Analysis and Classification of Crime Tweets

Citation for published version:
Lal, S, Tiwari, L, Ranjan, R, Verma, A, Sardana, N & Mourya, R 2020, 'Analysis and Classification of Crime Tweets', *Procedia Computer Science*, vol. 167, pp. 1911-1919. https://doi.org/10.1016/j.procs.2020.03.211

Digital Object Identifier (DOI):
10.1016/j.procs.2020.03.211

Link:
Link to publication record in Heriot-Watt Research Portal

Document Version:
Publisher's PDF, also known as Version of record

Published In:
Procedia Computer Science

Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 167 (2020) 1911–1919

Procedia Computer Science

www.elsevier.com/locate/procedia

International Conference on Computational Intelligence and Data Science (ICCIDS 2019)

Analysis and Classification of Crime Tweets

Sangeeta Lal[a], Lipika Tiwari[a], Ravi Ranjan[a], Ayushi Verma[a], Neetu Sardana[a], Rahul Mourya[b]

[a] *Jaypee Institute of Information Technology, Noida, Uttar Pradesh, India*

[b] *Heriot-Watt University, Edinburgh*

Abstract

Nowadays social Networking and micro-blogging sites like Twitter are very popular and millions of users are registered on these websites. The users present on these website use these websites as a platform to express their thoughts and opinions. Our analysis of content posted on Twitter shows that users often post crime related information on Twitter. Among these crime related tweets some tweets are the crime messages that need police attention. Detection of such tweets can be beneficial in utilizing pattroling resources. The analysis of the data present on these websites can have an enormous impact. In this paper,the work is done on analyzing Twitter data to identify crime tweet that need police attention. Text mining based approach is used for classification of 369 tweets into crime and not-crime class. Classifiers such as Naive Bayesian, Random Forest, J48 and ZeroR are used. Among all of these four classifiers, Random forest classifier give the best accuracy of 98.1%.

Keywords: Twitter; Crime Detection; Random Forest; J48; ZeroR

1. Introduction

1.1 Introduction

Social Networking sites like Facebook, Twitter are very popular these days. Millions of users are registered on these websites. These users are exchanging their thoughts, opinions, news, personal information in different forms like text, photos, video on these platforms. Social media acts as a source of public opinion and thus helps in knowing and understanding what the public is talking about. The data present on social media is beneficial in studies where public opinion is required. Earlier, organizations used to pay a lot of money to market research companies to organize polls and conduct focus group studies to get the sort of information that consumers now readily post on

10.1016/j.procs.2020.03.211

social media platforms. The analysis of this information has helped to achieve several business goals and has also led to the building of new tools and technologies. For example, in past these, data from social media have been analyzed for various purposes like to predict box office results [7], election results [8], and stock market trends [9].

Past studies [5][12] on twitter data show that people have started using twitter to express and report various types of crimes. For example, Table 1 show several crime related tweets posted by users on Twitter. In Example 1, the user has posted information about a fake Facebook ID and has requested the police to arrest the specified person that is behind creating this Facebook ID. This is a form of cybercrime that is reported on Twitter. In example 2, the user has posted information about forgery and harassment by a Nigerian-Indian couple. These examples show, how users use twitter for posting data related to crime. The analysis of such data can have a huge impact to improve the crime handling. For example, a classification tool that can separate crime tweet from non-crime tweet can be beneficial in effectively channelizing the resource from police department and in handling the crime on time. As the police department can focus on crime tweets and ignore other irrelevant tweets. This kind of tool can be of great importance in developing nations. Because developing nations like India have huge population but limited resources to address all the crimes.

Table 1 Example of crime-related tweets on Twitter.

S. No	Crime Type	Tweet Content
1	Cyber crime	Cyber Branch of @DelhiPolice must take notice of this Facebook ID and **catch hold** this person immediately. @Facebook also needs to suspend this particular ID. I wonder why Facebook has not done this till now [1]
2	Forgery/Fraud	**Forgery** & **harassmen**t, please look into this @DelhiPolice @DCPEastDelhi [4]

Figure 1: Non-crime related tweet sent to @DelhiPolice Twitter handle

1.2 Challenges

Crime tweet classification is an important problem. But manually identifying crime tweets is quite challenging. There are three main challenges in crime tweet identification:

- **Large number of tweets**: Quite a large number of tweets are posted on twitter every minute. Identifying crime tweet manually from these large number of tweets in quite tiresome.
- **Large number twitter handles:** One can think that tweets posted on a particular twitter handle, for example, @DelhiPolice, handle are crime related Tweets in Delhi. Our analysis reveals that there are

several such twitter handles for example @DelhiPolice (Official Twitter account of Delhi Police), @CPDelhi (Official Twitter Handle of the Commissioner of Police, Delhi), @LtGovDelhi (Official Twitter Account of Lieutenant Governor of Delhi) and many more. Identifying all such twitter handle is quite challenging.

