CASE REPORT

Predicting Cryptocurrency Prices During Periods of Conflict: A Comparative Sentiment Analysis Using SVM, CNN-LSTM, and Pysentimento

Muhammad Nabil Rateb[1] · Sameh Alansary[2] · Marwa Khamis Elzouka[3] · Mohamad Galal[4]

Abstract

In this paper, we propose a method for predicting the prices and trends of cryptocurrencies using sentiment analysis and time series forecasting. In this study, more than one million tweets spanning 3 months (March, June, and December 2022) regarding three cryptocurrencies, Bitcoin (BTC), Ethereum (ETH), and Binance Coin (BNB) during the Russian-Ukrainian War, are considered. Two models, a convolutional neural network with long short-term memory (CNN-LSTM) and a support vector machine (SVM) with GloVe and TF-IDF features, are trained on a labeled dataset of more than fifty thousand tweets about Bitcoin labeled as positive, negative, and neutral. A pretrained model (Pysentimento) for sentiment analysis is also employed to compare the performances of the three models. The models are tested on the labeled dataset and then evaluated on the unlabeled tweets, revealing that Pysentimento's level of accuracy outperforms the other two models. Google Trends, along with the opening and closing prices, and the volume of the three cryptocurrencies, in addition to the results of Pysentimento sentiment classification, are employed to apply the Pearson correlation coefficient and conduct price prediction analysis using the SARIMA model. It is found that Bitcoin may appeal to those seeking stability and a known record of accomplishment, while Binance Coin and Ethereum may attract investors looking for more diverse opportunities. A data-centric approach that can provide valuable insights and predictions for the cryptocurrency market, especially in the context of the Russian-Ukrainian War, which poses significant challenges and uncertainties for investors and traders, is demonstrated.

Keywords Sentiment analysis · Google Trends · Russian-Ukrainian War · Pysentimento · CNN-LSTM · SVM

Extended author information available on the last page of the article

🖄 Springer

1 Introduction

Cryptocurrency forecasting has become an increasingly prevalent research topic, with sentiment analysis being one of the key techniques implemented to predict cryptocurrency prices. Political events have been shown to significantly impact cryptocurrency prices and sentiment analysis of social media. Google Trends data is an emphatic method of predicting these fluctuations. The Russian-Ukrainian War represents a substantial occurrence that potentially influences the oscillations of cryptocurrency prices.

Sentiment analysis of social media and Google Trends data necessitates the employment of computational methods to identify and extract subjective information, such as opinions and attitudes, from these sources of data. In the context of cryptocurrency price forecasting, sentiment analysis can postulate valuable apprehension regarding how the masses perceive a particular cryptocurrency or the cryptocurrency market as a whole. By analyzing the sentiment expressed in social media and Google Trends data, it is possible to gain a better understanding of how the market may behave in the short or long term [1]. However, there are challenges to conducting sentiment analysis of social media and Google Trends data for cryptocurrency price forecasting. These include the need for accurate and representative datasets, the difficulty in accurately classifying sentiment, and the potential for bias in the data. Despite these challenges, sentiment analysis of social media and Google Trends data remains a promising approach for improving the accuracy of cryptocurrency price forecasting [2].

Behavioral finance refers to the study of how psychological biases and emotions influence financial decision-making. In relation to cryptocurrency price forecasting, sentiment analysis of social media and Google Trends data can provide insight into the emotions and psychological biases of cryptocurrency market participants. For example, social media sentiment analysis can reveal how investors feel about a particular cryptocurrency and how sentiment fluctuates over time. Google Trends data can uncover search trends related to a particular cryptocurrency, which can offer insights into the level of interest and attention the cryptocurrency is receiving [3].

Cryptocurrency prices are highly volatile and can be influenced by various factors, including global events. During times of crisis, such as the Russian-Ukrainian War or the COVID-19 pandemic, traditional market indicators may not produce an accurate reflection of the market sentiment. This is where sentiment analysis of social media and Google Trends data becomes even more significant, as it can capture the real-time sentiment of market participants. This information can be exploited by traders, investors, and analysts to make informed decisions about their investments [4].

Presently, countless subscribers use social media platforms such as Twitter (presently X), Instagram, and Facebook to share their thoughts about their daily lives and express their emotions. X as a social media platform, is distinguished with a number of privileges to be chosen for this research. First of all, according to Matthew Woodward [5], globally, there are 436 million active Twitter subscribers per

month. Secondly, Twitter has varied access accounts using application programming interfaces (APIs). The Twitter API is a collection of programmatic endpoints that can be exploited to comprehend or create the Twitter discourse and scrap tweets [6]. Finally, Drus and Khalid [7] maintain that the majority of social media subscribers worldwide use Facebook. However, due to the unstructured nature of the data, its poor organization, frequent usage of short forms, and high rate of spelling errors, sentiment analysis is not commonly conducted using these data, but rather more frequently by means of X.

In the context of the Russian-Ukrainian War, sentiment analysis of social media and Google Trends data can reveal how geopolitical tensions are impacting the cryptocurrency market. For instance, if social media sentiment around Bitcoin is overwhelmingly positive during the war, it may be an indication that investors could view it as a safe-haven asset. Conversely, if sentiment is negative, it may indicate that investors are selling off their cryptocurrency holdings and moving to more conventional haven assets such as gold or the US dollar [8].

This paper comprises thirteen sections that flow logically and are organized according to their significance to the overall study. The second section begins by setting the groundwork, outlining the research objectives, and presenting the attending results, followed by the third section, which takes a step further by discussing the research questions addressed in the study. This is followed by the fourth section, which introduces the limitations of the study, providing a context for the interpretation of the results. The fifth section delves into the heart of the research by reviewing the literature on sentiment analysis in relation to cryptocurrencies, while the sixth section presents the research methodology in detail, including the procedures utilized and a comparison of SVM, CNN-LSTM, and Pysentimento, which forms the crux of the study. The seventh section presents the analysis and discussion of the study's findings, while the eighth section provides a second evaluation of the results. The ninth section conducts a comparative analysis, providing a broader perspective on the research findings. The tenth section pivots toward applying the research findings in real-world scenarios, focusing on Pysentimento and Google Trends, while the eleventh section delves into correlation analysis, deepening the understanding of the research findings. The twelfth section discusses price forecasting using SARIMA, further expanding the practical application of the research. Finally, the paper concludes with the thirteenth section, providing future research directions and summarizing the study's findings, which not only wraps up the research but also suggests avenues for further exploration.

2 Objectives and Attending Results

The study embarks on a comprehensive evaluation of predictive models—support vector machine (SVM) classifier, convolutional neural network-long short-term memory (CNN-LSTM) model, and Pysentimento sentiment analyzer—to identify the most accurate method for analyzing cryptocurrency markets during wartime. It aims to enhance the precision of market trend predictions and price forecasting. Via introducing an innovative sentiment analysis approach via Pysentimento, it outperforms

conventional machine learning and deep learning models in accuracy. In conjunction with this, the research utilizes Google Trends data, alongside the closing prices for March, June, and December 2022, to evaluate the trends of Bitcoin, Ethereum, and Binance cryptocurrencies. This involves correlating predictor variables like normalized closing prices and volumes with sentiment-related target variables, providing a detailed view of market dynamics and sentiment trends.

The research concludes with the validation of SARIMA model predictions through the calculation of RMSE, offering insights into the profitability and stability of the cryptocurrencies during times of conflict. The findings aim to guide investment strategies, supported by rigorous data analysis and statistical validation. The research also suggests future directions for refining predictive models and integrating advanced computational techniques to improve the precision and adaptability of analytics tools. Additionally, in addressing the gaps identified in the literature, this study extends the work of Abraham et al. [9] by not only analyzing Twitter sentiment but also incorporating Google Trends data and advanced machine learning models to predict cryptocurrency prices. Unlike Abraham's study, which found a predominance of neutral sentiment, our approach leverages the Pysentimento model to capture a broader range of sentiments, leading to more nuanced insights into market behavior. Abraham's study and the following ones will be discussed later in literature review.

Building on the findings of Hyunyoung and Varian [10], this study confirms the value of Google Trends data in forecasting cryptocurrency prices. However, we advance their approach by integrating this data with sentiment analysis, providing a more comprehensive understanding of market dynamics. Our methodology also differs from Valencia et al. [11] and Xin Huang [12, 13] in that we employ a pretrained sentiment analysis model, Pysentimento, which demonstrates superior performance compared to the machine learning methods they used. Our case study analysis reveals that Pysentimento outperforms traditional models like SVM and CNN-LSTM, achieving higher accuracy in sentiment classification, which in turn enhances the reliability of our price predictions.

Throughout a case study analysis during a significant geopolitical event—the Russian-Ukrainian War—our study offers new perspectives on the cryptocurrency market's response to external shocks. This period, represented by the months of March, June, and December 2022, provides a unique context for examining the resilience and volatility of cryptocurrencies, contributing fresh insights to the field. The empirical nature of this study is grounded in the analysis of over one million tweets related to Bitcoin (BTC), Ethereum (ETH), and Binance Coin (BNB), alongside the application of advanced sentiment analysis and time series forecasting methods. It provides a holistic understanding of the cryptocurrency market by integrating the best of the three models with other factors such as price and volume, and it establishes a novel and robust framework for cryptocurrency research that combines the most accurate algorithm with other techniques like correlation analysis and time series forecasting. This framework is designed to yield insights into sentiment and market trends and to predict future prices and network dynamics for a year marked by significant events, starting with March (the initiation of the war) and ending with December (the end of 2022).

3 Research Questions

(A) What are the differences between SVM, CNN-LSTM, and Pysentimento? Which method yields the best level of accuracy?
(B) Can sentiment analysis on tweets and Google Trends produce accurate predictions about the emotional tendencies and prices of Bitcoin, Ethereum, and Binance coin?
(C) If predictions based on social media are possible, do they correlate with the trend of the actual interest of the chosen cryptocurrency retrieved from Google Trends?

4 Limitations of the Study

The study's methodology, which leverages sentiment analysis and price forecasting to predict cryptocurrency prices during the Russian-Ukrainian War, not only provides valuable insights but also exhibits several limitations that merit careful consideration. The exclusive reliance on Twitter data may introduce a bias, as it does not necessarily reflect the sentiments of the entire cryptocurrency investor population, which is diverse and global. The use of public datasets, while accessible, brings into question the reliability and cleanliness of the data, potentially introducing noise that could skew the results. The Pysentimento model showed promise in sentiment analysis; however, its effectiveness across various datasets has not been thoroughly tested, raising concerns about its generalizability. Furthermore, the application of Google Trends data in sentiment analysis is limited by its inability to discern the complex motivations behind search queries, such as the intent and context of the searches, the presence of noise like irrelevant searches or bot activity, and the volatility of search trends over time. These factors, along with the study's narrow focus on a specific geopolitical event and the inherent unpredictability of cryptocurrency markets, suggest that the conclusions drawn may not be broadly applicable. Despite these challenges, the study underscores the potential utility of combining sentiment analysis with price forecasting and Pearson correlations in the analysis of cryptocurrency markets, while also highlighting the importance of cautious interpretation of the results and the need for ongoing research to refine the methodologies used.

5 Review of Literature

Despite the growing research interest in sentiment analysis of cryptocurrency price prediction in recent years, both Pysentimento as sentiment analysis classifier and Binance Coin have been paid less attention to by other studies, possibly for being relatively recent. As far as sentiment analysis for cryptocurrency price forecasting is concerned, there are various studies with variant objectives. The present review

covers them as they highlight the significant role of sentiment analysis in predicting cryptocurrency and stock market prices.

Abraham et al. [9] employs a sentiment analysis of Twitter to predict cryptocurrency prices. The data is analyzed to determine whether they can be useful information to the final model. VADER (Valence Aware Dictionary and sentiment Reasoner) is a sentiment analysis tool employed in natural language processing. It assigns polarity scores to words to discern whether they convey positive, negative, or neutral sentiment, thereby facilitating the analysis of sentiments in social media texts and informal communication. In this study, VADER finds tweets to be more neutral, which reduces the accuracy of the results if the general public's sentiment is neutral because neutral sentiment usually does not indicate a buying or selling trend. The ratio of tweets and Google Trends is strongly correlated with prices. To estimate the final daily price of Bitcoin, a linear regression technique is applied. Despite potential price fluctuations, sentiment on cryptocurrency in Twitter discussions tends to be positive. However, future research should consider employing more sophisticated models than linear regression to enhance results, particularly since the data used in this study was collected during a period of price increases.

Another study has conducted specific research on Google Trends data, which concludes that basic seasonal autoregressive models that incorporates Google Trends data as input data surpassed other models which do not involve Google Trends data by 5 to 20% [10]. However, it is challenging for such models to explain situations where the direction changes. In these situations, Google Trends data can be beneficial.

