Bangalore Institute of Technology

M.Tech, Department of Computer Science and Engineering
K R Road, V V Pura, Bengaluru-560004

Mini Project Synopsis on

Tesla Stock Price Prediction

Submitted as Mini Project for the M.Tech Lab Component of IPCC subject

Artificial Intelligence and Machine Learning (22SCS22)

Submitted by
Spoorthy UK 1BI22SCS06

For academic year 2022-23

Under the guidance of
Dr. M S Bhargavi
Associate Professor
Dept. of CSE

Introduction to the problem:

The problem at hand is predicting the price movements of Tesla (TSLA) stock. Stock price prediction is a common challenge in the financial domain, and accurate predictions can have significant implications for investors and traders. The goal of this project is to develop a machine learning model that can forecast whether the price of Tesla stock will increase or decrease in the future based on historical data. By leveraging historical price patterns and other relevant features, the model aims to capture underlying trends and patterns that may influence future price movements. To address this problem, we will use various machine learning algorithms such as Logistic Regression, Support Vector Classifier (SVC), and XGBoost Classifier. These algorithms will be trained on historical data, including features such as opening price, closing price, high and low prices, trading volume, and other engineered features derived from the original data. The dataset used for training and evaluation contains historical Tesla stock data, including the target variable indicating the price movement (increase or decrease) on the following day. The data will be preprocessed, including handling missing values, feature scaling, and creating additional features to enhance the predictive power of the models.

Problem Statement:

The primary focus of this project is to apply machine learning algorithms to learn from the historical patterns and trends in the Tesla stock data. The trained model will then be used to make predictions on unseen data, providing valuable insights for investors and traders. To accomplish this, we will use a dataset that includes historical information about Tesla stock, such as opening price, closing price, high and low prices, and trading volume. Additionally, we will engineer additional features from the available data to potentially improve the predictive power of the model.

Objective:

The objective of this project is to develop a machine learning model that can accurately predict the future price movements of Tesla (TSLA) stock. The model will analyze historical stock data and utilize various features to forecast whether the stock price will increase or decrease in the next time period.

Dataset description with Snapshot of dataset

The dataset used in this project contains historical data of Tesla (TSLA) stock. It includes various features related to the stock's performance over time. Here is an updated description of the columns present in the dataset:

1.Date: The date corresponding to the stock price data point.
2.Open: The opening price of Tesla stock on a given day.
3.High: The highest price reached by Tesla stock during the day.
4.Low: The lowest price reached by Tesla stock during the day.
5.Close: The closing price of Tesla stock on a given day.
6.Adj Close: The adjusted closing price of Tesla stock on a given day. It accounts for factors such as dividends, stock splits, and other corporate actions that impact the stock price.
7.Volume: The trading volume, representing the number of shares traded on that day.

	A	B	C	D	E	F	G
1	Date	Open	High	Low	Close	Adj Close	Volume
2	29-06-2010	19	25	17.540001	23.889999	23.889999	18766300
3	30-06-2010	25.790001	30.42	23.299999	23.83	23.83	17187100
4	01-07-2010	25	25.92	20.27	21.959999	21.959999	8218800
5	02-07-2010	23	23.1	18.709999	19.200001	19.200001	5139800
6	06-07-2010	20	20	15.83	16.110001	16.110001	6866900
7	07-07-2010	16.4	16.629999	14.98	15.8	15.8	6921700
8	08-07-2010	16.139999	17.52	15.57	17.459999	17.459999	7711400
9	09-07-2010	17.58	17.9	16.549999	17.4	17.4	4050600
10	12-07-2010	17.950001	18.07	17	17.049999	17.049999	2202500
11	13-07-2010	17.389999	18.639999	16.9	18.139999	18.139999	2680100
12	14-07-2010	17.940001	20.15	17.76	19.84	19.84	4195200
13	15-07-2010	19.940001	21.5	19	19.889999	19.889999	3739800
14	16-07-2010	20.700001	21.299999	20.049999	20.639999	20.639999	2621300
15	19-07-2010	21.370001	22.25	20.92	21.91	21.91	2486500
16	20-07-2010	21.85	21.85	20.049999	20.299999	20.299999	1825300
17	21-07-2010	20.66	20.9	19.5	20.219999	20.219999	1252500

