Predicting Ethereum Prices Using Machine Learning by Addan Shahwaiz

Table of Contents

Predicting Ethereum Prices Using Machine Learning by Addan Shahwaiz

Abstract:

Ethereum, a leading blockchain platform enabling smart contracts and decentralized applications, has witnessed significant growth since its inception. This research investigates the potential of machine learning to predict Ethereum's **price movements**. By employing a linear regression model and analyzing historical data, we aim to assess the accuracy and reliability of such predictions. Our findings demonstrate a high level of model performance, with an **R-squared** score of **0.94447**, indicating that the model explains a substantial portion of the price variance. The **low Mean Squared Error (MSE)** and **Root Mean Squared Error (RMSE)** further support the model's accuracy in predicting Ethereum prices.

Introduction:

Ethereum, launched in **2015**, has revolutionized the blockchain landscape by introducing smart contracts and enabling the development of decentralized applications (dApps). Unlike Bitcoin, which primarily functions as a digital currency, Ethereum provides a platform for developers to build and deploy complex applications on a decentralized network. The native cryptocurrency, Ether (ETH), fuels these operations and serves as the backbone of the Ethereum ecosystem.

With the transition to **Ethereum 2.0** and the adoption of the Proof-of-Stake (PoS) consensus mechanism, the platform has enhanced its scalability, security, and energy efficiency. This has propelled Ethereum to the forefront of **decentralized finance (DeFi)** and blockchain innovation, impacting various industries.

Given the **volatility** and dynamic nature of the cryptocurrency market, accurate price prediction is crucial for investors, traders, and researchers. This research aims to explore the feasibility of using machine learning techniques to **forecast Ethereum's price** movements.

Methodology:

This study employs a "**Linear Regression"** model to predict Ethereum's price. The methodology encompasses the following steps:

1. Data Collection and Preprocessing:
    - Historical daily price data of Ethereum from **2017 to 2024** was collected.
    - Data cleaning was performed to handle **missing values** and **outliers**.
    - Feature engineering techniques were applied to extract relevant information from the raw data.

2. Model Development and Training:

    - A linear regression model was selected for its simplicity and interpretability.
    - Feature selection techniques were employed to identify the most influential factors impacting Ethereum's price.
    - The model was trained on historical data to establish relationships between the selected features and the target variable (price).

3. Model Evaluation and Optimization:

    - The model's performance was evaluated using key metrics:
        - **R-squared score:** Measures the proportion of variance in the target variable explained by the model.
        - **Mean Squared Error (MSE):** Quantifies the average squared difference between predicted and actual prices.
        - **Root Mean Squared Error (RMSE):** Represents the standard deviation of the residuals.
    - Techniques such as **hyperparameter tuning** and **cross-validation** were employed to optimize the model's performance and prevent overfitting.

Results:

The analysis yielded the following results:

- Descriptive Statistics:

    - Average Price: **$1423.59**
    - Price Volatility (Standard Deviation): **$1206.07**
    - Maximum Price: **$4812.09** (Date: 2021-11-08)
    - Minimum Price: **$84.31** (Date: 2018-12-14)
    - Average Daily Volume: **12424343615.48 ETH**

- Model Performance:

    - R-squared Score: **0.94447**
    - MSE: $\mathbf{1.588 \times 10^{-21}}$
    - RMSE: $\mathbf{3.985 \times 10^{-11}}$

Discussion:

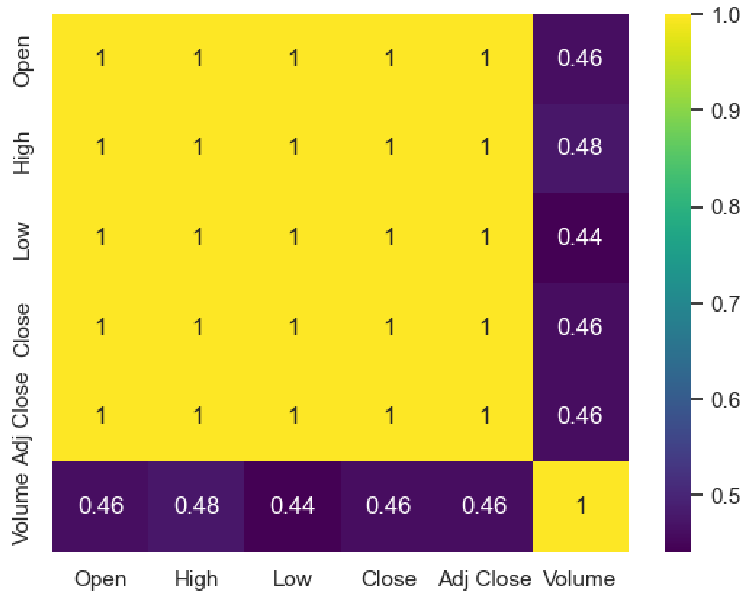
The high **R-squared score** indicates that the linear regression model effectively captures the underlying patterns and trends in Ethereum's price movements. The low **MSE and RMSE** further demonstrate the model's accuracy in predicting future prices. These findings suggest that machine learning techniques, particularly linear regression, can be valuable tools for forecasting Ethereum's price with a reasonable degree of confidence.

