Cryptocurrency Stock Price Movement Detection using Sentiment and Persuasion Analysis

Team Members: Ankit Chaurasia, Janvi Lakhani, Komal Wavhal, Prasoon Parashar

Abstract—**This project develops a Real-Time Cryptocurrency Trends Analysis Platform that leverages sentiment analysis from social media and news articles to predict cryptocurrency price movements. By utilizing various machine learning models and natural language processing (NLP) techniques, the platform aims to provide real-time insights into the potential market impact of social media trends. This work integrates data sources such as Google News, Yahoo Finance, and Twitter datasets, employing sentiment and persuasion detection methods to analyze textual data. Our platform's goal is to help investors and traders make informed decisions based on current market sentiment.**

Index Terms—**Cryptocurrency, Web Scrapping, Sentiment Analysis, Persuasion Analysis, Classification**

I. Introduction

The cryptocurrency market is known for its high volatility, with prices often reacting more to public sentiment and social media trends than traditional economic indicators. Unlike conventional financial markets, where price changes are generally driven by corporate performance and economic factors, cryptocurrencies are highly susceptible to social media influences from platforms such as Twitter, Reddit, and news outlets. This dynamic suggests that not only general sentiment but also the persuasiveness of language and influence of prominent profiles play a role in shaping investor decisions, leading to rapid price fluctuations.

A notable example of this phenomenon is the impact of Elon Musk's tweets on Dogecoin. In early 2021, Musk's tweets featuring playful and positive mentions of Dogecoin included highly persuasive language that resonated with his vast follower base. This led to an immediate surge in interest and a rapid increase in Dogecoin's value, exemplifying how social media-driven sentiment and influential messaging can significantly affect cryptocurrency prices.

The project, aims to investigate this nuanced relationship by analyzing the impact of general sentiment and persuasive language from a range of social media posts on cryptocurrency price movements. Specifically, the platform will examine sentiment shifts across two critical timeframes: before market openings and after market closures. This dual analysis enables us to evaluate whether specific influencers or persuasive messages hold greater sway during these windows and thereby impact trading behaviors and price trends.

II. Literature Review

The cryptocurrency market has experienced rapid growth, attracting the attention of investors, traders, and researchers. Its unique dynamics, driven by decentralization and the influence of online platforms, have made it a fertile ground for financial forecasting using innovative techniques like sentiment analysis [2]. Social media platforms and news outlets serve as significant sources of information that can influence market trends and investor behavior. While significant progress has been made in understanding these relationships, gaps remain in exploring the combined impact of multiple data sources and advanced analytical techniques.

Sentiment analysis has emerged as a critical tool in predicting cryptocurrency price movements. Studies have demonstrated that platforms like Twitter and Reddit provide valuable data reflecting public sentiment, which can influence price dynamics (Garcia & Schweitzer, 2015; Phillips & Gorse, 2017). Recent work by Loginova et al. (2021) introduced aspect-based sentiment analysis, leveraging fine-grained topic-sentiment features to improve predictive accuracy. This approach incorporated textual data from forums like Reddit, Bitcointalk, and CryptoCompare, highlighting the potential of combining topical and sentiment features for interpretable and effective prediction models.

Cryptocurrency-related information is disseminated across various platforms, each offering unique data characteristics. News articles, for instance, provide in-depth and structured content, while social media platforms reflect immediate public sentiment (Schoen et al., 2013). The inclusion of topic modeling techniques, such as dynamic topic modeling (Phillips & Gorse, 2018b), has shown promise in identifying market-relevant discussions over time. However, as Loginova et al. (2021) pointed out, the combination of data sources like news articles, forums, and social media has not been fully explored, leaving questions about their cumulative impact on predictive performance.

Machine learning techniques have been extensively applied to financial markets, including cryptocurrency forecasting. Prior studies primarily focused on Bitcoin, using historical time series data and basic sentiment analysis (Guo & Antulov-Fantulin, 2018; Amjad & Shah, 2017). Loginova et al. (2021) extended this research by integrating advanced topic-sentiment models, such as JST and TS-LDA, demonstrating improved predictive performance for directional returns. However, there is a lack of comprehensive exploration of alternative cryptocurrencies and their sensitivity to sentiment across diverse data sources.

- Research Gap:
  While previous studies have shown the effectiveness of sentiment analysis in predicting cryptocurrency price movements, they primarily focus on individual platforms (e.g., Twitter or Reddit) and a limited set of cryptocurrencies. The role of news articles remains under-explored, and the integration of multiple data sources for holistic analysis is still in its infancy. Additionally, most studies emphasize predictive accuracy but fail to address the practical implications for investors and traders in real-time decision-making.

- Contribution:
  This project contributes to the field of cryptocurrency market analysis by developing a Real-Time Cryptocurrency Trends Analysis Platform that combines sentiment and persuasion analysis from multiple social media sources with financial data to predict cryptocurrency price movements. The key contributions of this project are:

  1. Multi-source Data Integration: The platform integrates sentiment and persuasion analysis from various social media platforms such as Google News, HuggingFace (Twitter Datasets), and News API, alongside stock numerical data (finance data). This approach allows for a comprehensive analysis of how sentiment, persuasion, and market data together influence cryptocurrency prices.

