ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

USING DATA MINING ALGORITHM FOR SENTIMENT ANALYSIS OF USERS' OPINIONS ABOUT BITCOIN CRYPTOCURRENCY

MOHAMMED ALGHOBIRI
MIS DEPARTMET,
KING KHALID UNIVSITY,
ABHA, SAUDI ARABIA
Email: maalghobiri@kku.edu.sa

ABSTRACT

Cryptocurrency has turned out to be one of the most significant currencies in the recent times due to their secureness, ease and value. Among the other cryptocurrencies available in the market, Bitcoin cryptocurrency is the most valuable and famous currency. A large number of people discuss about the Bitcoin currency on the internet and social media platforms as well. These discussions help in determining the importance of Bitcoin in terms of users' discussions about Bitcoin and can help in determining the value of Bitcoin in terms of people point of views about the topic. In this paper, sentiment analysis of the tweets of users on the topic of Bitcoin has been carried out. For this purpose, real-world twitter data set of Bitcoins is used. The data set has been divided into five separate sections for better comparative analysis, including overall extensive data analysis regarding tweets, retweets, tweets with mentions, tweets containing external links and also about the users who discuss regarding cryptocurrency of Bitcoin. A framework for sentiment analysis is proposed on the basis of Naïve Bayes sentiment classification algorithms which is widely used as a better option for text data. The proposed framework is capable to perform sentiment analysis of the tweets data. The results resemble that users' opinions are on the high positive side about Bitcoins and people mostly people represent the positive sentiment about Bitcoins. The results are evaluated using the standard performance evaluation measures including precision, recall, f1-score and accuracy.

Keywords: *Sentiment Analysis; Bitcoin; Classification; Naïve Bayes; Machine Learning; Proposed Framework; Users' Opinions;*

1. INTRODUCTION

Cryptocurrency is a digital asset that is designed to work as a medium of exchange. cryptocurrency uses cryptography to secure the transactions. Due to the secure nature of cryptocurrencies, it has become one of the most used currencies for transactions among the internet users for different purposes. Although the cryptocurrency prices are not strong in the market [1], people use it for their transactions due to secureness. Bitcoin cryptocurrency, being introduced in 2008, is based upon the digital transaction between the owner and receiver that are broadcasted through the p2p network [2]. Bitcoin has turned out to be one the most valuable currency as it has a worth over 150 billion US dollars [3]. The Bitcoin users can be divided into four categories including computer programmers, investors, Libertarians and criminals [4]. Due to the importance and enormous usage of Bitcoin currency, people all around the world discuss the details of Bitcoin currency on the internet. Among this, people around the social media platforms talks about the advantages and disadvantages of Bitcoin currency, and presents their likings and dislikes about the cryptocurrencies and Bitcoin. These people discussions express the importance of Bitcoin and general public opinions about Bitcoins and whether the people like or dislikes the Bitcoin. Among the other social media platform Bitcoin is regularly discussed by the twitter users. Analyzing the peoples' opinions about Bitcoin through different perspective can help in identifying the public behavior towards Bitcoin and importance of Bitcoin in terms of peoples' opinions. It also helps in accessing the insight knowledge of people's opinion about

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Bitcoin that can be used to find the flaws and security problems in the Bitcoin systems. Similarly, the sentiment analysis of tweets on Bitcoin can help in understanding the peoples' concerns regarding the currency that can be used to improve the Bitcoin structure therefore analyzing people opinions about Bitcoins is of excessive importance.

The social web analysis is usually divided into three main types: Web Content Mining, Web Structure mining and Web Usage Mining. The web content mining is related to the content analysis which is generated by the social web users. The main difference between conventional web (also known as World Wide Web or Web 1.0) and the Social Web (also known as Web 2.0) is that in the Web 1.0, the web owners are content generators while the common people are the common users who are content consumers only whereas in the Web 2.0, the common people are content consumers as well as content generators. The main and widely used social web channels include Facebook, Twitter, LinkedIn, Flicker, YouTube, etc. The web structure mining deals in such areas where the web pages linked to each other are analyzed. The social structure also deals how the users of the social web are linked with each other. How the social web users' use the social web channels is analyzed in the social web usage mining. The vast significance of the social web has motivated the researchers to work regarding the social web and a number of new research areas has emerged. One of the main such research domains is the opinion mining [5] which is also known as sentiment analysis. Opinion mining can be further divided into sub domains of subjectivity analysis [6], sentiment polarity as well as sentiment valence analysis [7], mixed opinion classification [8]. The sentiment analysis is also widely used in other related research areas of the social web such as the identification of influential bloggers [9].The identification of the influential users and bloggers has itself become an active research area in the social web [10] thanks to its significance that top people can be found who may help others in taking better decisions in life in diverse fields such as politics, social issues, e-marketing and online business [11]. These issues and applications make the significance of opinion mining even more importance in the field of data mining.

Furthermore, as the data around is increasing at a very rapid rate on the internet [12], data processing of textual data for sentiment analysis and other complex natural language processing tasks become very important. These tasks involves complex natural language processing tasks including information extraction and processing through different techniques [13]. This helps in understanding the natural language and helps in further processing of the data. Usually data processing tasks are applied by using machine learning and statistical approaches. These approaches have their own pros and cons although machine learning approaches perform better in different natural language processing tasks [14].

