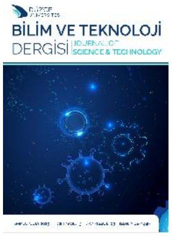

Düzce University
Journal of Science & Technology

Research Article

Performance Comparison of Traditional and Contextual Representations for Cryptocurrency Sentiment Analysis on Twitter

Melisa ATEŞ [a], Muhammet Sinan BAŞARSLAN [a*]

[a] *Department of Computer Engineering, Faculty of Engineering and Natural Sciences, Istanbul Medeniyet University, Istanbul, TURKEY*
** Corresponding author's e-mail address: muhammet.basarslan@medeniyet.edu.tr*

DOI: 10.29130/dubited.1673097

ABSTRACT

In recent years, discussions about cryptocurrencies, particularly on platforms such as Twitter, have become increasingly prevalent. This study focuses on conducting a sentiment analysis (SA) of tweets related to cryptocurrencies, applying machine learning (ML) and deep learning (DL) methodologies based on natural language processing (NLP). This research used a total of 10,000 tweets collected from open sources between 2020 and 2021. Prior to analysis, the dataset underwent detailed pre-processing, during which non-textual elements such as emojis, links, and HTML codes were removed. TF-IDF was initially employed to generate text representations. Various traditional ML models were applied, including Naïve Bayes (NB), Decision Tree (DT), Support Vector Machine (SVM). Advanced DL models were also used, including Bidirectional Long Short-Term Memory (BiLSTM) and Bidirectional Gated Recurrent Unit (BiGRU). To capture contextual relationships more effectively, text embeddings generated by the Bidirectional Encoder Representations from Transformers (BERT) model were also utilised. When performance was evaluated, the BERT-based BiGRU model achieved the highest Accuracy (Acc) of 93% and the best F1 score. This demonstrates the effectiveness of combining deep contextual embeddings with models capable of learning from sequential patterns. Overall, the findings suggest that DL approaches, particularly those that incorporate advanced representation methods such as BERT, can significantly outperform traditional models in sentiment classification tasks.

Keywords: *Cryptocurrency, sentiment analysis, text representation, ML, DL*

Twitter'da Kripto Para Duygu Analizi için Geleneksel ve Bağlamsal Temsillerin Performans Karşılaştırması

ÖZ

Son yıllarda, özellikle Twitter gibi platformlarda kripto para birimleri hakkındaki tartışmalar giderek yaygınlaşmaktadır. Bu çalışma, doğal dil işleme (NLP) tabanlı makine öğrenimi (ML) ve derin öğrenme (DL) metodolojilerini uygulayarak kripto paralarla ilgili tweetlerin duygu analizini (SA) yapmaya odaklanmaktadır. Bu araştırmada 2020 ve 2021 yılları arasında açık kaynaklardan toplanan toplam 10.000 tweet kullanılmıştır. Analiz öncesinde veri kümesi, emojiler, bağlantılar ve HTML kodları gibi metin dışı öğelerin kaldırıldığı ayrıntılı bir ön işlemden geçirilmiştir. Metin temsilleri oluşturmak için başlangıçta TF-IDF kullanılmıştır. Naïve Bayes (NB), Karar Ağacı (DT), Destek Vektör Makinesi (SVM) dahil olmak üzere çeşitli geleneksel makine öğrenimi modelleri uygulanmıştır. Çift Yönlü Uzun Kısa Süreli Bellek (BiLSTM) ve Çift Yönlü Geçitli Tekrarlayan Birim (BiGRU) dahil olmak üzere gelişmiş DL modelleri de kullanılmıştır. Bağlamsal ilişkileri daha etkili bir şekilde yakalamak için, Transformatörlerden Çift Yönlü Kodlayıcı Temsilleri (BERT) modeli tarafından üretilen kelime gömmede

Received: 10/04/2025, Revised: 22/06/2025, Accepted: 29/06/2025

kullanılmıştır. Performans değerlendirildiğinde, BERT tabanlı BiGRU modeli en yüksek doğruluğu (%93) ve en iyi F1 puanını elde etmiştir. Bu, derin bağlama dayalı kelime gömmeleri sıralı örüntülerden öğrenme yeteneğine sahip modellerle birleştirmenin etkinliğini göstermektedir. Genel olarak bulgular, DL yaklaşımlarının, özellikle de BERT gibi gelişmiş temsil yöntemlerini içerenlerin, duygu sınıflandırma görevlerinde geleneksel modellerden önemli ölçüde daha iyi performans gösterebileceğini göstermektedir.

Anahtar Kelimeler: *Kripto para, duygu analizi, kelime temsili, ML, DL*

I. INTRODUCTION

The emergence of cryptocurrencies has transformed the financial landscape, offering a decentralised, transparent, and reliable alternative to traditional currencies and financial systems. In 2008, Satoshi Nakamoto published a white paper entitled “Bitcoin: A Peer-to-Peer Electronic Cash System” [1], after which thousands of cryptocurrencies emerged in this field. Crypto assets have particularly revolutionised the world of finance recently. As the use of cryptocurrencies has grown and become widespread, investors have adapted to these new assets and started developing strategies to track their value and trends. Social media platforms, particularly Twitter, have played a significant role in this process. As the market expands, the volume of discussion and posts about crypto assets on social media platforms increases, particularly on Twitter. Today, Twitter provides a wealth of data on crypto assets, making it easy to analyse investor sentiment.