  2. Real-Time Decision-Making Tool: By providing real-time insights, the platform enables investors, traders, and market analysts to make informed decisions based on the latest social media trends and market data. This addresses the need for timely decision-making in the fast-paced cryptocurrency market.

  3. Advanced Sentiment and Persuasion Analysis: The platform implements both sentiment analysis and persuasion analysis to understand not just the mood of the public but also the persuasive impact of social media messages on market behavior. This dual approach enhances the accuracy and depth of the predictions.

  4. Machine Learning for Future Price Predictions: Different machine learning models are used to predict future cryptocurrency prices based on historical data. These models leverage both sentiment and persuasion signals, along with numerical financial data, to make predictions that help users anticipate market movements.

III. MOTIVATION

The idea of choosing the cryptocurrency topic is because the cryptocurrency market is known for its big price changes and unpredictability, which are mostly influenced by social media and people's feelings, rather than usual economic factors and company performance.

Cryptocurrencies are greatly affected by discussions and trends on sites like Twitter, Reddit, and news from around the world. These platforms play an important role in shaping investor opinions and trading behaviors, which leads to swift price changes. This dynamic offers an intriguing opportunity to examine the relationship between social media sentiment and market fluctuations.

This project aims to develop a Real Time Crypto-currency Trends Analysis Platform that analyzes the impact of social media sentiment on cryptocurrency price movements. In today's word the data is the key, and we can get the data from multiple resources like – social media news, articles and so on.

The platform offers investors, traders, and market analyze real-time insights into the potential impact of social media trends on the cryptocurrency market, giving them a tool for informed decision-making. The volatility of the cryptocurrency market presents both significant risks and opportunities for investors and traders. The rapid price swings, often triggered by viral social media posts or trending discussions, create challenges for those trying to make informed decisions based on traditional market analysis. Social media platforms play an outsized role in shaping public opinion and investor behavior, making them key factors influencing price fluctuations. The motivation behind this project is to explore and quantify the relationship between social media sentiment and cryptocurrency prices. By developing a tool that can predict price movements based on social media discussions, we aim to empower traders with real-time insights that can improve their investment strategies, minimize risks, and take advantage of short-term price movements influenced by online trends.

IV. METHODOLOGY

Our project is designed as an integrated system that combines data collection, sentiment analysis, persuasion analysis and machine learning to predict cryptocurrency price movements. The platform uses a graphical user interface (GUI) built with Tkinter for ease of use and interaction. Data is sourced from multiple platforms, including Google News, Yahoo Finance, Twitter, and News API, through web scraping and API calls. Sentiment and persuasion analysis are performed using advanced natural language processing (NLP) techniques, such as VADER, FinBERT, and GPT-4o-mini. Machine learning models, including Random Forest,

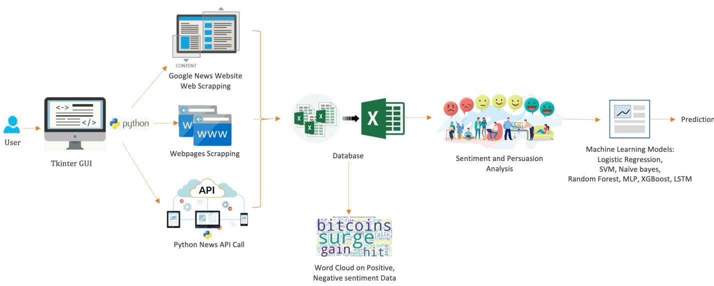

Fig 1. Project Framework

SVM, MLP, etc. are employed to predict price trends based on historical data and real-time sentiment. The framework is designed to provide investors and traders with actionable insights and forecasts by leveraging the power of data analytics and AI.

A. Data Collection

For this project, data was collected from multiple sources to analyze the impact of social media sentiment and news on cryptocurrency price movements. The data spans from 2014 to 2024 and includes sentiment analysis from various platforms and historical price data.

1. Google News:

- Source: Cryptocurrency-related news articles are scraped using Pandas, Selenium, and VADER Sentiment Analysis.
- Data Collected: Articles are scraped based on specific keyword queries and a date range (2014-2024) [1].
- Purpose: This data helps analyze news sentiment related to cryptocurrency and its potential impact on market trends.

2. Stock Numerical Data (Finance Data):

- Source: Historical cryptocurrency prices are retrieved from Yahoo Finance using Pandas and Selenium.
- Data Collected: Historical prices, including daily opening and closing prices, highs, lows, and volume, for Bitcoin from 2014 to 2024 [2].
- Purpose: This data is used to track the price movements of Bitcoin and to align with sentiment analysis for a comprehensive market trend analysis.

3. Twitter Data:

- Source: Financial tweet datasets related to cryptocurrency are accessed from Hugging Face.
- Data Collected: Historical tweets related to cryptocurrency, including sentiment data, are fetched for the analysis.
- Sentiment Analysis: Pandas, NLTK, and the NLTK Sentiment Analysis Toolkit are used for data processing and sentiment analysis.
- Purpose: This data helps track public sentiment from Twitter discussions and its potential influence on cryptocurrency price movements.

