

Article

# LLMs and NLP Models in Cryptocurrency Sentiment Analysis: A Comparative Classification Study

Konstantinos I. Roumeliotis <sup>1,\*</sup>, Nikolaos D. Tselikas <sup>1</sup> and Dimitrios K. Nasiopoulos <sup>2</sup>

<sup>1</sup> Department of Informatics and Telecommunications, University of Peloponnese, Akadimaikou G. K. Vlachou Street, 22131 Tripoli, Greece; ntsel@uop.gr

<sup>2</sup> Department of Agribusiness and Supply Chain Management, School of Applied Economics and Social Sciences, Agricultural University of Athens, 11855 Athens, Greece; dimnas@hua.gr

\* Correspondence: k.roumeliotis@uop.gr; Tel.: +30-271-037-2216

**Abstract:** Cryptocurrencies are becoming increasingly prominent in financial investments, with more investors diversifying their portfolios and individuals drawn to their ease of use and decentralized financial opportunities. However, this accessibility also brings significant risks and rewards, often influenced by news and the sentiments of crypto investors, known as crypto signals. This paper explores the capabilities of large language models (LLMs) and natural language processing (NLP) models in analyzing sentiment from cryptocurrency-related news articles. We fine-tune state-of-the-art models such as GPT-4, BERT, and FinBERT for this specific task, evaluating their performance and comparing their effectiveness in sentiment classification. By leveraging these advanced techniques, we aim to enhance the understanding of sentiment dynamics in the cryptocurrency market, providing insights that can inform investment decisions and risk management strategies. The outcomes of this comparative study contribute to the broader discourse on applying advanced NLP models to cryptocurrency sentiment analysis, with implications for both academic research and practical applications in financial markets.



**Citation:** Roumeliotis, K.I.; Tselikas, N.D.; Nasiopoulos, D.K. LLMs and NLP Models in Cryptocurrency Sentiment Analysis: A Comparative Classification Study. *Big Data Cogn. Comput.* **2024**, *8*, 63. <https://doi.org/10.3390/bdcc8060063>

Academic Editors: Fabrizio Marozzo and Riccardo Cantini

Received: 22 April 2024

Revised: 30 May 2024

Accepted: 3 June 2024

Published: 5 June 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

In recent years, cryptocurrencies have emerged as a significant asset class in the global financial landscape, attracting substantial interest from investors seeking diversification and alternative investment opportunities. The decentralized nature and technological innovations underpinning cryptocurrencies, such as Bitcoin and Ethereum, have reshaped traditional financial paradigms, introducing new dynamics driven by market sentiment and information dissemination.

The volatility of cryptocurrency markets is often influenced by a myriad of factors, including news articles, social media discussions, crypto signals, and investor sentiment [1–3]. Understanding and accurately gauging sentiment in this rapidly evolving environment is essential for making informed investment decisions and managing risks effectively. As a result, sentiment analysis techniques using advanced natural language processing (NLP) models have gained traction as valuable tools for extracting sentiment from textual data related to cryptocurrencies.

By understanding public sentiment for a specific cryptocurrency in advance and monitoring market tension, we can predict price movements. Negative sentiment in the crypto market often leads to price drops, while positive sentiment tends to drive price increases. For instance, the recent fear of potential conflict between Israel and Iran caused a marginal 10% decrease in the value of all cryptocurrencies.

In this context, our study focuses on the comparative analysis of large language models (LLMs) and NLP techniques in the domain of cryptocurrency sentiment analysis. Specifically, we investigate the efficacy of fine-tuned models such as GPT-4, BERT, and FinBERT for discerning sentiment from cryptocurrency news articles. By leveraging the capabilities of these sophisticated language models, we aim to explore how effectively they can capture nuanced sentiment patterns, thus contributing to a deeper understanding of sentiment dynamics within the cryptocurrency market.

The primary objectives of this study are twofold: firstly, to evaluate the performance of LLMs and NLP models in cryptocurrency sentiment analysis through a comparative classification study, providing insights that can inform investment strategies and risk management practices within cryptocurrency markets; and secondly, to address specific research inquiries that previous studies have not adequately covered.

- Q1: Among the fine-tuned models of GPT-4, BERT, and FinBERT, which one demonstrates superior predictive capabilities in cryptocurrency news sentiment analysis and classification?
- Q2: What is the impact of fine-tuning LLMs and NLP models for specific tasks?

To fulfill the goals of our study, we undertake an extensive review of the existing literature on sentiment analysis and classification pertaining to cryptocurrencies in Section 2. Section 3 details the materials and methodology utilized in this study, while Section 4 provides the prediction outcomes for all three models and their variations before and after fine-tuning. In Section 5, we delve into a detailed discussion of the results, extracting valuable insights derived from our research findings.

## 2. Literature Review

Cryptocurrency sentiment analysis involves employing sophisticated NLP models to analyze textual data sources like news articles, social media posts, forums, and blogs related to cryptocurrencies. The primary objective is to extract and quantify sentiment—whether positive, negative, or neutral—expressed in these texts to comprehend the prevailing market sentiment towards specific cryptocurrencies or the broader market as a whole.

The significance of cryptocurrency sentiment analysis in financial markets is underscored by several compelling reasons:

- Market sentiment understanding: Cryptocurrency markets are highly sensitive to sentiment influenced by news events, social media trends, regulatory developments, and investor perceptions [4]. Sentiment analysis aids in assessing the prevailing sentiment and mood of market participants, offering insights into potential market movements.
- Insights into investor behavior: By analyzing sentiment, researchers and investors gain valuable insights into investor behavior and sentiment dynamics [5,6]. Understanding investor sentiment facilitates predicting market trends, identifying sentiment shifts, and assessing associated risks.
- Decision-making support: Sentiment analysis plays a pivotal role in guiding investment decisions. Traders and investors leverage sentiment-driven insights to make informed decisions; positive sentiment may signal buying opportunities, while negative sentiment could prompt caution or sell signals [7,8].
- Risk management: Assessing sentiment contributes to managing risks linked with cryptocurrency investments [9]. Sudden shifts in sentiment towards specific cryptocurrencies could indicate potential price volatility or market downturns.

In the cryptocurrency ecosystem, sentiment analysis offers unique advantages owing to the decentralized and rapidly evolving nature of cryptocurrencies. The wealth of online data sources, including social media platforms and cryptocurrency news websites, renders sentiment analysis particularly relevant and valuable for comprehending market behavior and investor sentiment in this domain.

By harnessing sentiment analysis techniques, researchers and market participants attain deeper insights into the factors driving cryptocurrency markets, empowering them to make more informed decisions based on sentiment-driven intelligence. This underscores the pivotal role of sentiment analysis in enriching market intelligence and facilitating effective decision-making within the dynamic and volatile cryptocurrency landscape.

### 2.1. NLP in Sentiment Analysis

NLP is integral to the field of sentiment analysis, offering tailored techniques to effectively extract sentiment from textual data. In sentiment analysis, NLP methods are essential for discerning the emotions and opinions conveyed through language. Some of the most widely used techniques and applications within financial contexts include:

- **Sentiment lexicons:** Sentiment lexicons are curated dictionaries featuring words or phrases annotated with sentiment scores (e.g., positive, negative, neutral) [7], aiding sentiment analysis by quantifying sentiment from specific word usage. Widely used lexicons like SentiWordNet and Vader are foundational resources. In financial sentiment analysis, specialized lexicons capturing financial terminology enhance accuracy by accounting for domain-specific expressions and market conditions.
- **Deep learning approaches:** Deep learning architectures, such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), and bidirectional encoder representations from transformers (BERT) are extensively applied in sentiment analysis [10]. CNNs are effective for feature extraction, while RNNs excel in sequence modeling [11], and BERT specializes in capturing contextual information from language. Their application demonstrates remarkable accuracy in deciphering intricate patterns within text. These models undergo fine-tuning with financial datasets to better understand the nuances of financial language, thereby improving sentiment analysis in financial markets. Furthermore, integrating domain-specific features through NLP techniques enhances their effectiveness, enabling the extraction of meaningful signals from financial text data [12].

### 2.2. Previous Studies on Sentiment Analysis in Cryptocurrencies

Our literature review presents a comprehensive overview of 49 papers on the evolving role of sentiment analysis in understanding and predicting cryptocurrency market dynamics. The studies showcase a diverse array of methodologies, ranging from traditional sentiment analysis to cutting-edge deep learning models, that highlight the significant impact of sentiment-driven factors on cryptocurrency price movements and investor behavior. By leveraging data sources such as social media platforms, news articles, and market indicators, researchers demonstrate the potential of sentiment analysis to provide valuable insights for navigating the volatile cryptocurrency landscape and making informed investment decisions. Despite the progress made, challenges such as bias mitigation in sentiment data and the nuanced interplay between sentiment and fundamental market factors underscore the ongoing need for further exploration and refinement in this burgeoning field of research. Through innovative approaches and practical applications, the reviewed studies collectively contribute to advancing sentiment analysis as a powerful tool for cryptocurrency market analysis and decision-making.

