Crypto social media analysis: Deep learning framework for Text Classification and Question Answering

Jay Vaghela[1], Siba Sankar Sahu[1]

[1]*Department of Computer Science and Engineering, Sardar Vallabhbhai National Institute of Technology, Surat, India*

Abstract

In the current era, people frequently use social media platforms such as Facebook, Twitter, and Reddit to express their opinions. The opinion contains cryptocurrency-related content such as posts, tweets, and discussions. However, user-generated content contains frequent informal discussions. As part of the FIRE 2025 CryptoQA shared task, we investigate different deep learning models, where the goal is classification and question answering of cryptocurrency-related social media content. In this study, our team SVNIT_CSE investigated different deep learning models for text classification and Q&A. Among the models evaluated, the GRU architecture provides an average Macro-F_1 score of 0.52 on the Reddit, Twitter, and YouTube datasets. Similarly, GRU with Fasttext embedding provides optimal performance and offers a Macro-F_1 score of 0.75 in Q&A. The deep learning models evaluated provide competitive performance in cryptocurrency social media classification and Q&A.

Keywords

Cryptocurrency, multiclass classification, opinion extraction, question answering, social media

1. Introduction

The advent of cryptocurrencies has revolutionized financial systems and introduced decentralized digital assets that work independently of old banking infrastructures. The rapid rise of cryptocurrencies has transformed global financial landscapes, creating opportunities and uncertainties for investors, researchers, and policymakers. The cryptocurrency ecosystem is heavily shaped by user-generated content on social media platforms such as Twitter, Reddit, and YouTube. The presence of on-line content provides a good source of information, but it also presents challenges. Social media texts are often noisy, unstructured, and context-dependent, making it a challenging task to extract meaningful opinions and reasonable information. Posts may include slang, abbreviations, emojis, sarcasm, or multimedia elements, complicating automated analysis. In addition, content can range from factual reports and objective analyses to highly subjective opinions, promotional spam, or outright misinformation, necessitating sophisticated methods to discern valuable insights.

In the rapid growth of digitization, the cryptocurrency data on social networks have grown tremendously. Initially, the research relied on basic word matching techniques, using tools such as VADER or SentiWordNet to tag posts as positive or negative. These approaches were simple and struggled to understand context, informal language, and crypto-specific terms. So, few researchers explored machine learning techniques to overcome limitations. They used different machine learning models such as support vector machine (SVM), naive bayes, or random forest [9]. These models mainly focused on text features, such as word frequency or short phrases, to classify sentiments or predict market trends. These models are effective in some cases; however, the model provides poor performance in handling noisy and informal data. The deep learning model provides better performance by capturing semantic meaning of sentences and makes the model more suitable for sentiment analysis in social media data.

In this study, we explore various deep learning models, including RNN, LSTM and GRU along with different embedding techniques like Word2Vec [2], Glove [3] and FastText [7] to classify a cryptocurrency post on social media into various categories. In addition, the model addresses various challenges by focusing on opinion extraction and question answering from cryptocurrency-related social media

content. We found that the proposed deep learning models improve the performance in cryptocurrency text classification and question answering.

2. Problem Statement

FIRE[1] (Forum for Information Retrieval Evaluation) organised a shared task called CryptoQA[2] on text classification and question answering from cryptocurrency-related social media posts [1]. The shared task includes two different tasks. The objective of task 1 is to classify the social media post into multiple classes. Similarly, the objective of task 2 is the question-answering from the QnA dataset, given a set of question-comments pair, the task is to identify the relevance of the comment with respect to a question. Together, these tasks aim to benchmark and advance the development of robust natural language processing (NLP) systems that can understand, classify, and retrieve meaningful information from noisy large-scale cryptocurrency data. Building an efficient NLP model can be directly implemented for financial decision-making, market research, and the development of reliable information retrieval systems in the domain of cryptocurrencies.

The remainder of the paper is organized as follows. In Section 3 we present the state-of-the-art work in the domain of cryptocurrency text classification and question & answering. Section 4 describes the statistics of the data set used in the experiment. Section 5 presents the methods used in the text classification and QA system. The experimental results and their analysis are presented in Section 6. Finally, we conclude with the direction of future work in Section 7.

