OCEAN TOKEN SENTIMENT ANALYSIS CHALLENGE

@descentralízate

Table of Contents

1. **Challenge Description**

The goal of this challenge centers around the development of data analysis reports and machine learning models aimed at accurately assessing investor sentiment regarding the OCEAN token.

In between all different social networks Twitter has turn out to be an invaluable repository of insights and viewpoints on cryptocurrencies. Through tweets, investors articulate their perspectives, analyses, emotions and they generate a constant stream of market-sentiment data.

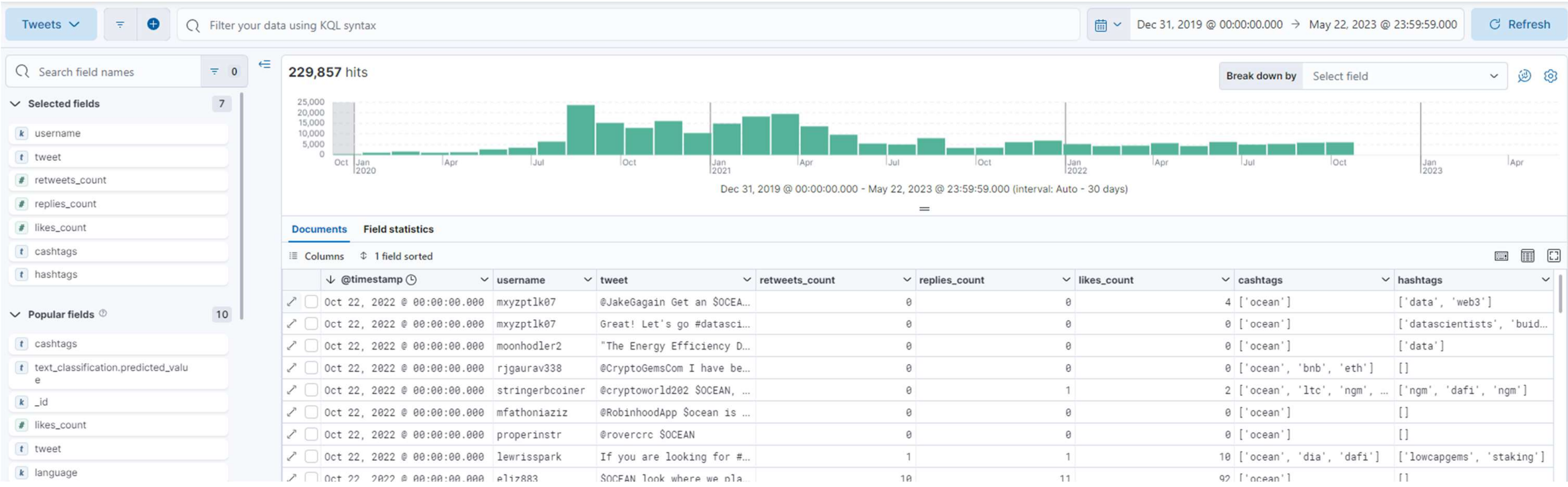
By tracking and meticulously analyzing this data, we can uncover noteworthy trends and fluctuations in sentiment that wield considerable influence over the token's value and appeal.

This analysis assumes a critical role in empowering investors to astutely manage their investments and make informed decisions in light of market volatility.

We have opted to publish this paper in A3 format to enhance the readability of the graphs included.

2. **Data Preparation**

With given CSV dataset , we have imported them into ELASTIC into different indexes, an index for the tweets CSV and another index for the price CSV.



Once all fields were supervised, we utilized Kibana graphs for showing graphs that could answer the questions of the challenge.

After supervision we considered we had to do a little data scrubbing, we took off tweets containing other token cashtags, because they were not related to OCEAN sentiment. There were some tweets in different languages, but they were not interfering data quality, so at this point we decided data was ready to make our exercise.

We created individual graphs in order to analyze every question separately. Doing it like this we had a better view of each metric, to let us distinguish which metrics were more relevant to estimate different events.