Conclusion:

This research provides evidence that machine learning models, specifically linear regression, can be effectively employed to **predict Ethereum's price**. The model's strong performance, as evidenced by the high R-squared score and low error metrics, indicates its potential for practical applications in **investment decision-making**, **risk management**, and **market analysis**.

Additional information

Correlation Matrix:



(Figure 1.1)

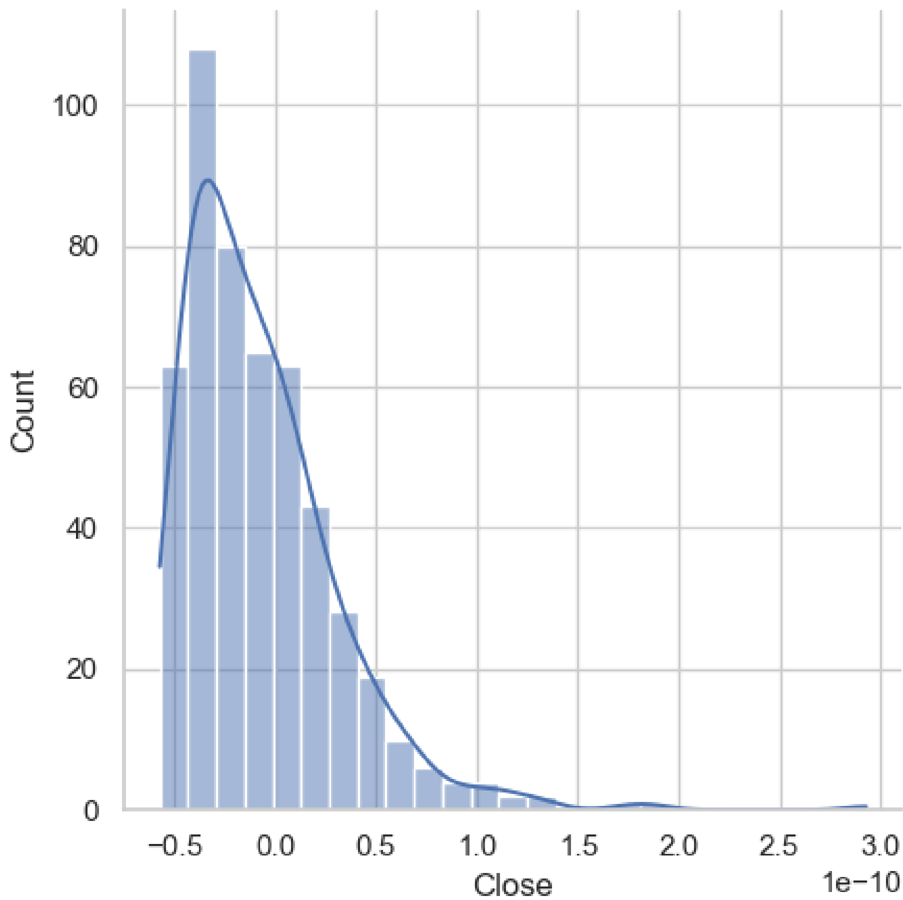
It displays a **correlation heatmap**, which shows the correlation coefficients between different financial metrics of a specific asset, possibly Ethereum. The **colors** indicate the strength of the correlation, with **yellow** representing a strong positive correlation and **purple** indicating a strong negative correlation. The heatmap reveals that the asset's **"Open," "High," "Low," "Close,"** and **"Adj Close"** prices are highly positively correlated, which is typical for financial data. The **"Volume"** metric shows a moderate positive correlation with the price metrics, suggesting that trading volume might influence the asset's price, although other factors likely play a larger role.

