International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-9 Issue-1, October, 2019

Detection of FAKE NEWS on SOCIAL MEDIA using CLASSIFICATION Data Mining Techniques

Deepak Sharma, Shilpa Singhal

Abstract: ***In today's world social media is one of the most important tool for communication that helps people to interact with each other and share their thoughts, knowledge or any other information. Some of the most popular social media websites are Facebook, Twitter, Whatsapp and Wechat etc. Since, it has a large impact on people's daily life it can be used a source for any fake or misinformation. So it is important that any information presented on social media should be evaluated for its genuineness and originality in terms of the probability of correctness and reliability to trust the information exchange. In this work we have identified the features that can be helpful in predicting whether a given Tweet is Rumor or Information. Two machine learning algorithm are executed using WEKA tool for the classification that is Decision Tree and Support Vector Machine.***

Keywords: Decision tree, Support Vector Machine (SVM), Data Mining, WEKA.

I. INTRODUCTION

In Today's world Social media is currently a place where huge amount of data is generated continuously. Nowadays, any breaking news appear first on microblogs, before making it through to other traditional media . Hence, social media websites are rich sources of information which have been successfully considered for the analysis of daily social phenomena, such as belief, opinion, and sentiment in online communication and Twitter is one of the most popular social media platforms with more than 250 million users. Accessibility, speed and ease-of-use have made it a valuable platform to read and share information. There are several fake news in past that increased the complexities on internet. A message "UNESCO declares new Rs. 2000 note to be the best currency in the World" widely spread on several Whatsapp groups and later on it was confirmed that no such statement was made by UNESCO. Another news propagated rapidly in November,2016 saying "A nano GPS tracking device will be embedded in the new Rs. 2000 notes which would alert the authorities if black money was acquired and also RBI was using radioactive ink to print the new Rs. 2000 and Rs. 500 notes." Later RBI clarified that no such tracking device or ink would be used in the new notes. All such messages creates false believes and untruths among people and must be identified.

Fig. 1. Fake News examples

We focus on messages and data from Twitter since it is regarded as one of the top social networks .Particularly, in emergency events, Twitter is the first choice of many for updated information, due to its continuous live feed and short length of the messages. Also, the majority of messages on Twitter are publicly available and Twitter's API allows us to collect the high volume of data, e.g. Messages, user's information, etc., required to build a rumor classifier.

Fig. 2. Fake News examples on currency

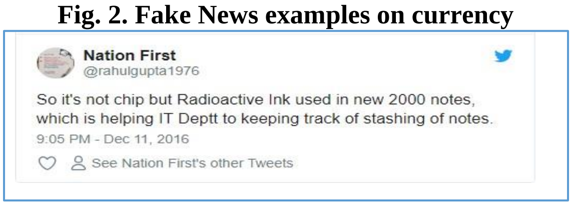

Fig. 3. Fake News examples on currency

We define a rumor to an unverified assertion that starts from one or more sources and spreads over time from node to node in a network. In the social sciences, a rumor involves some kind of a statement whose veracity is not quickly or ever confirmed.

Revised Manuscript Received on October 30, 2019.

* Correspondence Author
Deepak Sharma*, Research scholar, MD University, Rohtak, India.
Shilpa Singhal, PG student, C-DAC, GGSIP University, New Delhi, India.

With the pervasiveness of online media data as a source of information, verifying the validity of this information is becoming even more important yet quite challenging. Rumors spread a large quantity of misinformation on microblogs. Social media users spend several hours a day to read, post and search for news on microblogging platforms. Social media is becoming a key means for discovering news. However, verifying the trustworthiness of this information is becoming even more challenging. There are several negative effects of rumors on our daily lives. False rumors help Scammers to trick victims with serious scams. Rumors can be damaging, distressing and dangerous to individuals. Posting false rumors can stir up racial hatred which causes isolation towards ethnic minorities. False rumors induce more false rumors and untruths among people. Every false rumor circulated devalues the extent to which social networking and the internet acts as an effective method of spreading true information.

