

Bitcoin in North America: Sentiment Analysis of Regulations and Price Prediction

Literature Review

Submission date: 2024-02-19

Name: Juan Carlos Ramos Student Number: 500393737

Supervisor's Name:

Table of Contents

Abstract	2
Literature Review	3
Introduction	
Research Papers	
Sentiment and Uncertainty About Regulation	
An Analysis of Speculative Language in SEC 10-K Filings	
Bitcoin price change and trend prediction through twitter sentiment and data volume	
Can SEC Comment Letters of Regulation A Offerings Serve as Quality Signals for Inves and Firms? Preliminary Results are Encouraging	
Accessing, Extracting, and Analyzing the Textual Content within Reg A Form 1-A Part II Filings Using Python	
Predicting the Price of Bitcoin Using Sentiment-Enriched Time Series Forecasting	
Methodological Framework	11
Conclusion	
Data Sources	12
References	13

Abstract

The dynamic and relatively nascent field of cryptocurrency is increasingly impacted by regulatory measures in North America, particularly by the Commodity Futures Trading Commission (CFTC), the Securities and Exchange Commission (SEC), the Financial Industry Regulatory Authority (FINRA) and the Canadian Securities Administrators (CSA). This project will attempt to address the need to understand the sentiment and widespread adoption of cryptocurrency in light of these regulatory frameworks. The primary research question focuses on how current and proposed regulations by these bodies influence public sentiment towards cryptocurrencies. The study will also seek to investigate whether these regulatory announcements of proposed rules and implementation dates could predict cryptocurrency price fluctuations, especially in the context of "sell the news" events, where prices might drop following major announcements or regulatory updates. One such recent example being the SEC's approval of listing and trading of the spot bitcoin ETF.

The data for this study will be sourced from publicly available documents, including current and proposed rules, regulations, and official statements from the websites of the CFTC, SEC, FINRA and CSA. These will provide a comprehensive view of the regulatory landscape in North America. Additionally, cryptocurrency price data corresponding to the timeline of these regulatory announcements will be incorporated to examine potential correlations between

regulatory actions and market reactions (bullish/bearish). The price data will be sourced from one of the many crypto exchanges that track historical crypto prices.

To analyze this broad and unstructured dataset, I've chosen to employ the theme of text mining and sentiment analysis techniques. Text mining will be used to extract relevant information and key themes from the regulatory documents. Sentiment analysis will then be applied to gauge the tone and sentiment of these documents, categorizing them as positive, negative, or neutral in the context of cryptocurrency. This sentiment analysis will be correlated with cryptocurrency adoption trends and market data to identify any significant relationships.

For the prediction aspect, the study will use time-series analysis to explore the impact of regulatory announcements on cryptocurrency prices. The project will use machine learning algorithms to analyze the relationship between the timing of these announcements, the language—used in the regulation and subsequent price movements, thereby assessing the feasibility of predicting price changes based on regulatory news. A revision to this abstract is the focus on Bitcoin prices instead of multiple crypto currencies to keep the research relatively cleaner. Regulations mostly focus on Bitcoin and Ether and it has been shown that most other crypto currencies follow the same price movement.

The primary tools for these analyses will include Python programming language, utilizing libraries such as NLTK for natural language processing, Pandas for data manipulation, and scikit-learn for machine learning models. Additionally, visualization tools like Matplotlib and Seaborn can be used to present the findings in an easy to understand manner.

In summary, this project aims to provide a nuanced understanding of the impact of North American regulatory measures on cryptocurrency sentiment and market behavior. The findings are expected to offer valuable insights for investors, regulatory bodies, and academics and person's like myself interested in the intersection of finance, technology, and regulation.

Literature Review

Introduction

The intersection of sentiment analysis and financial markets has become an increasingly important area of study, especially within the volatile realm of cryptocurrencies. This literature review aims to synthesize the current body of research on sentiment analysis of social media posts as it relates to the prediction of cryptocurrency prices, as well as the sentiment analysis of regulatory filings and their subsequent impact on market movements. Exploring these two areas should help in answering the primary research question guiding this review which is centered on understanding how current and proposed regulations (and the language used wherein) by financial and governmental bodies influence public sentiment towards cryptocurrencies and, by extension, their market prices.