- **Noisy tweets:** Several times people post tweets to @DelhiPolice and similar handle to congratulate them on their achievements. These tweets create noise while finding the crime tweets that needs immediate attention. Fig.1 is showing such a tweet where the user has sent a tweet to Delhi police. In this tweet, the user is giving details about a session conducted in a school. This tweet is not crime related and it is important to filter out such tweets. Filtering out such tweets manually is not only difficult but it is also a time-consuming task.

An automated tool that can classify crime and non-crime tweets can be of great importance. The aim of this paper is to build a model that can separate crime tweet and non-crime tweets. In context of this paper, crime tweets are considered as the tweet that need police action whereas non-crime tweets are the other general tweets. In this paper, the work is done on applying text mining-based approach for crime tweet classification. Text mining-based approaches have been found useful in several application such election results prediction from tweets. Hence, it has been hypothesized that content present in tweet can be used to train machine learning classifiers to identify crime tweets. For example, tweets presented in Table 1 show presence of words like "catch hold", "Forgery & harassment". These words can be used as an input to machine learning classifiers.
In this paper, content of tweets and text mining approach were used for tweet classification. Classifiers such as Naive Bayesian, Random Forest, J48 and ZeroR are used for classifying 369 tweets into crime and not-crime class. All the classifiers found to be effective in classifying crime tweets with Random forest classifier giving the best accuracy of 98.1%.

2. Related Work and Research Contribution

In this section, previous research carried out by other researchers closely related to the work presented in this paper. After that, the unique and novel research contributions made by this work in comparison to the previous studies has been described.

2.1 Related Work

In this subsection, the review of the research papers is presented which is closely related to the defined area.

- <u>Twitter analysis in general</u>
  Meyer et al. [10] utilized the information posted on twitter to find breaking news. The author identifies various events like attack, earthquake, fire etc. Longueville et al. [11] analyze tweet related to forest fire in occurred in France in the year 2009. The author report that the analysis of the content posted on twitter was accurate and synchronized to fire event. Asur et al. [13] proposed a linear regression-based model for box office results prediction. The author conducted experiments on 3 Million tweets from Twitter. Experimental results showed that the linear regression model outperformed the Hollywood stock exchange. Bollen et al. [14] analyze content from twitter using OpinionFinder and GoogleProfiile tools. Their experimental results show that inclusion of public moods is very beneficial in improving the accuracy of DIJA predictions.

- <u>Twitter analysis for crime</u>
  Beiji et al. [5] work on predicting crime hot-spots. It is observed that crime tends to occur in clusters. The proposed approach consists of three phases: analysis of videos, prediction of crime, and mapping of crime. The author performs analysis of their proposed approach on violent scene detection (VSD) 2014 dataset. The neuro-fuzzy method has been used for event prediction. Wang et al. [12] perform prediction of hit and

run cases from social media. The author has used semantic analysis, LDA and dimensionality reduction approach for model building. The experiment is conducted on real world dataset. The proposed model outperformed the baseline classifier and give better ROC curve ad compared to baseline approach.

2.2 Research Contributions

In context to related work, this work makes following novel and unique research contributions:

- The authors perform analysis of crime tweets and non-crime tweets to identify if the content of the Tweet can be used for automated classification of crime and non-crime tweets of not. Using information from this analysis, authors apply text mining-based approach for classification of crime and non-crime tweets.

- The authors collect real world dataset from the Twitter interface. The authors present experimental results of four machine learning classifiers for the task of crime tweet classification. The authors compare the performances of these classifiers using four different parameters.

3. Data Visualization

Initially data visualization of crime-related tweets has been performed. Since the area of crime is so vast, primarily the analysis of the content random crime related tweets from twitter was done. The map in Fig. 2 demonstrates the heat map for crime-related tweets. The map has been made using the free online tool available on Trendsmap.com [6]. Relevant hashtags like (#crime #murder #rape #harassment #violence #robbery #criminal #fraud #police). Different color represents different hashtags spread across India. For example, the graph shows that several cases of “robbery and rape”, “Robbery and murder” are reported in places near to Delhi and southern part of India, respectively.