II. LITERATURE REVIEW

With the increase in the use of Social Media work has been done to detect the rumors but no such system came into existence yet due to which still it's a big issue and everyday a lot of false information is generated on Social Media Websites.Colin Singletonet et.al [1] tried to determine the veracity of the rumors on Twitter. For every rumor four sets of data was collected tweets e.g. Text, timestamp, retweet/quote/reply information etc., the users' information e.g. user id, number of posts/followers/friends etc. The user's followers such as friends and the users' most recent 400 tweets prior the start of the rumor. To extract the linguistic characteristics and sentiment of text LIWC version LIWC2015 were used which consists of up to date dictionaries of words that reflect different emotions and thinking styles. All the tweets collected were analyzed using this software and the attributes were extracted such as positive, negative sentiment score, fraction of words which represent, insight, cause, tentative, certainty, swearing. This work suggested that the Decision Tree, Random Tree and Logistic Regression are the best classifiers.

Another research was made by Quanzhi Li et.al [2] in which interaction of users and their belief on rumor propagation was studied. In this research first data was collected from the websites Snopes.com and Emergent.info that identified the present rumors on twitter and then belief was divided into four major categories Support, Deny, Question and Neutral. Those who support the rumor believes in rumor. Those who denies or question express the doubt. Those under Neutral Category doesn't make any statement about it. By this various user's beliefs were understood. On the basis of this study the users were found to fall in any of the four categories of News organization. Highly visible accounts, Low Credibility accounts and verified accounts. Using a set of 421 distinct rumors various aspects of rumor were captured and characterization of user belief were made with time. Figure 4 shows the User belief distribution and evolution over time.

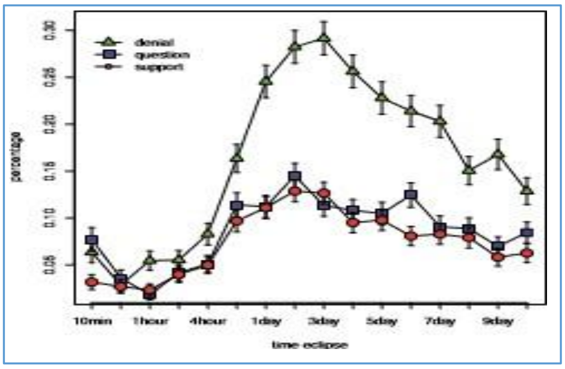

Fig. 4. User belief distribution and evolution over time

In first half hour the percentage of deny or question is less than percentage of tweets that support the rumor and after half an hour more evidence was collected so percentage of deny continues to grow larger than support tweets.

Sardar Hamidian et.al [4] tried to identify the rumor and then to classify it. They used the standard dataset for this research. A list of annotated twitter dataset for five different established rumor was published for this work in table I. Multistep classification with different set of features was performed on this dataset and the pipeline doesn't benefit much to the system also the use of limited amount of dataset was found.

Table- I: List of Annotated Tweets

Rumor	Rumor Reference	No of Tweets
Obama	Is Barack Obama muslim?	4975
Michele	Michelle Obama hired many staff members?	299
Cellphone	Cell phone numbers going public?	215
Palim	Sarah Palim getting divorced?	4423
AirFrance	Air France mid-air crash photos?	505

In other research [5],[6],[7] different rumor spreading models were studied to show how rumor propagate from one node to another in a network. The first Model was given by Daley and Kendall known as the basic DK Model in 1960s.Maki and Thomson focused on the analysis of the rumor spreading model based on mathematical theory and developed the MK Model. Both were used for the quantitative studies of rumor spreading but there are some major shortcomings of these models. They have not taken into account the topological characteristics of social networks such as connectivity, node degree. Not suitable for describing the rumor spreading mechanism on large-scale social networks. To deal with large scale networks new models SIR,SIHR,SICR were introduced and the results were compared. The SIR model was subdivided into three groups S:Spreader, I: Ignorant and S:Stifler. Spreader were the People who actively spread Rumors. Ignorant were People who are not aware of rumor and Stiflers were People who are aware of rumor but are not interested to spread it further. In SIHR model a new term Hibernators (H) was introduced to consider the forgetting and remembering mechanisms in networks. It was a temporary inactive state achieved by the spreader when they forget after spreading the rumor.