SA, which is the analysis of social media posts, is becoming an increasingly important part of financial market analysis.

SA is a NLP technique used to determine the emotional tone of text, i.e. whether it is positive, negative or neutral. This technique can be used to identify the potential market impact of social media posts. Previous studies have shown that the sentiment expressed in Twitter users' posts can predict stock market movements [2]. The researchers stated that 'Twitter sentiment can predict stock market movements, and social media analysis can be a powerful tool in financial forecasting'. Similarly, another study found a relationship between Bitcoin price and Twitter user sentiment, concluding that Twitter data could be used for crypto market predictions [3, 4]. The high level of investor sentiment surrounding cryptoassets provides a suitable environment for SA. The fact that investors react emotionally to market movements yields sharper results in the analysis process.

Cryptocurrencies have emerged as a transformative force in the global financial ecosystem, characterized by their decentralized structure and high market volatility. With the increasing popularity of platforms such as Twitter, social media has become a vital channel through which investors and the public share opinions, disseminate information, and shape market sentiment. Understanding these sentiments is crucial, as emotionally driven reactions to market events can significantly influence cryptocurrency price fluctuations.

Despite growing interest from both academia and industry, there remains a need for comprehensive studies that integrate large-scale social media data with advanced SA techniques to effectively capture and interpret public opinion dynamics. Although several studies have examined SA in the context of cryptocurrencies, many suffer from limitations in dataset size, temporal coverage, or methodological diversity. A considerable portion of prior research has focused on narrow timeframes or small datasets, frequently relying on a single ML algorithm without conducting comparative evaluations. Moreover, few have systematically assessed the performance of DL models—especially transformer-based architectures such as BERT—alongside traditional approaches.

To address these gaps, this study analyzes the sentiment of approximately 10,000 tweets related to cryptocurrencies [5], collected from Kaggle and spanning the 2021–2022 period—a time marked by intense market fluctuations and heightened public interest. The tweets were preprocessed and transformed using TF–IDF [6] for classical ML models [7], while BERT-based embeddings [8] were used to capture contextual meaning in transformer and deep neural architectures. A wide range of models—including (NB, DT, SVM, BiLSTM, BiGRU, and BERT—were trained and evaluated with class-weighted handling to address potential class imbalances in sentiment labels.

The primary contributions of this paper are threefold:

- It presents a large-scale and temporally relevant SASE using real-world social media data;
- It compares the performance of six distinct models, ranging from traditional ML algorithms to advanced DL architectures (e.g., BiGRU, BERT), under two different text representation strategies (TF-IDF and BERT embeddings);
- It offers insights into how social media sentiment can be leveraged to better understand, and potentially anticipate, trends in cryptocurrency markets.

One notable study in this field is that of Konstantinos et al., who examine the effectiveness of large language models for analysing the sentiment in cryptocurrency news articles. Their fine-tuning and comparative performance evaluations of these models provide insights that can inform investment decisions and risk management strategies. Using GPT, the researchers achieved scores of 0.867 Acc and 0.873 F1 [9]. In their Master's thesis, Klaudia Byc and Stefania-Cornelia Ilinca examined the relationship between Twitter sentiment and Bitcoin price movements, finding no significant correlation. This suggests that social media sentiment may have a limited impact on price prediction. This review evaluates the effectiveness of DL models in predicting cryptocurrency prices and analyses the performance of new DL models [10]. In Indonesia, Ramaputra et al. predict that the number of cryptocurrency investors will reach 18.83 million by January 2024, which indicates a growing interest in this market. The researchers conducted a SA of user reviews on Indodax and Tokocrypto, which are the leading cryptocurrency trading platforms in Indonesia. Following TF-IDF processing, the NB method achieved an Acc of 82.97% and an F1 of 77.52% on Indodax data and an Acc of 81.82% and an F1 of 77.59% on Tokocrypto data [11]. These studies demonstrate the importance of SA in cryptocurrency markets and the various ways in which it is being applied. Tools such as social media SA, DL models, and large language models provide valuable insights into understanding and predicting cryptocurrency price movements. Table 1 gives literature summary information.

Table 1. *Literature works regarding the Cryptocurrency SA*

Author	Field	Models	Performance Metrics	Key Finding
Konstantinos et.al. [9]	SA of cryptocurrency news	Large Language Models (LLMs)	Acc:0.867 F1:0.873	LLMs are effective for supporting investment decisions and risk analysis
Klaudia Byc & Stefania-Cornelia Ilinca [10]	Relationship between Twitter sentiment and Bitcoin prices	Social media SA	-	No significant correlation was found between Twitter sentiment and Bitcoin price movements.
Ramaputra et al. [11]	Investor growth prediction & SA of user reviews	TF-IDF + NB	Indodax: Acc 82.97%, F1 77.52% Tokocrypto: Acc 81.82%, F1 77.59%	Indonesia is projected to reach 18.83 million crypto investors by Jan 2024. User reviews reveal platform satisfaction insights
Mai et al.[12]	Bitcoin price prediction using social media SA	LSTM, GRU, SVM	-	Sentiment scores enhance prediction accuracy. LSTM outperformed other models.