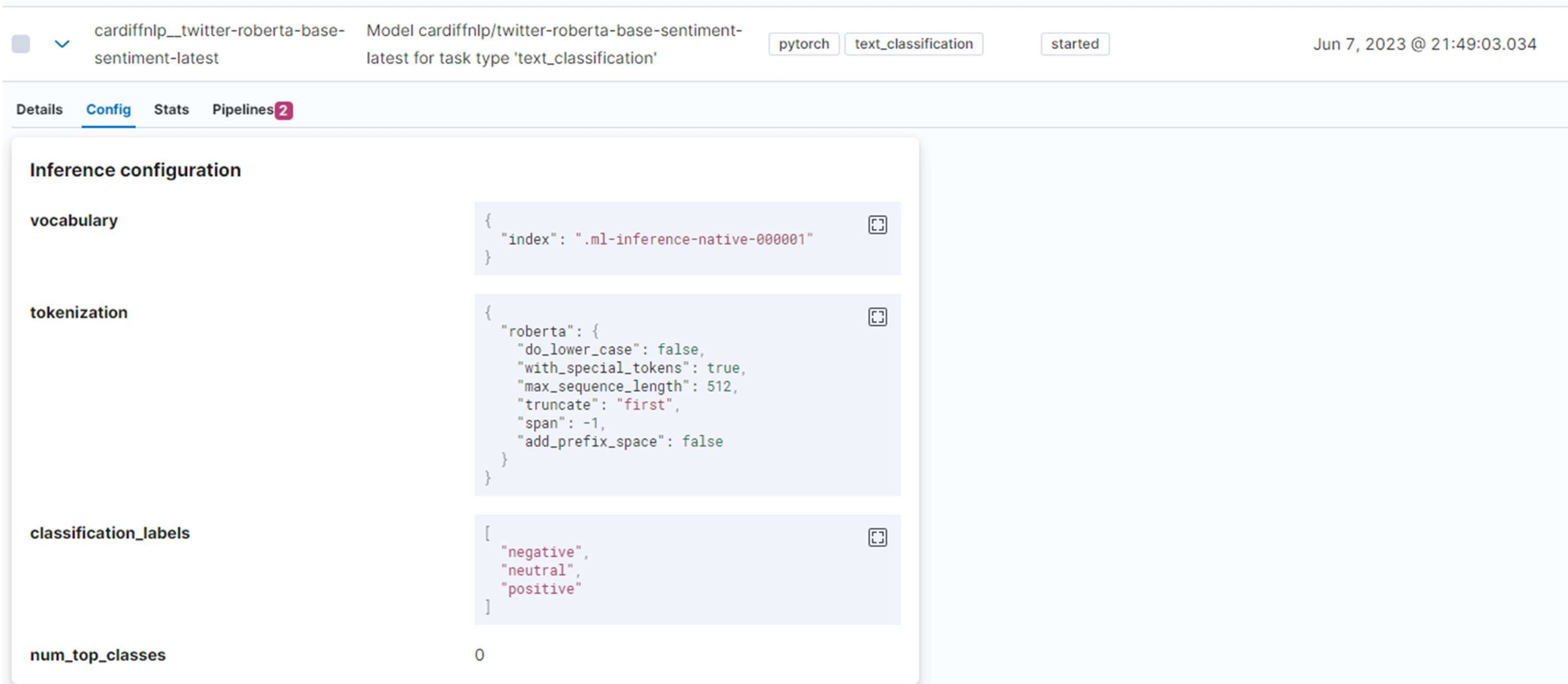
This was the code used to plot our different metrics:

1. #Tweets vs Price: moving_average(count(kql='id : *') / median('Adj Close'), window=18)
2. #Likes vs Price: moving_average(sum(likes_count) / median('Adj Close'), window=18)
3. #Retweets vs Price: moving_average(sum(retweets_count) / median('Adj Close'), window=18)
4. #Unique Users vs Price: moving_average(unique_count(user_id)/ median('Adj Close'), window=7)
5. #Influential Tweets vs Price: moving average(count(kql='likes_count > 350 and retweets_count > 50 and replies_count > 20') / median('Adj Close'), window=18)

