

Cryptocurrency Tweet Sentiment Analysis: Predicting Bitcoin Volatility Using Social Media Signals

Lucas Little

University of Colorado - Boulder

CSCA 5522: Data Mining Project

Contact: lucas.little@colorado.edu

July 5, 2025

Author **Note:**
cas.little@colorado.edu

Contact:

els trained on financial text for sentiment analysis, with temporal aggregation over 15-minute windows. Seasonal-Trend decomposition using LOESS (STL) decomposes the signals, while z-score based anomaly detection identifies significant changes in aggregated historical data that may be predictive of price movements. The research employs historical backtesting to determine whether the proposed approach can effectively distinguish between normal sentiment patterns and potential market-moving anomalies. Performance evaluation utilizes precision, recall, F1 score, and simulated Sharpe ratio metrics, with target thresholds of F1 score ≥ 0.50 and Sharpe ratio > 0 . The results serve as a validation study and proof of concept to assess the viability of the proposed approach for future real-time prediction applications.

Abstract

This project conducts a historical backtesting feasibility study to determine whether short-term Bitcoin volatility events can be predicted using cryptocurrency-related sentiment analysis from Twitter. The research utilizes a comprehensive Kaggle dataset containing 16 million cryptocurrency-related tweets spanning January 1, 2016 to March 29, 2019, alongside high-frequency Bitcoin price data, to develop and validate a volatility prediction framework. The methodology employs fine-tuned Sentence-Bidirectional Encoder Representations from Transformers (SBERT) mod-

1 Introduction

Cryptocurrency markets are known for their extreme volatility and unpredictability. Price swings in Bitcoin, for instance, can result in significant gains or losses within minutes. This, coupled with the 24/7 real-time nature of cryptocurrency trading, has led to a market that is very different from traditional financial markets, where market sentiment on social media may have a more direct impact on prices.

One major difference between cryptocurrency and traditional financial markets is the significant influence of retail investors in the cryptocurrency space. While traditional markets are heavily influenced by institutional investors, thousands of individual traders, enthusiasts, and skeptics are expressing their emotions, opinions, and trade decisions in real-time on social media during market movements. This presents an opportunity to capture a "group thought" at scale that is unfiltered and may have first-mover signals of imminent directional moves that are more visible and stronger than in traditional equity markets.

This research question emerged from systematic observation of the rapid and directional nature of sentiment movements in cryptocurrency communities, which frequently appear to precede rather than follow price changes. The hypothesis

posits that if retail investors possess the capacity to influence market movements through coordinated sentiment expression, and if the signal-to-noise ratio is more favorable in cryptocurrency markets compared to traditional equity markets, then this domain may present an optimal environment for demonstrating the predictive potential of sentiment-based analysis.

2 Problem Statement

While several studies have explored the potential for social media-based price prediction, few have rigorously demonstrated predictive capability using robust data-driven methodologies. Most existing research has identified correlations that are either expected or already well-documented, often as post hoc findings, or has focused on trend detection after the fact. Such approaches offer limited practical value for forward-looking prediction, particularly in the context of financial data analysis and modeling.

The critical gap addressed by this research is the statistical validation of sentiment anomalies as leading indicators for volatility events in cryptocurrency markets. This study investigates whether collective sentiment expressed on social media possesses predictive power that is more pronounced in cryptocurrency markets than in traditional financial markets. The analy-

sis focuses on the potential for retail-driven sentiment to be directly linked to cryptocurrency volatility, thereby providing a unique setting in which sentiment-based prediction may demonstrate tangible value.

2.1 Research Questions

To do this systematically, the study is split into four primary questions of if and how social sentiment can forecast volatility:

First, can sentiment anomalies in cryptocurrency Twitter discussions reliably precede Bitcoin volatility events within a 2-hour window? This fundamental question determines whether the approach has practical merit.

Second, what temporal resolution optimally captures these signals? The research examines whether 15-minute aggregation windows can capture rapid sentiment shifts more effectively than hourly windows, or if higher frequency introduces excessive noise.

Third, can this sentiment-based approach outperform traditional technical indicators? Demonstrating correlation with volatility is insufficient; the research must show superior predictive power compared to established methods.

Finally, how effective is the combination of STL decomposition and z-score analysis for distinguishing between normal sentiment fluctuations and genuinely anomalous patterns? This

methodological question is crucial for practical implementation.

