**EXPENSE TRACKER**

***Submitted by***

**ANUSHA PATRA[RA2111027010022]**

**ABHIGNYA PRIYADARSHINI[RA2111027010067]**

***Under the Guidance of***

**Dr. E SASIKALA**

***In partial satisfaction of the requirements for the degree of***

## **BACHELOR OF TECHNOLOGY**

**in**

**COMPUTER SCIENCE ENGINEERING**

**with specialization in Big Data Analytics**

## Logo, company name Description automatically generated

**SCHOOL OF COMPUTING**

# **COLLEGE OF ENGINEERING AND TECHNOLOGY**

**SRM INSTITUTE OF SCIENCE AND TECHNOLOGY**

**KATTANKULATHUR - 603203**

**OCTOBER 2023**

|  |  |
| --- | --- |
| Logo, company name  Description automatically generated | COLLEGE OF ENGINEERING & TECHNOLOGY  SRM INSTITUTE OF SCIENCE & TECHNOLOGY  S.R.M. NAGAR, KATTANKULATHUE – 603 203  Chengalpattu District |

**BONAFIDE CERTIFICATE**

Register No. **RA2111027010022** Certified to be the bonafide work done by **ANUSHA PATRA** of II Year/IV Sem B.Tech Degree Course in the **Practical Machine Learning 18CSE392T** in **SRM INSTITUTE OF SCIENCE AND TECHNOLOGY,** Kattankulathur during the academic year 2023 – 2024.

**LAB INCHARGE Head of the Department**

**E Sasikala**

**Associate Professor**

**Department of DSBS**

**SRMIST – KTR.**

|  |
| --- |

**TABLE OF CONTENTS**

| **CHAPTER NO** | **TITLE** | **PAGE NO** |
| --- | --- | --- |
|  |  |  |
| **1** | **ABSTRACT** |  |
| **2** | **PROBLEM STATEMENT** |  |
| **3** | **IDENTIFYING REQUIREMENTS** |  |
| **4** | **ALGORITHM** |  |
| **6** | **PERFORMANCE MATRIX** |  |
| **7**  **8**  **9**  **10**  **11** | **CODE**  **OUTPUT**  **FUTURE ENHANCEMENTS**  **CONCLUSION**  **REFERENCE** |  |

**ABSTRACT**

In this machine learning project, we investigate stock market prediction focusing on the stock prices of Google and Tesla, two prominent technology companies. Through the utilization of historical stock price data, financial indicators, and sentiment analysis from news and social media sources, our primary goal is to develop and evaluate predictive models that can offer insights into the future trends of Google and Tesla stocks.

The project encompasses comprehensive data collection, including historical stock prices and key financial metrics, along with the analysis of market sentiment from news articles and social media content. Features are thoughtfully engineered from this data, encompassing technical indicators, fundamental metrics, and sentiment scores. Multiple machine learning models, including regression and time series forecasting, are evaluated for their accuracy in predicting stock prices. Evaluation metrics like R2 and MSE are used to gauge model performance, and trading simulations are conducted to assess the practical profitability of the predictions.

Ultimately, this project seeks to reveal the potential of machine learning in stock market prediction and offer insights into the specific dynamics of Google and Tesla stock prices.

**PROBLEM STATEMENT**

Stock market prediction has long been a challenging and financially significant endeavor, with potential applications spanning from investment decisions to risk management. In this context, we address the problem of accurately forecasting the stock prices of two major technology companies, Google and Tesla, using machine learning techniques.

This problem statement entails the following key challenges:

**Data Complexity:** Historical stock price data, financial indicators, and sentiment analysis from diverse sources introduce complexity. The challenge is to handle, preprocess, and extract meaningful information from this diverse data while dealing with inconsistencies and outliers.

**Model Selection:** Given the multitude of available machine learning algorithms, the problem is to determine the most suitable models for predicting Google and Tesla stock prices accurately. These models should account for the unique dynamics of each company's stock.

**Evaluation and Real-World Applicability:** Developing predictive models is not sufficient; the challenge lies in evaluating their performance rigorously using appropriate metrics. Additionally, it is crucial to assess the practical applicability of these models in real-world investment scenarios.

**Market Dynamics and Sentiment Analysis:** Capturing the dynamic and ever-changing nature of stock markets, as well as the influence of market sentiment from news and social media, is a substantial challenge. Developing models that adapt to these dynamics is essential.

This project aims to address these challenges, providing valuable insights into the potential of machine learning in stock market prediction, ultimately assisting stakeholders in making more informed financial decisions regarding Google and Tesla stocks.

**IDENTIFYING REQUIREMENTS**

Tosuccessfully implement a machine learning project for stock market prediction, particularly for Google and Tesla, a set of clear requirements must be established. These requirements encompass data, technology, and project management aspects:

**Data Requirements:**

**a. Historical Stock Price Data:** Access to historical daily, weekly, or intra-day stock price data for Google and Tesla, covering a significant time period.

**b. Financial Indicators:** Comprehensive financial metrics, including trading volume, moving averages, price-to-earnings ratios, and dividend yields.

**c. News and Social Media Data:** A substantial corpus of news articles and social media content, with relevant timestamps, to conduct sentiment analysis.

**d. Feature Engineering Data:** Access to additional datasets for feature engineering, such as economic indicators, market indices, and other related financial data.

**Data Preprocessing Requirements:**

**a. Data Cleaning:** Robust data cleaning processes to handle missing values, outliers, and data inconsistencies.

**b. Feature Engineering:** Implementation of feature engineering techniques to create relevant features for the machine learning models.