International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-9 Issue-1, October, 2019

In SICR model a new group Counterattack (C) was also introduced to refute the rumor and also persuade the neighbors not to believe and further spread it. After that results were compared and SICR model give the best results.

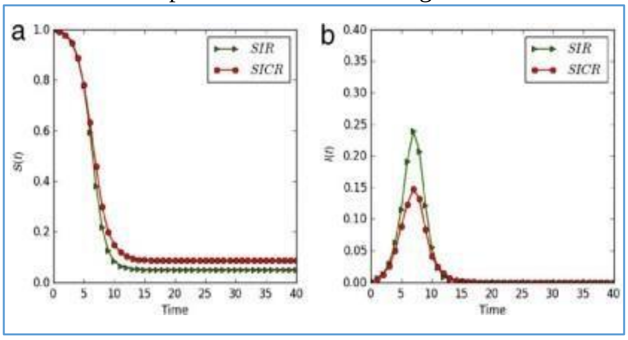

Fig. 5. Comparison of SICR and SIR model

The Susceptible scale of SICR model was higher than that of SIR Model so less number of people would be affected by rumor in SICR model also the Speed of rumor spreading was less in SICR model than SIR model due to counterattack mechanism. In research [3] work has been done on information about Potential Danger which is the central component of several rumors. Linear Algorithm such as logistic and linear regression were used for classification. All the Participants were asked to first read the study procedures about risks and discomforts and after that further steps were followed. Six products were taken for the Material Pre –Test and for each product 3 versions were created i.e. TRI – Threat related information, NEG- Negative feature of product, NEU-Neutral feature. In Study 1 participants were asked to read and compare descriptions of 3 products. There were 3 trials, one for each situation. At each trial participants were shown the two versions of the story, in parallel on computer screen. After reading the two parallel versions the answers were recorded.

Table- II: Difference between Threat content and negative content

	Threat content (TRI)	Negative content (NEG)
Trekking	[…] There are leeches that cling to your feet and can give you very serious deep burns. […]	[…]The Amazon is the poorest region of Brazil, with fewer schools, cities and roads than any of the other regions. […]
Computer program	[…] If you press control keys during installation, the software may damage your hard disk. […]	[…] The program can take a long time to master because the instruction manual is very complex. […]
Cooking	[…] If left to simmer too long the wildebeest meat will turn very bitter. […]	[…] Some people don't like this kind of stew because it looks gray, which they don't find appetizing. […]

Table- III: Results of Study

	Threat	Negative	Don't Know
Trekking	71.3	24.8	3.9
Computer program	72.9	24.0	3.1
Cooking	60.5	34.9	4.7

Again the maximum chosen was the threat related information than negative information in all above studies we studied about the privilege of threat related information over positive and neutral material and more easily people believe and transmit the threat related information. So by this we concluded that an information with more Potential Danger is more likely to be rumor.

III. PROPOSED APPROACH

In this section proposed approach is discussed in detail with the work flow. First we will be discussing the idea behind the approach followed by the working principal and a flow chart describing the iterative steps for the implementation of the approach.

A. Idea behind the Approach

The idea behind this research is to design an algorithm and train our machine to predict whether a given Tweet is a rumor or Information. In our proposed approach we will follow the following steps. Firstly, the collecting the Tweet dataset which are labeled as Rumor/Non Rumor for the work is to be done. After that Features Extraction Process for rumor followed by Features Selection for rumor based on Hypothesis. After that creation of a Machine Learning Algorithm to detect whether the Tweet is Rumor or Information which is also the main part of the work. After successful development of the model use the training dataset to optimize the function.

