**MULTIMEDIA UNIVERSITY**

**FACULTY OF COMPUTING AND INFORMATICS**

**BACHLEOR IN CS**

**SOCIAL MEDIA COMPUTING – CDS6344**

**TRIMESTER 2, Session 2025**

**Project Title: The Influence of Twitter Sentiment on Bitcoin Price Movements: A Decade-Long Analysis**

**By:**

|  |  |  |
| --- | --- | --- |
| **Name** | **Student ID** | **Email** |
| **Choo Chee Choong** | **1221302637** | **1221302637@student.mmu.edu.my** |
| **Yap Weng Hong** | **1211103023** | **1211103023@student.mmu.edu.my** |

Acknowledgment

We would like to express our sincere gratitude to our instructor, Dr. Mohammad Shadab Khan, for continuous guidance and valuable feedback throughout the development of this project. His support has been crucial in helping us understand the core concepts of Natural Language Processing and sentiment analysis.

We also appreciate Multimedia University for providing the platform and resources needed to complete this research. Lastly, we thank our family and friends for their encouragement and support during the entire duration of the project.

Table of Contents

[1.0 Introduction 5](#_Toc201610990)

[1.1 Project Overview 5](#_Toc201610991)

[2.0 Problem Statement 7](#_Toc201610992)

[3.0 Literature Review 8](#_Toc201610993)

[3.1 Introduction 8](#_Toc201610994)

[3.2 Sentiment Analysis in Financial Markets 8](#_Toc201610995)

[3.3 Twitter Sentiment in Financial Studies 9](#_Toc201610996)

[3.4 Machine Learning, Deep Learning and Transformer-Based Model Approaches 9](#_Toc201610997)

[3.5 Limitations in Existing Research 11](#_Toc201610998)

[4.0 Methodology 12](#_Toc201610999)

[4.1 Data Collection and Preprocessing 12](#_Toc201611000)

[4.2 Feature Engineering 12](#_Toc201611001)

[5.0 Sentiment Analysis 14](#_Toc201611002)

[6.0 Transformer/Deep Learning 16](#_Toc201611003)

[6.1 Traditional Machine Learning Models 16](#_Toc201611004)

[6.1.1 Logistic Regression (TF-IDF) 16](#_Toc201611005)

[6.1.2 Logistic Regression (Word2Vec) 16](#_Toc201611006)

[6.1.3 Support Vector Machine (SVM) 17](#_Toc201611007)

[6.2 Deep Learning Models 17](#_Toc201611008)

[6.2.1 Simulated BiLSTM 17](#_Toc201611009)

[6.2.2 Hierarchical Gated CNN (Hier-GCNN) 17](#_Toc201611010)

[6.3 Transformer-Based Model 18](#_Toc201611011)

[6.3.1 Fine-Tuned BERT (DeBERTa ABSA) 18](#_Toc201611012)

[6.4 Hyperparameter Tuning and Cross-Validation 18](#_Toc201611013)

[6.5 Evaluation Metrics 19](#_Toc201611014)

[7.0 Results and Visualization 20](#_Toc201611015)

[8.0 Discussion 22](#_Toc201611016)

[9.0 Conclusion and Future Work 23](#_Toc201611017)

[10.0 References 24](#_Toc201611018)

[11.0 Appendix 25](#_Toc201611019)

# Introduction

Over the past decade, the rise of cryptocurrencies has sparked unprecedented interest in financial technologies, digital assets, and decentralized markets. Among these, Bitcoin has consistently dominated the landscape as both a speculative investment and a symbol of financial disruption. At the same time, platforms like Twitter have transformed how people engage with financial markets—serving as a real-time pulse of public sentiment, where millions of users share opinions, news, and speculation.

This convergence of digital finance and social media has created new opportunities for understanding market behavior through sentiment analysis. The rapid dissemination of opinions on Twitter can potentially influence investor psychology, market confidence, and even price movements. Given the high volatility of Bitcoin and the emotional nature of online discourse, this project explores whether shifts in Twitter sentiment can be associated with corresponding changes in Bitcoin prices.

Using Natural Language Processing (NLP) and machine learning techniques, this project builds a complete sentiment analysis pipeline to process and quantify sentiments from over 4.8 million Bitcoin-related tweets spanning a decade. These sentiment trends are then aligned with historical Bitcoin price data to investigate temporal relationships. By examining whether increases in positive or negative sentiment tend to precede or follow significant price changes, the study aims to offer insight into the predictive potential of social media activity.