2.2 Problem Scope and Boundaries

To maintain focus and manageability, the scope is deliberately narrowed to Bitcoin (BTC/USD) as the primary cryptocurrency. Bitcoin’s market dominance and extensive social media coverage make it an ideal testing ground for sentiment-based prediction models. The temporal focus centers on short-term volatility events within a 2-hour prediction window—sufficiently long to be actionable for traders, but short enough to minimize the influence of external market factors that could confound the sentiment signal.

The data foundation relies on Twitter cryptocurrency discussions from a comprehensive Kaggle dataset, analyzed through historical validation using retrospective analysis. Volatility events are defined as price movements of $\pm 5\%$ within 2 hours following detected sentiment anomalies—a threshold that captures meaningful market movements while filtering out normal price fluctuations.

Several important limitations shape the study’s boundaries. The research explicitly excludes real-time trading implementation or live system deployment, focusing instead on historical validation that can inform future real-time applications. The analysis remains Bitcoin-

specific rather than attempting to generalize across other cryptocurrencies, each of which may have distinct sentiment dynamics. Additionally, the research avoids long-term price prediction beyond 2-hour windows and excludes fundamental analysis or technical chart patterns, maintaining focus on pure sentiment-based signals.

2.3 Limitations of Existing Approaches

Current sentiment-based cryptocurrency prediction methods face several critical limitations:

- **Delayed Signal Detection:** Most existing approaches identify trends after price movements have already begun, limiting practical utility for preemptive action
- **Signal-to-Noise Challenges:** Inability to effectively filter meaningful sentiment signals from the considerable noise inherent in social media data
- **Platform Isolation:** Lack of integration between different social platforms that may capture distinct investor demographics and sentiment patterns
- **Temporal Resolution Gaps:** Insufficient granularity to capture the rapid pace of cryptocurrency markets, with most studies focusing on daily or longer timeframes
- **Validation Methodology:** Limited rigorous backtesting frameworks that simulate real-world implementation constraints

2.4 Research Contribution

This research aims to validate the viability of sentiment-based volatility prediction through:

- **High-Frequency Analysis:** Implementing fine-grained temporal analysis using 15-minute aggregation windows on retrospective data
- **Anomaly Detection Framework:** Developing a methodology that distinguishes between normal sentiment fluctuations and potential market-moving anomalies using STL decomposition and z-score analysis
- **Rigorous Validation:** Creating a comprehensive backtesting framework that provides evidence for future real-time implementation viability
- **Practical Implementation Roadmap:** Establishing foundational evidence and methodology for operational deployment in live trading environments

3 Related Work

Having established the research framework and methodology, it is important to situate this work

within the broader landscape of sentiment-based cryptocurrency prediction research. The intersection of social media sentiment and cryptocurrency price movements has evolved from a niche curiosity to a serious area of academic investigation, though significant gaps remain.

This research builds upon several key areas that have shaped understanding of both the opportunities and limitations in this field:

3.1 Social Media Sentiment in Cryptocurrency Markets

Long et al. (2025) studied the relationship between social media sentiment, market volatility, and the activities of larger traders in cryptocurrency markets. Their findings suggest that sentiment shifts often precede whale movements, which correlate with volatility events. This work provides a foundation for the hypothesis that social sentiment can serve as an early indicator of market movements.

Previous research has provided evidence of statistically significant associations between social media sentiment and cryptocurrency returns. However, they are often limited to a daily time scale. It remains uncertain whether and how this approach can help detect intraday volatility patterns and incorporate information from multiple social platforms.

3.2 Multi-Platform Sentiment Analysis

Youssfi Noura et al. (2023) combined sentiment analysis from Twitter and Google News to predict Bitcoin price movements. Their methodology demonstrated that multi-source sentiment analysis outperformed single-source approaches and achieved higher accuracy in directional prediction. This supports the proposed high-frequency Twitter analysis approach.

3.3 Advanced Volatility Prediction Methods

Brauneis and Sahiner (2024) conducted a comprehensive comparison of volatility forecasting methods, including traditional heterogeneous autoregressive (HAR) models, sentiment-enhanced approaches, and machine learning techniques. Their findings suggest that hybrid models using both technical indicators and sentiment features resulted in the best performance in predicting cryptocurrency volatility.