#### 2.2.1. Traditional Supervised Learning

Seven papers have explored cryptocurrencies using traditional supervised learning techniques. Notably, Rouhani et al. [7] identified a prevailing positive sentiment towards cryptocurrencies on Twitter, indicating the potential for accurate sentiment prediction through supervised methods like SVM. Colianni et al. [13] achieved notably high prediction accuracies in algorithmic trading strategies based on Twitter sentiment analysis, highlighting the feasibility of utilizing sentiment data for effective trading strategies. Loganova et al. [14] made significant advancements in directional cryptocurrency price prediction by integrating fine-grained sentiment analysis into online textual data, showcasing

the effectiveness of innovative sentiment analysis techniques. Sureshbhai et al. [9] introduced KaRuNa, a blockchain-based sentiment analysis framework addressing risks in cryptocurrency schemes, achieving remarkable accuracy in risk assessment. Köse Ozan [15] automated the classification of crypto assets and sentiment analysis of news, providing valuable tools for informed decision-making based on the sentiment analysis of news sources. Ghazanfar et al. [16] successfully integrated social media sentiment analysis into cryptocurrency price prediction models, emphasizing its importance for financial forecasting. Seroyizhko et al. [17] introduced a sentiment and emotion annotated dataset for Bitcoin price forecasting based on Reddit posts, advocating for sophisticated social media integration techniques in crypto market analysis.

### 2.2.2. Deep Learning

Thirteen of the papers have utilized deep learning to enhance cryptocurrency price prediction. Huang et al. [1] pioneered the integration of sentiment from Chinese social media platform Sina-Weibo into LSTM models, demonstrating superior performance over traditional methods in predicting cryptocurrency prices. Building upon this approach, Parekh et al. [18] introduced DL-GuesS, a framework combining deep learning and sentiment analysis to forecast Dash and Bitcoin Cash prices based on historical data and Twitter sentiment. Raju et al. [19] further emphasized the significance of public sentiment by correlating Twitter and Reddit data with Bitcoin prices, achieving enhanced accuracy through supervised learning algorithms. Pant et al. [10] extended this sentiment analysis to predict Bitcoin price volatility using RNNs, offering insights into market sentiment dynamics. Pang et al. [20] highlighted the superiority of sentiment data models over conventional approaches in capturing market dynamics and predicting price fluctuations through machine learning techniques. Farimani et al. [21] contributed a methodology leveraging latent economic concepts and sentiments in financial news for market prediction, offering superior performance compared to baselines. Gadi et al. [22] investigated annotator selection's impact on sentiment corpus creation, while Farimani et al. [23] integrated sentiment analysis with technical indicators for real-time market prediction tools. Zamani et al. [2] developed an XLNet-GRU sentiment regression model for cryptocurrency news, showcasing effectiveness across languages. Nasekin et al. [24] constructed sentiment indices from StockTwits data, informing cryptocurrency market returns and volatility prediction. Wołk et al. [25] analyzed social media sentiment's impact on cryptocurrency prices, emphasizing psychological and behavioral attitudes. Vo [26] focused on the sentiment analysis of news for Ethereum price prediction, providing actionable insights for investors. Aslam et al. [27] emphasized understanding public sentiment towards cryptocurrency, achieving high accuracy through ensemble deep learning models.

### 2.2.3. Lexicon-Based Sentiment Analysis

Six of the examined papers utilized lexicon-based sentiment analysis for cryptocurrency price forecasting. Chen et al. [4] utilized sentiment data from StockTwits to investigate sentiment-induced bubbles, employing a smooth transition autoregression model to identify speculative bubbles and linking sentiment dynamics with market volatility. Their findings suggested that sentiment-driven exuberance can lead to locally explosive price dynamics, characteristic of speculative bubbles. Erdogan et al. [8] focused on the sentiment analysis of social media data, specifically Twitter, to discern user perceptions of cryptocurrencies. Through big data tools, they analyzed Twitter data to extract insights and classify user attitudes, providing valuable insights into public sentiment towards crypto assets, which can inform market strategies. Dwivedi et al. [28] contributed to understanding sentiment evolution pre- and post-COVID-19 using topic modeling techniques on Twitter data, extracting key themes in public sentiment towards Bitcoin and cryptocurrency. Şaşmaz et al. [29] delved into tweet sentiment analysis for cryptocurrencies like NEO, exploring correlations between sentiment and daily prices through an RF classifier trained on labeled tweet data. They demonstrated the feasibility of using automated sentiment

analysis for cryptocurrency trading decisions based on Twitter data. Kraaijeveld et al. [30] explored the predictive capacity of Twitter sentiment across multiple cryptocurrencies, finding predictive power for returns in specific cryptocurrencies such as Bitcoin, Bitcoin Cash, and Litecoin. Additionally, they detected cryptocurrency-related Twitter bot accounts, highlighting the complexity of sentiment analysis in this domain. Gurrib et al. [31] introduced a machine learning approach, combining linear discriminant analysis and sentiment analysis to predict Bitcoin price movements, leveraging BTC price data and crypto-related news sentiment to forecast next-day price directions.

#### 2.2.4. BERT and Transformer Models

Four studies have utilized BERT and other transformer-based models to delve into the understanding and prediction of cryptocurrency market behavior. Ider et al. [32] employed BERT classifiers and weak supervision techniques to improve the predictive power of text-based features for cryptocurrency return prediction. Their incorporation of weak labeling techniques addresses the challenge of unlabeled text data, showcasing the potential of NLP advancements in financial forecasting models. Similarly, Kulakowski et al. [33] developed sentiment classification solutions tailored to cryptocurrency-related social media posts, such as CryptoBERT and LUKE sentiment lexicon. Their study fine-tuned BERT models and integrated emoji sentiment analysis, offering investors tools for sentiment-based trading strategies. Nguyen Thanh et al. [34] explored ChatGPT for the sentiment analysis of Twitter data and its impact on Bitcoin returns, underscoring the influence of AI-generated sentiment indicators on market dynamics. Lastly, Raheman et al. [35] examined various NLP models for sentiment analysis in the cryptocurrency domain, emphasizing the importance of interpretable AI methods in practical financial prediction tasks. Their study highlighted correlations between sentiment metrics and Bitcoin price movements, shedding light on sentiment-driven market dynamics.

#### 2.2.5. Time-Series Analysis

In their study, Georgoula et al. [36] utilized a combination of time-series analysis and sentiment analysis to probe the determinants of Bitcoin prices. Leveraging machine learning techniques such as SVMs, they scrutinized sentiment data sourced from Twitter feeds, alongside economic and technological factors. Their findings unveiled noteworthy correlations: a positive relationship between Twitter sentiment ratio, public interest gauged through Wikipedia searches, and Bitcoin prices. Conversely, they underscored the negative influences of exchange rates and stock market indices on Bitcoin's valuation. Meanwhile, Chalkiadakis et al. [37] introduced a novel framework aimed at evaluating multimodal statistical causality within cryptocurrency markets. By employing multiple-output Gaussian processes and delving into sentiment time-series data, their research delved into the intricate causal relationships between sentiment, price dynamics, and blockchain indicators. Through their investigation, they shed light on sentiment-driven market behaviors and the consequential impacts of investor sentiment on cryptocurrency market dynamics.

#### 2.2.6. Computational Text Analysis

Anamika et al. [38] focused on assessing the influence of investor sentiment, particularly towards Bitcoin, on cryptocurrency returns. Using a direct survey-based measure from the sentix database, they found that optimistic sentiment among investors correlates with Bitcoin price appreciation. This sentiment measure also holds predictive power for Bitcoin prices even after accounting for other relevant factors. Moreover, they noted that Bitcoin sentiment can influence the prices of other cryptocurrencies, suggesting a broader impact on the market beyond Bitcoin itself. In a similar vein, Bouteska et al. [39] explored the predictive power of investor sentiment for Bitcoin returns during the COVID-19 pandemic. Using computational text analysis on social media messages, they constructed a sentiment index and analyzed its impact on cryptocurrency market returns through vector autoregression.

sive analysis. Their findings suggest that investor sentiment significantly influences Bitcoin returns, especially in the short term, demonstrating practical implications for investors.