A different study introduces a system that forecasts the prices of four chosen cryptocurrencies—Bitcoin, Ethereum, Ripple, and Litecoin—using social media data and machine learning methods. Random forests, support vector machines, and neural networks (NNs) are applied in this model. The results of this model demonstrate the applicability of sentiment analysis and machine learning methods for cryptocurrency prediction. Furthermore, the prices of certain cryptocurrencies, typically those with a substantial follow-up, can be anticipated solely by Twitter data [11].

Xin Huang [12] proposes an LSTM model for sentiment analysis. Sina-Weibo, a famous social networking platform in China, provides the data to measure the sentiment. A recurrent neural network based on long short-term memory (LSTM) and historical cryptocurrency values is used to predict the price trend for cryptocurrencies. The results show an accuracy rate of 87%. This result is 15.4% better than that of the currently used traditional auto-regression method [13].

6 Research Methodology

6.1 Data Collection

In this research, data collection methods are dedicated to efficiently gathering and analyzing selected tweets and Google Trends. The process involves utilizing various tools, including the Twitter API, sentiment analysis, and Google Trends. To address the research inquiries effectively, these methods will be elaborated on

in the subsequent section. The primary data source for this study is an academic Twitter Application Programming Interface (API) account. This account facilitates the extraction of tweets over a 3-month period (March, June, and December 2022), focusing on keywords such as "Bitcoin," "Ethereum," and "Binance." The Twitter API, comprising programmatic endpoints, enables the comprehension and extraction of Twitter discourse [6]. Subsequently, the collected tweets are filtered based on language (English) and location (worldwide), resulting in a total collection ranging from 700,000 to over a million tweets per month. By utilizing an academic API account, researchers gain access to ten million tweets monthly, ensuring an adequate supply of necessary data. Simultaneously, Google Trends data is gathered using the PyTrends library in Python [14], focusing on the same keywords and timeframe. Additionally, Twitter data scraping is conducted using the Tweepy library [15].

Two distinct datasets are utilized for training and evaluation purposes. The first dataset consists of 50,859 tweets that have been categorized as ["positive"], ["negative"], and ["neutral"]. This dataset is obtained from the Kaggle website, which is a renowned platform and online community for data scientists and machine learning practitioners [16]. For the secondary assessment, a distinct dataset is employed, encompassing 100 annotated tweets categorized as positive, negative, and neutral. These tweets are drawn randomly from the researcher's collection of tweets pertaining to the three cryptocurrencies. Additionally, historical price data for the three cryptocurrencies (Bitcoin, Ethereum, and Binance) is obtained from Yahoo Finance, a comprehensive platform offering cryptocurrency market information encompassing market capitalization, trading volume, price trends, and news. It also provides an API that facilitates access to historical and real-time cryptocurrency data, with price data collected for the same periods as the tweets (March, June, and December 2022) along with the real prices for the first 3 months of 2023 (January, February, and March) for the sake of price forecasting [17].

Google Trends, in turn, serves as a tool to gauge the relative popularity of specific search phrases relative to others over time [18]. These data provide insights into the popularity of cryptocurrencies in 2022, aligning with the price charts and sentiment analysis results for Bitcoin, Ethereum, and Binance. The Google Trends data is accessed through the Google Trends API, which offers insights into the relative popularity of Bitcoin, Ethereum, and Binance for the same periods as the tweets (March, June, and December 2022).

6.2 Support Vector Machine (SVM)

According to Cortes and Vapnik [19], support vector machines (SVMs) are a type of supervised machine learning algorithm that can be used for both classification and regression tasks. SVMs work by finding a hyperplane that separates the data points of different classes with the maximum margin. The data points that are closest to the hyperplane are called support vectors, and they determine the position and orientation of the hyperplane. SVMs can handle both linear and nonlinear problems by using different kernel functions. The chosen model for this study is SVM with (GloVe) and Term Frequency-Inverse Document Frequency (TF-IDF). GloVe is a

word embedding technique capturing semantic and syntactic information in vector spaces. These word embeddings serve as numerical representations of words for machine learning algorithms, capturing word meanings, contexts, similarities, analogies, and polarities [20].

According to Ramos [21], TF-IDF is a statistical measure evaluating the importance of a word within a document or corpus. TF-IDF assigns weights to words based on their frequency in a document and rarity across the corpus. Higher TF-IDF scores indicate greater word relevance to the document. TF-IDF is utilized in this study to reduce word vector dimensionality and eliminate noise and stopwords. In SVM, classification into positive, negative, or neutral categories is accomplished by identifying the hyperplane that best separates data points into different classes.

6.3 Advantages and Disadvantages of SVM

One notable advantage of SVMs is their ability to attain high accuracy and exhibit strong generalization performance by identifying a hyperplane that effectively separates data points into distinct classes while maximizing the margin between them [19]. This unique attribute enables SVMs to mitigate the issues of overfitting and underfitting often encountered by other machine learning algorithms like neural networks or decision trees. In the context of sentiment analysis within the cryptocurrency domain, characterized by its volatile and unpredictable data, SVM's ability to provide a reliable and robust model becomes crucial for addressing the inherent uncertainty and variability of the data [22]. SVMs effectively classify text into positive, negative, or neutral categories based on emotional tone,accommodate various text lengths and formats; and employ features such as word counts (TF-IDF, word embeddings, or *n*-grams. These methods can synergize with other techniques, such as feature selection, dimensionality reduction, or ensemble methods, to enhance performance and accuracy [23, 24].

Despite their strengths, SVMs face certain limitations within the domain of cryptocurrency sentiment analysis. Primarily, their inherent binary classification nature necessitates extensions for multiclass sentiment classification, often required in cryptocurrency discussions characterized by multifaceted sentiment landscapes. The computational intensity of SVMs poses challenges when dealing with large volumes of real-time cryptocurrency data, where timeliness is crucial [25]. Additionally, tuning hyperparameters and kernel functions can be time-consuming and resource intensive, complicating SVM implementation [26]. SVMs' sensitivity to different kernel functions and hyperparameter settings introduces variability in performance, necessitating extensive trial-and-error approaches or grid searches to optimize results [23].

6.4 CNN-LSTM

In the present research, a hybrid model that combines convolutional neural networks (CNNs) and long short-term memory (LSTM), originally adapted from Kaggle [16], has been employed with the primary aim of enhancing the accuracy of outcomes.

For a comprehensive exploration of the CNN component, it is imperative to delve deeper into the characteristics of convolutional neural networks (CNNs). They are a category of deep learning algorithms specifically designed to address intricate tasks, including image recognition and natural language processing [27].

The fundamental operation of CNNs involves the application of filters or kernels to input data, which can encompass images or textual content, thereby facilitating the extraction of localized features or discernible patterns. Notably, these filters or kernels are acquired through network learning during the training phase, and they possess the capacity to encapsulate diverse aspects or dimensions of the data, encompassing attributes such as edges, shapes, colors, linguistic constructs, *n*-grams, and sentiment expressions [28]. It is pertinent to mention that CNNs may additionally employ pooling layers for the purpose of diminishing data intricacy and dimensionality, alongside fully connected layers to cater to classification or regression objectives [29].

Within this integrated framework, CNNs assume the role of extracting localized features or patterns, while LSTM specializes in capturing temporal dynamics and protracted dependencies inherent in the data [30]. This harmonious fusion of layers imparts to CNN-LSTM the ability to deliver superior performance and address more intricate challenges, surpassing the efficacy of deploying each layer in isolation [31]. The CNN outputs a vector encompassing all these features. LSTM then processes this vector to generate a sentiment score. LSTM learns to utilize the vector's features to compute a numerical sentiment value for the tweet, taking into account the tweet's context and tone. The final outcome is a sentiment score for each tweet, reflecting its positivity or negativity toward a particular cryptocurrency. A high sentiment score suggests optimism about the cryptocurrency, potentially indicating an expected price increase, while a low score signifies pessimism, possibly signaling a price decrease [32].

6.5 CNN-LSTM Privileges and Drawbacks

CNN-LSTM has some merits and demerits that affect its performance and applicability for sentiment analysis of cryptocurrency. As for the first, CNN-LSTM can capture both local and global features along with patterns from the data, while other models may only capture one or the other. This means that CNN-LSTM can handle more diverse and heterogeneous data, and handle more fine-grained and nuanced sentiments [33]. It can handle variable-length inputs and outputs, which is useful for dealing with data of different lengths and formats. This means that CNN-LSTM can process texts of different word counts, languages, or styles, while other models may require fixed-length inputs or outputs [34]. It can be trained end-to-end, which means that it can learn the optimal parameters for both the convolutional and recurrent layers without requiring any manual feature engineering or preprocessing. This approach can reduce the complexity and cost of the model development and deployment.

Regarding the latter, CNN-LSTM requires considerable training data and computational resources to learn effectively, as it has many parameters and operations to

optimize [35]. This means that CNN-LSTM can be slow and expensive to train and test and may need many trails and errors or grid search to find the best settings [23]. It may not be able to capture the long-range dependencies and context information in the data, as it relies on fixed-size filters or kernels for the convolutional layers. This means that CNN-LSTM may miss some important clues or signals that affect the sentiment of the text, such as the word order, the sentence structure, or the discourse relations [36].

6.6 Pysentimento

Pysentimento, a Python library designed for accessibility, provides a user-friendly avenue for conducting sentiment analysis and various social (NLP) tasks. The library achieves this by harnessing the capabilities of models such as BERT and others. Within Pysentimento, users can effortlessly load pretrained sentiment analysis models, including BERT, DistilBERT, RoBERTa, XLM-RoBERTa, or custom models tailored to specific domains or languages. Pysentimento's versatility extends to the application of these models on text inputs or data frames, generating sentiment scores that serve as quantitative indicators of the textual content's positivity or negativity. Pérez et al. [37] developed Pysentimento, making it conveniently accessible on both GitHub and PyPI.

In the field of sentiment analysis, Pysentimento opts for BERT as its principal model for conducting evaluations in both English and Spanish. The toolkit extends its capabilities by incorporating models that have undergone fine-tuning to address the specific requirements of distinct datasets or languages. A notable example is Robertuito, a specialized variant within the Pysentimento toolkit crafted explicitly for the analysis of sentiment in Spanish tweets sourced from social media. This adaptation is rooted in RoBERTa, an improved version of BERT distinguished by augmented data and extended training iterations. Pérez et al. [38] introduced Robertuito to enhance sentiment analysis within the Pysentimento framework. This comprehensive toolkit is designed to provide users with the capability to perform sentiment analysis on their proprietary data or make use of alternative models available in Pysentimento or the Hugging Face Transformers library. In doing so, a versatile and adaptable solution is established [38, 39]. Pysentimento has been evaluated on several datasets and benchmarks for different languages and tasks. According to its authors, it achieves state-of-the-art results for sentiment analysis in Spanish and English and competitive results for other languages and tasks. For example, for sentiment analysis in Spanish, Pysentimento achieves an accuracy of 91.3% on the TASS 2017 dataset, which is a collection of tweets annotated with positive, negative, or neutral labels. For sentiment analysis in English, Pysentimento achieves an accuracy of 90.9% on the SemEval 2017 Task 4A dataset, which is a similar collection of tweets [40].

6.7 Pysentimento Strengths and Weaknesses

In terms of merits, the utilization of BERT models within Pysentimento enables high accuracy and robustness for sentiment analysis tasks, surpassing previous models on various natural language processing (NLP) benchmarks and tasks [39]. Furthermore,

Pysentimento offers models that are fine-tuned on specific datasets or languages, thereby enhancing performance and adaptability [38]. This approach is particularly advantageous as sentiment analysis can aid in predicting cryptocurrency trends by capturing public sentiment and opinions, shedding light on the emotions and attitudes expressed in cryptocurrency-related text data [41, 42]. Additionally, Pysentimento leverages established and well-maintained NLP libraries and frameworks, such as the Hugging Face Transformers library, which supports numerous models and frameworks. Moreover, it utilizes popular libraries like pandas, numpy, and scikit-learn, ensuring reliability and widespread usage [43]. Pysentimento does have some demerits. Firstly, users are required to install Pysentimento and its dependencies, which may necessitate considerable time and storage space, depending on the user's system and internet connection. Furthermore, compatibility issues or errors may be encountered during the installation process [38, 43].

7 Analysis and Discussion

In this section, the primary objective is to conduct a comparative analysis of the performance of (SVM), (CNN-LSTM), and the Pysentimento framework utilizing BERT for sentiment analysis within the domain of cryptocurrencies. The aim is to determine the most accurate model for sentiment analysis. Subsequently, Google Trends data, volume, and the results of sentiment analysis (positive, negative, and neutral) are amalgamated into a table and normalized to obtain an overall sentiment sum. Following this, Pearson correlation coefficients are computed, and a time series prediction utilizing the SARIMA model is executed to identify which cryptocurrency is the safest for investment during periods of geopolitical tension.