3. **Sentiment Data Prediction**

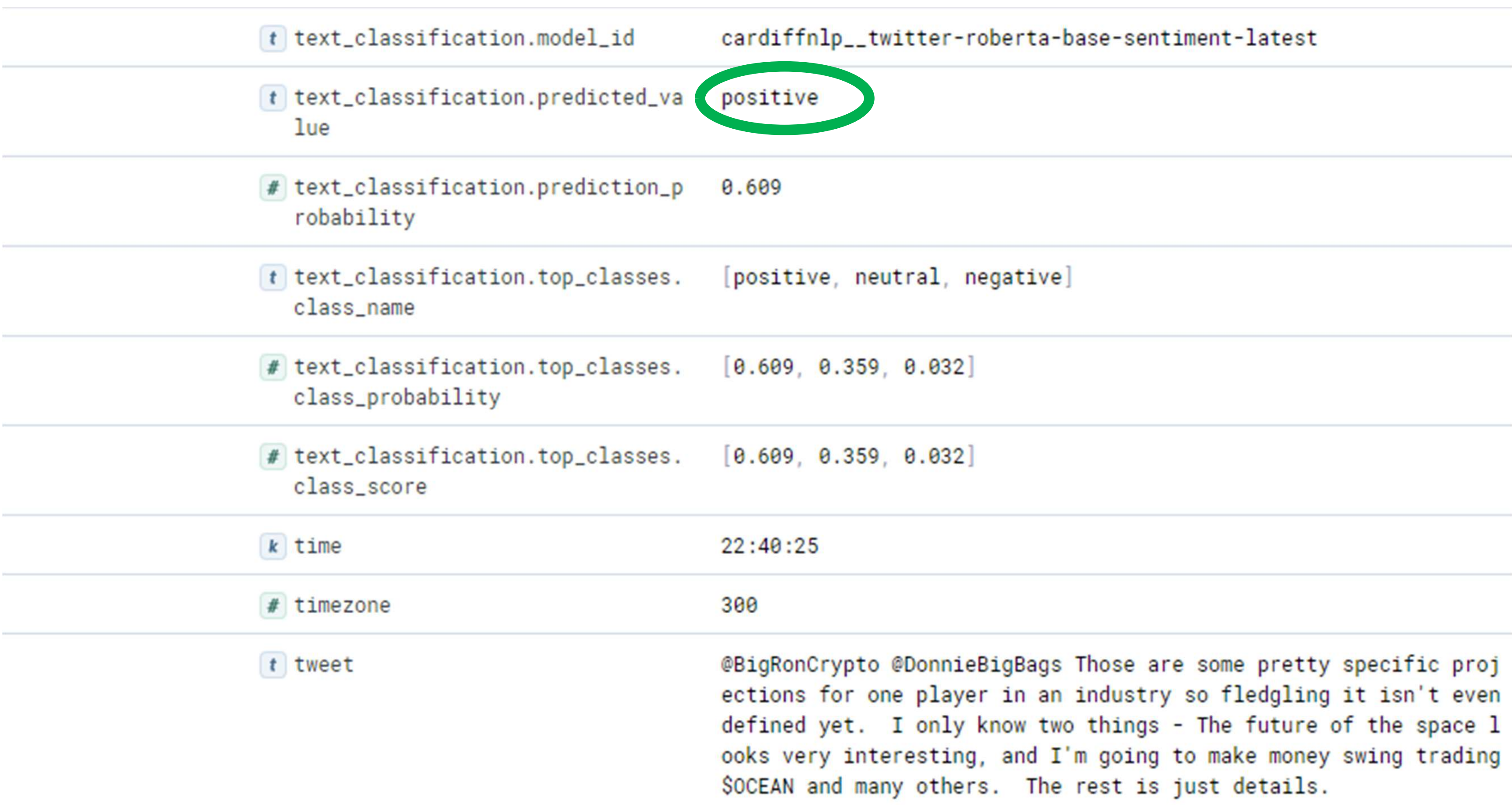
In order to make our prediction model we have deployed a machine learning model capable of classifying tweets sentiment. We employed a RoBERTa-base model that underwent training on approximately 124 million tweets spanning from January 2018 to December 2021. This model was further fine-tuned using the TweetEval benchmark to specialize in sentiment analysis. We used ELASTIC machine learning tools to import this RoBERTa based model.

