Impact of Social media sentiments on Cryptocurrency market volatility

Open Source Economics project

Matriculation Number: 3395916

September 2023

1 Introduction

Cryptocurrencies, particularly those based on blockchain technology, have garnered significant attention in recent years due to their inherent volatility and unpredictability. In this landscape, understanding the role of news sentiment in influencing cryptocurrency price movements has become a compelling area of research. Two of the most prominent cryptocurrencies, Bitcoin and Ethereum, have a combined market capitalization exceeding 160 billion US dollars, making them central subjects for investigation.

Previous research in this domain has extensively employed Natural Language Processing (NLP) techniques such as VADER, BERT, and Roberta to assess the sentiment of news and social media content. Machine learning and deep learning models, including Support Vector Machines, Random Forests, and Neural Networks (RNN), have been used to predict cryptocurrency price fluctuations based on sentiment analysis.

This research project aims to investigate the ability of sentiment labels derived from news and social media content to predict the direction of Bitcoin price changes over a more extended timeframe. Unlike many existing studies that collect data over a short time frame, this project employs web scraping techniques, specifically the Selenium package, to gather tweets from Twitter spanning from January 1, 2019, to December 31, 2021. Additionally, this study utilizes one-minute historical price data during the same period, allowing for a more granular analysis of sentiment-price relationships.

2 Analysis

2.1 Data Cleaning and Pre-processing

To assess sentiment, this project employs a pre-trained Roberta model obtained from Hugging Face, alongside softmax to obtain sentiment scores. The analysis reveals that 6821 tweets are classified as having a neutral sentiment, 1945 as negative, and 1182 as positive. Given that Twitter is often used as a platform for sharing news, which tends to have a neutral sentiment, an additional feature called "polarity" is introduced. This feature calculates the geometric mean of positive and negative sentiment scores for each tweet, providing a more nuanced sentiment representation.

2.2 Sentiment Analysis:

This paper employs a pre-trained Roberta model from Hugging Face and softmax to get the sentiment scores. From the result, 6821 tweets returned Neutral respone, 1945 tweets returned Negative sentiment and 1182 tweets returned positive sentiment. This is because most of the people use Twitter as a platform to share news which generally have a neutral sentiment. Therefore, we also created an additional feature called polarity that calculates the geometric mean of positive and negative sentiment scores of each tweets.

2.3 Impact of News Sentiment in Cryptocurrency market Volatility:

To explore the influence of news sentiment on cryptocurrency price movements, this study employs a logistic regression model. Initially, a binary target variable called "Price Direction" is created based on whether returns calculated from closing prices are negative (indicating a price decline) or positive (indicating a price increase). Other price-related features such as open, low, and high prices are omitted to mitigate multicollinearity issues.

The initial logistic model demonstrates that the "polarity" feature plays a significant role in predicting the direction of price movement, with a coefficient parameter of 97%. However, the model's accuracy score stands at 51%. In an effort to refine the model's performance, GridSearchCV is applied for fine-tuning, but it yields no

improvement in accuracy. This suggests the need to explore alternative models to comprehensively assess the impact of news sentiment on cryptocurrency prices.

3 Future Research idea

Data Cleaning and Pre-processing: Refine the keyword list and optimize pre-processing techniques to further reduce data noise.

Sentiment Analysis: Explore alternative sentiment analysis models and investigate the use of sentiment time series analysis to capture evolving sentiment trends. Especially employ semantic based sentiment analysis.

Model Selection: Assess the performance of various machine learning and deep learning models beyond logistic regression, such as recurrent neural networks (RNNs), convolutional neural networks (CNNs), and transformer-based models, to determine which best predicts cryptocurrency price movements based on sentiment.

Evaluation and Interpretation: Evaluate model performance using relevant metrics and interpret results to gain insights into the relationship between news sentiment and cryptocurrency price movements.

Discussion and Conclusion: Summarize findings, discuss their implications, and suggest potential avenues for further research.