Predicted Values:

([1892.51281738, 1908.91699219, 1876.92431641, 1831.95483398,1870.78930664, 1904.65185547, 1877.70410156, 1995.06091309,1900.22180176, 1873.07641602, 1849.04272461, 1848.60314941,1842.40148926, 1796.49060059, 1808.01977539, 1796.11486816,--

,2427.90234375, 2538.18725586, 2420.60375977, 2448.97705078, 2367.73754883, 2223.87646484, 2274.10717773, 2297.29296875]).

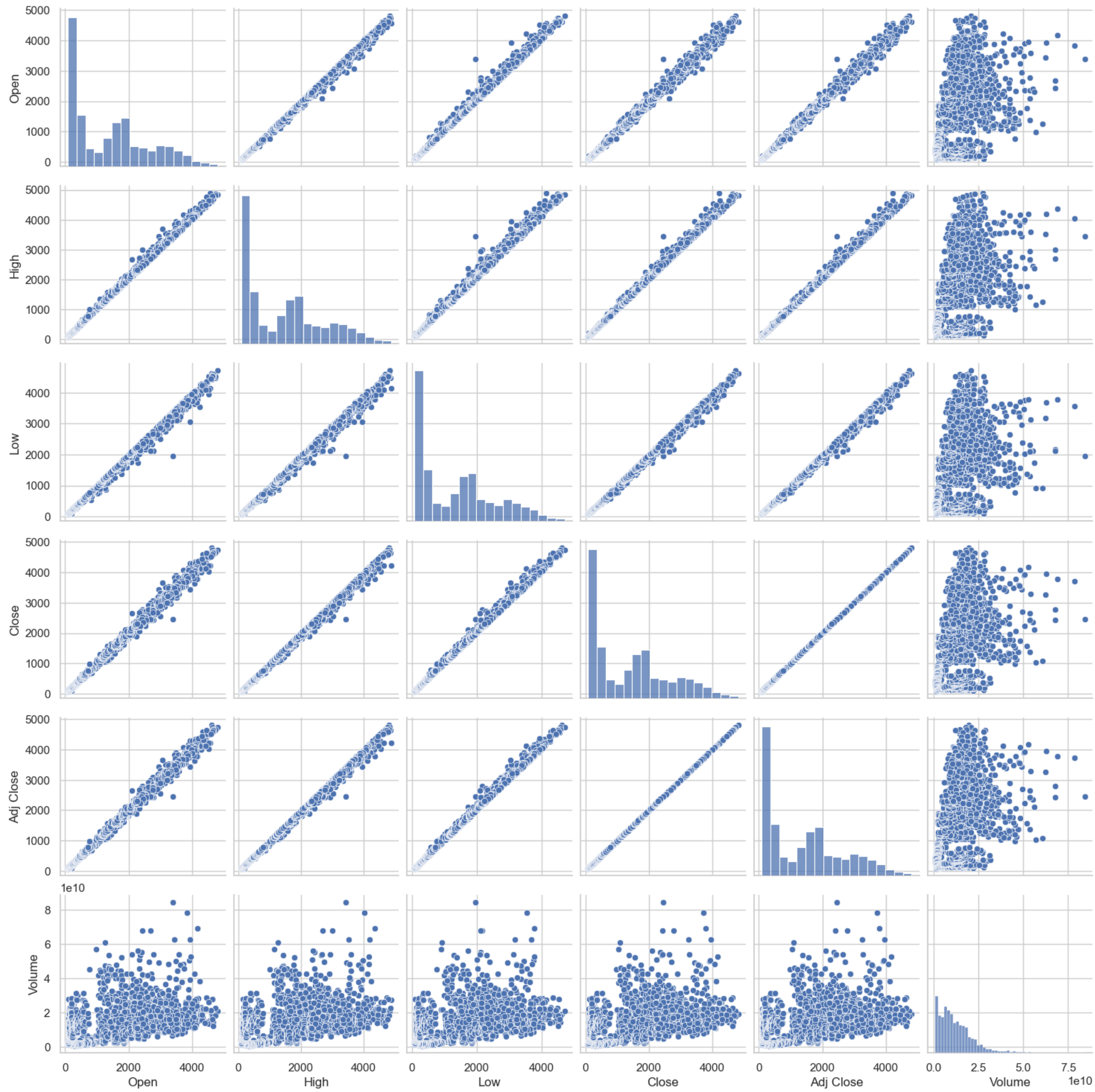
Histogram:



(Figure 2.1)

The graph shows a **histogram** of the dataset, likely representing the distribution of the "Close" variable. The histogram displays the **frequency** of data points within specific intervals. The distribution is **skewed to the right**, meaning most data points are concentrated at the lower values of the variable. The **curve** overlaid on the histogram represents the **probability density function**, providing a continuous view of the data distribution.

