International Journal of Applied Information Systems (IJAIS) – ISSN: 2249-0868
Foundation of Computer Science FCS, New York, USA
Volume 12– No. 43, April 2024– www.ijais.org

Bitcoin Price Prediction using Tweet Sentiment and User Interaction Behaviour

Qinan Zhu
University of Manchester
Manchester
United Kingdom

Rotimi Ogunsakin
University of Manchester
Manchester
United Kingdom

Laud Charles Ochei
University of Port Harcourt
Port Harcourt
Nigeria

ABSTRACT
Sentiment analysis has been used to predict Bitcoin prices, with results indicating relatively low prediction accuracy. User interaction behaviour such as likes, retweets, and replies have received little to no consideration as potential price prediction signals. Consequently, this paper uses both user interaction behaviour and Twitter sentiment to predict Bitcoin closing prices using multiple linear regression and Extreme Gradient Boosting (XGBoost). The predictive model outcomes are investigated using regression analysis and Shapley Additive Explanations (SHAP). Our findings indicate that using both sentiment score and user interaction behaviours significantly improves prediction accuracy, mostly during price volatility, but fails to capture Bitcoin price movement trends. Analysis of feature importance and impact on prediction outcome reveals sentiment score and user interaction behaviour play a lesser role in Bitcoin price prediction during price volatility. However, when Bitcoin prices are relatively stable, the improved accuracy is primarily due to the incorporation of user interaction behaviour. Therefore, the sentiment score is insufficient because the majority of Bitcoin-related tweets come from Bitcoin enthusiasts whose opinions are unaffected by market fluctuations. Whereas, the improved prediction accuracy observed during Bitcoin's price volatility is attributable to increased interactions from new sets of users.

Keywords
Sentiment analysis, bitcoin price prediction, explainable AI, Twitter sentiment

1. INTRODUCTION
The cryptocurrency was introduced in 2008 as a potential alternative to central bank-issued currency [25]. Since their introduction, cryptocurrencies have increased in both public recognition and value over time, not only due to their growing popularity but also because of the enormous potential returns generated by the investment activity associated with them. According to one of the most recent securities reports from Bank of America, a $1 investment in Bitcoin made at the beginning of 2010 is worth approximately $90,000 at the end of 2020 [32].

Despite the high returns, investment in cryptocurrencies comes with high risk due to their volatile prices. This is due to the absence of global regulation, as cryptocurrencies are yet to be recognised as a mature asset class. As a result, institutional investors with social and financial influence can disrupt the entire market without serious legal consequences until recently [34]. Consequently, in the absence of legal safeguards comparable to those in place for other securities, the prices of cryptocurrencies are susceptible to manipulation, posing a significant risk to those who choose to hold them. The over-enthusiasm caused by the news and reporting about the unprecedented return of cryptocurrencies also contributes to its volatility, leading to a ”Gold Rush”-like influx of new entrants into cryptocurrency trading [19].

Despite the fact that the volatility of cryptocurrency prices makes it difficult to identify price indicators and predict the price, there is still significant academic interest in the subject. With the support of a vast body of research on stock price prediction, many researchers believe that it is possible to predict the Bitcoin price and identify price indicators by employing techniques similar to those used for stock price prediction.

Similar to traditional financial markets, research in Bitcoin price prediction has shown that a relationship exists between human sentiment and Bitcoin price [16]. Moreover, the cryptocurrency market is still in its infancy. Therefore, traditional media outlets are unable to report cryptocurrency-related news in a timely manner compared to messages and news on social media. Thus the new media have become the primary source of cryptocurrency reporting and price volatility.

Consequently, this research aims to forecast the daily closing price of Bitcoin using Twitter sentiment and user interaction behaviour, such as retweets, replies, and likes [2], [10]. Twitter is one of the most popular platforms where people express and share opinions that indicate their feelings about specific topics. Therefore, Bitcoin-related Twitter posts, also known as ”tweets,” are gathered for sentiment analysis. In addition, the number of retweets, replies, and likes for each post are collected as potential Bitcoin price indicators that measure the impact of Twitter user interaction on price dispersion.

Twitter data from the past three years was extracted for sentiment analysis. The contents of posts are cleaned by removing unnecessary text, such as URLs and retweet handles. A lexicon-based method known as VADER (Valence Aware Dictionary and Sentiment Reasoner) [11] is used to assign sentiment polarity scores to tweets, while multiple linear regression and XGBoost are used for model development [6]. The results indicate that the user interaction behaviour has strong predictive power, whereas the sentiment score alone contributes little to the prediction outcome. Using both the sentiment score and user interaction behaviour significantly improves prediction accuracy but fails to capture Bitcoin price movement trends. However, using regression analysis and the Shapley Additive Explanations (SHAP) [21] value to analyse feature importance and impact on prediction outcomes shows sentiment score, and user interaction behaviour plays a lesser role in predicting the Bitcoin price during periods of price volatility. But when Bitcoin prices are relatively stable, however, improved accuracy is primarily due to the incorporation of user interaction behaviour. We conclude that the sentiment score is insufficient because the majority of Bitcoin-related tweets come from Bitcoin enthusiasts whose opinions are unaffected by market fluctuations, and the improved prediction accuracy observed during Bitcoin's price volatility is solely attributable to an increase in the total number of interactions from new sets of users and not the direction of the interaction that signifies user behaviour.

The remainder of the paper is organised as follows. Section 2 contains a review of research in the financial market, blockchain, bitcoin and bitcoin prediction. Section 3 presents the research methodology, which includes the data collection and analysis methods, and analysis and modelling of user behaviour and bitcoin prices. The result and interpretation of the model result are presented in Section 4. While the discussion and conclusion are presented in Sections 4 and 5, respectively.

2. LITERATURE REVIEW

2.1 Financial Markets and Public Sentiment

Human emotions are believed to influence financial activities and decisions in addition to monetary value [16]. Therefore, numerous studies have been conducted to investigate the viability of using human sentiment to forecast financial market activities and fluctuations.

A relationship has been established between investor sentiment and stock returns, where public fear and emotion are significant indicators of returns and the volatility index established as a good indicator of public sentiment [36]. Sometimes, investors and other market participants rely on the media for forecasts and decisions. Consequently, media sentiment has been studied alongside market trading volume and found to influence market prices. A high level of media pessimism has been established to exert downward pressure on market prices, while an abnormally high or low level of media pessimism explains temporarily high market trading volume [38]. In addition, [4] assessed the effect of social media-measured public sentiment on the Dow Jones Industrial Average (DJIA) and identified seven dimensions of public mood where a 3–4 day time lag is observed between some mood dimensions and the value of DJLA.

New research has focused on the collective public mood on Twitter as tweet sentiment has been shown to reflect public opinion on a broad range of issues and correlate with real-world emotions [7], [29]. [31] compared the sentiment extracted from a large number of Tweets about a specific company or index to its short-term market performance. After assigning positive or negative sentiments to tweets, the results indicate a strong cause-and-effect relationship between the negative and positive sentiment dimensions and the price movements of individual stocks. Similar conclusions can be drawn from [28] research, which demonstrated that Twitter sentiment indicates the trend of stock price movement for individual companies.

In addition to investigating the predictive power of tweet sentiment alone, some research incorporates additional variables to enhance the overall predictive power of Twitter sentiment. [43] combined hourly tweet volume and sentiment to forecast hourly stock prices. However, the addition of tweet volume did not improve the accuracy of predictions. A possible explanation is that tweets with a company's keywords do not necessarily contain opinions about the company. [30] narrowed the scope even further to investigate the relationship between tweet sentiment and stock return during Twitter's peak volume. During the peak period, the results indicate a significant relationship between sentiment polarity and the direction of cumulative abnormal returns. [37] accounted for the impact of the number of followers of Twitter users sending the messages. They concluded that the sentiment of tweets posted by users with a larger number of followers had a greater impact on same-day stock returns, but that this emotional impact was short-lived and quickly incorporated into the stock price.

