CSE-6242 Team 37 Final Report

Social Analytics Dashboard For Bitcoin Related Discussion

Ke Xin Chong Wei Hong Low Alex Noah Koh Joshua Caleb Rauf
Jingyun Sun Eason Xinran Wang
{kchong43, wlow7, akoh34, jrauf3, jsun477, easonxinranwang}@gatech.edu

1 INTRODUCTION

Bitcoin, the first decentralized digital currency, has become a powerful force in both financial markets and online discourse. As interest in cryptocurrencies grows, so too does the impact of social media on investor sentiment and market behavior. Platforms like X (formerly Twitter) have become hotspots for real-time information exchange, where news breaks, opinions form, and collective mood can swing prices within hours.

Yet, the complexity of this social layer often goes unnoticed in traditional crypto analysis. While financial indicators and price trends remain essential, they tell only part of the story. The narratives, opinions, and sentiments expressed in tweets offer valuable signals—early warnings, crowd sentiment shifts, or emerging concerns—that quantitative metrics alone cannot capture.

This project aims to bridge that gap by developing an interactive dashboard that uses natural language processing (NLP) and data visualization to analyze Bitcoin-related discussions on X in real time. By making social sentiment and emerging topics visible and trackable, we aim to provide a richer, narrative-aware view of the Bitcoin ecosystem.

2 PROBLEM DEFINITION

Despite the evident influence of social media on Bitcoin's market movements, existing tools fall short in capturing the qualitative aspects of online discussions. Most current platforms focus on numerical data such as trading volume, engagement metrics, and price trends, overlooking the content and tone of conversations that shape public perception.

This creates a significant blind spot for traders, analysts, and researchers who rely on timely insights. Without tools that extract meaning from social narratives, users miss out on crucial shifts in sentiment, trending topics, or coordinated movements that could signal volatility or opportunity.

To address this, our project introduces a social analytics dashboard tailored specifically for Bitcoin-related content on X. By applying NLP techniques to identify sentiment trends, keyword shifts, and evolving narratives, the dashboard empowers users to monitor the social dimension of the crypto market with greater clarity and depth.

3 LITERATURE SURVEY

Several studies have explored the correlation between social media sentiment and Bitcoin price movements. [1][2] found that sentiment, even from less active users, affects Bitcoin prices, highlighting the need for real-time sentiment analysis. [3][4] showed that NLP techniques and deep learning enhance cryptocurrency price trend predictions by capturing sentiment shifts. The role of visualization in linking sentiment to market volatility was highlighted by [5][6], and [7] found positive tweets can predict Bitcoin prices. However, traditional sentiment tools like VADER struggle with cryptocurrency jargon [8][9][10][11], while LSTM models [12] and BERT-based approaches [14] improve accuracy. [13] proposed frameworks for integrating market data into financial dashboards, and [17] used text visualization to track Bitcoin trends. [15] noted social media's role in market manipulation, while [16] showed search interest complements sentiment analysis. [18] found domain-specific models like CryptoBERT and FinBERT improve real-time sentiment analysis.

In summary, the gaps in current research are:

(1) **Limited Sentiment Analysis Tools**: Existing models often fail to fully capture cryptocurrency-specific jargon, impacting prediction accuracy; (2) **Lack of Real-Time Integration**: Many systems do not integrate real-time market data with sentiment analysis, limiting their potential to improve market insights; (3) **Insufficient Visualization Methods**: While some efforts exist to visualize Bitcoin trends, more advanced techniques are needed to provide dynamic, actionable insights, especially for non-expert users.

4 PROPOSED METHOD

4.1 Intuition While traditional sentiment analysis tools like VADER are quick and interpretable, they fall short in domains with specialized language—such as cryptocurrency. Our intuition is that transformer-based models, especially domain-specific variants like CryptoBERT, offer a distinct advantage by capturing nuanced semantics and context within tweets. Additionally, traditional dashboards often display raw metrics or shallow analytics; in contrast, our NLP-powered dashboard interprets narrative evolution and emotional shifts in real-time, creating actionable insights for both casual and professional users. This combination of deep language understanding and dynamic visualization gives our approach an edge over existing tools, enabling users to perceive not just what is being said about Bitcoin, but why and how conversations evolve.