Kleitsikas et al. (2025) took an innovative approach by extracting sentiment signals embedded in blockchain transactional data itself, demonstrating that on-chain activity contains predictive information about future price movements. Although the focus is on social media-based sentiment, their approach provides a use-

ful complement that can be incorporated into the model in future research.

3.4 Cross-Modal Analysis for Financial Prediction

Sawhney et al. (2020) pioneered cross-modal analysis combining textual sentiment, temporal patterns, and inter-stock relationships from social media and market data to predict stock movements. Their attentive graph neural network (MAN-SF) demonstrated improved performance when compared to unimodal models.

3.5 Research Gaps Addressed by This Work

This research addresses several gaps in the existing literature:

- Most studies focus on daily or longer timeframes, while this work targets short-term (2-hour) volatility events
- Few studies combine high-frequency Twitter data with minute-level price data
- Limited research exists on distinguishing between normal sentiment fluctuations and anomalous patterns that precede volatility events
- Practical implementation considerations for real-time systems are often overlooked

4 Data Description

This project utilizes two complementary Kaggle datasets that provide comprehensive coverage of Bitcoin-related social media sentiment and high-frequency price movements, enabling robust analysis of sentiment-volatility relationships:

4.1 Bitcoin Tweets Dataset

The foundation of the sentiment analysis comes from the Kaggle "Bitcoin Tweets" dataset collected by alaix14, containing an impressive 16 million tweets spanning January 1, 2016 to March 29, 2019. This comprehensive collection captures Bitcoin-related discussions through keywords, hashtags, and mentions during a particularly fascinating period of Bitcoin's evolution—from relative obscurity to mainstream financial headlines.

What makes this dataset especially valuable is its rich feature set beyond just tweet text. Each entry includes precise timestamps, tweet IDs, user metadata, and engagement metrics like retweets, likes, and replies. This additional social interaction data could prove crucial for understanding which sentiment signals actually resonate with the broader crypto community versus isolated opinions.

The temporal coverage is particularly well-

suited for this research, capturing multiple major Bitcoin market cycles including the dramatic 2017 bull run and subsequent crash. However, this historical focus also presents limitations—the 2016-2019 timeframe means the data predates some significant market developments and may not reflect current social media dynamics.

Preprocessing this dataset will require careful attention to data quality. The research will implement text normalization, duplicate removal, timestamp standardization to UTC, language filtering, and removal of spam/bot accounts. The challenge lies in maintaining signal integrity while filtering out noise—a balance that could significantly impact the final results.

4.2 Bitcoin Historical Price Data

- **Source:** Kaggle "Bitcoin Historical Data" dataset by mczielinski
- **Size:** High-frequency OHLCV (Open, High, Low, Close, Volume) data with 1-minute resolution
- **URL:** <https://www.kaggle.com/datasets/mczielinski/bitcoin-historical-data>
- **Content:** Minute-level and daily Bitcoin price data aggregated from major cryptocurrency exchanges

- **Features:** Timestamp, Open, High, Low, Close prices, Volume, and additional market indicators
- **Exchange Sources:** Data aggregated from leading exchanges including Coinbase, Binance, and other major trading platforms
- **Preprocessing Needs:** Timestamp synchronization with tweet data, missing value interpolation, volatility calculation, and exchange-specific anomaly detection
- **Advantages:** High temporal resolution enabling precise sentiment-price alignment, comprehensive market coverage, reliable exchange-sourced data
- **Limitations:** Potential exchange-specific price discrepancies, occasional data gaps during market disruptions

4.3 Data Alignment and Temporal Synchronization

The datasets will be temporally aligned to enable high-frequency sentiment-volatility analysis using the overlapping period from January 1, 2016 to March 29, 2019. The alignment strategy includes:

- **Temporal Resolution:** Aggregating tweet sentiment into 15-minute windows to match the high-frequency nature of cryptocurrency

markets while maintaining sufficient signal strength

- **Timestamp Standardization:** Converting all timestamps to UTC to ensure precise temporal alignment between social media sentiment and price movements
- **Data Quality Assurance:** Implementing robust filtering mechanisms to remove low-quality tweets, spam, and bot-generated content that could introduce noise into sentiment analysis
- **Missing Data Handling:** Developing interpolation strategies for periods with sparse tweet activity or missing price data to maintain temporal continuity
- **Volatility Event Definition:** Calculating rolling volatility metrics from minute-level price data to identify significant price movements ($\pm 5\%$ within 2-hour windows) that can be correlated with sentiment anomalies