**Machine Learning Models:**

**a. Algorithm Selection:** Evaluation and selection of machine learning algorithms, including regression models, time series forecasting methods, and deep learning approaches.

**b. Model Training:** Development and training of the chosen models using the historical data.

**c. Hyperparameter Tuning:** Fine-tuning model hyperparameters to optimize their performance.

**Evaluation Metrics:**

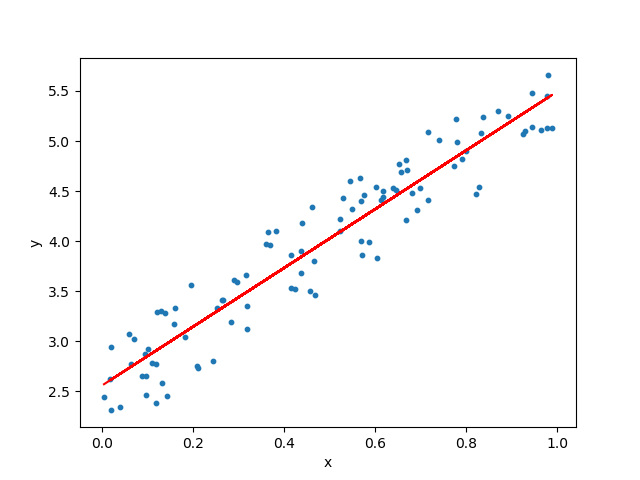
**a. Metric Selection:** Define appropriate evaluation metrics, such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and others, to assess model performance.

**b. Cross-Validation:** Implement cross-validation techniques to ensure robust model evaluation.

**ALGORITHM USED**

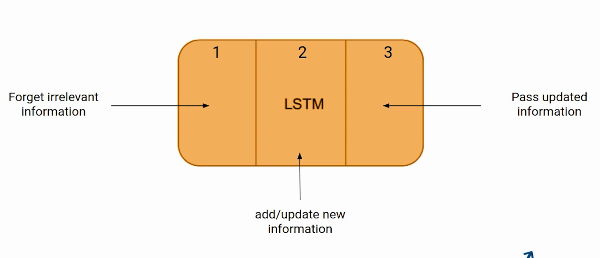
**Linear Regression**

Linear regression is a type of supervised machine learning algorithm that computes the linear relationship between a dependent variable and one or more independent features. When the number of the independent feature, is 1 then it is known as Univariate Linear regression, and in the case of more than one feature, it is known as multivariate linear regression. The goal of the algorithm is to find the best linear equation that can predict the value of the dependent variable based on the independent variables. The equation provides a straight line that represents the relationship between the dependent and independent variables. The slope of the line indicates how much the dependent variable changes for a unit change in the independent variable



LSTM: LSTM (Long Short-Term Memory) is a recurrent neural network (RNN) architecture widely used in Deep Learning. It excels at capturing long-term dependencies, making it ideal for sequence prediction tasks.

Unlike traditional neural networks, LSTM incorporates feedback connections, allowing it to process entire sequences of data, not just individual data points. This makes it highly effective in understanding and predicting patterns in sequential data like time series, text, and speech.



**PERFORMANCE METRICS USED**

**Mean Square Error:**

In statistics, the mean squared error (MSE) is defined as the mean or average of the squared differences between the actual and estimated values. Mean Squared Error (MSE) measures the amount of error in a statistical model. Evaluate the mean squared difference between observed and predicted values. If the model has no errors, the MSE is zero. Its value increases as the model error increases. The mean squared error is also known as the mean squared deviation (MSD). For example, in regression, the mean squared error represents the mean squared residual.

**R2 Error:**

R squared (R2) is a regression error metric that justifies the performance of the model. It represents the value of how much the independent variables are able to describe the value for the response/target variable.

Thus, an R-squared model describes how well the target variable is explained by the combination of the independent variables as a single unit.

The R squared value ranges between 0 to 1 and is represented by the below formula:

**R2= 1- SSres / SStot**

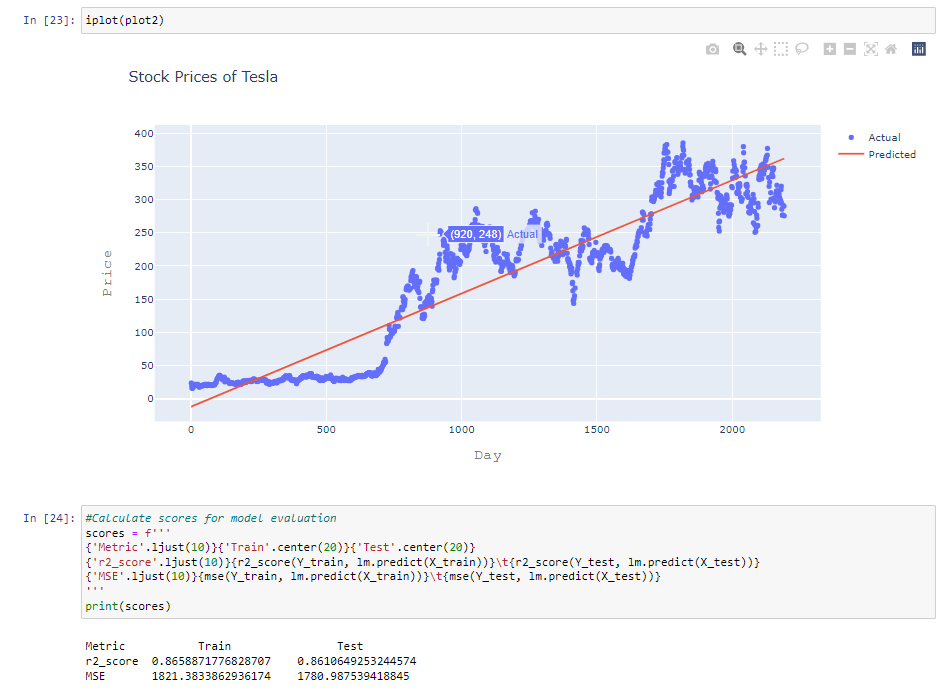
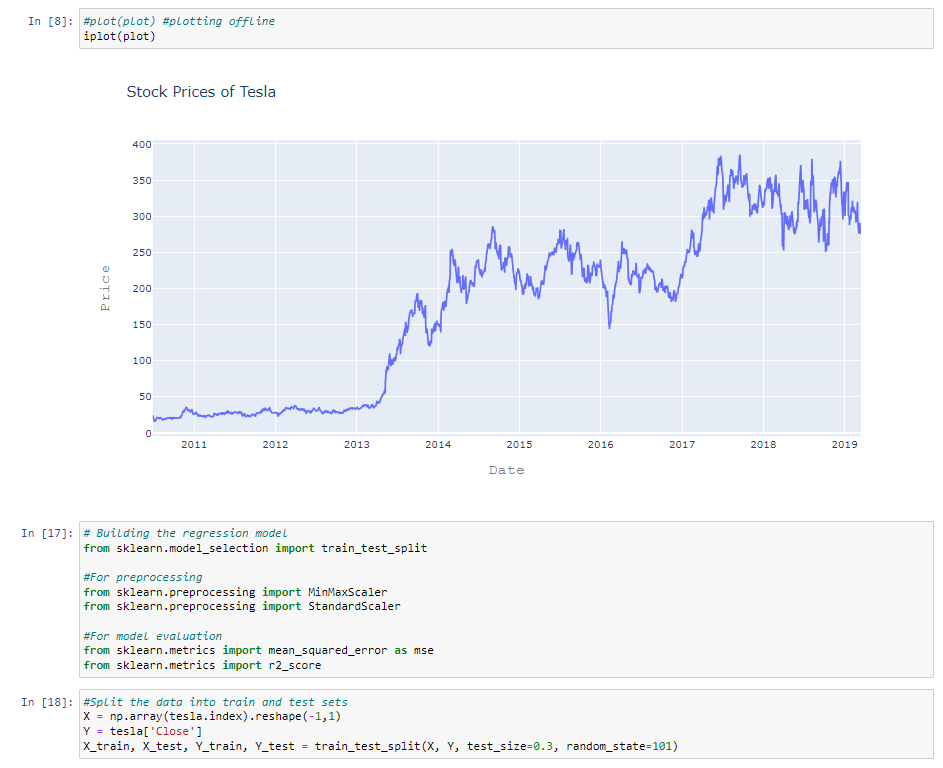
