Cryptocurrency Analysis: Price Prediction of Cryptocurrency Using User Sentiments and Quantitative Data

Dayan Perera, Jessica Lim, Shuta Gunraku and Wern Han Lim
School of Information Technology, Monash University Malaysia, Malaysia

Keywords: Cryptocurrency, Price Prediction, User-Generated Content (UGC), Long Short-Term Memory (LSTM), Recurrent Neural Network (RNN), Gated Recurrent Unit (GRU), Bidirectional-LSTM, Deep Learning.

Abstract: This research introduces an innovative approach to forecasting cryptocurrency prices by combining user-generated content (UGC) and sentiment analysis with quantitative data. The primary goal is to overcome limitations in existing methods for market forecasting, where accurate forecasting is crucial for informed decision-making and risk mitigation. The paper suggests a robust prediction methodology by integrating sentiment analysis and quantitative data. The study reviews prior research on sentiment analysis and quantitative analysis of cryptocurrency and stock price prediction. It explores the integration of machine learning and deep learning techniques, an area not extensively explored before. The methodology employs Long Short-Term Memory (LSTM), Recurrent Neural Network (RNN), Bidirectional LSTM and Gated Recurrent Unit (GRU) models to capture temporal dependencies. Prediction accuracy is assessed using metrics including Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and a confusion matrix. Results show that GRU models excel in prediction, while RNN models outperform in predicting price movements; with an emphasis on the significance of a suitable data preprocessing pipeline towards improving model performance. In summary, this study demonstrates the effectiveness of integrating sentiment analysis and quantitative data for cryptocurrency price forecasting using UGC data.

1 INTRODUCTION

Cryptocurrencies have disrupted the financial landscape, ushering in a new era of digital assets that captivate investors and traders worldwide. Cryptocurrencies recorded peak total market capitalization at USD 2,953 billion in November 2021[1]. As these decentralised digital currencies gain popularity and are under the observation of regulators, predicting their prices accurately becomes crucial for making informed investment decisions and optimising trading strategies.

The volatility [Mužić and Gržeta(2022)] and unpredictability of the cryptocurrency market [Boukhers et al.(2023)] pose formidable challenges to analysts and investors alike. Current approaches for traditional fiat currencies or stock markets [Tang and Chen(2018)] including quantitative analysis and news article reactions on their own struggled to capture the dynamics of these digital assets. Users do not consume traditional news of cryptocurrencies much while discussions were found to be prevalent on social media [Beck et al.(2019)].

In this paper, we delve into a novel methodology that integrates UGC sentiment analysis with quantitative data to overcome the limitations of existing prediction methods. By combining (1) sentiment analysis, which reflects the emotions and opinions of market participants; and (2) with quantitative data representing market fundamentals and price patterns – we seek to create a more holistic, accurate and robust prediction model. The synergistic effects of these two distinct information sources can lead to enhanced predictions, better risk assessment, and improved decision-making for investors and traders. Given the unstructured and noisy data, this research also propose a data preprocessing pipeline.

This paper is structured as follows – section 2 provides a succinct overview of relevant works and their findings, contextualising the purpose of this study. In section 3, we outline our research objectives and elaborate on the approach taken, complemented by exploratory data analysis (EDA) on a collected dataset. Subsequently, section 4 elucidates the experimental

[1]As reported by Statista https://www.statista.com/statistics/730876/cryptocurrency-maket-value/.

Perera, D., Lim, J., Gunraku, S. and Lim, W.
Cryptocurrency Analysis: Price Prediction of Cryptocurrency Using User Sentiments and Quantitative Data.
DOI: 10.5220/0012315100003636

In *Proceedings of the 16th International Conference on Agents and Artificial Intelligence (ICAART 2024) - Volume 3*, pages 210-217
ISBN: 978-989-758-680-4; ISSN: 2184-433X

setup and rationale. The results and discussion are presented in section 5, critically evaluating the outcomes of the experiment. Finally, section 6 presents a cohesive summary of key findings, achievements, and implications for future research.

2 RELATED WORKS

Most research on cryptocurrency and stock price prediction can be categorised into one of two main approaches: (1) market sentiment analysis and (2) quantitative analysis. Another notable point of contrast involves the comparison between machine learning and deep learning techniques; however, only a few papers have explored the combination of both approaches. In this systematic literature review, we organise the sections to form an integral part of the system architecture to provide a comprehensive overview of the relevant research.