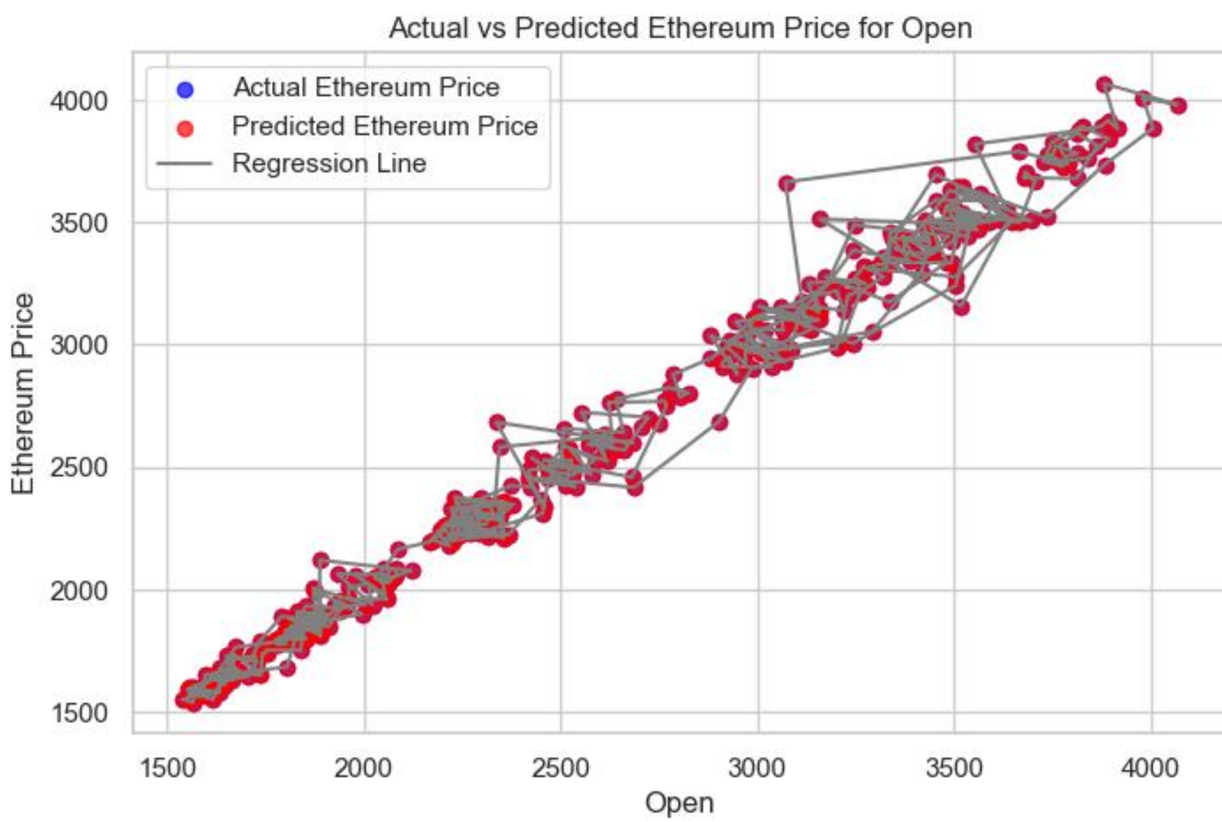
Pair Plot:



(Figure 3.1)