Table 1 (cont). *Literature works regarding the Cryptocurrency SA*

Alam et al.[13]	Crypto SA	CNN+LSTM	Acc: 85%	Combining social media and news sources yields better forecasting results
McNally et al. [14]	Bitcoin price forecasting	RNN & LSTM	Acc: 52%, 58%	LSTM slightly outperformed traditional models but overall accuracy remained limited.
Abraham et al.[15]	Sentiment from Twitter & Reddit for price prediction	VADER, BERT	F1, RMSE	Reddit provided more meaningful sentiment data than Twitter. BERT-based models performed better.

Our aim is to investigate the most effective classification method following frequency-based TF-IDF and transformer-based BERT on unbalanced tweet data. To this end, we will use different ML (NB, SVM, and DT) and DL (BiLSTM and BiGRU) methods to determine the most effective SA approach. Section 2 provides the background, text representation, ML and DL. Section 3 presents the experimental setup and results, and section 4 contains the conclusions and discussion.

II. BACKGROUND

This section outlines the methods of text representation and ML that were employed in this study.

A. TEXT REPRESENTATION

In this study, text data is represented using the TF-IDF method. TF-IDF is a statistical measure that assesses the importance of a term in a document [6]. It is used to determine the importance of terms in documents. Term frequency (TF) indicates how often a particular term occurs in a document, while inverse document frequency (IDF) indicates how common or rare a term is in the entire dataset. This representation method helps determine the importance of terms in text data, converting them into numerical features that can be used by ML models.

A. 1. Tf-Idf

The TF-IDF method emphasises important words that need to be emphasised in documents, while de-emphasising less important words. This ensures that important words are highlighted. The TF-IDF score is calculated as the product of the TF-IDF scores of a term t in a document j and the corpus D as a whole. TF is the method used to calculate the weights of terms in a document. Equation (1) can be seen [16].

$$TF(i,j) = \frac{Term\ i\ frequency\ in\ document\ j}{Total\ \ words\ in\ document\ j} \tag{1}$$

IDF tries to find the number of words that occur in more than one document and determine whether the word is a term or not. This is done by dividing the absolute value of the logarithm of the number of documents containing the term by the number of documents. Equation (2) can be seen [16].

$$IDF(i) = log\left(\frac{Total\ documents}{documents\ with\ term\ i}\right) \tag{2}$$

In this study, the cleaned tweet texts are transformed into feature vectors for classifying using the TF-IDF representation.

A.2. Bert

BERT is a language representation model proposed by Google in 2018, which has revolutionised the field of NLP. Based on the Transformer architecture, BERT processes text bi-directionally, meaning that it simultaneously analyses the context to the left and right of a word. This structure enables BERT to represent words in much deeper and more meaningful contexts. In the pre-training process, it learns the structure of the language using the masked language modelling and next sentence prediction tasks. In this way, it builds a strong basic model that enables transfer learning. BERT has achieved highly successful results in various NLP tasks (SA, name recognition, text classification, question-answer systems, etc.) and has outperformed most of the traditional methods used in this field. In this study, we aim to use the contextual representations provided by BERT to more accurately analyse the sentimental states contained in tweets [8].

B. DATASET

The dataset used in this study [5] is a dataset of approximately 10,000 tweets related to cryptocurrencies between 2021 and 2022. This dataset was obtained from the Kaggle platform and contains sentiment features related to crypto assets. The dataset was cleaned by pre-processing and made ready for SA. Table 2 presents a detailed summary of the dataset's characteristics and structure.

Table 2. *Data set information.*

Feature	Feature Description
Content	Raw text content of the tweet
Cleaned_Tweet	Cleaned version of the tweet content
Sentiment	Sentiment tag of the tweet (positive, neutral, negative)

In this study, preprocessing is applied to the texts to obtain cleaner and more meaningful data for NLP models. First, the emoji characters in the text are detected and cleaned. This was done by removing characters including facial expressions, symbols, transport signs and flags according to Unicode ranges using regular expressions.

Line breaks in the text are then merged by replacing them with the space character. The hashtag (#) and underscore (_) characters are then removed. Mentions (for example, @kullanici) are simplified by removing only the @ sign. Characters such as &amp; which are common in HTML content, are replaced with the more readable and. Also, URLs (starting with http, https, www) and HTML tags (e.g. <b>, </div>) are completely removed from the text using regular expressions. Finally, spaces at the beginning and end of the text are removed.

As a result of these processes, the resulting text is stripped of unnecessary elements that make it difficult for the model to understand. This results in more reliable and accurate results in NLP tasks such as SA. This provides cleaner and more meaningful data for NLP models. The class distribution of the cleaned dataset is shown in Figure 1.