2.1 Price Prediction Models

Various methods have been employed to forecast cryptocurrency values, including Logistic Regression and LSTM models [Ammer and Aldhyani(2022)]. With the advancements in machine learning, particularly the deep learning models of today, attention has shifted towards the use of such technologies to build complex predictive models. One prominent technology in this domain is a recurrent neural network (RNN) variant known as the Long Short-Term Memory (LSTM) [Hochreiter and Schmidhuber(1997)]. LSTM is commonly used due to its effectiveness in learning from long-term data, overcoming the vanishing gradient problem found in traditional RNNs. Furthermore, it has been discovered to excel in forecasting price alterations [Armin et al.(2022)]. As a result, LSTM is widely used to train predictive models, often in conjunction with other approaches [Bin Mohd Sabri et al.(2022)] such as logistic regressions [Ammer and Aldhyani(2022)], Ridge regression [Armin et al.(2022)] and ARIMAX [Serafini et al.(2020)]. Many of these approaches utilise LSTM, with various architectures, particularly in the dropout layers. Despite cryptocurrency prices being more volatile than stock prices, especially in the absence of fixed trading windows or strict regulations [Pervaiz et al.(2020)], an extension of LSTM known as Stochastic LSTM can effectively account for the randomness and fluctuations in prices [Jay et al.(2020)].

2.2 Predictive Model Features

Various research studies make use of a wide variety of features when attempting to predict stock or cryptocurrency prices. Traditionally, financial attributes such as the opening price, peak price, number of transactions, and other related financial indicators are commonly employed as features to train predictive models [Awoke et al.(2020)]. The internet has made information easily accessible, leading to the utilisation of new sources like Google Trends [Pervaiz et al.(2020)] in predictive models. Cryptocurrency-specific attributes, such as blockchain data, can also be incorporated as features in the prediction process, in addition to the traditional financial attributes [Ji et al.(2019)].

The advent of social media has emerged as a significant catalyst, generating a substantial amount of commotion through the prolific creation of consumable content on platforms like Twitter [Jay et al.(2020)]. This phenomenon can serve as a feature to assess the influence of social media on both predicting and driving price movements especially when generated by key opinion leaders (KOL) [Jiang(2022)].

It is important to acknowledge the consistent findings across various studies, revealing a correlation between stock prices and public sentiment expressed on both traditional and social media platforms [Smith and O'Hare(2022)]. Consequently, this research extends prior investigations to integrate public sentiments from social media [Sattarov et al.(2020)] as features for price prediction models. This extension is particularly pertinent for cryptocurrencies, characterised by high volatility and inherent difficulty in prediction. The extraction of such sentiments will be facilitated through the utilisation of state-of-the-art sentiment analysers as detailed in subsection 2.3.

2.3 Sentiment Analysis on Social Media Content

Sentiment analysis is a field that has been thoroughly researched and has its own set of established approaches, including a variety of highly effective lexicon-based models [Adwan et al.(2020)]. These advancements have led to the widespread use of popular pre-trained models such as VADER [Hutto and Gilbert(2014)], enabling swift sentiment analysis computation without compromising accuracy [Ibrahim(2021), Sattarov et al.(2020), Mohapatra et al.(2019)]. It is worth noting that the inclusion of sentiment analysis [Smith and O'Hare(2022)] has the potential to enhance the models' prediction performance of market movements. Since the primary focus of this research is not on enhancing sentiment analysis itself, we will utilise existing pre-trained models for extracting sentiment features.

A challenge in extracting sentiments from social media is the unstructured nature of UGC on such platforms [Sasmaz and Tek(2021)]. Data preprocessing is often necessary, which may include techniques like stemming and removal of stop words. [Ibrahim(2021)]. Moreover, platform-specific additions need to be handled with care such as the use of hashtags as annotations on Twitter [Sasmaz and Tek(2021)] or mentions creates a complex network of content on the platform. Thus, researchers such as Sebestião H et al. [Sebastião and Godinho(2021)] have employed varying statistical methods, including the Dickey-Fuller test, to perform enhanced data preprocessing.