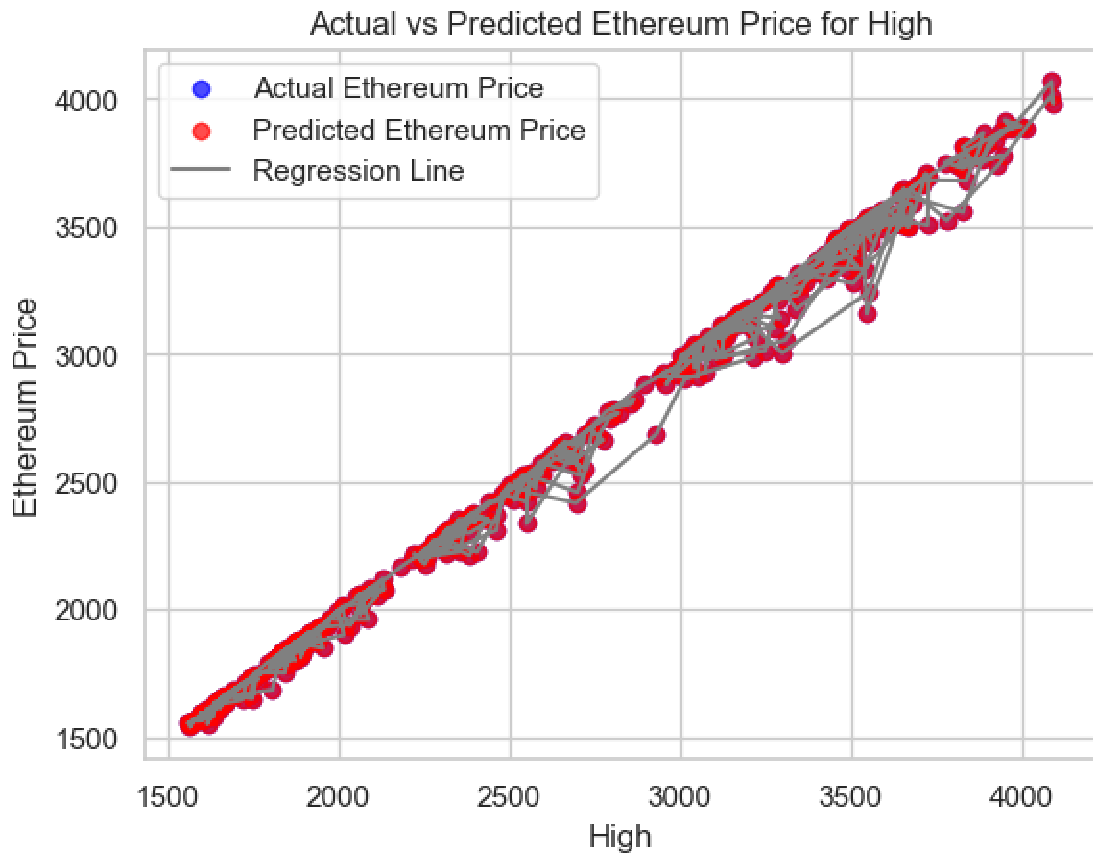
The image displays a **pair plot**, a visualization technique that helps explore relationships between multiple variables in a dataset. In this case, the pair plot shows the relationships between various financial metrics for an asset, likely Ethereum. Each subplot represents a pair of variables. The **diagonal subplots** display the distribution of each individual variable (e.g., histograms for "Open," "High," "Low," "Close," "Adj Close," and "Volume"). The **off-diagonal subplots** show **scatter plots**, illustrating the relationship between two variables. For example, the subplot in the second row and first column shows the scatter plot between "High" and "Open" prices. By examining these plots, we can identify **correlations** and potential **patterns** between the metrics, which are valuable for analysis or prediction.

Graph Plots:



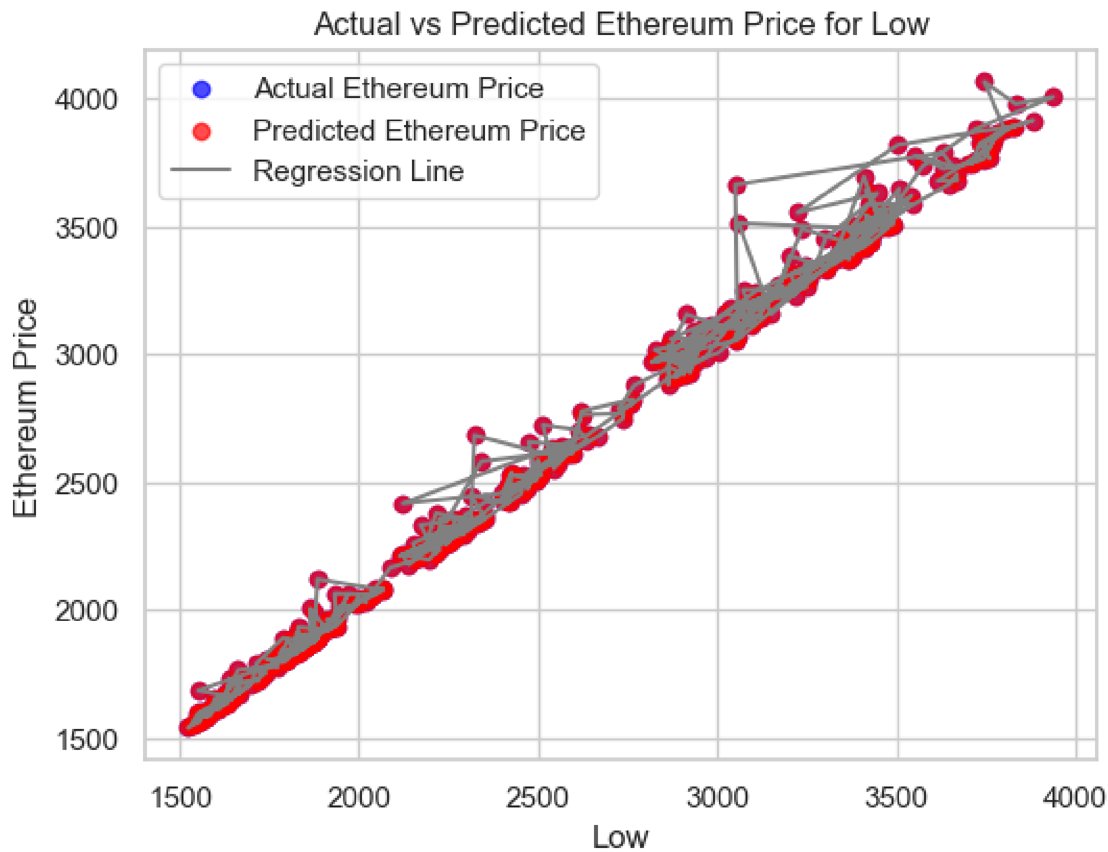
(Figure 4.1)