Analyzing peoples' opinions about different topics is one of the trending topics for the researchers. For this purpose, discussions among different social media platforms are used for analysis because social media platforms are useful resource to analyze the public views about different topics [15]. However, twitter is one of the major platforms where users among different fields of life discuss numerous topics, thus it has become one of the most helpful platform for sentiment analysis [16]. Twitter has also been used by the researchers to find the influential users for different purposes [17]. Sentiment analysis includes several Natural Language Processing (NLP) tasks because NLP tasks helps in understanding the language for different information retrieval purposes [12]. In this case, sentiment analysis is used to gain knowledge about the people opinions about Bitcoin through the help of their discussion about Bitcoin on the twitter.

For this purpose, real-world twitter dataset has been used for sentiment classification. The dataset is divided into 5 separate sections including overall tweets containing Bitcoin discussions, retweets, tweets in which the user mention other users, tweets that contains links to outside sources, and tweets containing discussions about cryptocurrencies. These sections help in determining the people opinion in a better way because this can help in determining how people behave towards Bitcoin in different perspectives and at different times. For sentiment analysis Naïve Bayes classification model is used to classify the tweets into positive and negative. Although Naïve Bayes is a simple classification model that perform classification on the basis of probability, it has been used by the by many researchers for sentiment classification [18][19].

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

After extracting the sentiment analysis results, different performance evaluation measures are used to measure the performance of the system.

The rest of the paper is divided as follows: Section II discusses the related work in the field of sentiment analysis and Bitcoin, Section III discusses the research methodology and framework proposed for the research, including performance evaluation methods, Section IV discusses the results and discussion of the research before finally concluding the paper in Section V.

2. RELATED WORK

Bitcoin and cryptocurrency is one of the trending topics for the researchers in the past five years. Researchers focused social media platform for different data mining and knowledge extraction tasks for Bitcoin. Mai F investigate the impact of social media in determining the values of Bitcoin. according to the author, social media sentiments are important source in determining the importance of Bitcoins [20]. Georgoula I et al performed the sentiment analysis to investigate the change in price of Bitcoins. they used machine learning model to perform sentiment analysis and investigate the reason behind the change in prices of Bitcoins [21]. Matta M et al investigate about the Bitcoin whether the positive sentiment of the people helps in increasing the prices of Bitcoins or not. according to their results there is a significant resembles between Bitcoin and google trends data however they didn't find detailed results of sentiment analysis and the effect of people opinions on the Bitcoin prices [22]. Researchers also work upon the sentiment analysis to predict the price of Bitcoin during a day based on the people sentiment towards the Bitcoins [23]. Kaminski J analyzed the correlation and causalities between Bitcoin and twitter posts in terms of people emotions. the paper investigates about the people emotions when the Bitcoin prices changes [24]. Bhargava GM et al performed sentiment analysis of cryptocurrencies including Bitcoin and other crypto currencies and analyzed the people opinions about different terms of cryptocurrencies [25]. Mai F et al discusses the impact of social media on Bitcoin and shows that the effect of social media platforms on the Bitcoin are effecting Bitcoins on hourly basis [26]. Moreover, user comments and replies are also helpful in determining the fluctuations in Bitcoin and other cryptocurrencies [27]. Kim YB et al [28] discusses the effects of users opinions towards predicting the fluctuations in Bitcoin currency. The authors proposed a currency value fluctuation system on the basis of user sentiments that predict the fluctuation in currency by analyzing the users sentiment towards a particular currency [29].

Now let us focus on how Naïve Bayes is important and how widely it has been used as the important data mining algorithm for classification. Naive Bayes classification is used for identifying seismic event and nuclear explosion [30]. self-adapting attribute weighting for Naive Bayes classification have also proposed by researchers using Artificial immune system [31]. Moreover, Naive Bayes classification techniques is also used with feature weighting. the experiments shows these weights rarely degrades the quality compared to simple Naive Bayes classification algorithm[32] . Naive Bayes classification techniques are also used for detecting DDOS attach by using the frequency based approach [33]. Naive Bayes classification is also used for ischemic stroke classification by using T1 weighted MRI scans[34]. passive indoor localization based classification is also performed by using Naive Bayes classification while the final results shows the algorithm performs as accurately as 86% [35]. negative class information in text classification is also performed by using naive Bayes classification and performed very good in the final results [36]. privacy preserving naive Bayes classification techniques are applied against substitution-then-comparison attack and calculating the server offline phases for the overall overhead of the computation[37].The use of the data mining algorithms has been done extensively in the past for sentiment analysis.

3. RESEARCH METHODOLOGY

The research methodology consists of several steps. These steps formed a framework that is presented in Figure 1. The proposed framework consists of 4 modules: dataset preparation, Dataset filtering, Naïve Bayes Sentiment Classification and Performance Evaluation. In the first module, the dataset is collected and cleaned for the dataset filtering. In the second module, the dataset is filtered based on five types. In the third module, Naïve Bayes sentiment classification model is applied to the dataset to classify the tweets on the basis of probability model. Finally,

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

in the fourth module, performance is evaluated for each of the output results in terms of standard performance measurement techniques.