3 METHODOLOGY

Based on the literature review discussed in section 2, it was hypothesised that sentiment analysis on social media content could serve as a reliable predictor for cryptocurrency price trends. However, there is a need for additional concrete data regarding the implementation and performance of sentiment analysis on social media content within the dataset. To confirm this hypothesis, two-tailed t-tests were conducted to compare the means of two groups, and simple linear regression was employed to evaluate relationships between continuous variables.

Figure 1 visualised the correlations observed between the variables on the collected datasets outlined in subsection 4.1 [2]. Past research indicates that sentiment analysis can offer valuable insights into market sentiment and its impact on cryptocurrency prices, as discussed in the related works section. However, sentiments alone may not be sufficient, as the findings from Figure 1 reveal a positive relation yet weak correlation (-0.12) between sentiments and open or close price. Similarly, a weak potential (0.037) was observed in the volume variable for predicting the Bitcoin prices. Thus in our research, both volumes and sentiments have been employed in Bitcoin prediction with the aim of enhancing and maximising the accuracy of our predictions.

Consequently, a new approach is proposed for our experiment, combining sentiment analysis and volume to predict cryptocurrency trends. Equation 1 describes the output gate of LSTM layers used in our primary algorithms designed for sequence prediction. These models are engineered to process input sequences and generate predictions by leveraging patterns and dependencies within the data, utilising their internal states. The LSTM model, equipped with its specialised memory cell and gating mechanisms, excels at capturing long-term dependencies in sequences, making it particularly effective for modelling intricate temporal relationships [Sak et al.(2014)]. We posit that its consideration of long-term data contributes to improved predictions, aligning with our hypothesis.

$$o_t = \sigma(W_{xo} \cdot x_t + W_{tho} \cdot th_{t-1} + b_o) \tag{1}$$

Conversely, the RNN model relies on recurrent connections to propagate information across time steps and formulate predictions based on both current and previous inputs. This characteristic is advantageous when predicting price trends based on Volume [Valendin et al.(2022)]. Additionally, the RNN is well-suited for price prediction when employing sentiment analysis, as it emphasizes the use of current data. The GRU and the Bidirectional-LSTM are variants of the RNN and LSTM that perform more efficiently and make use of other tricks to improve performance. Therefore our approach incorporates RNN, LSTM, Bidirectional-LSTM and GRU models optimised through a basic Grid Search. Table 1 outlines our best-performing models and their associated parameters.

4 EXPERIMENT SETUP

4.1 Datasets

The data used in this study was obtained from two primary sources: Twitter and Kaggle datasets. Specifically, our data training approach encompassed the time period from '05/02/2021 10:00:00' to '05/10/2021 23:00:00,' marked by a significant surge in demand and the growing popularity of cryptocurrencies. We selected this timeframe with the expectation that it would provide a wealth of data and relevant variables for our research. The combination of these two sources resulted in a substantial dataset, comprising more than 5799 observations. To maintain consistency and coherence in the data, we conducted a series of pre-processing steps to align their temporal aspects. Given the diversity of data sources, this process involved thorough data cleaning.

The Twitter data provided valuable insights into sentiment dynamics, while the Kaggle dataset offered

[2] The dataset mainly retrieved and sourced from https://www.kaggle.com/datasets/ilariamazzoli/3-million-tweets-cryptocurrencies-btc-eth-bnb