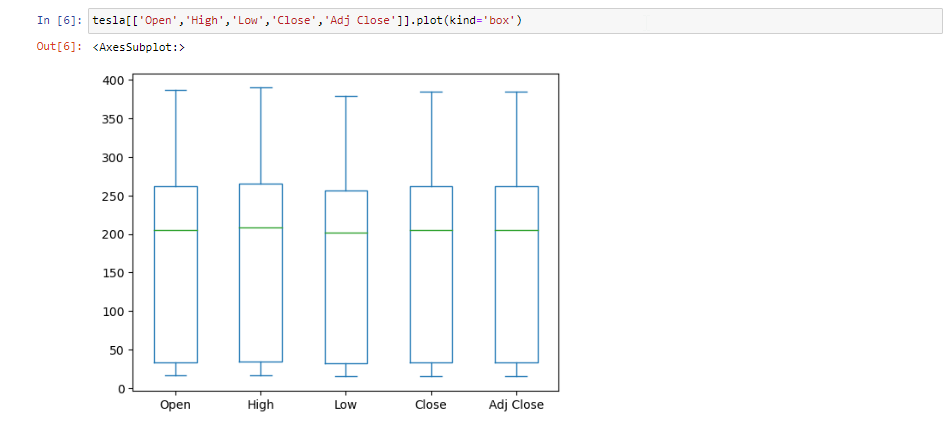
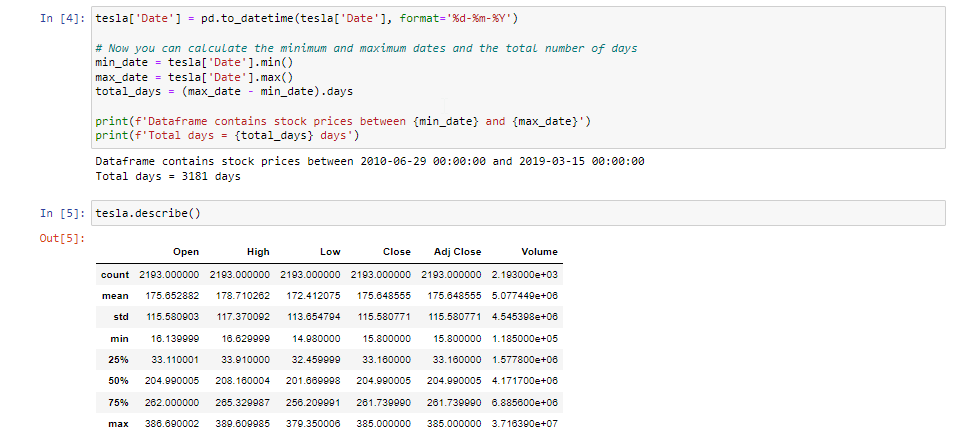
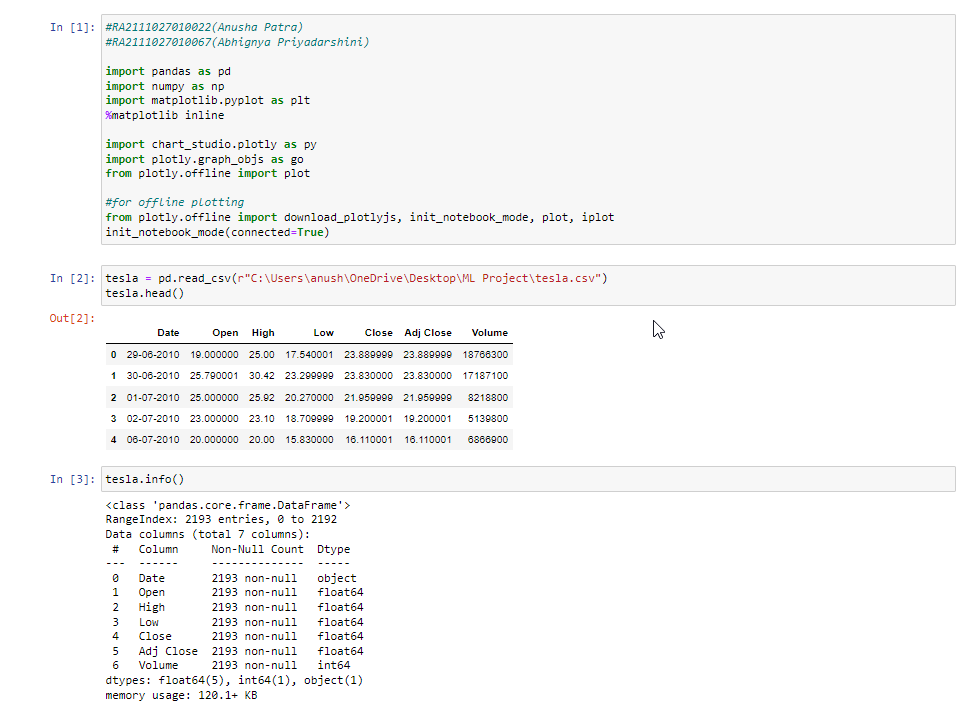
Here,