4. News API:

- Source: Real-time news coverage and articles related to cryptocurrencies are aggregated using the News API.
- Data Collected: Articles on specific cryptocurrency topics are retrieved, with sentiment analysis performed using tools like the Sentiment Intensity Analyzer from the NLTK library.
- Purpose: The aggregated news articles are analyzed for sentiment to assess how public perception and news coverage influence cryptocurrency markets over time.

This multi-source data collection approach allows for a comprehensive understanding of the factors influencing cryptocurrency price movements, with a specific focus on sentiment analysis and its correlation with market trends.

B. Exploratory Data Analysis (EDA)

The EDA aims to uncover meaningful insights and patterns within the cryptocurrency dataset, spanning various cryptocurrencies and stock data from 2012 to 2024. Our analysis focuses on descriptive statistics, trends, and visualizations to understand the data better and inform subsequent analytical stages.

Dataset Overview

The analysis is based on a consolidated dataset comprising 26,500 rows of cryptocurrency-related titles and 20,000 rows of stock data spanning the years 2012 to 2024. Key components of the dataset include:

Cryptocurrency	Number of Articles
Bitcoin (BTC)	11000
Tether (USDT)	6500
Binance (BNB)	4500
Ethereal (ETH)	4500

Table 1. Cryptocurrencies VS Number of Articles

Cryptocurrency Data: Titles, dates, and types for major cryptocurrencies:

Stock Data: Daily price metrics (open, close, high, low, and volume) primarily for Bitcoin.

	Date	Open	High	Low	Close	Adj Close	Volume	Name
0	Nov 2, 2024	69,476.78	69,855.11	69,476.78	69,551.83	69,551.83	44,786,638,848	BTC
1	Oct 31, 2024	72,335.05	72,662.31	69,590.50	70,215.19	70,215.19	40,627,912,076	BTC
2	Oct 30, 2024	72,715.37	72,905.30	71,411.73	72,339.54	72,339.54	40,646,637,831	BTC
3	Oct 29, 2024	69,910.05	73,577.21	69,729.91	72,720.49	72,720.49	58,541,874,402	BTC
4	Oct 28, 2024	67,922.67	70,212.27	67,535.13	69,907.76	69,907.76	38,799,856,657	BTC

Fig 2. BTC's Stocks Data

Key Analytical Components

1. Daily Return Analysis:

Daily returns represent the percentage change in the closing price from one day to the next, providing critical insights into volatility and risk.

The analysis revealed periods of heightened fluctuations, indicating volatile market conditions, alongside periods of relative stability.

These patterns are crucial for understanding the risk profile and predicting price trends for cryptocurrencies.

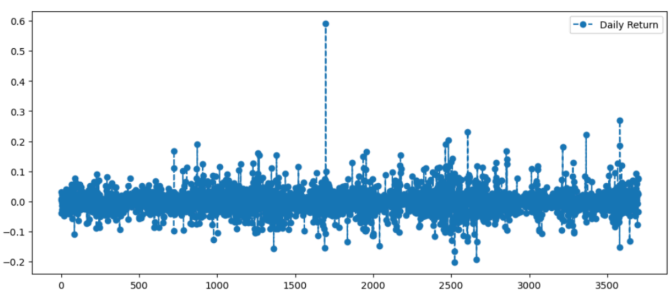

Fig 3. BTC's Daily Return

2. Moving Average (MA) Analysis:

We computed 10-day, 20-day, and 50-day moving averages (MAs) to smooth short-term price variations and observe long-term trends.

The comparison between closing prices and MAs revealed market trends, potential support, and resistance levels.

The 10-day MA highlighted short-term trends, while the 50-day MA provided a broader perspective on market movements.

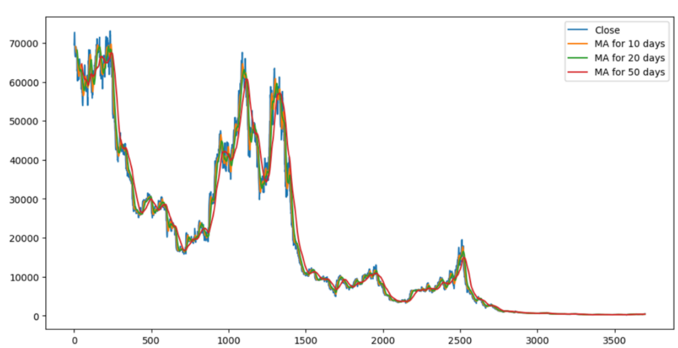

Fig 4. BTC's Moving Average

3. Title Length Distribution:

Titles predominantly ranged around 60 characters, indicating a focus on concise yet descriptive headlines.

By analyzing title lengths and removing common stop words, we identified clusters of domain-specific terms, reflecting key topics and themes in the cryptocurrency space.