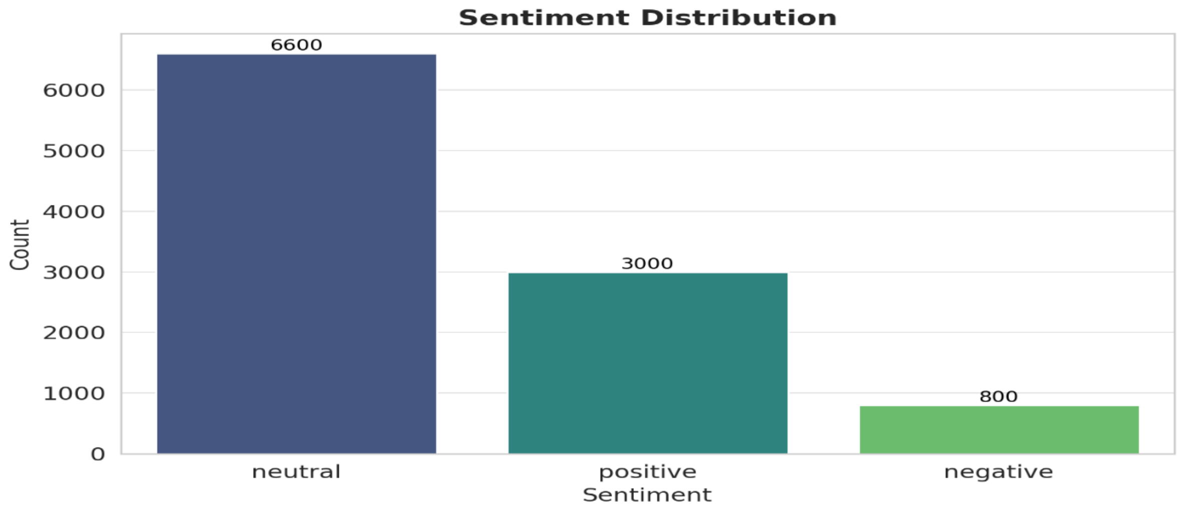

Figure 1. *class distribution*

In the dataset shown in Figure 1, 6657 tweets were labelled as neutral, 2967 as positive and 814 as positive. The Textblob library was used for this labelling. This distribution indicates that the dataset is unbalanced and the neutral class is dominant. This was taken into account during model training to ensure learning balance between classes, and class-weighted metrics were also reported during performance evaluation.

C. MACHINE LEARNING

NB aims to classify, i.e. categorise, data using computations based on probabilistic principles. It assumes that features with class labels are conditionally independent, and because of this simplicity and efficiency it performs well in a variety of text classification tasks [7].

C.1. Naïve Bayes

NB aims to classify, i.e. categorise, data using computations based on probabilistic principles. It assumes that features with class labels are conditionally independent, and because of this simplicity and efficiency it performs well in a variety of text classification tasks [17].

C.2. Decision Tree

DT is a predictive model that recursively splits the data based on feature values. It consists of internal nodes that perform tests on features, and leaves that provide final predictions. This model is commonly used in both classification and regression due to its ease of interpretation [18].

C.3. Support Vector Machine

SVM is a powerful classification method that stands out among supervised learning algorithms for its effectiveness, especially on high-dimensional data sets. SVM determines an optimal decision boundary that maximises the margin between classes to best separate different classes. SVM, which can be applied directly to linearly separable datasets, can also effectively solve non-linear problems thanks to its kernel functions [19].

D. DEEP LEARNING

DL, as a subfield of ML, focuses on learning patterns and representations in data using neural networks. In contrast to the need for manual feature engineering common in ML algorithms, DL models can automatically learn hierarchical and multi-layered features from data. This feature allows DL algorithms to build robust models that provide high Acc for complex tasks [18].

D.1. Long Short-Term Memory

LSTM [18] to solve the vanishing gradient problem commonly encountered in classical Recurent Neural Network (RNN). The LSTM cell has three basic gate mechanisms that control the flow of information: the input gate, the forgetting gate and the output gate. The input gate determines the extent to which incoming information is incorporated into the cell state, while the forget gate decides how much information from the previous time step is retained and how much is discarded. The output gate controls how much of the information stored in the cell is transferred to the next layers [20]. Each LSTM cell contains the current input (X_t), the previous output (h_{t-1}), the previous cell state (C_{t-1}) as input values. Current output (h_t), Current cell state (C_t) as output values. Information about the LSTM architecture can be found in Table 3.

Table 3. *Information about the LSTM architecture.*

Variables	Description
$X_t \in \mathbb{R}^d$	Vector representing the input fed into the LSTM unit
$f_t \in \mathbb{R}^d$	Activation value associated with the forget gate mechanism
$i_t \in \mathbb{R}^h$	Output gate activation signal
$h_t \in \mathbb{R}^h$	Hidden state vector, which also serves as the output representation of the LSTM cell
$\tilde{C}_t \in \mathbb{R}^h$	Input modulation vector contributing to the cell's internal computation
$C_t \in \mathbb{R}^h$	Internal memory representation, commonly referred to as the cell state
$W \in \mathbb{R}^{hXd}$	The parameters to be learned during the training process include weight matrices and bias vectors. These parameters are used in The superscripts h and d denote the number of hidden units and the number of input features, respectively.

Deciding what information to discard is the first step in LSTM. This is done by the layer with a sigmoid activation function called f_t. According to X_t and h_{t-1}, the cell state C_{t-1} produces an output between zero and one for each number. If the output is one, the input value is retained if the output value is one and discarded if the output value is zero. The next step of the LSTM cell is to decide whether to keep the new information or not. This step has two parts. The first is the input layer, where the sigmoid is used to decide whether to update the data. Then, in the next layer, tanh is applied and creates a new candidate vector, $\tilde{C}_t$, which can be added to the new state. In the next step, i_t and $\tilde{C}_t$, are combined to update the state. This produces a candidate value, which is scaled to decide how much each value should be updated. C_{t-1} and f_t are multiplied to forget the previous decision. Finally, it decides what to send to the output. This output depends on the state of the cell. The output of the sigmoid in this layer o_t is multiplied by the new value between (+1) and (-1) generated by the tanh layer from the cell state C_t.