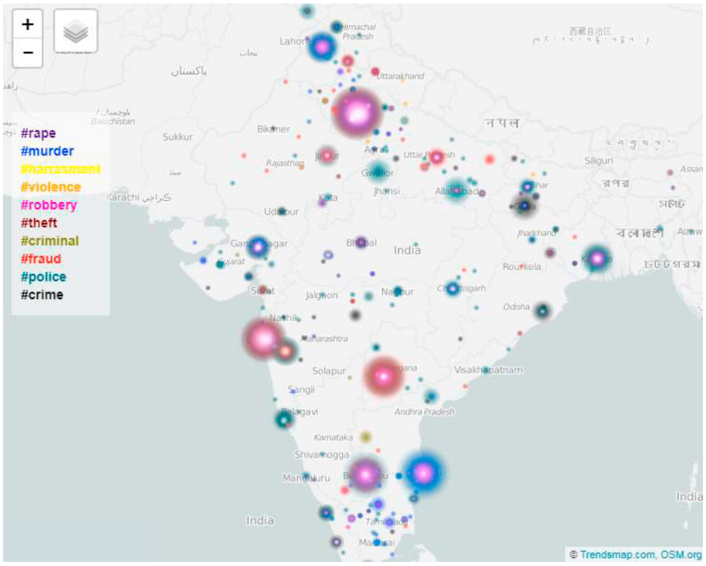

Fig. 2: Heat map showing a huge presence of crime-related tweets in India

4. Model Building

This section explains the details about the prediction model. Firstly, a motivating example is given, to show uses of text mining in crime and non-crime tweet classification.

4.1 Motivating Examples

Several examples of tweets on crime has been collected. It is noticed that these tweets have distinguishing words like fraud, arrested, fake, kidnapped. Fig. 3 shows a user tweet. In this tweet the user is complaining about a fraud recruiter. This tweet consists of a word fraudulent, which indicates that this tweet can be related to crime. Hence, it is hypothesized that that can be used to predict crime and non-crime tweets. Now to further analyse our hypothesis, a database has been created manually consisting of 20 crime tweets and compute their tf-idf [15] score. Table 2 shows the tf-idf score of crime tweets. This Table gives an indication that crime tweets has certain distinguishing words for example, harassment, forgery, fraud. These words can be used to train machine learning classifiers about crime tweets. Motivated by this fact, in this work, machine learning is used to predict crime related tweets using its content.

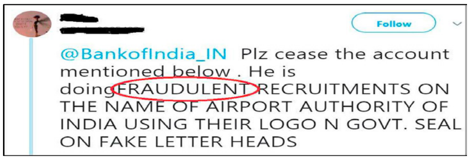

Fig. 3: A user tweet showing distinguishing words for crime tweet classification

4.2 Experimental Dataset Collection

The real time dataset from Twitter has been collected. The 500 tweets consisting of 230 (non-crime) or 270 (crime tweets). posted on @DelhiPolice has been collected.

Table 2 TF-IDF scores of crime tweets

S. No	Word	TF-IDF Score
1	Harassment	0.532
2	Forgery	0.532
3	Fraud	0.407

4.3 Data Pre-Processing

Data cleaning or pre-processing is an important step of prediction model building. In this work, the textual data from twitter to classify crime and non-crime tweets is used. Textual data consists of several noise like #, slangs, stop words. Using textual data directly can lead to misleading classification results. Hence, data pre-processing is applied and cleaning approaches before applying machine learning algorithms. The following six data pre-processing steps in our work has been applied:

- Sentence Segmentation: In this, the tweet is split into tokens or individual words. For example, 'Forgery & harassment, please look into this @DelhiPolice @DCPEastDelhi' after applying sentence segmentation changed to ['Forgery', '&', 'harassment,', 'please', 'look', 'into', 'this', '@DelhiPolice', '@DCPEastDelhi']. With the help of this, other cleaning techniques can be applied on the tokenized sentence.