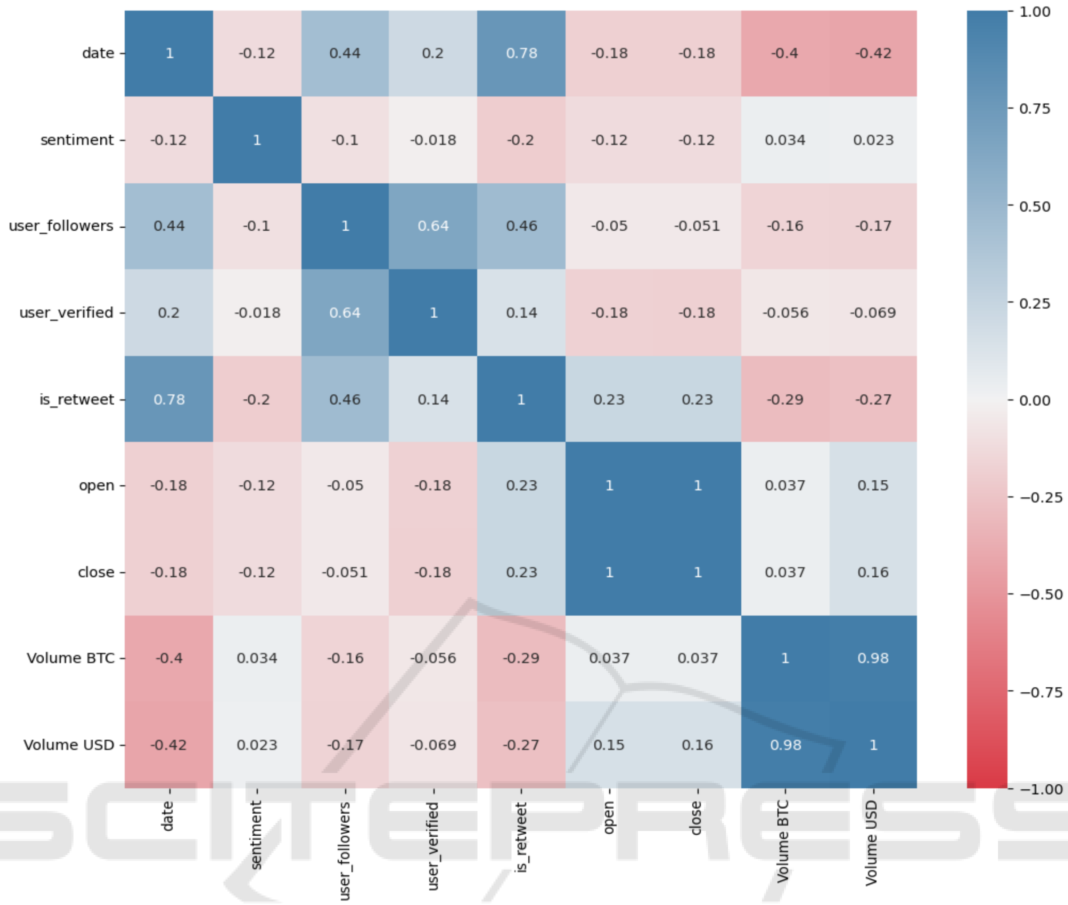


Figure 1: Correlation between Variables retrieved from Twitter and Kaggle.

information on cryptocurrency pricing trends. We focused on Twitter, motivated by its potential as a platform housing reliable sources and extensive discussions about cryptocurrencies, thereby enhancing the value of our sentiment analysis. Employing a well-structured pipeline, we utilised lemmatisation models to systematically process the content of tweets, ultimately generating sentiment scores for individual tweets. The resulting dataset, which formed the foundation for subsequent training and testing, underwent rigorous model training and testing procedures, constituting the core of our analytical endeavours. The dataset consisted of a total of nine variables, as indicated in Table 2.

4.2 Data Cleaning

A proper data splitting method needed to be utilised as simply using Scikit-learn's TrainTestSplit led to erroneous results, as the method randomised the data. This is incorrect given the sequential ordering nature of the time series data. Moreover, handling missing values required careful consideration due to the significant number of data points with missing values. Using a conventional imputer proved ineffective, as it imputed the same value for all missing values in the column. This approach is not suitable for time-series datasets, which are prone to high variance. Therefore, this research suggested the use of iterative imputer to be used.

4.3 Data Scaling

Scaling is crucial, especially when the range of values in columns differs. If left uncorrected, this discrepancy can result in some variables having a disproportionately greater impact on the results simply because they have larger values. Scaling brings all the variables to a similar range, allowing their true effect on the results to be observed. Both Min-Max Normalisation and Standardisation were utilised, and the results for each are presented in Table 3 for comparison.

Table 1: Best Performing Models as Identified through Grid Search.