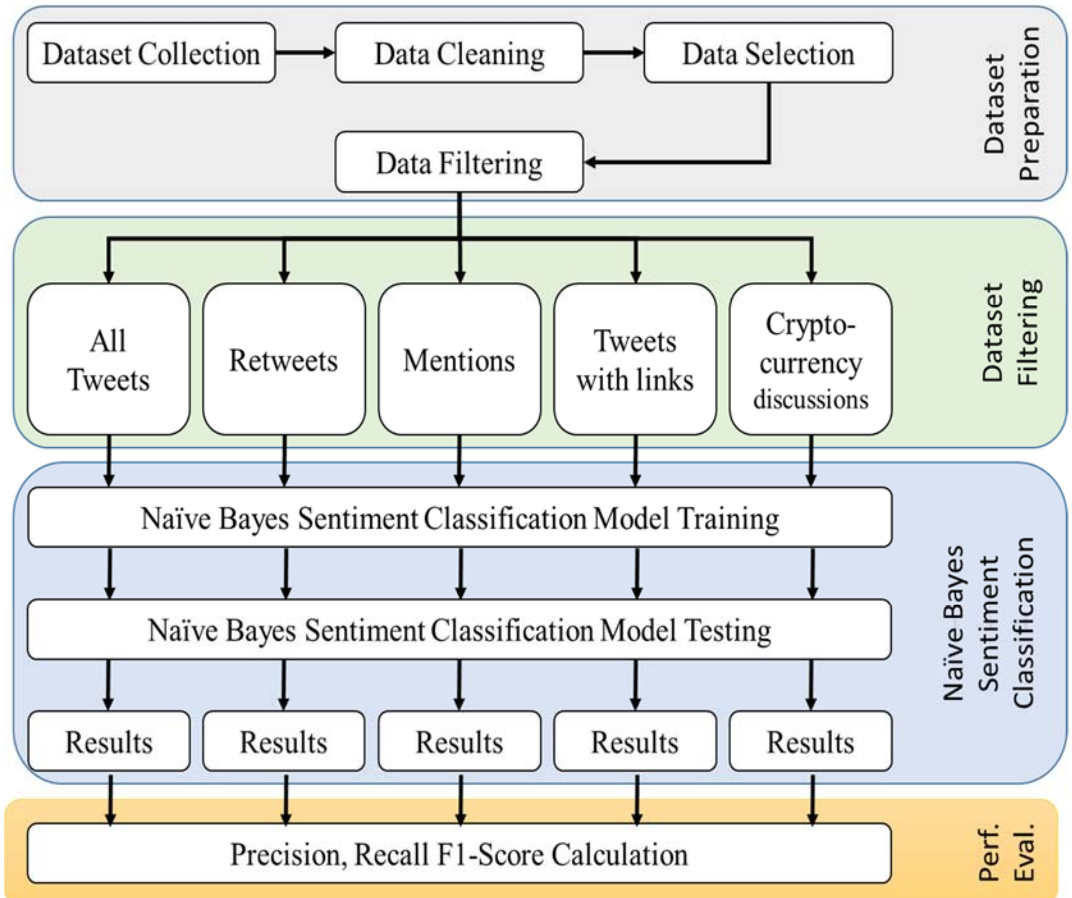

Figure 1 Proposed Research Methodology for the Sentiment Analysis

3.1. Dataset Preparation

The data set used in this research has been taken from Kaggle[1]. The dataset contains real-world dataset of tweets containing discussions related to Bitcoins. The dataset contains 50,859 tweets of the users from different backgrounds. The dataset contains complete details about the time when the tweet was made, the user is involved in tweet creation, text of the tweet and the sentiment score of the tweet. During the process of dataset preparation, the data are cleaned by performing stop-word removal and data selection. Tweets dataset contains a huge number of punctuations and stop-words; therefore, this process helps in better assessment of the tweets data. The data set selected for the research based upon the original dataset that contains all the information and no missing values. The details attributes of tweets dataset is shown in Table 1.

Table 1 Details of Tweets dataset attributes

Attribute	Details
id	Unique ID of tweet
text	Text of tweet
user	Username who made the tweet
score	Sentiment score for tweet

3.2. Dataset Filtering

The data filtering methods are applied to the data set to collect the data on the basis of different factors. The dataset is filtered on the basis of five types. The details are mentioned as follows.

I. Tweets collection containing all the tweets

[1] https://www.kaggle.com/skularat/bitcoin-tweets retrieved on 22-06-2018.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

II. Tweets collection containing tweets in which the users retweet the other tweets
III. Tweets collection containing tweets in which the users mentioned other users
IV. Tweets collection containing tweets in which the users shared outside links
V. Tweets collection containing tweets in which users' discus cryptocurrency

The major reasons behind the dataset division is to enable the system to perform sentiment analysis on the basis of different kinds of people discussions in which they discussed the Bitcoin through different angles. During the data analysis (I) tweet collection is useful to analyze the overall sentiment of the users towards Bitcoin. The total number of tweets for this analysis is 50,859. In Section II analysis the system analyzes the user's discussions that are actually retweets of the other tweets that shows the users' opinions when the users are responding to the other users and discuss the Bitcoin. The total number of tweets for this analysis is 27,544. In Section III analysis those tweets are discussed in which the users mentioned other users that shows how the tweeter users mention other users and discuss the Bitcoin. The total number of tweets for this analysis is 31,743. In Section IV, analysis those tweets are discussed in which the users shared the outside links that shows whether the users mention positive links about Bitcoins or negative links. This kind of analysis is also helpful in determining the user's tweets in which they advertise different services of the Bitcoins. The total number of tweets for this analysis is 33,077. In Section V, analysis those tweets are discussed in which the users are discussing the Bitcoin in context of cryptocurrency. The total number of tweets for this analysis is 5792. All these analyses are helpful in determining the user's opinions about Bitcoin from different contexts. Results and performance evaluation measures are gathered for each section separately that helps in the analysis of each section separately.