4.4 Dataset Suitability for High-Frequency Analysis

These datasets satisfy the requirements for rigorous data mining analysis by providing:

- **Scale:** 16 million tweets provide sufficient volume for robust statistical analysis and model training

- **Temporal Granularity:** Minute-level price data enables precise detection of short-term volatility events that align with the 15-minute sentiment aggregation windows
- **Historical Depth:** The 2016-2019 timeframe captures multiple Bitcoin market cycles, including significant bull and bear markets, providing diverse market conditions for model validation
- **Feature Richness:** Comprehensive metadata from both datasets enables sophisticated feature engineering for sentiment analysis and volatility prediction
- **Reproducibility:** Publicly available Kaggle datasets ensure research reproducibility and enable comparison with future studies

The combination of high-volume social media data with minute-level price information creates an ideal foundation for investigating the predictive relationship between sentiment anomalies and Bitcoin volatility events within the proposed 2-hour prediction window.

5 Methodology

The approach combines natural language processing for sentiment extraction with time series analysis for anomaly detection—a hybrid strategy that addresses the core challenge of distin-

guishing meaningful signals from social media noise.

5.1 Sentiment Analysis

The core of the sentiment analysis will be conducted using the FinBERT model (Araci, 2019), a Bidirectional Encoder Representations from Transformers (BERT) model fine-tuned on a large corpus of financial text. This choice is critical, as generic sentiment analyzers consistently misinterpret cryptocurrency discussions. When someone tweets "Bitcoin is going to the moon," a standard sentiment analyzer might miss the bullish implication, while FinBERT understands this as strongly positive sentiment. A potential challenge to address is the uneven velocity of tweets, which can lead to periods of sparse data and may require special handling during temporal aggregation.

The financial domain specificity provides crucial advantages: better recognition of trading jargon ("diamond hands," "HODL"), proper interpretation of directional language ("bearish" vs. "bullish"), and nuanced understanding of market-specific expressions that pervade crypto Twitter.

5.2 Temporal Aggregation

Once sentiment scores are obtained for individual tweets, the next challenge becomes tempo-

ral aggregation—how to transform thousands of individual sentiment data points into a coherent time series that can be meaningfully compared with price movements. After considerable experimentation with different window sizes, 15-minute aggregation windows were selected as the optimal balance between signal preservation and noise reduction.

This choice serves multiple purposes: it reduces the noise inherent in individual posts and tweets, creates a consistent time series that aligns naturally with high-frequency price data, and captures rapid sentiment shifts that might precede volatility events without being overwhelmed by minute-to-minute fluctuations.

For each 15-minute window, four key metrics are calculated that together paint a comprehensive picture of sentiment dynamics. The mean sentiment score provides the overall directional bias, while sentiment volume (the number of posts/tweets) indicates the intensity of discussion. Sentiment momentum, calculated as the rate of change between consecutive windows, helps identify accelerating trends. Finally, sentiment dispersion (variance within the window) reveals whether the community is unified in its sentiment or deeply divided—both potentially significant signals for market movements.

5.3 Anomaly Detection

To identify significant sentiment shifts that may precede volatility events, the methodology will:

- Apply Seasonal-Trend decomposition using LOESS (STL) to separate the sentiment time series into seasonal, trend, and residual components
- Calculate z-scores for the residual component to identify statistically significant deviations
- Flag anomalies when z-scores exceed predetermined thresholds

STL hyperparameter tuning will be performed to optimize for the seasonal window length (24 hours to 7 days) and the smoothing parameter for the trend component (0.1 to 0.5). The tuning process will use grid search and 5-fold cross validation to select parameters that best maximize anomaly detection performance. Additionally, hyperparameter tuning will be performed for the z-score threshold for anomaly detection with values ranging from 1.5 to 3.0 standard deviations to balance sensitivity and specificity for volatility event prediction.

The z-score is calculated as:

$$z = \frac{x - \mu}{\sigma} \quad (1)$$

where x is the observed value, μ is the mean, and σ is the standard deviation.