- Hashtag Expansion: In this, the joint hashtags are expanded. The hashtags are split into individual words. Using 'NLTK' and 're', a simple algorithm has been created to split the hashtags. For example, in the tweet "Special night checking for #IndependenceDay yields 2 snatchers/auto-lifters", "#IndependenceDay" has been split into "Independence Day".
- Stop Words Removal: Here, non-content bearing terms like `the', `is', `what' from the tweets are removed. These terms are called as stop-words and these terms do not provide any sentiment for crime analysis.
- Acronym and Slang Treatment: In this, the acronyms and common slangs which are used on twitter are expanded. For example, acronyms like 'asap' was expanded into 'as soon as possible' and slangs like 'b4' was expanded into 'before'.
- Punctuation Removal: The non-content bearing terms like `the', `is', `what' from the tweets are removed. These terms are called as stop-words and they do not provide any sentiment for crime analysis.
- TF-IDF conversion: Using term frequency and data frequency measurements to highlight those words that may be related to crime:(Currently, some words other than those that are crime related are also being displayed as the algorithm works only on term frequency and data frequency.).The idea behind the usage of this function is that crime-related words constitute a very small fraction of characters in the entire tweet text. Hence, the lower frequency fraction would mean that the word is related to crime.

4.4 Algorithm

//Algorithm 1: This Algorithm to performs data pre-processing

Input: Raw Tweet (t)
Output: Cleaned Tweet (t_c)
Function dataPreProcessing(t)
  words[] =**SentenceSegmentation**(t) //tokenize the sentence
  WordsExpandedHash[] = **ExpandHashTag**(words[])
  WordStopRemoved[] = **removeStopWords**(WordsExpandedHash[])
  WordsExpandSlang[] = **FindAndReplaceSlang**(WordStopRemoved[], SlangFile)
  CleanedTweet = **FindAndRemovePuctuation** (WordsExpandSlang [], punctuationArr[])
  TFIDF-Vector [] = **TFIDFConversion** (CleanedTweet)
return TFIDF-Vector []
//This function is used to perform hash tag expansion
Input: tweetWords []
Output: WordsHashTagsRemoved []
Aim: Tweet content without hashtags
ExpandHashTag(tweetWords []):
    for i in len(tweetWords []):
      if(**isHashTag**(tweetWords [i]))
        temp = **removeHash**(tweetWords [i])
        tempHashTagArray[]=**split_hashtag**(temp)
        WordsHashTagsRemoved[]=**insertHashTagToOrginalPosition**(tweetWords[], tempHashtagArray[])
return WordsHashTagsRemoved[]

5. Experimental Environment

The authors manually collected tweets from the Twitter interface. Two of the authors verified content of the tweets and labelled them as crime or not crime. The authors consider the tweets which are labelled same by both of the authors to conduct our experiments. The authors use the WEKA for conducting all of their experiments. The authors use version 3.8 of the WEKA tool in this work. For all the classifiers the authors use default parameters set in the WEKA tool. The authors run experiment on a system having Windows 10 operating system, 8GB RAM and Corei5 processor.

Table 3 Pre-processing output

S. No	Pre-processing Step	Initial	After pre-processing step
1	Sentence Segmentation	"The police caught the thieves with for murdering a man for revenge, further investigation is in progress!"	'The','police','caught','the','thieves ','with','for','murdering','a','man',' for','revenge',',','further','investiga tion','is','in','progress','!'
2	Hashtag Expansion	#MumbaiPolice #SafeIndia	'Mumbai', 'Police', 'Safe', 'India'
3	Acronym and Slang treatment	"The police caught him b4 he could run away"	'The','police','caught','him','before ','he', 'could','run','away'
4	Stop word removal	"The police caught the thieves for murdering a man for revenge, further investigation is in progress!"	'police','caught','theives','murderin g','man','revenge',',','further','inve stigation','progress','!'
5	Using tfidf function to predict roughly those words that may be crime related	This method is not perfect as it just gives us those words whose frequency is less in the tweet considering the fact that crime-related words appear a smaller number of times in a tweet. Also, it is obvious that a tweet would contain other non-crime related words to properly express something.	harassment = 0.532

6. Results

In this section, the results obtained by various classifiers are discussed. Two RQ's in this section have been answered. In the RQ1, the results of different classifier have been compared with respect to four different metrics. In RQ2, the ROC analysis of different classifiers has been performed.

RQ 1: What is the performance of various classifiers for classifying crime and non-crime tweets?

Motivation: Analysis of various classifiers can provide insights which type of classifier is most suitable for detecting crime tweets.