The graph compares the **real Ethereum prices** (blue dots) and **predicted prices** (red dots) from a machine learning model, with a **regression line** illustrating the overall trend. While the model successfully captures the **positive correlation** between actual and predicted prices, the scattered points around the line indicate that the predictions are not perfect. This suggests that, although the model reflects the general trend, there is still **room for improvement** in its accuracy.



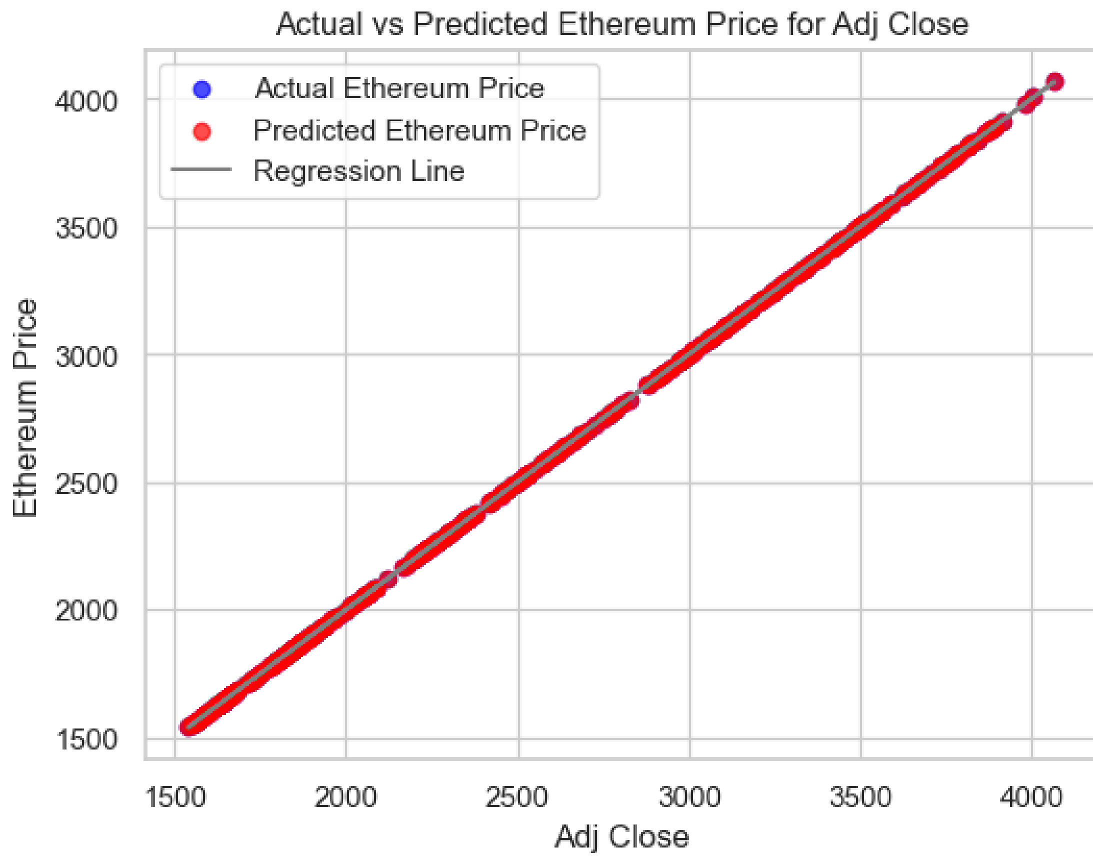
(Figure 4.2)

