approach that simulates the connection between one or more independent variables and one or more dependent variables using one or more dependent variables. Regression analysis is useful for gaining an understanding of how the value of a dependent variable shifts in response to changes in the value of an independent variable, particularly in situations in which other independent variables are maintained constant. It makes predictions based on actual, ongoing data such as temperature, age, pay, and cost, amongst others.
2. Regression Model The method of regression, which is part of supervised learning, enables us to make predictions about the continuous output variable based on one or more predictor variables and provides assistance in determining the relationship between the variables.
3. The majority of the time, it is used for making predictions, modeling time series, making forecasts, and figuring out the link between different variables. It is used to figure out patterns in the market. It is used to detect trends in the data. It is useful for forecasting real or continuous data. The method of regression, which is part of supervised learning, enables us to make predictions about the continuous output variable based on one or more predictor variables and provides assistance in determining the relationship between the variables.

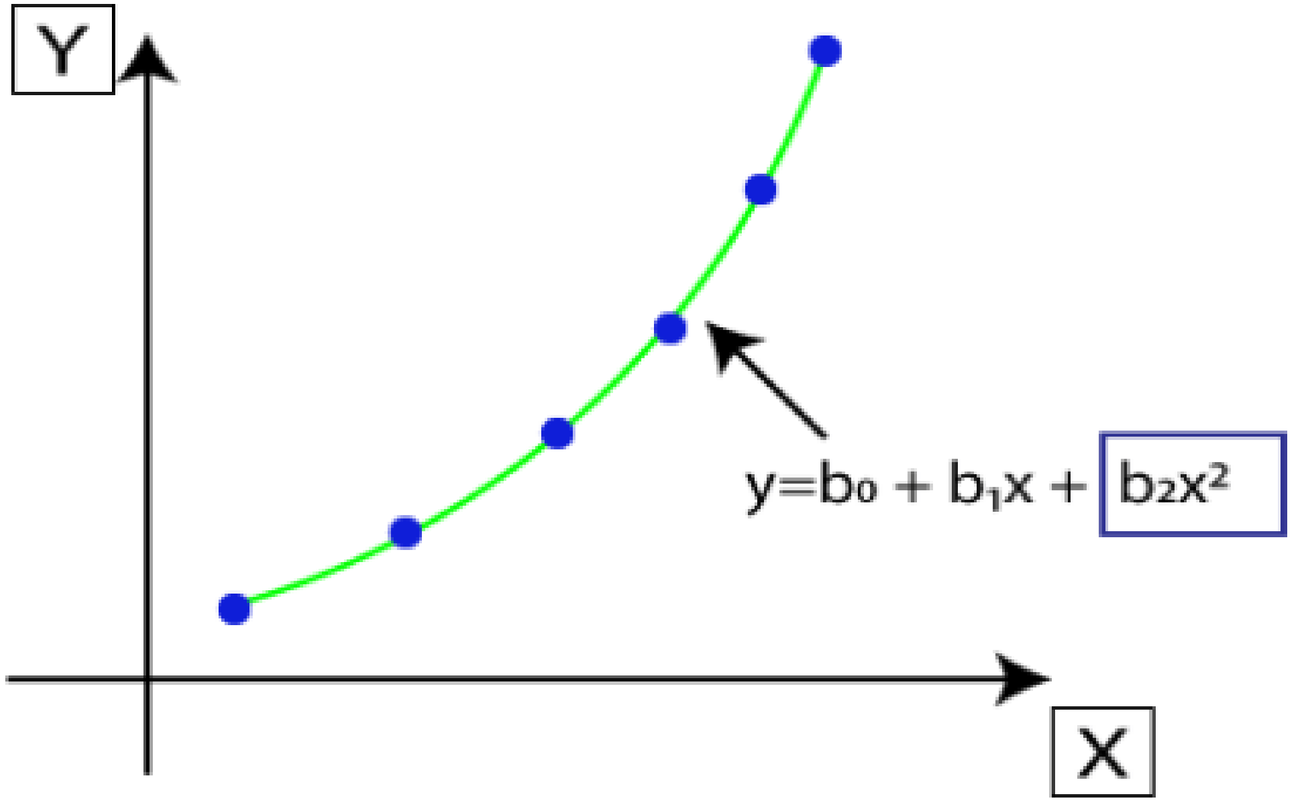


Figure 3. Regression Model

PROPOSED SYSTEM

A generalized dataset would be formed by combining datasets coming from a variety of different sources. In this part of the report, the data will be loaded, the cleanliness of the dataset will be checked, and then the dataset will be trimmed and cleaned so that it may be analyzed. The data set that was gathered for the purpose of making predictions based on supplied data was divided into two parts:

the Training set and the Test set. In most situations, ratios of 7:3 are used for dividing the Training set and the Test set. The data model, which was created with the help of machine learning algorithms, is applied to the Training set, and then, based on the accuracy of the test results, a prediction is made about the Test set.

The ML prediction model is successful in predicting the Tesla stock issue because it is good in preprocessing outliers and unimportant factors, and it also contains a combination of continuous, categorical, and discrete variables.