3.3. Naïve Bayes Supervised Learning Algorithm

In the third module Naïve Bayes Sentiment Classification model is applied on the tweets dataset to classify the tweets into positive (4) and negative (0). The unigram features of text analysis are used in sentiment classification for the sake of simplicity. The document term matrix is built on the basis of tokens present in each tweet. The term frequency is calculated for each of the matrix, then on the basis of the frequency the naive Bayes classification model is trained to predict the sentiment for the testing dataset. In the example, 80% data set is used for training purpose, while 20% data set is used for testing purposes. The Naïve Bayes classification model works on the basis of probabilities and the possibilities that a particular condition will occur or a particular item belongs to a class. The naïve Bayes classification model based upon the naive Bayes theorem and can be given as shown in the eq.1.

$$P(C_k|X) = \frac{p(C_k)\, p(x|C_k)}{p(x)} = \frac{\text{prior x likelihood}}{\text{evidence}} \quad (1)$$

The above equation shows that p (C_k, x_1,x_2,….,x_n). Therefore, the chain rule for repeated values and the conditional probabilities can be given as shown in eq.2.

$$p(C_k, x_1, x_2, \dots, x_n) = p(x_1|x_2, x_3, \dots . x_n, C_k)p(x_2|x_3, \dots . x_n, C_k)$$

$$\dots p(x_{n-1}|x_n|C_k)p(x_n|c_k)p(C_K) \quad (2)$$

On the basis of the equation 1 and equation 2, the naïve Bayes model is built to classify the sentiments into positive and negative.

3.4. Performance Evaluation Measures

The performance of the system is evaluated using the standard performance evaluation techniques. For this purpose, Precision, Recall, F1-measure and accuracy are used as standard performance evaluation technique [38]. If TP denotes the sentiments that are correctly classified by the system as positive, TN denotes the sentiments that are correctly classified by the system as negative, FP denotes the sentiment that is identified by system as positive but are actually negative, FN denotes the sentiments that are identified by the system as negative, but are actually positive then the precision, recall, f1-score and accuracy can be given as shown in equation 3,4,5 and 6 respectively.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (3)$$

Journal of Theoretical and Applied Information Technology
30th April 2019. Vol.97. No 8

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (4)$$

$$\text{F1} - \text{Score} = 2 * \frac{\text{Precision. Recall}}{\text{Precision} + \text{Recall}} \quad (5)$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (6)$$

4. RESULTS AND DISCUSSIONS

After applying the framework for sentiment analysis in the tweets, the results are generated on the basis of separate sections. Different sections demonstrate the different results of sentiments. The results of different subsets of datasets are separately discussed.

4.1. The Analysis covering Whole Data Set

During the first experiments, the sentiment analysis is performed over the whole data set. During this analysis, the positive and negative sentiments are computed. The results of this section are shown in

Figure 2. The results show that the total number of negative tweets about Bitcoin is 12%, which are very less than the positive tweets. Furthermore, the number of positive tweets about Bitcoin is 45%. This demonstrates that the about half of the people who talk about Bitcoin, actually have positive sentiments about it.

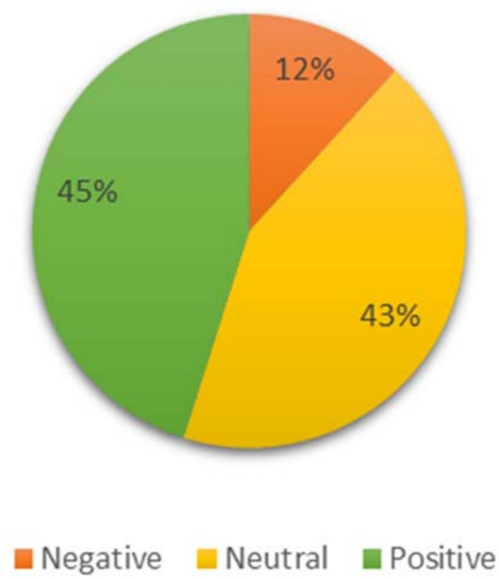

Figure 2 Results of overall Dataset

4.2. Retweets Analysis

During these experiments, the dataset used for experiments contains those tweets that are retweets of the other tweets. This section helps in determining the users' behavior and opinions towards Bitcoin when replying to the other users. The results of this section are shown in Figure 3. The results demonstrate that the number of negative tweets is 10%. Retweets are the tweets that actually contain replies or retweets of the other tweets. Although sentiment analysis of retweets is a complex domain, the results demonstrate that the number of tweets with negative comments is 10%. This resembles that 90% tweets about Bitcoin are not negative.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