### 2.2.7. Hybrid Models

Seven papers adopt hybrid approaches to improve cryptocurrency price prediction. Critien et al. [40] investigated the relationship between Twitter sentiment and Bitcoin price changes, employing neural network models to enhance prediction accuracy. Similarly, Girsang et al. [41] proposed a hybrid LSTM and GRU model integrating social network sentiment analysis to forecast cryptocurrency prices, surpassing benchmark models in performance. Serafini et al. [11] explored the sentiment-driven price prediction of Bitcoin, comparing statistical and deep learning approaches, highlighting the significance of sentiment in predicting market stocks. Chalkiadakis et al. [42] developed a hybrid ARDL-MIDAS-Transformer model, emphasizing the interplay between sentiment, price dynamics, and technology factors in cryptocurrency markets. Li et al. [43] focused on the sentiment analysis of Twitter data to predict ZClassic price fluctuations, revealing a strong correlation between predicted and actual price movements. Valencia et al. [44] leveraged machine learning techniques to predict cryptocurrency price movements across multiple coins, showcasing the effectiveness of sentiment analysis in forecasting market behavior. Finally, Amirshahi et al. [45] proposed a hybrid neural network model utilizing sentiment extracted from tweets for cryptocurrency price prediction, shedding light on the role of social media sentiment in financial forecasting.

### 2.2.8. Other Techniques

Finally, eight papers employed a diverse set of technologies, including sentiment analysis, statistical methods, and text preprocessing techniques, to comprehend and forecast cryptocurrency market behavior. Lamon et al. [5] leveraged labeled news and social media data for cryptocurrency price prediction, utilizing sentiment analysis as a core technology to directly correlate sentiment with market behavior. Sattarov et al. [46] applied sentiment analysis specifically to Twitter data to forecast Bitcoin price movements, showcasing the predictive power of social media sentiment analysis. Oikonomopoulos et al. [47] utilized Granger causality testing alongside the sentiment analysis of Twitter data to identify causal relationships between sentiment and cryptocurrency price fluctuations, demonstrating the integration of statistical techniques with sentiment analysis technologies. Pano et al. [48] delved into text preprocessing techniques as a technology to enhance sentiment analysis accuracy, emphasizing the importance of refining preprocessing methods for effective prediction models during market volatility. Kyriazis et al. [49] employed sentiment analysis measures from Twitter to analyze the influence of social media sentiment on cryptocurrency returns and volatility, highlighting the utilization of sentiment analysis tools for quantitative analysis in financial research. Gaies et al. [50] utilized sentiment analysis to examine the interactions between investor sentiment and Bitcoin prices, showcasing the application of sentiment analysis technologies to understand sentiment-driven dynamics in cryptocurrency markets. Naeem et al. [51] utilized sentiment analysis tools, including the FEARS index and Twitter happiness sentiment, to predict cryptocurrency returns, underscoring the utilization of sentiment analysis technologies for predictive modeling in financial markets. Rognone et al. [52] employed sentiment analysis to compare news sentiment between the cryptocurrency market and Forex, demonstrating the application of sentiment analysis technologies for comparative analysis across financial markets.

## 3. Materials and Methods

This study initially focuses on utilizing the innovative GPT-4 LLM, alongside a parallel comparison with BERT and FinBERT NLP models. All the models have demonstrated proficiency in comprehending human text. In this proposed approach, the models will undergo fine-tuning on a dataset using few-shot learning, and their sentiment analysis capabilities will be evaluated through direct comparisons before and after fine-tuning.

In the early stages of our research, we rigorously assess GPT-4's ability to accurately identify sentiments within a crypto news article. This involves conducting an exhaustive performance evaluation by comparing the base GPT-4 model with its fine-tuned counterpart. Subsequently, we proceed to fine-tune the BERT and FinBERT models for cryptocurrency sentiment analysis. We compare the fine-tuned LLM and NLP models, demonstrating how the fine-tuning process enhances their capability to effectively address the complex challenges of sentiment analysis with improved precision.

Our research findings confirm that both the LLMs and NLP models are highly accurate in their predictions, with the fine-tuned GPT demonstrating significantly higher performance compared to its base version and the BERT model. The complete codebase used in this study, including dataset cleaning classes, fine-tuning procedures, and datasets, is openly available in a GitHub repository under the MIT open-source license [53].

To meet this paper's goals, a tailored research methodology was crucial, chosen to facilitate data cleaning, feature engineering, model implementation, and fine-tuning. Subsequent subsections elaborate on this framework, offering a comprehensive strategy for the study.

### 3.1. Dataset Splitting, Cleaning, and Preprocessing

For our study, we utilized the Crypto News + dataset compiled by Oliviervha from various sources including cryptonews.com, cryptopotato.com, and cointelegraph.com [54]. Available on Kaggle under the Database Contents License (DbCL) v1.0, this dataset offers comprehensive data for cryptocurrency sentiment analysis. With a usability rating of 8.24, and containing 31,037 rows and six columns, it holds promise for uncovering insights into cryptocurrency sentiment analysis.

To ensure the quality and effectiveness of our predictive modeling and fine-tuning procedures, we meticulously prepared the dataset through a systematic approach comprising several essential steps designed to enhance its suitability for classification and sentiment analysis tasks.

Initially, we parsed the 'sentiment' column, extracting solely the class label from dictionaries like `{'class': 'negative', 'polarity': -0.01, 'subjectivity': 0.38}`. Following this, we focused on data preprocessing to enhance dataset quality, specifically removing special characters and unnecessary white spaces to maintain data integrity and coherence. This step was crucial to ensure compatibility for modeling purposes and prevent unwanted noise that could affect NLP model performance.

Additionally, text normalization played a crucial role in ensuring uniformity and standardization of text data. Operations included converting accented characters to their base forms, ensuring consistent treatment of words with accent variations. Furthermore, to achieve case-insensitivity, we systematically converted all text to lowercase.

For our few-shot learning study, we randomly selected 5000 rows from the dataset of 31,037, ensuring equal distribution of labels across the three sentiment categories.

In NLP tasks, dataset division is pivotal for effective model development, refinement, and evaluation. Our methodology involved splitting the dataset into training, validation, and test sets, facilitating model learning from input data during fine-tuning, tuning hyperparameters for enhanced generalization, and ultimately evaluating model performance through predictions on the test set.

Fine-tuning BERT or its variants (like FinBERT) typically requires label encoding due to the nature of the underlying frameworks (like PyTorch and TensorFlow). These frameworks expect numerical labels for classification tasks. Although higher-level libraries (such as Hugging Face's Transformers) allow working with string labels at a more abstract level, internally, these labels are converted to numerical values. For example, when using Hugging Face's Transformers library, you can specify labels as strings in the configuration or during dataset preparation, but the framework will encode these labels into integers for processing by the model. For our classification task, we selected transforming the string sentiment labels into integers using nominal encoding: positive as 1, negative as 0, and

neutral as 2, without implying any order. Additionally, we employed cross-entropy loss, the default loss function in `BertForSequenceClassification`, which is well suited for classification tasks with nominally encoded labels.

### 3.2. GPT Prompt Engineering

Our objective was to develop a prompt that seamlessly integrates with various LLMs, enhancing the accessibility of their results through our code. We focused not only on the prompt's content but also on its output formatting. To provide further clarity on the process of deriving the prompt design, we elaborate on several key aspects:

1. Initial research and understanding: The process began with a thorough examination of various language model architectures, including but not limited to GPT-3, GPT-4, LLaMA-2, and others. This involved understanding their capabilities, limitations, and unique features. This foundational research ensured that our prompt design would be compatible and effective across a diverse range of models.
2. Identification of challenges: We identified two primary challenges in crafting the prompt: (a) creating content that would be model-agnostic and (b) ensuring the accessibility of the output format [55]. The former required developing a prompt that could be understood and responded to by any LLM, regardless of its architecture. The latter involved designing an output format that would facilitate ease of use and integration into code, particularly emphasizing JSON formatting for versatility and compatibility.
3. Finalization and validation: After rigorous testing and refinement, we arrived at a finalized prompt that effectively elicited responses in the desired output format. Validation through Source Code 1 demonstrates the prompt's ability to generate meaningful responses comprehensible to all models, thus fulfilling our objectives of accessibility and compatibility.

```
conversation.append({'role': 'system', 'content': "You are a crypto expert."})
conversation.append({'role': 'user',
'content': 'Evaluate the sentiment of the news article. Return your response in JSON format
{"sentiment":'
"negative"} or {"sentiment": "neutral"} or {"sentiment": "positive"}. Article:\n' +
input['text'] + "})
```

- Source Code 1. Model-agnostic prompt.