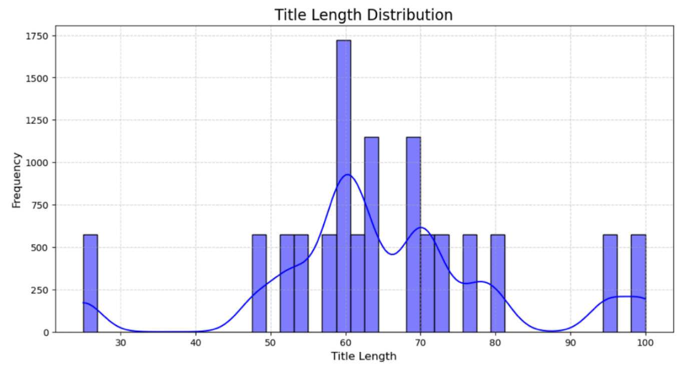

Fig 5. Article Title Length Distribution

C. Sentiment Analysis

To assess the correlation between the sentiment of news and social media articles and cryptocurrency stock price movements, we employed advanced Natural Language Processing (NLP) techniques. This section details the data preprocessing steps and the sentiment analysis methods used to quantify sentiment and its potential impact on market behavior.

Sentiment Scores per Article:

Our sentiment analysis resulted in a structured output for each article, providing valuable insights into its emotional tone:

JSON file:

```
{
// Sentiment category
    'class': 'positive/negative/neutral',

 // Probability of belonging to a sentiment class
    'score': 0.0 to 1.0 (probability),

 // Degree of positivity or negativity (-ve to +ve)
    'polarity': -1.0 to 1.0,

// Level of subjectivity (0 = objective, 1 = subjective)
    'subjectivity': 0.0 to 1.0
}
```

Sentiment Analysis Techniques:

We explored two primary methodologies for sentiment analysis, each offering unique advantages:

1. FinBERT-tone + TextBlob:

- *FinBERT-tone:* This deep learning model, available on Hugging Face, leverages pre-training and fine-tuning specifically for sentiment analysis within the financial sector. FinBERT-tone provided a strong foundation for capturing the sentiment of cryptocurrency-related news articles.
- *TextBlob:* We employed TextBlob, a popular NLP library, to complement FinBERT-tone by providing additional polarity and subjectivity scores. This comprehensive approach enhanced the granularity of our sentiment analysis.
- Baseline Scores: To evaluate the effectiveness of this approach, we tested it on the cryptocurrency news dataset from Kaggle. The FinBERT-tone + TextBlob combination achieved an accuracy of 80% in classifying sentiment categories.

2. GPT-4o-mini:

- OpenAI's GPT-4o-mini: To explore alternative approaches and potentially improve accuracy, we experimented with OpenAI's GPT-4o-mini model. This powerful large language model was leveraged within a few-shot prompting framework to extract sentiment scores directly from article titles.
- Comparison and Potential Application: The aim of this experiment was to compare the performance of GPT-4o-mini with FinBERT-tone + TextBlob and identify the most suitable methodology for our application.
- Baseline Scores: Upon testing GPT-4o-mini on the same Kaggle dataset, it achieved an accuracy of 84% in sentiment classification.

- Prompt used for GPT-4o-mini:

  "You are a machine learning bot with expertise in financial news, stock market trends, political science, cryptocurrencies, and international relations that impact stock prices.

  Analyze the following news title and identify its sentiment.

  Provide the result in the specified format, which includes:
  Class: Sentiment category (positive, negative, or neutral).
  Score: Sentiment confidence score as a float (0 to 1).
  Polarity: Sentiment polarity ranging from -1 (negative) to 1 (positive).
  Subjectivity: Subjectivity score between 0 (objective) and 1 (highly subjective).

  Input and Output constraints:
  Input: title: string
  Output: A JSON object with the following structure:
  'class': <sentiment>, 'score': <float>, 'polarity': <float>, 'subjectivity': <float>

  Notes:
  Always provide output strictly in the JSON format specified above.
  Do not include any explanations, comments, or text outside of the JSON object in your response."

These results indicate that GPT-4o-mini offers a slight improvement in accuracy compared to FinBERT-tone + TextBlob. However, the choice of the optimal approach may depend on factors such as computational resources and the specific requirements of the application.