LSTM	LSTM (1st)	Units: 64, Return Sequences: True, Activation: tanh
	LSTM (2nd)	Units: 64, Return Sequences: True, Activation: tanh
	LSTM (3rd)	Units: 64, Activation: tanh
	Dense	Units: 1
RNN	RNN (1st)	Units: 64, Return Sequences: True, Activation: tanh
	RNN (2nd)	Units: 64, Return Sequences: True, Activation: relu
	RNN (3rd)	Units: 64, Activation: tanh
	Dense	Units: 1
GRU	GRU (1st)	Units: 64, Return Sequences: True, Activation: tanh
	GRU (2nd)	Units: 64, Return Sequences: True, Activation: relu
	GRU (3rd)	Units: 64, Activation: relu
	Dense	Units: 1
Bi-directional LSTM	Bi-directional (1st)	Units: 64, Return Sequences: True, Activation: relu
	Bi-directional (2nd)	Units: 64, Return Sequences: True, Activation: relu
	Bi-directional (3rd)	Units: 64, Activation: relu
	Dense	Units: 1

4.4 Evaluation Measures

The models are evaluated using the following metrics: Root Mean Squared Error(RMSE), Mean Squared Error(MSE) and a confusion matrix. This is achieved through a row-by-row comparison, with the value of 0 indicating a price increase and the value of 1 indicating a decrease. Meanwhile, RMSE and MSE are utilised to assess the actual prices themselves. The confusion matrix serves as a visual performance assessment of the classification algorithm, evaluating how well the model predicts the price changes [Ibrahim(2021)]. RMSE and MSE were chosen for their effectiveness with regression-type data.

Table 2: Datasets Training Variables.

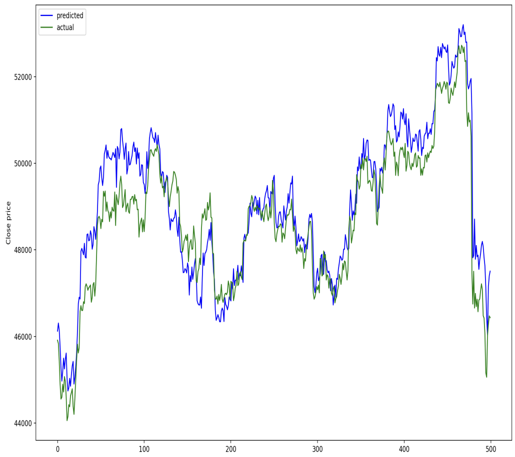
Variable	**Description**
Date	Date and time of the data point (e.g., 2021-02-05 10:00:00)
Sentiment	Sentiment score (range: 0 to 1)
User Followers	Number of followers of the user
User Verified	Whether the user is verified (0 or 1)
Is Retweet	Whether the data point is a retweet (0 or 1)
Open	Opening price of the financial instrument
Close	Closing price of the financial instrument
Volume BTC	Trading volume in Bitcoin (BTC)
Volume USD	Trading volume in U.S. dollars (USD)

5 RESULTS AND ANALYSIS

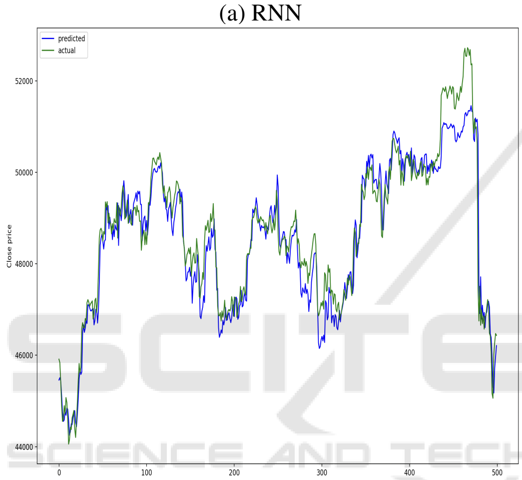
We conducted experiments on four models: LSTM, GRU, RNN, and Bidirectional-LSTM, evaluating their performance through a combination of error measures, a confusion matrix and graphs. The results are displayed in Table 3, indicating that the GRU performed the best when considering both normalisation and standardisation with normalisation performing outperforming standardisation. This superiority can be attributed to the GRU's computational efficiency compared to the other models, along with its ability to better remember short-term data, which is crucial for predicting cryptocurrency prices greatly influenced by short-term events as well as a reactionary market on social network sentiment. The visualisations of the actual and predicted values over time by RNN and GRU are illustrated in Figure 2, showcasing GRU's behaviour to be less volatile.