3. Related Work

The analysis of social media content for cryptocurrency insights has gained significant attention as researchers seek to understand public sentiment, predict market trends, and extract actionable knowledge from different social media platforms such as YouTube, Twitter, and Reddit. In this section, we present some of the existing work on opinion extraction and question answering in the context of social networks.

Bakliwal et al.[1] focused on the analysis of political sentiment on Twitter, specifically during the Irish general elections. They classify tweets as positive, negative, or neutral sentiment. The authors explored lexicon-based approaches and supervised machine learning methods. They found that lexicon-based methods improved performance. However, the support vector machine provides the best effectiveness. Rajadesingan & Liu [4] introduced ReLP, a semi-supervised framework for detecting user opinions on Twitter data. The framework combines a retweet-based label propagation algorithm with a supervised classifier. ReLP demonstrates that leveraging retweet cooccurrence patterns for semi-supervised label propagation can drastically reduce labeling effort while outperforming traditional supervised and clustering baselines in identifying opposing stances on contentious political debates.

Balikas, Moura & Amini et al. [6] addressed the fine-grained Twitter sentiment classification by leveraging the relationship between multiple sentiment analysis tasks. The author proposed a multitask learning framework based on a bidirectional LSTM (biLSTM) network, which jointly learns both tasks to improve classification performance. Bagdori & Oard [5] explored cross-platform question answering by linking unanswered questions on Twitter to potential answers from Yahoo! answers. They demonstrated that a significant portion of twitter questions could be successfully answered using Yahoo! answers. Askari et al. [11] introduced LegalQA, a new benchmark dataset for the legal domain. The data set comprises 9,846 questions and 33,670 lawyer-verified answers. The author addressed two challenges: the knowledge gap between lawyers and common people and the mixed formal or informal style of legal QA forums. Nguyen et al. [8], introduces BERTweet, the first large-scale pre-trained language model specifically designed for English Tweets. Built with the same architecture as BERTbase but trained using the RoBERTa procedure, BERTweet addresses the challenges of informal and noisy Twitter text.

---

[1]https://fire.irsi.org.in/fire/2025/home
[2]https://sites.google.com/view/cryptoqa-2025/task-description

It was pre-trained on a massive corpus of 850M Tweets. The model outperformed RoBERTabase and XLM-Rbase in all tasks.

Roumeliotis, Tselikas & Nasiopoulos [13] explored the capabilities of different large language models (LLMs) to analyze the sentiment of cryptocurrency-related news articles. They fine-tuned models such as GPT-4, BERT, and FinBERT. All models achieved strong performance with the Adam optimizer, with GPT-4 showing the best results, closely followed by FinBERT and BERT. In another similar approach employed by Kulakowski & Frasincar [10] focuses on developing two new approaches to the sentiment analysis of cryptocurrency-related social media posts. CryptoBERT - a Twitter oriented fined-tuned BERTweet model improved performance, and the Language-Universal Cryptocurrency Emoji (LUKE) sentiment lexicon, using Support Vector Machine, also enhanced performance. Deroy & Maity [12] investigated prompt-based learning with GPT-4 Turbo using zero-shot prompting for text classification and few-shot learning for Q&A. The model improved performance in different text classification and Q&A domain. Sarkar et al. [14] created a multilevel dataset CryptOpiQA with inclusion of gold standard (manually annotated) and silver standard (inferred from the gold standard) labels. SVM with BERT and GPT-3.5 Turbo improve both the performance of Twitter and Reddit platforms, and GPT-3.5 offers optimal results.

From the above study, we found that the various methods have been explored in the cryptocurrency-related social media data. In this study, we investigated GRU-based models with FastText and Word2Vec embeddings to address text classification and question answering in the social media domain.