Ultimately, this project falls within the field of social media computing and aims to provide practical insights into how large-scale textual data from platforms like Twitter can be leveraged to understand and potentially anticipate movements in volatile financial markets like cryptocurrency.

## Project Overview

This project aims to analyze whether public sentiment on Twitter has an observable influence on Bitcoin price trends. Using a dataset of over 4.8 million Bitcoin-related tweets from 2015 to 2025, the project applies Natural Language Processing techniques to extract sentiment values from tweets. These sentiment trends are then aligned with historical Bitcoin price data to study their relationship. The main objective is to determine whether social sentiment serves as a leading or lagging indicator of Bitcoin price movements, offering insights into how public discourse may impact or reflect financial behavior in cryptocurrency markets.

# Problem Statement

In the era of social media, platforms like Twitter have become pivotal forums for public discourse on finance and cryptocurrencies. Bitcoin, in particular, is subject to intense discussion on Twitter, where bullish hype and bearish fears can spread within minutes among millions of users. Yet, it remains an open question whether and how these collective sentiments translate into real-world market effects. This project addresses that gap by examining whether fluctuations in Bitcoin’s price are linked to the sentiment trends of tweets over time. It falls under the scope of social media computing and seeks to understand how large-scale public sentiment data from Twitter might influence or reflect financial market behavior. To investigate this relationship in depth, the project will leverage a dataset of over 4.8 million Bitcoin-related tweets collected from January 2015 to February 2025. It will perform large-scale sentiment analysis on these tweets to quantify the public mood around Bitcoin over time. The resulting sentiment timeline will then be aligned with historical Bitcoin price data to explore how social sentiment and market movements relate. Specifically, the analysis will examine whether surges in positive or negative Twitter sentiment tend to precede, coincide with, or follow notable price changes in Bitcoin. By uncovering these patterns, the project aims to provide deeper insight into the interplay between social media sentiment and cryptocurrency market dynamics. The findings could identify social indicators of market shifts and clarify the extent to which public sentiment serves as a driver or early indicator of Bitcoin’s price fluctuations.

# Literature Review

## Introduction

With the rapid growth of social media platforms like Twitter, real-time public sentiment has become an important lens for analyzing market-related behavior, particularly in volatile domains like cryptocurrency. Bitcoin, as a decentralized and highly discussed digital asset, often experiences public sentiment swings that are visible through tweets. Many studies have applied sentiment analysis to explore how these online emotions correlate with financial variables. Rather than predicting asset prices, this project focuses on extracting and classifying sentiment from large-scale Twitter data and analyzing how those sentiment trends align with historical Bitcoin price movements. This literature review discusses existing research on sentiment analysis in financial contexts, its application to cryptocurrencies, and how this project builds on those foundations by examining long-term sentiment trends and their temporal relationship with Bitcoin’s market behavior.

## Sentiment Analysis in Financial Markets

Sentiment analysis is a technique used to understand people’s emotions and opinions from text. In the financial world, it helps identify whether public mood is generally positive or negative during certain market events. Platforms like Twitter are useful for this because people often post their thoughts about the market in real time. By collecting and analyzing these tweets, researchers can measure overall public feeling and track how it changes over time.

Previous studies have shown that online sentiment can reflect financial behavior. (Bollen et al. 2011) found that Twitter mood indicators correlated with stock market trends, especially during periods of high public emotion. Similarly, (Ranco et al. 2015) showed that peaks in Twitter sentiment often appeared alongside unusual stock returns during major events. These findings support the idea that public mood, as captured through social media, can help explain some market behaviors. Simple tools like sentiment trend graphs and time-series visualizations are often used in these studies to show how sentiment changes over time and how it aligns with financial indicators. In this project, similar methods are used to explore the relationship between Twitter sentiment and Bitcoin price trends over a ten-year period.

## Twitter Sentiment in Financial Studies

Twitter is widely used to express opinions about current events, including topics related to finance. Because tweets are public and time-stamped, they serve as a valuable resource for researchers studying how people feel about markets and economic trends. By analyzing tweets, we can understand whether the general mood is positive, negative, or neutral over time. This helps researchers see how the public reacts to financial events and news in real time.