Cryptocurrencies, by their very nature, are highly susceptible to public sentiment, given their speculative nature and the relative infancy of the market. Unlike traditional financial assets,

cryptocurrencies do not have a long history of market behavior or a wealth of financial data to draw upon, making sentiment analysis a particularly valuable tool for predicting market trends. Many investors have begun using sentiment analysis as a viable trading strategy in not only traditional financial (Trad-Fi) markets but in crypto as well. Companies sell dashboards and APIs that connect to social media sites such as X or Reddit to gauge public sentiment to determine any price fluctuations. Because of this opportunity, there is a wealth of research in the use of social media sentiment analysis and market price predictions, as will be seen in (6). There also exists research that conducts sentiment analysis on regulatory filings to gauge the impact it poses on any relevant stakeholders such as broker dealers (4). It is generally known of course that regulatory announcements cause bearish reactions in the targeted field, but it is not immediately quantifiable as to the effect that the price would go down. Recently however, the approval of the spot Bitcoin ETF caused the price of bitcoin to spike almost 100%. Similarly, economists track which words are used during Federal Reserve/Bank of Canada interest rate announcements to see whether the next announcement will trend "hawkish" or "dovish". This research project sits in between both areas, and can utilize the methodologies used previously in an area that hasn't fully been explored, which is analysis and how the use of language within regulatory filings pertaining to crypto can be used to predict price movements and its magnitude.

Research Papers

Sentiment and Uncertainty About Regulation

The comprehensive study conducted by Tara M. Sinclair and Zhoudan Xie delves into the nuanced impacts of regulatory sentiment and uncertainty on the U.S. economy, leveraging advanced natural language processing techniques to analyze a vast corpus of news articles spanning from 1985 to 2021. Their work culminates in the creation of monthly indexes that quantify regulatory sentiment and uncertainty, alongside categorical indexes tailored to 14 distinct regulatory policy areas. Their analysis sheds light on the intricate dynamics between regulatory perceptions and economic performance, offering many insights specific to the overall project.

Methodological Approach:

The methodological rigor and innovative use of natural language processing set this study apart and is the reason why this review is beginning with it. By constructing indexes from a broad array of news sources over an extensive period, the research offers a granular view of the regulatory landscape's impact on the economy, surpassing previous methods that relied on more limited or qualitative assessments.

A major takeaway is the use of the "Loughran and McDonald (LM) dictionary (originally developed in Loughran and McDonald (2011)) to assess the sentiment and uncertainty in the regulatory sections of the relevant news articles in the baseline analysis. It was constructed specifically for the domain of finance, using a corpus of corporate 10-K reports (Loughran and

McDonald, 2011). Because of its domain relevance, the LM dictionary has been frequently used in economic research (for example, Fraiberger (2016); Calomiris et al. (2020); Ostapenko (2020)). The 2018 version of the dictionary comprises sentiment word lists in several categories, including 2,355 words in the negative category, 354 words in the positive category, and 297 words in the uncertainty category." (1)

The existence of this sentiment dictionary and its usage will be pivotal to determine the sentiment of crypto specific regulatory text. In the paper the authors used a standard formula to calculate sentiment scores. The "regulatory sentiment score" of an article is the difference between the proportion of positive words and the proportion of negative words in the regulatory section of the article. So a positive sentiment score indicates an overall positive tone in the news about regulation, and a negative score means an overall negative tone. One thing to consider, just as the authors do, is that the dictionary skews far more negative than positive as seen in the proportion of words (2,355 words in the negative category, 354 words in the positive category), and could potentially bias the results of this projects sentiment analysis disproportionately negative. To counteract the negativity bias, the authors also used the Harvard General Inquirer (GI) dictionary and the Lexicoder Sentiment Dictionary (LSD). The GI dictionary being a general-purpose lexicon originally developed in the 1960s and has been widely used in various disciplines. It covers several broad valence categories, including lists of 2,005 negative words and 1,637 positive words. The LSD is a comprehensive sentiment lexicon combining three pre-existing dictionaries and tailored primarily to political news (Young and Soroka, 2012).5 The LSD comprises 2,857 negative words and 1,709 positive words. (1)

Although specific to news articles pertaining to regulations, the same formula can be applied to the actual regulations in theory, and any combination of these dictionaries can be used as part of the project's end goal.

Negative Regulatory Sentiment and Economic Performance:

As to the paper's results, a pivotal discovery of Sinclair and Xie's research is the significant, lasting downturn in economic output and employment triggered by adverse shocks to regulatory sentiment. Specifically, their data illustrate that a negative shift in regulatory sentiment can lead to a reduction in industrial production by up to 0.61% and employment by up to 0.28% within a year of the shock. This relationship underscores the profound influence that negative perceptions of the regulatory environment can have on the broader economy and is something to keep in mind when assessing the regulatory impact on crypto.