4. Dataset

The CryptoQA shared task is part of the FIRE (Forum for Information Retrieval Evaluation) evaluation campaign comprising two subtasks. Task 1 is about text classification, and task 2 is about question and answer. In task 1, three training data sets were provided from different social media platforms such as Reddit, Twitter, and YouTube. Each data set consists of posts related to cryptocurrency and annotated at following three hierarchical levels. The statistics of the training and testing data set are shown in Tables 1 and 11. In task-2, a question and answering (Q&A) dataset was provided. The statistics of Q&A dataset is shown in Table 6. The Q&A dataset comprises queries corresponding to their comments as columns. Moreover, the training data comments have a binary relevance. The statistic of relevance and non-relevance comments is shown in Table 7.

- Level 1: 0: NOISE, 1: OBJECTIVE, 2: SUBJECTIVE
- Level 2: Only for SUBJECTIVE posts → 0: NEUTRAL, 1: NEGATIVE, 2: POSITIVE
- Level 3: Only for NEUTRAL posts → 0: NEUTRAL SENTIMENT, 1: QUESTION, 2: ADVERTISEMENT, 3: MISCELLANEOUS

The class distribution for the three training datasets is mentioned in Tables 3, 4 and 5.

Table 1
Statistics of training dataset

Platform	Total samples
Reddit	5000
Twitter	4288
YouTube	5000

5. Methodology

We implement three steps to perform text classification and question-answering tasks on social media. The three steps include data pre-processing, model design, and evaluation. In the preprocessing step,

Table 2

Statistics of testing dataset

Platform	Total samples
Reddit	500
Twitter	500
YouTube	500

Table 3

Data distribution for text classification in Reddit

Level	Class	Total Samples
1	Noise	645
	Objective	503
	Subjective	3852
2	Neutral	3183
	Negative	410
	Positive	259
3	Neutral Sentiments	476
	Questions	2390
	Advertisements	105
	Miscellaneous	212

Table 4

Data distribution for text classification in Twitter

Level	Class	Total Samples
1	Noise	1338
	Objective	1702
	Subjective	1248
2	Neutral	902
	Negative	78
	Positive	270
3	Neutral Sentiments	178
	Questions	136
	Advertisements	544
	Miscellaneous	46

Table 5

Data distribution for text classification in YouTube

Level	Class	Total Samples
1	Noise	786
	Objective	32
	Subjective	4182
2	Neutral	2401
	Negative	1574
	Positive	207
3	Neutral Sentiments	1391
	Questions	1000
	Advertisements	1
	Miscellaneous	9

the raw text is tokenized, segmented, and presented in a structured format. The raw text data are cleaned and transformed into numerical sequences in the following ways.

Table 6

Statistics of QnA dataset

Type	Total samples
Training	25369
Testing	6323

Table 7

QnA training Dataset Statistics

Target	Total samples
Relevant	3704
Not Relevant	21531

- Data cleaning: The raw data contains missing values, null values, and the data set is imbalanced. Hence, we need necessary cleaning steps to provide consistency in the data. We remove missing and null value data and balance the data set by implementing the class weights [3] method.
- Tokenization: A Keras tokenizer [4] was used to convert text into numerical sequences, limiting the vocabulary to 10,000 words, and handling out-of-vocabulary words with a special token. Texts were tokenized according to word frequency, and sequences were padded to a fixed length of 100 tokens to ensure uniform input for models.

In the preprocessing step, the data set is cleaned and presented in a structured format. We used deep learning models such as LSTM and GRU for text classification. Moreover, the QA model is designed using deep learning methods with modern embedding-based approaches. Finally, we evaluate the text classification and QA system using standard evaluation metrics.

5.1. Opinion classification

We built a unified pipeline model that classified social media content into different categories. In text classification, we used long-short-term memory (LSTM) unit[5] and gated recurrent unit (GRU)[6]. A single cell LSTM and GRU unit is shown in Figs. 1 and 2. The architecture of the LSTM unit comprises two layers. First, the embedding layer with 128 dimensional dense vectors then two LSTM layers with 64 neurons and tanh activation function each, both having dropout of 0.3 to reduce overfitting. Ending with a dense layer of 32 neurons with tanh activation and an output layer comprising the softmax activation function. The GRU architecture contained three layers. First, the embedding layer with 128 dimensional dense vectors then three GRU layers with 128 neurons and tanh activation function each, and uses dropout rate of 0.3 to reduce overfitting. Ending with a dense layer of 32 neurons with tanh activation and an output layer comprising the softmax activation function. The hyperparameter used in the training process is shown in Table 8.

