Scientific Journal of Impact Factor (SJIF): 5.71

e-ISSN (O): 2348-4470
p-ISSN (P): 2348-6406

International Journal of Advance Engineering and Research Development

Volume 5, Issue 02, February -2018

Rumor Detection with Twitter and News Channel Data Using Sentiment Analysis and Classification

Dr. Dinesh B. Vaghela[1], Divya M. Patel[2]

Information Technology, Shantilal Shal Engineering College, Bhavnagar, Gujarat, India
Information Technology, Shantilal Shal Engineering College, Bhavnagar, Gujarat, India

Abstract —*Nowadays peoples are going towards social media increasingly to fetch the information and to share their opinion on social media. As there is rapid diffusion of information on social media, the information posted on social media spread so fast and easy. This information posted on social media not always right or not truthful to make sense. An advantage of social media is that all the people can share information and also gives their opinions on that platform. The drawback of such rapid diffusion of information is that false information are also spread. As the rumors are spreading on Twitter and other social media so fast and easier. We need to provide some solutions to detect such rumors. In this paper, our detection approach is based on the classification. Our detection approach is divided into three parts: Preprocessing, Sentiment Analysis and Classification. Also we are comparing different supervised learning techniques/methods for getting better and accurate detection of rumors. We are using one more external module i.e. news websites verification and comparing sentiment score of our proposed method and sentiment score of this external module.*

Keywords- *Twitter, Rumor detection, Sentiment analysis, Sentiment Score, News Websites, and Rumor Analysis.*

I. INTRODUCTION

A. What is Rumor?[2]

Before we proceed further, first we need to understand the definition of Rumor. Rumors are the "unverified and untrusted events" which harms the people's emotions. We define a rumor to an unverified statement that starts from one or more sources and spreads over time.

A rumor can end in three ways: it can be resolved as either true, false or remain unresolved.

B. How it will affect?[1]

As the information spread on social media so fast and easy by one or more users/sourcesover time. This diffusion of information can change the scenario towards the real world because of some rumors are affecting to people's life. Most of people believe that whatever information spread on social media it has to be true rather than to check it on verified news sites or news channel. For example, a millions of peoples believe that President of US Barack Obama was injured during the Boston Marathon. A lot of money and time had spent by US government to recover from this rumor.

C. Introduction about Twitter[1][2]

As the users on social media increasing day by day we know that, nowadays peoples are more reliable on social media to get/access news than the tradition media because it's fast and easy. It is faster and easily accessible because of wider users on social platform.

The main motive behind why I had chosen Twitter platform is that, twitter is most popular platform in which common peoples are also share their information and their opinions with each other. All the celebrities, politicians, etc. are very powerful peoples which are comfortable to share their status and their updates with this social media.

A Traditional systems like television channels, radio channels, etc. are the media to disseminate information but the social media takes the place of traditional media nowadays because of the social media is faster and easily accessible. This makes Twitter a major improvement over the existing news dissemination systems. Twitter is thus an interesting service to scan in order to detect news before they are published by more traditional news media.

The rest of the paper organized as follow: Section II describes the previous related work carried out by different researchers which is useful for this paper. Section III describes the comparative analysis of literature we have referred in section II. Section IV describes proposed method to detect the rumors on Twitter by us. Section V describes Experimental results. Section VI and VII provide the conclusion and the future work of our paper respectively.

II. RELATED WORK

A lot of work has been done in the research field to detect the rumors on Twitter. In this section, we provide the review of the latest literature developments on rumors detection on Twitter.

Sardar Hamidian et al. [6] focus on the problem of detecting rumors in Twitter data. They used label-independent methods to generate features that depends on the tweet content.Their experiment is based on two condition: Single step Rumor Detection and classification (SRDC) and Two step (TRDC). In both SRDC and TRDC, features are divided into classes and according to that classes tweets it apply for classification. In the experiment they used WEKA platform for training and testing their proposed method using the J48 classifier.

SuchitaJain et al. [7] focus on automatic detection of rumors on twitter in real time. In their approach they said that the verified News Channel accounts gives the more credible information than the general public accounts. Their approach is based on sentiment and semantic analysis to detect the rumors.