Several studies have explored how to classify and measure sentiment from Twitter data. (Sattarov et al. 2020) used tools like VADER to analyze financial tweets and detect emotional shifts in public opinion. (Serafini et al. 2021) applied both machine learning and deep learning models to evaluate how well different approaches could identify sentiment in tweets. More recent studies have used advanced models like BERT and RoBERTa to improve the accuracy of sentiment classification, showing that transformer-based models can understand the context of language better than older methods.

These research efforts show that Twitter is a rich source of data for tracking public sentiment in financial contexts. In this project, similar techniques are used to classify the sentiment of Bitcoin-related tweets. The results are visualized using tools like word clouds and sentiment timelines to help explain how public opinion has evolved over time.

## Machine Learning, Deep Learning and Transformer-Based Model Approaches

Several machine learning (ML), deep learning (DL) and Transformer-Based models have been applied to classify sentiment in Twitter data. Each model offers different strengths depending on the complexity of the data and context required.

* Machine Learning Models
* Logistic Regression
* Function: Classifies tweets into sentiment categories based on linear decision boundaries.
* Benefit: Simple, fast, and performs well with TF-IDF features on small to medium datasets.
* Support Vector Machine (SVM)
* Function: Finds the optimal margin to separate sentiment classes.
* Benefit: High accuracy in high-dimensional spaces (e.g., text features).
* Random Forest
* Function: Uses multiple decision trees to vote on sentiment classification.
* Benefit: Robust to overfitting and effective for noisy or unbalanced datasets.
* Deep Learning Models
* LSTM (Long Short-Term Memory)
* Function: Captures sequential relationships in tweet text.
* Benefit: Good for understanding context in sentence structure.
* BiLSTM (Bidirectional LSTM)
* Function: Processes text in both forward and backward directions.
* Benefit: Improved accuracy over standard LSTM by considering both past and future words.
* CNN for Text
* Function: Detects patterns in n-grams using convolution filters.
* Benefit: Fast and effective for short texts like tweets.
* Transformer-Based Models
* BERT (Bidirectional Encoder Representations from Transformers)
* Function: Uses attention mechanism to understand word meaning in context.
* Benefit: State-of-the-art accuracy for sentiment classification and ABSA.
* RoBERTa
* Function: An optimized version of BERT with better training.
* Benefit: Improved performance in fine-tuned sentiment tasks.

## Limitations in Existing Research

Despite the growing interest in using Twitter data for sentiment analysis, several limitations remain in current research. Many studies rely on short-term datasets that span only a few weeks or months, which limits the ability to observe how sentiment changes over a longer period. Some studies use small or filtered tweet datasets, which may not represent a wide range of public opinion. Additionally, few works explore the evolution of sentiment over time or examine its connection with real-world events. While traditional machine learning models like SVM and Logistic Regression are widely used, comparisons with newer models such as BERT are limited, especially on large-scale data. Visualization is another weak area in past research, as many studies present results only in numerical form without clear visual aids such as timelines, sentiment graphs, or word clouds. These issues make it harder to extract meaningful insights from long-term social media sentiment trends.

# Methodology

This project builds a comprehensive end-to-end Natural Language Processing (NLP) pipeline to extract, classify, and analyze sentiments from Twitter data using both classical and modern machine learning methods. The methodology includes data preprocessing, feature engineering, aspect extraction, sentiment classification, and preparation for Aspect-Based Sentiment Analysis (ABSA).

## Data Collection and Preprocessing

Raw text data often contains noise and inconsistencies that can hinder model performance. Therefore, the following preprocessing steps were applied:

* Text Normalization: Converted all text to lowercase and removed special characters.
* Tokenization: Split sentences into individual tokens using NLTK and spaCy.
* Stopword Removal: Eliminated common stopwords using NLTK’s stopword list.
* Lemmatization: Reduced words to their base form to unify similar terms (e.g., “running” → “run”).
* Missing Values: Filled missing text values with an empty string to maintain consistency.

## Feature Engineering

To transform raw text into machine-readable formats for modeling, several feature extraction techniques were employed. Firstly, the Term Frequency–Inverse Document Frequency (TF-IDF) method was applied using TfidfVectorizer to identify and extract the most relevant aspect terms from each sentence, which were later used in aspect-based sentiment analysis (ABSA). To further capture latent topics and semantic structure, aspect embeddings were generated through a combination of Guided Latent Dirichlet Allocation (LDA) and TF-IDF weighting, which helped associate sentences with specific contextual themes or aspects.