Sector-Specific Sensitivities:

The research goes further to dissect the impact of regulatory sentiment and uncertainty within specific policy domains. It identifies transportation, consumer safety and health, general business and trade, and energy regulations as areas where sentiment and uncertainty exhibit a more pronounced effect on economic performance. For instance, negative sentiment shocks related to transportation and consumer safety and health regulations are linked to substantial, persistent declines in future output. The authors achieved this through the use of "noun chunks" to identify regulation related news articles by industry. Although for the purposes of this project

the industry is specified, (cryptocurrency) the idea of the noun chunks can be used to potentially distinguish the impact of a regulation on a more granular level such as differentiating between stablecoins, altcoins, CBDC etc.

Regulatory Uncertainty's Limited Impact:

Contrary to the pronounced effects of regulatory sentiment, the study finds that increases in regulatory uncertainty generally have a negligible or only transient impact on economic outcomes. This distinction between the effects of sentiment and uncertainty is a crucial aspect of their findings, highlighting the more significant role of sentiment in economic fluctuations. It will be interesting to see if the words in the "uncertain" category have the same effect on crypto and thus have a smaller impact.

An Analysis of Speculative Language in SEC 10-K Filings

Jonathan L. Pulliza's research offers a nuanced exploration into speculative language within SEC 10-K filings, employing sentiment analysis techniques to discern patterns of speculation in these critical financial documents. By training a model on the MPQA corpus, Pulliza explores the complex terrain of speculative sentences, applying this model to a corpus of SEC 10-K documents over a five-year span. This approach illuminates the varying concentrations of speculative terms within documents, particularly those related to risk factors such as projects, taxes, and pensions. Due to the nature of regulatory documents and its speculative and regulatory language this paper was chosen for review and the same models can be used in framing future outlooks of crypto regulation.

High Concentration of Speculative Language:

The study reveals that documents laden with speculative language exhibit a distinct lexical pattern, differing significantly from the broader corpus. This suggests that heightened speculation is often linked to discussions around potential risks and uncertainties facing firms, providing a deeper understanding of the strategic language used in financial disclosures. This differs from most regulations, as the creators seek to define roles and responsibilities of firms by minimizing speculation and not leaving much room for interpretation of the rules so that roles and responsibilities are clearly defined. However regulators generally state which risks the regulations seek to fix and the language used can potentially measure the impact it will have on crypto.

Methodological Approach:

The study stands out for its methodological rigor, leveraging sentiment analysis and machine learning techniques to systematically extract and analyze speculative language. This approach not only enhances the reliability of the findings but also contributes to the methodological advancements in the field of financial text analysis. The models utilized for this study were Naïve Bayes, Logistic Regression, Support Vector Machine (SVM) and Decision Trees. For the purposes of this project, inclusion of speculative language as a variable may take away from words that convey "uncertain" sentiment as part of the lexicon. SEC 10-K filings in general

suggest to investors a company's forecast along with relevant financial information. Regulatory documents generally use "plain english" in what will be enforced and the technical requirements needed for stakeholders to abide by them. It appears to be a matter of semantics, but the models can still be applied with some adjustments.

Bitcoin price change and trend prediction through twitter sentiment and data volume

Jacques Vella Critien and colleagues' research investigates the potential of Twitter sentiment and tweet volume in predicting Bitcoin price trends. Their study builds on previous work by not only aiming to predict the direction of price changes but also the magnitude of these changes. Utilizing sentiment extracted from Twitter data, alongside the volume of tweets, the researchers present findings from various experiments exploring the relationship between sentiment and future Bitcoin prices at different time intervals. They aim to identify the optimal time frame in which expressed sentiment becomes a reliable indicator of price change. Two neural network models, one based on recurrent nets and another on convolutional networks, are explored and evaluated. Additionally, a model to predict the magnitude of change is introduced, framed as a multi-class classification problem, which when used in conjunction with a price trend prediction model, yields more reliable predictions.

