

Project Report: Analysis of Cryptocurrency Trading Behavior and Market Sentiment

1. Abstract

This report details an analysis of cryptocurrency trading data to uncover the relationship between trader behavior and market sentiment. By merging historical trading records with the Fear and Greed Index, we identify distinct patterns in trading activity across different sentiment classifications (Extreme Fear, Fear, Neutral, Greed, Extreme Greed). The analysis reveals that periods of "Fear" are associated with the highest net daily profits, while "Extreme Greed" correlates with increased risk-taking. To further explore these relationships, an ElasticNet regression model is developed to predict daily profit and loss, demonstrating the potential for building predictive models based on market sentiment. The key findings are presented in a comprehensive dashboard that visualizes the behavioral differences between trader groups.

2. Introduction

The cryptocurrency market is notoriously volatile and heavily influenced by investor sentiment. Understanding how traders behave under different market conditions is crucial for developing effective trading strategies and managing risk. This project aims to provide a data-driven analysis of this relationship by answering the following questions:

- How do trading volume, volatility, and profitability differ across various market sentiment levels?
- Are there distinct behavioral patterns associated with "Fear" and "Greed"?
- Can we build a predictive model for daily profit and loss based on trading activity and market sentiment?

To address these questions, we perform a comprehensive analysis that includes data preprocessing, feature engineering, exploratory data analysis, and predictive modeling.

3. The Data

The analysis is based on two primary datasets:

- **historical_data.csv**: This dataset contains anonymized records of individual trades, including details such as the coin traded, execution price, trade size (in tokens and USD), and the profit or loss (PnL) from each trade.
- **fear_greed_index.csv**: This dataset provides the daily Fear and Greed Index, a widely used metric for gauging market sentiment. The index is classified into five categories: "Extreme Fear," "Fear," "Neutral," "Greed," and "Extreme Greed."

4. Methodology

The project is implemented in a two-notebook workflow:

1. **notebook2.ipynb (Data Preprocessing and Feature Engineering)**: This notebook is responsible for cleaning and preparing the raw data. Key steps include:
 - Converting timestamp columns to a consistent datetime format.
 - Handling categorical data by converting relevant columns to the `category` dtype.
 - Engineering a set of daily aggregated features from the raw trade data, such as:
 - `net_daily_pnl`: The total daily profit or loss.
 - `pnl_volatility`: The standard deviation of daily PnL, representing risk.
 - `total_trade_volume_usd`: The total dollar value of all trades in a day.
 - `avg_trade_value_usd`: The average size of a trade in USD, as a proxy for risk appetite.
 - `active_traders`: The number of unique traders per day.
 - Saving the processed data to intermediate files (`historical_data_inter.csv` and `fear_data_inter.csv`).
2. **notebook.ipynb (Analysis, Modeling, and Visualization)**: This script loads the preprocessed data and performs the main analysis:
 - **Data Merging**: The processed historical data is merged with the Fear and Greed Index data based on the timestamp.
 - **Exploratory Data Analysis (EDA)**: The merged data is grouped by the sentiment `classification` to calculate summary statistics for each group.
 - **Predictive Modeling**: An ElasticNet regression model is trained to predict the `net_daily_pnl` based on the engineered features. ElasticNet is chosen for its ability to handle multicollinearity and perform feature selection by combining L1 and L2 regularization.
 - **Visualization**: A comprehensive dashboard is generated to visualize the key findings and compare the behavior of different trader groups.

5. Results and Discussion

5.1. Dashboard Insights

The `my_dashboard.png` in the folder `outputs/my_dashboard.png` file provides a visual summary of the key findings:

- **Profitability by Sentiment**: The analysis reveals that the highest average daily net P&L is observed during periods of "Fear." This counter-intuitive finding suggests that periods of market fear, while risky, may present the most profitable trading opportunities.
- **Risk and Volume**: Trading activity, as measured by `total_trade_volume_usd`, is highest during periods of "Fear." Similarly, `pnl_volatility` is also elevated during these times, indicating that fearful markets are both active and unpredictable.
- **Risk Appetite**: The `avg_trade_value_usd`, a proxy for risk appetite, is highest during periods of "Extreme Greed." This suggests that traders are more willing to make larger bets when the market is optimistic.
- **Top Coins**: The analysis of "top coins" shows which cryptocurrencies are most popular during different sentiment periods, providing insight into where market attention is focused.

5.2. ElasticNet Model Performance

The ElasticNet model was trained to predict `net_daily_pnl`. The model's performance was evaluated using Mean Squared Error (MSE) and R-squared (R^2). The results indicate that while the model can identify some relationships between the features and the target variable, its predictive power is limited, as evidenced by a negative R-squared value. This suggests that a linear model may not be sufficient to capture the complex, non-linear dynamics of the cryptocurrency market.

The model's coefficients, however, provide some interesting insights. For example, `avg_daily_pnl` has a large positive coefficient, indicating that past profitability is a strong predictor of future profitability.

6. Conclusion

This project successfully demonstrates a comprehensive workflow for analyzing cryptocurrency trading data and its relationship with market sentiment. The analysis reveals distinct behavioral patterns associated with different levels of fear and greed, providing valuable insights for traders and market analysts. The implementation of a predictive model, while not highly accurate, serves as a proof-of-concept for how machine learning can be applied to this domain.

Future work could explore more sophisticated non-linear models (e.g., gradient boosting, neural networks) to improve predictive accuracy. Additionally, incorporating a wider range of features, such as social media sentiment or on-chain data, could provide a more holistic view of the market and lead to more robust trading strategies.

7. How to Run the Project

To replicate the analysis, follow these steps:

1. Install Dependencies:

This project uses `uv` for environment management. The dependencies are listed in the `pyproject.toml` file. To install them, run:

```
uv pip install -e .
```

2. Run the Notebooks/Scripts:

The project consists of a data preprocessing notebook and an analysis script. They should be run in the following order:

- First, run `notebook2.ipynb` to perform data preprocessing and generate the intermediate data files. This can be done using Jupyter Lab or Jupyter Notebook.
- Then, run `notebook.ipynb` to perform the final analysis, train the model, and generate the visualization dashboard. This can be run from the command line:

```
uv run notebook.py
```

The final output will be a dashboard image named `outputs/my_dashboard.png`.