4.2 Approach Description Our approach involves a modular, end-to-end pipeline that integrates data preprocessing, domain-specific natural language processing, and interactive visual analytics to surface insights from Bitcoin-related social media activity. We presented the key components below:

4.2.1 Data Preprocessing and Cleaning

The raw dataset, sourced from Kaggle's "Bitcoin Tweets (2016–2019)" collection, consists of millions of tweets containing mentions of Bitcoin. To prepare this data for meaningful analysis, we performed the following cleaning steps:

- **Language Filtering**: We removed all non-English tweets using the langdetect library to focus our analysis on a linguistically consistent subset.
- **Noise Removal**: URLs, user mentions (username), hashtags (#topic), emojis, and special characters were stripped from tweet text using regular expressions and text preprocessing utilities to reduce model noise.
- **Tokenization and Normalization**: For downstream NLP tasks, we applied standard tokenization, lowercasing, and lemmatization to normalize text for embedding models.
- **Data Storage**: After preprocessing, the cleaned data—along with computed sentiment scores, engagement metrics, and topic modeling outputs—was stored in CSV files. This format allowed for flexible manipulation during development and easy integration into the visualization pipeline.

4.2.2 Sentiment Analysis using Transformer Models

Recognizing the limitations of lexicon-based tools like VADER, especially in the cryptocurrency domain where slang, sarcasm, and abbreviations are prevalent, we opted for a deep learning-based sentiment classifier.

- We fine-tuned CryptoBERT, a transformer model pretrained on cryptocurrency-related text, to classify tweets into positive, negative, or neutral sentiment.
- This model was chosen for its ability to capture semantic nuances and contextual cues unique to the Bitcoin community.
- Each tweet was passed through the model to generate sentiment scores, which were then stored alongside metadata in CSV files. These files serve as the primary data source for the dashboard and allow for efficient loading and filtering during visualization, even without a dedicated database.

4.2.3 Topic Modeling with BERTopic

To uncover latent themes in Bitcoin-related tweets, we used **BERTopic**, a transformer-based topic modeling framework that integrates contextual embeddings, dimensionality reduction, clustering, and advanced topic labeling techniques:

- **Embeddings**: We encoded each tweet using the `BAAI/bge-small-en-v1.5` SentenceTransformer model, which balances performance and efficiency. Computation was accelerated using GPU.
- **Dimensionality Reduction**: UMAP reduced the high-dimensional embeddings to 5 components, making clustering more tractable while preserving semantic relationships.
- **Clustering**: HDBSCAN was applied to the reduced embeddings to form topic clusters without requiring a predefined number of groups. Parameters such as `min_samples=30` ensured robustness to noise.
- **Prompt-Based Labeling**: To generate clear, human-readable topic labels, we used a few-shot prompt with `meta-llama/Llama-3.1-8B-Instruct`. Each prompt included sample tweets and keywords for the topic, and the model returned a short label. The full prompt format is provided in Appendix A.
- **Representation**: We further enhanced label quality by integrating three methods—KeyBERT, Maximal Marginal Relevance (MMR), and LLaMA—into BERTopic's `representation_model`.
- **Output**: Topic labels and assignment scores were appended to the dataset and saved as `combined_df.csv`, with full topic metadata exported to `topic_detail.csv`. The trained model was saved for reuse.

This approach enabled us to surface coherent themes and emerging narratives in Bitcoin discussions, improving interpretability over traditional keyword-based methods.

4.2.4 Dashboard Design

The dashboard was explicitly designed to provide comprehensive insights into Bitcoin sentiment analysis through an intuitive and visually engaging interface. It comprised several interactive visualizations: a sentiment line chart, a stacked area chart, a heatmap, and a word cloud.

The visualization design was tailored to maximize interpretability and insight extraction. The sentiment line chart was selected for its ability to clearly depict overall trends in public opinion over time. Constructed using D3.js, it maps sentiment scores against temporal markers, enabling users to detect shifts in sentiment and correlate them with real-world events. The line was rendered in steel blue to maintain visual consistency and leverage the intuitive connotation of calm, trustworthy tones. Axis labels, gridlines, and tick marks were clearly formatted using a sans-serif font to ensure legibility without distraction.

The stacked area chart was included to show the relative proportions of sentiment categories over time, offering a complementary perspective to the line chart. Each sentiment type was represented with distinct, high-contrast colors—green for positive, grey for neutral, and red for negative—chosen to align with user expectations and emotional associations. This visualization emphasized not only fluctuations but also the magnitude of sentiment shifts.

