News Sentiment and Cryptocurrency Trading How to Use Sentiment Analysis to Improve Investment Decisions

TL; DR

After collecting more than 28k news articles dating back to 2018, I utilized a neural nets to evaluate the sentiment of each article. Through this analysis, I discovered a correlation between the sentiment of the articles and the price of Bitcoin, suggesting that news sentiment may be an important factor in understanding and predicting changes in the cryptocurrency market.

Introduction:

My interest in the cryptocurrency market began during the 2017 bull run and over all these years I've seen a lot, and news has been an integral part of my immersion into the world of cryptocurrencies. In 2022, I developed an interest in data analytics and decided to combine my passion for crypto with my newfound skills to conduct a research project.

Main question of this research was:

***Is there a relationship between the sentiment of news articles and the price of BTC?***

Others questions:

* Can news sentiment serve as a useful indicator for making informed investment decisions?
* Among the media sources included in the study, which ones were found to have a more favorable view of cryptocurrencies?
* Which topics were the most frequently discussed or popular?

Methodology:

For data collection I’ve used news aggregation site <https://cryptonews.net>, so at the end I had 28061 article from 103 website. I’ve collected data from it using Python, requests and BeatifulSoup library, saved it to a SQL database, using sqlite3 library, which allows to make SQL queries inside a Python script.

I used two neural nets through <https://huggingface.co> API, to evaluate articles:

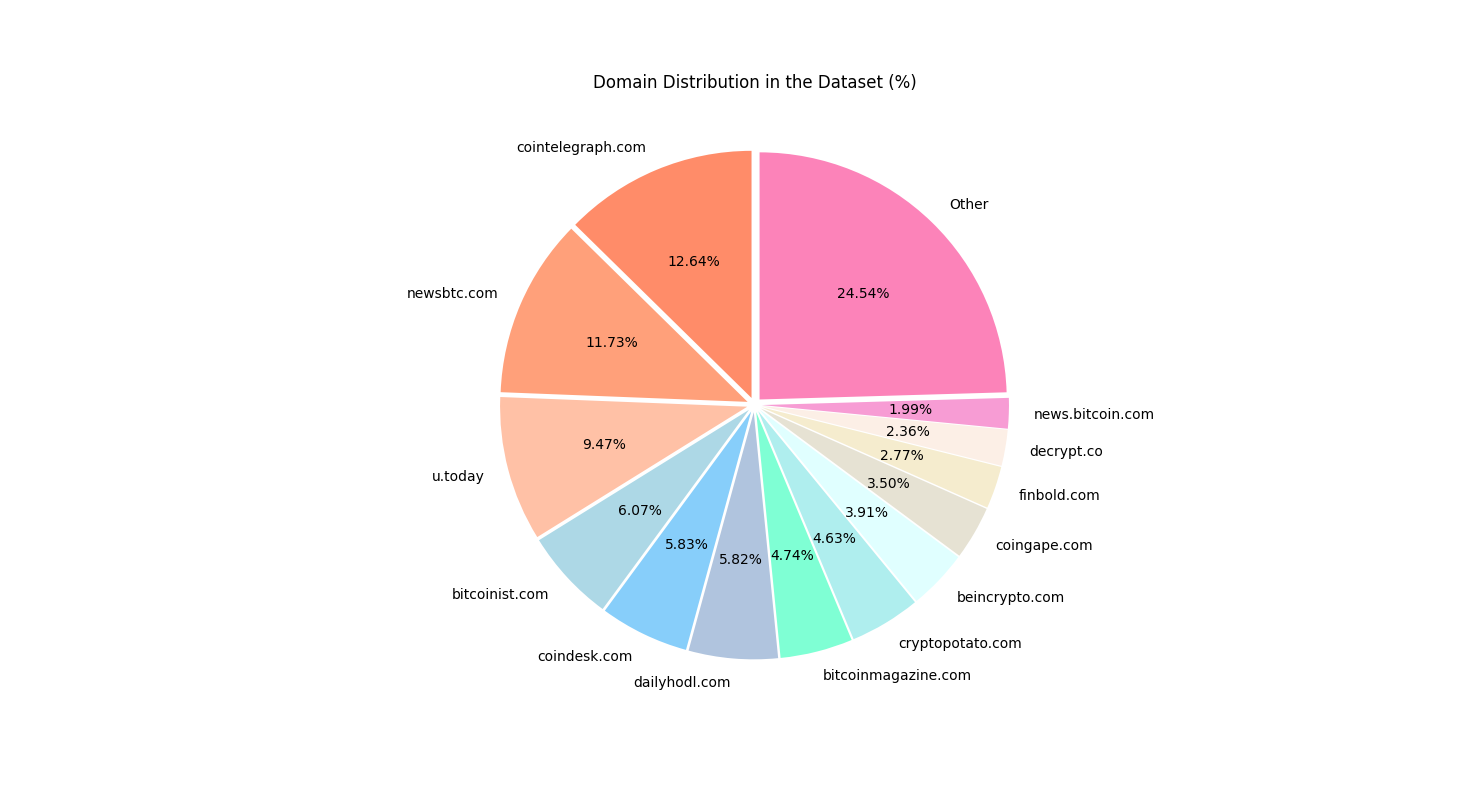
1. <https://huggingface.co/cardiffnlp/twitter-roberta-base-sentiment-latest>
2. <https://huggingface.co/ProsusAI/finbert>

Some of the articles are longer than the maximum input length of 512 tokens (≈2000 symbols), in which case I split these articles into equal parts and then calculated the average sentiment score for each part. This approach allowed me to account for the sentiment expressed throughout the entire article, while still being able to process the data within the constraints of the neural network.