### 3.3. Model Deployment, Fine-Tuning, and Predictive Evaluation

As discussed earlier in this study, we employ two distinct models to address sentiment analysis and classification tasks within the cryptocurrency sector. While previous research has explored cryptocurrency sentiment analysis using various algorithms and techniques, there is a significant gap in utilizing LLMs for this purpose. In this study, we have decided to extensively test the GPT-4, BERT, and FinBERT models both before and after fine-tuning. Below, we present an overview of how each model is deployed, showcasing our innovative approach to utilizing them for this specific task.

#### 3.3.1. GPT Model Deployment and Fine-Tuning

In this stage, we utilized the GPT-4 Turbo base model, also referred to as `gpt-4-0125-preview`, for the sentiment analysis and classification of news articles. [56]. Leveraging the extensive pre-training of large language models (LLMs), our approach enabled the model to discern nuanced cues within the textual content of news articles. This section details the fine-tuning process to adapt the pre-trained model for our specific task.

##### Conceptual Background of LLM Fine-Tuning

Fine-tuning an LLM involves taking a pre-trained language model and training it further on a specific dataset to adapt it for a particular task. This process allows the model

to learn task-specific patterns and nuances that may not be covered during its general pre-training phase.

The GPT-4 model, a closed-source LLM, is exclusively accessible through its API. Users can directly interact with the model or fine-tune it for specific tasks via the API or the user interface/website. They do not need to handle the fine-tuning process or import any libraries; instead, they simply provide the appropriate hyperparameters and ensure that the training and validation sets are in the appropriate format. According to Azure CTO Mark Russinovich [57], the model will be fine-tuned using default training techniques, including:

- Low-rank adaptation (LoRA): this technique reduces the number of trainable parameters by decomposing the weight matrices into lower-rank matrices, making the training process more efficient.
- Parameter-efficient fine-tuning (PEFT): PEFT methods focus on fine-tuning only a small subset of the model's parameters, reducing the computational load and memory usage.
- DeepSpeed: an optimization library that facilitates efficient large-scale model training by improving GPU utilization and memory management.

### Fine-Tuning Phase

We fine-tuned the gpt-4-0125-preview base model using the official OpenAI API, incorporating the following steps:

1. Data preparation: We generated two JSONL files for training and validation, containing prompt-completion pairs (Source Code 2). As described in Section 3.1, these pairs were derived from our training and validation CSV files, which contain the textual content of news articles and their corresponding sentiment labels.
2. Training configuration: the fine-tuning was conducted with the following hyperparameters:
  3. Job ID: `ft:gpt-4-turbo-0125:personal::9Aa7kY0h`.
  4. Total tokens: 765,738.
  5. Epochs: 3.
  6. Batch size: 6.
  7. Learning rate (LR) multiplier: 8.
  8. Seed: 65426760.
9. Training process: The model underwent a multi-epoch training strategy, iteratively refining its understanding and capabilities. The initial training loss was 0.9778, which gradually decreased to nearly 0.1, indicating significant improvement in model performance over time.

### Sentiment Analysis and Evaluation Phase

Following fine-tuning, the model was deployed to perform sentiment analysis on a test set of news articles. The process included the following steps:

1. Prediction generation: The fine-tuned model was tasked with predicting the sentiment of each article in the test set by making calls to the OpenAI API, specifying the ID of the fine-tuned model. Analyzing the textual content, the model examined the articles and generated sentiment labels based on the patterns learned during fine-tuning, presenting the results in a JSON format.
2. Comparison with original labels: The predicted sentiment labels were rigorously compared against the original labels in the dataset. This comparison facilitated a comprehensive analysis to assess the model's effectiveness in capturing article sentiments and its alignment with human judgments.
3. Integration of results: the outcomes of the sentiment analysis were integrated into the `test_set.csv` file, providing a consolidated view for subsequent comparative analyses.

4. Performance metrics: To evaluate the model's performance, standard metrics such as accuracy, precision, recall, and F1-score were computed. These metrics provided insights into the model's ability to generalize and perform accurately on unseen data.

The same prediction and evaluation procedure was applied to the GPT-4 base model using zero-shot training. This allowed for a direct comparison between the base model and the fine-tuned model, underscoring the necessity of fine-tuning for specific tasks.

```
{"messages": [  
    {"role": "system", "content": "You are a crypto expert."},  
    {"role": "user", "content": "Evaluate the sentiment of the news article."  
    "Return your response in JSON format {\\"sentiment\\": \\"negative\\"}  
    "or {\\"sentiment\\": \\"neutral\\"} or {\\"sentiment\\": \\"positive\\"}."  
    "Article:\n ..."},  
    {"role": "assistant", "content": "{\"sentiment\": \"negative\""}  
]
```

- Source Code 2. Fine-tuning training and validation samples in JSONL format.

### 3.3.2. BERT and FinBERT Model Deployment and Fine-Tuning

In this phase, we utilized both the `bert-base-uncased` and FinBert models [58,59]. FinBert is specifically designed for financial sentiment classification, achieved by further training the BERT language model on a large financial corpus [60].

#### Training Phase

1. Data preparation: the dataset, as described in Section 3.1, was prepared for our task, resulting in the creation of three files: test, train, and validation CSV files.
2. Training configuration: both `bert-base-uncased` and FinBert models were trained with the following hyperparameters:
  3. Optimizers: Adam and AdamW.
  4. Epochs: 3.
  5. Batch size: 6.
  6. Learning rate: 6.
  7. Maximum sequence length: 512.
8. Training process: The training procedure took place in the Google Colab environment. Each model underwent three epochs of training, with progress monitored using the `tqdm` library. The training involved backpropagation, optimization, and validation, leveraging an A100 GPU for accelerated computations. Post-training, the fine-tuned models and tokenizers were saved to a directory in Google Drive for future use.

#### Sentiment Analysis and Evaluation Phase

Following training, both models were deployed to perform sentiment analysis on the same test set previously used by the GPT models for predictions. The resulting predictions were saved in two separate columns within the same test set. Subsequently, the evaluation process was conducted in a manner similar to that described in Section 3.3.1. The code, along with training and validation metrics, is available in an `ipynb` Jupyter file on GitHub [53].

## 4. Results

In Section 3, we explore the methodological approach used to assess the predictive abilities of the GPT-4 LLM, BERT, and FinBERT NLP models in crypto sentiment analysis and classification, seeking to understand their underlying complexities. This section reveals the results of the comparative analysis conducted between these three models at various stages of our research.

Before presenting our findings, it is crucial to emphasize the importance of model evaluation. In the fields of machine learning and NLP, evaluating models is crucial for understanding their performance. This is especially true in finance and cryptocurrencies,

where informed decisions are paramount. Table 1 summarizes a comprehensive range of evaluations for each model, incorporating a variety of essential evaluation metrics. In this table, the “Pr” represents the precision metric, “Re” is the recall metric for each class, “F1” signifies the F1-score, and “T” indicates the time in seconds required to fine-tune a model. Notably, the 0 class denotes the negative, the 1 class signifies the positive, and the 2 class indicates neutrality.

**Table 1.** Model performance metrics comparison.

Model	Accuracy	Pr_0	Pr_1	Pr_2	Re_0	Re_1	Re_2	F1_0	F1_1	F1_2	T
base:gpt-4-0125-preview	0.829	0.832	0.855	0.803	0.85	0.796	0.841	0.841	0.824	0.822	-
ft:gpt-4	0.867	0.873	0.888	0.842	0.865	0.859	0.877	0.869	0.873	0.859	5518
ft:bert-adamw	0.801	0.713	0.817	0.917	0.88	0.793	0.731	0.788	0.805	0.813	107.38
ft:bert-adam	0.833	0.797	0.849	0.856	0.838	0.826	0.835	0.817	0.837	0.845	74.94
ft:finbert-adamw	0.836	0.817	0.867	0.828	0.856	0.784	0.868	0.836	0.823	0.848	132.29
ft:finbert-adam	0.843	0.863	0.804	0.865	0.829	0.85	0.85	0.846	0.826	0.858	91.76

#### 4.1. GPT Base Model Evaluation Phase

In the initial phase of our study, we used the GPT-4 LLM base model, specifically gpt-4-0125-preview, to conduct sentiment analysis and classification on cryptocurrency news articles. We employed the test set consisting of 1000 articles with the goal of assessing the accuracy of the GPT base model in categorizing each article’s sentiment (positive, neutral, or negative).