LSTMs are the storage form of the RNN algorithm. They are used in various fields such as text processing, handwriting recognition, machine translation [20-23].

D.2. Bidirectional Long Short-Term Memory

BiLSTM is an advanced RNN architecture with external iterations providing both forward and backward information flow, as well as internal information transfer processes controlled by gate mechanisms. The LSTM gates include the forget gate (f_t), which decides what information should be forgotten, the input gate (i_t), which determines what information should be added, and the output gate (o_t), which decides what information should be passed to the next state. The weight matrices determine how much influence each gate has. Below, the processing steps of this system are explained by the LSTM equations, provided in Equation (3) and Equation (4) [20, 21].

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i) \tag{3}$$

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f) \tag{4}$$

$$c_t = f_t c_{t-1} + i_t tanh\,(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \tag{5}$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o) \tag{6}$$

$$h_t = o_t tanh(c_t) \tag{7}$$

During the training of the BiLSTM model, since the LSTM operates bidirectionally, the equations between Equation (3) and Equation (7) are executed in both forward and backward directions.

D.3. Gated Recurrent Unit

Gated Recurrent Unit (GRU) is a next-generation algorithm that uses RNN. GRU is faster than LSTM because it uses fewer tensor operations. It is very similar to the LSTM model. The difference with LSTM is that the cell state is removed from the model. Whereas in the LSTM model the cell state provides the information transfer, here the information transfer is done by the hidden state. Another difference in the model is that there are only two gates. These gates are the reset gate and the update gate [22]. The update gate is the gate that decides whether to add or discard new information. The Reset Gate is the gate that decides how much of the past data to forget. GRU structures perform fewer vector operations than other types of RNN due to their simpler gate mechanism. This makes GRUs faster in training and inference processes [24].

A certain t For time step t, input data $x_t \in R^{n*d}$ is defined as n is the number of samples, d is the number of features in each sample. The latent state of the previous time $h_{t-1} \in R^{n*h}$ where h represents the number of neurons in the hidden layer. The GRU architecture has two main components that regulate the flow of information: the update gate and the reset gate. Of these components, the update gate decides how much of the past information to preserve or update [24].

$$U_t = \sigma(W_U h_{t-1} + W_U x_t + b_U) \tag{8}$$

The calculation method of the update gate is expressed in equation (8). This gate acts as the component that decides to what extent the cell state should be updated. On the other hand, the reset gate plays an active role in determining how much of the past information should be forgotten [24].

$$R_t = \sigma(W_R h_{t-1} + W_R x_t + b_R) \tag{9}$$

The calculation process of the reset gate is described by equation (9). If the output of this gate takes a value close to 0, it means that the information of the previous time step is ignored in the current memory state, i.e. forgotten. Conversely, a value closer to 1 indicates that the previous information is retained in the current memory [25].

$$\widehat{h_t} = tanh(W.[R_t * h_{t-1}] + W_{x_t}) \tag{10}$$

The state of the memory content at time t is expressed by equation (10). After determining how much information to forget and how much to keep in memory through reset and update gates, this information is scaled by the activation function “tanh” [25]. Equation (11) is then used to calculate the amount of information stored in the hidden layer at time t.

$$h_t = (1 - U_t) * h_{t-1} + U_t * \widehat{h_t} \tag{11}$$

D.4. Bidirectional Gated Recurrent Unit

BiGRU is a structure created by combining two separate GRU layers. Thanks to this architecture, the model can obtain information from both past and future context at each time step [23]. The BiGRU processes inputs bidirectionally, with one processing the time sequence from past to future and the other from future to past. Thanks to this bidirectional approach, the context in both directions is preserved at each time step, resulting in a unified latent state. Thus, the model can perform more comprehensive learning by taking into account not only the information from previous steps, but also the information from subsequent steps [26].

E. EXPERIMENTAL SETTING

80%-20% training to test separation when experimenting on the dataset. The data was cleaned by removing links, user tags, hashtags, and special characters, and fully converted to lower case. Text Representation was performed using the TF-IDF, BERT method. The experiments were coded in Python using the scikit-learn library and run on the Google Colab platform.

F. PERFORMANCE METRICS

Within the scope of the study, Acc, F1, Rec, Pre performance criteria were used to evaluate the performance of the experiments. Calculation of these performance criteria [27].
When an evaluation is made based on the positive and negative sentiments in the texts,

True positive (Tp) is text whose sentiment class is predicted to be positive and is labelled positive. True Negative (Tn) is text that is predicted to have a negative sentiment class and is labelled as negative. False positive (Fp) is text whose sentiment class is predicted to be positive, but whose sentiment class is labelled negative. False Negative (Fn) is the text whose sentiment class is predicted to be negative but labelled as positive. These values are described in this section [27].

F.1. Accuracy

Acc is the ratio of correctly predicted to total dataset sum in classification models [27]. This metric is calculated according to equation (12) [27].

$$Acc = \frac{Tp + Tn}{Tp + Fp + Fn + Tn} \tag{12}$$

It alone is not sufficient to measure model performance, especially in datasets without a balanced class distribution. If the dataset is not evenly distributed, other performance measures should also be considered [28].