Description of the dataset

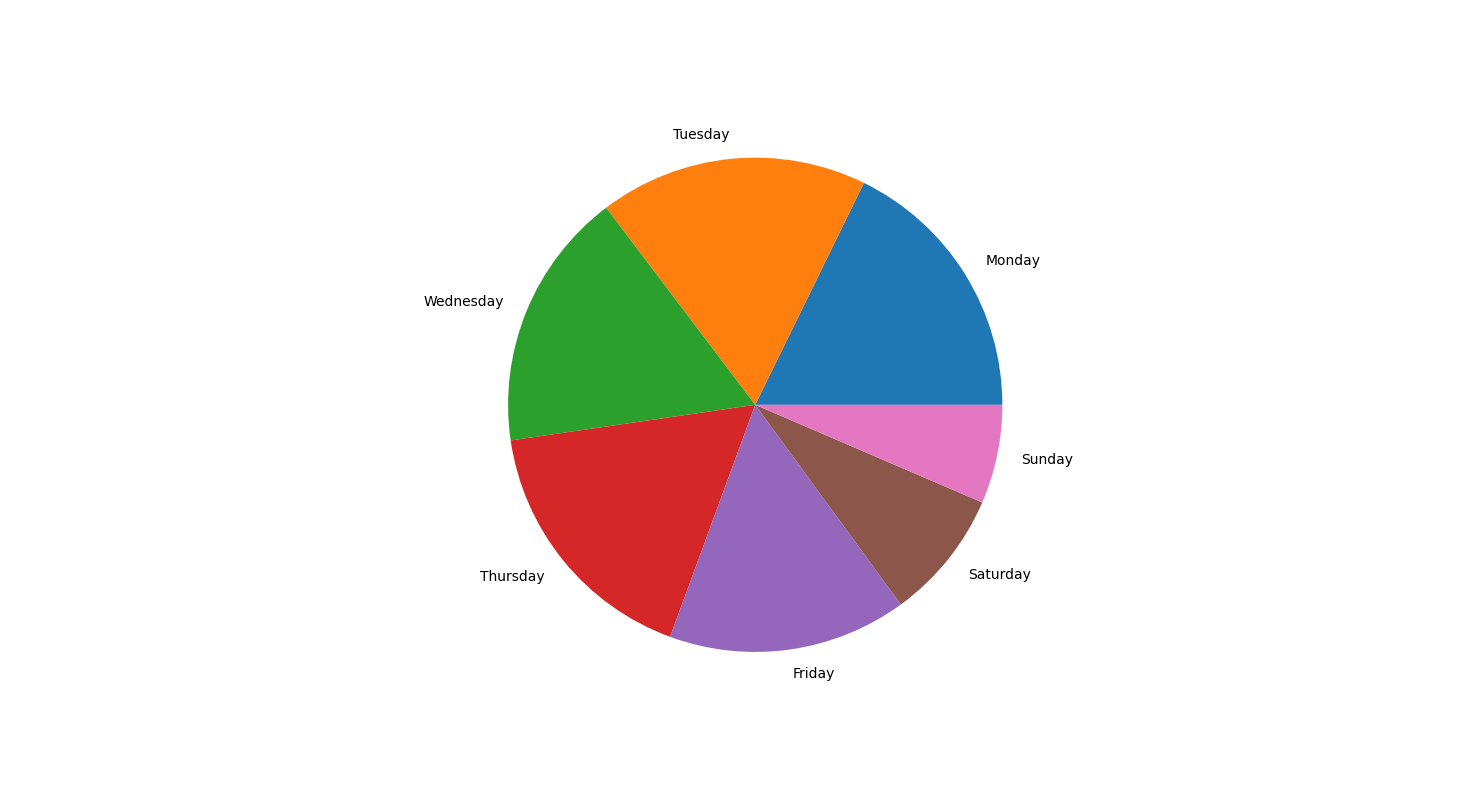
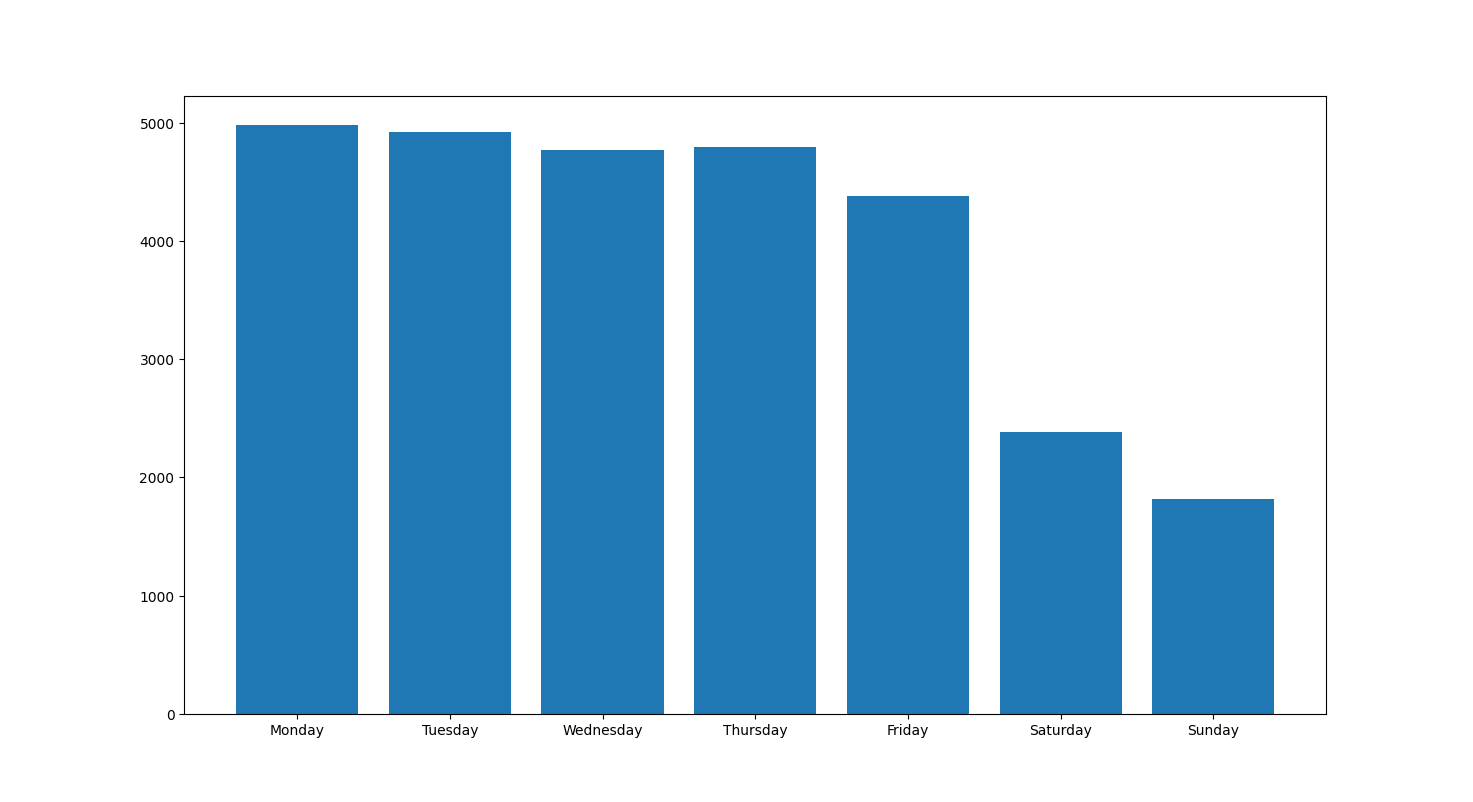
The raw dataset includes information about articles, such as their unique ID used in cryptonews.net, the title of the article, the domain which leads to the news resource, links to the article, the timestamp when the article was posted, the text of the article, a link to an image used in the article, the time required to read the article, and a tag indicating whether the article is related to Bitcoin.

Analysis of the dataset



The top 13 domains make up 75% of the dataset, while the remaining 90 domains only account for 25%.

Distribution by the days of the week



On Mondays, there are more articles published compared to other days of the week, while the number of articles published on Sundays is generally lower. Saturdays and Sundays have approximately half the number of articles posted than on other weekdays.



The words that appear most frequently in the titles are represented by larger font size, indicating their higher frequency of occurrence.