Upon thorough analysis of the model’s outputs and comparison with the original user-provided labels, we made several notable observations. The gpt-4-0125-preview model exhibited a commendable level of accuracy, correctly predicting the sentiment class for 82.9% of the articles. This equates to successful predictions for 829 out of the 1000 articles in our test dataset. These results demonstrate that even the base model, without specific fine-tuning, shows a significant ability to infer sentiment from content based on contextual clues.

It is important to highlight that achieving an 82.9% accuracy rate in predictive modeling is quite high, indicating that the model can be confidently applied to sentiment analysis and classification tasks. Nonetheless, further improvements can be achieved by fine-tuning the models to better suit the specific task at hand. This refinement process can enhance their ability to capture and understand complex patterns and relationships within the data.

#### 4.2. Fine-Tuned Models Evaluation Phase

In the subsequent phase of our research, we focused on fine-tuning the GPT-4, BERT, and FinBERT models using two different optimizers. This fine-tuning process was conducted using a training set of 3200 cryptocurrency articles, with the primary objective of enhancing the models’ performance and their ability to analyze cryptocurrency sentiment.

From Table 1, we draw valuable insights showcasing the significant applicability of both LLMs and NLP models for the specific task, particularly after fine-tuning. The fine-tuned GPT-4 model emerges as the top performer, boasting an impressive accuracy rate of 86.7%, closely followed by FinBERT with 84.3%, and BERT with 83.3%. Notably, it is crucial to highlight that both the BERT and FinBERT models achieved higher accuracy when using the Adam optimizer instead of AdamW.

Precision values, indicating the models’ ability to accurately classify instances within each class, vary across models and classes. Interestingly, the GPT base and fine-tuned models demonstrate relatively higher precision in predicting positive labels (class 1), while both the BERT and FinBERT models excel in predicting neutral labels (class 2), suggesting a higher tendency for false positives in the class 0 and 1.

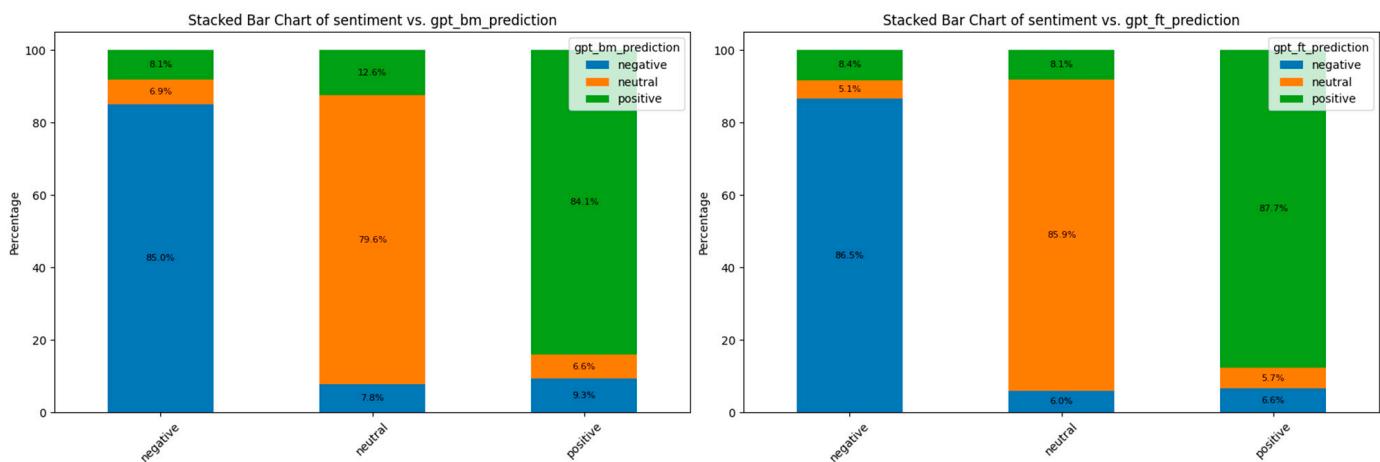
The recall values reflect the models’ proficiency in correctly capturing instances of each class. For instance, the ft:finbert-adamw model exhibits high recall for class 2, highlighting its effectiveness in identifying instances belonging to that class.

The F1-score, striking a balance between precision and recall for each class, showcases the `ft:gpt-4` model's superior performance with the highest F1-score for class 1, indicating its balanced precision and recall for that class.

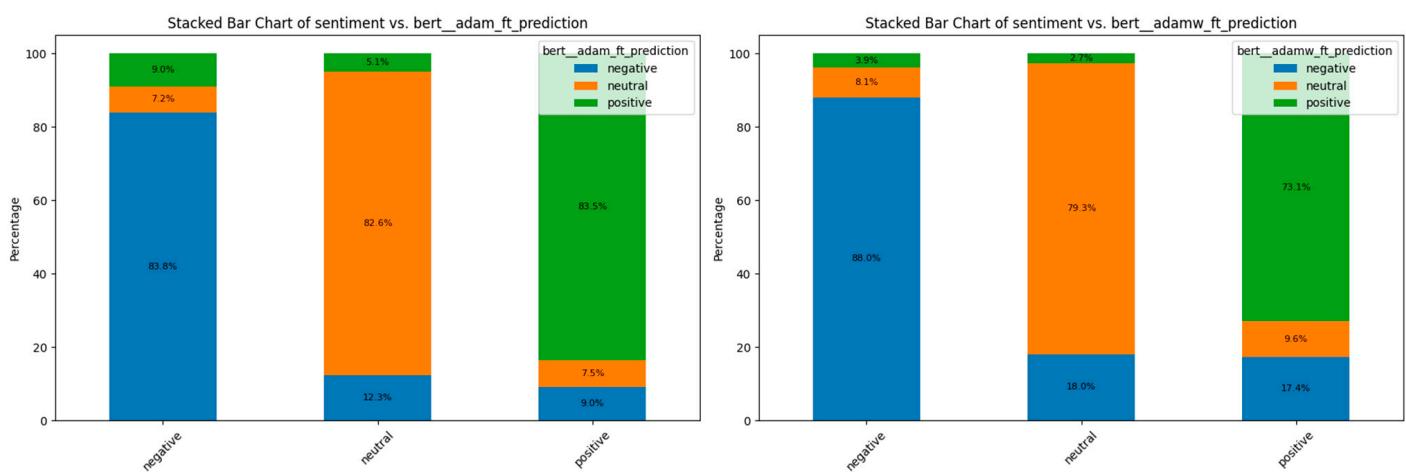
In summary, while each model displays strengths in specific metrics, the fine-tuned GPT-4 model emerges as the standout performer with high accuracy, balanced precision and recall, and low mean absolute error, suggesting its superior performance across multiple evaluation criteria. Nevertheless, the choice of model may vary depending on specific use cases or priorities, where other models may offer advantages in certain aspects such as recall for specific classes or precision in particular scenarios.

#### 4.3. Assessing Models' Performance and Proximity with Original Labels

To visualize our findings, we chose to use stacked bar charts generated by the `crosstab` function from the `pandas` library (Figures 1–3). This method will provide readers with more detailed insights into the distribution of predictions compared to the corresponding sentiment labels given by users.



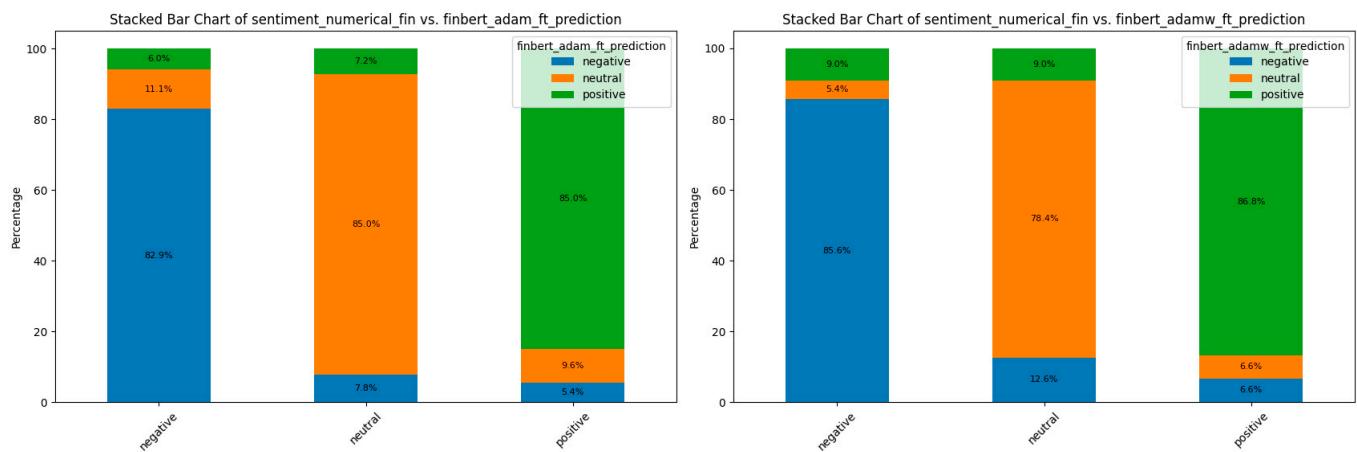
**Figure 1.** Stacked bar charts comparing the GPT-4 base model with its fine-tuned counterpart.