The graph titled "Actual vs Predicted Ethereum Price for High" shows how well a machine learning model predicts Ethereum's price. **Blue dots** represent the actual "High" prices, while **red dots** show the predicted prices. The **regression line** shows a positive trend, indicating a good correlation between actual and predicted prices. However, the **scattered red dots** suggest that the model's predictions are not perfect, highlighting that there is still **room for improvement** in accuracy.



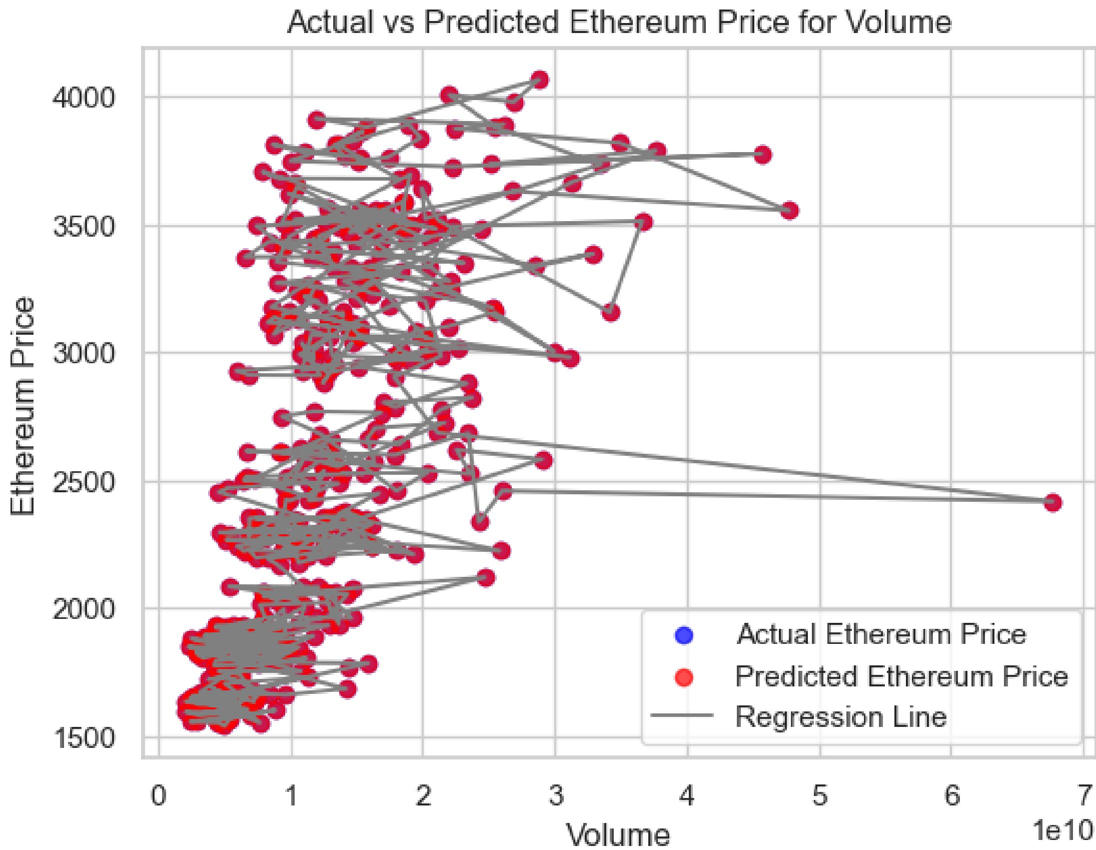
(Figure 4.3)

The graph titled "Actual vs Predicted Ethereum Price for Low" compares **actual Ethereum "Low" prices** (blue dots) with the **predicted prices** (red dots) from a machine learning model. The **regression line** shows a positive trend, indicating a good overall match between actual and predicted values. However, the **scattered red dots** around the line suggest that the model's predictions are not always accurate, showing there is **room for improvement** in its precision.