Table 3: RMSE values for RNN and LSTM variants. The best-performing result is in **bold**.

Model	Scaling	RMSE	MSE
LSTM	Standardisation	1866	3482464
LSTM	Normalisation	1214	1472836
RNN	Standardisation	1185	1403514
RNN	Normalisation	1099	1207855
Bi-LSTM	Standardisation	1239	1534512
Bi-LSTM	Normalisation	1020	1041181
GRU	Standardisation	937	877861
GRU	**Normalisation**	**659**	**433958**



(a) RNN



(b) GRU

Figure 2: Price prediction for RNN and GRU. Blue for the predicted value and green for the actual value.

The preference for normalisation arises from the fact that standardisation assumes a Gaussian distribution in the data. Additionally, our dataset lacks extreme outliers, as Bitcoin prices in a short time window generally fall within a small range. This condition favours normalisation, as standardisation typically excels when dealing with datasets containing extreme outliers.

Figure 3 depicts experiments performed to determine the optimal window size, where window size refers to the number of time steps (hours) the models predict in advance. The number of time steps tested ranged from 1 to 24 hours, and the results revealed that the best-performing window size varied depending on the model. Interestingly, the window sizes of 24 and 17 appeared twice each in the optimal configuration of models. This variability in optimal window sizes is attributed to the models encoding different amounts of long-term and short-term information.

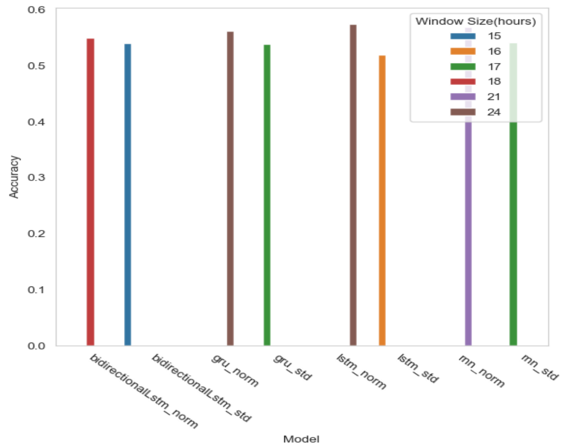


Figure 3: Prediction accuracy for price increase or decrease for each model. Only the best performing window size and accuracy is shown due to space constraint.

In Figure 4, the confusion matrix for the optimal models, identified by Table 3 and using a window size of 17, is depicted. It is evident that when predicting the price changes based on the polarity of prices, all four models exhibit similar performance, with RNN standing out in accurately predicting price increases. This success can be attributed to the RNN's reliance on short-term memory, aligning well with the nature of cryptocurrency prices that are predominantly influenced by short-term events.

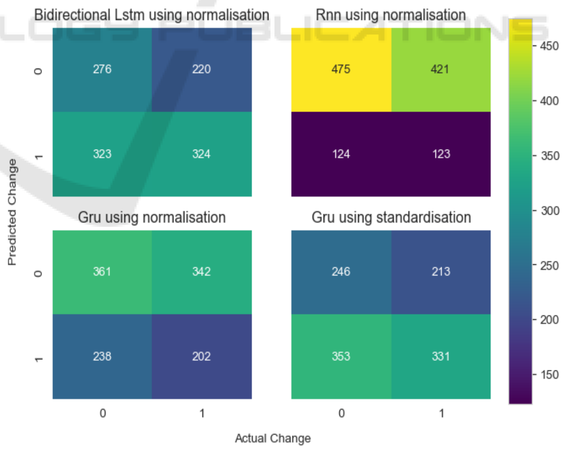


Figure 4: Confusion matrices for the price change prediction (increase or decrease) of optimal models using a window size of 17.

This study also investigates the time of day when the model performs the best. This categorical variable is derived from the timestamps, corresponding to specific segments of the day in Coordinated Universal Time (UTC), spanning from morning to night. Significantly, the 'Afternoon' category emerges as the most accurate, indicating a notable surge in data volume during this time frame. This effectiveness can be hypothesised to stem from its alignment with morning hours in US time zones, particularly significant markets for BTC, where heightened trading activity and increased Twitter engagement are prevalent.