The heatmap was implemented to provide granular insights into temporal patterns, especially daily and hourly fluctuations in sentiment, which are crucial for real-time analysis and market timing. Constructed as a two-dimensional grid, it used a sequential color scale (from pale yellow to deep red) to denote tweet frequency. The scale enabled users to instantly identify high-activity periods. Textual overlays and tooltip layers were carefully designed to prevent visual clutter, with lighter fonts used in low-density cells.

The word cloud visualization, positioned for thematic exploration, distilled key discussion topics by frequency and prominence, derived from keyword extraction techniques. Keywords were scaled by frequency and colored based on associated sentiment categories, allowing users to interpret not only what was being discussed but the underlying tone. Interactive zooming and hover-enabled metadata provided depth without overwhelming the initial view.

The interface design was informed by user-centered principles, focusing on simplicity, responsiveness, and aesthetic coherence. Layouts were arranged in a responsive grid format, aligning with natural visual scanning patterns and facilitating comparison across views. Visual hierarchy was preserved through consistent use of type weights, white space, and font sizes. A modern, sans-serif typeface was selected for all text to support screen readability.

Filters for sentiment category and date range were positioned prominently at the top of the dashboard to encourage data slicing and hypothesis testing. The color palette was kept neutral and low-saturation to reduce eye strain and support sustained exploration. Interactive elements—including hover states, tooltips, and zoom/pan features—were designed with soft transitions and subtle animations to reinforce user control while maintaining a low cognitive load.

5 EVALUATION

To assess the effectiveness and usability of our dashboard, we conducted a user study involving six participants representative of our intended audience. Participants varied in their familiarity with data analytics and visualization tools, ranging from novice to moderately experienced users. Specifically, three participants had minimal prior experience with data visualizations, while the remaining three participants possessed moderate familiarity with analytics platforms. Participant demographics included two graduate students majoring in computer science, one undergraduate student majoring in finance, two professionals in the technology sector with moderate analytical experience, and one marketing professional with limited exposure to data visualization. This diverse composition allowed us to assess the accessibility of our design for beginners and its depth of insight for more experienced users.

The primary objectives of our evaluation were guided by several research questions:

How intuitively can users navigate through the dashboard and interact with various filtering functionalities?

To what extent are users able to accurately interpret trends and insights from each visualization type—line charts, stacked area charts, heatmaps, and word cloud visualizations?

What features do users find most helpful or challenging, and why? How effectively does the dashboard facilitate actionable insights regarding Bitcoin sentiment data?

How do participants perceive the dashboard's overall aesthetic and information density? Does the dashboard support exploratory analysis across both macro (weekly/monthly) and micro (hourly/daily) temporal scales?

What improvements would increase the interpretability or perceived utility of each visual module?

How do prior levels of visualization experience affect users' interpretation and interaction behaviors?

Are users able to connect sentiment patterns with hypothetical market timing strategies based on visualized data?

Participants were individually asked to perform specific tasks designed to test various dashboard functionalities comprehensively. Tasks were meticulously crafted to reflect realistic user scenarios and included filtering data by custom date ranges, adjusting sentiment categories, interpreting overall sentiment trends from the Bitcoin sentiment line chart, analyzing sentiment distributions from the stacked area chart, identifying hourly and weekly sentiment patterns using the heatmap visualization, and extracting key themes from the word cloud. Tasks were progressively structured from simpler interactions, such as basic navigation and filtering, to more complex interpretive tasks, thereby gauging the dashboard's usability at varying levels of complexity.

Participants were individually asked to perform specific tasks designed to test various dashboard functionalities comprehensively. These tasks were crafted to mirror real-world use cases and included: Filtering data by custom date ranges

Adjusting sentiment categories to explore shifts in narrative

Interpreting overall sentiment trends from the Bitcoin sentiment line chart

Analyzing relative distributions of sentiment from the stacked area chart

Identifying hourly and weekly sentiment patterns using the heatmap visualization

Extracting and interpreting key discussion themes from the word cloud

Each session was conducted in a quiet, controlled environment using a 15-inch MacBook Pro with Google Chrome in full-screen mode. Sessions lasted approximately 30 minutes and began with a guided introduction to the dashboard's purpose and features. Participants then received a set of printed tasks with contextual prompts, designed to simulate authentic analytical goals (e.g., "Imagine you're preparing a market timing strategy based on recent sentiment changes—how would you proceed?"). Observations were made unobtrusively, and screen recordings were used to capture interaction flow and timing.