SahanaV P et al. [8] focus on automatically detect the rumors spreading on Twitter and identify its source. They took topic "London Riots in 2011" and used some of the rumored tweets posted and some non-rumored tweets. They used Weka tool for classification. They achieves best accuracy for J48 classification algorithm. Also they propose an algorithm to finding the origin of spreading rumors on Twitter.

ZhiweiJin et al. [9] focus on to detect rumors spreads in political events. They propose an algorithm to detect rumors on topic 2016 U.S. presidential election. They analyzed rumors tweets from the followers of two presidential candidates: Hillary Clinton and Donald Trump. They detect rumor tweets by matching large amount of tweets related to president election with verified rumor articles. For classification they use different word matching method i.e TF-IDF, BM25, Word2Vec and Doc2Vec.

QiaoZhang et al. [10] focus on the detection of rumors automatically by using the combination of previously used literature's shallow features and the implicit features that they have newly proposed. They said that in many previous literature, Shallow features is the features that cannot distinguish between rumor messages and normal messages in many cases. For the classification of label they used supervised learning method such Support Vector Machine, Random Forest, etc.

Yan Zhang et al. [11] focus on autoencoder to perform rumor detection. They used Sina Weibo which is the most popular microblog in China. They use several self-adapting thresholds which are calculated based on the property of each recent Weibo set, which can help in rumor detection. In addition, they also discuss how the different number of hidden layers of autoencoder can affect the detection performance.

III. COMPARATIVE ANALYSIS

Following table shows summary of papers which are referred for literature review:

Sr. No.	Title	Algorithm/Technique Used	Observations	Limitations
1.	Rumor Detection and Classification for Twitter Data [6]	J48 decision tree Classifier with SRDC and TRDC	Output with Preprocessing data gives the low accuracy compared with the output without Preprocessing data. Achieved F-Measure more than 0.82 and 0.85 on a mixed and Obama rumor data sets which is one of the rumor topic they selected, respectively.	Real- time data were considered. Need to better preprocessing process.
2.	Towards Automated Real-Time Detection of Misinformation on Twitter [7]	sentiment and semantic analysis	Used some example of rumors and according to verified news channel and general public tweets ratio it detects rumors by using sentiment and semantic analysis.	Feature selection/extraction part is missing.

3.	Automatic detection of Rumored Tweets and finding its Origin [8]	J48 decision tree Classifier	Recall rate is given high accuracy 0.877	Focused only on one specific rumor topic. Real-time twitter data were not considered.
4.	Detection and Analysis of 2016 US Presidential Election Related Rumors on Twitter [9]	TF-IDF and BM25, Word2Vec and Doc2Vec, Lexicon matching	For rumor classification task: F1-measures BM25 achieved best accuracy of 0.82. For rumor identification task: BM25 algorithm gives accuracy i.e. 0.799.	Focused only on specific topic i.e. "2016 US President Election related rumors".
5.	Automatic Detection of Rumor on Social Network [10]	Support vector machine	First, they used content-based, user-based, content-user-based features and the method combined with content and user features is better than others, with 7.1% improvement in precision and 6.3% improvement in recall rate. Second, they used Shallow-Content-Based, Implicit-Content-Based, Shallow-User-Based and implicit-User-Based features and Implicit-Content-Based method had improvement compared to Shallow-Content-Based method with 10.5% improvement in precision and 4.7% in recall rate.	User credibility. Detection of rumors on the Chinese micro-blogging services.
6.	Detecting Rumors on Online Social Networks Using Multi-layer Autoencoder [11]	Autoencoder (Artificial Neural Network)	They used some threshold based on their original dataset and among those threshold, the threshold med+1.5(Q3 - Q1) achieve the accuracy of 88%, f1 of 82% and FPR of 7%.	Detection of rumors on the chienese micro-blogging services. Performance of autoencoder with 2 hidden layer gives best performance.

Table 1: Comparative Analysis of Literature

IV. PROPOSED METHODOLOGY

From inspiring with the among research papers, we thought that the rumors are spreading so fast and easy. The people connected on social media might not get that the information are shared on the social platform is true or not. From this, we proposed a method for detecting rumors on one of the social media i.e. Twitter.