7.1 Data Preprocessing and Training of CNN-LSTM

The training and testing dataset is comprised of 50,859 tweets on Bitcoin that have been categorized as ["positive"], ["negative"], and ["neutral"]. This dataset has been obtained from the Kaggle website, which is a renowned platform and online community for data scientists and machine learning practitioners [16]. Eighty percent of the data are devoted to training and the remaining portion is dedicated for testing where the training set consists of (40687) tweets, while the testing set comprises (10172) tweets. A larger training set allows the model to learn from a substantial amount of data, capturing patterns and relationships in the tweets that are essential for making accurate predictions. The bar plot indicates that 22,937 tweets are labeled as positive, 21,939 as neutral, and 5983 as negative. This exploratory analysis offers an initial understanding of sentiment distribution prior to model development (Fig. 1) conveys such a distribution.

Preprocessing initiates with the column "tweet" turning into "clean_tweet" after preprocessing and the column "label" is the one for sentiments. The word count is a common feature employed in text analysis tasks to assess the complexity and content of text data whereas text length column quantifies the length of each tweet in
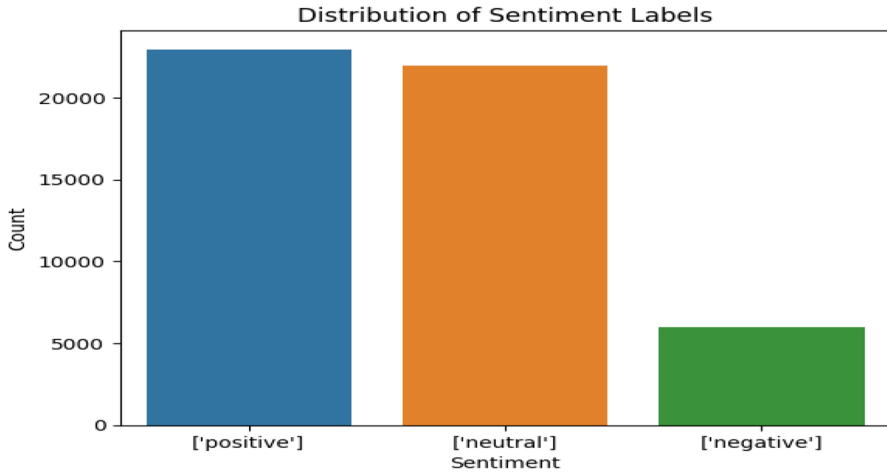
Fig. 1 Distribution of sentiment labels

terms of character count. This metric can be valuable for examining the distribution of tweet lengths within the dataset. The mean and standard deviation enable the researcher to determine the correct average of tweet length employing Freedman-Diaconis rule, which was 30.

The training and testing shapes become (40687, 30) (10172, 30). Figure 2 shows the data after preprocessing with both metrics, while Fig. 3 uncovers the word cloud, a powerful visual representation of frequently occurring words in text data, generated at this stage since it offers a visual summary, highlighting the most prevalent words within the dataset. It provides insights into the dataset's underlying themes and characteristics. The epochs' number is 5 and the batch size is 128.

The model in Fig. 4 undergoes five processing stages. To begin with, an embedding layer maps each of 30 input tokens to a 200-dimensional vector, forming a three dimensional tensor of shape (30, 200). It involves two crucial parameters: maximum features and embedding dimensions. These parameters grant the model flexibility in encoding word information, making it adaptable to the dataset's specific characteristics. The model in Fig. 4 undergoes five processing stages. To begin with, an embedding layer maps each of 30 input tokens to a 200-dimensional vector, forming a three dimensional tensor of shape (30, 200). It involves two crucial parameters:

Fig. 2 Pre-processing with text length and word count

Fig. 3 Word cloud

maximum features and embedding dimensions. These parameters grant the model flexibility in encoding word information, making it adaptable to the dataset's specific characteristics. Subsequently, the architecture incorporates two convolutional 1D layers, which introduce a spatial understanding of the text data.

These layers employ a set of 1D convolutional filters, alike to sliding windows, to detect local patterns and feature representations within the tweet sequences. Throughout utilizing filters with varying receptive field sizes, the model becomes proficient at recognizing both fine-grained details and broader textual features. The

Fig. 4 CNN-LSTM sequential model components

Rectified Linear Unit (ReLU) activation function adds a critical element of non-linearity, enabling the model to capture complex relationships between words. Following the convolutional layers, two MaxPooling 1D layers act as a dimensionality reduction mechanism. These layers systematically downsample the output from the convolutional layers preserving the most salient and informative features while mitigating computational complexity. This process allows the model to concentrate on the most relevant elements of the data, enhancing its efficiency and focus.

The neural network architecture further evolves with the inclusion of an LSTM layer as it excels at capturing sequential dependencies and long-range contextual information within text data. This layer is vital in modeling the temporal dynamics of tweet sequences, understanding how words relate to each other over time, and capturing intricate patterns that might span across the entire sequence. Finally, the dense layer maps the LSTM output to a three-dimensional vector using linear activation, forming a tensor of shape (3) where the softmax function transforms the model's internal representations into probability distributions across the three sentiment classes: negative, neutral, and positive [44].

7.2 Classification Report of CNN-LSTM

Based on Sharma and Sharma [45], classification reports are essential instruments for assessing the efficacy of models intended to infer attitudes (positive, negative, or neutral) from textual data. These reports provide crucial metrics, each of which provides unique information on how well the model works. The first indicator, precision, measures how well positive predictions are made in comparison to false positives, or how frequently the model accurately predicts positive feelings. The other one, recall, gauges how well the model can recognize real-world positive examples, demonstrating its accuracy in expressing good emotions. Eventually, the third metric, the F1-score, finds a middle ground between recall and accuracy. This middle ground is especially helpful when the distribution of emotion classes is not uniform. Figure 5 uncovers CNN-LSTM classification report.

The results demonstrate the ability of the proposed method to classify tweets into three sentiment categories: positive (pos), negative (neg), and neutral (neu). As for negative (neg), the precision for the negative sentiment class is approximately 0.92. This indicates that when the model predicts a tweet as negative, it is correct proximately 92% of the time. The recall, at around 0.95, demonstrates that the model captures about 95% of the actual negative tweets. The F1-score, which harmonizes these metrics, is approximately 0.93, suggesting a strong balance between precision and recall for the negative sentiment class. This means that the model excels in identifying and correctly classifying tweets expressing negative sentiments.

Regarding the neutral sentiment class, the precision is roughly 0.98, indicating that the model's predictions of neutrality are highly accurate. The recall, at approximately 0.97, signifies that the model correctly identifies about 97% of the actual neutral tweets. The F1-score of about 0.98 reaffirms the model's exceptional performance in classifying neutral sentiment, with a balanced combination of precision

	precision	recall	f1-score	support
neg	0.92	0.95	0.93	1197
neu	0.98	0.97	0.98	4388
pos	0.98	0.98	0.98	4587
accuracy			0.97	10172
macro avg	0.96	0.97	0.96	10172
weighted avg	0.97	0.97	0.97	10172

Fig. 5 CNN-LSTM classification report

and recall. This signifies the model's proficiency in distinguishing neutral tweets from others.

The positive sentiment class exhibits similar excellence, with a precision of approximately 0.98, indicating highly accurate positive predictions. The recall, around 0.98, indicates that the model captures about 98% of the actual positive tweets. The F1-score, approximately 0.98, reflects the strong balance between precision and recall for the positive sentiment class. This underscores the model's ability to effectively identify and classify positive sentiment in tweets. The overall model performance is impressive, with an accuracy of approximately 97%. This accuracy demonstrates the model's proficiency in classifying tweets across all sentiment categories.

The macro-average F1-score, at about 0.96, signifies that the model maintains a robust balance between precision and recall for all sentiment classes, considering their individual support levels. Additionally, the weighted average F1-score, also around 0.97, indicates the model's consistency in performance across different sentiment classes, considering their varying proportions in the dataset.

To conclude, the model demonstrates high proficiency in classifying tweets into positive, negative, and neutral sentiment categories, with an overall accuracy of 97%. It upholds robust balanced performance across sentiment classes, as indicated by macro and weighted average F1-scores of 0.96 and 0.97, respectively. The results emphasize the model's effectiveness in sentiment analysis of social media texts.

7.3 Data Preprocessing and Training of SVM

All the steps of this phase are similar to those in the previous model apart from minor differences. First of all, tokenization breaks the text into individual words or tokens, enhancing the model's ability to comprehend and analyze the content. Secondly, stop-words removal is essential for eliminating common but uninformative words. Thirdly, lemmatization reduces words to their base or root forms, ensuring

consistency in the data. Variations like "running" and "ran" are both transformed to "run," streamlining feature extraction and pattern recognition. Finally, Part-of-Speech Tagging (POS) enriches the data by assigning grammatical labels to each word, allowing for the capture of specific linguistic patterns. This step can aid in discerning verbs, adjectives, or other parts of speech relevant to sentiment analysis.

For this model, the Hugging Face 200-dimensional GloVe feature is imported, and it has been widely implemented in NLP tasks. The 200 dimensions of the word vectors refer to the fact that each word is represented by a vector of 200 numerical values to provide a rich and informative representation of words. The feature works by converting words into a continuous vector space, where the similarity between words can be measured using cosine similarity [20].

On the positive side, this feature is built upon a vast and diverse corpus of tweets, enabling it to effectively capture informal, colloquial language, slang, and even emoticons commonly found in social media conversations and tweets. It excels at capturing both global and local information from the corpus, including word frequency, word order, and word context. Additionally, the model can unveil intriguing linear relationships between words, such as analogies, antonyms, and synonyms. However, there are certain restrictions associated with the Hugging Face GloVe embeddings. Firstly, its performance is contingent upon the vocabulary size and coverage of the underlying corpus, which may not encompass rare or domain-specific (cryptocurrency) words, potentially limiting its applicability in specialized contexts. Secondly, the feature may struggle to capture intricate and nonlinear relationships between words, including polysemy, homonymy, irony, and sarcasm. Lastly, it may encounter difficulties in handling out-of-vocabulary words or common typographical errors commonly encountered in the informal language of tweets. Understanding these merits and demerits is crucial when considering the application of the Hugging Face GloVe embeddings 200d in various natural language processing tasks [20].

Conversely, TF-IDF is a fundamental technique in (NLP). Being part and parcel of this model, TF-IDF addresses a critical challenge in NLP which is how to represent the inherent information and nuances of text data in a way that algorithms can effectively process [46]. TF-IDF vectorization offers several advantages when compared to alternative text representation methods like bag-of-words or word embeddings.

First of all, it addresses the issue of high dimensionality by considering only words that appear in at least one document within the dataset. TF-IDF aids in solving the obstacle of having too many different words to deal with by only looking at words that appear in at least one document in the dataset. This reduction in dimensionality makes it computationally efficient and manageable. Secondly, TF-IDF captures both local and global information about words. It accounts for word frequency within a document and across all documents in the dataset, providing a holistic view of word importance. Thirdly, TF-IDF assigns higher weights to words that carry more informative or distinctive characteristics for a document or topic, while assigning lower weights to common or generic words. This feature is particularly valuable for highlighting the significance of words in context. Finally, TF-IDF is known for its simplicity in implementation and interpretation, as it does not necessitate complex mathematical operations or external resources [47].

Based on Xue [48], there exist certain limitations and challenges associated with TF-IDF vectorization. To begin with, TF-IDF assumes that words are independent of each other, disregarding their order and context within a document. This can lead to the loss of valuable sequential information. Moreover, TF-IDF does not capture semantic or syntactic relationships between words, such as synonyms, antonyms, or grammatical structures, which limit its ability to understand the deeper meaning of language. Additionally, TF-IDF may assign low weights to words that are relevant but infrequent across all documents, such as proper nouns or domain-specific terms. Eventually, TF-IDF can be sensitive to outliers or noisy data, including spelling errors or typos. To mitigate some of these limitations, the researcher employs complementary techniques like stemming, lemmatization, NLTK stop-word removal, and adding (pos) to tokens with TF-IDF.

These enhancements improve its overall performance and accuracy, making it a versatile choice for text analysis tasks and that has already been put into practice at the first step of pre-processing. Figure 6 uncovers the tweets after preprocessing steps executed. The input data shapes are examined, with train GloVe at 40,687 samples by 200 features and test GloVe at 10,172 by 200, encoding tweets in a semantic space. Meanwhile, train TF-IDF and test TF-IDF have the same sample counts and 9991 features, representing tweets high-dimensionally. This conveys the input data shapes of GloVe and TF-IDF. GridSearchCV is a tool that performs an exhaustive search over specified parameter values for an estimator using cross-validation. Cross-validation is a technique that splits the data into k-folds, utilizes one-fold as the test set and the rest as the training set, and repeats this process k times, averaging the results. This approach can provide a more reliable estimate of the performance of the estimator on unseen data.