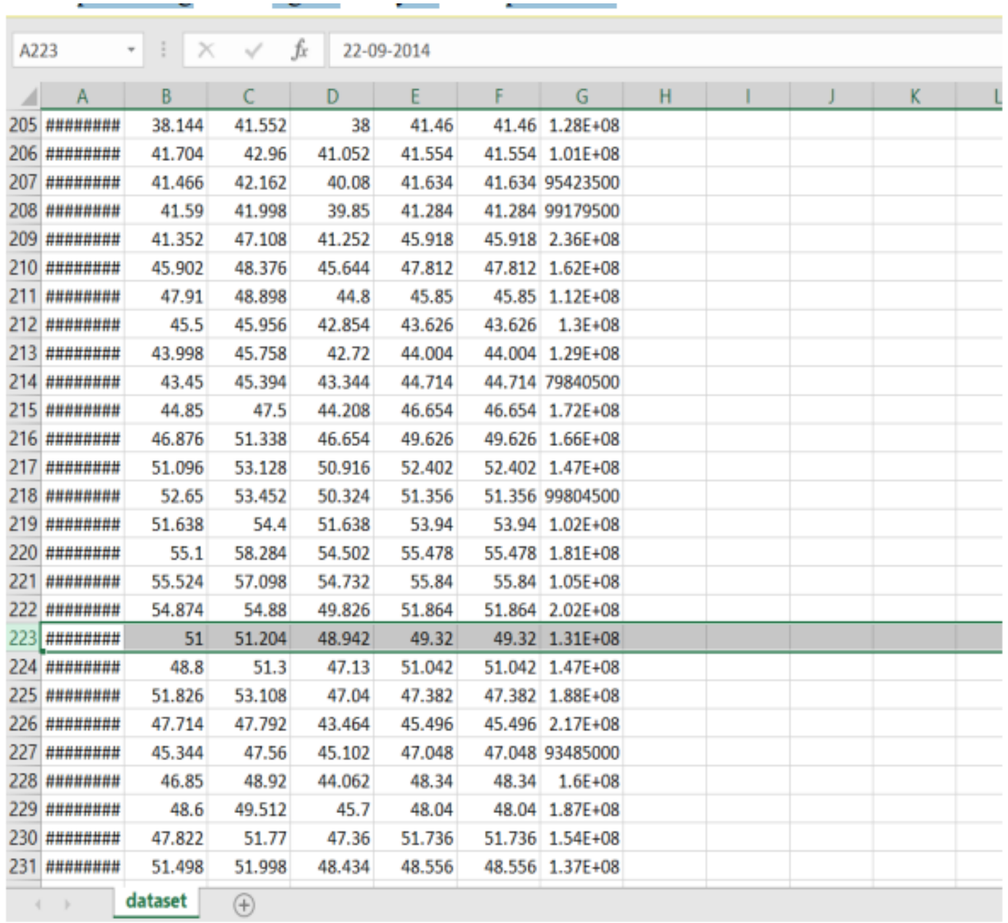
LEARNING ALGORITHM

Assume we have some data, we have a exact task that we want to accomplish, we have a proposed model that will hopefully predict the values of the output for every input and undertake the model is given by the function If we have certain values for the parameters a,b and c, then we can also discover a loss function relating how off the forecast function is from the actual values. One way to do it is to assume 1. 2. # change y from {0,1} to {-1,1} y [where (y == 0)] = -1 22 a range for a, b and c, like -20 to +20 and plug in values for all the different groupings and find out by brute force, for which values of a, b and c is the value of the loss function reduced. And to do this resourcefully, we make use of a couple of algorithms.

SRM INSTITUTE OF SCIENCE & TECHNOLOGY
(Deemed to be University u/s 3 of UGC Act, 1956)

RESULTS

The Ridge regressor has a higher level of accuracy in forecasting future stock prices for Tesla. Below are more facts pertaining to the figure that you have presented



A223 | 22-09-2014

	A	B	C	D	E	F	G
205	########	38.144	41.552	38	41.46	41.46	1.28E+08
206	########	41.704	42.96	41.052	41.554	41.554	1.01E+08
207	########	41.466	42.162	40.08	41.634	41.634	95423500
208	########	41.59	41.998	39.85	41.284	41.284	99179500
209	########	41.352	47.108	41.252	45.918	45.918	2.36E+08
210	########	45.902	48.376	45.644	47.812	47.812	1.62E+08
211	########	47.91	48.898	44.8	45.85	45.85	1.12E+08
212	########	45.5	45.956	42.854	43.626	43.626	1.3E+08