5.2. Question Answering

The main objective of the Q&A is to check if the comments for the corresponding question in the data set are relevant or not. The Q&A model comprises RNN, LSTM, and GRU in conjunction with TF-IDF [7], Word2Vec, GloVe and FastText embeddings. The hyperparameters used in the training process are shown in table 9. To effectively represent textual data in the cryptocurrency Q&A task, distributed word embeddings were employed as word embeddings effectively keep semantic and syntactic relationships

[3]https://scikit-learn.org/stable/modules/generated/sklearn.utils.class_weight.compute_class_weight.html

[4]https://www.tensorflow.org/api_docs/python/tf/keras/preprocessing/text/Tokenizer

[5]https://shorturl.at/JyzA8

[6]https://shorturl.at/kZSwP

[7]https://www.learndatasci.com/glossary/tf-idf-term-frequency-inverse-document-frequency/
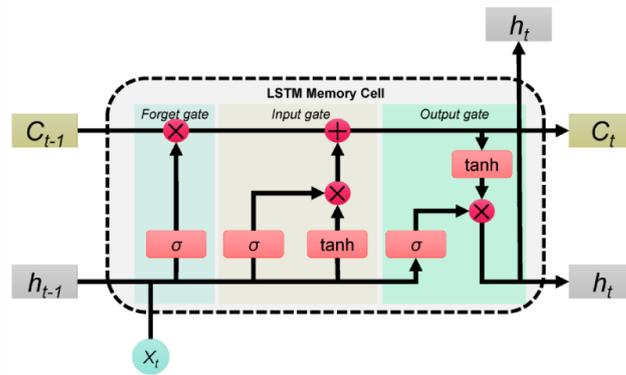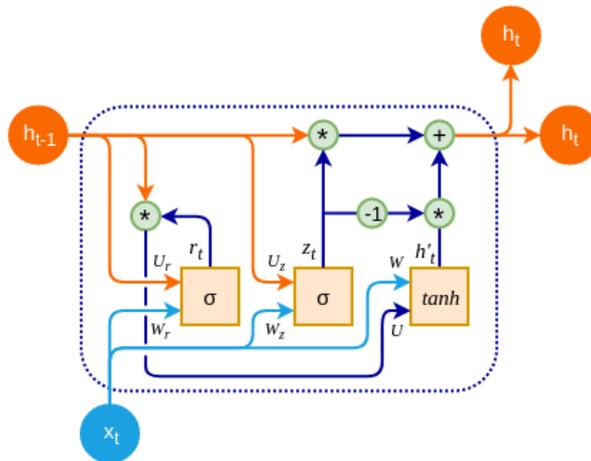
Figure 1: A single Cell LSTM Architecture

Figure 2: A single Cell GRU Architecture

Table 8
Hyperparameters used in text classification

Hyperparameter	Value
MAX_WORDS	10000
VOCAB_SIZE	10000
MAX_LEN	100
batch_size	64
epochs	20

between words, so models such as LSTM and GRU to better understand the underlying contextual meaning in social media posts.

- Word2Vec Embedding: It is a widely used word embedding technique that uses a fixed-dimension dense vector to represent words. It has two main architectures, that is, the continuous bag-of-words (CBOW) and skip-gram approach. In the CBOW approach, it predicts a word based on context and the skip-gram approach predicts the surrounding word provided a target word. Word2Vec capture the semantically similar words by putting words with similar meanings closer together in vector space. It helps the model understand the domain-specific relationships between

Table 9
Hyperparameters used in Q&A

Hyperparameter	Value
embedding_dim	300
MAX_LEN	100
batch_size	64
epochs	15

words that frequently co-occur in cryptocurrency Q&A task.