**Figure 2.** Stacked bar charts comparing the fine-tuned BERT models using different optimizers.

Analyzing the stacked bar charts offers valuable insights into model behavior after fine-tuning. In Figure 1, both the base and fine-tuned GPT-4 models demonstrate a balanced distribution of predictions for positive and negative labels. However, the base GPT-4 model exhibits significantly more errors in predicting neutral labels compared to its fine-tuned

counterpart, highlighting the necessity of fine-tuning to improve accuracy in predicting neutral sentiment.



**Figure 3.** Stacked bar charts comparing the fine-tuned FinBERT models using different optimizers.

Continuing this analysis in Figure 2, focusing on the fine-tuned BERT models, we observe similar challenges with the ADAMW optimizer in accurately identifying positive labels, despite achieving 88% accuracy in predicting negative labels.

Armed with these insights, further optimizations can be pursued through adjustments to enhance overall model performance.

## 5. Discussion

In earlier sections, we delved into the methodology and outcomes related to the predictive performance of the GPT-4 LLM, BERT, and FinBERT NLP models before and after fine-tuning for cryptocurrency news sentiment analysis and classification. This section unveils the research discoveries and insights gained by the authors concerning the effectiveness of LLMs and NLP models as valuable tools for sentiment analysis in the cryptocurrency domain, addressing pertinent research inquiries.

### 5.1. Research Findings

In Section 4.2, we presented the outcomes of our study on the effectiveness of the fine-tuned GPT-4, BERT, and FinBERT models in analyzing crypto sentiment. We used a set of 1000 crypto news articles for this investigation. Remarkably, the models demonstrated almost flawless prediction accuracy post-fine-tuning with few-shot learning, achieving a maximum of 86.7%. Specifically, the fine-tuned GPT model achieved 86.7%, while BERT achieved 83.3% accuracy when using the ADAM optimizer. Interestingly, using the ADAMW optimizer led to a slight decrease in accuracy, around 3.2% lower compared to the ADAM optimizer. The FinBERT model presented slightly higher accuracy, with 83.6% when trained using the ADAMW optimizer and 84.3% when trained using the ADAM optimizer. These results imply that even subtle adjustments in fine-tuning parameters can yield diverse outcomes, underscoring the complexities of model optimization.

Moreover, both the BERT and FinBERT models with the ADAM optimizer nearly matched the predictive capability of a robust LLM, differing by only 3.4% and 2.4%, respectively. This small variation could be attributed to differences in their pre-training phases, where the GPT-4 model had the advantage of exposure to a wider and more varied set of public datasets. This extensive training likely facilitated more comprehensive and targeted fine-tuning, granting the GPT model a deeper comprehension of nuanced cryptocurrency themes. Consequently, the GPT model might be better positioned to produce precise and contextually relevant responses.

It is important to note that each model exhibits higher accuracy for specific labels. For instance, the fine-tuned GPT-4 model is more accurate in predicting positive labels,

while BERT and FinBERT excel in predicting neutral labels. This highlights that each model has its own strengths and weaknesses, suggesting that a hybrid approach would be more effective in maximizing results in a production environment.

Additionally, there are more specialized variants of BERT models, such as CryptoBERT, which are trained on a broader range of data sources, including crypto news articles, Telegram, and social media [33,61].

Finally, it is imperative to highlight that even without fine-tuning, the GPT-4 base model achieved 82.9% accuracy. This demonstrates that LLMs, due to their extensive pre-training with billions of parameters, can perform tasks with high accuracy whether fine-tuned on a specific dataset or not. Such capabilities could lead to the development of zero-shot, simple yet accurate tools for non-specialized teams and organizations.

### 5.2. The Impact of Fine-Tuning

Customizing LLMs and NLP models via fine-tuning for specialized tasks within domains like finance and cryptocurrency is essential to optimize their effectiveness in practical applications. Models such as OpenAI's GPT-4 or BERT are initially trained on vast text datasets from diverse sources, which equips them to understand and generate human-like text across various domains. However, fine-tuning these models for specific tasks like sentiment analysis significantly boosts their performance and relevance in focused contexts. For instance, the FinBERT model, a variant of BERT, is fine-tuned on specialized financial label and unlabeled data. Moreover, even after fine-tuning, these models can undergo further fine-tuning for specific tasks.

Based on the results derived from this study, fine-tuning an LLM or an NLP model on large and representative datasets allows it to gain a deeper comprehension of the nuances and patterns unique to sentiment analysis, such as in the crypto domain, leading to more accurate predictions. This adaptation involves learning to recognize crypto-related terminology, grasp context, and extract relevant features from textual inputs. Through iterative adjustments in the fine-tuning process, using both training and validation data, the model becomes increasingly proficient at capturing these domain-specific intricacies, resulting in improved accuracy and other performance metrics.

### 5.3. Exploring the Cost and Usability of LLMs and NLP Models

Our research revealed that both the GPT and NLP models achieved high accuracy in the sentiment analysis of cryptocurrency news articles. While the fine-tuned GPT model showed slightly higher accuracy compared to BERT and FinBERT, it is important to note that BERT is an open-source model available for free use in a self-hosted environment. On the other hand, GPT-4 requires access through the official OpenAI API, which incurs a cost.

For stakeholders or investors seeking accurate and immediate predictions, the GPT API, despite its cost, offers a more convenient and sustainable solution due to its user-friendly interface. Conversely, for companies equipped with an AI team, deploying a self-hosted solution with BERT or similar NLP models may be a more cost-effective choice.

Specifically, within the cryptocurrency sector, where prices can fluctuate within minutes, software equipped with NLP technologies can detect market trends through news blogs, Telegram, or socials, providing crypto investors with information to adjust their strategies and potentially maximize profits. For instance, if a global event like a conflict between two countries is imminent, an investor could rely on the model to analyze user comments on platforms like Telegram and preemptively sell certain cryptocurrencies to avoid a significant drop in their value. Conversely, investors could also take advantage of low-cost cryptocurrencies that, based on crypto signals, might rapidly increase in value, leading to substantial gains.

Nevertheless, it is prudent to caution investors about spam comments or articles deliberately created to stimulate investment. For example, many articles are written daily speculating that Ethereum could reach \$10,000 by 2025. Such articles serve two purposes:

generating clicks and enticing inexperienced investors to invest in the cryptocurrency with the expectation of substantial value multiplication.

## 6. Conclusions

In summary, the integration of LLMs and NLP models for cryptocurrency sentiment analysis represents a powerful toolset that enhances investment decision-making in the dynamic cryptocurrency market. This study showcases the efficacy of state-of-the-art models like GPT-4 and BERT in accurately interpreting and categorizing sentiments extracted from cryptocurrency news articles. The key strength of these models lies in their ability to capture subtle shifts and complexities in market sentiments, enabling investors to navigate the volatile cryptocurrency landscape with greater insight and confidence.

By leveraging advanced NLP capabilities, including few-shot fine-tuning processes, this study highlights the adaptability and robustness of LLMs and NLP models in analyzing sentiment data within the cryptocurrency domain. These findings underscore the transformative potential of advanced NLP techniques in empowering investors with actionable insights, enabling them to proactively manage risks, identify emerging trends, and optimize investment strategies to maximize returns.

Specifically, the application field of the research results extends to cryptocurrency investment and risk management. By providing a deeper understanding of sentiment dynamics in cryptocurrency markets, this study facilitates a more informed and data-driven approach to cryptocurrency investment, enabling investors to make well-informed decisions based on the real-time sentiment analysis of news articles and other relevant sources.