Approach: In this RQ 4 different classifiers are selected: Naive Bayesian (NB), Random Forest, J48 and ZeroR. The accuracy of all the classifiers is tested on the imbalanced dataset. Further, 10-fold cross-validation is used.

Results: Table 4 shows the results obtained by various classifiers. Results show that RF classifier give the best accuracy value of 98.1%. The ZeroR classifier performs the worst and give accuracy of 61.5%. RF classifier took the maximum time of 7.24 Seconds in model building.

Table 4: Classification results on the imbalanced dataset

Classifier	Precision	Recall	Accuracy	F-measure	Time (Seconds)
NB	97.9	97.8	97.83	97.8	1.2
RF	**98.2**	**98.1**	**98.1**	**98.1**	**7.24**
J48	97.0	97.0	97.0	97.0	2.5
ZeroR	37.8	61.5	61.5	49.9	0.01

RQ2: Which classifier performs the best in ROC analysis?

Motivation: The ROC analysis of classifier is important as it can provide insights about TP rate and FP rate on the Imbalanced dataset. Because of imbalanced dataset a classifier can give good accuracy just be predicting every datapoint as negative. ROC analysis of the dataset classifier provide insight about how accurately a class is detecting positive and negative class instances.

Approach: The ROC curve analysis of all the four classifiers have been performed using WEKA tool.

Results: Fig. 4 show the ROC analysis of all the four classifiers. This figure shows that classifiers NB, J48, and RF give good ROC score, and hence, are able to classify the crime tweet with good accuracy. Whereas, the classifier ZeroR performs random prediction and is not effective the detecting crime and non-crime tweets.

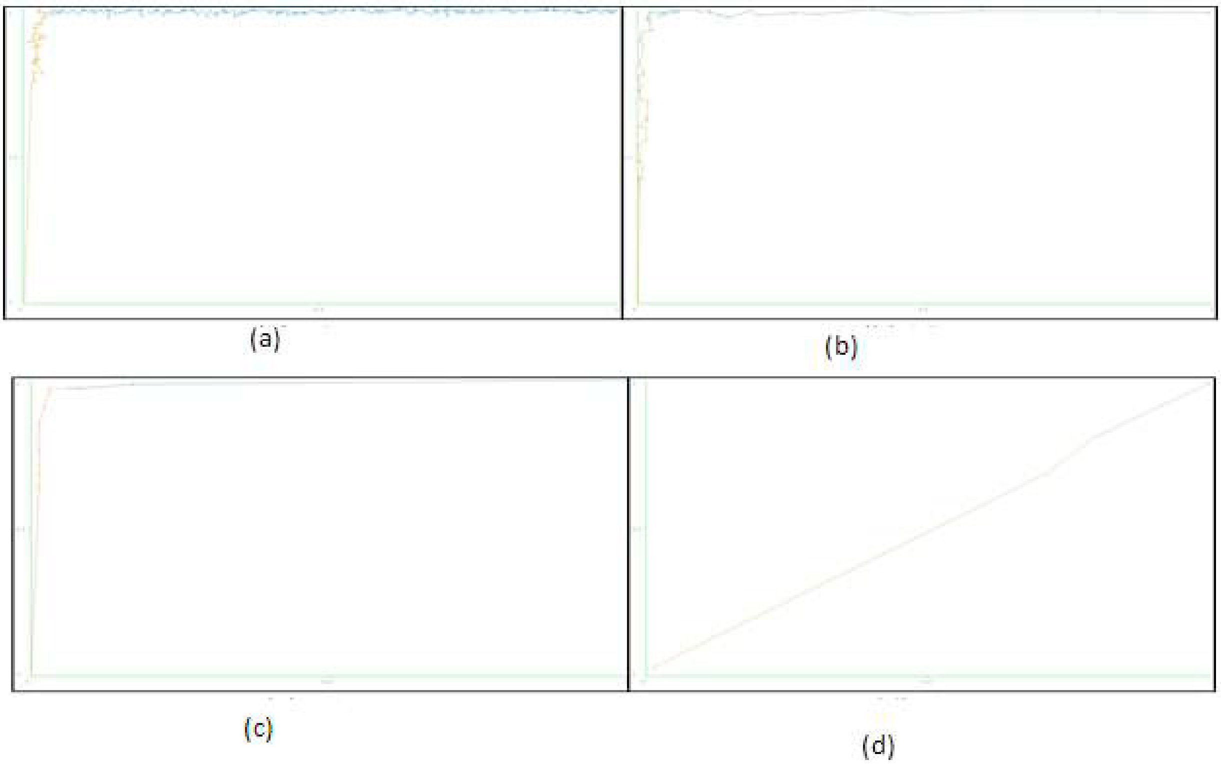

Fig.4: ROC curve for all the four classifiers: (a) NB, (b) RF, (c) J48, (d) ZeroR