2.2 Bitcoin and Price Prediction

Satoshi Nakamoto introduced Bitcoin in 2008, and it's primarily regarded as a revolutionary digital currency that solves the problems associated with constructing a secure and robust digital currency system [25]. Currently, numerous investors purchase Bitcoins as financial instruments and resell them speculatively for profit. Due to the high level of speculation and the absence of intrinsic value, the bitcoin market is irrational, making it difficult to predict the Bitcoin price [35]. In addition, the high volume of speculative transactions on the Bitcoin market adds volatility to the price and makes it difficult for investors to predict the price movement with precision [41].

A relatively large amount of research has been conducted to forecast the Bitcoin price using multiple predictive and machine-learning models. [22] compared the performance of binominal logistic regression, support vector machine, and random forest algorithms after selecting 16 features to forecast Bitcoin price at various intervals. The results indicate that random forest outperforms other algorithms over an extended period of time. [23] tested the predictive power of the ARIMA model for the Bitcoin price, as it was widely utilised in price prediction problems [7]. They discovered that ARIMA was less accurate than the deep learning model. [15] compiled 13 Bitcoin demand and supply-related features into a Bayesian neural network (BNN) and other deep learning algorithms for Bitcoin price forecasting. They discovered that BNN described Bitcoin's log price and volatility more accurately than other methods.

In recent times, some researchers have paid special attention to the correlation between Bitcoin and human emotion, and a number of studies have revealed and supported this correlation. [?] discovered that Google Trends, which reflects the realtime public interest in a topic, is a reliable indicator of the Bitcoin price movement. This discovery supports the opinion that polarity and emotional valence adequately explain Bitcoin price fluctuations [9]. The impact of tweet signals on Bitcoin exchange rates has been investigated by [8]. They concluded that the accuracy of exchange rate predictions improved after the sentiment score was included as input data, suggesting that tweet sentiment data may be utilised further in developing Bitcoin trading strategies. Furthermore, the scope of conventional sentiment analysis was broadened by [18]. Where they analyse Bitcoin forum comments using historical data spanning 2.8 years and obtained a higher level of Bitcoin price and transaction volume prediction accuracy, indicating that the proposed sentiment source is promising.

Various sentiment extraction and modelling techniques, ranging from multiple linear regression to advanced machine learning, have been implemented for cryptocurrency price prediction in order to effectively utilise the predictive power of sentiment. Using multiple linear regression models, [14] predicted the 2-hour Bitcoin price. Using Textblob sentiment polarity, they extracted sentiments from tweets tagged with ”Bitcoin” and ”BTC” and categorised each tweet as either positive, neutral, or negative. The number of tweets within each class is used as an input variable in the multiple linear regression model. The model’s performance suggested that sentiment is not the only factor influencing Bitcoin’s price. It is dependent on other variables, including political factors and mining costs.

The Extreme Gradient Boosting Regression Tree Model (XGBoost) was used to investigate alternative methods for predicting the prices of cryptocurrencies [20]. Tweets are not only scored based on sentiment but are further quantified using the dispersion of original tweets and the generation of secondary sentiment caused by retweets. The p-value of the final prediction indicates statistically significant results, suggesting that the consideration of secondary influence from retweets using the XGBoost model could result in a more accurate prediction of cryptocurrency prices. The primary concern of this study is the failure to exclude Twitter bot accounts that retweet promoted content with little influence on end users.

A real-time platform for predicting the price of cryptocurrencies based on Twitter sentiment score and historical price was developed by [24]. VADER (Valence Aware Dictionary and Sentiment Reasoner), a lexicon-based approach widely used to determine the sentiment score of social media messages, was used to collect and score Cyrpto-related Tweets, while each tweet’s composite score is multiplied by the poster’s number of followers and the number of likes and retweets to reflect its impact on end users. Historical prices were chosen as a time series characteristic to improve the accuracy of predictions. The results demonstrate the feasibility of developing a real-time platform capable of accurately predicting the Bitcoin price one minute in advance.

[5] compared Na¨ıve Bayes (Bernoulli and Multinominal) and a logistic regression model to predict the hourly and daily movement of the Bitcoin price. Text-processing.com API is applied to each tweet to return a vector containing the positivity, negativity, and neutrality sentiment scores, while the Bitcoin price movement is binomially classified as an increase (1) or decrease (0). The model’s outcome shows that logistic regression outperforms Na¨ıve Bayes, achieving a daily accuracy of 86.00% and an hourly accuracy of 98.58%. [27] examined whether certain Twitter accounts are more influential than others in terms of their ability to predict the Bitcoin price. They compared the accuracy of prediction between a dataset comprised of all Bitcoin-related Tweets and a dataset comprised of Tweets from the top 50 cryptocurrencies’ Twitter accounts. The findings indicate that the most influential Twitter accounts are the ones that drive returns, whereas other Twitter accounts merely contribute noise and volatility. This could imply that investors only need to monitor the most influential accounts to obtain necessary information for analysis.

Extensive research has validated the predictive power of sentiment, but some studies have cast doubt on the applicability of such methods under different circumstances and the causal relationship between sentiment and Bitcoin price. [17] conducted a Granger causality analysis and discovered no evidence of an emotional tweet’s causal effect on the Bitcoin market, and suggested that sentiment more frequently reflects the market than anticipates it. [1] discovered that regardless of potential price fluctuations, Twitter users are generally positive about cryptocurrencies. Even when the value of cryptocurrencies declines, those who tweet about them have an interest in them that transcends the investment opportunity. Thus Twitter sentiment appears inconsistent with the falling price of Bitcoin. [39] tested the relationship between sentiment and Bitcoin price using a multiple linear regression model. The model fails the significance test, leading to the conclusion that sentiment and Bitcoin price have no statistically significant relationship.

One of the most recent studies by [27] demonstrated that only collecting tweets from the most influential cryptocurrency Twitter accounts makes it possible to eliminate noisy data and reduce the size of the dataset while maintaining the same level of prediction accuracy. Therefore, this study extracts data from only the most influential Twitter users and explores new methods for measuring the chain effect caused by user interactions on Twitter. In addition to extracting sentiment from tweets, we collect user interaction behaviour of the subsequent impact of original tweets, such as the number of likes, retweets, and replies. The sentiment score and these interaction behaviours are selected as modelling features, and their combined predictive power for the Bitcoin price is examined.