	Title	Date	Open	High	Low	Close	Adj Close	Volume	class	score	polarity	subjectivity
0	NaN	2024-11-02	69,476.78	69,855.11	69,476.78	69,551.83	69,551.83	44,786,638,848	neutral	0.967353	0.000000	0.000000
1	MicroStrategy Announces $42 Billion Bitcoin In...	2024-10-31	72,335.05	72,662.31	69,590.50	70,215.19	70,215.19	40,627,912,076	neutral	0.998327	0.000000	0.000000
2	BlackRock CEO Issues Stark Fed Warning Amid $3...	2024-10-31	72,335.05	72,662.31	69,590.50	70,215.19	70,215.19	40,627,912,076	negative	0.555574	-0.200000	0.600000
3	MicroStrategy Announces Third Quarter 2024 Fin...	2024-10-30	72,715.37	72,905.30	71,411.73	72,339.54	72,339.54	40,646,637,831	neutral	0.999553	0.000000	0.000000
4	Bitcoin (BTC/USD) Takes a Pause, $75k Still On...	2024-10-30	72,715.37	72,905.30	71,411.73	72,339.54	72,339.54	40,646,637,831	negative	0.966950	0.000000	0.000000
5	Bitcoin teases all-time high days from preside...	2024-10-30	72,715.37	72,905.30	71,411.73	72,339.54	72,339.54	40,646,637,831	neutral	0.999959	0.160000	0.540000
6	Election optimism pushes bitcoin near new high...	2024-10-30	72,715.37	72,905.30	71,411.73	72,339.54	72,339.54	40,646,637,831	positive	0.999777	0.132121	0.464848
7	Bitcoin ETFs Record $870M Inflows as BTC Flirt...	2024-10-30	72,715.37	72,905.30	71,411.73	72,339.54	72,339.54	40,646,637,831	positive	1.000000	0.000000	0.000000
8	Michael Saylor's MicroStrategy Plans to Raise ...	2024-10-30	72,715.37	72,905.30	71,411.73	72,339.54	72,339.54	40,646,637,831	neutral	0.999910	0.250000	0.250000
9	World's Lowest Fee Bitcoin and Ether ETPs (Tic...	2024-10-29	69,910.05	73,577.21	69,729.91	72,720.49	72,720.49	58,541,874,402	neutral	0.998083	0.000000	0.000000
10	Bitcoin Prices Break Past $70,000 To Reach Hig...	2024-10-29	69,910.05	73,577.21	69,729.91	72,720.49	72,720.49	58,541,874,402	positive	0.999798	-0.250000	0.250000
11	This Japanese Company Is Playing The Michael S...	2024-10-29	69,910.05	73,577.21	69,729.91	72,720.49	72,720.49	58,541,874,402	neutral	0.999968	0.000000	0.000000
12	Bitcoin Tops $73,000—A 7-Month High—As Electio...	2024-10-29	69,910.05	73,577.21	69,729.91	72,720.49	72,720.49	58,541,874,402	neutral	0.999010	0.000000	0.000000
13	First Mover Americas: BTC Jumps Above $71K, DO...	2024-10-29	69,910.05	73,577.21	69,729.91	72,720.49	72,720.49	58,541,874,402	positive	1.000000	0.125000	0.216667
14	How Do CPI Announcements Affect Bitcoin Price?	2024-10-29	69,910.05	73,577.21	69,729.91	72,720.49	72,720.49	58,541,874,402	neutral	0.999119	0.000000	0.000000
15	Should You Trade Options on Spot Bitcoin ETFs?	2024-10-28	67,922.67	70,212.27	67,535.13	69,907.76	69,907.76	38,799,856,657	neutral	0.999983	0.000000	0.000000
16	Bitcoin May Shift To A Super Cycle, Breaking T...	2024-10-28	67,922.67	70,212.27	67,535.13	69,907.76	69,907.76	38,799,856,657	neutral	0.999981	0.333333	0.666667
17	Bitcoin Is on the Verge of Topping Record High...	2024-10-28	67,922.67	70,212.27	67,535.13	69,907.76	69,907.76	38,799,856,657	positive	0.992436	0.160000	0.540000
18	BTC and the GOP: How Trump learned to stop wor...	2024-10-27	67,023.48	68,221.31	66,847.23	67,929.30	67,929.30	16,721,307,878	neutral	0.968642	0.500000	0.600000
19	Tesla Bitcoin Moves May No Longer Shake The Ma...	2024-10-26	66,628.73	67,317.92	66,360.59	67,014.70	67,014.70	19,588,098,156	neutral	0.819006	0.000000	0.000000

Fig 6. Sentiment Analysis Done for BTC

D. Persuasion Analysis

To assess the use of persuasive language in news and social media articles and its potential influence on cryptocurrency price movements, we developed a scoring system to quantify how likely persuasive techniques are used in each text.
In order to attain these persuasion scores, we made use of GPT-4o-mini along with official persuasion techniques documentation from SemEval [1]

GPT-4o-mini :
- Made use of few-shot prompting by providing GPT-4o-mini with definitions and examples of all 23 techniques to ensure accurate scoring and detection. The LLM gave a score from (0-1) that defines how likely it is for a technique out of the mentioned 23 techniques to be present in the given texts
- The given techniques were: Appeal to Authority, Fear-Prejudice, Loaded Language, Repetition etc.

Example Usage of the Persuasion module:

Article title: 'HOW TO EARN $10 TO $20 DAILY ON BINANCE AS A BEGINNER😎 👇🏻'
Persuasion Score: 0.9

Prompt used for GPT-4o-mini:
"You are a machine learning bot that has background in financial news, stock news, political science, cryptocurrencies, and international relations that can affect stock prices
Analyze the following news title and identify the persuasion techniques used from the list provided below. Give a confidence score (on a scale of 0 to 1) how likely

it is to be using one of the persuasion techniques mentioned below.
The confidence metric should indicate how confident you are that the identified technique is present in the title.
Consider the specific clues related to each technique to aid in your analysis.