**SSres:** The sum of squares of the residual errors.

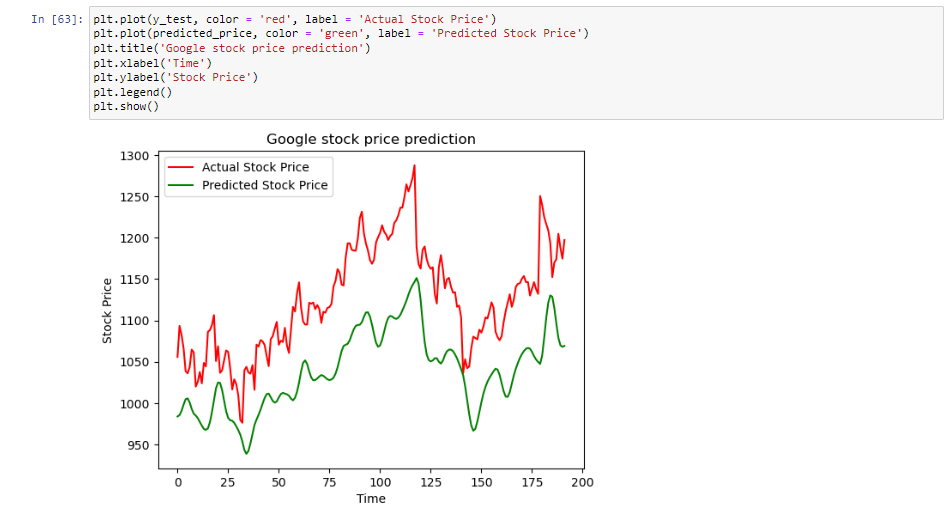
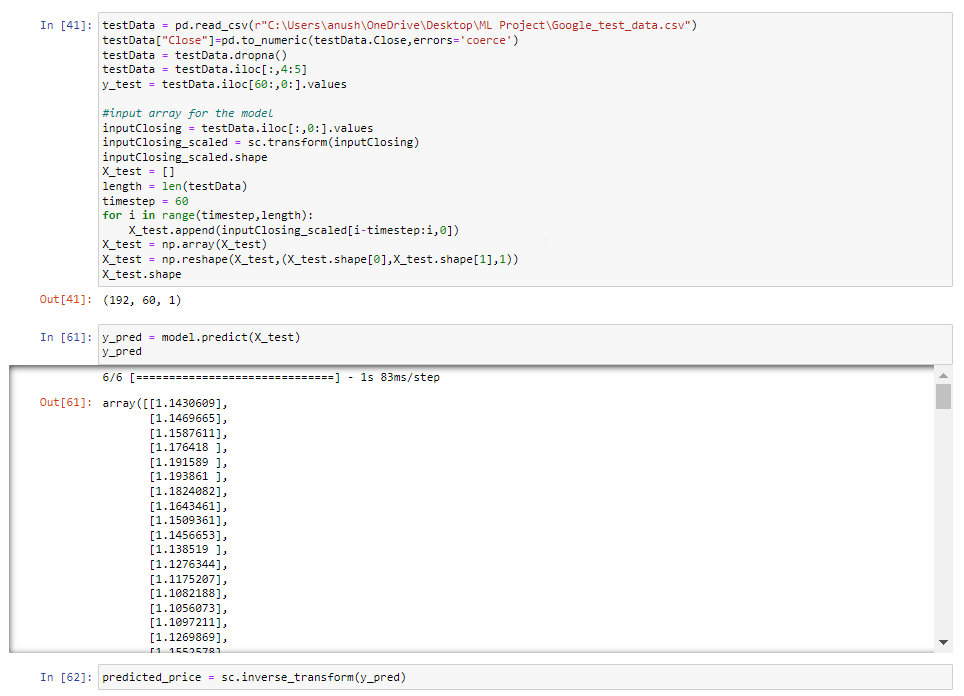
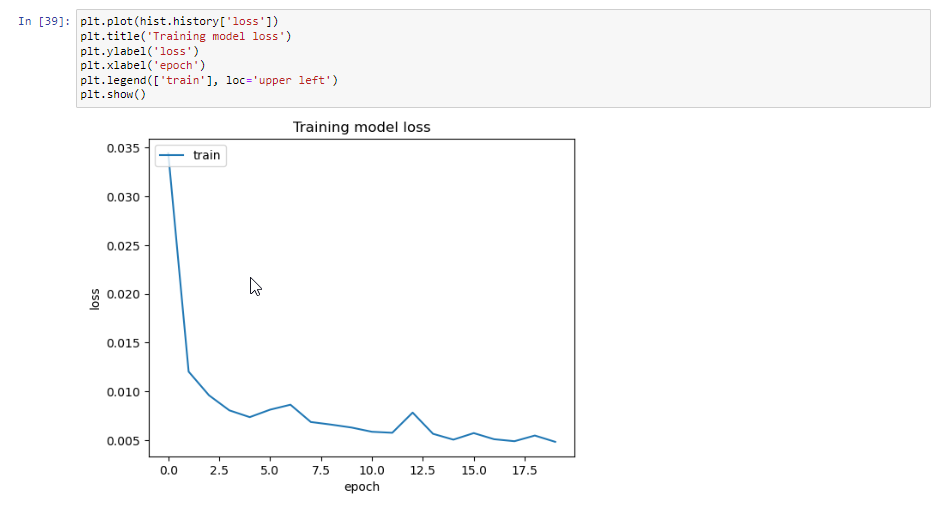