3. METHODOLOGY

This section describes the techniques for data collection, preparation, model construction and analysis. The sentiments and the user interaction behaviour, which are the number of replies, retweets, and likes, are derived from tweets and serve as the input variable to the multiple linear regression and Extreme Gradient Boosting (XGBoost) models, while the daily close price of Bitcoin serves as the target variable. An overview of the research methodology is presented in Figure 1

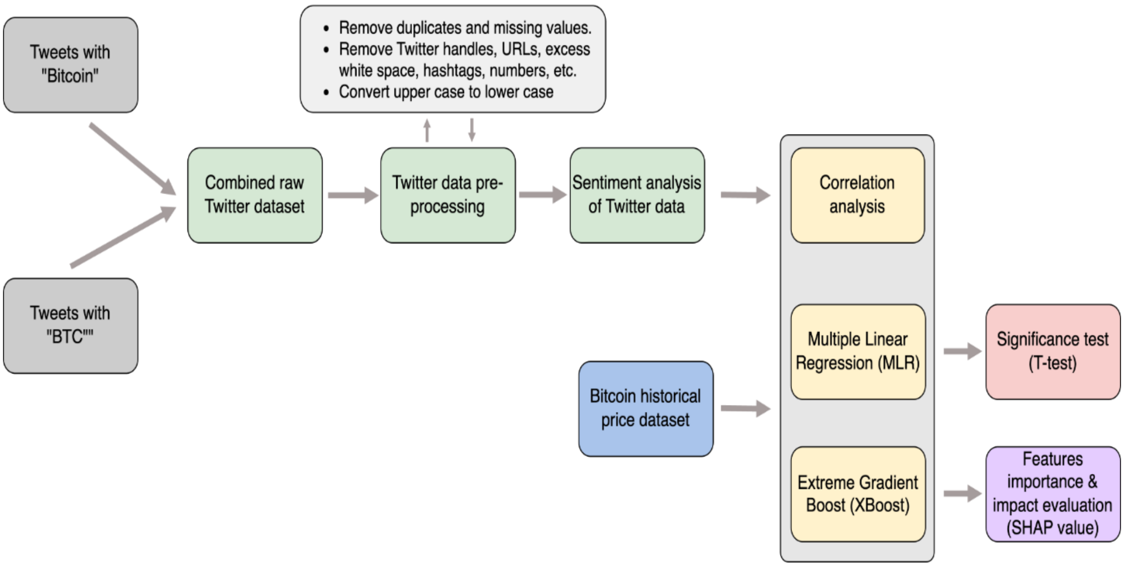

Fig 1: Methodology for Bitcoin price prediction using sentiment analysis and user interaction behaviour

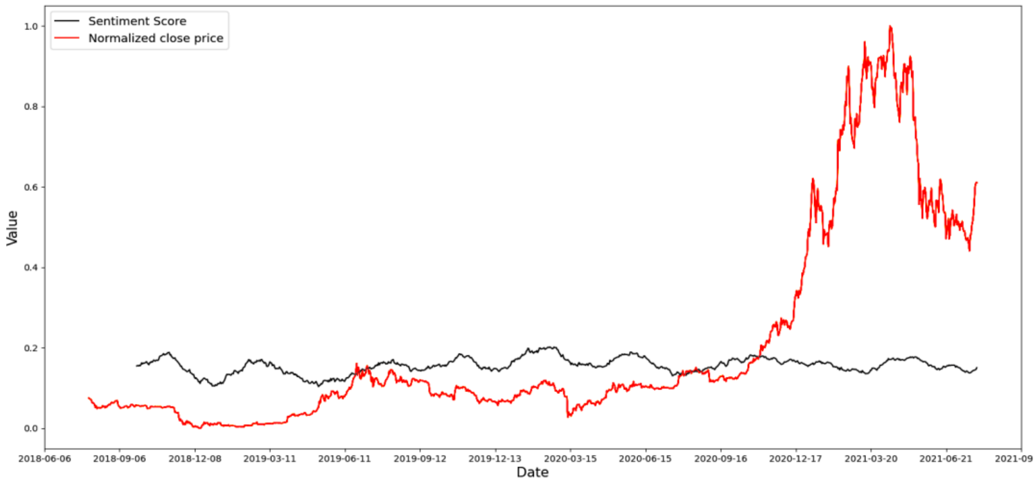

Fig 2: Line graph for sentiment score and Bitcoin price

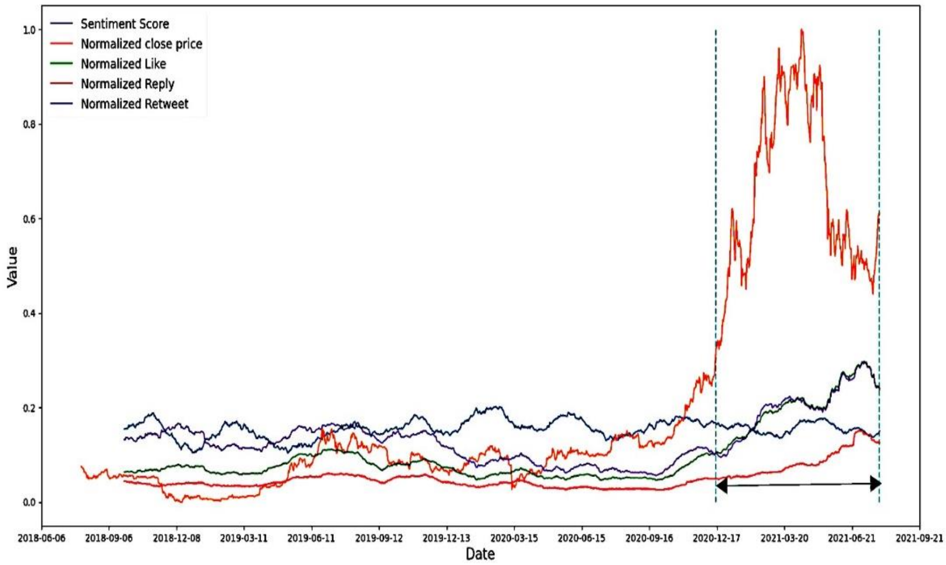

Fig 4: Comparing Bitcoin price, sentiment score and user interaction behaviour

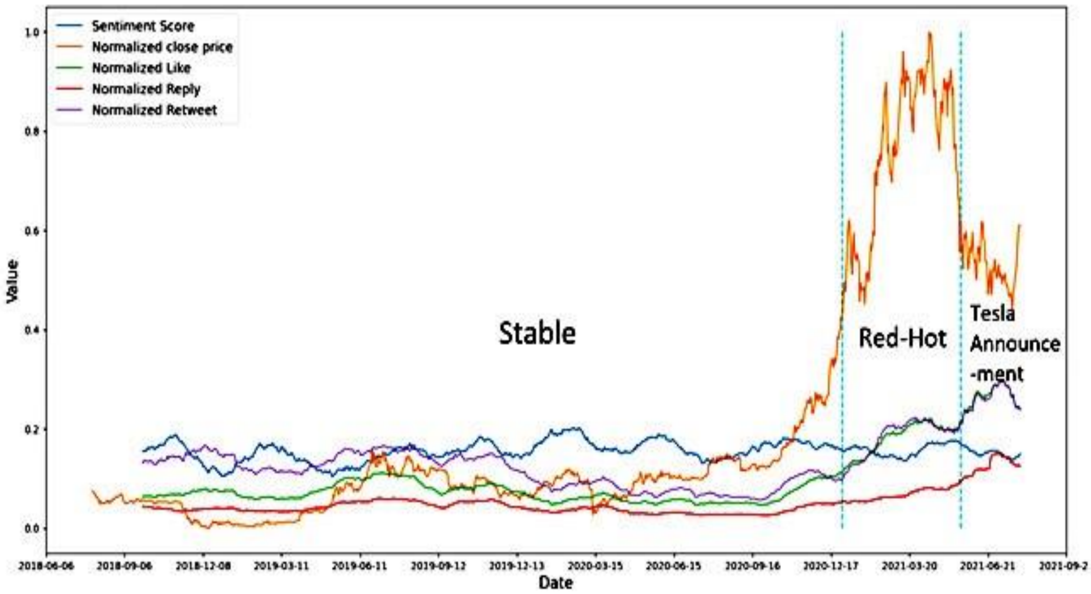

Fig 7: Line graph for each cluster

Table 4. Model result based on the entire dataset

Independent variable	Coefficients (un-normalized)	Coefficients (Normalized)	p-value	Significant?	RMSE
Sentiment Score (x_s)	120.86	0.1285	0.258	No	
Number of Like (x_l)	3.57	0,0020	< 2e−16	Yes	*Linear Regression* 35519.18
*Number of Retweet (x_{rt})	−11.96	3.948	< 2e−16	Yes	
Number of Reply (x_{rp})	-3.16	-2.7621	11e−15	No	*XGBoost*: 35032.36
Intercept	10937.21	−1.0571	2e−16	Yes	