Description of the neural networks used

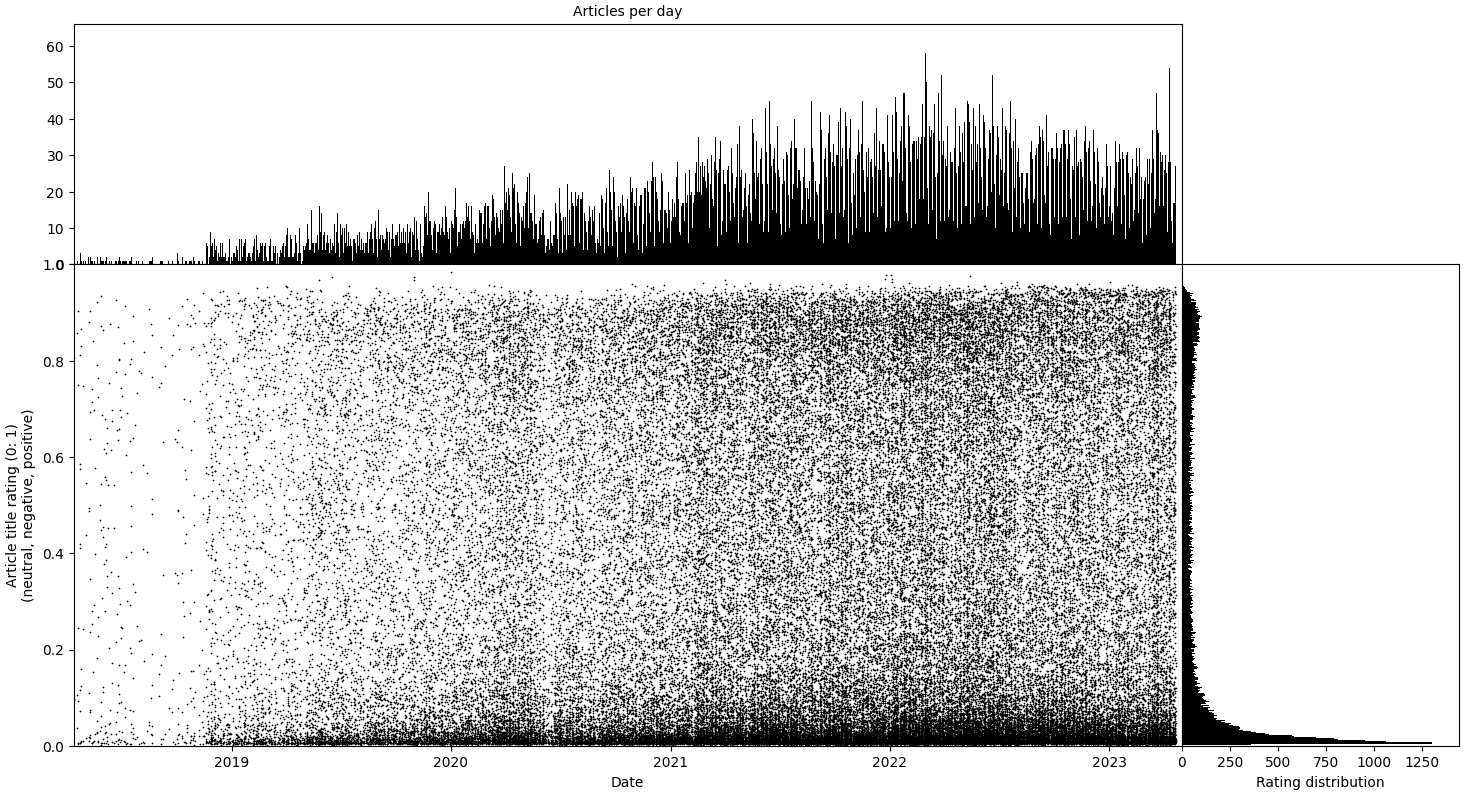
The neural network takes text as input and generates three outputs: a neutral, negative, and positive sentiment rating of the text. These values are represented as floating-point numbers within the range of 0 to 1, and can be interpreted as the network's confidence in assigning the provided text to each category.

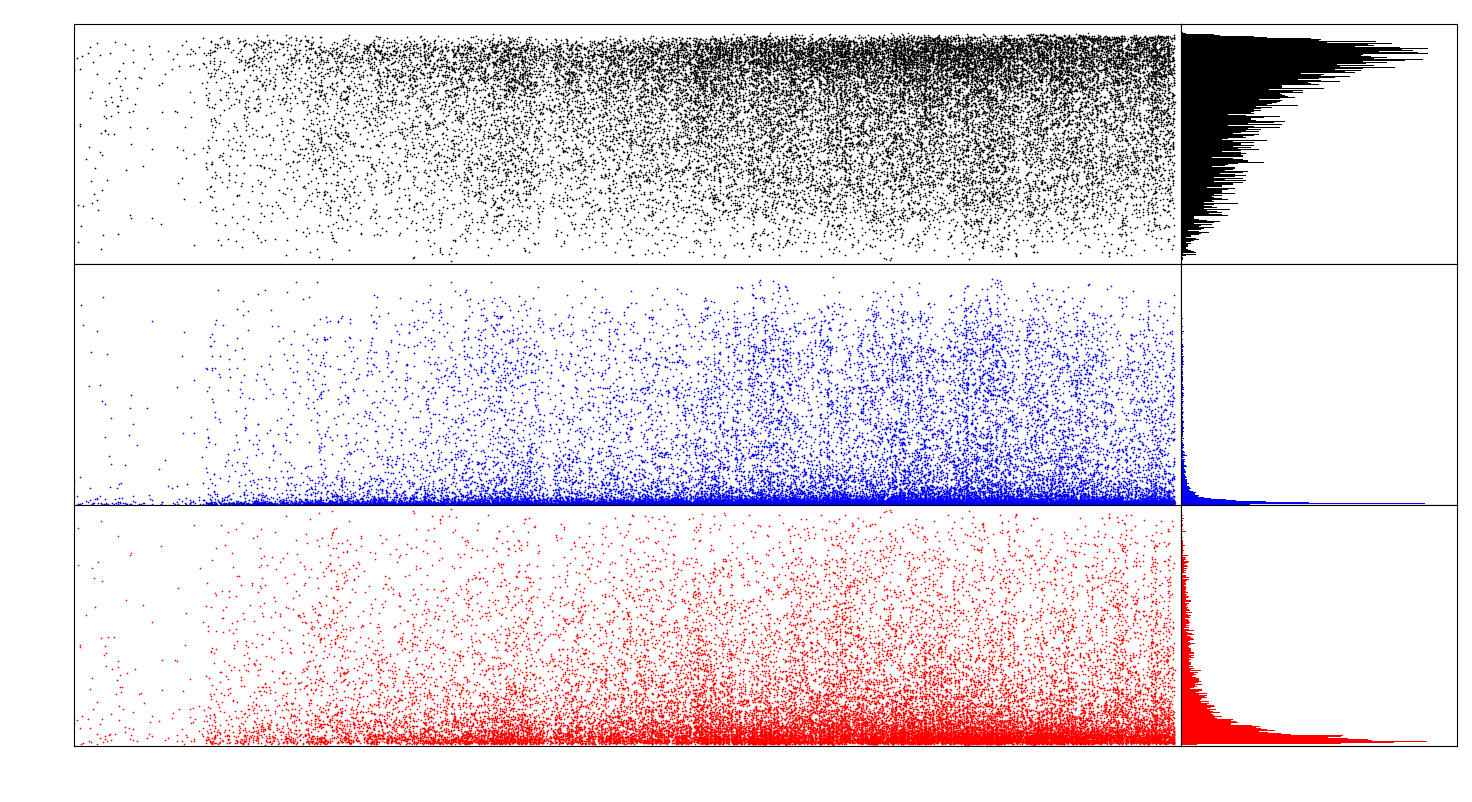
If the sentiment analysis model outputs the numbers {neutral: 0.450, negative: 0.527, positive: 0.024} for a given input text, it suggests that the text is more likely to have a negative sentiment, as the negative value is higher than the others.