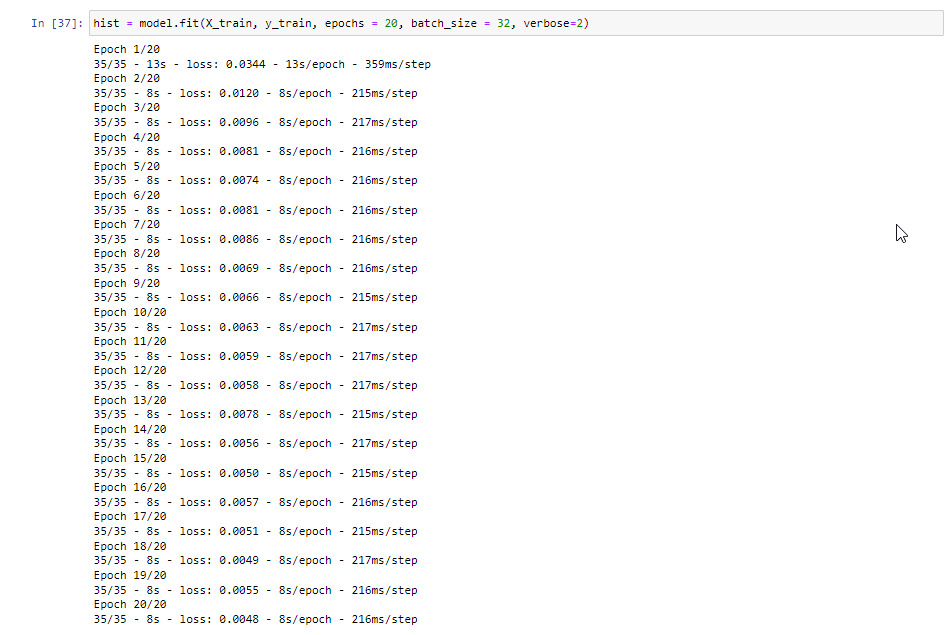
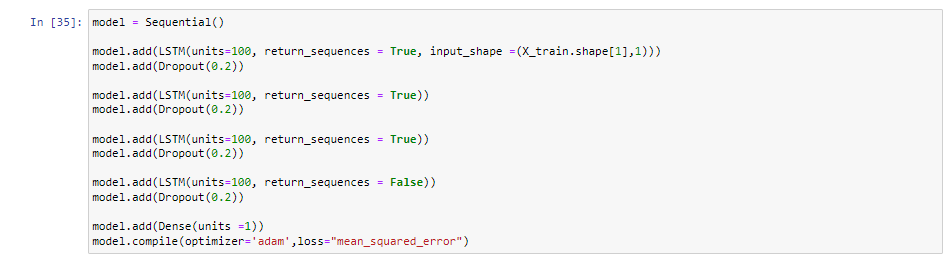
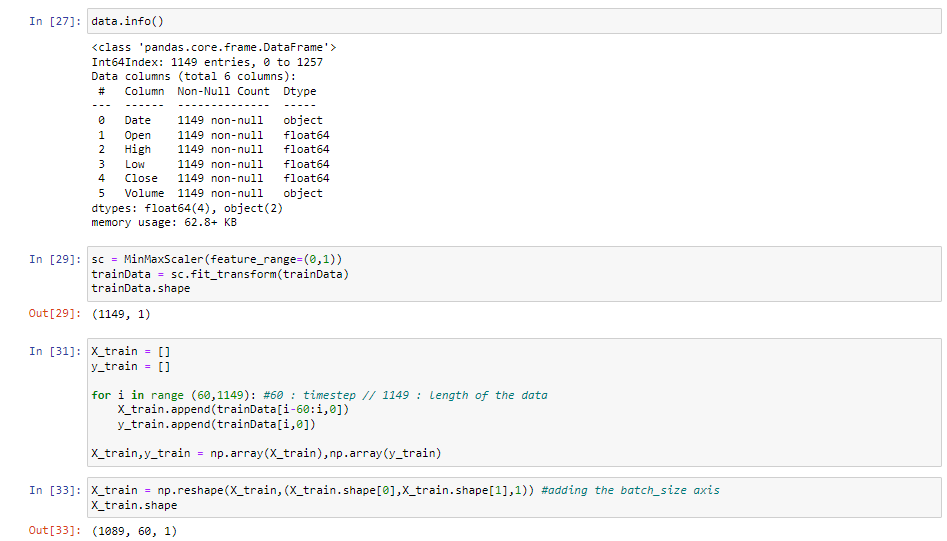
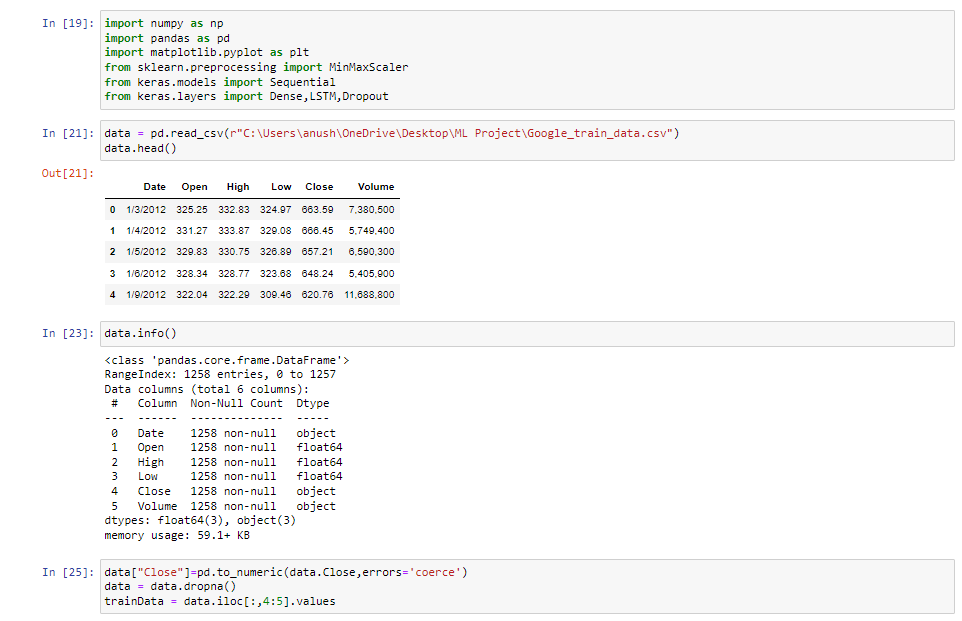
**SStot:** It represents the total sum of the errors.