∗ significant variable with the highest coefficient

Table 5. Model result based on the entire dataset

Independent variable	Coefficients (un-normalized)	Coefficients (Normalized)	p-value	Significant?	RMSE
Sentiment Score (x_s)	116.2551	0.0019	0.0029	Yes	
Number of Like (x_l)	0.5011	0.4910	2.91e−15	Yes	*Linear Regression* 97.00
*Number of Retweet (x_{rt})	−1.7789	−0.4109	2.42e−16	Yes	
Number of Reply (x_{rp})	0.2066	0.0690	0.5524	No	XGBoost: 9678.98
Intercept	17633.1324	0.0738	< 2e−16	Yes	

Table 6. Results for the Red-hot Cluster

Independent variable	Coefficients (un-normalized)	Coefficients (Normalized)	p-value	Significant?	RMSE
Sentiment Score (x_s)	526.9385	0.0087	0.0621	No	
Number of Like (x_l)	0.2832	0.2774	0.0543	No	*Linear Regression:7919.28*
*Number of Retweet (x_{rt})	1.4873	0.3436	0.0329	Yes	
Number of Reply (x_{rp})	2.9201	0.9756	0.1563	No	XGBoost: 8357.14
Intercept	46911.4589	0.7242	$< 2e^{-16}$	Yes	

Table 7. Model result based on the entire dataset

Independent variable	Coefficients (un-normalized)	Coefficients (Normalized)	p-value	Significant?	RMSE
Sentiment Score (x_s)	83.40	0.0014	0.864	No	
Number of Like (x_l)	- 0.0610	- 0.0598	0.197	No	*Linear Regression:3913.14*
*Number of Retweet (x_{rt})	0.0948	0.0219	0.761	No	
Number of Reply (x_{rp})	- 0.1293	- 0.0432	0.200	No	*XGBoost: 1927.14*
Intercept	36300	0.5484	$< 2e^{-16}$	Yes	

3.1 Data Collection

Instead of the official Twitter API, which has a monthly limit on the number of tweets that can be collected, snscrape is utilised to scrape tweets from Twitter. It is an open-source Python library that allows users to collect an infinite number of tweets for any given time period. The data collection method introduced by [27] is adopted. Therefore, only tweets posted by the top 35 influential Twitter users are collected for the study. These Twitter users or accounts are chosen based on the number of their followers, and their inclusion on Bitcoin.com as the most influential ”Bitcoiners” in 2020 [33]. These 35 accounts' tweets containing the keywords ”Bitcoin” and ”BTC” from July 31 2018, to July 29 2021 are collected and analysed. A further benefit of extracting tweets only from influential users is that it automatically excludes meaningless tweets from Twitter bot accounts, ensuring all tweets retrieved are authentic and not promotional. The daily closing Bitcoin prices in US dollars are sourced from Bitstamp, one of the most well-known Bitcoin trading platforms in the world. The historical Bitcoin prices from July 31, 2018, to July 29, 2021 are available to the public and were collected.

3.2 Data preprocessing

Sentiment analysis requires extensive preprocessing of the raw Twitter data because it is in an unstructured format and contains a high level of noise. The initial CSV dataset consists of 37,339 tweets with their respective text contents and numbers of likes, retweets, and replies. The following procedures are used to clean the data:

1. Remove duplicate tweets: tweets containing the terms ”Bitcoin” and ”BTC” are separately searched and then combined. There is a possibility that some tweets with both keywords are counted twice.
2. Determine missing values: there are 16 tweets with missing values for the number of replies, so the value zero is imputed.
3. Remove unwanted words and characters: words and characters, such as Twitter handles, retweet handles, URLs, excess white space, hashtags, and numbers, are removed.
4. Convert case: finally, capital letters are converted to lowercase prior to analysis.

3.3 Twitter sentiment analysis using VADER

The sentiment score is extracted from tweets using a lexicon-based method [11]. A lexicon is a dictionary in which each word is assigned a predetermined positive or negative sentiment value. In lexicon-based approaches, a piece of text is represented as a bag of words. Following this message representation, all positive and negative words and phrases are assigned sentiment values from the lexicon. The final score of the message's overall sentiment is determined by combining functions such as sum and average.

The [11] developed Valence Aware Dictionary for Sentiment Reasoning (VADER) is chosen for Twitter sentiment analysis. Gilbert and Hutto (2014) demonstrate that VADER outperforms both human annotators and the majority of benchmark classifiers. In addition, it is specifically attuned to sentiments expressed in social media and is, therefore, a good fit for analysing sentiment from tweets. Not only does VADER classify text as positive, negative, or neutral, but it also measures the polarity using a normalised score between -1 and 1, with scores close to -1 indicating extreme negativity and scores close to 1 indicating extreme positivity.

Similarly, VADER takes punctuation and semantics into account as it assigns a higher score to exclamation marks, accounts for changes in sentiment after conjunctions such as ”but,” and, most importantly, it updates its lexicon to include the most recent emojis, abbreviations, and slang such as ”lol,” ”lmao,” and ”:P,” which indicate extremely strong emotions. It also identifies hidden emotions in contractions like ”haven’t” and ”don’t.” Due to VADER’s capability to interpret punctuation, uncommon expressions, abbreviations, and stop words, they are not removed in this study, as they improve the accuracy of the sentiment score rather than adding noise. Python’s nltk library’s SentimentIntensityAnalyzer() is used to implement the VADER.

3.4 Correlation between Sentiment Score and Bitcoin Price

Prior to engaging in modelling, it is necessary to evaluate and visualise the approximate relationship between sentiment score and Bitcoin price, which is depicted in Figures 2 and 3 as a line graph and scatter plot, respectively. Normalising the Bitcoin price improves the visual representation of the relationship between the Bitcoin price and the sentiment score.

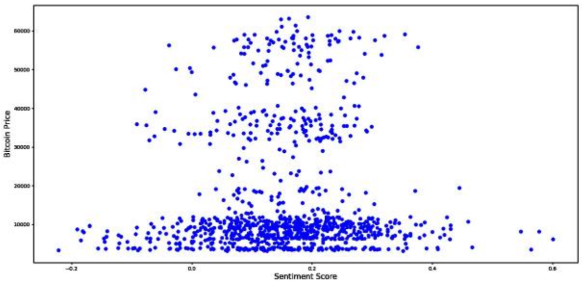

Fig. 3. Scatter plot for sentiment score and bitcoin price

The correlation between the Bitcoin price and sentiment score is 0.025. This low correlation is evident in the line graph illustrating the evolution of the daily average sentiment score and the Bitcoin price over the past three years. It demonstrates that the daily sentiment score averaged around 0.2, indicating that Twitter users tend to post more positive tweets about Bitcoin, consistent with the findings of [1]. The sentiment score does not show any discernible trend over the past three years. However, the price of Bitcoin experienced a tremendous increase at the end of 2020, after a long period of stability from 2018 to 2020. Despite a significant decrease in May 2021, the price is still relatively higher than it was previously.

The scatter plot presents the relationship between the bitcoin price and sentiment score in greater detail (Figure 3). At first glance, a linear relationship between sentiment and Bitcoin price does not appear to exist. For example, observations with a daily sentiment score between 0 and 0.3 correspond to Bitcoin prices ranging from 5,000 to 60,000 USD. Thus, it appears Bitcoin’s price is unlikely to be explained solely by sentiment score, consistent with the findings of [39]. To increase the prediction accuracy of the Bitcoin price, other indicators are therefore strongly needed as input features that bring extra informational contributions to the final model.