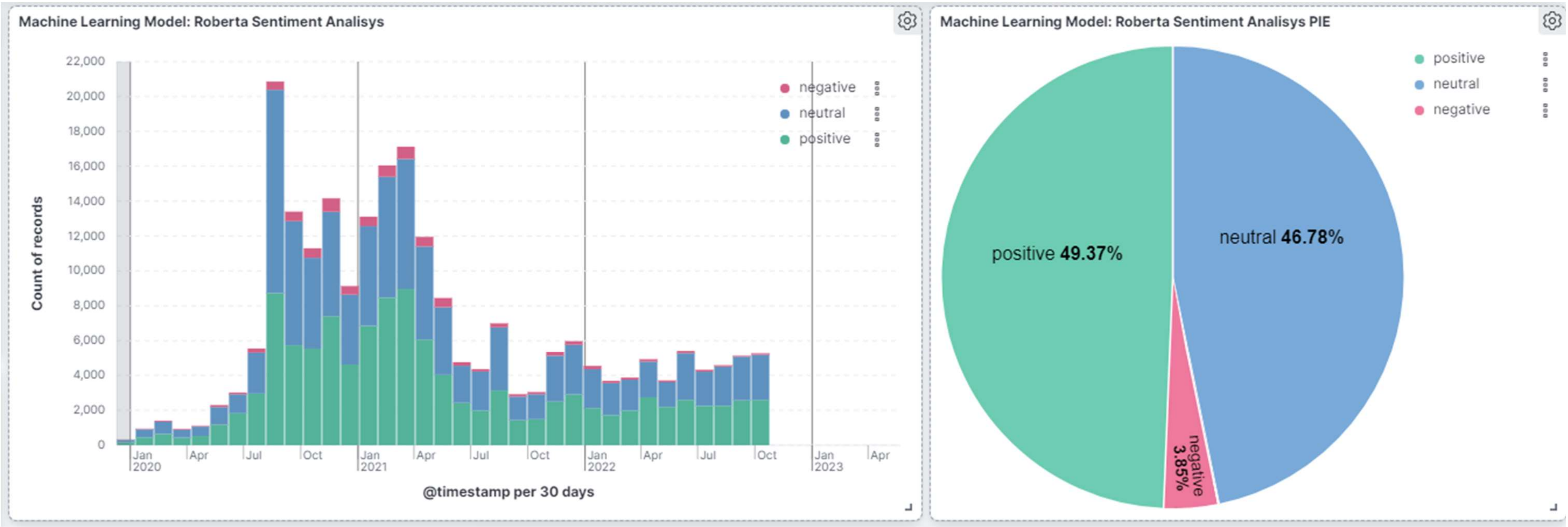
The model is designed to categorize sentiment into three distinct clusters: Bullish (positive), Neutral, and Bearish (negative).



The model exhibits impressive performance, effectively capturing intricate language patterns and comprehending contextual connections. Its contextual awareness proves advantageous for sentiment analysis tasks, as sentiment can often be influenced by the broader context. Although the model was trained in English, the non-English posts were perfectly classified.

This is how results look once they were process, showing for every tweet different percentage of sentiment, then the highest value is classified as Predicted Sentiment.





The image displayed above illustrates the distribution of sentiment categorization. After conducting a supervised analysis, we made the decision to filter out neutral sentiment. This filtering allowed us to focus solely on comparing bullish sentiment against bearish sentiment, enabling a clearer understanding of market sentiment.

This is the code we used to get the sentiment drawn in our chart:

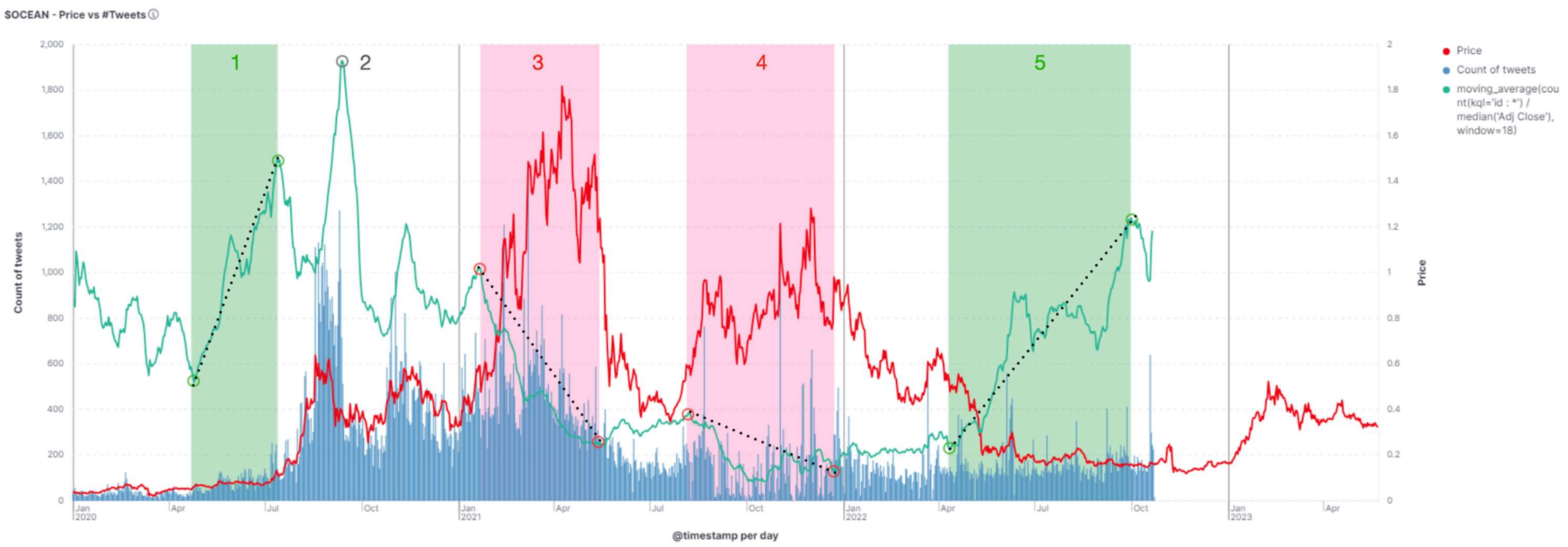
moving_average(count(kql='text_classification.predicted_value.keyword : "positive" ')/count(kql='text_classification.predicted_value.keyword : "negative" '), window=14)