In our proposed methodology, following basic steps are required to detect the rumors [5].

A. Dataset Collection

The collection of data will be fetched using tweepy library available in python programming language. We will use Twitter Streaming API provided by twitter. The data collected will be in JSON format.

The collection of News sites RSS feed will be used to fetch news related the topic given by the user.

B. Data Preprocessing

In this step,unwanted noise like not needed words in dataset and general words such as stop wordswill removed. Also punctuation marks, URLS and stop words will removed using various library available in python. Also, usingStreamer algorithm for all the similar words like connect, connecting, connected, connects are stemmed and only the prefix or root word is taken from that word. Using this, avoid confusion for algorithm and canimprove accuracy. Thisprocessed data will further used as feature selection process and this features are used further for classification.

C. Feature Selection

Feature selection is a process where you can select those features automatically in your data that contribute most to the prediction variable or output in which you are interested.

Based on our requirements, we have pre-selected features. We are not applying any methods or any algorithms to select the features.

D. Classification

The various classification techniques are used to classify the data. Also very important to measure the accuracy of data. We will compare output of different classification techniques to measure accuracy of our system.

As in the literature, they are using some of the classification techniques i.e. Decision Tree, Naïve Bayes and Support Vector Machine (SVM) to decide the class label as Rumor or Not Rumor. So, we also take this classification techniques to classify class label and trying to get better accuracy to detect rumors.

The proposed method for the Rumor Detection is shown in figure 1.Our proposed approach is divided into three step: 1) Pre-processing, 2) Sentiment Analysis, and 3) Classification.

- **In first step**, we are going to **preprocess on the real-time tweets** to determine the topic about which the given input tweet is posted.
- **In second step**, we are finding **tweet's sentiment polarity** of each tweets by using sentiment score.
- **In final step**, we are going to apply this sentiment score as an input to the different classification algorithm.

We are using one more external module i.e news websites verification and comparing sentiment score of our proposed method and sentiment score of this external module which is score of news websites. If both gives the same result then we can say that our approach gives the better accuracy. This comparison approach also provide the verification about the rumor topic. This flow is shown in following figure 2.

For dataset collection, we are going to collect tweets from twitter using Twitter streaming API. We are preprocessing on tweets and going to decide the features for classification. For feature classification, we are going to use Weka tool with different classifier and compare result among them. So, we are trying to get better result than the existing one.