**CODE**

**STOCK MARKET PREDICTION OF TESLA:**

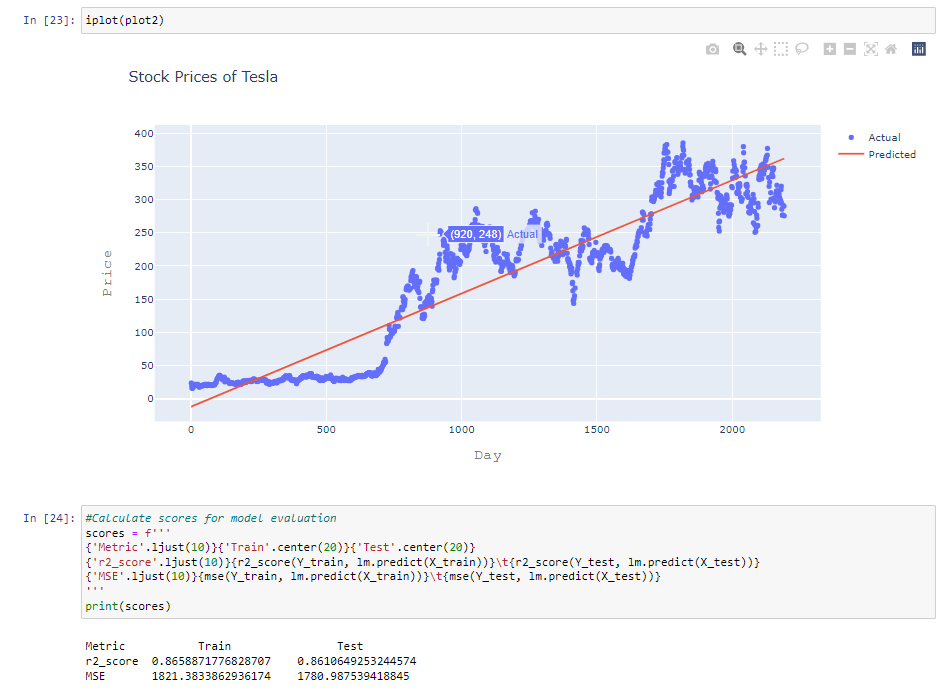
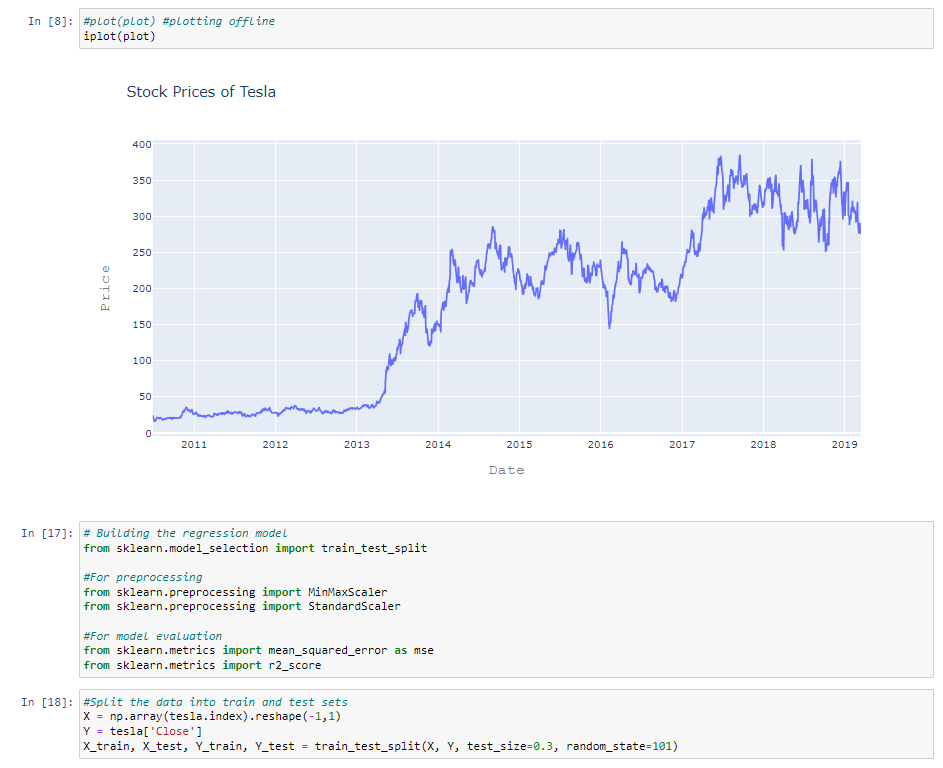
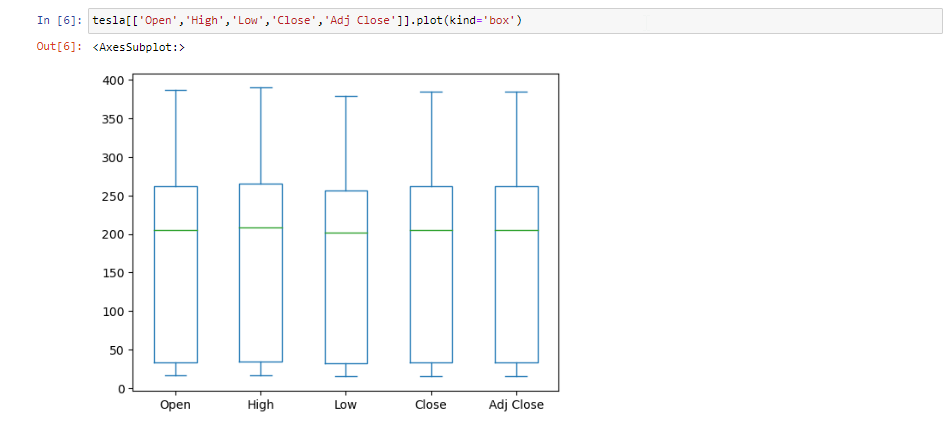
****