Participants were encouraged to verbalize their decision-making and any points of confusion using the think-aloud method. After the task completion, we conducted a brief semi-structured interview to probe deeper into their experiences and perceptions. Additionally, a 10-item questionnaire rated aspects such as usability, visual appeal, and confidence in derived insights on a 5-point Likert scale.

Evaluations employed qualitative and quantitative techniques. Observational data documented task completion times, interaction patterns, and error frequencies. Post-task interviews delved deeper into participant experiences, clarifying observed behaviors, and gathering detailed feedback on specific features. Additionally, a structured questionnaire was administered at the end of each session, quantitatively assessing satisfaction, perceived usability, aesthetics, and information density on a Likert scale.

Throughout the study, several notable observations emerged. Participants unanimously found the Bitcoin sentiment line chart easy to understand, clearly conveying general sentiment trends. However, some participants encountered challenges when selecting custom date ranges beyond preset defaults, suggesting the need for clearer instructions or visual enhancements for date selection.

The sentiment category filter was highly praised for its straightforward functionality, significantly aiding participants in data exploration. Reactions to the stacked area chart varied; while it effectively illustrated sentiment distribution for several participants, others experienced initial confusion due to overlapping colors. Participants recommended better color differentiation or the addition of interactive legends.

Participants generally appreciated the heatmap visualization for its detailed hourly and weekly sentiment breakdown. However, dense tooltip information initially overwhelmed some users, indicating a need for simpler presentation or incremental information disclosure. The word cloud visualization was particularly effective in summarizing key themes, although participants suggested enhancements in interactivity, such as improved zoom functionality and keyword highlighting.

Questionnaire findings supported observational insights and offered quantitative perspectives. Participants expressed overall high satisfaction, averaging 4.2 out of 5. Usability scored 4.0 out of 5, indicating ease of use with minor difficulties linked to specific interactive components. Aesthetic appeal was notably high, with an average score of 4.5, reflecting strong visual clarity and professional design. Information density received a moderate average rating of 3.8, highlighting potential improvements in complexity management and content organization.

Overall, feedback emphasized the dashboard's strengths in aesthetics, intuitive interactions, and clarity in depicting sentiment trends. Concurrently, the evaluation identified clear opportunities for improvement in visual differentiation, interactivity refinements, and information presentation. These findings will guide targeted enhancements aimed at optimizing user experience and improving the dashboard's effectiveness in data-driven decision-making.

6 CONCLUSIONS AND DISCUSSION

In this project, we developed a prototype for an interactive social analytics dashboard that visualizes sentiment trends and topical dynamics in Bitcoin-related discussions on X (formerly Twitter). Our approach combined fine-tuned transformer models, topic modeling with BERTopic, and an interactive visualization pipeline to create a tool that goes beyond conventional crypto dashboards by focusing on narrative and sentiment rather than just price or engagement metrics.

Key innovations include the use of CryptoBERT for domain-specific sentiment analysis, which enabled more accurate interpretation of cryptocurrency slang and jargon, and BERTopic, which surfaced high-level discourse themes by leveraging semantic embeddings. The processed results were stored in CSV format, enabling lightweight integration with our D3.js-powered dashboard, which visualizes rolling sentiment scores, activity heatmaps, and emerging topics.

6.1 Impact and Significance This work highlights the growing importance of social context in cryptocurrency markets. By capturing and visualizing how sentiment and discourse evolve over time, the dashboard empowers traders, analysts, and enthusiasts to better anticipate shifts in public opinion, identify community-driven events, and spot early signals of market movements.

6.2 Limitations While our results are promising, the project has several limitations:

- The dataset is historical (2016–2019) and static, limiting real-time insights.
- Tweets in languages other than English were excluded, which may reduce global sentiment coverage.
- CSV-based storage, while practical during prototyping, is less scalable for large or live-streaming datasets.
- Sentiment classification was limited to a three-class system (positive, neutral, negative); a more granular emotional spectrum could provide deeper insights.

6.3 Future Work To address these limitations and extend the project's impact, we plan to:

- Integrate live data streaming via the X API or third-party tools to support real-time sentiment monitoring.
- Expand multilingual support to include major global languages relevant to the crypto community.
- Transition from CSVs to a cloud-hosted database or data warehouse to support scalable querying and analysis.
- Explore multi-label sentiment models or emotion classifiers (e.g., anger, excitement, fear) for finer-grained emotional tracking.
- Incorporate price and volatility data to explore correlations between social sentiment and market performance.