Twitter-roBERTa-base  
This is a RoBERTa-base model trained on ~124M tweets from January 2018 to December 2021, and finetuned for sentiment analysis with the TweetEval benchmark.

Article title rating (neutral, negative, positive).

On X axis – date of the publication, on Y axis – rating. Graph above – distribution of publications by date, graph on the right – distribution by sentiment rating.

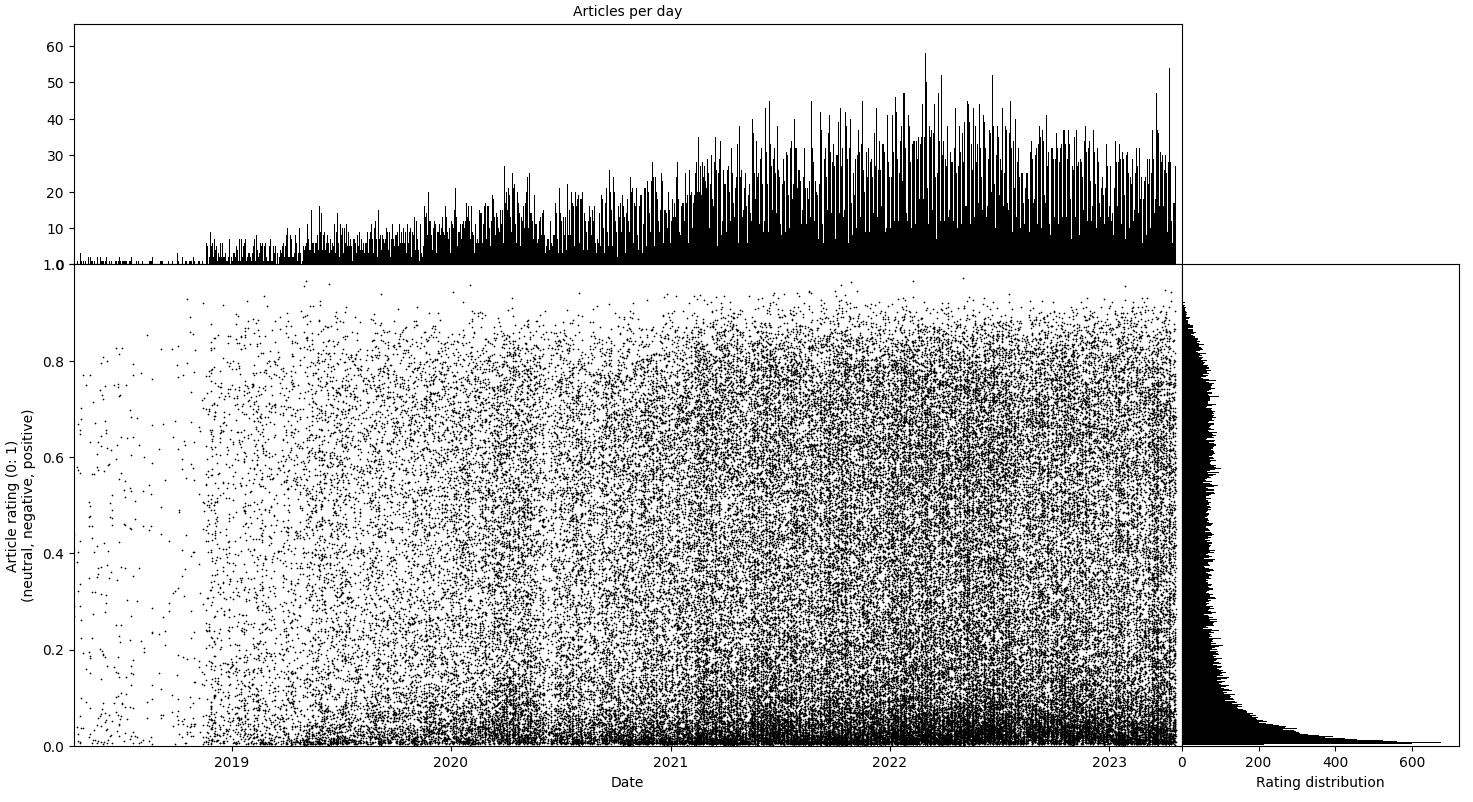




Black – neutral rating distribution, blue – negative rating distribution, red – positive rating distribution.