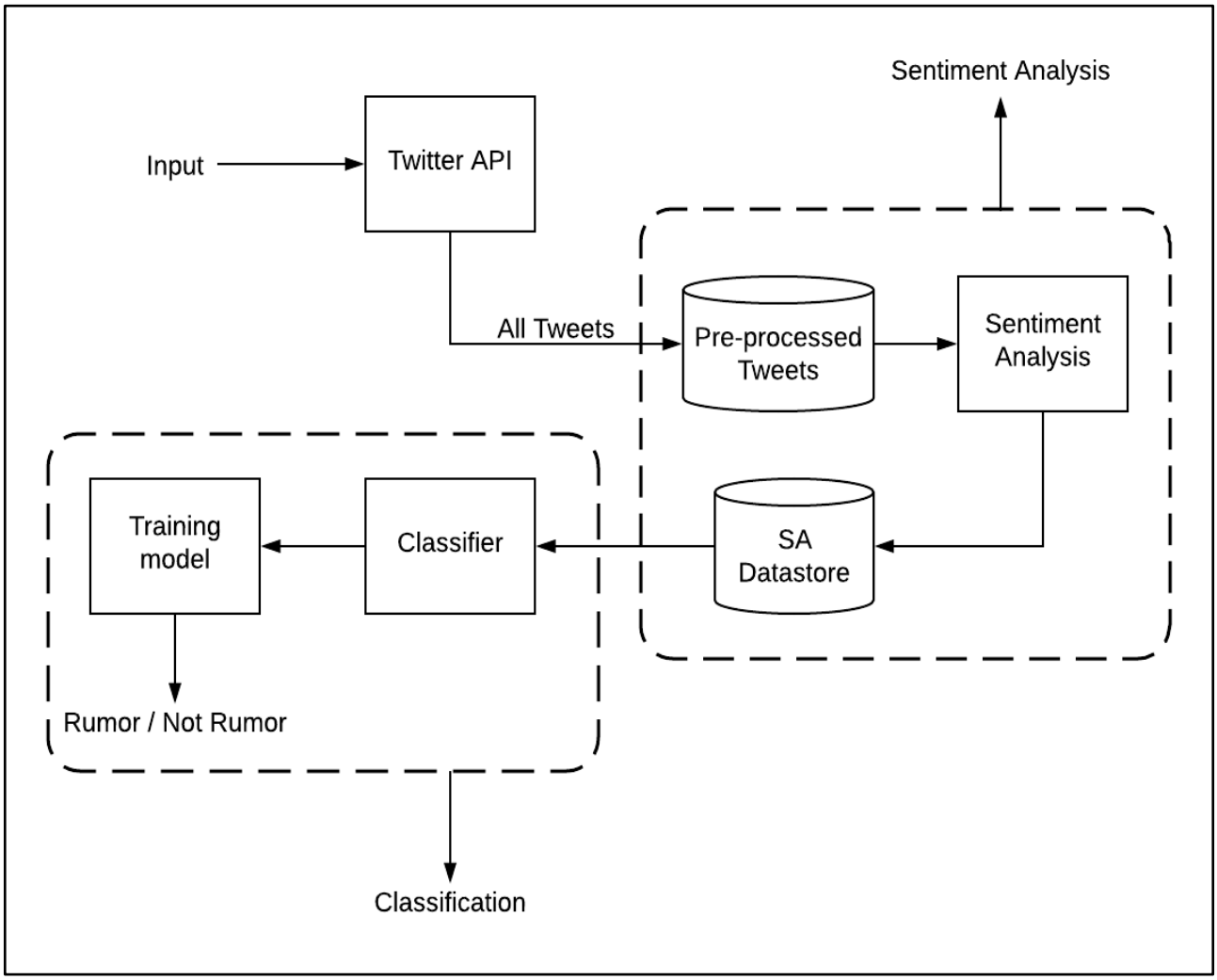

Figure 1. **Proposed Method**

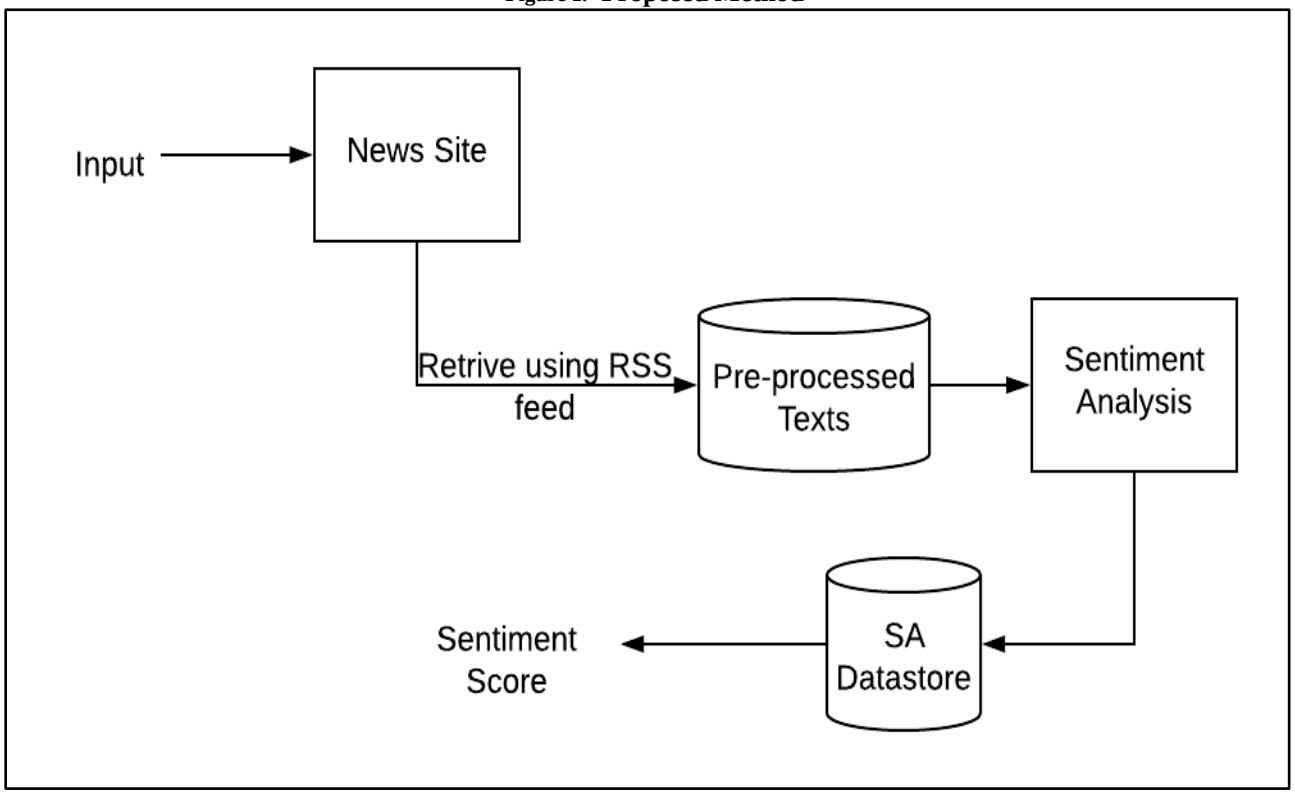

Figure 2. **Using News Website Approach**

V. EXPERIMENTAL RESULTS

Following shows the experimental results about some topics and the sentiments related to that topic.

Tweets in Table format

	Tweets	len	ID
0	Warm birthday greetings to my valued colleague...	106	946205354123190272
1	In Shimla, relished coffee at the Indian Coffe...	139	945935906518605824
2	I thank the people of Shimla for the warm welc...	74	945928082946265089
3	The swearing in ceremony of the Council of Min...	138	945927462088622080
4	Congratulations to Shri Jairam Thakur and all ...	140	945926733202571265
5	I would once again like to thank the people of...	139	945571209063759873
6	Attending today's oath taking ceremony in Guja...	140	945567041750048768
7	Political leaders, Chief Ministers of various ...	139	945565955656007680