**STOCK MARKET PREDICTION OF GOOGLE:**

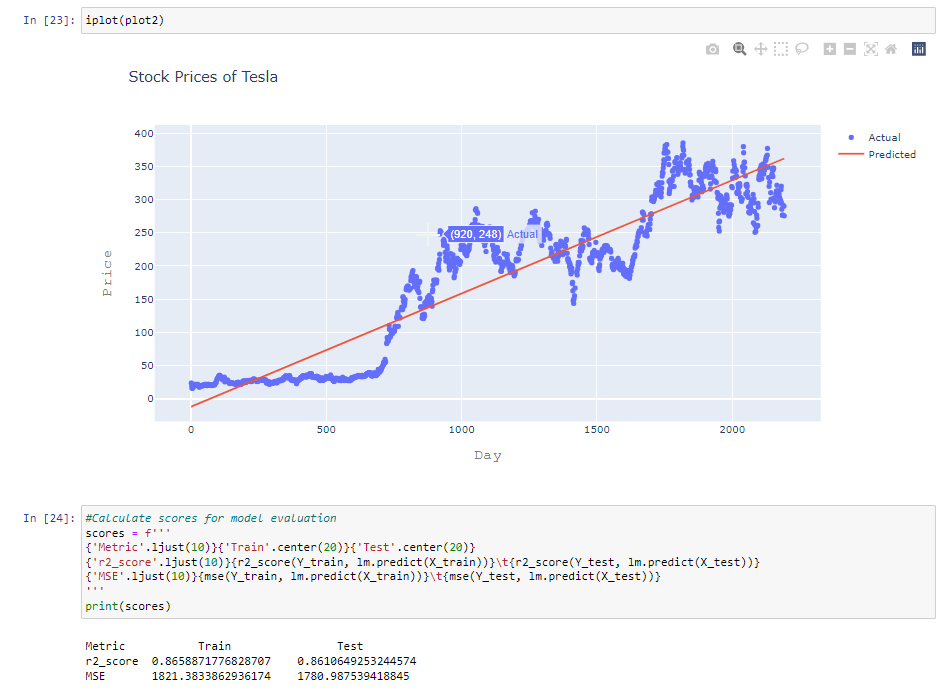
****

**OUTPUT**

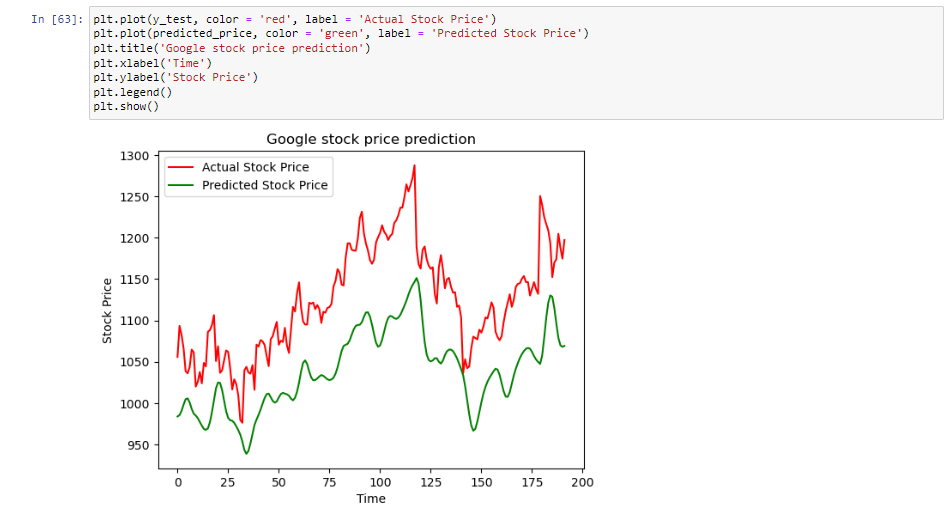
**Stock Market Prediction for Tesla:**

****

**MSE and R2 Scores:**

****

**Stock Market Prediction for Google:**

****

**FUTURE ENHANCEMENTS**

Continuing to improve and expand a machine learning project for stock market prediction is essential to stay relevant and adapt to evolving market dynamics. Here are some potential future enhancements:

**Incorporate Alternative Data Sources:**

Include alternative data sources such as satellite imagery, weather data, and consumer sentiment indices to provide a more comprehensive view of the factors influencing stock prices.

**Ensemble Learning:**

Implement ensemble learning techniques that combine predictions from multiple models to enhance accuracy and robustness.

**Reinforcement Learning for Trading Strategies:**

Develop and integrate reinforcement learning algorithms to automate and optimize trading strategies based on model predictions.

**Real-time Data Streaming:**

Transition to real-time data streaming and analysis to make predictions and decisions on the most up-to-date information.

**Natural Language Processing (NLP) Advancements:**

Improve sentiment analysis with advanced NLP techniques, including sentiment-specific word embeddings and context-aware sentiment analysis.

**Deep Reinforcement Learning:**

Explore deep reinforcement learning for portfolio management and risk assessment, considering multiple assets and their interdependencies.

**Interpretability and Explainability:**

Enhance model interpretability and explainability to make predictions more transparent and understandable to users.

**Automated Hyperparameter Tuning:**

Implement automated hyperparameter tuning methods like Bayesian optimization to optimize model performance efficiently.

**Market Microstructure Analysis:**

Incorporate market microstructure analysis to understand the impact of order book dynamics, liquidity, and trading activity on stock prices.

**CONCLUSION**

The stock market prediction project focusing on Google and Tesla, empowered by machine learning techniques and data-driven insights, represents a significant step towards understanding and harnessing the dynamic world of financial markets. This project has strived to provide a robust framework for predicting the stock prices of these technology giants, offering valuable contributions to investors, traders, and financial analysts.

This project's findings underscore the importance of adapting to market dynamics and sentiment analysis derived from news and social media content. Such insights have the potential to guide investment decisions and risk management strategies effectively. Moreover, the research has highlighted the role of ethical considerations in machine learning projects involving financial data, ensuring fairness and compliance with industry regulations.

As we conclude this project, it's essential to recognize that the ever-evolving nature of financial markets necessitates continuous improvement and adaptation. Future enhancements, including the incorporation of alternative data sources, advanced NLP techniques, and real-time analysis, will further enhance the accuracy and relevance of our predictions. The project stands as a testament to the power of machine learning in navigating the complexities of stock markets and serves as an open door to further innovations and insights in the field of stock market prediction.

**REFERENCES**

1. [**https://scholarworks.lib.csusb.edu/cgi/viewcontent.cgi?article=1435&context=jitim**](https://scholarworks.lib.csusb.edu/cgi/viewcontent.cgi?article=1435&context=jitim)
2. [**https://www.sciencedirect.com/science/article/pii/S1877050922021937**](https://www.sciencedirect.com/science/article/pii/S1877050922021937)
3. [**https://link.springer.com/chapter/10.1007/978-981-16-3915-9\_10**](https://link.springer.com/chapter/10.1007/978-981-16-3915-9_10)
4. [**https://www.ijeast.com/papers/258-262,Tesma508,IJEAST.pdf**](https://www.ijeast.com/papers/258-262,Tesma508,IJEAST.pdf)