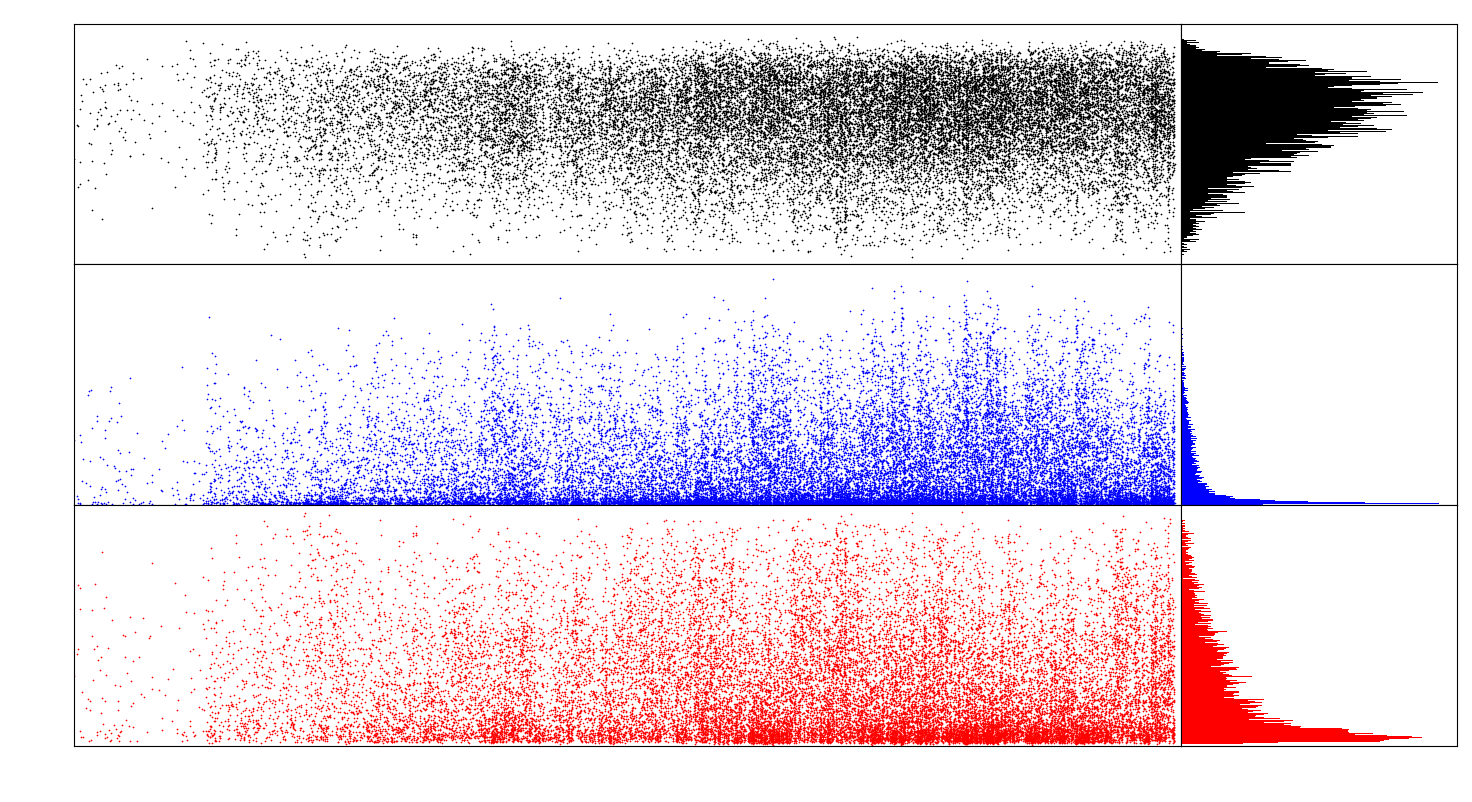
It is noticeable from the graphs that the roBERTa model tends to assign a neutral rating to a significant portion of the article titles, while the negative and positive ratings are relatively low.

Article text rating (neutral, negative, positive).



On X axis – date of the publication, on Y axis – rating. Graph above – distribution of publications by date, graph on the right – distribution by sentiment rating.

Black – neutral rating distribution, blue – negative rating distribution, red – positive rating distribution.