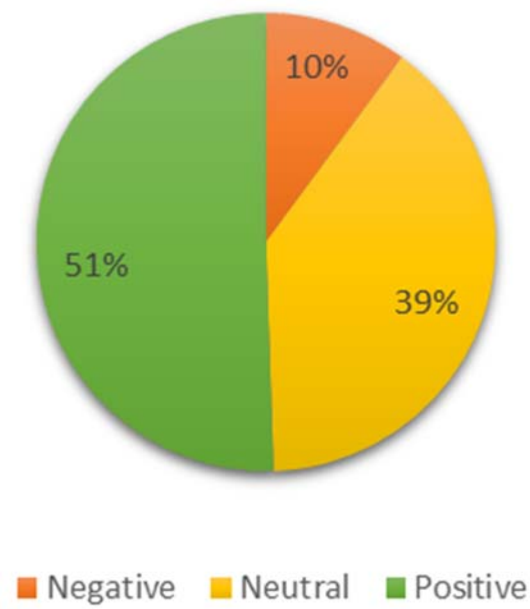

Figure 3 Results of Retweets

4.3. Mentions Analysis

During these experiments, the dataset used for experiments contains those tweets in which the users mention other users. This section helps in determining the users' sentiments when mentioning other users and talking about Bitcoin. The results of this section are shown in [

Figure 4. The results demonstrate that this section has similar results with respect to the results of the previous section. 51% people have positive sentiments about Bitcoin. Moreover, the results show that 10% people's opinions are negative when users mention other users and in their tweets.

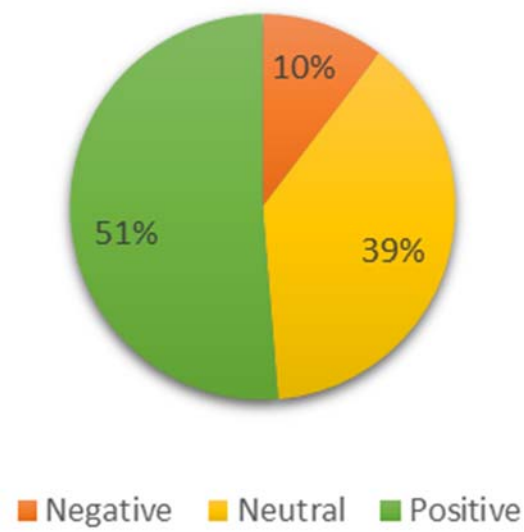

Figure 4 Results of Tweets with mentions

4.4. Outer links Analysis

Sharing outside links is considered an important factor in social media analysis. During these experiments, the dataset used contains the tweets in which the users shared outside links in their tweets. This section also helps in determining the users that are actually advertising the Bitcoin through outside links. The results of this section are shown in

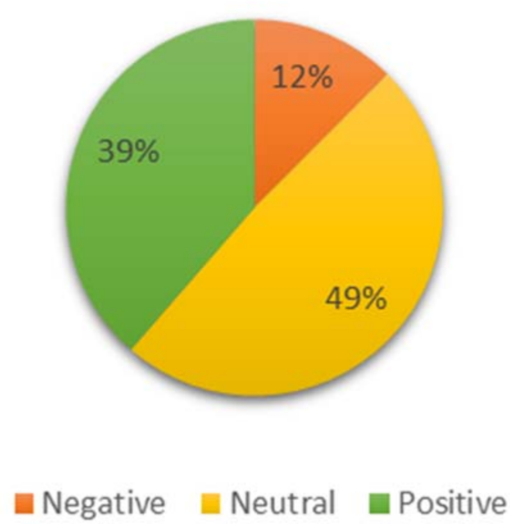

Figure 5. The tweets with outer links are mostly those tweets in which the users advertise different websites in the domain of Bitcoin.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

The results show that in this case, as expected, the number of positive sentiments is far less than the other sections. This is due to the tweets with advertisements in which usually there are no sentiments. As a result, the number of neutral tweets is far more than the other sections. Although the negative tweets with the outer links are 12%, the overall highest opinions of the users also lie in this dataset as the 49% users shows positive tweets while mentioning the outer links. Some of these users mention other website links that provide different bitcoin services or in order words these tweets contain high number of advertisements that is the major reason behind the highest number of positive opinions in this dataset.

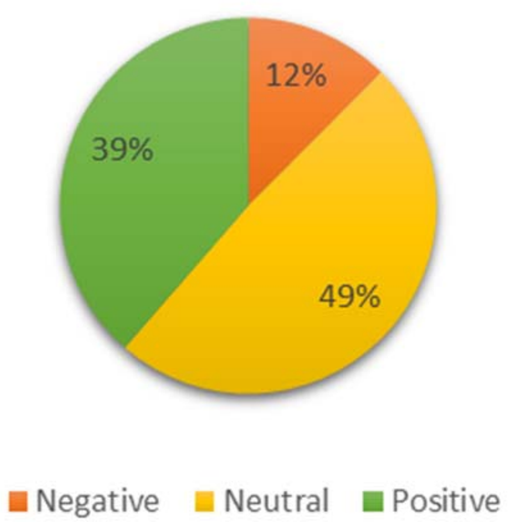

Figure 5 Result of Tweets with links

4.5. The Analysis of the Context of Cryptocurrency

In the final section, the experiments are performed on the data set that contains the tweets in which users discussed the Bitcoin in context of cryptocurrency. This dataset is retrieved using the word "cryptocurrency" in the users' tweets. The results of this section are shown in

Figure 6. The results of this section show that peoples' opinions are really positive towards cryptocurrencies. The number of negative comments is 7%, while the number of positive comments is 54%. These types of tweets usually contain discussions in which the word 'cryptocurrency' is present. This means, these discussions have technical background and users discussed in technical sense. Therefore, less negative sentiment results of these types of tweets show that overall users have positive opinions about bitcoin as a cryptocurrency.

