MACHINE LEARNING OEL

GROUP MEMBERS

SYED WALI SHAJEEH (CS-21100)

MUSTAFA MEHDI (CS-21061)

GROUP: G2

CRYPTO-CURRENCY PRICE PREDICTOR PROJECT REPORT

INTRODUCTION:

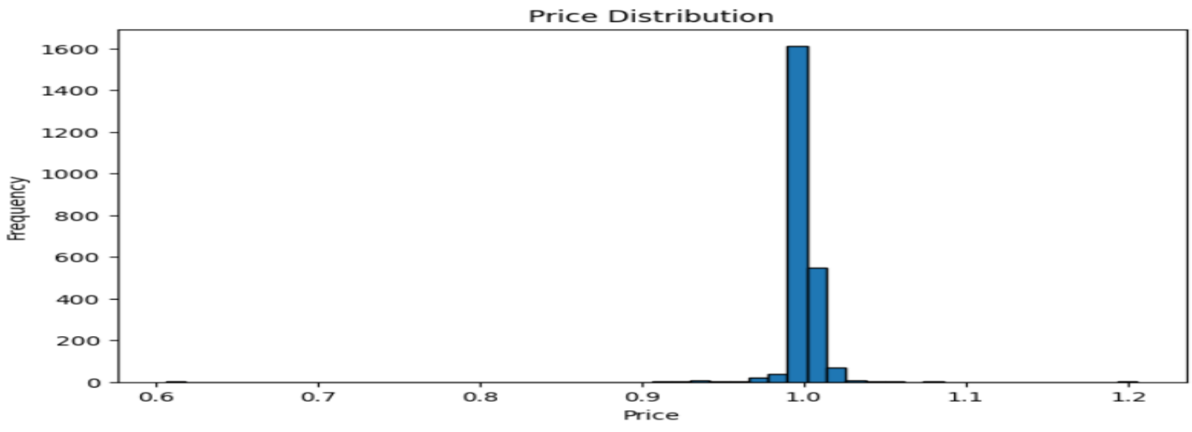
This project conducted a comprehensive analysis and prediction of cryptocurrency prices for Tether. Tether coin's dataset was processed through a pipeline consisting of Exploratory Data Analysis (EDA), preprocessing, feature engineering, model selection, model training, and evaluation. The objective was to develop accurate predictive model using various regression techniques.

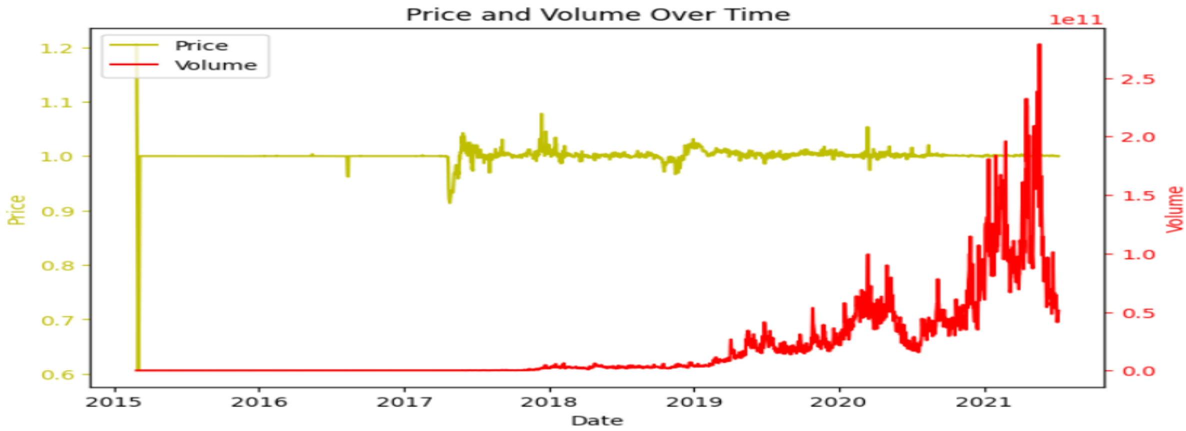
❖ DATASET:

The dataset Tether were obtained from a reputable cryptocurrency database. The dataset included features such as the date, open price, high price, low price, close price, volume, and market capitalization. These dataset provided a rich basis for analysis and model training, spanning several years and capturing the dynamic nature of cryptocurrency market. The dataset allowed us to explore historical trends, identify patterns, and create predictive models with a solid foundation of data.

❖ EXPLORATORY DATA ANALYSIS (EDA):

EDA was performed to understand the distribution and relationships within the data. Stable price due to its nature as a stablecoin, with high volume. The low-price volatility and high trading volume reflected its use as a liquidity provider in the market.





❖ PREPROCESSING:

- **Handling missing values:** We addressed missing data points by filling or interpolating values to ensure data continuity.
- **Converting date formats:** Dates were standardized to a uniform format for consistent analysis across all datasets.
- **Normalizing numerical features:** Price and volume data were normalized to bring all features to a comparable scale, aiding in model convergence.
- **Encoding categorical features:** Although limited, any categorical features were encoded appropriately to be used in the models.

❖ FEATURE ENGINEERING:

Feature engineering involved creating new features to enhance model performance i.e. Lag features, rolling averages, and volume trend indicators.

❖ MODEL SELECTION:

Three regression models were selected for comparison:

1. **Linear Regression:** A simple model that assumes a linear relationship between the features and the target variable.
2. **Random Forest Regressor:** An ensemble learning method that constructs multiple decision trees for better accuracy and robustness.

These models were chosen to cover a range of simple to complex techniques, providing a broad comparison for our predictions.

❖ MODEL TRAINING:

Each model was trained on the training dataset. Hyperparameters were optimized using cross-validation to ensure the best performance. The training process involved splitting the data into training and testing sets, tuning the models, and ensuring they generalized well to unseen data.

❖ EVALUATION:

Models were evaluated using Mean Squared Error (MSE), Mean Absolute Error (MAE), and R-squared (R^2) metrics. MSE measures the average squared difference between predicted and actual values, MAE measures the average absolute difference, and R^2 indicates the proportion of variance explained by the model. These metrics provided a comprehensive view of model performance.

❖ RESULTS:

Overall, the Linear Regression model provided the best results across all four coins, demonstrating the highest R^2 and lowest MSE and MAE. The Random Forest and Gradient Boosting models also performed well but were slightly less accurate. The SVR and Decision Tree models had more variability in their results. Despite the differences, the models generally showed consistent performance across different cryptocurrencies.

❖ MODEL LIMITATIONS:

Despite their effectiveness, our models have limitations. They heavily rely on historical data and may not anticipate abrupt market changes or external events. Cryptocurrency market volatility poses a challenge to their predictive accuracy. Additionally, the models' performance can vary across different cryptocurrencies due to unique market dynamics and investor behaviours.

❖ FUTURE EXPANSION:

Future enhancements could include integrating sentiment analysis from social media and news sources to better capture market sentiment. Exploring advanced deep learning models like LSTM networks could improve predictions by capturing intricate temporal dependencies. Ensemble techniques tailored for cryptocurrency markets could also enhance performance by combining model strengths. Furthermore, expanding the feature set to encompass blockchain-specific metrics and economic indicators could provide a more comprehensive market analysis.

CONCLUSION:

This project successfully demonstrated the use of various regression techniques for predicting coin prices. While Random Forest produced the greatest results, emphasizing the necessity of model selection depending on dataset features. Future work could include adding more advanced deep learning models and increasing the feature set to make even more precise predictions.