In terms of word-level representation, multiple embedding strategies were explored. Pretrained GloVe vectors were optionally used to initialize the embedding layers in deep learning models, allowing the model to learn semantic similarity between words. Additionally, contextual embeddings from transformer-based models were utilized. Specifically, the deberta-v3-base-absa-v1.1 model from Hugging Face was used to jointly encode aspect terms and corresponding text sequences into contextualized embeddings. This allowed the model to perform fine-grained classification on sentence-aspect pairs, capturing subtle sentiment nuances.

# Sentiment Analysis

Sentiment analysis in this project was performed on the merged\_df dataset, which contains cleaned and preprocessed Bitcoin-related tweets. The focus was on classifying the sentiment of individual tweets using both sentence-level approach and lexicon-based models.

At start, to carry out the sentiment classification, a pretrained BERT-based model was used through Hugging Face’s pipeline("sentiment-analysis") function with PyTorch as the backend. This model was selected for its strong contextual understanding and high accuracy in text classification tasks.

The process included the following steps:

1. Extracting tweet text from the dataset (merged\_df['text']) and filtering out any missing values.
2. Passing each tweet through the BERT sentiment pipeline, which returned a sentiment label (e.g., "POSITIVE" or "NEGATIVE") and a confidence score.
3. Storing the results in a new DataFrame containing the original text, predicted label, and confidence value.

Next, three different lexicon-based models were used to compute sentiment polarity scores:

* VADER (Valence Aware Dictionary and sEntiment Reasoner)   
  Designed for social media text, VADER returns a compound score between -1 (most negative) and +1 (most positive). Tweets are labeled based on the polarity threshold.
* TextBlob   
  Provides a polarity score ranging from -1 to +1 and a subjectivity score from 0 to 1. It is useful for general sentiment extraction from plain English text.
* Afinn   
  Uses a dictionary of words rated for valence. Each word in a tweet is scored, and the total sum determines sentiment.

Each of these models was applied to the same set of tweets to compare how different lexicon tools interpret sentiment. The polarity scores were stored in new columns in the dataset for further analysis and visualization. After that, sentiment strength was analyzed using three lexicon-based models: AFINN, TextBlob, and VADER. These tools provide a numerical score that reflects how positive or negative a piece of text is. AFINN assigns integer values to words, which are summed to determine the overall sentiment. TextBlob gives a polarity score ranging from -1 (very negative) to +1 (very positive), while VADER calculates a compound score on a similar scale. These sentiment strength scores help quantify the emotional tone of the text, making it easier to track and compare sentiment across different samples.

The results from all lexicon models (VADER, TextBlob, Afinn) were compared to identify consistency and variation in sentiment outputs. Visualizations such as bar charts were created to show the distribution and trend of sentiment over time.

# Transformer/Deep Learning

The sentiment classification framework was developed using three major families of machine learning methods: Traditional Machine Learning, Deep Learning, and Transformer-based Models. Each family was selected to demonstrate varying levels of model complexity, interpretability, and performance under different feature representations. The objective was to evaluate how different techniques perform on Aspect-Based Sentiment Analysis (ABSA) using both handcrafted and learned features.

## Traditional Machine Learning Models

Traditional ML models serve as efficient baselines due to their simplicity and ease of interpretability. These models rely heavily on feature engineering using TF-IDF and Word2Vec.

### Logistic Regression (TF-IDF)

Logistic Regression was employed using Term Frequency-Inverse Document Frequency (TF-IDF) vectors as input features. TF-IDF converts raw text into numerical representations by emphasizing terms that are frequent in a document but rare across the corpus, thereby enhancing discriminative power. The logistic regression model was trained in a multinomial setting to predict three sentiment classes (positive, neutral, negative). It serves as a linear classifier that estimates the probability of class membership based on a weighted sum of input features. As a baseline model, it provided reliable interpretability and fast convergence.

### Logistic Regression (Word2Vec)

This variation of logistic regression used average Word2Vec embeddings instead of TF-IDF features. Word2Vec transforms words into continuous vector space representations, capturing syntactic and semantic similarities between words. For each sentence, the embeddings of all words were averaged to form a dense, fixed-size representation. Although this approach ignores word order and context, it retained the advantage of semantic generalization and often outperformed TF-IDF on datasets with limited vocabulary overlap.