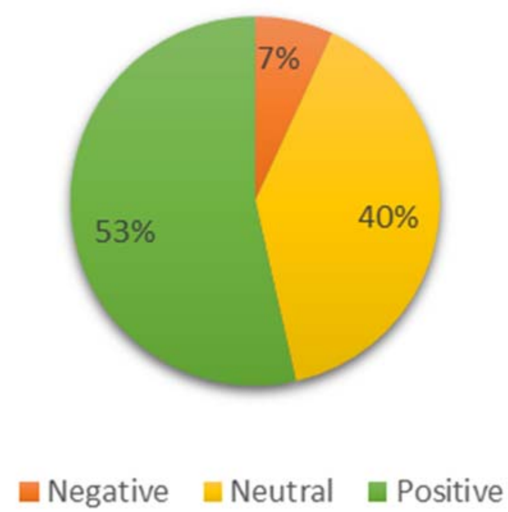

Figure 6 Results of Tweets with cryptocurrency Discussions

4.6. Comparative Analysis of the Proposed Framework

This section demonstrates the performance of the system during sentiment classification. The evaluation measure techniques are applied in all the sections so the difference in performance measure can be computed and compared with the others. The results of performance evaluation of all the sections are presented in

Figure 7**Error! Reference source not found.**. The results show that the performance of overall dataset is good as the system is able to achieve an accuracy of 77%. Moreover, the recall rates are higher because there are more number of positive sentiments in the overall data set. During the performance evaluation of the system, there are some tweets in which text analysis is easy and the system is able to produce good results while there are some tweets in which the system doesn't produce good results. Furthermore, the results of the overall dataset produce good results because the system is able to train the model much better in this case.

Although the results of precision, recall, f1-score and accuracy are highly identical, the results of all datasets of tweets show there is a little difference between different types of tweets dataset. However, the overall accuracy remains between 67% to 77% among which the results of tweets having cryptocurrency is 67% while the overall results of accuracy is 77% probably due to the better training and testing dataset. Moreover, as the overall sentiments of the dataset are positive and there are very few negative opinions of the users about the bitcoin, the recall results are overall very good as the results remain from 87% to 93%.

This is due to the very low number of negative opinions thus very low impact of false negative classification. While in case of precision, the results are a little low as it remains from 70% to the 77% because there are high number of wrong prediction of positive and negative opinions. These results are reflected in f1-score as the results remain from 78% to 84% that are very good considering the complexity of the domain and simplicity of the classification algorithm.

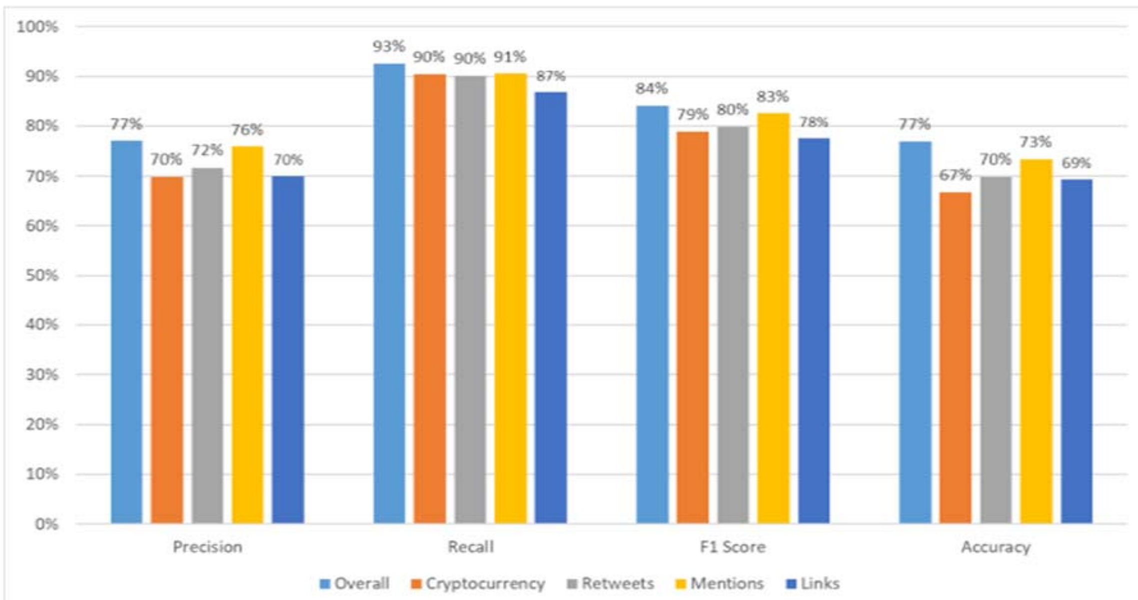

Figure 7 The comparative analysis of the results using Performance Evaluation Measures

5. CONCLUSION

The sentiment analysis framework for the tweets and used real-world tweets dataset of Bitcoins is proposed. Detailed text analysis is performed on the dataset by dividing the dataset further into five parts and performing analysis of each part separately. The results show that the common users 'sentiments are mostly positive towards Bitcoin and the twitter users are mostly commenting positive tweets while discussing Bitcoin. Although the Bitcoin cryptocurrency is not considered as the most liked currency by financial experts, the sentiment analysis results show that generally the people are positive towards it.