This approach allows distinguishing between normal sentiment patterns (e.g., daily cycles) and potentially market-moving anomalies.

5.4 Volatility Prediction

The validation framework will identify historical sentiment anomalies and compare them against price movements. Specifically, the approach will:

- Define a "volatility event" as a $\pm 5\%$ price move within 2 hours following a sentiment anomaly
- Calculate precision (proportion of sentiment anomalies that preceded actual volatility events) and recall (proportion of volatility events that were preceded by sentiment anomalies)
- Compare performance against baseline models (30-day realized volatility SMA and random signals) using historical data

5.5 Implementation Considerations

The system will be implemented in Python, utilizing:

- HuggingFace Transformers for sentiment analysis
- Pandas and NumPy for data manipulation

- StatsModels for time series analysis
- Scikit-learn for evaluation metrics

The implementation will maintain a modular architecture to allow for:

- Easy substitution of different sentiment analysis models
- Experimentation with various anomaly detection parameters
- Integration of additional data sources in future iterations
- Handling of uneven tweet velocity and data sparsity through techniques like forward-filling or interpolation.

6 Evaluation Plan & Success Criteria

The evaluation framework represents perhaps the most critical aspect of this research—it is where theoretical promise meets practical reality. The evaluation is designed to focus on both statistical performance and practical trading application, recognizing that academic metrics do not always translate to real-world utility.

6.1 Performance Metrics

Four complementary metrics have been chosen that together provide a comprehensive view of

model performance. Precision measures the proportion of sentiment alerts that correctly predict volatility events—essentially, how often the model is correct when it sounds an alarm. Recall captures the proportion of actual volatility events that were predicted by sentiment alerts, answering whether significant market movements are being missed.

The F1 Score provides the harmonic mean of precision and recall, offering a balanced measure that prevents gaming either metric in isolation. However, particular interest lies in the Sharpe Ratio, which measures risk-adjusted returns if the model were used in a simple long/flat trading strategy. This metric bridges the gap between academic performance and practical utility—a model might have impressive F1 scores but still lose money in practice.

The choice to assume a risk-free rate of 0% reflects the current low-interest environment and focuses attention on the model’s ability to generate absolute returns rather than relative performance.

6.2 Experimental Setup

The experimental design reflects one of the most challenging aspects of financial prediction research: how to validate a model’s performance without falling into the trap of hindsight bias. A rolling window approach has been chosen that

trains and tunes the model on historical data up to time T , then evaluates performance on out-of-sample data from time T to $T + \Delta$, before rolling forward and repeating the process.

This methodology serves multiple purposes beyond just preventing look-ahead bias. It simulates the real-world constraints of live trading where only information available at the time of prediction can be used. More importantly, it reveals how model performance degrades over time as market conditions evolve—a critical consideration for any practical implementation.

The choice of window size presents an interesting trade-off: longer training windows provide more data for model stability but may include outdated patterns that no longer reflect current market dynamics. Shorter windows capture recent patterns but may be more susceptible to overfitting. After preliminary testing, a 30-day training window with 7-day evaluation periods has been selected, though experimentation with different configurations will assess sensitivity.

6.3 Statistical Significance Testing

To test the statistical significance of the result, the Diebold-Mariano test will be used to measure the predictive accuracy of the proposed model against the baseline models. This test is specifically designed to determine whether the difference in forecast accuracy between two models is

statistically significant, taking into account serial correlation in the forecast errors.

6.4 Baseline Comparisons

The model will be compared against two baselines:

- **Technical Indicator Baseline:** 30-day realized volatility simple moving average (SMA), which generates alerts when current volatility exceeds the moving average by a predetermined threshold
- **Random Baseline:** Randomly generated alerts with the same frequency as the model's alerts

6.5 Success Thresholds

Based on the project specifications and comparable research, success will be defined as:

- F1 score ≥ 0.50 , indicating balanced performance in terms of precision and recall
- Sharpe ratio > 0 , demonstrating positive risk-adjusted returns
- Statistically significant outperformance of both baseline models

7 Timeline

The project will follow a 6-week timeline outlined in Table 1.

This timeline provides sufficient time for iterative development and thorough validation of the proposed approach. The use of Kaggle datasets eliminates potential API access issues and ensures reproducible results.