4. Data Analysis

In this study, we will examine five distinct sentiment metrics derived from Twitter data and explore their correlations with price action in order to gain insights into market dynamics.

4.1. Correlation between the price of $OCEAN and the number of tweets containing "$OCEAN".

The correlation analysis offers valuable insights into the relationship between the price of $OCEAN and the corresponding tweet volume. By observing the three-month continuous divergence of the green indicator (tweet volume) and price, investors can anticipate potential changes in the price trend. It has been observed that when the green line diverges from the price while trending upwards, it signifies a favorable time to buy. Conversely, if the green line diverges while trending downwards, it indicates a suitable time to sell.

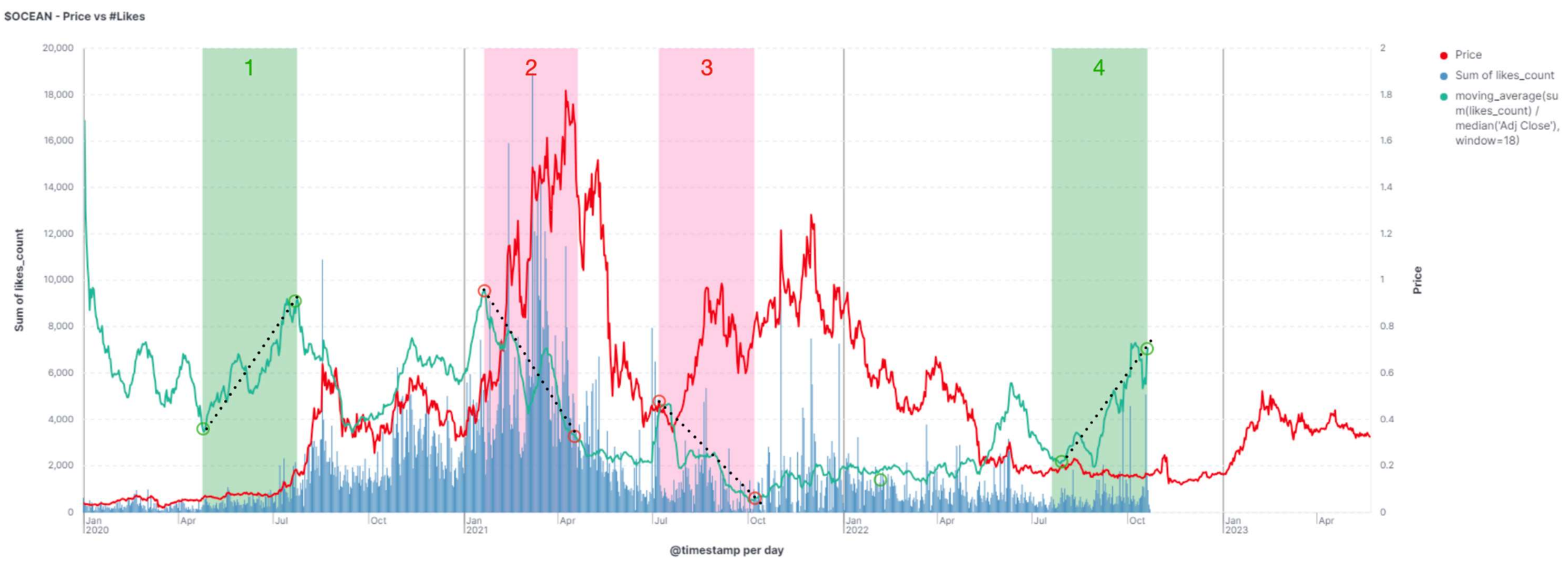


1. 3 month divergence: Despite a significant increase in new tweet activity related to price, the price action has not yet responded accordingly. Buying opportunity.
2. Disregard this peak is noise it is an anomaly surged due to the 18-day moving average plus sudden stop and huge sudden start over Twitter activity plus the effect of the substantial increase in price, 100x in less than 6 months for some investors, as result then a minor 2x correction drop has occurred.
3. 4 month divergence warning: The price is experiencing an excessive surge compared to Twitter activity. Selling opportunity.
4. 4 month divergence warning: Be cautious of the disparity between Twitter activity and price. Selling opportunity.
5. + 6 month divergence: Twitter activity shows a pump in relation to the price, but the price has not yet reflected this increase. Buying opportunity.

By understanding and leveraging this correlation between $OCEAN price and tweet volume, investors can make more informed decisions regarding their investment strategies.