8	People from all walks of life joined the oath ...	140	945565266573791233
9	Congratulations to Shri @vijayrupanibjp, Shri ...	140	945564622735646720

	Date	Source	Likes	RTs	SA
0	2017-12-28 02:25:10	Twitter for iPhone	39530	5319	1
1	2017-12-27 08:34:29	Twitter Web Client	39308	6774	1
2	2017-12-27 08:03:24	Twitter Web Client	15810	2876	1
3	2017-12-27 08:00:56	Twitter Web Client	20904	4026	1
4	2017-12-27 07:58:02	Twitter Web Client	22333	3460	1
5	2017-12-26 08:25:18	Twitter Web Client	35511	5989	0
6	2017-12-26 08:08:45	Twitter Web Client	30944	5302	0
7	2017-12-26 08:04:26	Twitter Web Client	16019	3167	0
8	2017-12-26 08:01:41	Twitter Web Client	11382	2520	0
9	2017-12-26 07:59:08	Twitter Web Client	10141	2396	1

Percentage of positive tweets: 36.5%
Percentage of neutral tweets: 58.0%
Percentage de negative tweets: 5.5%

Figure 3. Sentiment score about rumor topic.

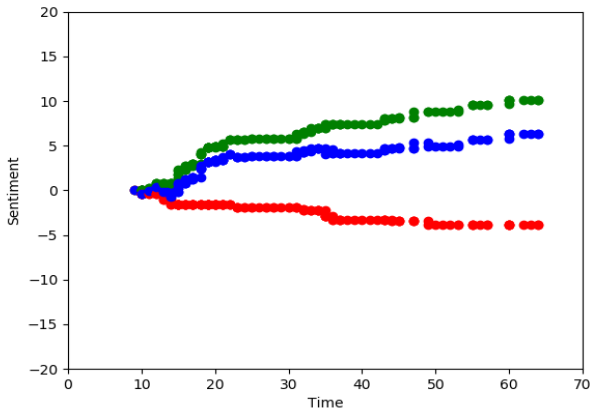

Figure 4. Graph about rumor topic.