Ultimately, this project lays the foundation for a comprehensive social analytics platform for the cryptocurrency domain—one that leverages NLP and visualization to make the collective voice of the crypto community transparent and actionable.

7 DISTRIBUTION OF TEAM MEMBER EFFORT

All six team members contributed collaboratively across various stages of the project, with tasks allocated based on expertise and workload balance. Joshua and Eason handled data consolidation and cleaning, ensuring the dataset was properly filtered and preprocessed for analysis. Ke Xin and Wei Hong focused on mining insights from the Twitter data, including sentiment analysis and topic modeling. Eason, Alexander, and Jingyun led the design and development of the interactive dashboard using D3.js, transforming the initial Figma prototypes into functional, dynamic visualizations.

In the later stages, Eason conducted six user studies and performed analysis to evaluate the dashboard's usability and interpretability. All team members actively participated in creating the final report, poster, and presentation. Overall, the effort was distributed equitably, with each member contributing significantly to the project's technical development, evaluation, and communication components.

References

[1] Anthony Awad. Impact of social media sentiment on cryptocurrency price movements. 2022. 1

[2] Johan Bollen, Huina Mao, and Xiaojun Zeng. Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1):1–8, Mar. 2011. 1

[3] David Garcia and Frank Schweitzer. Social signals and algorithmic trading of bitcoin. *Royal Society Open Science*, 2(9):150288, Sept. 2015. 1

[4] Andrew Ibekwe. Using sentiment analysis of social media posts in cryptocurrency price prediction. 2021. 1

[5] Achyut Jagini, Kaushal Mahajan, Namita Aluvathingal, Vedanth Mohan, and Prajwala Tr. Twitter sentiment analysis for bitcoin price prediction. *2023 3rd International Conference on Smart Data Intelligence (ICSMDI)*, pages 32–37, 2023. 1

[6] Jermain Kaminski. Nowcasting the bitcoin market with twitter signals, 2016. 1

[7] Ladislav Kristoufek. Bitcoin meets google trends and wikipedia: Quantifying the relationship between phenomena of the internet and bitcoin market dynamics. *Scientific Reports*, 3:3415, 2013. 1

[8] Wen Li, Lingfeng Bao, Jiachi Chen, John Grundy, Xin Xia, and Xiaohu Yang. Market manipulation of cryptocurrencies: Evidence from social media and transaction data. *ACM Trans. Internet Technol.*, 24(2), Mar. 2024. 1

[9] T. Low, J. Smith, R. Doe, and A. Lee. Deep learning with reddit sentiment for cryptocurrency price prediction. In *Proceedings of the 2023 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, pages 55–62. IEEE, 2023. 1

[10] Feng Mai, Zhe Shan, Qing Bai, Xin (Shane) Wang, and Roger H. L. Chiang. How does social media impact bitcoin value? a test of the silent majority hypothesis. *Journal of Management Information Systems*, 35:19 – 52, 2018. 1

[11] Ross C. Phillips and Denise Gorse. Predicting cryptocurrency price bubbles using social media data and epidemic modelling. In *2017 IEEE Symposium Series on Computational Intelligence (SSCI)*, pages 1–7, 2017. 1

[12] Khizar Qureshi and Tauhid Zaman. Social media engagement and cryptocurrency performance. *PLOS ONE*, 18(5):e0284501, 2023. 1

[13] J. Smith and L. Chen. Designing interactive financial dashboards: Enhancing decision-making in volatile markets. *Information Visualization*, 20(3):202–217, 2021. 1

[14] Lars Steinert and Christian Herff. Predicting altcoin returns using social media. *PloS one*, 13(12):e0208119, 2018. 1

[15] Evita Stenqvist and Jacob Lönnö. Predicting bitcoin price fluctuation with twitter sentiment analysis. 2017. 1

[16] Moein Shahiki Tash, Olga Kolesnikova, Zahra Ahani, and Grigori Sidorov. Psycholinguistic and emotion analysis of cryptocurrency discourse on x platform. *Scientific Reports*, 14:8585, 2024. 1

[17] Natkamon Tovanich, Nicolas Soulié, Nicolas Heulot, and Petra Isenberg. Miningvis: Visual analytics of the bitcoin mining economy. *IEEE Transactions on Visualization and Computer Graphics*, 28(1):868–878, 2022. 1

[18] Yajing Zhi, Lanfang Kong, Samed Atouati, T-H. Hubert Chan, and Mauro Sozio. Explaining cryptocurrency price trends: Statistical analysis of social media posts vs market prices. In *Proceedings of the 2024 3rd Asia Conference on Algorithms, Computing and Machine Learning*, CACML '24, page 234–239, New York, NY, USA, 2024. Association for Computing Machinery. 1

Appendix A: Prompt Template for LLaMA-2 Topic Labeling To generate concise topic labels, we used the following prompt template with the `meta-llama/Llama-3.1-8B-Instruct` model. The prompt includes a system instruction, a concrete example, and a templated format in which representative documents and topic keywords are inserted dynamically.