4.2 Correlation between the price of $OCEAN and the number of likes received by tweets containing "$OCEAN".

This correlation provides noteworthy findings regarding the interplay between the price of $OCEAN and the number of tweets. In this case we can also observe a three month continuous divergence of the green indicator vs price will anticipate a change in price trend, if it diverges going up it is favorable time to buy, while a decreasing correlation indicates a good time to sell. This correlation indicator alerts when it diverges from price action it accurately represents market optimism or pessimism towards a change of trend in the price action.



1. 3 month divergence: Despite a significant increase in new tweet activity related to price, the price action has not yet responded accordingly. Buying opportunity.
2. 3 month divergence warning: The price is experiencing an excessive surge compared to Twitter activity. Selling opportunity.
3. 3 month divergence warning: Be cautious of the disparity between Twitter activity and price. Selling opportunity.
4. 3 month divergence: Twitter activity shows a pump in relation to the price, but the price has not yet reflected this increase. Buying opportunity.

This metric draws similar conclusions to the previous one, where observed divergences over a three-month period indicate a potential impending risk of a change in price action. The correlation between the variables is clear.

4.3 Correlation between the price of $OCEAN and the amount of retweets containing "$OCEAN".

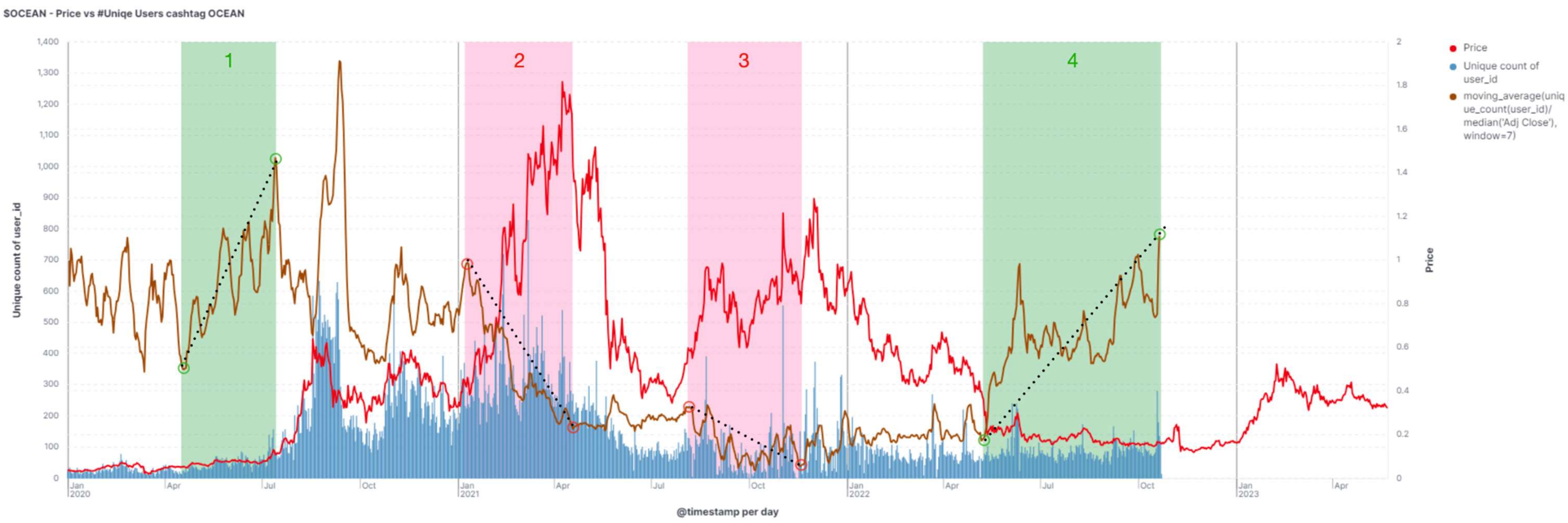
Shown in green, we can observe correlation between the total number of retweets and the price of $OCEAN. This correlation reflects the impact of campaigns and signifies a similar sentiment regarding investment terms as the previous analysis. If the price increases while the campaign force remains stable or decreases, it suggests that the campaign is unable to drive the price further, potentially indicating a signal to sell.