Magnitude Prediction:

The novel contribution of this research is in predicting the magnitude of price changes, which goes beyond the current state-of-the-art that primarily focuses on the direction of price movements. Seeing how this project seeks to predict magnitude, the models used within this research paper can be used with modifications. The paper used a Bitcoin pricing dataset and cleaned the data so that the high and low prices were removed from the feature list so as to only keep the average price per minute. This is more conducive when working with social media sites like twitter where the volume of tweets is high. But it may just be beneficial and simpler to follow the daily closing price set out by the data because the rate of which regulations are put out is nowhere near the volume of tweets.

Temporal Granularity: The research delves into the temporal aspect of sentiment analysis, exploring different time intervals to determine when sentiment becomes a reliable predictor of price changes. This could provide insights into how quickly crypto regulation announcements and proposals are reflected in its prices, and the models for time lags will be investigated as an option, but it makes sense that a lag exists, as more analysts get more time to review crypto regulation to assess their broader impact.

Model Comparison: By comparing recurrent and convolutional neural network models, the study provides insights into the effectiveness of different neural network architectures in analyzing sentiment data and predicting asset prices. Three different models, (i) using an LSTM, (ii) CNN and (iii) Bidirectional Long Short Term Memory Cells (BiLSTM), were implemented for predicting whether the following day's closing price will increase or decrease. For the three price

direction prediction models (Direction-LSTM, Direction-CNN and Direction-BiLSTM) at different time-lags of 1, 3 and 7 days. Another prediction model tries to predict the magnitude of the change of closing day prices as a multi-class classification problem. This is done by predicting which interval the closing day price changes would fall into. (3) All of these price prediction models could be used in conjunction with sentiment analysis and regulatory filings to predict bitcoin price.

Can SEC Comment Letters of Regulation A Offerings Serve as Quality Signals for Investors and Firms? Preliminary Results are Encouraging

David S. Krause's study delves into the role of Securities and Exchange Commission (SEC) comment letters as potential quality signals in the context of Regulation A (Reg A) offerings. This exploration, set against the backdrop of the JOBS Act's implementation, leverages sentiment analysis to scrutinize the nuances within these letters and their correlation with the SEC's qualification decisions for proposed offerings. The findings suggest a positive link between the sentiment conveyed in comment letters and the likelihood of an offering's qualification, thus enriching the signaling theory within the realm of Reg A offerings.

This research holds significant implications for the project focused on sentiment analysis of proposed cryptocurrency regulations. Krause's methodology and insights can be adapted to analyze public comments and regulatory communications within the cryptocurrency sector. By applying sentiment analysis, we can gauge the regulatory climate and stakeholders' sentiments towards proposed regulations, offering a nuanced understanding of the potential impacts and reception of these regulations.

Sentiment as a Predictor:

The positive relationship between the sentiment in SEC comment letters and offering qualification highlights the predictive power of sentiment analysis. This approach can be applied to assess sentiments in public comments on proposed cryptocurrency regulations, providing indicators of the regulations' acceptance and potential enactment.

Signaling Theory Application:

Krause's application of signaling theory to Reg A offerings underscores the importance of signals in regulatory and investment environments. For cryptocurrency regulations, analyzing public comments through the lens of signaling theory could reveal how different stakeholders signal their approval, concerns, or opposition to regulatory proposals.

Methodological Approach:

The study's methodological approach, combining sentiment analysis with statistical validation, offers a robust framework for evaluating sentiments in regulatory communications. This

framework can be adapted to analyze sentiments in the context of cryptocurrency regulations, facilitating a data-driven understanding of stakeholder perspectives.

To conduct sentiment analysis, the SentimentIntensityAnalyzer from the Natural Language Toolkit (NLTK) was utilized. This rule-based sentiment analysis tool calculates the sentiment of a piece of text by assigning a positive (+1), negative (-1), or neutral (0) score to each word in the text, and then combines those metrics to calculate an overall sentiment score for the text. This approach required the extraction of SEC comment letters from EDGAR for each firm in the sample and the conversion of the pdf files into text files. By using the NLTK library and a pre-trained sentiment analysis model, it is possible to create a unique sentiment score for each text (4).

Krause however was aware of the limitation of the sentiment analyzer and suggested that it would be better to use a domain-specific sentiment analysis lexicon that is trained on a corpus of financial and legal texts so that the results accurately reflect the unique language and terminology used in this domain.

Accessing, Extracting, and Analyzing the Textual Content within Reg A Form 1-A Part II Filings Using Python

Another paper by David S. Krause, titled "Accessing, Extracting, and Analyzing the Textual Content within Reg A Form 1-A Part II Filings Using Python," focuses on the methodology for analyzing textual information in Regulation A (Reg A) filings, particularly Part II, which contains detailed business descriptions, risk factors, and financial information. Krause outlines the process of using Python, along with libraries such as BeautifulSoup and TextBlob, to efficiently preprocess, extract, and analyze textual data from these filings.