Three-fold cross-validation is employed to evaluate each combination of parameter values. This conveys that data will be split into three parts, and each part is used as a test set once, while the other two parts are employed as a training set. The average score across the three folds is used as the performance metric for each combination. The parameters that are tuned are as follows: C, which is the regularization parameter for SVM that controls the trade-off between margin maximization and error minimization; and kernel, which is the kernel function for SVM that determines the type of transformation applied to the data.

```
Processed Tweets (Training Set):
 28589    much money bitcoin miner
 15363    laptop wifi network adapter fritz
 10609    new airdrop free y
 29157    rt blockchain amp bitcoin conference tallinn interesting case great networking
 9683     new airdrop huobi exchange airdrop tron user base e

Processed Tweets (Test Set):
 48220    rt use code hfbday purchase get discount contract offer ghs
 14090    name nubits usnbt hour change price rank total supply volume
 46986    rt checkout
 48596    bitcoin future organization token sale protect worldwide economy blockchain technology
 9580     bitcoin money
```

Fig. 6 Tweets after SVM preprocessing

```
Best hyperparameters for SVM with GloVe features: {'select__k': 'all', 'svm__C': 10, 'svm__kernel': 'rbf'}
Best score for SVM with GloVe features:  0.9194091495998823
Best hyperparameters for SVM with TF-IDF features: {'select__k': 'all','svm__C': 10, 'svm__kernel': 'rbf'}
Best score for SVM with TF-IDF features:  0.882299687677226
```

Fig. 7 Best hyperparameters for SVM with GloVe and TF-IDF

After that, the researcher designates "all" as the value for *k*, indicating the selection of all features by SelectKBest. The "SVM C" parameter, which represents several levels of regularization strength, is set to a range of values, including 0.1, 1, and 10, in the SVM classifier. Furthermore, two kernel options "linear" and "RBF" (Radial Basis Function) are defined for the "SVM kernel" parameter, which affects how the SVM maps input data. The first one finds a linear hyperplane that separates the data points based on their features (positive, negative neutral). The second maps the data points into a higher-dimensional space where a linear hyperplane can separate them better than in the original space. The grid search process evaluates the model's performance in a systematic manner across different sets of these hyperparameter values in order to determine which set best improves model performance. The results are illustrated in Fig. 7.

The figure conveys that utilizing "all" features, setting *C* to 10, and employing the RBF kernel provide the best performance for SVM with GloVe and TF-IDF features on the data. $C = 10$ indicates a reasonably high value of this hyperparameter for the SVM classifier, balancing margin maximization and error minimization effectively. The RBF kernel is frequently implemented for capturing non-linear relationships in the data, indicative of the data's complexity and non-linearity. The model achieves accuracy scores of approximately 91.94% with GloVe and 88.22% with TF-IDF, reflecting its capability to classify tweets into their respective sentiment categories accurately.

7.4 Classification Reports of SVM

7.4.1 Classification Report of SVM-GloVe

As revealed in Fig. 8, with regard to sentiment analysis, the classification report thoroughly provides a thorough summary of the performance data obtained from an established support vector machine (SVM) model that makes use of GloVe features. This assessment methodology includes support, recall, *F*1-score, and accuracy measures that are specific to the three sentiment categories (positive, neutral, and negative). Central measures of the model's prediction accuracy for every sentiment class are precision values, which are shown as noteworthy percentages. For the negative sentiment category, the accuracy is 91.83%, for the neutral sentiment category, it is 94.46%, and for the positive sentiment category as 91.25%.

The recall percentages surface as 95.17% for negative sentiments, 93.54% for neutral sentiments, and 82.39% for positive sentiments. The accuracy, a pivotal metric gauging correctness in classifying cases crosswise all sentiment subclasses, is 92.96%. This overarching accuracy metric provides a holistic viewpoint on the model's effectiveness in sentiment examination. Macro and weighted averages, integral portions

	precision	recall	f1-score	support
negative	0.9183	0.9517	0.9347	4345
neutral	0.9446	0.9354	0.9400	4663
positive	0.9125	0.8239	0.8659	1164
accuracy			0.9296	10172
macro avg	0.9251	0.9037	0.9135	10172
weighted avg	0.9297	0.9296	0.9292	10172

Fig. 8 Classification report of (SVM-GloVe)

of the report, offer nuanced appraisals accounting for stabilizing and circulation of sentiment categories. The macro-averaged precision, recall, and $F1$-score communicate as 92.51%, 90.37%, and 91.35%, respectively. Conversely, the weighted-averaged precision, recall, and $F1$-score arise at 92.97%, 92.96%, and 92.92%, correspondingly. These averages offer a more comprehensive assessment regarding the impact of class imbalances on the level of performance held by the model.

To summarize, the classification report delivers comprehensive and illuminating information about the SVM model's performance, providing subtle insights into its accuracy, recall, and $F1$-score for various sentiment categories. While macro and weighted averages offer a comprehensive picture of the model's performance in sentiment analysis tasks, the support metric and overall accuracy add more context. This thorough examination is essential for identifying the model's advantages and possible areas for improvement, enabling well-informed sentiment analysis decision-making.

7.4.2 Classification Report of SVM-TF-IDF

The deployed (SVM) model's sentiment analysis performance is painstakingly evaluated in the classification report in Fig. 9. The assessment is carried out via TF-IDF characteristics, a commonly employed method for determining a word's relevance inside a collection of documents. The report provides a detailed overview of the discriminative skills of the model by covering precision, recall, $F1$-score, and support metrics for each sentiment category (positive, neutral, and negative).

The precision values signify the degree of accuracy with which the model predicts each sentiment class. In the present case, the accuracy of the negative sentiment class is 86.54%, indicating that the model is capable of correctly classifying negative attitudes. The model's accuracy in predicting neutral attitudes is demonstrated by the neutral sentiment class's precision of 94.47%, whereas the positive sentiment class's precision is 90.70%.

Metrics for recall offer valuable information on how well the model captures examples from each sentiment category. Recall for the negative class is 95.63%, highlighting the model's capacity to accurately identify a significant percentage of real

```
Classification Report for SVM with TF-IDF features:
              precision    recall  f1-score   support

    negative     0.8654    0.9563    0.9085      4345
     neutral     0.9447    0.8849    0.9138      4663
    positive     0.9070    0.7814    0.8395      1164

    accuracy                         0.9035     10172
   macro avg     0.9057    0.8742    0.8873     10172
weighted avg     0.9065    0.9035    0.9030     10172
```

Fig. 9 Classification report of (SVM-TF-IDF)

negative cases. With an 88.49% recall rate, the neutral class demonstrates a strong sensitivity to real-world neutral mood occurrences. Nevertheless, the recall of the positive class is 78.14%, indicating a somewhat lower capture of real positive cases.

F1-scores, which exhibit a harmonic mean of recall and precision, shed more light on the model's well-balanced performance. With an F1-score of 90.85%, the negative emotion class exhibits a well-balanced trade-off between recall and precision. Likewise, the class representing neutral sentiments has an F1-score of 91.38%, signifying a well-balanced approach to forecasting neutral sentiments. On the other hand, the F1-score of 83.95% for the positive sentiment class indicates a trade-off between recall and precision for positive sentiments.

The SVM model with TF-IDF features has an overall accuracy of 90.35%, which provides a global indicator of how accurate it is at categorizing instances in all sentiment categories. Weighted averages and macro statistics offer a thorough analysis that takes the impact of class disparities into account. The macro-averages for recall, F1-score, and precision are 90.57%, 87.42%, and 88.73%, in that order. On the contrary, the F1-score, weighted-averaged precision, and recall are 90.35%, 90.65%, and 90.70 percent, correspondingly.

To conclude, this thorough assessment provides a detailed overview of the SVM model's advantages and disadvantages in sentiment analysis tasks, offering insightful information that can be used to develop the model and make well-informed decisions.

The comparative study of SVM model that employs both of TF-IDF and GloVe characteristics has several implications. The TF-IDF feature is regularly outperformed by the SVM model combined with GloVe embeddings in terms of accuracy, precision, and F1-score for a range of attitudes. This consistency implies that GloVe embeddings play a major role in building a more sophisticated and reliable sentiment analysis model due to their capacity to capture complex semantic information.

Both features have a high level of competence when it comes to neutral emotions, demonstrating their ability to navigate through a wide range of expressions in this area. The F1-scores accurately reflect the subtle nature of neutral sentiment

expressions, which forces models to strike a careful balance between recall and precision. With the highest recall in this category, the SVM model which utilizes TF-IDF features demonstrates a unique ability to efficiently capture occurrences of negative sentiment efficiently. This discovery suggests that TF-IDF could be vastly advantageous in recognizing manifestations of negative emotion, providing insightful information, especially in applications where detecting negative sentiments is critical. These findings highlight the significance of contemplating the intricacies of sentiment expressions and carefully evaluating the trade-offs between precision and recall when selecting a sentiment analysis model.

7.5 Pysentimento

This particular model does necessitate neither a preprocessing nor a training phase, as it is a pre-trained model equipped with a tokenizer. The process commences with the installation of the Pysentimento library (version 0.7.2) and its prerequisites. It procures the following dependencies: "accelerate" library (version 0.22.0), "datasets" (version 2.14.5), and "emoji" (version 1.7.0). The model architecture, "robertuito," suggests that this model is based on the RoBERTa architecture. RoBERTa is a variant of the BERT (Bidirectional Encoder Representations from Transformers) model and is known for its effectiveness in a wide range of NLP tasks. The model is fine-tuned specifically for sentiment analysis. Fine-tuning involves training a pre-existing model on a task-specific dataset. In this case, the model has been trained on a dataset of text samples with associated sentiment labels (positive, negative, and neutral). This fine-tuning process supports the model to learn the patterns and features relevant to sentiment analysis. Pre-trained models for sentiment analysis have become increasingly popular due to their effectiveness in capturing nuances in sentiment across various domains [40].

The "accelerate" library, designed for Python, focuses on optimizing and expediting computations, particularly within the realms of numerical and scientific computing. Moreover, the "emoji" library, also a Python library, augments the project with capabilities centered on "emoji" processing and management. This library enables tasks such as "emoji" detection, extraction, and manipulation within textual data. Its functionality extends to "emoji" identification, conversion between "emoji" and Unicode representations, and "emoji" visualization. Incorporating the "emoji" library is a sensible decision aimed at enhancing the project's capacity to adeptly manage emojis within the sphere of NLP endeavors, encompassing applications such as sentiment analysis, text classification, and text generation [49]. The dataset dedicated to training purposes consists of 50,859 tweets related to Bitcoin. It is partitioned into an 80% training set and a 20% testing set previously. The testing dataset from the previous models, encompassing 10,174 tweets, is stored in a CSV file. This dataset will be employed to assess the accuracy of the Pysentimento model, ensuring uniformity across all three models under evaluation. Figures 10 and 11 demonstrate the stages of Pysentimento model and the results of Pysentimento application on a portion of the testing dataset correspondingly. In Fig. 11, the "label" column represents the classification of the training dataset whereas the "sentiment" is the one for Pysentimento.
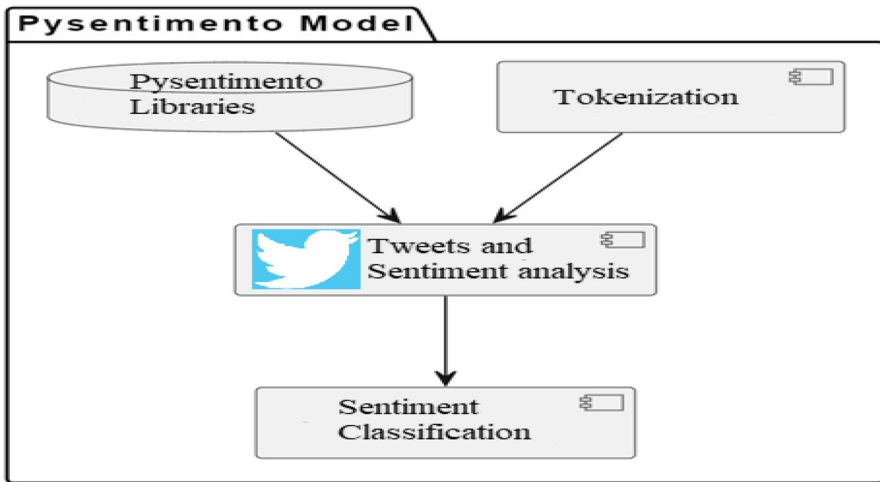
Fig. 10 Pysentimento model stages

7.5.1 Classification Report of Pysentimento

The presented classification report concerns the "Pysentimento" model's performance following its application to test data. This report evaluates the model's capabilities in classifying text-based sentiments into three categories: "positive," "negative," and "neutral." It is crucial to analyze each aspect of the report to gain a comprehensive understanding of the model's performance as conveyed in Fig. 12.