8 Future Work: Real-Time Implementation

If this historical validation proves successful, the natural next step would be building a live system. This project explores sentiment-based market prediction techniques in cryptocurrency markets. The transition from retrospective analysis to real-time prediction introduces new challenges, but also represents the culmination of years of investigation into sentiment-based market prediction.

The technical architecture would need to handle live data streaming from Twitter API v2, processing thousands of posts per minute while maintaining the 15-minute aggregation windows that proved effective in historical testing. Apache Kafka could be explored for the streaming pipeline, though the real challenge will be managing the inevitable API rate limits and data quality issues that do not appear in clean Kaggle datasets.

A notable consideration for real-time implementation is the potential feedback loop,

wherein widespread adoption of sentiment-based trading strategies may alter the very sentiment patterns being analyzed. This phenomenon could lead to evolving market dynamics, potentially diminishing the effectiveness of predictive models over time—a challenge that may be particularly salient in cryptocurrency markets.

The system would also need robust risk management controls—too many algorithmic trading disasters have occurred to approach this casually. Configurable alert thresholds, automatic circuit breakers, and continuous model performance monitoring would be essential safeguards. But perhaps most importantly, this validation study will help determine whether the computational and infrastructure investment required for real-time implementation is justified, and whether sentiment-based prediction can be effectively applied in the cryptocurrency domain.

References

- Araci, D. (2019). FinBERT: Financial sentiment analysis with pre-trained language models. *arXiv preprint arXiv:1908.10063*.
- Brauneis, A., & Sahiner, M. (2024). Cryptocurrency volatility forecasting: Mounting a HAR, sentiment, and machine learning horserace. *Asia-Pacific Financial Markets*. <https://doi.org/10.1007/s10690-024-09510-6>

- Kleitsikas, C., Korfiatis, N., Leonardos, S., & Ventre, C. (2025). Bitcoin's edge: Embedded sentiment in blockchain transactional data (arXiv Preprint No. 2504.13598). *arXiv*. <https://arxiv.org/abs/2504.13598>
- Long, S. C., Xie, Y., Zhou, Z., Lucey, B., & Urquhart, A. (2025). From whales to waves: Social media sentiment, volatility, and whales in cryptocurrency markets. *The British Accounting Review*. <https://doi.org/10.1016/j.bar.2025.101682>
- Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using Siamese BERT-networks. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics.
- Sawhney, R., Agarwal, S., Wadhwa, A., & Shah, R. (2020). Deep attentive learning for stock movement prediction from social media text and company correlations. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 8415–8426.
- Youssfi Noura, A., Bouchakwa, M., & Jamoussi, Y. (2023). Bitcoin price prediction considering sentiment analysis on Twitter and Google News. In *Proceedings of the 27th International Database Engineering & Applications Symposium* (pp. 71–78). Association for Computing Machinery. <https://doi.org/10.1145/3589462.3589494>

Table 1: Project Timeline

Week	Tasks	Deliverables
1	<ul style="list-style-type: none"> - Download Kaggle cryptocurrency tweets dataset - Download Bitcoin price data from Kaggle - Perform initial data exploration and cleaning - Conduct exploratory data analysis 	<ul style="list-style-type: none"> - Clean, aligned datasets - Initial data visualizations - Data quality assessment report
2-3	<ul style="list-style-type: none"> - Implement sentiment analysis pipeline using FinBERT - Develop temporal aggregation methodology - Create feature engineering framework - Build prototype volatility prediction models 	<ul style="list-style-type: none"> - Sentiment extraction code - Feature engineering pipeline - Initial model prototypes - Preliminary performance metrics
4	<ul style="list-style-type: none"> - Tune STL and z-score hyperparameters - Implement cross-validation framework - Conduct backtesting on historical data - Compare against baseline models 	<ul style="list-style-type: none"> - Optimized model parameters - Backtesting results - Comparative performance analysis
5	<ul style="list-style-type: none"> - Analyze results and identify patterns - Create visualizations of key findings - Assess model limitations and strengths - Document lessons learned 	<ul style="list-style-type: none"> - Detailed results analysis - Performance visualizations - Model evaluation report
6	<ul style="list-style-type: none"> - Draft final report - Create presentation slides - Review and refine all deliverables - Prepare for final submission 	<ul style="list-style-type: none"> - Complete written report - Presentation slide deck - All supporting code and documentation