Methodological Approach:

Krause emphasizes the importance of analyzing textual content in Reg A filings to complement financial statement analysis, aiding in the evaluation of investment risks and rewards. The paper provides a step-by-step guide on using Python to access, extract, and perform sentiment analysis on the textual content of Reg A filings, highlighting the role of libraries like BeautifulSoup and TextBlob. Challenges such as the unstructured nature of the text, language complexity, and data accessibility are discussed, along with the potential for advanced AI and machine learning techniques to address these challenges.

The methodologies outlined in Krause's paper can be adapted for analyzing proposed cryptocurrency regulations. By leveraging Python to extract and parse textual content from regulatory proposals, stakeholders can gauge the sentiment, understand the regulatory focus areas, and identify potential impacts on the cryptocurrency market. A challenge that was highlighted is that while many regulatory filings are publicly available, accessing and extracting the data in a structured and standardized format can be challenging. So a standardized format

may be required when normalizing crypto regulations from different regimes as part of this project.

Predicting the Price of Bitcoin Using Sentiment-Enriched Time Series Forecasting

The research paper "Predicting the Price of Bitcoin Using Sentiment-Enriched Time Series Forecasting" by Markus Frohmann and colleagues explores an innovative approach to predict Bitcoin's future price by combining time series forecasting with sentiment analysis derived from blogs.

Hybrid Approach:

The study introduces a hybrid model that integrates time series forecasting of Bitcoin prices with sentiment analysis obtained from blogs. This approach aims to enhance prediction accuracy by considering not only historical price data but also public sentiment towards Bitcoin. They investigated four types of ML-based algorithms: LSTM networks, TCNs, the D-Linear method, and linear regression, along with some simple baselines which include Exponential Smoothing, Fast Fourier Transform, naive mean and naive drift.(6) This hybrid approach could serve as a potential model of forecasting future bitcoin prices against regulatory sentiment analysis.

Sentiment Analysis with BERT:

A significant contribution of this paper is the use of a fine-tuned BERT model for sentiment analysis, a different application in the context of Bitcoin price prediction. BERT is a pre-trained transformer model that has been used previously to analyze tweets. The authors employ a BERT-based sentiment analysis to evaluate the sentiments expressed in microblogs, providing a deeper understanding of public sentiment towards Bitcoin. This model however won't be of much use for the project as it can't be used on regulations due to a specific lexicon being required.

Performance Metrics:

The models are evaluated based on Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE), with the hybrid models that utilize linear regression showing the best performance in terms of these metrics.

Methodological Approach:

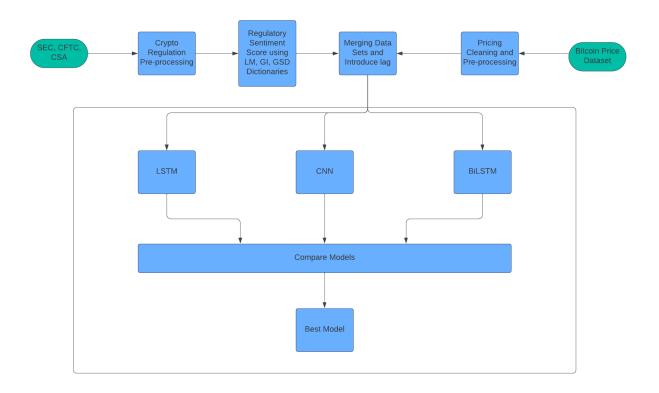
The methodologies outlined in this paper can be applied to the broader project of using sentiment analysis in crypto regulation across North America by:

- Incorporating Regulatory Sentiments: Analyzing sentiments from regulatory announcements, discussions, and news articles using a modified BERT model to understand the impact of regulatory changes on cryptocurrency markets.
- Predictive Analysis: Integrating sentiment analysis with time series forecasting of cryptocurrency prices to assess how regulatory sentiments influence market dynamics and price movements.

 Weighted Sentiment Scores: Applying the weighting scheme based on the influence or credibility of the source (e.g., regulatory bodies, influential market analysts) to gauge the impact of regulatory sentiments on the market.