When it comes to assessing the "positive" emotion category, the model performs admirably. For this category, the precision is 0.89, meaning that 89% of the feelings that the model predicted to be "positive" are indeed true. The recall for the "positive" category is equally impressive, with a score of 0.88, indicating that the model effectively captures 88% of actual positive sentiments present in the dataset. An $F1$-score of 0.88 in this situation indicates that the model is remarkably well-balanced, with a strong equilibrium between recall and precision. This equilibrium suggests

Unna	tweet	label	text_clean	sentiment
0	"French Regulatory Agency Blacklists 15 Cryptocurrency And Crypto-Asset W(	negative	"French Regulatory Agency Blacklists 15 Cryptocurrency And Crypto-Asset Web	NEG
1	: #Bethereum makes any sports event more fun! âš¼	positive	: hashtag bethereum makes any sports event more fun! emoji balÃ³n de	POS
2	(AUD/USD Technical Analysis: Two-Year Uptrend Threatened) has been publi	neutral	(AUD/USD Technical Analysis: Two-Year Uptrend Threatened) has been publish	NEU
3	â–¶ High paying #Bitcoin #faucet ; 50% lifetime referral commissionÂ	neutral	emoji reproducir emoji High paying hashtag bitcoin hashtag faucet ; 50%	NEU
4	The creator of Bitcoin chose a male name of Japanese origin most proâ€¦	neutral	The creator of Bitcoin chose a male name of Japanese origin most pro	NEU
5	'Child porn links could make Bitcoin blockchain illegal' ... #Bitcoin	negative	'Child porn links could make Bitcoin blockchain illegal' ... hashtag bitcoin	NEG
6	Itâ€™s time #bitcoin #2A	neutral	Its time hashtag bitcoin hashtag 2 a	NEU
7	Bitcoin Gold (BTG) do become popularity ; seemly new expensive, just You	positive	Bitcoin Gold (BTG) do become popularity ; seemly new expensive, just You sti	POS
8	: Bitcoin Could Become Illegal Almost Everywhere, After Shocking Discovery	negative	: Bitcoin Could Become Illegal Almost Everywhere, After Shocking Discovery in	NEG

Fig. 11 Pysentimento application results

Fig. 12 Pysentimento classification report

	precision	recall	f1-score	support
positive	0.89	0.88	0.88	3031
negative	0.83	0.69	0.76	1479
neutral	0.92	0.97	0.94	5663
accuracy			0.90	10173
macro avg	0.88	0.85	0.86	10173
weighted avg	0.90	0.90	0.90	10173

that the model performs well in predicting "positive" feelings and in capturing the majority of genuine positive sentiments seen in the data.

The model's precision score in the "negative" sentiment category is 0.83, which means that 83% of the feelings it identified as "negative" are actually "negative." However, the recall for the "negative" category is 0.69, demonstrating that the model only accurately classifies 69% of the real negative attitudes in the dataset. For the "negative" category, the F1-score, a measure of precision and recall equilibrated, is recorded as 0.76. The model appears to be operating in this category quite harmoniously, based on its F1-score of 0.76. With a precision of 0.92 in the "neutral" sentiment category, the "Pysentimento" model does exceptionally well, correctly predicting 92% of the "neutral" feelings among its predictions.

Moreover, the recall for the "neutral" category is an impressive 0.97, highlighting the model's capacity to correctly classify 97% of actual neutral sentiments in the dataset. The F1-score for "neutral" is 0.94, denoting a balance between precision and recall. This score highlights the model's ability to discern neutral sentiments. The "Pysentimento" model has an overall accuracy of 90%, demonstrating its competence in sentiment prediction across all categories.

In conclusion, the "Pysentimento" model performs admirably when tested utilizing test data. It maintains a balanced F1-score, demonstrating its efficacy in capturing various sentiment categories, and demonstrates notable strengths in precision, especially in the "neutral" category. The model's overall 90% accuracy rating specifies how consistently it can predict "positive," "negative," and "neutral" sentiments.

8 The Second Evaluation

The researcher's choice of utilizing a second evaluation stems from a compelling need for rigorous accuracy validation. This imperativeness arises due to substantial doubts surrounding the observed accuracy levels of the three models. The core objective is to ascertain the authenticity of the attained accuracy, safeguarding against potential pitfalls like overfitting or underfitting, which may compromise the credibility of the results. Given the paramount prominence of this objective, the previously constructed models are saved and subsequently recalled to undergo a test against a dataset comprising 100 tweets of the randomly collected data regarding three

cryptocurrencies: BTC, ETH, and BNB, which represents unseen data manually categorized by the researcher into positive, negative, and neutral sentiments. Figure 13 unveils the results of the second evaluation.

The performance metrics, including precision, recall, $F1$ rating, and accuracy, for the four distinct NLP models become conveyed. They are "CNN-LSTM," "SVM (GloVe)," "SVM (TF-IDF)," and "Pysentimento." Such metrics function as fundamental indicators of the models' proficiency in accurately identifying positive, negative, and neutral sentiments within a specified dataset. The precision values reveal observable modifications across every model under examination. These indices serve as indications of the percentage of correctly expected positive cases among all cases expected as positive. "CNN-LSTM" uncovers a noteworthy precision of 90%, emphasizing its dexterity in productively categorizing positive sentiments with precision. In contrast, "SVM (GloVe)" and "SVM (TF-IDF)" demonstrate comparatively reduced precision values of 80% and 78%, respectively, indicating a relatively smaller accuracy in correctly distinguishing positive cases.

Contrastingly, "Pysentimento" separates itself with a remarkable precision of 97%, symptomatic of its heightened ability to outwit incorrect positives and discern positive sentiments with extraordinary accuracy. This discrepancy in precision values underscores the nuanced execution divergences among the models, thereby offering crucial insights into their efficacy in positive sentiment characterization.

The recall values, which signify the proportion of actual positive instances properly recognized by the models, unveil intricate patterns that shed light on the models' efficacy in capturing positive sentiments. "CNN-LSTM" achieves a commendable recall of 88%, affirming its ability to capture a considerable part of genuine positive sentiments within the dataset. In contrast, both "SVM (GloVe)" and "SVM
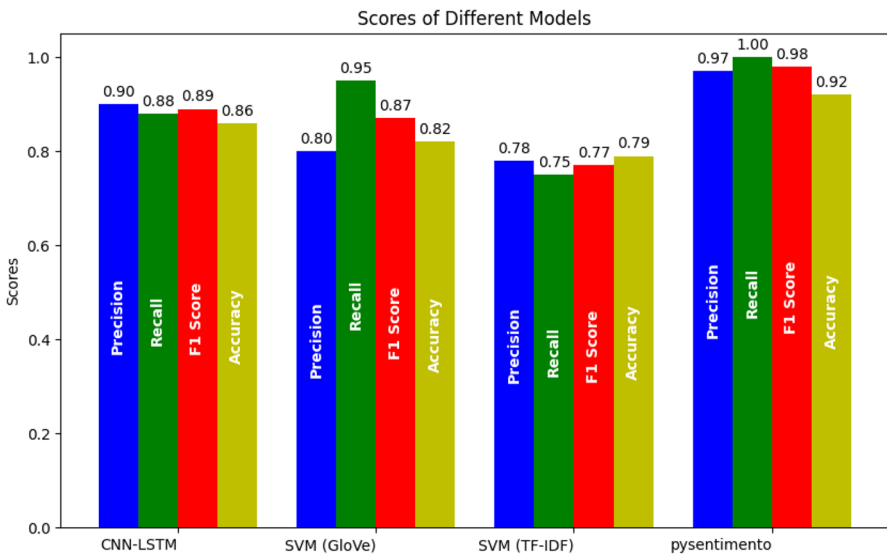

Fig. 13 Second evaluation scores

(TF-IDF)" present dissimilar recall values of 95% and 75%, respectively. This delineates varied abilities in correctly identifying positive instances, with "SVM (GloVe)" demonstrating a notably higher recall than "SVM (TF-IDF)." Noteworthy is the exceptional execution of "Pysentimento," boasting a perfect recall of 100%, indicating its unparalleled capability to comprehensively identify all positive sentiments present in the dataset. This divergence in recall values among the models accentuates nuanced distinctions in their ability to effectively identify and capture actual positive sentiments, providing valuable insights into their respective strengths in positive sentiment recognition.

The examination of the $F1$ rating provides a thorough assessment of the model accomplishment. A harmonic average derived from recall and precision is the $F1$ rating. "CNN-LSTM" becomes distinguished by an $F1$ rating of 89%, indicating a tuneful balance between precision and recall, thereby proposing a well-rounded effectiveness. In contrast, "SVM (GloVe)" and "SVM (TF-IDF)" present $F1$ scores of 87% and 77%, respectively, accentuating nuanced variations in their precision-recall equilibrium. The heightened $F1$ rating of 98% accomplished by "Pysentimento" attests to its exceptional proficiency in accomplishing a brilliant amalgamation of precision and recall, underscoring its intensified effectiveness in sentiment analysis tasks.

9 Comparative Analysis

Below is a comprehensive analysis of sentiment classification models, namely, "CNN-LSTM," "SVM (GloVe)," "SVM (TF-IDF)," and "Pysentimento." The focus is on $F1$ score values, accuracy, precision, and recall in both the first and second assessments. Subtle variations in the models' output that provide insight into how well they perform in sentiment analysis applications will be revealed through careful inspection of these metrics.

The precision values for sentiment classification models in the first and second evaluations exhibit slight variations. "CNN-LSTM" obtains a precision of 0.90 in the first evaluation, but in the second evaluation, it slightly drops to 0.89. Conversely, "SVM (GloVe)" exhibits a clear increase from 0.80 to 0.87, indicating an increase in its capacity to accurately detect pleasant emotions. The precision of "SVM (TF-IDF)" decreases from 0.78 to 0.77, suggesting a slight deterioration in its precise performance. In contrast, "Pysentimento" displays consistency in both evaluations, with a precision value of 0.97, highlighting its ability to prevent false positives.

Similarly, recall values for each of the sentiment categorization models disclose significant alterations between the two assessments. The recall of the "CNN-LSTM" model falls slightly from 0.88 to 0.87, suggesting a marginal decline in the model's ability to identify true positive cases. In contrast, "SVM (GloVe)" reveals a boost in its recall from 0.95 to 0.98, indicating an elevated degree of accuracy in identifying positive sentiments. "SVM (TF-IDF)" has a more substantial reduction in

recall from 0.75 to 0.68, indicating a shift in its proficiency in recognizing positive instances. It is noteworthy that "Pysentimento" consistently scores a perfect recall of 1.00 in both evaluations, demonstrating its unwavering ability to fully detect all positive sentiments in the dataset.

The $F1$ score offers a comprehensive view of model performance as a well-balanced combination of recall and precision. With an $F1$ score of 0.89 in both assessments, "CNN-LSTM" performs consistently and represents an equitable trade-off between recall and precision. The $F1$ score of "SVM (GloVe)" slightly increases from 0.87 to 0.88, demonstrating a better equilibrium between recall and precision. However, "SVM (TF-IDF)" experiences a more significant decrease in $F1$ score, from 0.77 to 0.73, suggesting a change in the precision-recall balance. With a consistent $F1$ score of 0.98 in both tests, "Pysentimento" performs exceptionally well, demonstrating its ongoing ability to achieve a pleasing balance between recall and precision.

The sentiment classification models' accuracy values reflect significant variations between the first and second assessments. In both evaluations, "CNN-LSTM" consistently scores an accuracy of 0.86, demonstrating stability of its overall prediction ability. From 0.82 to 0.81, "SVM (GloVe)" shows a slight decline, indicating a minor change in accuracy performance. SVM (TF-IDF) experiences a more notable decline from 0.79 to 0.71, signifying a notable change in its overall accuracy. Conversely, "Pysentimento" has a remarkable consistency, exhibiting a constant accuracy of 0.92, highlighting its continuous ability to anticipate sentiments correctly in all classes.

Potential overfitting or underfitting factors when examining performance fluctuations in sentiment analysis models such as "CNN-LSTM," "SVM (GloVe)," "SVM (TF-IDF)," and "Pysentimento" must be taken into account. When a model becomes overfitted to the training data, it captures irrelevant details and exhibits poorer performance on new, unseen data. This phenomenon is known as overfitting. Conversely, underfitting occurs when a model accomplishes poorly than expected across measurements due to its oversimplification. Overfitting may be indicated by abrupt declines in accuracy, recall, $F1$ score, or precision, while underfitting, which exposes the model's simplicity in capturing sentiment classifications, may be designated by persistent underperformance across metrics.

In conclusion, subtle variances in the performance metrics of the sentiment analysis models "CNN-LSTM," "SVM (GloVe)," "SVM (TF-IDF)," and "Pysentimento" are exposed through comparing their analyses between the first and second evaluations. In both assessments, "CNN-LSTM" continuously maintains lofted values for accuracy, recall, $F1$ score, and precision, indicating a strong and reliable performance. Recall and $F1$ score improvements for "SVM (GloVe)" demonstrate the system's capacity to recognize positive sentiments. A number of measures indicate a possible shift in the precision-recall balance of "SVM (TF-IDF)." It is noteworthy that "Pysentimento" consistently exhibits exceptional precision, recall, and $F1$ score, confirming its effectiveness regarding sentiment analysis accurate prediction and it will be employed on the data concerning the three cryptocurrencies.