Possible Persuasion Techniques:
1. Appeal_to_Authority: Cites authority to support a conclusion.
2. Appeal_to_Popularity: Cites popularity or majority support to back a claim.
3. Appeal_to_Values: Invokes widely shared values.
4. Appeal_to_Fear-Prejudice: Uses fear or prejudice to reject or promote an idea.
5. Flag_Waving: Refers to patriotism or group allegiance.
6. Causal_Oversimplification: Oversimplifies the causes of an issue.
7. False_Dilemma-No_Choice: Implies only two options when there may be more.
8. Consequential_Oversimplification: Oversimplifies the consequences of accepting a proposition.
9. Straw_Man: Misrepresents someone's position to make it easier to attack.
10. Red_Herring: Diverts attention from the main topic.
11. Whataboutism: Distracts from the topic by charging an opponent with hypocrisy.
12. Slogans: Uses a brief, catchy phrase to encapsulate a message.
13. Appeal_to_Time: Suggests that the time is ripe for a certain action.
14. Conversation_Killer: Discourages critical thought or discussion.
15. Loaded_Language: Uses emotionally charged words or phrases.
16. Repetition: Repeatedly reinforces the same idea.
17. Exaggeration-Minimisation: Downplays or exaggerates a subject.
18. Obfuscation-Vagueness-Confusion: Deliberately unclear, leaving room for varied interpretations.
19. Name_Calling-Labeling: Employs demeaning labels.
20. Doubt: Undermines credibility by questioning character or attributes.
21. Guilt_by_Association: Discredits by associating with a negatively viewed group.
22. Appeal_to_Hypocrisy: Accuses the target of hypocrisy.
23. Questioning_the_Reputation: Undermines the reputation of the target.

Input and Output constraints:
Input: text: string
Output: confidence_score: float ”

	Title	Date	Open	High	Low	Close	Adj Close	Volume	persuasion_score
0	NaN	2024-11-02	69,476.78	69,855.11	69,476.78	69,551.83	69,551.83	44,786,638,848	0.00
1	MicroStrategy Announces $42 Billion Bitcoin In...	2024-10-31	72,335.05	72,662.31	69,590.50	70,215.19	70,215.19	40,627,912,076	0.90
2	BlackRock CEO Issues Stark Fed Warning Amid $3...	2024-10-31	72,335.05	72,662.31	69,590.50	70,215.19	70,215.19	40,627,912,076	0.90
3	MicroStrategy Announces Third Quarter 2024 Fin...	2024-10-30	72,715.37	72,905.30	71,411.73	72,339.54	72,339.54	40,646,637,831	0.70
4	Bitcoin (BTC/USD) Takes a Pause, $75k Still On...	2024-10-30	72,715.37	72,905.30	71,411.73	72,339.54	72,339.54	40,646,637,831	0.60
5	Bitcoin teases all-time high days from preside...	2024-10-30	72,715.37	72,905.30	71,411.73	72,339.54	72,339.54	40,646,637,831	0.90
6	Election optimism pushes bitcoin near new high...	2024-10-30	72,715.37	72,905.30	71,411.73	72,339.54	72,339.54	40,646,637,831	0.60
7	Bitcoin ETFs Record $870M Inflows as BTC Flirt...	2024-10-30	72,715.37	72,905.30	71,411.73	72,339.54	72,339.54	40,646,637,831	0.60
8	Michael Saylor's MicroStrategy Plans to Raise ...	2024-10-30	72,715.37	72,905.30	71,411.73	72,339.54	72,339.54	40,646,637,831	0.70
9	World’s Lowest Fee Bitcoin and Ether ETPs (Tic...	2024-10-29	69,910.05	73,577.21	69,729.91	72,720.49	72,720.49	58,541,874,402	0.30
10	Bitcoin Prices Break Past $70,000 To Reach Hig...	2024-10-29	69,910.05	73,577.21	69,729.91	72,720.49	72,720.49	58,541,874,402	0.75
11	This Japanese Company Is Playing The Michael S...	2024-10-29	69,910.05	73,577.21	69,729.91	72,720.49	72,720.49	58,541,874,402	0.70
12	Bitcoin Tops $73,000—A 7-Month High—As Electio...	2024-10-29	69,910.05	73,577.21	69,729.91	72,720.49	72,720.49	58,541,874,402	0.60
13	First Mover Americas: BTC Jumps Above $71K, DO...	2024-10-29	69,910.05	73,577.21	69,729.91	72,720.49	72,720.49	58,541,874,402	0.20
14	How Do CPI Announcements Affect Bitcoin Price?	2024-10-29	69,910.05	73,577.21	69,729.91	72,720.49	72,720.49	58,541,874,402	0.60
15	Should You Trade Options on Spot Bitcoin ETFs?	2024-10-28	67,922.67	70,212.27	67,535.13	69,907.76	69,907.76	38,799,856,657	0.90
16	Bitcoin May Shift To A Super Cycle, Breaking T...	2024-10-28	67,922.67	70,212.27	67,535.13	69,907.76	69,907.76	38,799,856,657	0.60
17	Bitcoin Is on the Verge of Topping Record High...	2024-10-28	67,922.67	70,212.27	67,535.13	69,907.76	69,907.76	38,799,856,657	0.70
18	BTC and the GOP: How Trump learned to stop wor...	2024-10-27	67,023.48	68,221.31	66,847.23	67,929.30	67,929.30	16,721,307,878	0.90
19	Tesla Bitcoin Moves May No Longer Shake The Ma...	2024-10-26	66,628.73	67,317.92	66,360.59	67,014.70	67,014.70	19,588,098,156	0.60

Fig 7. Persuasion Analysis Done for BTC

E. Feature Engineering

Apart from the stocks numeric data and the analysis derived from the sentiment and persuasion analysis, we derived a couple of new features for enhanced predictions.