B. Working principle

In collecting the Tweet dataset which are labeled as Rumor/Non Rumor for the work, we collected the dataset for our work. We used the available Pheme Dataset of rumors and non-rumors that consists of collection of Twitter rumors and Non rumors that were posted during five breaking news in different countries. The dataset is as follows:

Table- IV: Pheme Dataset

S.No	News	Rumor	Non-Rumor
1.	Charlie Hebdo	458(22.0%)	1,621(78.0%)
2.	Ferguson	284 (24.8%)	859(75.2%)
3.	Germanwings Crash	238 (50.7%)	231(49.3%)
4.	Ottawa Shooting	470(52.8%)	420(47.2%)
5.	Sydney Siege	522(42.8%)	699(57.2%)

Each of the news consists of two subfolders: One for rumor and other for non-rumor. Inside each folder there are further sub folders consisting of 'Source Tweet' and 'Reactions 'on each Tweet. Each Tweet has a Tweet ID and is a JSON file.

In Features Extraction Process for rumor :In this Step we have identified the important features of Tweet and get the two categories of features Direct Features and Indirect Features. Direct features are those which are directly accessed from the crawled JSON Tweet. Indirect features are those which cannot be directly extracted from the crawled JSON tweet. It consists of mostly the historical data of user and profile based features.

Index	Feature	Comments
1	Source	Tweeting tools
2	Type	Regular, Replies, Mentions and Retweets.
3	Retweet_count	The number of times the tweet is retweeted.
4	Favorite_count	The number of times the tweet is favorited
5	Hashtags_count	The number of hashtags in the tweet.
6	Urls_count	The number of urls in the tweet
7	Mentions_count	The number of mentions in the tweet.
8	Media_count	The number of media in the tweet
9	Symbols_count	The number of cashtag in the tweet.
10	Possibly_sensitive	If the tweet possibly contains sensitive content
11	Location	If the location field of profile is null.
12	URL	If the URL field of profile is null.
13	Description_len	The length of the description field of.
14	Verified	If the user is verified by Twitter.
15	Ff_ratio	Followers_count / Friends_count
16	Followers_count	The number of followers of the user.
17	Friends_count	The number of friends of the user.

Fig. 6. List of Direct Features

Index	Feature	Comments
1	Source_count	No. of sources used for posting n latest tweets.
2	Type_count	No. of types of the latest n tweets posted.
3	Hashtags_proportion	% of tweets with hashtags in the latest n tweets.
4	Urls_proportion	% of tweets with urls in the latest n tweets.
5	mentions_proportion	% tweets with mentions in the latest n tweets.
6	Media_proportion	% tweets with media in the latest n tweets.
7	Symbols_proportion	% tweets with symbols in the latest n tweets.
8	Sensitive_proportion	% tweets possibly sensitive
9	Nonfriends_interaction	If the tweet is an interaction between non-friends.

Fig. 7. List of Indirect Features

In Features Selection for rumor based on Hypothesis: In this step we select a few features that are useful for our work and neglect the irrelevant features. For this first we have collected the dataset of 50 rumors and 50 non rumors. For all the 21 features extracted before we analyze the behavior and identify all the important features and make a hypothesis of considering that feature into our work or not to consider it.

In creation of a Machine Learning Algorithm to detect whether the Tweet is Rumor or Information, we create an algorithm that will take as input a Tweet and will be able to detect whether it is a Rumor or a Non Rumor. We classify our data into 2 classes of Rumor or Information. We use the Weka Platform for Classification Process.

This will be elaborated in the implementation Process. After this we use the training dataset to optimize the function: In this step we use the different dataset to optimize our results further. At last use the Function for Predicting the Tweets is used to detect whether tweet is rumor or information and find the results.