F.2. Precision

Pre is the measure of the predicted positive values being positive [28].

$$Pre \ = \frac{Tp}{Tp + Fp} \tag{13}$$

The Pre score is calculated as the ratio of Tp values that are positively labelled and positively predicted in the sentiment class to positive values that are always positively predicted as predictions, even though they are actually labelled positive or negative [28].

F.3. Recall

Rec is a measure of how many of the values predicted to be positive are predicted to be positive [26]. Equation (14) shows the Rec value [29].

$$Rec = \frac{Tp}{Tp + Fn} \tag{14}$$

F.4. F1

It is the harmonic mean of the other classification metrics, Pre and Rec. It is a balanced performance criterion as it is calculated with Rec and Pre criteria. F1 is given by Equation (15) [30].

$$F1 = 2 * \frac{Pre \ * Rec}{Pre \ + Rec} \tag{15}$$

This study uses the unbalanced Twitter dataset. In the subsections Experimental studies and results and Other studies conducted during the thesis, other performance metrics besides the Acc performance metric are included.

III. EXPERIMENTAL RESULT

This study was written on the Google Colab platform using Python. NB and RF algorithms are used to verify the performance of the TF-IDF method and classifiers on cryptocurrency related tweets. The results are based on the text representation techniques used, the Acc of the classifiers and the F1 scores. Table shows the performance results obtained after TF-IDF and vectorisation according to BERT.

Table 4. *Results of the performance of the class-weighted model according to text representation*

		Pre	**Rec**	**F1**	**Acc**
TF-IDF	DT	0.87	0.87	0.86	0.86
	NB	0.86	0.82	0.75	0.82
	SVM	0.86	0.70	0.73	0.87
	BiLSTM	0.91	0.88	0.89	0.90
	BiGRU	**0.92**	**0.89**	**0.90**	**0.91**
BERT	DT	0.89	0.88	0.87	0.88
	NB	0.88	0.84	0.81	0.83
	SVM	0.90	0.89	0.88	0.89
	BiLSTM	0.93	0.92	0.91	0.92
	BiGRU	**0.94**	**0.93**	**0.92**	**0.93**

As shown in Table 3, the classical ML models NB, DT, and SVM achieved 82%, 86%, and 87% Acc respectively with the TF-IDF representation. On the other hand, the DL based models BiLSTM and BiGRU achieved 90% and 91% Acc respectively with TF-IDF and were more successful than the classical methods.

When the BERT representation was introduced, performance generally improved. The NB, DT and SVM models achieved 83%, 88%, and 89% Acc respectively, while the BiLSTM and BiGRU models achieved the best results with 92% and 93% Acc respectively. Thanks to the better contextualisation of BERT, significant improvements in F1 were also observed, especially when used with DL models.

Overall, the BiGRU model performed best with both TF-IDF and BERT representations. This shows that BiGRU's strong learning ability on sequential data combined with BERT's context-based representations leads to superior performance. The results show that DL models outperform classical models and that advanced embedding methods such as BERT significantly improve performance.

IV. DISCUSSION AND CONCLUSION

In this study, we performed SA on 10,000 cryptocurrency-related tweets shared between 2020 and 2021. The tweets were labelled as positive, negative, or neutral using the TextBlob library. A range of ML and DL models were applied using both TF-IDF and BERT textual representations. Among all model combinations, the BiGRU model trained with BERT embeddings achieved the highest performance, with 93% Acc and a strong F1, highlighting the strength of combining context-aware language representations with sequential deep models. These results are summarized in Figure 2, which compares Acc and F1 across all models and feature representations.