**Author Contributions:** Conceptualization, K.I.R. and N.D.T.; methodology, K.I.R., N.D.T. and D.K.N.; software, K.I.R.; validation, K.I.R. and N.D.T.; formal analysis, K.I.R., N.D.T. and D.K.N.; investigation, K.I.R., N.D.T. and D.K.N.; resources, K.I.R.; data curation, K.I.R.; writing—original draft preparation, K.I.R. and N.D.T.; writing—review and editing, K.I.R., N.D.T. and D.K.N.; visualization, K.I.R.; supervision, D.K.N. and N.D.T. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** Data supporting the reported results can be found at ref. [53].

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Huang, X.; Zhang, W.; Tang, X.; Zhang, M.; Surbiryala, J.; Iosifidis, V.; Liu, Z.; Zhang, J. LSTM Based Sentiment Analysis for Cryptocurrency Prediction. In *Lecture Notes in Computer Science*; Springer: Berlin/Heidelberg, Germany, 2021; Volume 12683, pp. 617–621. [[CrossRef](#)]
2. Azmina, N.; Zamani, M.; Liew, J.; Yan, S.; Yusof, A.M. XLNET-GRU Sentiment Regression Model for Cryptocurrency News in English and Malay. *ACL Anthol.* **2022**, *24*, 36–42.
3. Sakas, D.P.; Giannakopoulos, N.T.; Margaritis, M.; Kanellos, N. Modeling Supply Chain Firms' Stock Prices in the Fertilizer Industry through Innovative Cryptocurrency Market Big Data. *Int. J. Financ. Stud.* **2023**, *11*, 88. [[CrossRef](#)]
4. Chen, C.Y.H.; Hafner, C.M. Sentiment-Induced Bubbles in the Cryptocurrency Market. *J. Risk Financ. Manag.* **2019**, *12*, 53. [[CrossRef](#)]
5. Lamon, C.; Nielsen, E.; Redondo, E. Cryptocurrency Price Prediction Using News and Social Media Sentiment. *SMU Data Sci. Rev.* **2017**, *1*, 1–22.
6. Sakas, D.P.; Giannakopoulos, N.T.; Kanellos, N.; Migkos, S.P. Innovative Cryptocurrency Trade Websites-Marketing Strategy Refinement, via Digital Behavior. *IEEE Access* **2022**, *10*, 63163–63176. [[CrossRef](#)]
7. Rouhani, S.; Abedin, E. Crypto-Currencies Narrated on Tweets: A Sentiment Analysis Approach. *Int. J. Ethics Syst.* **2020**, *36*, 58–72. [[CrossRef](#)]
8. Erdoğan, M.C.; Canayaz, M. Crypto-Currency Sentiment Analyse on Social Media. In Proceedings of the 2018 International Conference on Artificial Intelligence and Data Processing, IDAP 2018, Malatya, Turkey, 28–30 September 2018. [[CrossRef](#)]
9. Sureshbhai, P.N.; Bhattacharya, P.; Tanwar, S. KaRuNa: A Blockchain-Based Sentiment Analysis Framework for Fraud Cryptocurrency Schemes. In Proceedings of the 2020 IEEE International Conference on Communications Workshops, ICC Workshops 2020—Proceedings, Dublin, Ireland, 7–11 June 2020. [[CrossRef](#)]

10. Pant, D.R.; Neupane, P.; Poudel, A.; Pokhrel, A.K.; Lama, B.K. Recurrent Neural Network Based Bitcoin Price Prediction by Twitter Sentiment Analysis. In Proceedings of the 2018 IEEE 3rd International Conference on Computing, Communication and Security, ICCCS 2018, Kathmandu, Nepal, 25–27 October 2018; pp. 128–132. [[CrossRef](#)]
11. Serafini, G.; Yi, P.; Zhang, Q.; Brambilla, M.; Wang, J.; Hu, Y.; Li, B. Sentiment-Driven Price Prediction of the Bitcoin Based on Statistical and Deep Learning Approaches. In Proceedings of the International Joint Conference on Neural Networks 2020, Glasgow, UK, 19–24 July 2020. [[CrossRef](#)]
12. Jatoh, C.; Jain, R.; Fiore, U.; Chatharasupalli, S. Improved Classification of Blockchain Transactions Using Feature Engineering and Ensemble Learning. *Future Internet* **2022**, *14*, 16. [[CrossRef](#)]
13. Colianni, S.; Rosales, S.; Signorotti, M. Algorithmic Trading of Cryptocurrency Based on Twitter Sentiment Analysis. *CS229 Proj.* **2015**, *1*, 1–4.
14. Loginova, E.; Tsang, W.K.; van Heijningen, G.; Kerkhove, L.P.; Benoit, D.F. Forecasting Directional Bitcoin Price Returns Using Aspect-Based Sentiment Analysis on Online Text Data. *Mach. Learn.* **2021**, *113*, 4761–4784. [[CrossRef](#)]
15. Köse, O. Crypto Asset Taxonomy Classification and Crypto News Sentiment Analysis. Master’s Thesis, Middle East Technical University, Ankara, Turkey, 2020.
16. Bhatt, S.; Ghazanfar, M.; Amirkhosseini, M.H. Sentiment-Driven Cryptocurrency Price Prediction: A Machine Learning Approach Utilizing Historical Data and Social Media Sentiment Analysis. *Mach. Learn. Appl. Int. J.* **2023**, *10*, 1–15. [[CrossRef](#)]
17. Seroyizhko, P.; Zhexenova, Z.; Shafiq, M.Z.; Merizzi, F.; Galassi, A.; Ruggeri, F. A Sentiment and Emotion Annotated Dataset for Bitcoin Price Forecasting Based on Reddit Posts. In Proceedings of the FinNLP 2022—4th Workshop on Financial Technology and Natural Language Processing, Proceedings of the Workshop 2022, Abu Dhabi, United Arab Emirates, 8 December 2022; pp. 203–210. [[CrossRef](#)]
18. Parekh, R.; Patel, N.P.; Thakkar, N.; Gupta, R.; Tanwar, S.; Sharma, G.; Davidson, I.E.; Sharma, R. DL-GuesS: Deep Learning and Sentiment Analysis-Based Cryptocurrency Price Prediction. *IEEE Access* **2022**, *10*, 35398–35409. [[CrossRef](#)]
19. Raju, S.M.; Tarif, A.M. Real-Time Prediction of BITCOIN Price Using Machine Learning Techniques and Public Sentiment Analysis. *arXiv* **2020**, arXiv:2006.14473.
20. Pang, Y.; Sundararaj, G.; Ren, J. Cryptocurrency Price Prediction Using Time Series and Social Sentiment Data. In Proceedings of the BDCAT 2019—6th IEEE/ACM International Conference on Big Data Computing, Applications and Technologies 2019, New York, NY, USA, 2 December 2019; pp. 35–42. [[CrossRef](#)]
21. Farimani, S.A.; Jahan, M.V.; Fard, A.M.; Haffari, G. Leveraging Latent Economic Concepts and Sentiments in the News for Market Prediction. In Proceedings of the 2021 IEEE 8th International Conference on Data Science and Advanced Analytics, DSAA 2021, Porto, Portugal, 6–9 October 2021. [[CrossRef](#)]
22. Gadi, M.F.A.; Sicilia, M.A. Annotators’ Selection Impact on the Creation of a Sentiment Corpus for the Cryptocurrency Financial Domain. *IEEE Access* **2023**, *11*, 131081–131088. [[CrossRef](#)]
23. Anbaee Farimani, S.; Vafaei Jahan, M.; Milani Fard, A.; Tabbakh, S.R.K. Investigating the Informativeness of Technical Indicators and News Sentiment in Financial Market Price Prediction. *Knowl. Based Syst.* **2022**, *247*, 108742. [[CrossRef](#)]
24. Nasekin, S.; Chen, C.Y.-H. Deep Learning-Based Cryptocurrency Sentiment Construction. *Digit. Financ.* **2020**, *2*, 39–67. [[CrossRef](#)]
25. Wołk, K. Advanced Social Media Sentiment Analysis for Short-Term Cryptocurrency Price Prediction. *Expert Syst.* **2020**, *37*, e12493. [[CrossRef](#)]
26. Vo, A.-D. Sentiment Analysis of News for Effective Cryptocurrency Price Prediction. *Int. J. Knowl. Eng.* **2019**, *5*, 47–52. [[CrossRef](#)]
27. Aslam, N.; Rustam, F.; Lee, E.; Washington, P.B.; Ashraf, I. Sentiment Analysis and Emotion Detection on Cryptocurrency Related Tweets Using Ensemble LSTM-GRU Model. *IEEE Access* **2022**, *10*, 39313–39324. [[CrossRef](#)]
28. Dwivedi, D.N.; Vemareddy, A. Sentiment Analytics for Crypto Pre and Post COVID: Topic Modeling. In *Lecture Notes in Computer Science*; Springer: Berlin/Heidelberg, Germany, 2023; Volume 13776, pp. 303–315. [[CrossRef](#)]
29. Şaşmaz, E.; Tek, F.B. Tweet Sentiment Analysis for Cryptocurrencies. In Proceedings of the 6th International Conference on Computer Science and Engineering, UBMK 2021, Ankara, Turkey, 15–17 September 2021; pp. 613–618. [[CrossRef](#)]
30. Kraaijeveld, O.; De Smedt, J. The Predictive Power of Public Twitter Sentiment for Forecasting Cryptocurrency Prices. *J. Int. Financ. Mark. Inst. Money* **2020**, *65*, 101188. [[CrossRef](#)]
31. Gurrib, I.; Kamalov, F. Predicting Bitcoin Price Movements Using Sentiment Analysis: A Machine Learning Approach. *Stud. Econ. Financ.* **2022**, *39*, 347–364. [[CrossRef](#)]
32. Ider, D.; Lessmann, S. Forecasting Cryptocurrency Returns from Sentiment Signals: An Analysis of BERT Classifiers and Weak Supervision. *arXiv* **2022**, arXiv:2204.05781.
33. Kulakowski, M.; Frasincar, F.; Cambria, E. Sentiment Classification of Cryptocurrency-Related Social Media Posts. *IEEE Intell. Syst.* **2023**, *38*, 5–9. [[CrossRef](#)]
34. Nguyen Thanh, B.; Nguyen, A.T.; Chu, T.T.; Ha, S. Chatgpt, Twitter Sentiment and Bitcoin Return. *SSRN* **2023**. [[CrossRef](#)]
35. Raheman, A.; Kolonin, A.; Fridkins, I.; Ansari, I.; Vishwas, M. Social Media Sentiment Analysis for Cryptocurrency Market Prediction. *arXiv* **2022**, arXiv:2204.10185.
36. Georgoula, I.; Pournarakis, D.; Bilanakos, C.; Sotiropoulos, D.N.; Giaglis, G.M. Using Time-Series and Sentiment Analysis to Detect the Determinants of Bitcoin Prices. *SSRN Electron. J.* **2015**. [[CrossRef](#)]
37. Chalkiadakis, I.; Zaremba, A.; Peters, G.W.; Chantler, M.J. On-Chain Analytics for Sentiment-Driven Statistical Causality in Cryptocurrencies. *Blockchain Res. Appl.* **2022**, *3*, 100063. [[CrossRef](#)]