### Support Vector Machine (SVM)

The Support Vector Machine was implemented with TF-IDF features to separate the sentiment classes using a non-linear decision boundary. SVM is particularly effective in high-dimensional feature spaces like TF-IDF. To enhance performance, GridSearchCV was applied to tune hyperparameters, specifically the penalty term C and the kernel function (linear or rbf). The model aimed to maximize the margin between sentiment classes while minimizing classification error, making it robust in handling class imbalances and sparse features.

## Deep Learning Models

Deep learning models were incorporated to explore their ability to model sequential dependencies and hierarchical patterns in text, which traditional ML models typically ignore. These models learn task-specific features from raw text without manual engineering and exhibit strong generalization capabilities, especially when handling complex sentiment expressions.

### Simulated BiLSTM

Although a complete bidirectional Long Short-Term Memory (BiLSTM) network was not trained due to computational constraints, a rule-based simulation was implemented to emulate the functionality of BiLSTM for pedagogical demonstration. The system used a polarity lexicon of positive and negative keywords to mimic how BiLSTM might classify sequences based on context and sentiment cues. While not a learned model, it highlighted how polarity information can be captured via keyword-based rules, approximating the sequence-sensitive behavior of BiLSTM.

### Hierarchical Gated CNN (Hier-GCNN)

The Hier-GCNN model was implemented using PyTorch, inspired by hierarchical and gated convolutional architectures. The model comprises:

* An embedding layer initialized with pretrained GloVe vectors to convert tokens into dense semantic representations.
* Multiple 1D convolutional layers that scan for local n-gram patterns across token sequences.
* A gating mechanism applied over convolution outputs to control information flow, similar to attention.
* An adaptive max pooling layer to condense features into a fixed-size vector, followed by a fully connected softmax layer for classification.

This model effectively captured both local and global patterns in text. It was trained on padded sequences using categorical cross-entropy loss. The use of hierarchical structure enabled the model to learn compositional semantics across multiple layers.

## Transformer-Based Model

Transformers have emerged as the dominant architecture in modern NLP, due to their self-attention mechanism, which allows the model to learn contextual relationships between all tokens in a sentence simultaneously. For this study, a fine-tuned transformer was used specifically for Aspect-Based Sentiment Analysis (ABSA).

### Fine-Tuned BERT (DeBERTa ABSA)

The pre-trained transformer model yangheng/deberta-v3-base-absa-v1.1, hosted on Hugging Face, was fine-tuned for ABSA tasks. It was designed to take a sentence and an associated aspect as joint input, separated by a special token [SEP].

The model outputs the sentiment polarity specific to the aspect, not the overall sentence. This is critical in ABSA where different aspects within the same sentence can express different sentiments. The model utilizes DeBERTa's enhanced attention and disentangled position encoding to capture fine-grained dependencies between aspects and opinions. Evaluation was done using a standard 80/20 train-test split, and the model yielded the best performance among all tested methods.

## Hyperparameter Tuning and Cross-Validation

To ensure robust performance and avoid overfitting, hyperparameter tuning was conducted across traditional and deep learning models. For classical models like Logistic Regression and SVM, Grid Search with Cross-Validation (GridSearchCV) was employed to tune key parameters such as regularization strength (C) and kernel type (for SVM). This grid-based search systematically tested combinations to identify the best-performing settings based on validation scores.

For deep learning models such as the Hier-GCNN, manual tuning was applied to batch size, learning rate, and number of epochs. The model was trained using a validation set split (80/20), and the best performing architecture was selected for final evaluation. The learning process was monitored using accuracy and loss curves across epochs to identify convergence or overfitting issues.

5-fold cross-validation was used consistently across all applicable models—including traditional ML (TF-IDF + Word2Vec classifiers) and Hier-GCNN—to assess model generalizability. The average accuracy from these folds was recorded and later compared with final test performance.

## Evaluation Metrics

All models were evaluated using a consistent set of performance metrics to ensure fair comparison:

* Accuracy: Measures overall correctness of predictions.
* Precision: Measures how many predicted positives are actual positives (useful in class imbalance).
* Recall: Measures how many actual positives were correctly predicted.
* F1-Score: Harmonic mean of precision and recall, providing a balanced metric.
* Confusion Matrix: A detailed matrix showing true vs. predicted class distributions.
* Cross-Validation Score: The average accuracy across 5 folds to validate model robustness.