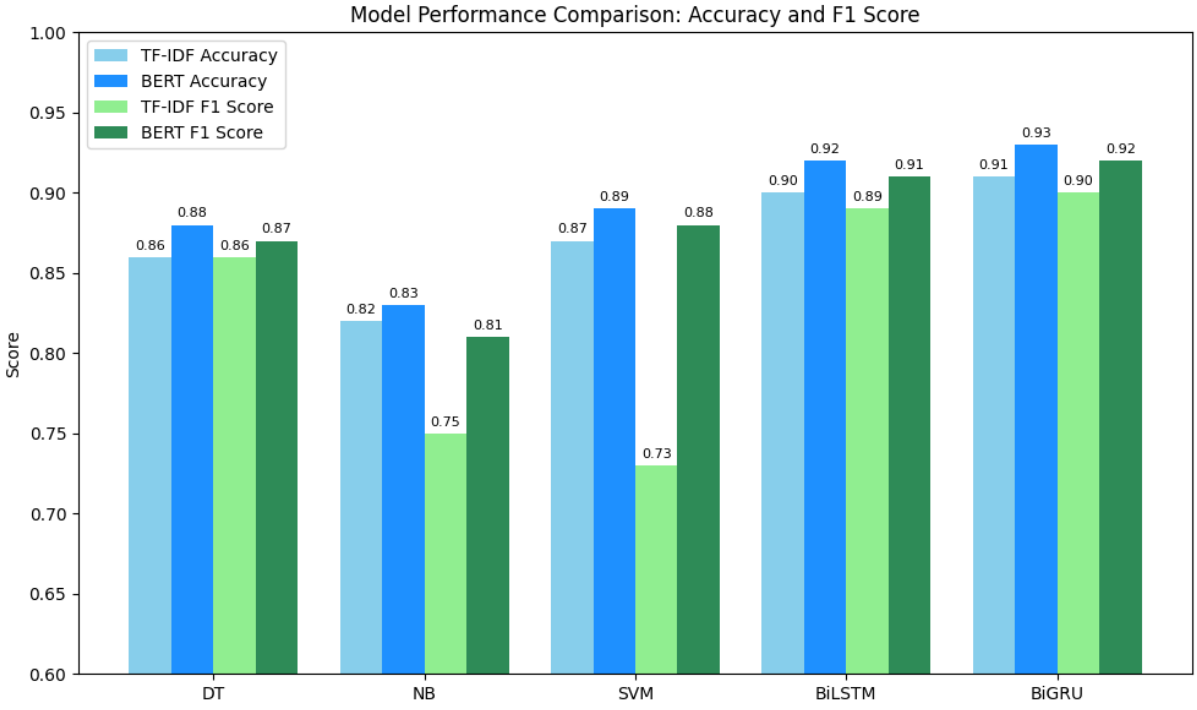

Figure 2. *class distribution.*

The findings indicate that while traditional models can offer moderate success in SA tasks, DL approaches using contextual embeddings (e.g., BERT) are significantly more effective—particularly for understanding nuanced textual data such as tweets. Moreover, the class imbalance in the dataset led to higher model performance for the neutral class, which is an important observation for future modelling strategies that seek to correct or compensate for class distribution biases.

When compared to the existing literature, our results demonstrate substantial improvements. For instance, Konstantinos et al. [9] reported an Acc of 86.7% and an F1-score of 87.3% using fine-tuned large language models for news-based cryptocurrency SA, which is lower than the 93% accuracy observed in this study. Similarly, Ramaputra et al. [11] achieved accuracy rates around 82–83% with TF-IDF and NB on user reviews, and Alam et al. [13] reported 85% accuracy using CNN+LSTM models. While Abraham et al. [15] highlighted the superiority of BERT over lexicon-based tools such as VADER in multi-platform SA (Twitter and Reddit), our findings reinforce this insight with concrete numerical superiority. These comparisons clearly suggest that BiGRU combined with BERT provides a new state-of-the-art baseline for tweet-based sentiment classification in the cryptocurrency domain.

Looking ahead, we aim to expand the dataset over a broader time range and improve label quality through manual annotation or crowdsourcing. In addition to TextBlob, other sentiment labelling tools such as VADER, Flair, and SentiWordNet will be used to enrich the sentiment labels. Moreover, the dataset will be tested with additional transformer-based models like RoBERTa, DistilBERT, and ALBERT, enabling a comprehensive comparative analysis of both labelling tools and model performances.

In conclusion, this study not only demonstrates the effectiveness of contextual language models in social media SA but also provides a roadmap for future research. It highlights which methodological combinations are most promising for extracting insights from noisy, short-form, user-generated text thus contributing to the broader goal of understanding market sentiment and social dynamics in the cryptocurrency space.

Article Information

Acknowledgments: The authors would like to express their sincere thanks to the editor and the anonymous reviewers for their helpful comments and suggestions.

Author’s Contributions: Preprocessing the dataset, data analysis, experiments and evaluations, manuscript draft preparation M.A. Defining the methodology, evaluations of the results and draft editing M.S.B, All authors have read and approved the final version of it.

Artificial Intelligence Statement: No any Artificial Intelligence tool is used in this paper.

Conflict of Interest Disclosure: No potential conflict of interest was declared by authors.

Plagiarism Statement: This article was scanned by the plagiarism program.

V. REFERENCES

[1] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," Bitcoin. Accessed: Apr. 9, 2025. [Online]. Available: https://bitcoin.org/bitcoin.pdf.

[2] J. Bollen, H. Mao and X. Zeng, “Twitter mood predicts the stock market,” *Journal of Computational Science*, vol. 2, no. 1, pp. 1–8, 2011.

[3] Y. Liu and A. Tsyvinski, “Risks and returns of cryptocurrency,” *The Review of Financial Studies*, vol. 34, no. 6, pp. 2689–2727, 2021.

[4] F. Mai, Z. Shan, Q. Bai, X. (Shane) Wang and R. H. L. Chiang, “How does social media impact bitcoin value? A test of the silent majority hypothesis,” *Journal of Management Information Systems*, vol. 35, no. 1, pp. 19–52, 2018.

[5] A.-U. Islam, "Crypto tweets," Accessed: Apr. 9, 2025. [Online]. Available: https://www.kaggle.com/datasets/leoth9/crypto-tweets

[6] G. Salton and C. Buckley, “Term-weighting approaches in automatic text retrieval,” *Information Processing & Management*, vol. 24, no. 5, pp. 513–523, 1988.

[7] J. R. Quinlan, “Induction of decision trees,” *Machine Learning*, vol. 1, no. 1, pp. 81–106, 1986.

[8] J. Devlin, M. W. Chang, K. Lee and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding*,”*in *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, vol. 1, Minneapolis, USA, 2019, pp. 4171–4186.

