Dataset Documentation

for

Developing Trading Strategies in Decentralized Trading Markets Prediction by Using AI, ML And Blockchain Technology

Researchers Name: Md Abu Sufian

MSc in Data analysis for business intelligence

Department: Mathletics and Computer Science

University of Leicester

Developing Trading Strategies in Decentralized Markets Prediction by Using AI, ML And Blockchain Technology

Data sources

In this section, I used two sets of datasets from the yahoo finance and Kaggle website to get a glimpse of real-life analysis using machine learning, artificial intelligence and blockchain tools by Python code.

- https://finance.yahoo.com/quote/LQDT/history?p=LQDT (Liquidity variable)
- https://finance.yahoo.com/quote/%255EVIX/history/ (Volatility variable)
- https://uk.finance.yahoo.com/quote/GDPHF/history?p=GDPHF (GDP variable)
- https://finance.yahoo.com/quote/%5ECARDS.REGA/history?p=%5ECARDS.REGA (trading volumes)
- https://www.kaggle.com/datasets/equinxx/stock-tweets-for-sentiment-analysis-and-prediction (Sentimental analysis)

Data format and structure:

The file format (CSV & Excel,) and the structure of the data (e.g., columns, rows) to facilitate further analysis.

Dataset procedure:

These datasets are the secondary source that has been collected from websites. The data presented are from secondary data but not from web scraping.

Dataset Overview

The datasets are one year from 15 March 2022 to 15 March 2023. The dataset focuses on developing AI, ML, and blockchain models to develop decentralized marketing strategies.

The datasets are divided into different parts, namely, AI, ML and sentimental analysis, to give the recent technologies used for decentralized marketing.

Variables

1.Market indicators: Variables that capture the overall market trends, such as market volatility, liquidity, and trading volume. These variables were collected from yahoo finance.
2. Social media sentiment: Variables that capture the sentiment and opinions of the public or experts about the prediction. This can be obtained from social media platforms, news outlets, or surveys. This variable was obtained from Kaggle.
3. Historical data: Variables that capture the past performance of the prediction, such as historical prices or trading volumes. This variable was obtained from yahoo finance.
4Economic indicators: Variables that capture the overall economic conditions and events that may impact the prediction, such as interest rates, GDP growth, or inflation. This variable was obtained from Yahoo finance
5User behaviour: Variables that capture the behaviour of users in the decentralized prediction market, such as user activity, user reputation, or user stakes. This variable was obtained from the Kaggle link dataset.
6. Other prediction markets: Variables that capture the performance of similar prediction markets, as this may provide insights into the market behaviour. It was collected from yahoo finance.

Data Collection:

Yahoo Finance and Kaggle were the two main places where the information for this project was gathered. As well as information on market volatility, liquidity, trading volume, GDP growth, and the success of comparable prediction markets, Yahoo Finance also supplied historical data and economic indicators. Data on user activity, reputation, and stakes in the decentralised prediction market were provided by Kaggle, together with information on social media sentiment and user behaviour. It's crucial to remember that the data wasn't scraped from the web but rather obtained from secondary sources. The dataset covers a one-year period from March 15, 2022, to March 15, 2023, and is divided into different parts to reflect

the various technologies used for decentralized marketing, including AI, ML, and sentimental analysis.

Data Usage

The public may freely utilize the blockchain dataset for non-commercial purposes concerning the limitations and constraints set up by the blockchain dataset accessibility. In any publications or software applications that use the data, users must give the data's creator credit. Review the dataset documentation and use instructions before using the data for any analysis or application; users are recommended.

Limitations

Although the dataset provides a complete summary of the decentralized market, there are certain constraints on the data. Certain data points may be restricted or anonymized, and some data may not be accessible for all years or periods to protect people's privacy. Users are advised to read the dataset description and use instructions to ensure they utilize the data correctly and within the permitted scope.

Recommendation

The large dataset provides more accurate results and accuracy. More research and analysis of the big data set would be beneficial for making decisions.