- FastText Embedding: It is introduced by Facebook, extends the Word2Vec model by representing each word as a bag of character n-grams. Instead of learning a single vector for each word, the model learns embeddings for character substrings, and the final representation of words is derived from the sum of its n-gram vectors. The model generates embeddings for out-of-vocabulary (OOV) words, which is highly useful for noisy social media data (misspellings, hashtags, slang, abbreviations, etc.). Cryptocurrency discussions often contain informal expressions like holding presented as 'hodl', bitcoin presented as 'btc', or ethereum as 'eth2.0'. FastText's subword modeling captures these variations and makes more robust for domain-specific contexts. The representation of words in FastText and Word2Vec [8] is shown in Figure 3.
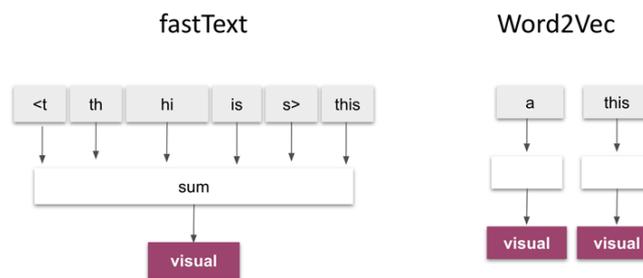

Figure 3: Representations of words in FastText and Word2Vec

6. Results & Discussion

In this study, we explore various deep learning models and embedding techniques for multiclass classification and question-answering for social media texts. To evaluate the effectiveness of the models, we used accuracy and macro-F_1 score. The evaluation was carried out on a masked test set of Reddit (500 instances), Twitter (500 instances), Youtube (500 instances), and Q&A (6323 instances). Table 10 presents the models submitted for text classification (Task 1) and Q&A (Task 2). The performance of the submitted models on different test data is shown in Table 11. From the evaluation results, we found that the GRU architecture and GRU with FastText embedding techniques provides optimal performance and outperformed other models in text classification and question answering. The evaluated model provides competitive performance in both text classification and Q&A task.

In text classification, we found that the LSTM model provides the best performance on YouTube (Macro-F_1 = 0.69) and Reddit (Macro-F_1 = 0.47), while the GRU architecture offers the best performance on Twitter (Macro-F_1 = 0.49) and attained the best average score (Macro-F_1 = 0.52). A similar observation was also found in the validation data set, as shown in Table 12. From the evaluation results, we found that both the models, i.e., LSTM and GRU provide a competitive score in text classification.

---

[8]https://shorturl.at/TPYp5

Table 10
Best performing submitted model

RUN	Task	Model
1	Task-1	LSTM
	Task-2	GRU with FastText
2	Task-1	GRU
	Task-2	GRU with Word2Vec

Table 11
Evaluation score on submitted models

Task	Dataset	RUN	Model	Macro-F_1	Rank
1	Reddit	run1	LSTM	0.47	7/8
	Twitter	run2	GRU	0.49	6/8
	YouTube	run1	LSTM	0.69	6/8
	Average	run2	GRU	0.52	7/8
2	Q&A	run1	GRU with FastText	0.75	5/7

In Q&A, the GRU architecture with FastText embeddings achieved the highest Macro-F_1 score (0.75) on the test set, as shown in Table 11. GRU with FastText embeddings emerged as the most balanced configuration, highlighting the importance of combining advanced recurrent models with subword-level embedding methods for handling cryptocurrency-related social media text. A similar observation was also found in the validation data set, as shown in Table 13. From the validation results, we found that word embedding models provide better accuracy than traditional TF-IDF based methods. The primary reason is that the word embedding model captures contextual meaning between words in the embedding space. Different word embedding models provide a similar performance in different social media data. In general, the findings indicate that the word embedding method with a deep learning architecture provides optimal performance in Q&A domain.

