**Stock Market Price Prediction Using Linear Regression**

**Abstract**

The stock market is a cornerstone of the global financial system, reflecting economic trends and investor sentiment. Accurate prediction of stock prices offers immense potential, helping traders and investors make informed decisions. This project leverages a machine learning approach, using the **Linear Regression** model to forecast the future closing prices of Apple Inc. (AAPL) stock. The model utilizes historical data to generate lagged features, allowing it to detect trends and patterns in price movements. The effectiveness of the model is evaluated through **Mean Absolute Error (MAE)** and **Root Mean Squared Error (RMSE)** metrics. Additionally, visual comparisons between actual and predicted prices validate the performance of the model. While this project demonstrates the utility of machine learning for financial forecasting, it also highlights the inherent limitations of predicting a highly volatile market. The results provide a foundation for future work that can incorporate more sophisticated models and external factors to improve accuracy.

**Introduction**

The stock market significantly impacts businesses, economies, and individual wealth management. Predicting stock prices is a complex yet crucial task, as accurate forecasts can mitigate risks and optimize investment strategies. Traditional methods like technical and fundamental analyses have been widely employed; however, they often rely on human expertise, making them prone to biases and errors. With the advent of machine learning, data-driven predictions have gained prominence, enabling automation and uncovering hidden patterns in financial datasets.

This project focuses on predicting the closing stock price of **Apple Inc. (AAPL)** using historical data. A **Linear Regression** model is used due to its simplicity, interpretability, and effectiveness in identifying linear relationships. By incorporating lagged features as predictors, the project aims to assess the feasibility of using this model for financial forecasting.

**Existing System**

Traditional approaches to stock market prediction include:

1. **Technical Analysis**:
   * Uses past market data like price and volume.
   * Relies on patterns such as trends, support/resistance levels, and moving averages.
   * Subjective and limited to historical trends without considering external factors.
2. **Fundamental Analysis**:
   * Focuses on a company’s financial health, including revenue, profits, and debt levels.
   * Time-consuming and requires domain expertise.
   * Does not account for short-term market fluctuations.
3. **Expert Opinions**:
   * Involves professional analysts offering insights based on experience.
   * Prone to biases and inconsistency due to human emotions.

These methods, while valuable, have limitations:

* **Subjectivity**: Influenced by individual opinions and interpretations.
* **Scalability**: Inefficient when dealing with large datasets.
* **Inaccuracy**: Unable to adapt to rapid market changes.

To address these challenges, automated systems powered by **machine learning** are being explored.

**Proposed System**

The proposed system utilizes machine learning, specifically Linear Regression, to predict stock prices, focusing on Apple Inc. (AAPL) stock data. This approach aims to analyze historical closing prices, identify patterns, and make short-term predictions. By employing data-driven techniques, the system provides a scalable, interpretable, and automated solution, addressing the limitations of traditional forecasting methods. The methodology is summarized as follows:

**1. Data Collection**

High-quality historical stock data for Apple Inc. was retrieved using the Yahoo Finance API, covering January 2023 to December 2024. The dataset includes:

* **Date:** Transaction date for chronological organization.
* **Close:** Closing price, used as the primary variable for prediction.

This dataset forms the foundation of the analysis, serving as input for preprocessing and modeling.

**2. Feature Engineering**

Lagged features were created to enhance predictive accuracy by capturing sequential dependencies:

* **Lagged Features (Lag\_1, Lag\_2, Lag\_3):** Represent the stock's closing prices from one, two, and three days prior, respectively.  
  These features enable the model to recognize trends and relationships between past and current prices, improving its forecasting capabilities.

**3. Data Splitting**

The dataset was split into training (80%) and testing (20%) subsets to ensure fair evaluation. To preserve the temporal nature of the data, shuffling was avoided. This ensures that the model trains on earlier data and predicts later values, mimicking real-world scenarios.

**4. Model Training**

Linear Regression was chosen for its simplicity and efficiency. This supervised algorithm assumes a linear relationship between lagged features and the target variable (current closing price). During training:

* The model learned the correlation between past and current prices.
* Parameters were adjusted to minimize errors, ensuring the line of best fit captures the underlying trends.

**5. Prediction and Evaluation**

The trained model was tested on unseen data, and its performance was evaluated using:

* **Mean Absolute Error (MAE):** Measures the average magnitude of prediction errors.
* **Root Mean Squared Error (RMSE):** Highlights larger errors, providing insight into prediction precision.

These metrics helped assess the model’s ability to generalize and identify any gaps in accuracy.