6 CONCLUSIONS

This paper has presented the findings and outcomes aimed at developing a predictive system for analysing price trends of the highly volatile cryptocurrencies such as Bitcoin using user sentiment from Twitter as a popular User-Generated Content (UGC) platform for discussion. The UGC dataset was generated from the scraping of Twitter; temporal-mapped to the cryptocurrency data from Kaggle. To do so, this paper explores and optimises four models – Long Short-Term Memory (LSTM), Recurrent Neural Network (RNN), bidirectional LSTM (bi-LSTM), and Gated Recurrent Unit (GRU) for the task. The accuracy and reliability of the predictions were then enhanced through machine learning models and appropriate evaluation techniques.

GRU is the best-performing model based on Root Mean Squared Error (RMSE), followed by Bi-LSTM. This is due to its capabilities in remembering short-term events. As such, the findings supported the hypothesis for public sentiment as a price prediction feature. Besides that, the models were found to best predict 17 to 24 hours in advance where the global market does react slower despite the volatile nature of cryptocurrency – thus investors are patient with a tendency to hold and observe further, or it can be interpreted as slow reactors to public sentiment on UGC.

As a future work, we aim to explore the inclusion of other UGC platforms and their sentiments to build a more robust model. If more micro-economic data is to be obtained, we would also like to explore smaller temporal windows for price prediction for a more sensitive model especially when there are anomalies in the market such as during a rug pull.

REFERENCES

Omar Yousef Adwan, Marwan Al-Tawil, Ammar Huneiti, Rawan Shahin, Abeer Abu Zayed, and Razan Al-Dibsi. 2020. Twitter sentiment analysis approaches: A survey. *International Journal of Emerging Technologies in Learning (iJET)* 15, 15 (2020), 79. https://doi.org/10.3991/ijet.v15i15.14467

Mohammed Abdullah Ammer and Theyazn H. Aldhyani. 2022. Deep learning algorithm to predict cryptocurrency fluctuation prices: Increasing investment awareness. *Electronics* 11, 15 (2022), 2349. https://doi.org/10.3390/electronics11152349

Atieh Armin, Ali Shiri, and Behnam Bahrak. 2022. Comparison of machine learning methods for cryptocurrency price prediction. *2022 8th Iranian Conference on Signal Processing and Intelligent Systems (ICSPIS)* (2022). https://doi.org/10.1109/icspis56952.2022.10043898

Temesgen Awoke, Minakhi Rout, Lipika Mohanty, and Suresh Chandra Satapathy. 2020. Bitcoin price prediction and analysis using Deep Learning Models. *Communication Software and Networks* (2020), 631–640. https://doi.org/10.1007/978-981-15-5397-4_63

Johannes Beck, Roberta Huang, David Lindner, Tian Guo, Zhang Ce, Dirk Helbing, and Nino Antulov-Fantulin. 2019. Sensing Social Media Signals for Cryptocurrency News. In *Companion Proceedings of The 2019 World Wide Web Conference* (San Francisco, USA) *(WWW '19)*. Association for Computing Machinery, New York, NY, USA, 1051–1054. https://doi.org/10.1145/3308560.3316706

Muhammad Husaini Bin Mohd Sabri, Amgad Muneer, and Shakirah Mohd Taib. 2022. Cryptocurrency price prediction using long short-term memory and Twitter sentiment analysis. *2022 6th International Conference On Computing, Communication, Control And Automation (ICCUBEA* (2022). https://doi.org/10.1109/iccubea54992.2022.10011090

Zeyd Boukhers, Azeddine Bouabdallah, Cong Yang, and Jan Jürjens. 2023. Beyond Trading Data: The Hidden Influence of Public Awareness and Interest on Cryptocurrency Volatility. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management* (Birmingham, United Kingdom) *(CIKM '23)*. Association for Computing Machinery, New York, NY, USA, 142–151. https://doi.org/10.1145/3583780.3614790

Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural Computation* 9, 8 (1997), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

C. Hutto and Eric Gilbert. 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the International AAAI Conference on Web and Social Media* 8, 1 (2014), 216–225. https://doi.org/10.1609/icwsm.v8i1.14550