7. Conclusion and Future Work

The analysis of the content posted on Twitter shows that users often post crime information on Twitter. It is believed that automated classification or identification of crime tweets from the Twitter can be beneficial in better management of the crime reported on Twitter. This can help in reducing the time it takes to resolve the crime which can be beneficial for both victim and police officials. Motivated by this fact, in this paper, the analysis of tweet contents is performed. In this paper, text mining-based approach has been used for crime tweet classification. Four classifiers have been used; Naive Bayesian, Random Forest, J48 and ZeroR. All the classifiers except ZeroR found to be effective in classifying crime tweets with Random forest classifier giving the best accuracy of 98.1%.

In Future, it is planned to extend this work on multiple dimension. First, use more classifiers to test the effectiveness of the proposed approach. Second, adding the location information to the crime tweet to help the police in identifying the crime location. Third, applying ensemble learning based approach for crime tweet classification. Four, applying NLP techniques like parts-of-speech tagging for improving the feature extraction. Fifth, performing comparative analysis of the approach used in this paper with the other approaches.

References

[1] Twitter, https://twitter.com/iAbhishekkBJP/status/1032306902862983168 [accessed on 2/10/2018]

[2] Twitter, https://twitter.com/dcp_southwest/status/1032685782085947392 [accessed on 2/10/2018]

[3] Twitter, https://twitter.com/DcpDwarka/status/1031157545191333888 [accessed on 2/10/2018]

[4] Twitter, https://twitter.com/Ankit_Tyagi01/status/1027471441149808640 [accessed on 2/10/2018]

[5] Beiji, Zou, Nurudeen Mohammed, Zhu Chengzhang, and Zhao Rongchang (2017). "Crime hotspot detection and monitoring using video based event modeling and mapping techniques." *International Journal of Computational Intelligence Systems***10 (1)**: 962-969.

[6] Map plottting tool, https://www.trendsmap.com/ [accessed on 21/1/2019]

[7] Gaikar, Dipak Damodar, Bijith Marakarkandy, and Chandan Dasgupta (2015). "Using Twitter data to predict the performance of Bollywood movies." *Industrial Management & Data Systems***115 (9)**: 1604-1621.

[8] Tumasjan, Andranik, Timm O. Sprenger, Philipp G. Sandner, and Isabell M. Welpe (2010). "Predicting elections with twitter: What 140 characters reveal about political sentiment." *In Fourth international AAAI conference on weblogs and social media.*

[9] Zhang, Xue, Hauke Fuehres, and Peter A. Gloor(2011). "Predicting stock market indicators through twitter "I hope it is not as bad as I fear"." *Procedia-Social and Behavioral Sciences***26**: 55-62.

[10]Meyer, Brett, Kevin Bryan, Yamara Santos, and Beomjin Kim(2011). "TwitterReporter: Breaking News Detection and Visualization through the Geo-Tagged Twitter Network." *In CATA*:84-89

[11] De Longueville, Bertrand, Robin S. Smith, and Gianluca Luraschi(2009) "Omg, from here, i can see the flames!: a use case of mining location based social networks to acquire spatio-temporal data on forest fires." *In Proceedings of the 2009 international workshop on location based social networks:*73-80

[12] Wang, Xiaofeng, Matthew S. Gerber, and Donald E. Brown(2012) "Automatic crime prediction using events extracted from twitter posts." *In International conference on social computing, behavioral-cultural modeling, and prediction, Springer, Berlin, Heidelberg*:231-238.

[13] Asur, Sitaram, and Bernardo A. Huberman. "Predicting the future with social media." In Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology **01**:492-499. IEEE Computer Society.

[14] Bollen, Johan, Huina Mao, and Xiaojun Zeng (2011)"Twitter mood predicts the stock market." *Journal of computational science***2 (1)**: 1-8.

[15] Han, Jiawei, Jian Pei, and Micheline Kamber(2011) Data mining: concepts and techniques. *Elsevier.*