[9] K. I. Roumeliotis, N. D. Tselikas and D. K. Nasiopoulos, "LLMs and NLP models in cryptocurrency sentiment analysis: a comparative classification study," *Big Data and Cognitive Comput*uting, vol. 8, no. 6, 2024, Art. no. 63.

[10] K. Byc, S. C. Ilinca and R. R. Mukkamala, "The relationship between social media sentiment and Bitcoin price volatility," M.S. thesis, Copenhagen Bus. Sch., Copenhagen, Denmark, 2021. [Online]. Available: https://research.cbs.dk/files/71300950/1303955_Master_Thesis_Sep_2021.pdf

[11] C. A. Ramaputra, M. H. Z. Al Faroby and B. R. Lidiawaty, "Sentiment analysis of user reviews on cryptocurrency application: Evaluating the impact of dataset split scenarios using multinomial naive Bayes," *The Indonesian Journal of Computer Science*, vol. 13, no. 4, 2024.

[12] C. Mai, M. R. A. Khan, A. Nicholson and S. A. R. Abu-Bakar, "Using Sentiment Analysis to Predict Bitcoin Price," in *2018 4th International Conference on Computer and Information Sciences (ICCOINS)*, Kuala Lumpur, Malaysia, 2018, pp. 1–6.

[13] F. Alam, S. K. Joty and M. Imran, "Deep learning for sentiment analysis of social media texts," *Information Processing & Management*, vol. 57, no. 1, pp. 102-106, 2020.

[14] S. McNally, J. Roche and S. Caton, "Predicting the Price of Bitcoin Using Machine Learning," in *2018 26th Euromicro International Conference on Parallel, Distributed and Network-Based Processing (PDP)*, Cambridge, UK, 2018, pp. 339–343.

[15] A. Abraham, A. Zeng and J. Zhang, "Twitter and Reddit sentiment analysis for cryptocurrency price prediction," in *2021 IEEE International Conference on Big Data (Big Data)*, Orlando, FL, USA, 2021, pp. 4560–4565.

[16] S. N. Başa and M. S. Basarslan, “Sentiment analysis using machine learning techniques on IMDB dataset,” in *2023 7th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*, Ankara, Türkiye, 2023, pp. 1–5.

[17] A. Triyono and A. Faqih, “Implementation of the Naive Bayes Method in sentiment analysis of Spotify application reviews,” *Journal of Artificial Intelligence and Engineering Applications (JAIEA)*, vol. 4, no. 2, 1091-1097, 2025.

[18] N. Hussain, A. Qasim, G. Mehak, O. Kolesnikova, A. Gelbukh and G. Sidorov, "Hybrid machine learning and deep learning approaches for insult detection in Roman Urdu text," *AI*, vol. 6, no. 2, pp. 33, 2025.

[19] S. Gayathri, A. Chandar R.S., Rithicagash J. and Guna A. "Integrating fuzzy approach in text mining and summarization," in *2025 6th International Conference on Mobile Computing and Sustainable Informatics (ICMCSI),* Goathgaun, Nepal, 2025, pp. 98-102.

[20] S. Hochreiter and J. Schmidhuber, "Long Short-Term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[21] M. B. Çakı and M. S. Başarslan, "Classification of fake news using machine learning and deep learning*," Journal of Artificial Intelligence and Data Science*, vol. 4, no. 1, pp. 22–32, 2024.

[22] D. Chi, "Research on electricity consumption forecasting model based on wavelet transform and multi-layer LSTM model," *Energy Reports*, vol. 8, pp. 220–228, 2022.

[23] Y. Xiong, N. Wei, K. Qiao, Z. Li and Z. Li, "Exploring consumption intent in live e-commerce barrage: a text feature-based approach using BERT-BiLSTM model," *IEEE Access*, vol. 12, pp. 69288-69298, 2024.

[24] M. S. Başarslan, "M-C&M-BL: a novel classification model for brain tumor classification: multi-CNN and multi-BiLSTM," *The Journal of Supercomputing*, vol. 81, no. 3, 2025, Art. no. 502.

[25] Y. Çelik, "Bellek tabanlı LSTM ve GRU makine öğrenmesi algoritmaları kullanarak BIST100 endeks tahmini", *Fırat Üniversitesi Mühendislik Bilimleri Dergisi,* vol. 36, no. 2, pp. 553–561, 2024.

[26] M. S. Islam and N. A. Ghani, "A novel BiGRUBiLSTM model for multilevel sentiment analysis using deep neural network with BiGRU-BiLSTM," *Lecture Notes in Electrical Engineering*, vol. 730, pp. 403–414, 2022.

[27] J. Yan, J. Liu, Y. Yu and H. Xu, "Water quality prediction in the Luan river based on 1-DRCNN and BiGRU Hybrid Neural Network Model," *Water (Basel),* vol. 13, no. 9, 2021, Art. no. 1273.

[28] Z. Turgut and G. Akgün, "Occupancy and occupant number detection for energy saving in smart buildings via machine learning techniques," *International Journal of Exergy*, vol. 44, no. 3, pp. 204–226, 2024.

[29] M. U. Etli et al., "Evaluating deep learning techniques for detecting aneurysmal subarachnoid hemorrhage: A comparative analysis of convolutional neural network and transfer learning models," *World Neurosurg*, vol. 187, pp. e807–e813, 2024.

[30] F. Pedregosa et al., "Scikit-learn: machine learning in Python," *The Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.