3.5 User Interaction Behaviour and Bitcoin Price

The preceding section elaborated on the need for additional features to improve Bitcoin price prediction accuracy. For this purpose, Twitter user interaction behaviour, such as the number of replies, likes, and retweets, was added. A normalised timeseries graph showing the relationship between Bitcoin price and user interaction behaviour is presented in Figure 4. A slightly upward movement in the number of replies, retweets, and counts during the Bitcoin price’s rapid increase at the end of 2020 is observed. Indicating a relationship between the user interactive features and the Bitcoin price, suggesting that the user interaction behaviour may have a direct influence on Bitcoin’s price.

To further quantify these relationships, a correlation heatmap between the sentiment score, user interaction behaviour and the Bitcoin close price is used (see Figure 5). The number of likes appears to be the most relevant feature to the Bitcoin price (r = .58) followed by the number of replies and the number of retweets (r = .58,r = .34), respectively. A low correlation coefficient (r = .025) between sentiment and Bitcoin price demonstrates a weak relationship between the two variables.

To further investigate the relationship between the sentiment score and the Bitcoin price, a scatter plot of the Bitcoin price versus the sentiment score was used. The result shows three clearly defined clusters (see Figure 6). Using the Kmeans clustering technique, the clusters and their centroids were determined. The red (A) cluster represents the extreme value of Bitcoin’s price, primarily from January 1 2021, to May 20 2021, when Bitcoin became a mainstream asset class. The majority of the yellow (C) clusters correspond to times when Bitcoin’s price was relatively stable. The majority of the black (B) cluster represents the Bitcoin price after May 21 2021, when Elon Musk announced the suspension of Bitcoin payments for Tesla vehicle purchases.

Therefore, we build a predictive model based on the entire dataset and on the individual clusters to investigate the impact of the different cluster characteristics on the accuracy of the predictive models. Building a predictive model based on individual clusters is intuitive, as extreme datasets from different clusters may introduce bias, especially in the presence of weakly correlated features.

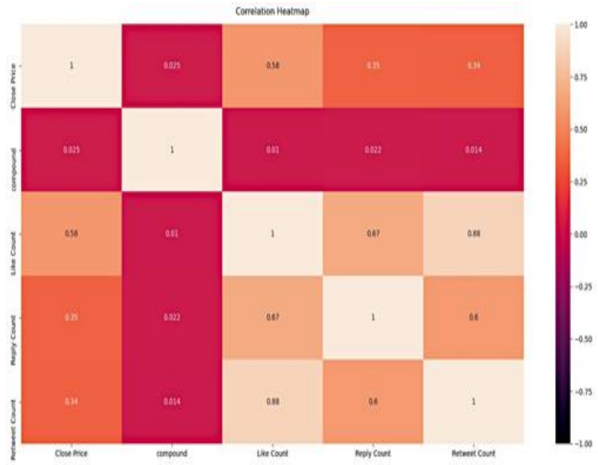

Fig. 5. Correlation heatmap

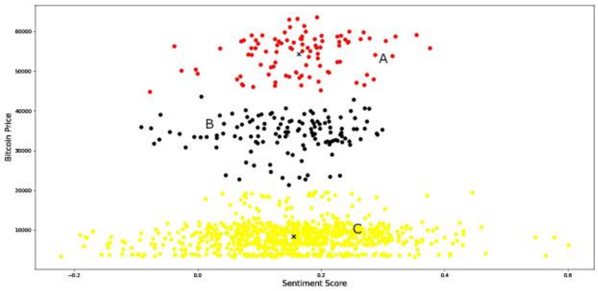

Fig. 6. Scatter plot for sentiment score after k-means clustering

The accuracy of these models is compared, afterwards. The names of the three clusters are derived from either the widespread acclaim or the significant events that occurred during the corresponding time period (see Table I and Figure 7). Additionally, the majority of the data points within each cluster have sentiment values between 0 and 0.3, indicating that sentiment values are independent of clusters, thus suggesting the inclusion of new features with strong predictive power.

Table 1. Period for each cluster

Name	Stable	Red-hot	Tesla announcement
Period	31/07/2018-31/12/2020	01/01/2021-20/05/2021	21/05/2021-29/07/2021
Cluster ID	C (Yellow)	A (Red)	B (Black)

3.6 Modelling

Model Selection: Multiple linear regression (MLR) and Extreme Gradient Boost (XGBoost) are utilised to forecast the Bitcoin price. Both models have been widely used in forecasting Bitcoin prices using the social media sentiment score. MLR models the linear relationship between independent and dependent variables. On the other hand, the XGBoost is an ensembled decision tree algorithm that constructs a strong classifier from a series of weak classifiers, thus able to deal with the bias-variance trade-off [6].

Model Building: Table II lists the dependent variables and target variables selected for this study. All variables retain their initial value in the absence of normalisation and standardisation, and each record in the dataset represents a tweet along with its sentiment score, number of likes, retweets, and replies, and the Bitcoin close price on the day the tweet was posted. To ensure that every tweet corresponds to a single Bitcoin price, the daily Bitcoin price is duplicated so that all tweets posted on the same day correspond to the same Bitcoin close price.

The R's lm() function was used to execute the multiple linear regression model, and the xgboost and sklearn packages are used to execute the XGBoost model in Python, while the accuracy of predictions is measured using the Root Mean Square Error (RMSE). For the XGBoost, a set of parameters are determined beforehand, where parameters with minimal impact on the model's predictions output retain their default values, while other parameters, such as eta, max depth, subsample, and colsample bytree are assigned reasonable values within the suggested range. The eta represents the number of steps required to arrive at the optimal prediction. The max depth parameter prevents over-fitting, which indicates that a greater depth will result in extremely specific relationships with a given sample. The subsample represents the fraction of observations to be sampled at random for each tree, and the colsample bytree represents the fraction of columns to be sampled at random for each tree (Jain, 2016). Each value is presented in Table III.

Table 2. Notation for each variable

Name of Variables	Notation
Sentiment Score (compound score)	x_s
Number of Like	x_l
Number of Retweet	xrt
Number of Reply	xrp
Daily Bitcoin Close Price (USD)	y

Table 3. Value of XGBOOST parameters

Types of parameter	Name of parameter	Value
General	booster Silent nthread	default default default
	num pbuffer	default
	num feature	default
Tree Booster	eta gamma max depth min child weight	0.1 default 12 default
	max delta step	default
	subsample	0.7
	colsample btree	0.7
Task	objective base score	reg:linear default
	eval metric	default

4. RESULT

4.1 Modelling the entire dataset

This section presents the outcomes of the multiple linear regression and XGBoost models utilising the entire dataset. The entire dataset is divided based on the timeline, with the training dataset set containing data from 31 July 2018 to 31 January 2021 and the test dataset set containing data from 1 February 2021 to 29 July 2021.

4.1.1 Multiple Linear Regression

The coefficients, p-values, and RMSE for each input variable are shown in Table IV. The p-value evaluates the null hypothesis that the coefficient is zero,i.e., it has no effect on the model outcome. Therefore, a significant p-value at 95% CI indicates that the null hypothesis is rejected and the variable contributes significantly to the model output. The result shows that the numbers of likes, retweets, and replies are all statistically significant predictors with a p-value less than 0.001 ($p < .001$) (see Table IV). However, sentiment's p-value is extremely large and nonsignificant ($p > .1$), indicating that changes in sentiment score are unlikely to have any influence on Bitcoin price.