1. Target Variable (Price Direction):
- A binary variable indicating the direction of price movement compared to the previous day.
- A value of 1 signifies an increase in price, while a value of 0 indicates a decrease or no change.

Cryptocurrency	No of. Price_Direction = 0	No of. Price_Direction = 1
BTC	1748	1951
ETH	1246	1304
USDT	1308	1242
BNB	1236	1314

Table 2. Distribution of Target Variable VS Cryptocurrency

2. Lag Features:
- Historical values (lags) of key variables such as sentiment scores (score, polarity, subjectivity) and persuasion scores.
- Incorporating past behavior helps capture time-series dependencies in financial markets.

3. Rolling Averages:
- Smoothed averages of features over defined time windows (e.g., 3 or 5 days).
- Reduces noise in the data and provides insights into short- and medium-term sentiment shifts influencing market behavior.

4. Daily Return:
- Calculates the percentage change in the adjusted closing price compared to the previous day.
- Captures the magnitude and direction of price movement on a daily basis.

5. Volatility:
- A statistical measure of price fluctuations over a given period (e.g., 5 days).

- High volatility often indicates uncertainty or significant market movement, providing valuable information for predicting price direction.

F. Model Building

Several machine learning models were implemented and evaluated for binary classification (price movement prediction). Each model was tested on the same dataset to ensure a fair comparison. The models include:

- Logistic Regression: A linear model that provides baseline predictions for binary outcomes.
- Support Vector Machine (SVM): Separates data into classes using an optimal hyperplane, effective for linear and non-linear separations.
- Naïve Bayes: A probabilistic classifier assuming feature independence, commonly used for text classification tasks.
- Random Forest Classifier: An ensemble method that combines multiple decision trees to enhance predictive accuracy and reduce overfitting.
- Multi-Layer Perceptron (MLP): A neural network capable of capturing complex, non-linear patterns.
- XGBoost: A gradient-boosting framework known for its speed and accuracy.
- Long Short-Term Memory (LSTM): A type of recurrent neural network designed to handle sequential data and long-range dependencies, making it well-suited for time series forecasting.

The models were evaluated using accuracy and ROC-AUC scores, focusing on their ability to correctly predict the price movement's direction. The results are summarized below:

Model	Accuracy	Max ROC-AUC score
Logistic Regression	0.57	0.58
SVM	0.57	0.41
Naïve Bayes	0.50	0.53
Random Forest	**0.64**	**0.65**
Multilayer Perceptron (MLP)	0.59	0.62
XGBoost	0.54	0.56
LSTM	0.53	0.53

Table 3. Performance of different ML Models

The Random Forest Classifier outperformed all other models in terms of both accuracy and ROC-AUC, making it the best-performing model for this task.

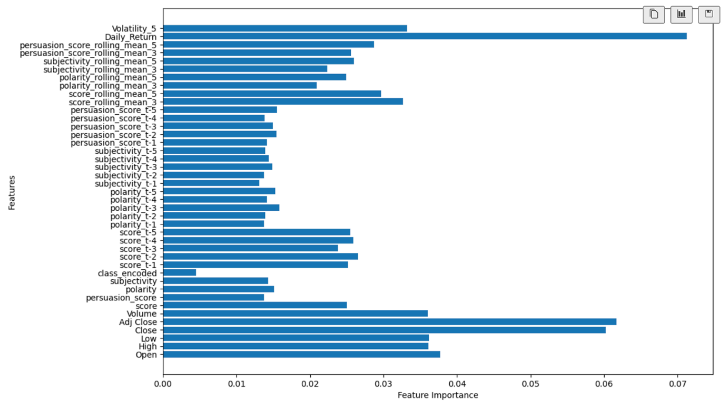

Fig 8. Feature Importance in Random Forest

The Random Forest Classifier's feature importance analysis provided key insights into the most influential factors in predicting price movements:

- Sentiment Polarity: Polarity scores reflecting the positive or negative tone of news and tweets were among the top predictors.
- Persuasion Scores: Scores quantifying the likelihood of persuasive techniques in text played a critical role, indicating a strong correlation between market sentiment and persuasive content.
- Lag Features: Historical values of sentiment and persuasion scores were significant, demonstrating that past trends heavily influence future market behavior.
- Rolling Averages: Smoothed averages over 3- and 5-day windows helped reduce noise and provided a clearer view of sentiment shifts.
- Daily Return: Percentage changes in closing prices showed a strong predictive relationship with price direction.