10 Pysentimento and Google Trends

This implementation of Pysentimento is preceded by a pre-processing step, which involves the removal of irrelevant tweets from the dataset pertaining to the three currencies. A pattern of words is designated and it serves the purpose of recognizing tweets that pertain to the context of cryptocurrencies, encompassing terms such as Bitcoin, Ethereum, and Binance, along with a multitude of associated hash tags and keywords. Tweets that deviate from this pattern, commonly referred to as "mismatches," are systematically identified and segregated, with a dedicated output file housing these mismatched tweets. As for Google Trends, Fig. 14 illustrates total searches for BTC, ETH, and BNB, respectively, in March, June, and December, 2022. The total searches on Google Trends for the three cryptocurrencies, Bitcoin, Ethereum, and Binance Coin in 2022, demonstrate varying levels of attention cross the given months. In March, Bitcoin reached 2075 searches, Ethereum garnered 2155 searches, and Binance Coin observed 2621 searches. In June, Bitcoin declined to 1506 searches, Ethereum reduced to 1371 searches, and Binance Coin held a substantial interest with 2150 searches. December notably saw a significant increase in searches for Bitcoin and Ethereum, reaching 2558 searches for each, indicating heightened fascination in these cryptocurrencies toward the end of the year. In contrast, Binance Coin observed a decrease in searches to 1899.

These search volume fluctuations imply dynamic shifts in public sentiment and interest regarding these cryptocurrencies over the course of the year 2022. The collective data reveal complex patterns in the online attention for these digital assets, potentially swayed by external elements such as market trends, regulatory developments, and overall market sentiment.

After the application of Pysentimento to the data, the sentiments sum total is collected for the three previously designated cryptocurrencies and months in 2022. The opening and closing of last days for each month in addition to the trading volume of the three cryptocurrencies are provided using Yahoo Finance.

The trading volume is the total quantity of a cryptocurrency that is traded across all exchanges over a specific timeframe. It exhibits the total quantity of cryptocurrency units that have been acquired and sold on the market during that
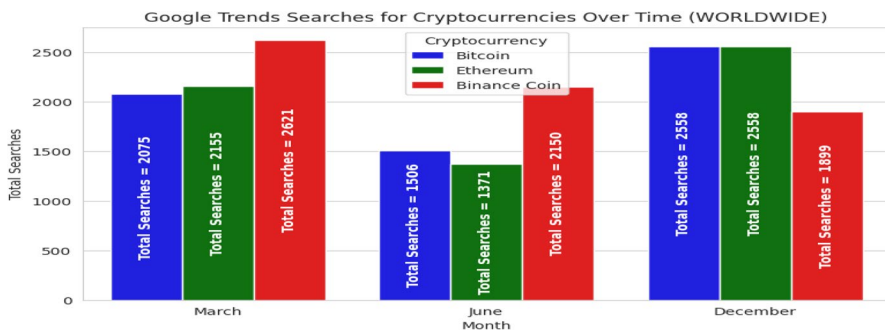

Fig. 14 Total searches of Google Trends

period. Trading volume is a crucial metric in the cryptocurrency market since it offers insights into the market activity and liquidity. Higher trading volumes commonly indicate increased market interest, liquidity, and the potential for more accurate price discovery. On the contrary, lower trading volumes could suggest reduced interest, potentially leading to higher price volatility and less reliable price information [50]. Table 1 displays how the collected data are organized.

The table provides a detailed overview of cryptocurrency data for March (MAR-22), June (JUN-22), and December (DEC-22) of the year 2022. The "Month" column categorizes data based on the recording month and year, allowing for a closer look at cryptocurrency performance during specific periods. Each entry in the "Coin" column acts as a unique identifier, indicating the cryptocurrency (e.g., BNB, BTC, and ETH) and simplifying the analysis process. Sentiment-related columns like "Positive %," "Negative %," "Neutral %," and "Overall Sentiment %" break down sentiment percentages for each cryptocurrency. The "Total Tweets%" column provides insights into social media activity.

For sentiment analysis, the "Positive %," "Negative %," "Neutral %," and "Overall Sentiment %" columns are crucial, revealing the percentage of positive, negative, neutral sentiments, and an overall sentiment indicator. "Opening ($)" shows the opening price on the first day of the month, while "Closing ($)" indicates the closing price on the last day. "Volume (billions)" records the total trading volume, and the "Google index" reflects Google searches. Normalization is applied to several columns, including "Positive %," "Negative %," "Neutral %," "Overall Sentiment %," "Volume (billions)," and "Google index." This ensures comparability and meaningful analysis across different data scales. The dataset is a valuable resource for analysts, investors, and researchers, offering insights into cryptocurrency behavior, performance, and sentiment analysis. It enhances understanding in correlation and time series forecasting, providing valuable indicators of market sentiment.

11 Correlation Analysis

The Pearson correlation coefficient serves as an instrument for calculating the linear association between two variables. It assumes a pivotal role in the assessment and quantification of relationships existing amidst diverse predictor variables and target variables found within the dataset. Precisely, this statistical measure aids in determining the extent of correlation between two variables, signifying whether their association is robust or feeble. The output of the Pearson correlation coefficient is a numerical value that falls within the range of −1 to 1, with distinct interpretations.

First of all, a coefficient of 1 indicates a perfect positive linear relationship, implying that as one variable exhibits an increase, the other demonstrates a synchronous increase at a constant rate. Secondly, a coefficient of −1 signifies a perfect negative linear relationship, suggesting that as one variable experiences growth; the other exhibits a parallel and consistent decrease. Thirdly, when the coefficient assumes a value of 0, it conveys the absence of a linear relationship, denoting that the variables under consideration lack correlation.

Table 1 Collected data organization

Month	Coin	Sentiment					Prices			
		Total tweets %	Positive %	Negative %	Neutral %	Overall sentiment %	Opening ($)	Closing ($)	Volume (billions)	Google index
MAR-22	BNB	51.78	43.19	5.03	51.78	48.22	395.62	428.92	54	2621
JUN-22	BNB	57.92	33.79	8.32	57.92	42.08	320.47	219.30	39	2150
DEC-22	BNB	54.74	32.46	12.83	54.74	45.26	300.75	246.35	21	1899
MAR-22	BTC	57.02	33.25	9.78	57.02	42.98	43,194.50	45,538.68	831	2075
JUN-22	BTC	60.56	23.98	15.44	60.56	39.44	31,792.55	19,784.73	924	1506
DEC-22	BTC	57.66	29.68	12.66	57.66	42.34	17,168.00	16,547.50	541	2558
MAR-22	ETH	59.50	36.81	3.68	59.50	40.50	2919.78	3281.64	437	2155
JUN-22	ETH	71.07	23.47	5.33	71.07	28.93	1942.05	1067.30	562	1371
DEC-22	ETH	60.07	37.00	2.93	60.07	39.93	1295.77	1196.77	161	2558

The calculation of Pearson correlation coefficients transpires between a multitude of predictor variables (such as "Normalized Closing of Last Day," "Normalized Volume," "Normalized SumTotal Google Trends") and target variables (including "Normalized Overall Sentiment," "Normalized Positive Sum," "Normalized Negative Sum," and "Normalized Neutral Sum"). The importance of computing Pearson connection coefficients in this specific setting stems from the need to comprehend how the anticipating variables interrelate with the target variables. Such an investigation demonstrates valuable in numerous regards.

Fundamentally, it assists in recognizing anticipating-target relationships, offering clearness on which anticipating variables exert a huge impact on the target variables. After that, inside the area of predictive displaying and machine learning, understanding the connections between elements including anticipating variables and the target variable is of most extreme significance. Elements showing strong connections regularly emerge as likely parts for prescient models. Next, it adds to the creation of bits of knowledge identifying with the potential outcomes of changes in particular anticipating variables on the target variables. These bits of learning hold worth for educated dynamic producing and the elucidation of information examples.

To realize such a purpose, a heatmap illustrating the correlation coefficients among various sentiment measures for three specific cryptocurrencies: BNB, BTC, and ETH. Within each cryptocurrency, sentiment measures encompass sub-measures related to predictor variables, while columns represent target variables in Fig. 15. Such coefficients result in priceless insights into the interplay between sentiment and the three cryptocurrencies market indicators can be comprehended. As for ETH, Normalized Overall Sentiment exhibits a correlation coefficient of (0.3773) with "Normalized Closing of Last Day." This indicates a correlation where Ethereum's overall sentiment tends to rise modestly with an increase in its closing price at the end of each month. Conversely, a substantial negative correlation (−0.8531) between "Normalized Volume" and Normalized Overall Sentiment implies that higher Ethereum trading volumes coincide with decreased overall sentiment. A robust positive correlation (0.9885) between "Normalized SumTotal Google Trends" and Normalized Overall Sentiment reveals that heightened Google search activity corresponds to increased overall sentiment for Ethereum.

Normalized Positive Sum proves a moderate positive correlation (0.3811) with "Normalized Closing of Last Day," indicating that as Ethereum's closing price rises at month-end, positive sentiment similarly tends to increase, though modestly. "Normalized Volume" exhibits a strong negative correlation (−0.851) with Normalized Positive Sum, showing that increased Ethereum trading volumes coincide with a significant decrease in positive sentiment. There's a solid positive correlation (0.9879) between "Normalized SumTotal Google Trends" and Normalized Positive Sum, illustrating that increased Google search activity for Ethereum aligns with elevated positive sentiment. Weak negative correlation (−0.3507) observed between Normalized Negative Sum and "Normalized Closing of Last Day." There is a solid positive correlation (0.9879) between "Normalized SumTotal Google Trends" and Normalized Positive Sum, illustrating that increased Google search activity for Ethereum aligns with elevated positive sentiment. Weak negative correlation (−0.3507) was

Fig. 15 Heatmap of correlation coefficients

observed between Normalized Negative Sum and "Normalized Closing of Last Day."

In essence, as Ethereum's month-end closing price increases, negative sentiment slightly decreases, indicating minor shifts in sentiment with closing price. "Normalized Volume" and Normalized Negative Sum illustrate a notably strong correlation (0.8677), demonstrating that increased Ethereum trading volumes align with a significant rise in negative sentiment. "Normalized SumTotal Google Trends" and Normalized Negative Sum have a strong negative correlation (-0.9925), indicating that increased Google search activity for Ethereum aligns with a significant decrease in negative sentiment. "Normalized Closing of Last Day" and Normalized Neutral Sum exhibit a strong negative correlation (-0.8295). "Normalized Volume" and Normalized Neutral Sum show a less pronounced correlation (0.4174), implying a relatively weak positive relationship.

In conclusion, the substantial influence of public interest on Ethereum's market dynamics is highlighted by the robust correlation between heightened Google search activity and diminished neutral sentiment. The nuanced nature of Ethereum's sentiment dynamics is emphasized by the varied impacts of predictor variables on different sentiment aspects. The significance of factors such as closing prices and trading volume in shaping market sentiment is suggested by positive and negative correlations, respectively. The strong association between Google search trends and sentiment implies the importance of information-seeking behavior. These collective

findings indicate that monitoring Google Trendsdata, closing prices, and trading volume is imperative for insightful market analysis, enabling informed investment decisions and forecasting.

Regarding BTC, "Normalized Closing of Last Day" strongly correlates (0.867) with overall sentiment, signifying a noteworthy association with Bitcoin's closing price. "Normalized Volume" has a negative correlation (-0.1305) with "Normalized Overall Sentiment," suggesting that fluctuations in trading volume have a limited negative impact on Bitcoin's overall sentiment. Moreover, "Normalized SumTotal Google Trends" reveals a moderate positive correlation of (0.45) with "Normalized Overall Sentiment."

Positive sentiment, particularly with "Normalized Closing of Last Day," shows a notably strong correlation of (0.8883). This robust connection implies that positive sentiment in the cryptocurrency market is highly responsive to Bitcoin's closing price, with a higher closing price corresponding to overwhelmingly positive sentiment. In opposition, "Normalized Volume" exhibits a weak negative correlation of (-0.0864) with "Normalized Positive Sum." Furthermore, "Normalized SumTotal Google Trends" proves a moderate positive correlation of (0.4099) with "Normalized Positive Sum."

Negative sentiment, especially with "Normalized Closing of Last Day," exhibits a significant negative correlation of (-0.4232). A rise in closing price tends to reduce negative sentiment, emphasizing the pivotal influence of price performance (increases) on negative sentiment in the cryptocurrency market. "Normalized Volume" shows a significant positive correlation of (0.6766) with "Normalized Negative Sum." "Normalized SumTotal Google Trends" shows a significant negative correlation of (-0.8814) with "Normalized Negative Sum." High Google search interest in Bitcoin corresponds to a notable decrease in negative sentiment. Neutral sentiment, particularly starting with "Normalized Closing of Last Day," displays a remarkably strong positive correlation of (0.9751). Additionally, "Normalized Volume" demonstrates a modest positive correlation of (0.1663) with "Normalized Neutral Sum." "Normalized SumTotal Google Trends" shows a similar modest positive correlation of (0.1679) with "Normalized Neutral Sum."