According to the shared task evaluation policy, we can submit up to two runs for each task. Hence, we evaluated the remaining models by ourselves based on the accuracy evaluation metric. Table 12 shows the accuracy achieved by different models in the validation data on text classification. Similarly, Table 13 shows the accuracy achieved by different embeddings on the RNN, LSTM and GRU models in the validation data on Q&A. The effectiveness of different models is presented graphically in Figures 4 and 5.
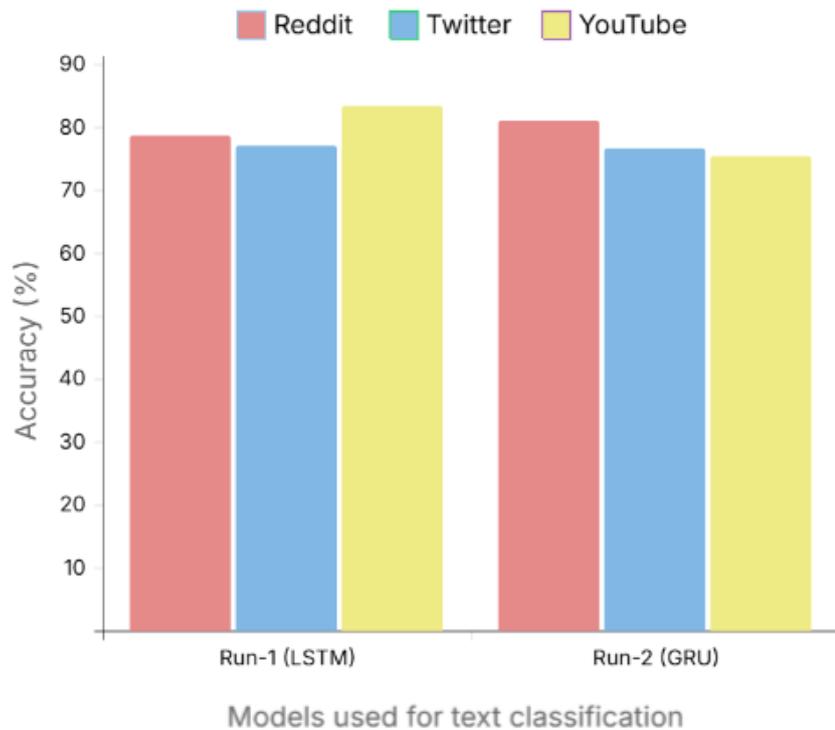
Table 12
Evaluation score in validation data for text classification

Model	Dataset	Accuracy
LSTM	Reddit	0.786
	Twitter	0.770
	YouTube	0.833
GRU	Reddit	0.810
	Twitter	0.766
	YouTube	0.753

Table 13

Evaluation score in validation data for Q&A

Embedding	Model	Accuracy
	LSTM	0.86
FastText	GRU	0.88
	RNN	0.71
	LSTM	0.85
Word2Vec	GRU	0.88
	RNN	0.86
	LSTM	0.84
GloVe	GRU	0.83
	RNN	0.77
	LSTM	0.72
TF-IDF	GRU	0.76
	RNN	0.62

Figure 4: Accuracy achieved by different models on validation data set

7. Conclusion

Text classification and Q&A is an important downstream task in the text analysis domain. In the cryptocurrency-related data, it presents unique challenges due to the noisy nature of user-generated content. In this study, we explore different deep learning models with embedding techniques to handle noisy content on social networks. From the evaluation results, we found that deep learning models are effective at capturing the nuances of crypto-related discourse in both classification and QA tasks. Models such as LSTM & GRU provide similar performance in opinion classification. GRU with the FastText model provides optimal performance and outperforms other models in Q&A. Although deep learning models provide promising results, several challenges remain. The performance of the models varied considerably in different datasets and evaluation metrics, indicating that no single architecture provides
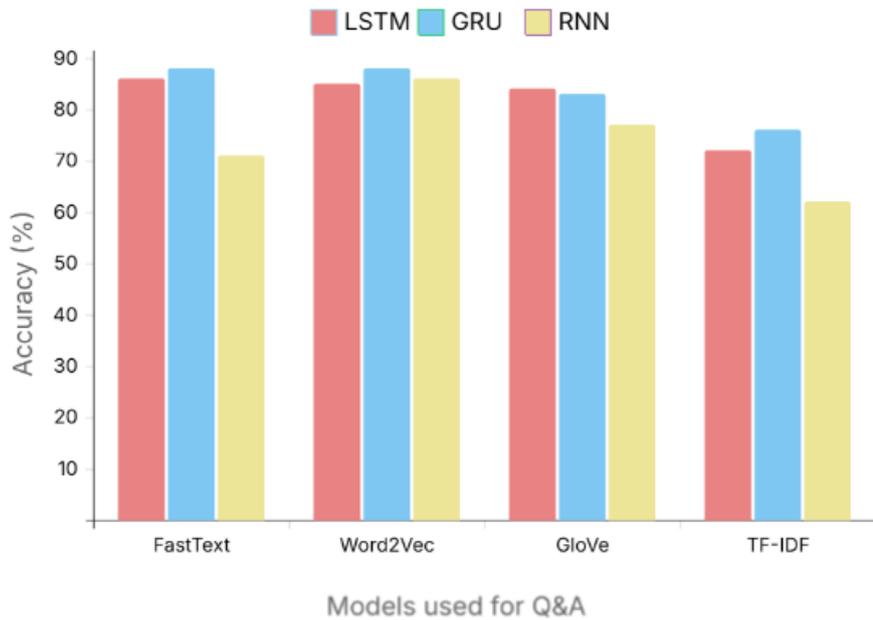
Figure 5: Accuracy achieved by different models on validation data set