1. 3 month divergence warning: The price is experiencing an excessive surge compared to Twitter activity. Selling opportunity.
2. 3 month divergence: Twitter activity increases in relation to the pric, but the price has not yet reflected this. Buying opportunity.

This indicator is a great achievement as it is another confirmation on previous sentiment conclusions, divergences going along during 3 months indicate an imminent risk that price action could change. It gives good insights about community engagement.

4.4 Correlation between the price of $OCEAN and the number of individuals tweeting with the cashtag "$OCEAN".

This correlation reveals similar analysis results of cases 4.1 and 4.2.



1. 3 month divergence: Despite a significant increase in new tweet activity related to price, the price action has not yet responded accordingly. Buying opportunity.
2. 3 month divergence warning: The price is experiencing an excessive surge compared to Twitter activity. Selling opportunity.
3. 3 month divergence warning: Be cautious of the disparity between Twitter activity and price. Selling opportunity.
4. + 5 month divergence: Twitter activity shows a pump in relation to the price, but the price has not yet reflected this increase. Buying opportunity.

The accuracy of this indicator allows for a proactive approach to capitalize on favorable buying or selling opportunities, aligning investment decisions with market sentiments.

4.5 Impact of influential tweets on the price of the OCEAN token.

This is the correlation we have found between the number of influential tweets and the price of $OCEAN. We have considered “influential tweet” as a specific set of criteria: Tweets with over 50 retweets, 350 likes and 20 replies.

Here we can observe this correlation normally follows price dynamics.

1. The most evident observation in this analysis is a significant rise in impactful tweets, particularly noticeable from Q3 2022 onwards, which is distinct due to its lack of reflection in the price dynamics. Buying opportunity.

We consider this indicator to be an intriguing achievement in evaluating community engagement concerning potential community expansion event. However, it may translate into a pump in price but it is not a reliable indicator for forecasting price declines.

Some of the most influential individuals among the entire population of tweets, many of these are big accounts that attract new investors.

