

Bitcoin sentiment analysis

Ambroise Thibault, Faune Blanchard, Maxime Lorenzo

Table of contents

| | | |
|----------|--|----------|
| 1 | Introduction | 1 |
| 2 | Generating sentiment score | 1 |
| 2.1 | Presentation of the database | 1 |
| 2.2 | Generating prompts | 1 |
| 2.3 | Scores methodology | 2 |
| 3 | Comparison with BTC returns | 2 |
| 3.1 | Extracting BTC returns | 2 |
| 3.2 | Comparing with scores | 2 |
| 3.2.1 | Initial results | 2 |
| 3.2.2 | Problems with our model | 2 |
| 3.2.3 | Interpretation | 2 |

1 Introduction

Write with chat

2 Generating sentiment score

2.1 Presentation of the database

We collected a dataset from GitHub entitled [Crypto News Dataset](#). It aggregates daily headlines from a variety of crypto news websites. Each headline is labeled with the cryptocurrencies to which it refers. They are also timestamped down to the second the news was posted on the site. We also have the number of likes, dislikes, comments and others, although these are often 0, as well as the link to the article. In this project, we will only be using the raw headlines in order to generate the scores that we need.

Here is a quick view of the pre processed dataset with our function “import_data” in the “prompts” module. This function reads the csv files from the GitHub repository linked above, then it removes unwanted columns and formats the date so that we only have the day at which it was posted. We can see that there can be multiple headlines in one day, or no headlines. This will cause issues that we will explain later on in the project.

The function also writes the prompts that we will later feed to the Large Language Model in order to properly generate the scores, which we will explain in the following section.

2.2 Generating prompts

For each headline, we generate a prompt which asks the AI its beliefs about the evolution of Bitcoin returns based on this headline. The prompt we decided to use is the exact prompt presented in Bybee [1]. The exact prompt reads :

```
Here is a piece of news:  
"%s"
```

```
Do you think this news will increase or decrease BTC?  
Write your answer as:
```

```
{increase/decrease/uncertain}:  
{confidence (0-1)}:  
{magnitude of increase/decrease (0-1)}:  
{explanation (less than 25 words)}
```

Where “%s” is replaced by the headline. After editing this prompt, we notice that removing part of this prompt changes the results drastically (from “increase” to “decrease” for the same headline). Therefore, we decided to keep the prompt from the paper as is, even though we will only use the results for “increase” or “decrease” for this project. This prompt has been adapted from traditional retail investor surveys or surveys of CFOs that are used to produce a sentiment indicator. However, these are costly and slow. As presented in the paper, using AI models to replace human answers mimics the same results on a much bigger and faster scale. We will also try to edit this prompt by adding context: since Bitcoin has mostly been popular among retail investors, we can add a simple sentence (e.g. “You are a retail investor”) at the beginning of this prompt, to see if this yields better results. We will conclude on this in the end of the project.

More information about the construction on this prompt can be found in Bybee [1] .

2.3 Scores methodology

3 Comparison with BTC returns

3.1 Extracting BTC returns

3.2 Comparing with scores

3.2.1 Initial results

3.2.2 Problems with our model

3.2.3 Interpretation

Conclusion

[1] J. Leland Bybee. *The Ghost in the Machine: Generating Beliefs with Large Language Models*. 2025.