V. Comparison With Previous Work

Our findings show a marked improvement over existing studies in the domain:

- The study [2] reported a maximum ROC-AUC score of 0.58 using MLP, while our Random Forest Classifier achieved a ROC-AUC of 0.65, representing a 12% improvement.
- In comparison to [3], which achieved an accuracy of 0.61, our model achieved an accuracy of 0.64, a 5% enhancement.
- Our integration of persuasion detection, which quantifies the use of persuasive language in market-related texts, provided an innovative angle not explored in prior studies, further improving prediction reliability.

VI. CONCLUSION

Our study highlights the significant influence of public sentiment and persuasive messaging in cryptocurrency markets. By combining sentiment analysis, machine learning models, and persuasion detection techniques, we developed a robust framework that advances prediction accuracy over prior works. This contribution equips traders, researchers, and market participants with more informed decision-making tools amidst the volatile and unpredictable cryptocurrency landscape. The proposed approach establishes a foundation for building real-time systems, driving further innovation at the intersection of sentiment analysis and financial forecasting.

It delved into the intricate relationship between cryptocurrency price movements, public sentiment, and persuasive language techniques. By analyzing a substantial dataset of 26,500 Bitcoin-related news articles and 20,000+ stock price records, we uncovered key insights:

- Data Analysis:

  We identified significant volatility in daily returns, particularly during market disruptions.
  Moving averages (10, 20, 50-day) highlighted support/resistance levels for both short- and long-term price behavior.
  Headline analysis revealed concise formats (~60 characters) with persuasive intent.

- Sentiment Analysis:

  Utilizing FinBERT-tone and GPT-4o-mini, we captured nuanced sentiment and subjectivity, enhancing predictive modeling.
  FinBERT-tone achieved 80% accuracy, while GPT-4o-mini, fine-tuned on financial datasets, improved accuracy to 84%.
  Generated polarity scores and rolling mean sentiment served as influential predictors of market fluctuations.

- Persuasion Detection:

  We successfully identified persuasive techniques (e.g., Appeal to Authority, Loaded Language) with high accuracy using SemEval methods and GPT-4o-mini.
  Persuasive language was shown to influence investor sentiment and, indirectly, price dynamics.

- Machine Learning Performance:

  The Random Forest Classifier emerged as the most effective predictor, achieving 64% accuracy and a 0.65 ROC-AUC score.
  This outperformed earlier research ([2], [3]) by 12% in ROC-AUC and 5% in accuracy, underscoring the contributions of sentiment and persuasive insights to predictive accuracy.

VII. FUTURE ENHANCEMENTS

To enhance our model and address limitations, future research will focus on:

- Incorporating visual sentiment analysis (image processing) to complement textual data insights.
- Leveraging the CoinGecko API to enable real-time data streaming for dynamic predictions.
- Developing a real-time trading signal system based on integrated NLP and machine learning pipelines.
- Building an accessible web application using Django to provide real-time insights and notifications.
- Implementing LSTM-based deep learning models for hourly or minute-level forecasts, addressing short-term volatility patterns.

VIII. REFERENCES

Data Collection links:

1. Google News https://news.google.com/search?q=btc%202014-2024&hl=en-US&gl=US&ceid=US%3Aen
2. Stock Numerical Data (Finance Data) https://finance.yahoo.com/quote/BTC-USD/history/?period1=1410912000&period2=1730509639
3. Twitter Data https://huggingface.co/datasets/StephanAkkerman/financial-tweets-crypto/viewer/default/train?q=bitcoin

Previous Research papers:

1. SemEval Documentation (Persuasion Detection): https://knowledge4policy.ec.europa.eu/sites/default/files/JRC132862_technical_report_annotation_guidelines_final_with_affiliations_1.pdf
2. “Forecasting directional bitcoin price returns using aspect-based sentiment analysis on online text data,” *Springer*, 2021. [Online]. Available: https://link.springer.com/article/10.1007/s10994-021-06095-3
3. “Price Movement Prediction of Cryptocurrencies Using Sentiment Analysis and Machine Learning,” *MDPI*, 2021. [Online]. Available: https://www.mdpi.com/1099-4300/21/6/589
4. “Bitcoin Price Prediction Using Sentiment Analysis and Empirical Mode Decomposition,” *Springer*, 2024. [Online]. Available: https://link.springer.com/article/10.1007/s10614-024-10588-3
5. “Predicting Bitcoin price movement through Sentiment Analysis: A Comprehensive Study,” *ACM*, 2023. [Online]. Available: https://dl.acm.org/doi/abs/10.1145/3663741.3664791
6. “Social Media Sentiment Analysis for Cryptocurrency Market Prediction,” *arXiv*, 2022. [Online]. Available: https://arxiv.org/abs/2204.10185
7. “Cryptocurrency price and volatility predictions with machine learning,” *Springer*, 2023. [Online]. Available: https://link.springer.com/article/10.1057/

s41270-023-00239-1

8. Predicting Cryptocurrency Value using Sentiment Analysis https://ieeexplore.ieee.org/document/9065838
9. Sentiment-Driven Cryptocurrency Price Prediction: A Machine Learning Approach Utilizing Historical Data and Social Media Sentiment Analysis https://repository.uel.ac.uk/item/8w48q