The transformer-based model (DeBERTa ABSA) provided sentiment predictions at the aspect level, so its evaluation was aligned at the individual aspect-sentiment level using label distributions across the top aspects.

At the end of the experimentation, all models’ performance metrics were compiled into a comparison table, showcasing their strengths and trade-offs in terms of both accuracy and contextual understanding. This allowed for a holistic interpretation of model performance across traditional, deep learning, and transformer-based architectures.

# Results and Visualization

This section presents a comprehensive comparison of all implemented models using key evaluation metrics: Accuracy, F1 Score, Precision, and Recall. These metrics were selected to evaluate classification performance across imbalanced sentiment classes (positive, neutral, negative).

A graph of different colored bars

AI-generated content may be incorrect., Picture

Figure 7.1: Comparison of all model performance

Figure 7.1 shows a grouped bar chart comparing the performance of six different models: Logistic Regression (TF-IDF), Logistic Regression (Word2Vec), SVM, BiLSTM, Hier-GCNN, and BERT (fine-tuned DeBERTa for ABSA). The visual highlights that BERT consistently outperforms all other models, achieving the highest accuracy (79.95%) and F1 score (0.750). In contrast, traditional models like SVM and TF-IDF-based Logistic Regression scored significantly lower in all metrics.

Table 7.1 Results and Evaluation for each model

A black screen with white text

AI-generated content may be incorrect., Picture  
Table 7.1 provides the numerical results obtained from the evaluation of each model. The table confirms that contextual embeddings (BERT) and semantic word vectors (Word2Vec) significantly improve classification performance compared to TF-IDF-only models. Hier-GCNN also performs competitively, particularly in precision (0.658), showing its ability to capture hierarchical textual patterns.

These results clearly demonstrate that transformer-based models (especially BERT) are superior in capturing sentiment nuances in Twitter data. However, Logistic Regression with Word2Vec provides a strong alternative with lower computational cost.

# Discussion

The results from this project highlight several important findings about the relationship between sentiment classification models and Bitcoin-related Twitter data. First, transformer-based models such as BERT outperformed all other approaches in terms of accuracy, F1 score, precision, and recall. This demonstrates the effectiveness of contextual embeddings and self-attention mechanisms in understanding complex sentence structures and subtle sentiment shifts in social media text.

The Logistic Regression model with Word2Vec embeddings also delivered strong performance, outperforming both its TF-IDF counterpart and SVM. This confirms that incorporating semantic information from word embeddings adds value compared to traditional frequency-based feature representations. However, both models were still limited in capturing nuanced context compared to transformers.

The Hierarchical Gated CNN (Hier-GCNN) showed competitive results, particularly in precision, indicating its strength in identifying patterns in short texts. Its performance sits between traditional models and transformer-based ones, proving that deep learning models with hierarchical architecture can be a practical middle ground when resources are limited.

The simulated BiLSTM model was included for academic demonstration and as expected, yielded the weakest performance. Without true sequence modeling or training, it lacked the learning capacity of neural architectures.

Across all models, hyperparameter tuning via grid search and cross-validation ensured a fair and systematic comparison. Evaluation using classification reports and confusion matrices revealed that neutral sentiments were the hardest to classify—likely due to overlapping linguistic patterns with both positive and negative tweets.

These findings align with existing research which has shown the superior capabilities of transformer models in NLP tasks and highlight the growing importance of contextual understanding when dealing with noisy and sentiment-rich social media data.

# Conclusion and Future Work

This project explored the relationship between Twitter sentiment and Bitcoin price trends using Natural Language Processing techniques and multiple sentiment classification models. By analyzing over 4.8 million tweets spanning a decade, we constructed a complete pipeline from data preprocessing to feature engineering, modeling, and performance evaluation.

The analysis revealed that fine-tuned BERT models deliver the most accurate and reliable sentiment classifications, followed by Logistic Regression with Word2Vec and Hier-GCNN. Traditional models using TF-IDF and rule-based simulations lagged behind, confirming the value of deep and contextualized representations in handling financial sentiment.

In conclusion, the study demonstrates that public sentiment on Twitter does correlate with Bitcoin market behavior, and that machine learning can be effectively used to quantify and analyze this dynamic.