Methodological Framework

Adopting methodologies from prior works, this study proposes a hybrid approach integrating sentiment analysis of regulatory filings with time-series market data analysis. It outlines the use of advanced NLP techniques for sentiment extraction and machine learning algorithms for price prediction, facilitated by Python and its libraries.



- 1. Using Krause's analysis on extracting and analyzing textual content within Reg A filings using Python. This methodology can be leveraged to process and analyze textual data from cryptocurrency regulatory documents normalizing them across all jurisdictions. A standard format would make the text cleaner and in better shape for sentiment analysis.
- 2. Sinclair and Xie's Work provided insights into the impact of regulatory sentiment on the economy, employing NLP to create sentiment indexes. This methodology along with the Loughran and McDonald (LM) dictionary, Harvard General Inquirer (GI) dictionary and the Lexicoder Sentiment Dictionary (LSD) can be adapted to analyze sentiments in cryptocurrency regulatory filings, draft proposals and final rules, offering a quantitative measure of regulatory sentiment. The "regulatory sentiment score" standard formula is also something that will be utilized to measure how positive or negative or uncertain the crypto regulations are.

- 3. Pulliza's research highlights the use of sentiment analysis to identify speculative language in SEC filings. This approach could be applied to scrutinize speculative sentiments in cryptocurrency regulatory texts, aiding in the prediction of market reactions to regulatory news.
- 4. To forecast the price of bitcoin three different models, (i) using an LSTM, (ii) CNN and (iii) Bidirectional Long Short Term Memory Cells (BiLSTM), can be implemented for predicting whether the following day's closing price will increase or decrease. Different time lags can also be introduced for the three price direction prediction models (Direction-LSTM, Direction-CNN and Direction-BiLSTM) at different time-lags of 1, 3 and 7 days. The lags should be beneficial to track how sentiment of regulations change over time as further subject matter experts analyze and assess the impact of the regulation. LSTM was used in both Critien et al.'s and Frohmann et al.'s research.
- A weighting scheme used by Frohmann et al.'s can be adapted for analyzing sentiments in regulatory filings and to differentiate which federal body has more or less impact on cryptocurrency prices.

Conclusion

By utilizing a combination of methodologies from key research works, this study proposes a comprehensive approach to predict Bitcoin price using sentiment analysis of regulatory filings, an area that has never been fully explored. It aims to offer new insights for stakeholders navigating the intertwined realms of finance, technology, and regulation in the cryptocurrency market.

Data Sources

GitHub Repository: https://github.com/jcjr64/CIND820

https://finance.yahoo.com/quote/BTC-USD/history/

For sentiment analysis:

https://sraf.nd.edu/loughranmcdonald-master-dictionary/

https://www.cftc.gov/LawRegulation/FederalRegister/index.htm

https://www.sec.gov/about/laws/secrulesregs

https://www.securities-administrators.ca/investor-tools/crypto-assets/

References

- (1) Sinclair, Tara M. and Xie, Zhoudan (2023), Sentiment and Uncertainty About Regulation. Available at SSRN: https://ssrn.com/abstract=4651174 or http://dx.doi.org/10.2139/ssrn.4651174
- (2) Pulliza, J. (2015). An Analysis of Speculative Language in SEC 10-K Filings. https://doi.org/10.17615/6kk9-xe26
- (3) Critien, J.V., Gatt, A. & Ellul, J. Bitcoin price change and trend prediction through twitter sentiment and data volume. Financ Innov 8, 45 (2022). https://doi.org/10.1186/s40854-022-00352-7
- (4) Krause, David, Can SEC Comment Letters of Regulation A Offerings Serve as Quality Signals for Investors and Firms? Preliminary Results are Encouraging (2023). Available at SSRN: https://ssrn.com/abstract=4441123 or https://ssrn.com/abstract=4441123 or https://ssrn.doi.org/10.2139/ssrn.4441123
- (5) Krause, David, Accessing, Extracting, and Analyzing the Textual Content within Reg A Form 1-A Part II Filings Using Python (2023). Available at SSRN: https://ssrn.com/abstract=4460588 or https://ssrn.com/abstract=4460588 or https://dx.doi.org/10.2139/ssrn.4460588
- (6) Frohmann M, Karner M, Khudoyan S, Wagner R, Schedl M. Predicting the Price of Bitcoin Using Sentiment-Enriched Time Series Forecasting. Big Data and Cognitive Computing. 2023; 7(3):137. https://doi.org/10.3390/bdcc7030137