Moreover, the separate analysis of different subsets of dataset presents a better picture of the sections where the people sentiments are positive or not. For instance, 54% people sentiments are positive when talking about Bitcoin in context of cryptocurrency while only 39% of peoples' sentiments are positive when talking about Bitcoin while sharing a link on twitter. During the experiments we face some problems in training the model by separate dataset and some limitations of the naïve Bayes model. The results can further be improved by proposing a sentiment analysis solution by using other machine learning models and perform further analysis. Therefore, in future, we aim to work in the same domain by further expanding the analysis in terms of users and analyze the sentiments with respect to the users. We also aim to generate ontology of the tweets in the topic and use other machine learning models to improve the results.

References

[1] R. C. Phillips and D. Gorse, "Predicting cryptocurrency price bubbles using social media data and epidemic modelling," in *2017 IEEE Symposium Series on Computational Intelligence (SSCI)*, 2017, pp. 1–7.
[2] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," 2008.
[3] M. Conti, S. K. E, C. Lal, and S. Ruj, "A Survey on Security and Privacy Issues of Bitcoin," *IEEE Commun. Surv. Tutor.*, pp. 1–1, 2018.
[4] A. Yelowitz and M. Wilson, "Characteristics of Bitcoin users: an analysis of Google search data," *Appl. Econ. Lett.*, vol. 22, no. 13, pp. 1030–1036, Sep. 2015.
[5] B. Liu and L. Zhang, "A Survey of Opinion Mining and Sentiment Analysis," in *Mining Text Data*, Springer, Boston, MA, 2012, pp. 415–463.
[6] "Using machine learning techniques for subjectivity analysis based on lexical and non-lexical features," *ResearchGate*. [Online]. Available: https://www.researchgate.net/publication/318225737_Using_machine_learning_techniques_for_subjectivity_analysis_based_on_lexical_and_non-lexical_features. [Accessed: 26-Jul-2018].
[7] F. G. Contratres, S. N. Alves-Souza, L. V. L. Filgueiras, and L. S. DeSouza, "Sentiment Analysis of Social Network Data for Cold-Start Relief in Recommender Systems," in *Trends and Advances in Information Systems and Technologies*, 2018, pp. 122–132.
[8] H. U. Khan, "Mixed-sentiment Classification of Web Forum Posts Using Lexical and Non-lexical Features," *J Web Eng*, vol. 16, no. 1–2, pp. 161–176, Mar. 2017.
[9] U. Ishfaq, H. U. Khan, and K. Iqbal, "Identifying the influential bloggers: a modular approach based on sentiment analysis," *J. Web Eng.*, vol. 16, no. 5–6, pp. 505–523, 2017.
[10] H. U. Khan *et al.*, "Modelling to identify influential bloggers in the blogosphere: A survey," *Comput. Hum. Behav.*, vol. 68, pp. 64–82, Mar. 2017.
[11] H. U. Khan, A. Daud, and T. A. Malik, "MIIB: A Metric to identify top influential bloggers in a community," *PloS One*, vol. 10, no. 9, p. e0138359, 2015.
[12] A. Mahmood, H. U. Khan, Zahoor-ur-Rehman, and W. Khan, "Query based information retrieval and knowledge extraction using Hadith datasets," in *2017 13th International Conference on Emerging Technologies (ICET)*, 2017, pp. 1–6.
[13] A. Mahmood, H. U. Khan, F. K. Alarfaj, M. Ramzan, and M. Ilyas, "A Multilingual Datasets Repository of the Hadith Content," *Int. J. Adv. Comput. Sci. Appl. IJACSA*, vol. 9, no. 2, pp. 165–172, 2018.
[14] F. Sebastiani, "Machine Learning in Automated Text Categorization," *ACM Comput Surv*, vol. 34, no. 1, pp. 1–47, Mar. 2002.
[15] R. Khan, H. U. Khan, M. S. Faisal, K. Iqbal, and M. S. I. Malik, "An Analysis of Twitter users of Pakistan," *Int. J. Comput. Sci. Inf. Secur.*, vol. 14, no. 8, p. 855, 2016.
[16] E. Kouloumpis, T. Wilson, and J. D. Moore, "Twitter sentiment analysis: The good the bad and the omg!," *Icwsm*, vol. 11, no. 538–541, p. 164, 2011.
[17] U. Ishfaq, H. U. Khan, and K. Iqbal, "Modeling to find the top bloggers using Sentiment Features," in *2016 International Conference on Computing, Electronic and Electrical Engineering (ICE Cube)*, 2016, pp. 227–233.