Fig 1: Snapshot

ML techniques that could be applied in predicting Tesla Stock price are:

Several techniques can be applied to tackle the problem of stock price prediction for Tesla (TSLA) stock. Here are some commonly used techniques:

1. **Time Series Analysis**: Since stock price data is typically sequential and exhibits temporal dependencies, time series analysis techniques can be employed. These include methods such as autoregressive integrated moving average (ARIMA), seasonal decomposition of time series (STL), and exponential smoothing models.

2. **Machine Learning Algorithms**: Various machine learning algorithms can be applied to stock price prediction. These include regression algorithms like Linear Regression and Ridge Regression, ensemble methods like Random Forest and Gradient Boosting, and neural network models like Long Short-Term Memory (LSTM) or Convolutional Neural Networks (CNNs).

3. **Feature Engineering**: Creating informative and relevant features from the available data can improve prediction accuracy. Feature engineering techniques involve transforming or combining existing features to extract meaningful information. For example, creating lagged variables (using past price or volume data), technical indicators (moving averages, relative strength index), or sentiment analysis of news and social media data.

4. **Technical Analysis Indicators**: Technical analysis techniques can be used to extract patterns and trends from historical price data. These include indicators such as moving averages, Bollinger Bands, relative strength index (RSI), and MACD (Moving Average Convergence Divergence). Incorporating these indicators as features can enhance the predictive power of the models.

Tools to be used

1. **Python**: The code is written in the Python programming language, which is widely used for data analysis and machine learning tasks.
2. **Numpy**: Numpy is a fundamental library for scientific computing in Python. It provides support for efficient numerical operations and arrays.
3. **Pandas**: Pandas is a powerful data manipulation and analysis library. It provides data structures like DataFrames, which allow for easy handling and manipulation of structured data.
4. **Matplotlib**: Matplotlib is a popular plotting library in Python. It is used for creating various types of visualizations, including line plots and bar charts.
5. **Seaborn**: Seaborn is a high-level data visualization library built on top of Matplotlib. It provides additional functionality and aesthetic enhancements for creating attractive statistical visualizations.
6. **Scikit-learn**: Scikit-learn is a comprehensive machine learning library in Python. It includes various algorithms and tools for tasks such as data preprocessing, model selection, and evaluation.
7. **XGBoost**: XGBoost is an optimized gradient boosting library used for machine learning tasks. It provides efficient implementations of gradient boosting algorithms, which are popular for handling structured data.

IDE Used: Visual Studio Code

Methodology:

1.**Data preprocessing**: This involves handling missing values, dropping irrelevant columns, and normalizing or scaling the data as necessary.

2.**Feature engineering**: Creating additional features based on domain knowledge or patterns observed in the data that could potentially improve the model's performance.

3.**Splitting the data**: Dividing the dataset into training and testing sets, ensuring that the model's performance is evaluated on unseen data.

4.**Model selection and training**: Experimenting with different machine learning algorithms such as Logistic Regression, Support Vector Classifier (SVC), and XGBoost Classifier. Training the models using the training data and optimizing their hyperparameters if necessary.

5.**Model evaluation**: Assessing the performance of the trained models using appropriate evaluation metrics such as ROC AUC score. Comparing the results to determine the best-performing model.

6.**Prediction and analysis**: Utilizing the selected model to make predictions on new, unseen data. Analyzing the results and evaluating the model's reliability and usefulness in predicting Tesla stock price movements

Final outcome expected:

The final outcome expected from the stock price prediction project for Tesla (TSLA) stock is a reliable and accurate model that can predict the future price movements of the stock. The model should be able to take in historical data as input and provide predictions on whether the stock price will increase or decrease in the next time period. Expected outcomes: Model Accuracy, Robust Generalization, Real-Time Predictions, Insights and Interpretability, Continual Improvement