To conclude, the correlation analysis stresses a consistent and substantial positive relationship between Bitcoin's closing price and all sentiment aspects, with correlation coefficients ranging from 0.867 to 0.9751. This implies that as Bitcoin's price increases, sentiment across positive, negative, and neutral aspects becomes notably more positive. In contrast, the impact of "Normalized Volume" on sentiment is relatively weak, suggesting a minor and inconsistent effect on overall sentiment and its components.

Concerning BNB, the coefficient between "Normalized Closing of Last Day" and "Normalized Overall Sentiment" is 0.9914, indicating a remarkably strong positive linear relationship. Likewise, the correlation coefficient between "Normalized Volume" and "Normalized Overall Sentiment" is 0.6927, demonstrating a positive correlation, though not as powerful as the closing price. The coefficient (0.8254) between "Normalized SumTotal Google Trends" and "Normalized Overall Sentiment" indicates a strong positive relationship. Increased Google search volume for Binance corresponds to a considerable rise

in overall sentiment. The correlation coefficient (0.7434) between "Normalized Closing of Last Day" and "Normalized Positive Sum" suggests a moderately strong positive linear relationship. The correlation coefficient between "Normalized Volume" and "Normalized Positive Sum" is (0.1635), denoting a positive correlation, though relatively weak.

The coefficient (0.3614) between "Normalized SumTotal Google Trends" and "Normalized Positive Sum" indicates a positive relationship. Higher Google search volume for Binance corresponds to an increase in positive sentiment. The correlation coefficient (−0.4831) between "Normalized Closing of Last Day" and "Normalized Negative Sum" suggests a moderate negative relation. An increase in Binance's closing price on the last day of the month corresponds to a noticeable decrease in negative sentiment.

Similarly, a burly negative correlation is indicated by the coefficient (−0.9239) between "Normalized Volume" and "Normalized Negative Sum." A substantial decrease in negative sentiment toward Binance is robustly linked to an increase in the trading volume. The coefficient (−0.8263) between "Normalized SumTotal Google Trends" and "Normalized Negative Sum" specifies a firm negative relationship. Higher Google search volume for Binance correlates strongly with decreased negative sentiment. The correlation coefficient (−0.4831) between "Normalized Closing of Last Day" and "Normalized Negative Sum" suggests a moderate negative relationship.

The coefficient (−0.9239) between "Normalized Volume" and "Normalized Negative Sum" conveys a solid negative correlation. Higher Binance trading volume is stoutly linked to a notable decrease in negative sentiment. The correlation coefficient (−0.8263) between "Normalized SumTotal Google Trends" and "Normalized Negative Sum" denotes a glaring negative relationship.

In conclusion, the noticeable patterns across these correlations indicate that the closing price, trading volume, and Google search trends are influential factors in shaping sentiment toward Binance. While overall sentiment and positive sentiment display positive relationships with these predictors, negative and neutral sentiments exhibit negative or weaker associations. This comprehensive understanding of correlations provides valuable insights for stakeholders in Binance, aiding in strategic decision-making, market analysis, and sentiment trend predictions. In considering investment decisions amid the Russian-Ukrainian War, Bitcoin emerges as an appealing choice due to its positive correlations with market variables, providing stability sought by investors during geopolitical uncertainties. Binance Coin, with its mixed correlations, demands a nuanced approach, considering potential impacts on its distinctive dynamics. Ethereum, showing positive associations akin to Bitcoin, may attract those seeking diverse opportunities. The choice among these cryptocurrencies, amidst the complexities of the war, should align with investors' risk tolerance and goals, necessitating thorough research, consideration of additional factors, and staying informed within the dynamic cryptocurrency market. Price forecasting analysis integration remains pertinent in navigating the evolving geopolitical landscape and making well-informed investment decisions.

12 Price Forecasting (SARIMA)

Cryptocurrency price forecasting has drawn escalating attention, as researchers and practitioners deploy advanced techniques to boost the meticulousness of predictions. The SARIMA (Seasonal Autoregressive Integrated Moving Average) library, a Python package that actualizes the SARIMA model for time series forecasting, is utilized for this purpose. Founded on the statsmodels library, which furnishes an extensive array of tools for statistical analysis in Python, the SARIMA library is instrumental in modeling and forecasting univariate time series data that exhibit non-stationarity [51, 52]. The SARIMA model extends the ARIMA model, standing for Autoregressive Integrated Moving Average, and introduces a seasonal component to capture periodic fluctuations occurring at fixed intervals, such as daily, weekly, monthly, or yearly.

SARIMA excels in capturing seasonal patterns, enhancing forecast precision. Addressing non-stationary data, it adeptly handles cyclical fluctuations and variations. SARIMA provides confidence intervals and diagnostic plots for forecasts, aiding in assessing uncertainty and reliability. Despite its benefits, SARIMA requires substantial historical data for accurate models and faces challenges in parameter determination and computational intensity [51, 53]. Figure 16 clarifies the predictions based on the prices spinning March, June, and December 2022. The prices in Table 1 are utilized in the training.

The figure illustrates price predictions for three cryptocurrencies—Ethereum (ETH), Bitcoin (BTC), and Binance Coin (BNB)—spanning from January to March 2023. Two lines are depicted: the blue line denotes actual prices, and the orange dotted line reflects predicted prices. Predictive outcomes are delineated for each cryptocurrency, encompassing Ethereum (ETH), Binance Coin (BNB), and Bitcoin (BTC) as shown above. These predictions offer valuable insights into anticipated price trends.

For Ethereum, the model forecasts $1727.21 on January 31, 2023, $1603.38 on February 28, 2023, and $1632.29 on March 31, 2023. Binance Coin is expected to
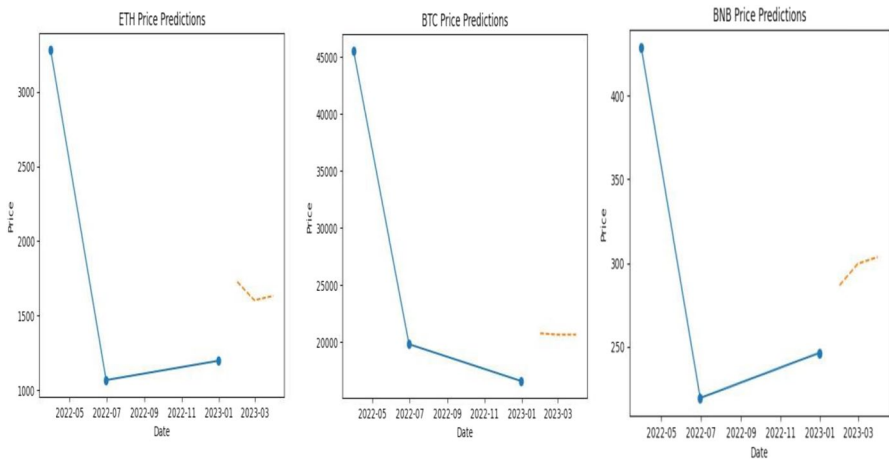

Fig. 16 Predicted vs. historical prices

reach \$286.71 on January 31, 2023, \$299.61 on February 28, 2023, and \$303.72 on March 31, 2023. Bitcoin's predicted prices are \$20,739.52 on January 31, 2023, \$20,630.29 on February 28, 2023, and \$20,633.13 on March 31, 2023. These projections offer a detailed insight into expected price movements, derived from the SARIMA model's analysis of historical data and temporal patterns.

To gauge prediction accuracy, the root mean squared error (RMSE) metric is utilized. It is applied to the model's predictions to compare them with the actual prices for the three cryptocurrencies in 2023 in (Fig. 17).

The aforementioned image comprises three line graphs, each delineating the price trends of distinct cryptocurrencies over a specified duration. The blue and orange lines in each graph represent the actual and predicted prices respectively. The first graph illustrates a representation of the real vs. predicted prices based on the root mean squared error (RMSE) for the Binance Coin (BNB) over January, February, and March 2023. The second graph presents a comparison of the actual versus predicted prices for Ethereum (ETH), with the accuracy of these predictions evaluated using the root mean squared error (RMSE) for the same period. The third graph offers a similar representation for Bitcoin (BTC), showcasing the real versus predicted prices based on the RMSE. The divergence between the actual and predicted prices in each graph signifies the prediction error. A greater divergence implies a higher RMSE, indicating a less accurate prediction, while a smaller divergence suggests a lower RMSE, indicative of a more accurate prediction.

The model predicts BTC closing prices at \$20,739.52, \$20,630.29, and \$20,633.13 for January 31, 2023, February 28, 2023, and March 31, 2023, respectively. BNB forecasts are \$286.71, \$299.61, and \$303.72, while ETH projections are \$1,727.21, \$1,603.38, and \$1,632.29 during the same period. Actual closing prices for BTC, BNB, and ETH on these dates are \$22,840.14, \$23,522.87, \$28,033.56, \$307.07, \$304.86, \$316.57, \$1,567.33, \$1,634.33, and \$1,792.74, respectively. RMSE values, assessing predictive accuracy, are 4745.03 for BTC, 131.99 for ETH, and 14.22 for BNB. Elevated BTC RMSE indicates challenges capturing its complex price patterns. ETH's lower RMSE suggests better predictive performance, while BNB's notably low RMSE indicates accurate predictions, hinting at more foreseeable price movements.

The distinct RMSE values underscore the variability in predicting different cryptocurrencies with the SARIMA model. The intricate field of predicting cryptocurrency prices requires continuous examination and adjustments to align with the
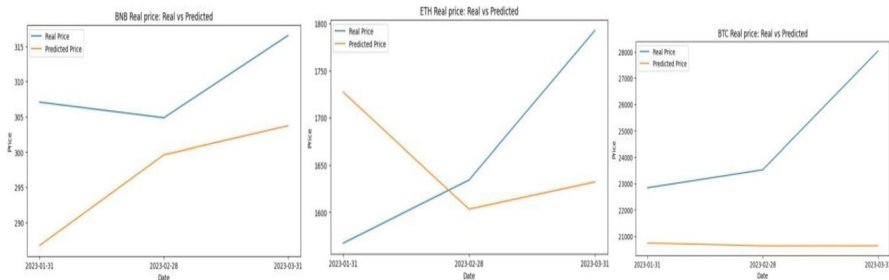

Fig. 17 Real vs. predicted prices RMSE

ever-changing dynamics of the market. This suggests that the predictability of each cryptocurrency differs, emphasizing the ongoing need to fine-tune model parameters for reliable predictions. Given the geopolitical uncertainties introduced by the Russian-Ukrainian War, accurate cryptocurrency price forecasting becomes increasingly crucial. While the SARIMA model holds promise, its effectiveness hinges on continuous adaptation to the unique characteristics of each cryptocurrency. Persistent efforts to improve forecasting techniques are essential, particularly in response to geopolitical shifts and market fluctuations linked to the ongoing conflict.

13 Conclusion and Future Research

In conclusion, this study has provided a comprehensive evaluation of predictive models—support vector machine (SVM) classifier, convolutional neural network-long short-term memory (CNN-LSTM) model, and Pysentimento sentiment analyzer—in the context of analyzing cryptocurrency markets during the Russian-Ukrainian War. The research aimed to identify the most accurate method for predicting cryptocurrency prices and trends, considering the influence of geopolitical events on market dynamics.

The study finds that the Pysentimento model outperformed the conventional machine learning and deep learning models. Despite not being specifically trained on cryptocurrency-related tweets, the Pysentimento model demonstrated robustness by achieving superior performance in sentiment classification. Regarding the effectiveness of sentiment analysis on tweets and Google Trends in predicting emotional tendencies and prices of Bitcoin, Ethereum, and Binance Coin, the findings confirmed the utility of sentiment analysis. Leveraging Pysentimento alongside Google Trends data, the study provided valuable insights into market sentiment and behavior during the Russian-Ukrainian War. The results indicated that sentiment analysis on social media and Google Trends data can produce near accurate predictions, enhancing our understanding of market dynamics. The research acknowledges the correlation between predictions based on social media sentiment and the actual interest of cryptocurrencies retrieved from Google Trends and other factors like Pearson Correlations. Furthermore, the study validated SARIMA model predictions through the calculation of RMSE, offering insights into the profitability and stability of cryptocurrencies during times of conflict. These findings provide valuable guidance for investment strategies and underscore the importance of incorporating sentiment analysis and time series forecasting in cryptocurrency market analysis.