(Figure 4.4)

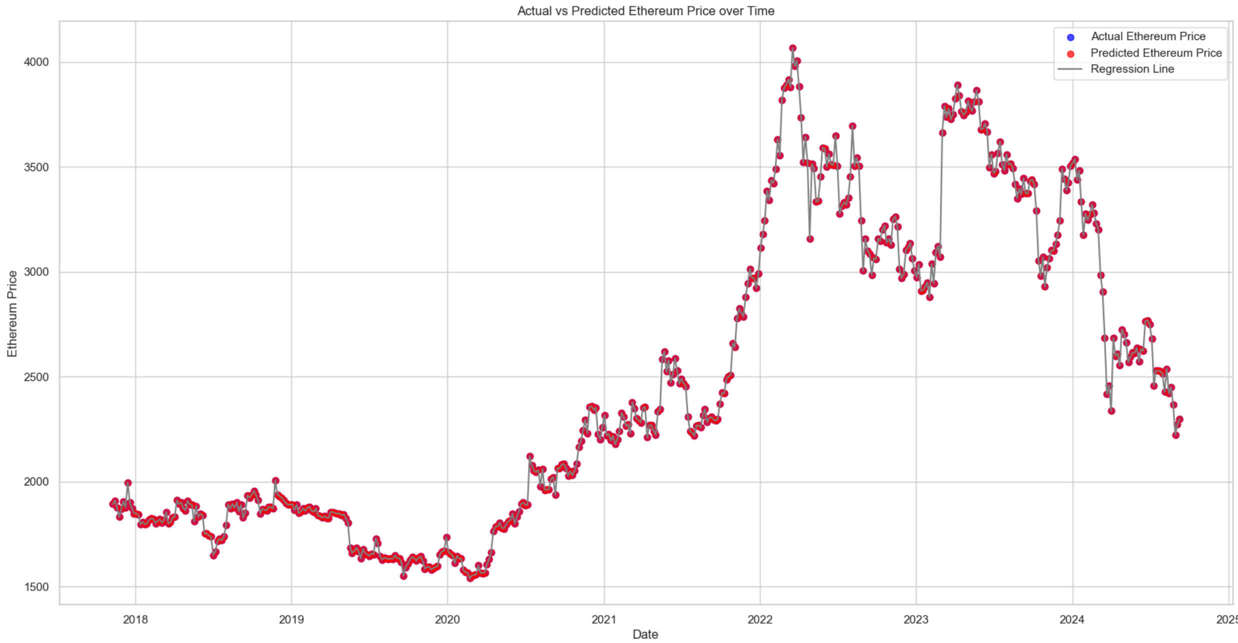
The graph titled "Actual vs Predicted Ethereum Price for Adj Close" compares **actual Ethereum "Adj Close" prices** (blue dots) with **predicted prices** (red dots) from a machine learning model. The **regression line** shows a strong positive correlation between actual and predicted values. The points are tightly clustered around the line, indicating that the model's predictions are **highly accurate** for the **"Adj Close" price** of Ethereum.



(Figure 4.5)

The graph titled "Actual vs Predicted Ethereum Price for Volume" compares **actual Ethereum prices** (blue dots) with **predicted prices** (red dots) based on trading volume. The **regression line** shows the general trend, but the relationship is less clear than with other price metrics. The **scattered points** suggest that the model's accuracy in predicting Ethereum price using only trading volume is limited, indicating that **other factors**, like market sentiment or news events, may have a stronger influence on price.

Past Data of Etherium Coin:



(Figure 5.1)

This graph shows how well a machine learning model predicts Ethereum's price over time. **Blue dots** represent the **actual Ethereum prices**, while **red dots** show the **predicted prices** from the model. The **gray regression line** captures the overall trend between actual and predicted prices.

Although the model generally follows the trend of actual prices, there are instances where the predicted values deviate, indicating some inaccuracies. The **scatter around the line** suggests that while the model captures the general trend, there is still **room for improvement** in predicting Ethereum's price fluctuations more precisely.

References:

Kim, H.M., Bock, G.W. and Lee, G., 2021. Predicting Ethereum prices with machine learning based on Blockchain information. *Expert Systems with Applications*, *184*, p.115480.

Chen, M., Narwal, N. and Schultz, M., 2019. Predicting price changes in ethereum. *International Journal on Computer Science and Engineering (IJCSE) ISSN*, pp.0975-3397.

Chen, M., Narwal, N. and Schultz, M., 2019. Predicting price changes in ethereum. *International Journal on Computer Science and Engineering (IJCSE) ISSN*, pp.0975-3397.