[18] J. Song, K. T. Kim, B. Lee, S. Kim, and H. Y. Youn, "A novel classification approach based on Naïve Bayes for Twitter sentiment analysis," *KSII Trans. Internet Inf. Syst. TIIS*, vol. 11, no. 6, pp. 2996–3011, 2017.
[19] S. Tan, X. Cheng, Y. Wang, and H. Xu, "Adapting Naive Bayes to Domain Adaptation for Sentiment Analysis," in *Advances in Information Retrieval*, 2009, pp. 337–349.
[20] F. Mai, Z. Shan, Q. Bai, X. (Shane) Wang, and R. H. L. Chiang, "How Does Social Media Impact Bitcoin Value? A Test of the Silent Majority Hypothesis," *J. Manag. Inf. Syst.*, vol. 35, no. 1, pp. 19–52, Jan. 2018.
[21] I. Georgoula, D. Pournarakis, C. Bilanakos, D. Sotiropoulos, and G. M. Giaglis, "Using Time-Series and Sentiment Analysis to Detect the Determinants of Bitcoin Prices," Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 2607167, May 2015.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[22] M. Matta, I. Lunesu, and M. Marchesi, “Bitcoin Spread Prediction Using Social and Web Search Media.,” in *UMAP Workshops*, 2015.
[23] V. Karalevicius, N. Degrande, and J. De Weerdt, “Using sentiment analysis to predict interday Bitcoin price movements,” *J. Risk Finance*, vol. 19, no. 1, pp. 56–75, 2018.
[24] J. Kaminski, “Nowcasting the Bitcoin Market with Twitter Signals,” *ArXiv14067577 Cs*, Jun. 2014.
[25] M. G. Bhargava and D. R. Rao, “Sentimental analysis on social media data using R programming,” *Int. J. Eng. Technol.*, vol. 7, no. 2.31, pp. 80–84, 2018.
[26] F. Mai, Q. Bai, Z. Shan, X. S. Wang, and R. Chiang, “The impacts of social media on Bitcoin performance,” 2016.
[27] Y. B. Kim *et al.*, “Predicting Fluctuations in Cryptocurrency Transactions Based on User Comments and Replies,” *PLOS ONE*, vol. 11, no. 8, p. e0161197, Aug. 2016.
[28] Y. B. Kim, J. Lee, N. Park, J. Choo, J.-H. Kim, and C. H. Kim, “When Bitcoin encounters information in an online forum: Using text mining to analyse user opinions and predict value fluctuation,” *PLOS ONE*, vol. 12, no. 5, p. e0177630, May 2017.
[29] Y. B. Kim, S. H. Lee, S. J. Kang, M. J. Choi, J. Lee, and C. H. Kim, “Virtual World Currency Value Fluctuation Prediction System Based on User Sentiment Analysis,” *PLOS ONE*, vol. 10, no. 8, p. e0132944, Aug. 2015.
[30] L. Dong, X. Li, and G. Xie, “Nonlinear Methodologies for Identifying Seismic Event and Nuclear Explosion Using Random Forest, Support Vector Machine, and Naive Bayes Classification,” *Abstract and Applied Analysis*, 2014. [Online]. Available: https://www.hindawi.com/journals/aaa/2014/459137/abs/. [Accessed: 26-Jul-2018].
[31] J. Wu, S. Pan, X. Zhu, Z. Cai, P. Zhang, and C. Zhang, “Self-adaptive attribute weighting for Naive Bayes classification,” *Expert Syst. Appl.*, vol. 42, no. 3, pp. 1487–1502, Feb. 2015.
[32] L. Jiang, C. Li, S. Wang, and L. Zhang, “Deep feature weighting for naive Bayes and its application to text classification,” *Eng. Appl. Artif. Intell.*, vol. 52, pp. 26–39, Jun. 2016.
[33] R. F. Fouladi, C. E. Kayatas, and E. Anarim, “Frequency based DDoS attack detection approach using naive Bayes classification,” in *2016 39th International Conference on Telecommunications and Signal Processing (TSP)*, 2016, pp. 104–107.
[34] J. C. Griffis, J. B. Allendorfer, and J. P. Szaflarski, “Voxel-based Gaussian naïve Bayes classification of ischemic stroke lesions in individual T1-weighted MRI scans,” *J. Neurosci. Methods*, vol. 257, pp. 97–108, Jan. 2016.
[35] Z. Wu, Q. Xu, J. Li, C. Fu, Q. Xuan, and Y. Xiang, “Passive Indoor Localization Based on CSI and Naive Bayes Classification,” *IEEE Trans. Syst. Man Cybern. Syst.*, pp. 1–12, 2017.
[36] Y. Ko, “How to use negative class information for Naive Bayes classification,” *Inf. Process. Manag.*, vol. 53, no. 6, pp. 1255–1268, Nov. 2017.
[37] C. Gao, Q. Cheng, P. He, W. Susilo, and J. Li, “Privacy-preserving Naive Bayes classifiers secure against the substitution-then-comparison attack,” *Inf. Sci.*, vol. 444, pp. 72–88, May 2018.
[38] Y. Yang, “An Evaluation of Statistical Approaches to Text Categorization,” *Inf. Retr.*, vol. 1, no. 1–2, pp. 69–90, Apr. 1999.