**6. Visualization**

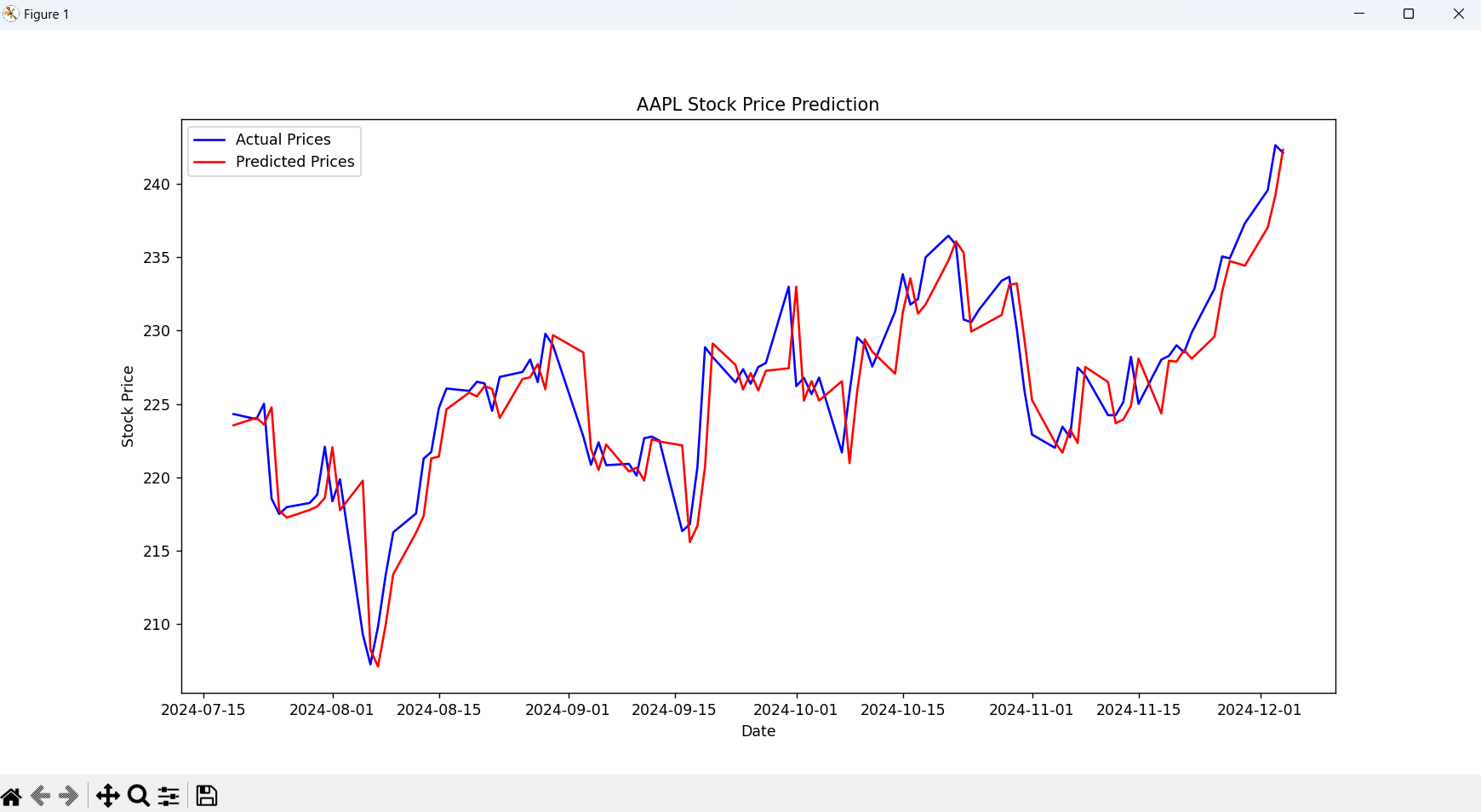
A line graph was created to visually compare the actual closing prices with predicted prices:

* **X-axis:** Timeline.
* **Y-axis:** Stock prices.  
  The visualization highlights how closely the predictions align with actual values, providing insights into the model's strengths and areas for improvement.

**Implementation and Code**

****

**Results :**



**Conclusion**

This project demonstrates the feasibility of using Linear Regression for short-term stock price prediction. The results indicate that while the model is effective in identifying basic trends, its accuracy is limited by the complexities of financial markets. Linear Regression offers a straightforward and interpretable approach, making it suitable for exploratory analysis and foundational research.

Future work could focus on improving prediction accuracy by incorporating additional features such as trading volume, news sentiment, and macroeconomic indicators. Advanced algorithms like Long Short-Term Memory (LSTM) networks or ensemble methods like Random Forests could also be explored for better performance. Additionally, expanding the dataset to include multiple companies and longer time frames could provide a more robust analysis.

In conclusion, this project highlights the potential of machine learning in financial analytics, providing a stepping stone for further exploration in stock market forecasting.