To compare the contribution of each independent variable, a normalised dataset is employed to ensure that all variables' values fall within the same numerical range. In this manner, the coefficient values have an equivalent effect on the predicted output. The compound score is adopted for the sentiment score, which is a metric that calculates the sum of all the lexicon ratings, which have been normalised between −1 (extremely negative) and +1 (extremely positive).

The results demonstrated that the number of retweets had the highest coefficient of 3.948, indicating that it has a higher predictive power in relation to the Bitcoin price. The positive value of its normalised coefficient indicates that the price of Bitcoin follows the same upward or downward trend as the number of retweets. In contrast, the number of replies has a negative effect on the Bitcoin price, meaning that when reply activity increases, the Bitcoin price is more likely to exhibit a downward trend.

The average of the Bitcoin price predictions from tweets posted on the same day yields the daily Bitcoin price prediction. The RMSE for MLR and XGBoost is similar and is computed to be 35519.18. Thus, if the actual Bitcoin price is 40,000, the predicted output is likely to be in the range of $75,519.18. This is an extremely large RMSE as it is equal to approximately half of the actual Bitcoin price range, which is between $3179 and $63564. Therefore, the MLR fails to estimate the average Bitcoin daily price performance.

4.1.2 XGBoost

Using the XGBoost, the final RMSE is 35032.36, which is comparable to the RMSE produced by MLR. Consequently, the Shapley Additive Explanations (SHAP) value summary plot is used to explain the innerworkings of the XGBoost model (see Figure 8). The SHARP summary plot illustrates the ranking of feature importance and whether each feature has a negative or positive effect on the predicted Bitcoin price. The top variables contribute more to the output of the model than the bottom variables, therefore having greater predictive power.

In the SHARP summary plot, the SHAP value is predominantly negative when the number of likes is low, indicating a low Bitcoin price output. Similarly, as the number of likes increases, the SHAP value increases significantly, indicating an increase in the Bitcoin price output. This suggests that the number of likes, as a feature, has a positive relationship with the Bitcoin price output. While the number of retweets and replies has a slightly negative effect on Bitcoin's output, suggesting that fewer retweets and replies result in a slightly higher Bitcoin output price.

In summary, when the training set ignores the difference among clusters, multiple linear regression or XGBoost shows similar prediction accuracy. Neither of them can develop accurate prediction outcomes that are close to the actual level of the Bitcoin price. A common finding is that the number of likes has a positive impact on the Bitcoin price, while the number of replies shows a negative impact. The impact of the sentiment score is insignificant, both visually and statistically.

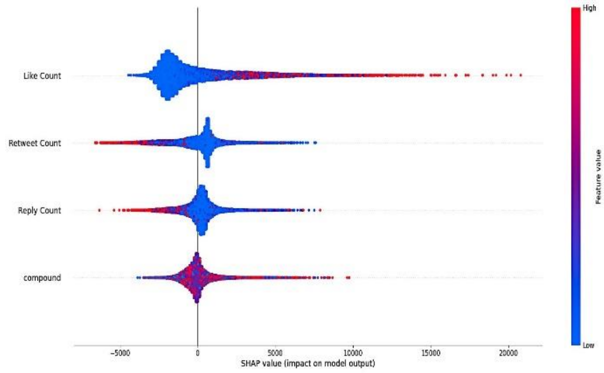

Fig. 8. SHAP value summary plot for entire dataset

4.2 Modelling on individual Clusters

We investigate the model performance on each individual cluster. To achieve this, the dataset is partitioned based on the timeline with a ratio of 5:1 for training and testing datasets respectively. This ratio is repeated for the "red-hot" and "Tesla announcements" clusters (see Table I).

4.2.1 Stable Cluster

A t-test is performed to examine the significance of individual regression coefficients in the MLR model to determine the significance and relative importance of each feature in the cluster. The p-value produced by linear regression for sentiment score coefficients is statistically significant ($p = .001$) (see Table V). However, the coefficients of sentiment score derived from the normalised dataset exhibit a weak correlation, indicating that sentiment score continues to contribute relatively less to price prediction. The number of likes is statistically significant and remains the most contributory and crucial feature ($p < .001$), followed by the number of retweets ($p < .001$). The number of replies contributes the least, with an insignificant regression coefficient ($p > .1$).

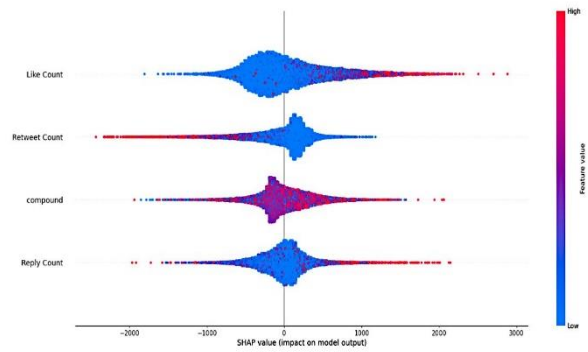

Fig 9: SHAP summary plot for stable cluster

Figure 9 illustrates the importance of the features based on the SHAP value from XGBoost. The SHAP value is more likely to be negative when the number of likes is low, suggesting that a low number of likes indicates a low predicted Bitcoin price. Consequently, the SHAP value increases relative to the number of likes, indicating that a larger number of likes corresponds to a higher value in the Bitcoin price prediction. Also, during the prediction period, the price experiences a slight increase, but neither of the two models captures this pattern

In contrast, the number of retweets has the opposite effect on the model output; fewer retweets result in a higher predicted price. On the other hand, the majority of points for the number of replies are located near the zero baselines, indicating a weak contribution to the model prediction. In addition, the influence of sentiment score indicated as the compound score, is relatively small, as the same sentiment score achieves varying SHAP values, indicating that sentiment score is a poor predictor. The aforementioned findings are consistent with the outcome of the linear regression.

4.2.2 Red-hot cluster

As with the stable cluster, a t-test is conducted to determine the significance of each feature in the red-hot cluster by examining the significance of individual regression coefficients in the multiple linear regression model. Due to the value consistency between the training set and test set, the accuracy of predictions is marginally enhanced (RMSE = 7919 and 8357 for the MLR and XGBoost, respectively).

However, only the number of retweets passes the significance test, with the exception of the intercept. Therefore, the higher accuracy may be primarily due to the cumulative increase in the number of user interactions during the price volatility period and not necessarily due to the direction of the interaction, which captures the user behaviour. This shows that user interaction behaviour may be less important during Bitcoin price volatility. Also, despite the fact that the coefficient of reply count is the largest of all, it is statistically insignificant ($p > 0.1$), so its impact is debatable.

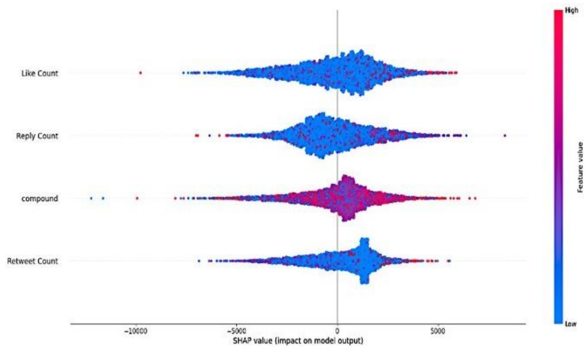

Fig 10: SHAP summary plot for red-hot cluster

The output of XGBoost’s feature importance rank for the red-hot cluster is opposite that of the stable cluster, as the average effect of the like count on the Bitcoin price is neutral. In general, the effect revealed by each individual observation appears to be more centralised than in the stable cluster, with the majority of values located near the zero baselines and smaller values at both the positive and negative SHAP value ranges, with the exception of the compound values. In terms of their contribution to price prediction, the number of likes, retweets, and replies in the red-hot cluster plays a lesser role compared to the stable cluster (see Figure 10).