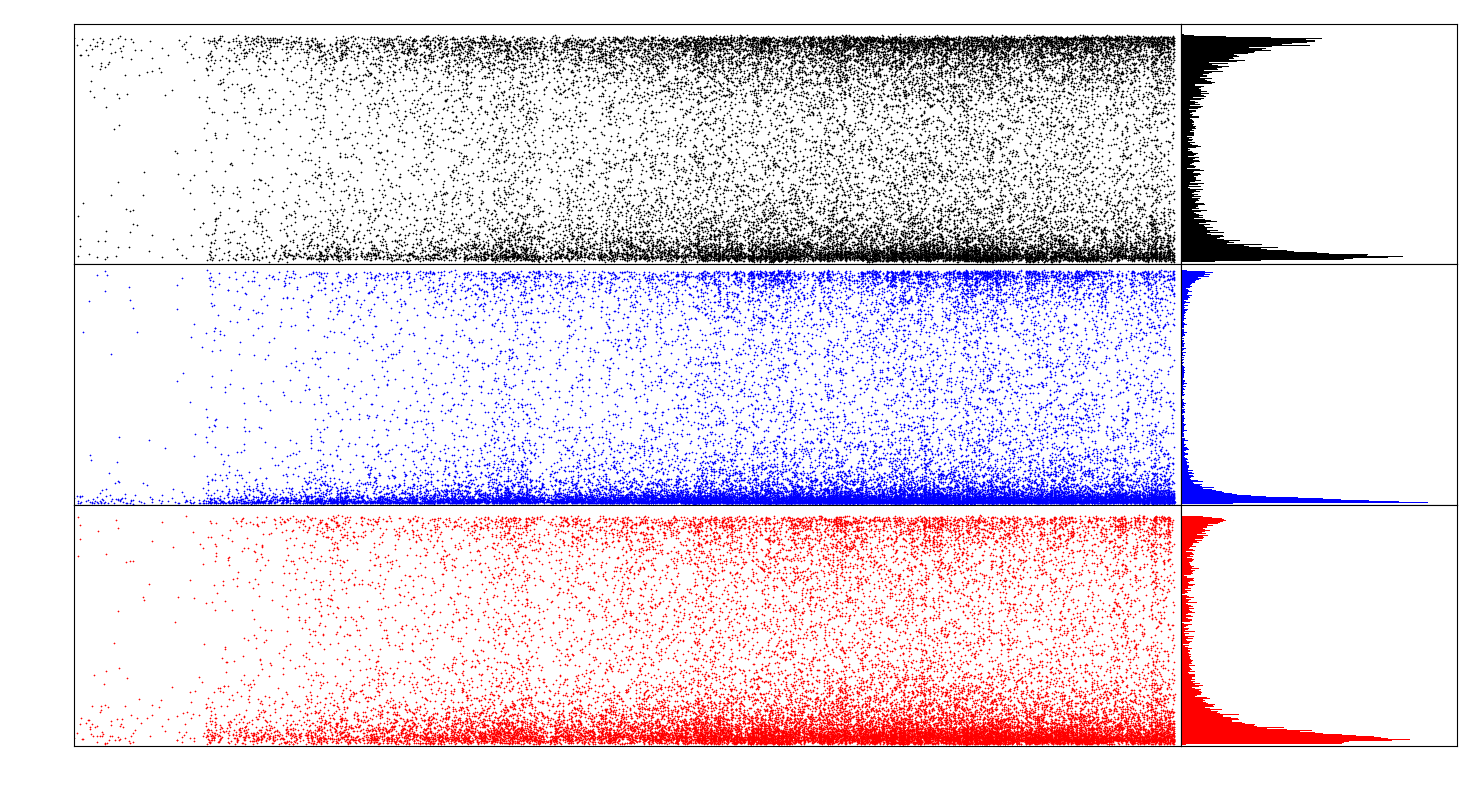


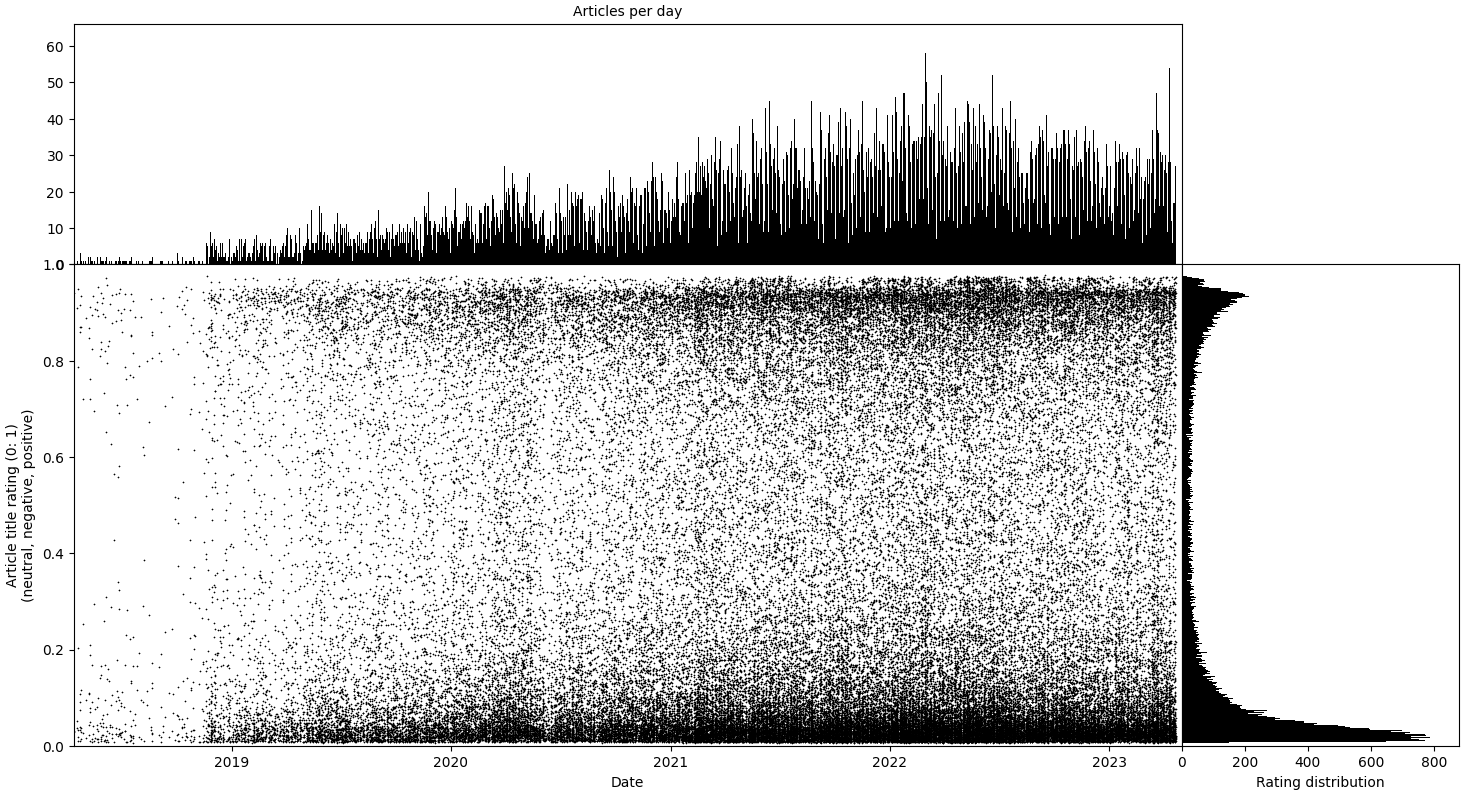
The distribution of text ratings shows a similar pattern, where neutral rating is assigned more frequently compared to positive and negative ratings, and positive rating is more frequent than negative rating.

One limitation of this neural network is that the sentiment analysis model was trained on tweets, which may have a different writing style and language compared to news articles. Therefore, the results obtained may not be entirely accurate. Moreover, the model has its own limitations and biases, which can affect the results.

FinBERT

Article title rating (neutral, negative, positive).