Looking ahead, future research will focus on refining these predictive models, exploring sophisticated SARIMA models for price forecasting, and integrating advanced Large Language Models (LLMs) and Pathways Language Model (PALMs) to process extensive datasets for more nuanced predictions. The research calls for a comprehensive approach that encompasses diverse data sources, including market demand, tokenomics, technological advancements, regulatory changes, economic indicators, social media sentiment, network metrics, and the influence of competing cryptocurrencies, alongside extended datasets and current market data such as trading volumes and values. This approach aims to address the challenges posed by the volatile

cryptocurrency market and to expand upon the insights gained, enhancing the precision and adaptability of predictive analytics tools in this rapidly evolving landscape. The study's findings and future directions offer a promising path forward, building on the questions posed and contributing to the broader field of predictive analytics in digital currencies. In summary, this study contributes to the field of machine learning and cryptocurrency research by demonstrating the effectiveness of sentiment analysis and time series forecasting in predicting market trends and prices, particularly during geopolitical events. The findings offer valuable insights for investors, traders, and analysts, paving the way for further advancements in predictive analytics in digital currencies.

Author Contribution As for the roles of the authors, they are provided as follows: M.R. authored the manuscript and conducted the research experiments and analysis. M.G. provided consultation on machine learning and deep learning approaches, guided the experimental methods and analysis, and reviewed the coding. M.K. and S.A. served as research supervisors.

Data Availability No datasets were generated or analysed during the current study.

Code Availability Codes cannot be shared openly, to protect the author's privacy due to the fact that some of such codes are directly involved in my Ph.D. thesis, which has not yet been defended.

Declarations

Competing Interests The authors declare no competing interests.

References

1. Wołk K (2020) Advanced social media sentiment analysis for short-term cryptocurrency price prediction. Expert Syst 37(2):e12493
2. Sattarov O, Jeon HS, Oh R, Lee JD (2020) Forecasting bitcoin price fluctuation by Twitter sentiment analysis. Science and Communications Technologies (ICISCT). IEEE, pp 1–4
3. Al-Mansour BY (2020) Cryptocurrency market: behavioral finance perspective. J Asian Finance Econ Bus 7(12):159–168. https://doi.org/10.13106/jafeb.2020.vol7.no12.159
4. Pano T, Kashef R (2020) A complete vader-based sentiment analysis of Bitcoin (BTC) tweets during the era of covid-19. Big Data Cogn Comput 4(4):33. https://doi.org/10.3390/bdcc4040033
5. Woodward M (2022) Twitter user statistics: how many people use twitter in 2022? Matthew Woodward. Retrieved November 1, 2022, from https://www.matthewwoodward.co.uk/work/twitter-user-statistics
6. Reitan J, Faret J, Gambäck B, Bungum L (2015) Negation scope detection for twitter sentiment analysis. In: Proceedings of the 6th workshop on computational approaches to subjectivity, sentiment and social media analysis. https://doi.org/10.18653/v1/w15-2914
7. Drus Z, Khalid H (2019) Sentiment analysis in social media and its application: systematic literature review. Procedia Comput Sci 161:707–714
8. Balasudarsun NL, Ghosh B, Mahendran S (2022) Impact of negative tweets on diverse assets during stressful events: an investigation through time-varying connectedness. J Risk Fin Manag 15(6):260. https://doi.org/10.3390/jrfm15060260
9. Abraham J, Higdon D, Nelson J, Ibarra J (2018) Cryptocurrency price prediction using tweet volumes and sentiment analysis. SMU Data Science Review 1(3):1
10. Hyunyoung C, Varian H (2018) Replicating "predicting the present with google trends" by Hyunyoung Choi and Hal Varian (the economic record, 2012). Economics. https://doi.org/10.5018/economics-ejournal.ja.2018-34

11. Valencia F, Gómez-Espinosa A, Valdés-Aguirre B (2019) Price movement prediction of crypto-currencies using sentiment analysis and machine learning. Entropy 21(6):589. https://doi.org/10.3390/e21060589

12. Huang X, Zhang W, Tang X, Zhang M, Surbiryala J, Iosifidis V, Zhang J (2021) Lstm based sentiment analysis for cryptocurrency prediction. In: Database systems for advanced applications: 26th international conference, DASFAA 2021, Taipei, Taiwan, April 11–14, 2021, proceedings, part III 26. Springer international publishing, pp 617–621

13. Liu Z, Lin W, Shi Y, Zhao J (2021) A robustly optimized Bert pre-training approach with post-training. Lect Notes Comput Sci 1:471–484. https://doi.org/10.1007/978-3-030-84186-7_31

14. Fonseca J (2020) PyTrends (Version 4.9.1) [Computer software]. GitHub. Retrieved December 15, 2022, from https://github.com/GeneralMills/PyTrends

15. Russel MA, Klassen M (2018) Mining the social web: data mining facebook, twitter, linkedin, instagram, github, and more. O'Reilly

16. Dundee (2002) Bitcoin tweets sentiment analysis: cnn-lstm. Kaggle. Retrieved November 5, 2022, from https://www.kaggle.com/code/dundee2002/bitcoin-tweets-sentiment-analysis-GloVe-cnn-lstm/log

17. Yahoo! (n.d.) Yahoo finance - stock market live, quotes, business & finance news. Yahoo! Finance. Retrieved January 8, 2023, from https://finance.yahoo.com/

18. Johnson R, Smith T, Williams K, Davis M (2020) Using Google Trendsdata to explore public interest in breast cancer screening. BMC Public Health 20(1):1–6

19. Cortes C, Vapnik VN (1995) Support-vector networks. Mach Learn 20(3):273–297

20. Pennington J, Socher R, Manning C (2014) GloVe: global vectors for word representation. In: Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP). https://doi.org/10.3115/v1/d14-1162

21. Ramos J (2003) Using tf-idf to determine word relevance in document queries. In: Proceedings of the first instructional conference on machine learning, vol 242, no. 1. pp 133–142

22. Bollen J, Mao H, Zeng X (2011) Twitter mood predicts the stock market. J Comput Sci 2(1):1–8. https://doi.org/10.1016/j.jocs.2010.12.007

23. Medhat W, Hassan A, Korashy H (2014) Sentiment analysis algorithms and applications: a survey. Ain Shams Eng J 5(4):1093–1113

24. Wang SI, Manning CD (2012) Baselines and bigrams: simple, good sentiment and topic classification. In: Proceedings of the 50th annual meeting of the association for computational linguistics, vol. 2. Short papers, pp 90–94

25. Schölkopf B, Platt JC, Shawe-Taylor J, Smola AJ, Williamson RC (2001) Estimating the support of a high-dimensional distribution. Neural Comput 13(7):1443–1471. https://doi.org/10.1162/089976601750264965

26. Hsu C-W, Chang C-C, Lin C-J (2003) A practical guide to support vector classification (Technical Report). Department of Computer Science and Information Engineering, National Taiwan University

27. LeCun Y, Bottou L, Bengio Y, Haffner P (1998) Gradient-based learning applied to document recognition. Proc IEEE 86(11):2278–2324

28. Kim Y (2014) Convolutional neural networks for sentence classification. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). https://doi.org/10.3115/v1/d14-1181

29. Yamashita R, Nishida Y, Kido R, Akita K (2018) Convolutional neural networks: an overview and applications in medical image analysis. In Medical Imaging Informatics. Springer, Cham, pp 449–483

30. Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9(8):1735–1780

31. Wang Y, Sun Y, Liu T, Huang X (2016) A CNN-LSTM based model for text classification. arXiv preprint arXiv:1611.07205. Retrieved from https://arxiv.org/abs/1611.07205. (ar5iv)

32. Shuang K, Ren X, Chen J, Shan X, Xu P (2017) Combining word order and CNN-LSTM for sentence sentiment classification. In: Proceedings of the 2017 international conference on software and e-business. https://doi.org/10.1145/3178212.3178230

33. Liu L, Shang J, Ren X, Xu F, Gui H, Peng J, Han J (2018) Empower sequence labeling with task-aware neural language model. In: Proceedings of the AAAI conference on artificial intelligence, vol. 32, No. 1

34. Sutskever I, Vinyals O, Le QV (2014) Sequence to sequence learning with neural networks. In: Advances in neural information processing systems, 27

35. Donahue J, Anne Hendricks L, Guadarrama S, Venugopalan S, Saenko K, Darrell T (2015) Long-term recurrent convolutional networks for visual recognition and description. IEEE Trans Pattern Anal Mach Intell 39(4):677–691. https://doi.org/10.1109/TPAMI.2015.2500292

36. Zhang Y, Marshall I, Wallace BC (2016) Rationale-augmented convolutional neural networks for text classification. In: Proceedings of the conference on empirical methods in natural language processing. Conference on empirical methods in natural language processing, vol 2016. NIH public access, p 795

37. Pérez JM, Rajngewerc M, Giudici JC, Furman DA, Luque F, Alemany LA, Martínez MV (2021) Pysentimiento: a python toolkit for opinion mining and social nlp tasks. arXiv preprint arXiv:2106.09462

38. Pérez JM, Furman DA, Alemany LA, Luque F (2022) RoBERTuito: a pre-trained language model for social media text in Spanish. arXiv preprint arXiv:2111.09453

39. Devlin J, Chang, MW, Lee K, Toutanova K (2019) BERT: pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805

40. Perez JM, Rajngewerc M, Giudici JC, Furman DA, Luque F, Alemany LA, Martínez MV (2023) Pysentimiento: a python toolkit for opinion mining and social NLP tasks. https://doi.org/10.21203/rs.3.rs-3570648/v1

41. Zhao H, Crane M, Bezbradica M (2022) Attention! transformer with sentiment on cryptocurrencies price prediction. In: Proceedings of the 7th international conference on complexity, future information systems and risk. https://doi.org/10.5220/0011103400003197

42. Chen H, Sun M, Tu C, Lin Y, Liu Z (2016) Neural sentiment classification with user and product attention. In: Proceedings of the 2016 conference on empirical methods in natural language processing. https://doi.org/10.18653/v1/d16-1171

43. Microsoft (n.d.) Microsoft/Huggingface-Transformers: transformers: state-of-the-art natural language processing for pytorch and tensorflow 2.0. GitHub. Retrieved May 28, 2021, from https://github.com/microsoft/huggingface-transformers

44. Gaber M, Ezzat M, Mokhtar M (2021) Hyperparameter optimization for deep learning-based sentiment analysis. IEEE Access 9:78030–78047

45. Sharma P, Sharma D (2022) Classification reports: essential tools for sentiment analysis model evaluation. arXiv preprint arXiv:2208.03906

46. Sharma N, Khosla A, Kim T, Gade A, Pagh R (2023) TF-IDF: a fundamental technique in natural language processing. GitHub Repository. Retrieved February 10, 2023, from https://github.com/GeneralMills/PyTrends

47. Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J (2013) Distributed representations of words and phrases and their compositionality. In: Advances in neural information processing systems, 26

48. Xue N (2010) In: Bird S, Klein E, Loper E (eds) Natural language processing with python. O'Reilly Media, Inc. 2009. Nat Lang Eng 17(3):419–424. https://doi.org/10.1017/s1351324910000306

49. Carpedm20 (2015) Emoji: emoji for Python (Version 1.7.0) [Software]. Retrieved March 12, 2022, from https://github.com/carpedm20/emoji

50. Bitsgap (2021) What is a trading volume in cryptocurrency and why is it important? Retrieved April 25, 2022, from https://bitsgap.com/blog/what-is-a-trading-volume-in-cryptocurrency-and-why-is-it-important

51. Brownlee J (2020) How to develop LSTM models for time series forecasting. Machine learning mastery. Retrieved December 3, 2022, from https://machinelearningmastery.com/how-to-develop-lstm-models-for-time-series-forecasting/

52. Aggarwal PK (2020) Powerful learning is all about retrieval. ASCD Education Update 62(1):1–5. Retrieved December 30, 2022, from https://www.ascd.org/el/articles/powerful-learning-is-all-about-retrieval

53. Hyndman RJ, Athanasopoulos G (2021) Forecasting: principles and practice, 3rd edn. OTexts. https://otexts.com/fpp3/

Authors and Affiliations

Muhammad Nabil Rateb[1] · Sameh Alansary[2] · Marwa Khamis Elzouka[3] · Mohamad Galal[4]

✉ Muhammad Nabil Rateb
mohamed.rateb@alexu.edu.eg

Sameh Alansary
s.alansary@alexu.edu.eg

Marwa Khamis Elzouka
m.khamis@alexu.edu.eg

Mohamad Galal
Mo.abdelaziz@nu.edu.eg

[1] Department of English Language and Translation, Alexandria University, El-Gaish Rd, Al Azaritah WA Ash Shatebi, Bab Sharqi, Alexandria 5424010, Arab Republic of Egypt

[2] Phonetics Department, Alexandria University, El-Gaish Rd, Al Azaritah WA Ash Shatebi, Bab Sharqi, Alexandria 5424010, Arab Republic of Egypt

[3] English , Language and Translation Department, Alexandria University, El-Gaish Rd, Al Azaritah WA Ash Shatebi, Bab Sharqi, Alexandria 5424010, Arab Republic of Egypt

[4] Department of Informatics, Nile University, 26th of July Corridor Sheikh Zayed City, Sheikh Zayed City, Giza 3247010, Arab Republic of Egypt