A different finding in this cluster is that the reply count outperforms sentiment and retweet count in terms of the feature’s importance and has a more negative impact on the Bitcoin price. This implies many users are more ready to comment on Bitcoin-related tweets when the prices are low than when the prices are high. Additionally, the sentiment score has a greater influence on Bitcoin’s price prediction in the red cluster, with a predominantly negative impact on Bitcoin’s price, suggesting that a high sentiment score indicates a low Bitcoin price prediction. This is in agreement with the impact of reply count on the model outcome, as more sentiment is likely to be observed when Bitcoin prices are low (more replies) compared to when Bitcoin prices are high.

4.2.3 Tesla announcement cluster

Multiple linear regression’s prediction accuracies is increased in the Tesla announcement cluster by approximately 50%, while XGBoost improves by approximately 75%, with RMSEs of 3913 and 1927, respectively. Unlike previous clusters, XGBoost significantly outperforms multiple linear regressions and appears to captures price movement to some extent.

Nevertheless, as indicated by the p-value from linear regression in Table VII, none of the features passes the significance test with the exception of the intercept ($p > 0.1$). Hence, the higher accuracy may be primarily attributable to the cumulative increase in the number of user interactions within the Tesla announcement cluster as opposed to the directions of interaction that signifies user behaviour. This shows that user interaction behaviour becomes less important during Bitcoin price volatility. The SHAP summary plot presented in Figure 11 reveal that the number of likes is no longer the most important feature; rather, the number of retweets is the most important and has a neutral impact on the Bitcoin price. All other characteristics also have a relatively neutral effect on the Bitcoin price. The sentiment score, meanwhile, becomes the second most important feature, indicating more sentiments are observable during Bitcoin’s price volatility, as indicated in the red-hot cluster.

In summary, although the accuracy of prediction varies from cluster to cluster, XGBoost generally outperforms the multiple linear regression model in accurately predicting the Bitcoin price and capturing the price trend. In addition, depending on the clusters, the significance and relative importance of each feature, as well as their ranks, change significantly.

5. DISCUSSION

The results provide valuable insights into Bitcoin price prediction using Twitter sentiment score and user interaction hehaviour. It has been demonstrated that the predictive power of sentiment is insufficient to forecast the Bitcoin price as a whole. However, there is a positive correlation between the Bitcoin price and user interaction behaviour, namely the number of retweets, replies, and likes. Consequently, a combination of the sentiment score and the user interaction behaviour as inputs to the multiple linear regression and the XGBoost significantly improves the accuracy of the prediction but fails to capture the trends in the Bitcoin price movement.

Comparing the results from each of the three clusters reveals that when there is a fluctuation or movement trend, the user interaction behaviour and sentiment score play a lesser role in predicting the Bitcoin price. This observation is supported by the decreased in variable coefficient and significance for the multiple linear regression, as well as the relatively neutral impact of the features in the XGBoost model. Therefore, the increased prediction accuracy for the volatile clusters (redhot and Tesla announcement) is not solely attributable to the additional features, but rather to the models’ overall predictive ability and the increase in user interaction. However, when the Bitcoin price is relatively stable, the improved accuracy is primarily due to the addition of corresponding user interaction behaviour, which have both a strong positive and negative impact on the models' outcomes.

The model analysis indicates that the impact of the sentiment score on the predicted Bitcoin price is significantly low, regardless of the price movement trend. This indicates that the majority of Bitcoin-related tweets originate from Bitcoin enthusiasts whose opinions are unaffected by market fluctuations. The relatively neutral to positive sentiment scores also support this conclusion. The results support Jain's (2018) assertion that Twitter sentiment alone is insufficient for predicting the Bitcoin price and that additional factors are strongly recommended to improve the accuracy of price forecasts.

The addition of user interaction behaviour to the sentiment score improves the overall accuracy of prediction for multiple linear regression and XGBoost but does not reflect the trends in bitcoin price movement. This is due to the fact that the addition of user interaction behaviour significantly contributes to the increases in the accuracy of model predictions only when the Bitcoin price is stable, as their respective impacts on the model outcome are significant, with either negative or positive effects. While the impact of user interaction behaviour on model output is negligible during upward or downward price volatility. Consequently, the Twitter user interaction behaviour are independent of the price movement, as their impacts on the model's output become negligible during periods of price volatility, despite the increase in prediction accuracy. This is also evident in the multiple regression model's non-significant regression coefficients.

Therefore, the improved prediction accuracy observed during the volatile periods (red-hot and Tesla's announcement) is solely attributable to an increase in the total number of interactions from new sets of users, as opposed to the interactions from existing users during the stable period. According to [22], there is a strong correlation between Google Trends and Bitcoin price. Consequently, it suggests that during Bitcoin's volatile price movement, more people (mostly non-Bitcoin enthusiasts) tend to react to Bitcoin tweets or search for Bitcoinrelated tweets on Twitter, similar to a Google search, thus bringing neutrality to the behavioural trends observed during the stable price period. This realisation also explains why the sentiment score is predominantly neutral during periods of price volatility as opposed to periods of price stability.

6. CONCLUSION

In conclusion, this paper predicts the closing price of Bitcoin using Twitter sentiments and user interaction behaviour as inputs to MLR and XGBoost models. Our research indicates that Twitter sentiment is insufficient for predicting the Bitcoin price as a whole. Moreover, the results of the predictive models indicate that using both sentiment score and user interaction behaviour as model inputs significantly improves prediction accuracy, but fails to capture Bitcoin price movement trends. The use of regression analysis and SHAP plot to analyse feature importance and impact on prediction outcome for both models indicates that sentiment score and user interaction behaviour play a lesser role in predicting the Bitcoin price when price volatility is present. When the Bitcoin price is relatively stable, however, the improved accuracy is primarily due to the addition of user interaction behaviour to the sentiment score as input into the predictive models.

We conclude that the sentiment score is insufficient because the majority of Bitcoin-related tweets come from Bitcoin enthusiasts whose opinions are unaffected by market fluctuations. We also conclude that the improved prediction accuracy observed during Bitcoin's price volatility is solely attributable to an increase in the total number of interactions from new sets of users and not the direction of interaction observable during the stable price period. This phenomenon is comparable to the correlation observed between the price of Bitcoin and Google Trends [22]. Thus, the increase in non-Bitcoin enthusiast user interaction during periods of price volatility neutralises the sentiment and directional user interaction trends observed during periods of stable price movement. As a result, despite the models' improved prediction accuracy during periods of price volatility, they fail to capture the trends in the Bitcoin price movement.

7. REFERENCES

[1] Abraham, J., Higdon, D., Nelson, J. and Ibarra, J. (2018) Cryptocurrency Price Prediction Using Tweet Volumes and Sentiment Analysis. SMU Data Science Review: [online] 1(3) Available from: https://scholar.smu.edu/datasciencereview/vol1/iss3/1 [accessed 13 September 2021].

[2] Benevenuto, F., Rodrigues, T., Cha, M., & Almeida, V. (2009, November). Characterizing user behavior in online social networks. In Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement (pp. 49-62).

[3] Blasco, N., Corredor, P. and Ferreruela, S. (2012). Market sentiment: a key factor of investors' imitative behaviour. Accounting & Finance, 52(3), 663-689.

[4] Bollen, J., Mao, H. and Zeng, X. (2011). Twitter mood predicts the stock market. Journal of Computational Science, 2(1), 1-8.

[5] Colianni, S., Rosales, S., and Signorotti, M. (2015). Algorithmic Trading of Cryptocurrency Based on Twitter Sentiment Analysis.