213	########	43.998	45.758	42.72	44.004	44.004	1.29E+08
214	########	43.45	45.394	43.344	44.714	44.714	79840500
215	########	44.85	47.5	44.208	46.654	46.654	1.72E+08
216	########	46.876	51.338	46.654	49.626	49.626	1.66E+08
217	########	51.096	53.128	50.916	52.402	52.402	1.47E+08
218	########	52.65	53.452	50.324	51.356	51.356	99804500
219	########	51.638	54.4	51.638	53.94	53.94	1.02E+08
220	########	55.1	58.284	54.502	55.478	55.478	1.81E+08
221	########	55.524	57.098	54.732	55.84	55.84	1.05E+08
222	########	54.874	54.88	49.826	51.864	51.864	2.02E+08
223	########	51	51.204	48.942	49.32	49.32	1.31E+08
224	########	48.8	51.3	47.13	51.042	51.042	1.47E+08
225	########	51.826	53.108	47.04	47.382	47.382	1.88E+08
226	########	47.714	47.792	43.464	45.496	45.496	2.17E+08
227	########	45.344	47.56	45.102	47.048	47.048	93485000
228	########	46.85	48.92	44.062	48.34	48.34	1.6E+08
229	########	48.6	49.512	45.7	48.04	48.04	1.87E+08
230	########	47.822	51.77	47.36	51.736	51.736	1.54E+08
231	########	51.498	51.998	48.434	48.556	48.556	1.37E+08

dataset

SRM
INSTITUTE OF SCIENCE & TECHNOLOGY
(Deemed to be University u/s 3 of UGC Act, 1956)

OUTPUT

Figure 3: Sample Output

REFRENCES

Multimed Tools Appl (2023).

https://doi.org/10.1007/s11042- 023- 15926-5

Venkateswara Reddy, B., Vellela, S. S., Sk, K. B., Roja, D., Yakubreddy, K., &Rao, M. V. Conceptual Hierarchies for Efficient Query Results Navigation. International Journal of All Research Education and Scientific Methods (IJARESM), ISSN, 2455-6211.

http://dx.doi.org/10.14569/IJACSA.2023.01406128

THANK YOU