Topic	Decision tree	Naïve Bayes	SVM
Changing your Facebook display picture to support DigitalIndia is equal to supporting internet.org	76.99%	74.33%	77.56%
Beef is not being served in Kerala House	60.78%	59.66%	61.30%

Table 2: Accuracy analysis of Rumor topic using classification technique

VI. CONCLUSION

After the study of different research paper on rumor detection, different methods to are used to detect rumors. We pre-identified features based on our requirement. There are many classifiers available for detecting rumors.We have taken Decision Tree, Naïve Bayes and Support Vector Machine classification techniques to classify class label as Rumor or Not Rumor. Using above results in both topic we achieves high accuracy of SVM classifier than the other two. So we can say that our proposed method gives the better accuracy for SVM classifier. Also used one external module i.e. News Websites verification to verify with our proposed method sentiment score to give better accuracy.

VII. FUTURE WORK

The news site verification module is under the construction. This work remaining for implementation. This rumor problem is seen in almost all major social networks not only on Twitter. Our proposed method/algorithm extended with some modifications.In future we plan to work on above directions.

REFERENCES

[1] Anubrata Das, Moumita Roy, Soumi Dutta, Saptarshi Ghosh, Asit Kumar Das. "Predicting Trends in the Twitter Social Network: A Machine Learning Approach", Springer International Publishing Switzerland, pp. 570-581, 2015.
[2] Soroush Vosoughi, PhD Thesis, "Automatic Detection and Verification of Rumors on Twitter", June 2015.
[3] Palash Sharma, Aishwarya Agrawal, Lalit Alai, Akshay Garg. "Challenges and Techniques in Preprocessing for Twitter Data", International Journal of Engineering Science and Computing, pp. 6611-6613, April 2017.
[4] Aditi Gupta, PhD Thesis, "A survey on Analyzing and Measuring Trustworthiness of User-Generated Content on Twitter during High-Impact Events",April 2013.
[5] Jiawei Han, Micheline Kamber.Data Mining: Concepts and Techniques, Second Edition.
[6] Sardar Hamidian and Mona Diab. "Rumor Detection and Classification for Twitter Data",The Fifth International Conference on Social Media Technologies, Communication, and Informatics, pp. 71-77, SOTICS 2015.
[7] Suchita Jain, Vanya Sharma and Rishabh Kaushal. "Towards Automated Real-Time Detection of Misinformation on Twitter", Intl. Conference on Advances in Computing, Communications and Informatics (ICACCI),pp. 2025-2020, IEEE 2016.
[8] Sahana V P, Alwyn R Pias, Richa Shastri, and Shweta Mandloi. "Automatic detection of Rumoured Tweets and finding its Origin", Intl. Conference on Computing and Network Communications (CoCoNet'15), pp. 607-612, IEEE 2015.

[9] Zhiwei Jin, Juan Cao, Han Guo, Yongdong Zhang, Yu Wang, and Jiebo Luo. "Detection and Analysis of 2016 US Presidential Election Related Rumors on Twitter", Springer International Publishing AG 2017, pp. 230–239, Springer 2017.

[10] Qiao Zhang, Shuiyuan Zhang, Jian Dong, Jinhua Xiong, and Xueqi Cheng. "Automatic Detection of Rumor on Social Network", Springer International Publishing Switzerland 2015, pp. 14-24, Springer 2017.

[11] Yan Zhang, Weiling Chen, Chai Kiat Yeo, Chiew Tong Lau, Bu Sung Lee, "Detecting Rumors on Online Social NetworksUsing Multi-layer Autoencoder", IEEE Technology & Engineering Management Conference (TEMSCON), pp. 1-5, IEEE 2017