Below is the full prompt structure (placeholders `[DOCUMENTS]` and `[KEYWORDS]` are replaced at runtime):

```
<s>[INST] <<SYS>>
You are a helpful, respectful and honest assistant for labeling topics.
<</SYS>>

I have a topic that contains the following documents:
- Bitcoin started as a decentralized digital currency but has grown
  into a speculative asset with significant energy demands.
- The process of mining Bitcoin consumes more electricity than some entire
  countries.
- Owning Bitcoin doesn’t inherently make you rich, nor does
  avoiding it make you poor.

The topic is described by the following keywords:
  ’bitcoin, crypto, mining, energy, blockchain, currency, transaction,
  wallet, decentralized, investment’.

Based on the information about the topic above,
please create a short label of this topic.
Make sure you to only return the label and nothing more.

[/INST] Environmental impacts of Bitcoin mining

[INST]
I have a topic that contains the following documents:
[DOCUMENTS]

The topic is described by the following keywords: ’[KEYWORDS]’.

Based on the information about the topic above, please
create a short label of this topic. Make sure you to only return the
label and nothing more.
[/INST]
```

Appendix B: HEILMEIER QUESTIONS **1. What are you trying to do?** The team wishes to develop a real-time analytics dashboard to analyze social media sentiment (Twitter/X, Reddit) on Bitcoin. It will track crypto narratives like #BitcoinHalving, highlight trending topics, and provide advanced keyword search and interactive visualizations to help users understand evolving public discussions around Bitcoin.

2. How is it done today; what are the limits of current practice? Current tools analyze Bitcoin sentiment using NLP on social media and market data but lack real-time interactivity, nuanced context analysis, and dynamic visualizations. Our solution enhances user interaction with real-time exploration of sentiment and trending keywords, offering a more detailed and engaging view of Bitcoin discussions.

3. What's new in your approach? Why will it be successful? We will display topic emergence and apply clustering techniques to group similar topics dynamically, based on semantic similarities and user-defined keywords. Additionally, users can input custom keywords, enabling the dashboard to deliver tailored insights and trends aligned with their specific interests and evolving discussions.

4. Who cares? Researchers, social media enthusiasts, and cryptocurrency community members care because Bitcoin discussions significantly influence market sentiment and price trends. Current tools overlook social sentiment and topic evolution, while our dashboard addresses this gap by providing clear insights into public sentiment and trending keywords in Bitcoin discussions.

5. If you're successful, what difference and impact will it make, and how do you measure them (e.g., via user studies, experiments, ground truth data, etc.)? If successful, our dashboard will offer users an innovative platform to explore and understand the evolving social narrative around Bitcoin. The impact will be measured through:

1. **User Engagement and Feedback:** Conducting surveys, interviews, and usability tests with researchers and crypto enthusiasts to gather qualitative insights.

2. **Quantitative Metrics:** Tracking dashboard usage statistics, such as active sessions, search frequency, and time spent on visualizations, to evaluate user engagement.

3. **Benchmark Comparisons:** Comparing our sentiment analysis outputs against established benchmarks to validate accuracy and relevance.

6. What are the risks and payoffs? The primary risk is the fragmented and noisy nature of social media data, which can hinder accurate sentiment analysis. However, the payoff includes a dynamic tool that transforms chaotic discussions into actionable insights, with success measured by user engagement, community feedback, and its ability to reflect real-time Bitcoin narratives effectively.

7. How much will it cost? There are no anticipated costs as the project will utilize existing free datasets available on platforms like Kaggle.

8. How long will it take? The project will be completed by Apr 14th, with an accompanying poster and presentation to be released along with the tool.

9. What are the midterm and final "exams" to check for success? How will progress be measured? Midterm success involves integrating and cleaning social media datasets, extracting insights, and finalizing design sketches for sentiment and keyword visualizations. Final success is the completion of the interactive dashboard, presentation, and report by mid-April.