**Ethereum Price Forecasting Model Report**

Description of the Datasets:

Data Collection

* Ethereum Price Data
* **Source:** Kaggle - https://www.kaggle.com/datasets/varpit94/ethereum-data
* Period: 2017-11-09 to 2022-03-25
* Features:
  + Date: Timestamp.
  + Open, High, Low, Close, Adj Close: Daily price indicators.
  + Volume: Trading volume.
* Cryptocurrency News Data.
  + **Source:** Kaggle - <https://www.kaggle.com/datasets/oliviervha/crypto-news>
  + Period: 2021-10-12 to 2023-04-05
  + Features:
    - Date: Timestamp.
    - Sentiment: Positive, Negative, Neutral.
    - Source: sources of news.
    - Subject: bitcoin, altcoin, nft, blockchain, ethereum.
    - Text: First paragraph of the article.
    - Title: Title of the news.
    - Url: URL of the news.

Data Preprocessing

* Ethereum Price Data
  + Convert the 'Date' column to datetime and set it as the index.
  + Check for missing values.
* Cryptocurrency News Data
  + Convert the 'Date' column to datetime in ‘Y-m-d’ format and set it as the index.
  + Sorting data by Date column “acending”.
  + Filter eth price data based on the date range (2021-2023).
  + Convert ‘sentiment’ column to numeric values by extracting the value of class .
  + Drop all columns in news data exclude ‘sentiment\_mapped’.
  + Group data by ‘date’ and calculate the majority vote sentiment.
  + Group data by ‘date’ and calculate the sum of sentiments.
  + Merge eth price data with news data.
  + Fill NA values with 0.
  + Create ‘next\_day\_close’ column then create the ‘price\_movement’ column from (close, next day close) to get the class, 1 if next close larger than close and 0 if next close smaller than close.

Description of the Used Models:

Time Series Forecasting Models

* + Chosen Models: LSTM, Bi-LSTM, Echo State Network.
  + LSTM predicts Ethereum prices well by remembering long-term patterns.
    - Parameters:
      * Time step: 14
      * Hidden layer sizes: 64
      * Loss: MSE
      * Optimizer: Adam
  + Bi-LSTM sees both past and future trends, improving its ability to predict Ethereum prices.
    - Parameters:
      * Time step: 14
      * Hidden layer sizes: 64
      * Loss: MSE
      * Optimizer: Adam
  + Echo State Network efficiently learns and predicts Ethereum price changes with a fixed structure.
    - Parameters:
      * Input size: 1
      * Reservoir size: 100
      * Output size: 1
      * Leak rate: 0.5
      * Spectral radius: 0.6

ClassificationModels

* + Chosen Models: Multi-Layer Perceptron (MLP), SVM, Random Forest.
  + MLP is a flexible learner adept at recognizing complex patterns in data.
    - Parameters:
      * Activation: tanh
      * Learning rate: 0.0001
      * Hidden layer sizes: (50, )
      * Max iteration: 500
      * Solver: SGD
  + SVM demonstrates a keen ability to draw clear boundaries between different data points, aiding in effective prediction.
    - Parameters:
      * C: 1
      * Kernal: linear
  + Random Forest is a collaborative model that combines insights from multiple decision trees for perfect predictions.
    - Parameters:
      * Bootstrap: True
      * Min samples leaf: 4
      * Min samples split: 2
      * N estimators: 200

Experiment setup and results:

Setup: The dataset was split into training and testing sets using 80% split. Time series models used a 14-day rolling window for training.

Time Series Forecasting models:

* Results
  + - LSTM outperformed with MSE of 0.017.
    - ESN outperformed with MSE of 0.0003.
    - Bi-LSTM outperformed with MSE of 0.022.

Classification Models:

* Results
  + SVM outperformed with accuracy of 55%.
  + Random Forest outperformed with accuracy of 63%.
  + MLP outperformed with accuracy of 64%.
* Additional Models
  + Xgboost outperformed with accuracy of 63%.
  + Catboost outperformed with accuracy of 60%.
  + KNN outperformed with accuracy of 60%.
  + GaussianNB outperformed with accuracy of 55%.
  + LSTM outperformed with accuracy of 72%.

Discussion of the results

**LSTM**: Performed well with an MSE of 0.017, showcasing its proficiency.

**Echo State Network (ESN):** Performed the best with the lowest MSE of 0.0003, making it a strong candidate**.**

**Bi-LSTM:** Showed decent performance with an MSE of 0.022.

**Multi-Layer Perceptron (MLP):** Led with the highest accuracy of 64%, demonstrating discerning ability in classifying Ethereum price movements.

**Random Forest:** Showcased strong classification performance with an accuracy of 63%, adapting well to intricate patterns.

**Support Vector Machine (SVM):** Showed competitive performance with an accuracy of 55%.

Python code:

The Python code for the implemented models and experiments can be found on GitHub. Please visit the following link for the complete code

<https://github.com/Huthayfa-Hodeb/Ethereum-Price-Prediction>