[6] Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining (pp. 785-794).

[7] Contreras, J., Espinola, R., Nogales, F. and Conejo, A. (2003). ARIMA models to predict next-day electricity prices. IEEE Transactions on Power Systems, 18(3), 1014-1020.

[8] Galeshchuk, S., Vasylchyshyn, O., and Krysovatyy, A. (2018). Bitcoin Response to Twitter Sentiments. ICTERI Workshops.

[9] Garcia, D. and Schweitzer, F. (2015) Social signals and algorithmic trading of Bitcoin. Royal Society Open Science, 2(9).

[10] Gotz, D., Zhou, M. X. (2008). An empirical study of user interaction behavior during visual analysis. IBM Research RC24525 (W0803-127).

[11] Hutto, C., & Gilbert, E. (2014, May). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In Proceedings of the international AAAI conference on web and social media (Vol. 8, No. 1, pp. 216-225).

[12] Jahjah, F.H., and Rajab, M. (2020). Impact of Twitter Sentiment Related to Bitcoin on Stock Price Returns. The Journal of Engineering, 26, 60-71.

[13] Jain, A. (2016). Complete Guide to Parameter Tuning in XGBoost with codes in Python. [online] Available from: https://www.analyticsvidhya.com/blog/2016/03/complete-guideparameter-tuning-xgboost-with-codes-python/ [accessed 13 September 2021].

[14] Jain, A., Tripathi, S., Dwivedi, H. and Saxena, P. (2018). Forecasting Price of Cryptocurrencies using Tweets Sentiment Analysis. Proceedings of 2018 Eleventh International Conference on Contemporary Computing (IC3), Noida, India.

[15] Jang, H. and Lee, J. (2017). An Empirical Study on Modeling and Prediction of Bitcoin Prices with Bayesian Neural Networks Based on Blockchain Information. IEEE, 6, 5427-5437.

[16] Kahneman, D. and Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. Econometrica, 47(2), 263-292.

[17] Kaminski, J. (2014), Nowcasting the bitcoin market with twitter signals. MIT Media Lab [online] Available from: http://arxiv.org/abs/1406.7577 [accessed 13 September 2021].

[18] Kim, Y., Lee, J., Park, N., Choo, J., Kim, J. and Kim, C. (2017). When Bitcoin encounters information in an online forum: Using text mining to analyse user opinions and predict value fluctuation. PLOS ONE, 12(5), e0177630.

[19] Kraaijeveld, O., Smedt, J. (2020). The predictive power of public Twitter sentiment for forecasting cryptocurrency prices. Journal of International Financial Markets, Institutions & Money. 65, 101188.

[20] Li, T., Chamrajnagar, A., Fong, X., Rizik, N. and Fu, F. (2019). Sentiment-Based Prediction of Alternative Cryptocurrency Price Fluctuations Using Gradient Boosting Tree Model. Frontiers in Physics, 7(98).

[21] Linardatos, P., Papastefanopoulos, V., & Kotsiantis, S. (2020). Explainable ai: A review of machine learning interpretability methods. Entropy, 23(1), 18.

[22] Madan, I., Saluja, S., & Zhao, A. (2015). Automated bitcoin trading via machine learning algorithms. URL: http://cs229. stanford.edu/proj2014/Isaac% 20Madan, 20.[accessed 12 January 2023].

[23] McNally, S., Roche, J. and Caton, S. (2018) Predicting the Price of Bitcoin Using Machine Learning. 26th Euromicro International Conference on Parallel, Distributed and Network-based Processing (PDP). 339-343.

[24] Mohapatra, S., Ahmed, N. and Alencar, P. (2019). KryptoOracle: A RealTime Cryptocurrency Price Prediction Platform Using Twitter Sentiments. IEEE Big Data.

[25] Nakamoto, S. (2008). Bitcoin: A peer-to-peer electronic cash system. Decentralized Business Review, 21260.

[26] 2010)]con01 O'Connor, B., Balasubramanyan, R., Routledge, B. and Smith, N. (2010). From tweets to polls: Linking text sentiment to public opinion time series. Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media, 122-129.

[27] Ozt¨ urk,¨ S., and Bilgic¸, M. (2021). Twitter & bitcoin: are the most influential accounts really influential?, Applied Economics Letters [online] Available from: https://doi.org/10.1080/13504851.2021.1904104 [accessed 13 September 2021].

[28] Pagolu, V., Challa, K., Panda, G. and Majhi, B. (2016). Sentiment Analysis of Twitter Data for Predicting Stock Market Movements. International conference on Signal Processing, Communication, Power and Embedded System (SCOPES). Paralakhemundi, India. 3th-5th October 2016.

[29] Panger, G.T. (2017) Emotion in Social Media. Ph.D. University of California, Berkeley.

[30] Ranco, G., Aleksovski, D., Caldarelli, G., Grcar, M., and Mozetiˇ c, I.ˇ (2015). The Effects of Twitter Sentiment on Stock Price Returns. PLoS ONE, 10 (9).

[31] Rao, T. and Srivastava, S. (2012) Analyzing Stock Market Movements Using Twitter Sentiment Analysis. Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining.

[32] Redman, J. (2019). Bitcoin Gained 8.9 Million Percent Over the Last Decade – Featured Bitcoin News. [online] Available from: https://news.bitcoin.com/bitcoin-gained-8-9-million-percent-over-the-lastdecade/ [accessed 13 September 2021].

[33] Redman, J. (2021). The 35 Most Influential Bitcoiners Dominating Crypto Twitter by Follower Count – Featured Bitcoin News. [online]. Available from: https://news.bitcoin.com/35-most-influentialbitcoin-crypto-twitter/ [accessed 13 September 2021].

[34] Runkevicius, D. (2021). Myth Busted: Institutional Investors Are Not Investing In Bitcoin [online] Available from:https://www.forbes.com/sites/danrunkevicius/2021/01/14/myth-bustedwall-street-is-not-investing-in-bitcoin/?sh=43efa1025510 [accessed 13 September 2021].

[35] Silverman, G., Murphy, H. and Authers, J. (2017). Bitcoin: an investment mania for the fake news era, [online] Financial Times. Available from: https://www.ft.com/content/c84caffc-d683-11e7-a303-9060cb1e5f44 [accessed 13 September 2021].

[36] Smales, L. (2017). The importance of fear: investor sentiment and stock market returns. Applied Economics, 49(34), 3395-3421.

[37] Sul, H., Dennis, A. and Yuan, L. (2016). Trading on Twitter: Using Social Media Sentiment to Predict Stock Returns. Decision Sciences, 48(3), 454-488.

[38] Tetlock, P. (2007). Giving Content to Investor Sentiment: The Role of Media in the Stock Market. The Journal of Finance, 62(3), 1139-1168.

[39] Verma, M. and Sharma, P. (2020). Money Often Costs Too Much: A Study to Investigate The Effect Of Twitter Sentiment On Bitcoin Price Fluctuation. Preprints [online] Available from: https://doi.org/10.20944/preprints202009.0216.v1 [accessed 13 September 2021].

[40] Weigel, Detlef, and Jane Glazebrook. 2002. Arabidopsis: A Laboratory Manual. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

[41] Yermack, D. (2013). Is Bitcoin a Real Currency? An Economic Appraisal. NBER Working Paper, [online] 19747. Available from:

[42] http://www.nber.org/papers/w19747 [accessed 13 September 2021].

[43] Yu, J. and Yuan, Y. (2011). Investor sentiment and the mean–variance relation. Journal of Financial Economics, 100(2), 367-381.

[44] Zheludev, I., Smith, R. and Aste, T. (2014). When Can Social Media Lead Financial Markets? Scientific Reports, 4(1).