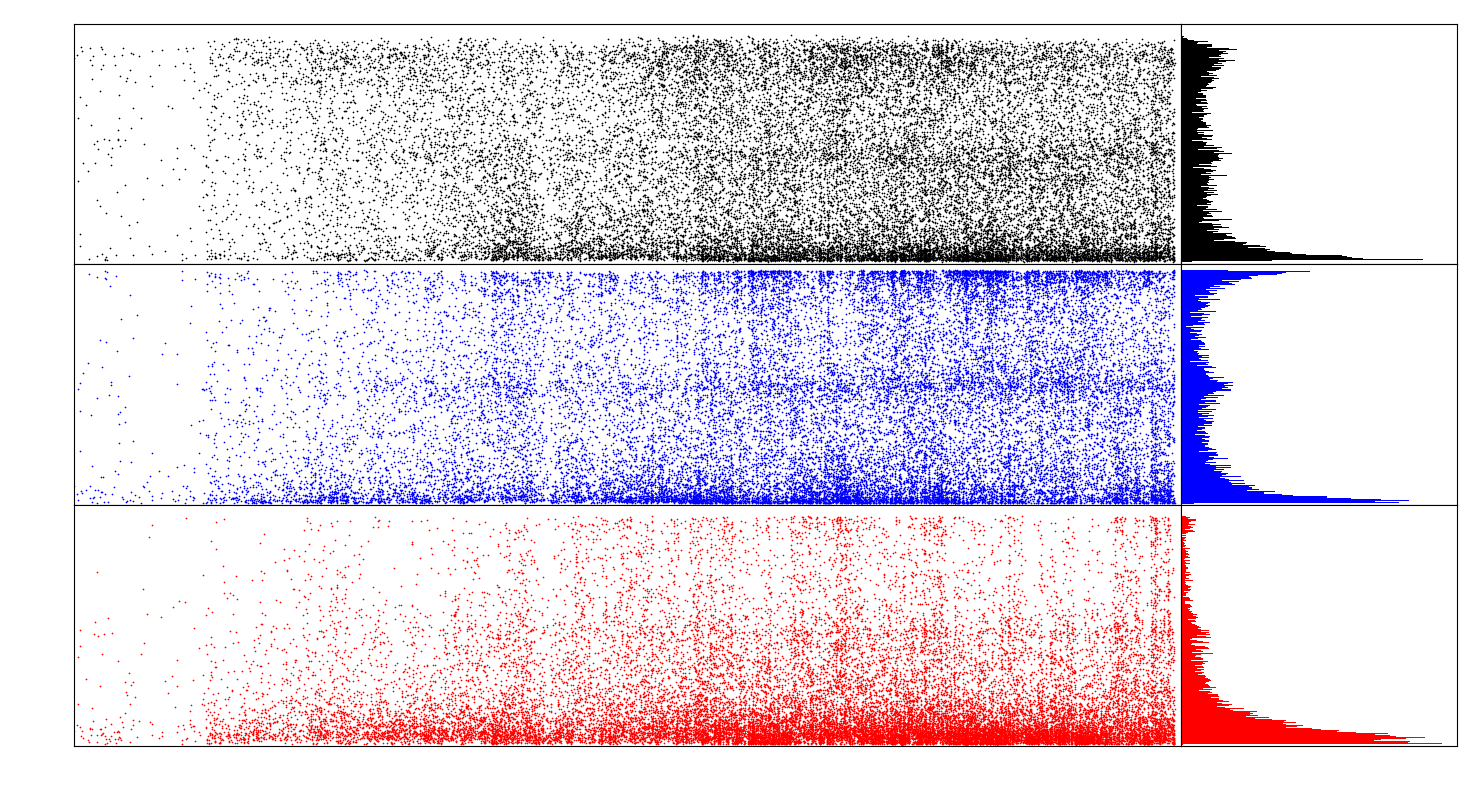
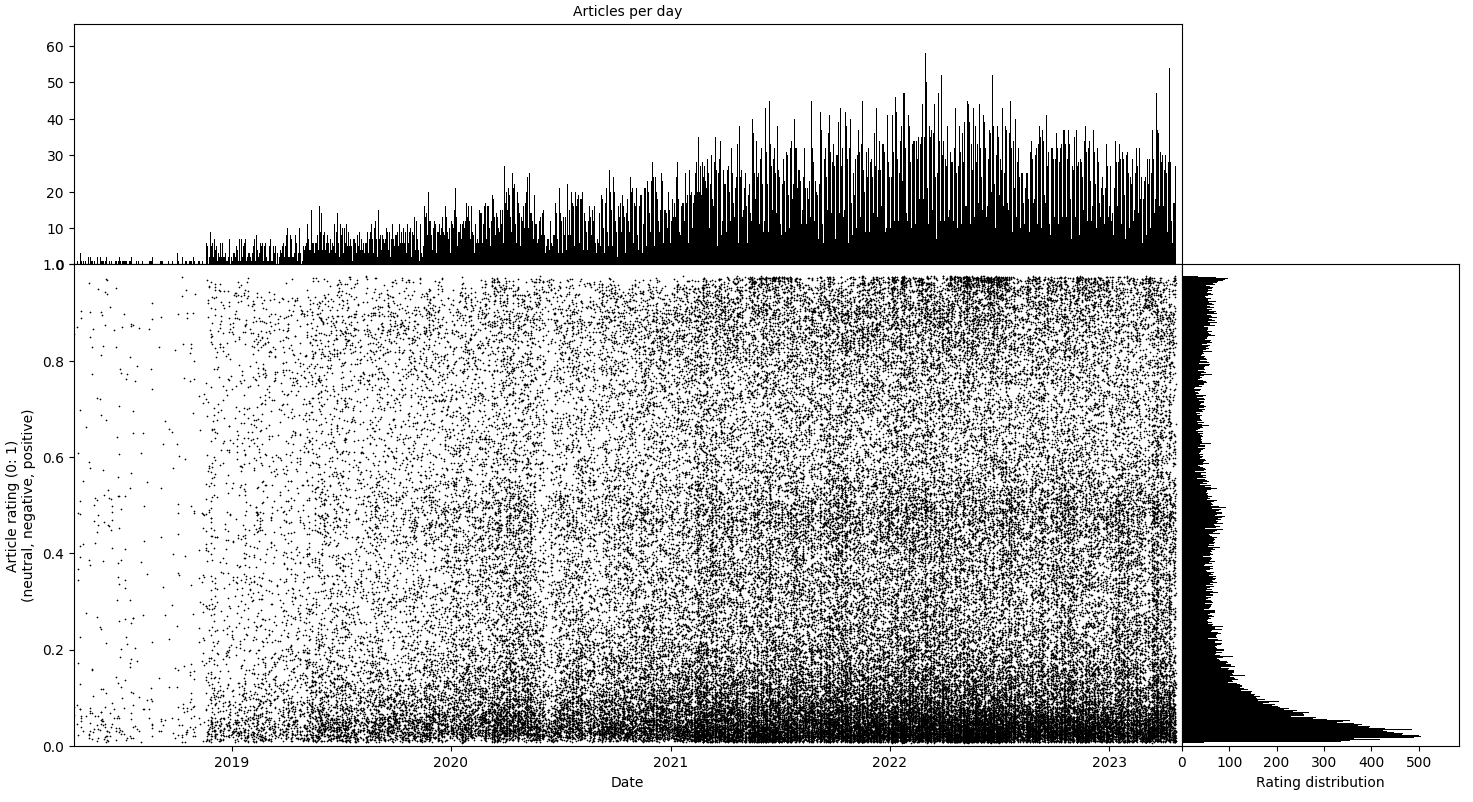
On X axis – date of the publication, on Y axis – rating. Graph above – distribution of publications by date, graph on the right – distribution by sentiment rating.Black – neutral rating distribution, blue – negative rating distribution, red – positive rating distribution.



FinBERT model have a tendency to assign news article titles ratings that are closer to the extreme ends of the scale [0; 1]. Furthermore, the model's output suggests that it gives higher weight to negative sentiment (maximum rating of 0.978) as compared to neutral (max = 0.956) and positive sentiment (max = 0.958).

Article text rating (neutral, negative, positive).

On X axis – date of the publication, on Y axis – rating. Graph above – distribution of publications by date, graph on the right – distribution by sentiment rating.  
  
Black – neutral rating distribution, blue – negative rating distribution, red – positive rating distribution.