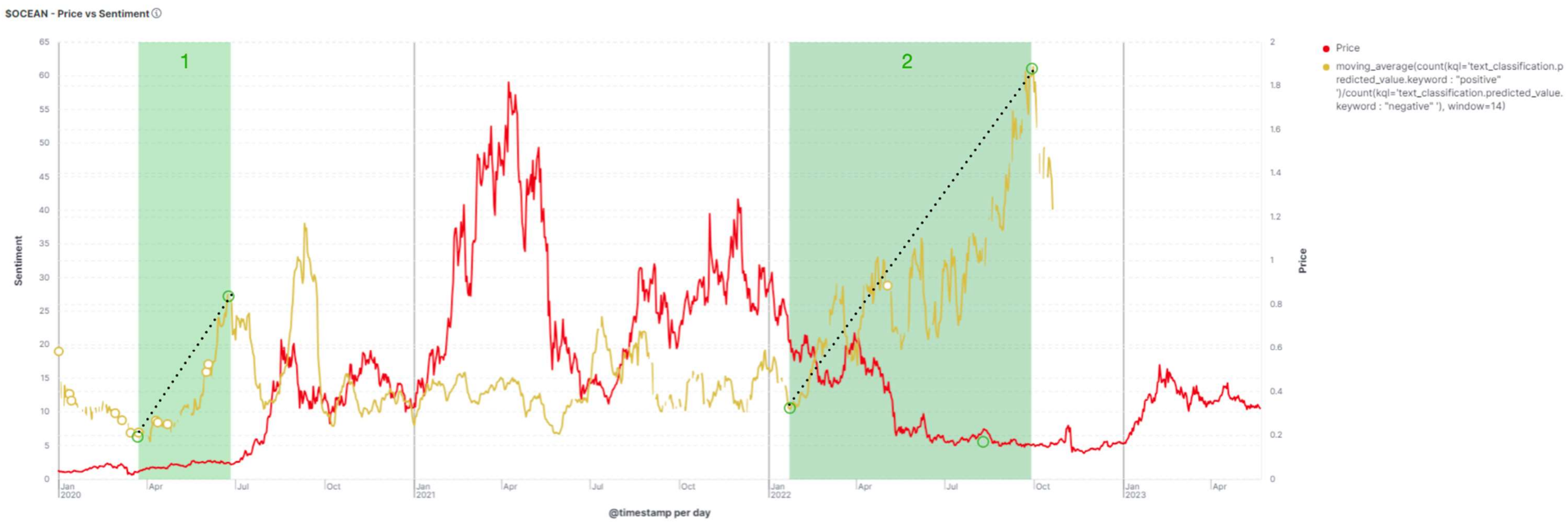
Performance of most influential accounts:

Top 15 values of username	Count of records	Median of likes_count	Median of retweets_count	Median of replies_count
cryptowizardd	61	589	78	58
donniebigbags	13	450	91	33
cryptogodjohn	12	816	81.5	62.5
cryptotroia	11	497	165	44
oceanprotocol	11	616	184	47
bitcointrapper	6	601	100.5	58
cryptodiffer	5	469	136	61
oddgems	4	413.5	62	72.5
scoinaldo	4	847	98.5	128
thechainmedia	4	780	129	247.5
binance	3	1,311	595	246
coin98analytics	3	458	152	43
krakenfx	3	430	100	65
middlechildpabk	3	663	106	63
petabytecapital	3	614	123	134
Other	81	571	113	84

5. **Prediction Model:**

In this study, we have deployed a machine learning model capable of classifying sentiments expressed in tweets as bullish, bearish, or neutral, thereby facilitating prediction in the OCEAN token market.

Notably, the sentiment analysis in yellow, unlike previous graphs, is not directly related to price action. This distinction enhances the value of market sentiment data as a separate indicator.



1. 3 month divergence: Despite a significant increase of bullish sentiment price action has not yet responded accordingly. Buying opportunity.
2. +8 month divergence: Twitter sentiment is at all time high, but the price has not yet reflected this increase. Buying opportunity.

This metric holds great importance as an increase in sentiment hype tends to translate into investment allocation.

However, it is clear that a significant increase in positive sentiment often foreshadows a surge in OCEAN price movement, reaffirming the deductions made in the previous analyses and offering an interesting point of view about market potential.

Of course, there are various other factors that exert influence on the market price of OCEAN, including market capitalization, trading volume, bitcoin dominance, regulatory developments, partnerships, integrations, and technology updates. Hence, conducting a thorough analysis necessitates considering multiple factors and indicators to obtain a comprehensive understanding of an altcoin's price chart.

6. **Conclusion**

This study employed diverse correlations and analyses to determine the market sentiment of OCEAN using data from Twitter feeds. To ensure a high level of accuracy, we carefully examined various metrics, including the number of tweets, likes, retweets, influential tweets and a market sentiment prediction model.

This process yielded valuable insights into investment behavior and market sentiment. The correlation between the different parameters of study revealed that could anticipate market behavior. We can conclude in all different cases a significant sentiment vs price divergence will normally indicate the change of market trend.

Overall, these findings suggest the significance of social media metrics and sentiment analysis in understanding market behavior and making informed investment decisions. However, it is important to note that further analysis and consideration of additional factors are necessary to obtain a comprehensive understanding of the altcoin's price chart and market dynamics.