38. Anamika; Chakraborty, M.; Subramaniam, S. Does Sentiment Impact Cryptocurrency? *J. Behav. Financ.* **2023**, *24*, 202–218. [[CrossRef](#)]
39. Bouteska, A.; Mefteh-Wali, S.; Dang, T. Predictive Power of Investor Sentiment for Bitcoin Returns: Evidence from COVID-19 Pandemic. *Technol. Forecast. Soc. Change* **2022**, *184*, 121999. [[CrossRef](#)]
40. Critien, J.V.; Gatt, A.; Ellul, J. Bitcoin Price Change and Trend Prediction through Twitter Sentiment and Data Volume. *Financ. Innov.* **2022**, *8*, 45. [[CrossRef](#)]
41. Girsang, A.S.; Stanley. Hybrid LSTM and GRU for Cryptocurrency Price Forecasting Based on Social Network Sentiment Analysis Using FinBERT. *IEEE Access* **2023**, *11*, 120530–120540. [[CrossRef](#)]
42. Chalkiadakis, I.; Peters, G.W.; Ames, M. Hybrid ARDL-MIDAS-Transformer Time-Series Regressions for Multi-Topic Crypto Market Sentiment Driven by Price and Technology Factors. *Digit. Financ.* **2023**, *5*, 295–365. [[CrossRef](#)]
43. Li, T.R.; Chamrajnagar, A.S.; Fong, X.R.; Rizik, N.R.; Fu, F. Sentiment-Based Prediction of Alternative Cryptocurrency Price Fluctuations Using Gradient Boosting Tree Model. *Front. Phys.* **2018**, *7*, 98. [[CrossRef](#)]
44. Valencia, F.; Gómez-Espinosa, A.; Valdés-Aguirre, B. Price Movement Prediction of Cryptocurrencies Using Sentiment Analysis and Machine Learning. *Entropy* **2019**, *21*, 589. [[CrossRef](#)] [[PubMed](#)]
45. Amirshahi, B.; Lahmiri, S. Investigating the Effectiveness of Twitter Sentiment in Cryptocurrency Close Price Prediction by Using Deep Learning. *Expert Syst.* **2023**, e13428. [[CrossRef](#)]
46. Sattarov, O.; Jeon, H.S.; Oh, R.; Lee, J.D. Forecasting Bitcoin Price Fluctuation by Twitter Sentiment Analysis. In Proceedings of the 2020 International Conference on Information Science and Communications Technologies, ICISCT 2020, Tashkent, Uzbekistan, 4–6 November 2020. [[CrossRef](#)]
47. Oikonomopoulos, S.; Tzafilkou, K.; Karapiperis, D.; Verykios, V. Cryptocurrency Price Prediction Using Social Media Sentiment Analysis. In Proceedings of the 13th International Conference on Information, Intelligence, Systems and Applications, IISA 2022, Corfu, Greece, 18–20 July 2022. [[CrossRef](#)]
48. Pano, T.; Kashef, R. A Complete VADER-Based Sentiment Analysis of Bitcoin (BTC) Tweets during the Era of COVID-19. *Big Data Cogn. Comput.* **2020**, *4*, 33. [[CrossRef](#)]
49. Kyriazis, N.; Papadamou, S.; Tzeremes, P.; Corbet, S. The Differential Influence of Social Media Sentiment on Cryptocurrency Returns and Volatility during COVID-19. *Q. Rev. Econ. Financ.* **2023**, *89*, 307–317. [[CrossRef](#)]
50. Gaies, B.; Nakhli, M.S.; Sahut, J.M.; Schweizer, D. Interactions between Investors' Fear and Greed Sentiment and Bitcoin Prices. *North Am. J. Econ. Financ.* **2023**, *67*, 101924. [[CrossRef](#)]
51. Naeem, M.A.; Mbarki, I.; Shahzad, S.J.H. Predictive Role of Online Investor Sentiment for Cryptocurrency Market: Evidence from Happiness and Fears. *Int. Rev. Econ. Financ.* **2021**, *73*, 496–514. [[CrossRef](#)]
52. Rognone, L.; Hyde, S.; Zhang, S.S. News Sentiment in the Cryptocurrency Market: An Empirical Comparison with Forex. *Int. Rev. Financ. Anal.* **2020**, *69*, 101462. [[CrossRef](#)]
53. GitHub—Kroumeliotis/LLM-and-NLP-Models-in-Cryptocurrency-Sentiment-Analysis: Crypto. Available online: <https://github.com/kroumeliotis/LLM-and-NLP-models-in-Cryptocurrency-Sentiment-Analysis> (accessed on 12 April 2024).
54. Crypto News +. Available online: <https://www.kaggle.com/datasets/oliviervha/crypto-news> (accessed on 12 April 2024).
55. Zhang, K.; Zhou, F.; Wu, L.; Xie, N.; He, Z. Semantic Understanding and Prompt Engineering for Large-Scale Traffic Data Imputation. *Inf. Fusion* **2024**, *102*, 102038. [[CrossRef](#)]
56. Models—OpenAI API. Available online: <https://platform.openai.com/docs/models/gpt-4-and-gpt-4-turbo> (accessed on 1 April 2024).
57. What Runs ChatGPT? Inside Microsoft's AI Supercomputer | Featuring Mark Russinovich—YouTube. Available online: <https://www.youtube.com/watch?v=Rk3nTUfRZmo> (accessed on 17 December 2023).
58. BERT—Transformers 3.0.2 Documentation. Available online: [https://huggingface.co/transformers/v3.0.2/model\\_doc/bert.html#berttokenizer](https://huggingface.co/transformers/v3.0.2/model_doc/bert.html#berttokenizer) (accessed on 17 December 2023).
59. ProsusAI/Finbert · Hugging Face. Available online: <https://huggingface.co/ProsusAI/finbert> (accessed on 18 May 2024).
60. Araci, D.T. FinBERT: Financial Sentiment Analysis with Pre-Trained Language Models. *arXiv* **2019**, arXiv:1908.10063.
61. ElKulako/Cryptobert · Hugging Face. Available online: <https://huggingface.co/ElKulako/cryptobert> (accessed on 15 April 2024).

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.