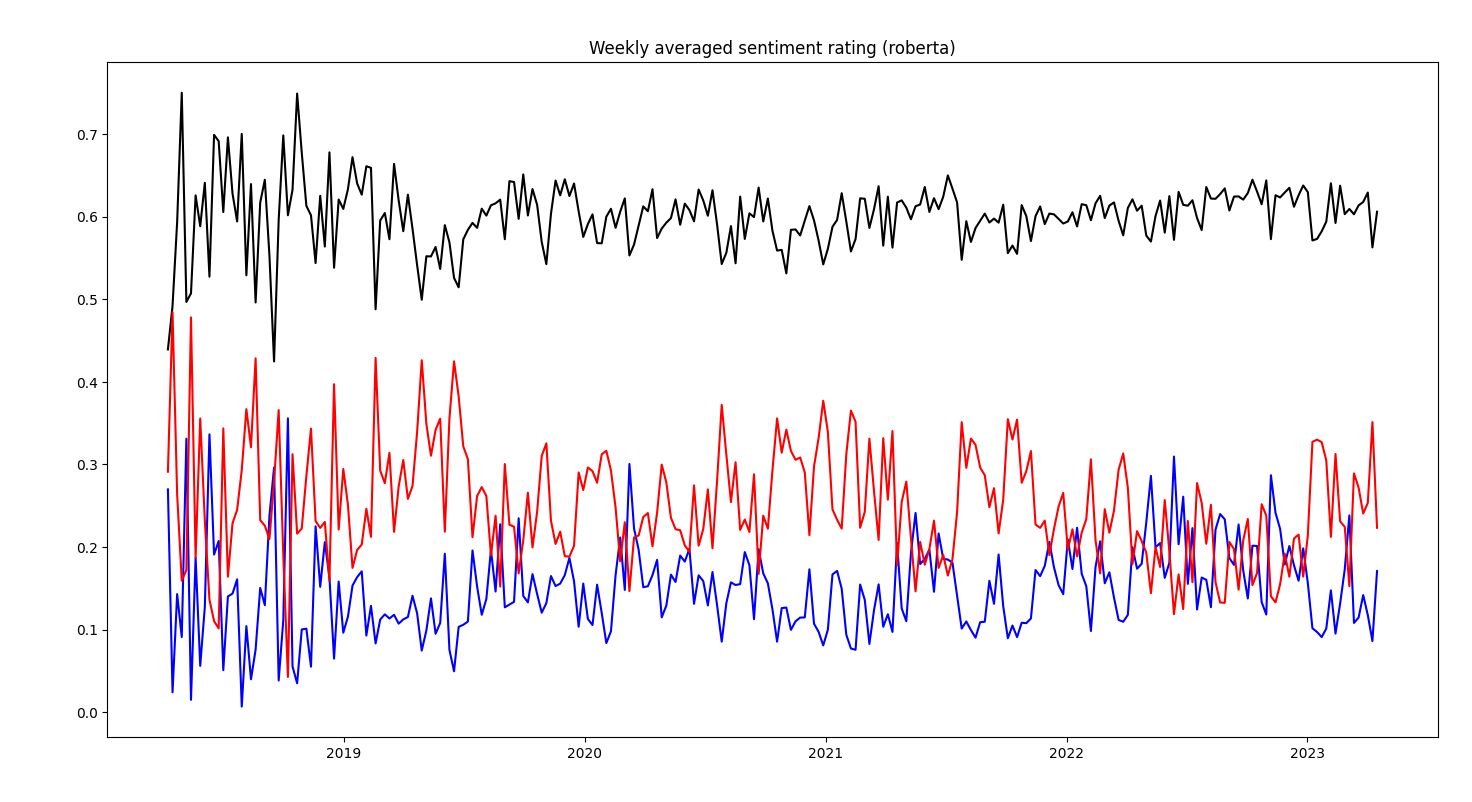
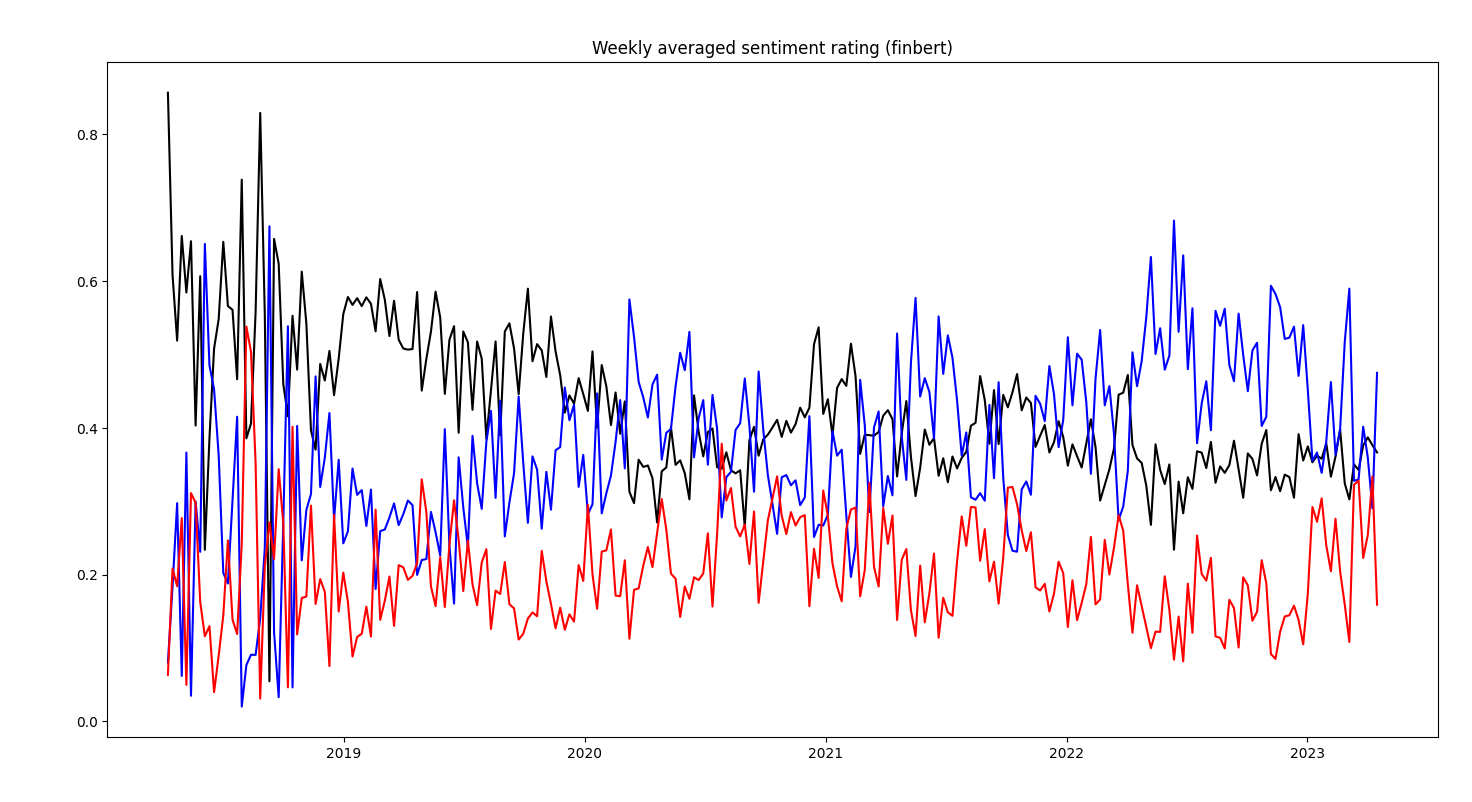


The analysis reveals that a limited number of news articles receive a positive rating above 0.5. However, a significant portion of negative ratings is distributed across the 0, 0.5, and 1 ranges. This indicates that utilizing this particular model may result in lower sentiment ratings compared to alternative models.

FinBERT is a pre-trained Natural Language Processing (NLP) model designed to analyze the sentiment of financial text. This model was created by further training the BERT language model in the finance domain using a large financial corpus, making it fine-tuned for financial sentiment classification.

Given its specific training to analyze sentiment, it may be more suitable for making price predictions for Bitcoin.

Analyzing sentiment ratings



This is the ratings averaged by weeks. Black – neutral rating, blue – negative rating, red – positive rating.

Conclusion:

In conclusion, this analysis of over 28,000 cryptocurrency news articles provides compelling evidence for a correlation between news sentiment and Bitcoin prices. Using neural networks like FinBERT to conduct sentiment analysis can uncover useful signals in the noisy crypto news landscape.

The findings suggest that negative news sentiment tends to precede declines in Bitcoin prices, while positive sentiment corresponds to price increases. This relationship allows us to make more informed investment decisions around timing and exposure to cryptocurrencies. Monitoring ongoing news developments and sentiment could provide an edge for investors.

While this analysis focused specifically on Bitcoin, the methodology could be expanded to other cryptocurrencies. Future research could also incorporate additional data sources beyond news articles, such as social media, expert forecasts, and blockchain data.

Overall, sentiment analysis of crypto news and information shows promising potential for developing profitable investment strategies. With the accelerating pace of news and data creation around cryptocurrencies, advanced analytics techniques will become increasingly relevant. Continued research in this area may lead to novel types of indicators and predictive signals to capitalize on the volatility of the emerging crypto asset class.