optimal performance in all datasets. In the future, we can explore the applicability of transformer models in cryptocurrency-related data.

Declaration on Generative AI

During the preparation of this work, the author(s) used ChatGPT, QuillBot in order to: Spelling Check, Paraphrase and reword. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication's content.

References

[1] A. Bakliwal, J. Foster, J. van der Puil, R. O'Brien, L. Tounsi, and M. Hughes, "Sentiment analysis of political tweets: Towards an accurate classifier," in *Proceedings of the 2013 Conference of the Association for Computational Linguistics*, 2013.

[2] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *arXiv preprint arXiv:1301.3781*, 2013.

[3] J. Pennington, R. Socher, and C. D. Manning, "Glove: Global vectors for word representation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014, pp. 1532–1543.

[4] A. Rajadesingan and H. Liu, "Identifying users with opposing opinions in twitter debates," in *International conference on social computing, behavioral-cultural modeling, and prediction*, Springer, 2014, pp. 153–160.

[5] M. Bagdouri and D. W. Oard, "Building bridges across social platforms: Answering twitter questions with yahoo! answers," in *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2017, pp. 1181–1184.

[6] G. Balikas, S. Moura, and M.-R. Amini, "Multitask learning for fine-grained twitter sentiment analysis," in *Proceedings of the 40th International ACM SIGIR conference on research and development in information retrieval*, 2017, pp. 1005–1008.

[7] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, "Enriching word vectors with subword information," *Transactions of the Association for Computational Linguistics*, vol. 5, pp. 135–146, 2017.

[8] D. Q. Nguyen, T. Vu, and A. T. Nguyen, "Bertweet: A pre-trained language model for english tweets," *EMNLP 2020*, p. 9, 2020.

[9] H. Pabuçcu, S. Ongan, and A. Ongan, "Forecasting the movements of bitcoin prices: An application of machine learning algorithms," *Quantitative Finance and Economics*, vol. 4, no. 4, pp. 679–692, 2020.

[10] M. Kulakowski and F. Frasincar, "Sentiment classification of cryptocurrency-related social media posts," *IEEE Intelligent Systems*, vol. 38, no. 4, pp. 5–9, 2023.

[11] A. Askari, Z. Yang, Z. Ren, and S. Verberne, "Answer retrieval in legal community question answering," in *European Conference on Information Retrieval*, Springer, 2024, pp. 477–485.

[12] A. Deroy and S. Maity, "Cryptollm: Unleashing the power of prompted llms for smartqna and classification of crypto posts," *arXiv preprint arXiv:2411.07917*, 2024.

[13] 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 Computing*, vol. 8, no. 6, p. 63, 2024.

[14] S. Sarkar et al., "Cryptopiqa: A new opinion and question answering dataset on cryptocurrency," in *Proceedings of the 31st International Conference on Computational Linguistics*, 2025.