Ahmed Ibrahim. 2021. Forecasting the early market movement in bitcoin using Twitter's sentiment analysis: An ensemble-based prediction model. *2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS)* (2021). https://doi.org/10.1109/iemtronics52119.2021.9422647

Patel Jay, Vasu Kalariya, Pushpendra Parmar, Sudeep Tanwar, Neeraj Kumar, and Mamoun Alazab. 2020. Stochastic Neural Networks for cryptocurrency price prediction. *IEEE Access* 8 (2020), 82804–82818. https://doi.org/10.1109/access.2020.2990659

Suhwan Ji, Jongmin Kim, and Hyeonseung Im. 2019. A comparative study of bitcoin price prediction using

Deep Learning. *Mathematics* 7, 10 (2019), 898. https://doi.org/10.3390/math7100898

Sihao Jiang. 2022. The Relationship between the Cryptocurrency and the KOL of Elon Musk. In *Proceedings of the 2022 13th International Conference on E-Education, E-Business, E-Management, and E-Learning* (Tokyo, Japan) *(IC4E '22)*. Association for Computing Machinery, New York, NY, USA, 454–458. https://doi.org/10.1145/3514262.3514341

Shubhankar Mohapatra, Nauman Ahmed, and Paulo Alencar. 2019. Kryptooracle: A real-time cryptocurrency price prediction platform using Twitter sentiments. *2019 IEEE International Conference on Big Data (Big Data)* (2019). https://doi.org/10.1109/bigdata47090.2019.9006554

Ivan Mužić and Ivan Gržeta. 2022. Expectations of Macroeconomic News Announcements: Bitcoin vs. Traditional Assets. *Risks* 10, 6 (2022). https://doi.org/10.3390/risks10060123

Faizaan Pervaiz, Christopher Goh, Ashley Pennington, Samuel Holt, James West, and Shaun Ng. 2020. Fear and volatility in Digital assets. *SSRN Electronic Journal* (2020). https://doi.org/10.2139/ssrn.3721304

H. Sak, Andrew Senior, and F. Beaufays. 2014. Long short-term memory recurrent neural network architectures for large scale acoustic modeling. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH* (01 2014), 338–342.

Emre Sasmaz and F. Boray Tek. 2021. Tweet sentiment analysis for cryptocurrencies. *2021 6th International Conference on Computer Science and Engineering (UBMK)* (2021). https://doi.org/10.1109/ubmk52708.2021.9558914

Otabek Sattarov, Heung Seok Jeon, Ryumduck Oh, and Jun Dong Lee. 2020. Forecasting bitcoin price fluctuation by Twitter sentiment analysis. *2020 International Conference on Information Science and Communications Technologies (ICISCT)* (2020). https://doi.org/10.1109/icisct50599.2020.9351527

Helder Sebastião and Pedro Godinho. 2021. Forecasting and trading cryptocurrencies with machine learning under changing market conditions. *Financial Innovation* 7, 1 (2021). https://doi.org/10.1186/s40854-020-00217-x

Giulia Serafini, Ping Yi, Qingquan Zhang, Marco Brambilla, Jiayue Wang, Yiwei Hu, and Beibei Li. 2020. Sentiment-driven price prediction of the Bitcoin based on statistical and deep learning approaches. *2020 International Joint Conference on Neural Networks (IJCNN)* (2020). https://doi.org/10.1109/ijcnn48605.2020.9206704

Stephen Smith and Anthony O'Hare. 2022. Comparing traditional news and social media with stock price movements; which comes first, the news or the price change? *Journal of Big Data* 9, 1 (2022). https://doi.org/10.1186/s40537-022-00591-6

Jinqi Tang and Xiong Chen. 2018. Stock Market Prediction Based on Historic Prices and News Titles. In *Proceedings of the 2018 International Conference on Machine Learning Technologies* (Jinan, China) *(ICMLT '18)*. Association for Computing Machinery, New York, NY, USA, 29–34. https://doi.org/10.1145/3231884.3231887

Jan Valendin, Thomas Reutterer, Michael Platzer, and Klaudius Kalcher. 2022. Customer base analysis with recurrent neural networks. *International Journal of Research in Marketing* 39, 4 (2022), 988–1018. https://doi.org/10.1016/j.ijresmar.2022.02.007