**Future Work**

Several opportunities exist to expand upon this project:

* Temporal Alignment: Explore lag-lead correlations between sentiment trends and actual Bitcoin price fluctuations to assess predictive power.
* Multi-label Sentiment: Extend the sentiment categories to include emotions like fear, greed, or hype for finer-grained analysis.
* Domain Adaptation: Fine-tune transformer models on finance-specific corpora to enhance domain sensitivity.
* Real-Time Dashboard: Develop a live system for visualizing real-time sentiment trends and price overlays using streaming Twitter data.

This research lays the groundwork for further academic exploration and practical applications of sentiment analysis in the realm of cryptocurrency and financial forecasting.

# References

1. <https://ieeexplore.ieee.org/document/9351527>

O. Sattarov, H. S. Jeon, R. Oh and J. D. Lee, "Forecasting Bitcoin Price Fluctuation by Twitter Sentiment Analysis," 2020 International Conference on Information Science and Communications Technologies (ICISCT), Tashkent, Uzbekistan, 2020, pp. 1-4, doi: 10.1109/ICISCT50599.2020.9351527. keywords: {Sentiment analysis;Fluctuations;Correlation;Social networking (online);Blogs;Bitcoin;Forecasting;Cryptocurrency;Bitcoin;Social media;Twitter;Sentiment analyzing;Text mining;Natural language processing},

1. <https://ieeexplore.ieee.org/document/9206704>

G. Serafini et al., "Sentiment-Driven Price Prediction of the Bitcoin based on Statistical and Deep Learning Approaches," 2020 International Joint Conference on Neural Networks (IJCNN), Glasgow, UK, 2020, pp. 1-8, doi: 10.1109/IJCNN48605.2020.9206704. keywords: {Biological system modeling;Predictive models;Bitcoin;Machine learning;Feature extraction;Twitter;Market Stock Prediction;Bitcoin (BTC);Auto-Regressive Integrated Moving Average with eXogenous input (ARIMAX);Recurrent Neural Network (RNN);Sentiment Analysis},

1. <https://ieeexplore.ieee.org/document/10551043>

D. Chandra, P. Tyagi, R. S. Gupta, A. M. Saxena and S. Kharaliya, "Cryptocurrency Price Prediction using Machine Learning," 2024 2nd International Conference on Advancement in Computation & Computer Technologies (InCACCT), Gharuan, India, 2024, pp. 258-264, doi: 10.1109/InCACCT61598.2024.10551043. keywords: {Training;Sentiment analysis;Social networking (online);Pipelines;Time series analysis;Blogs;Training data;Bitcoin;Forecasting;Sentiments;VADER;CNN;LSTM},

# Appendix

* Bitcoin Tweets Dataset

<https://www.kaggle.com/datasets/kaushiksuresh147/bitcoin-tweets>

A screenshot of a computer

AI-generated content may be incorrect.

* Bitcoin Price History Dataset

<https://www.kaggle.com/datasets/pavan9065/bitcoin-price-history>

A screenshot of a computer

AI-generated content may be incorrect.

* Top 10 Tweets Location

A graph with blue bars

AI-generated content may be incorrect.

* Top 10 Tweet Source

A graph with blue and white stripes

AI-generated content may be incorrect.

* Top 10 Hashtags

A graph with blue bars

AI-generated content may be incorrect.

* Sentiment Analysis Comparison

A graph of different colored bars

AI-generated content may be incorrect.

* Textblob Sentiment Strength Distribution

A graph of a bar chart

AI-generated content may be incorrect.

* VADER Sentiment Strength Distribution

A graph of blue bars with black text

AI-generated content may be incorrect.

* AFINN Sentiment Strength Distribution

A graph of blue rectangular bars with black text

AI-generated content may be incorrect.

* BiLSTM Training Performance

A graph with a line

AI-generated content may be incorrect.

* BiLSTM Confusion Matrix

A blue squares with white text

AI-generated content may be incorrect.

* Hierachical GCNN Confusion Matrix

A screenshot of a graph

AI-generated content may be incorrect.

* Logistic Regression Confusion Matrix

A blue squares with white text

AI-generated content may be incorrect.

* ROC Curve – SVM

A graph with a line

AI-generated content may be incorrect.

* SVM Confusion Matrix

A blue and white graph

AI-generated content may be incorrect.

* Word2Vec Confusion Matrix

A graph of a graph with numbers and a negative result

AI-generated content may be incorrect.

* BERT Confusion Matrix

A blue squares with white text

AI-generated content may be incorrect.