C. Flow charts

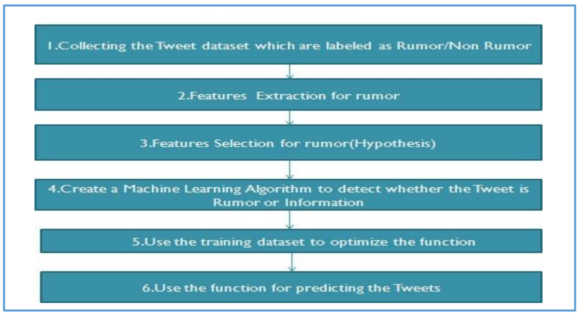

Fig. 8. Work Flow of System

IV. IMPLEMENTATION AND RESULTS

This section explains the techniques used in proposed approach, how the new proposed approach has been implemented in the WEKA and the results of that implementation.

A. Techniques Used

In this research work classification was used on our Dataset as it gives a Binary Classification. The various other algorithms we have used are the Naïve Bayes and Support Vector Machine Classifier. Random Tree Classifier is a supervised Classifier in which each node is split using the best among the subset of predicators randomly chosen at that node and No pruning is done in this and it contains all the features which are selected. On the dataset of 350 Tweets we have applied the Random Tree Classifier.

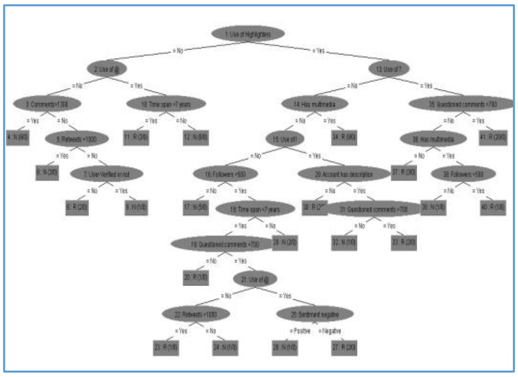

Fig. 9. Random Tree

In Decision Tree Classifier an information gain is calculated for each attribute and based on which features are selected and further classification at each node is done. After that Pruning is done to decrease the classification errors which are being produced by specialization in the training set for the instances which are not well defined and generalization of tree and the generated tree is:

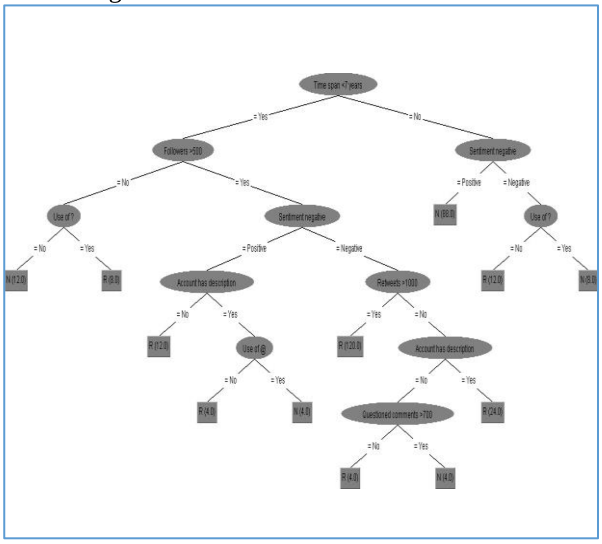

Fig. 10. Decision Tree

Support Vector Machine" (SVM) is a supervised machine learning algorithm which can be used for both classification and regression challenges. However, it is mostly used in classification problems. In this algorithm, we plot each data item as a point in n-dimensional space where n is number of features) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiate the two classes. Support Vectors are simply the co-ordinates of individual observation. Support Vector Machine is a frontier which best segregates the two classes (hyper-plane/ line).

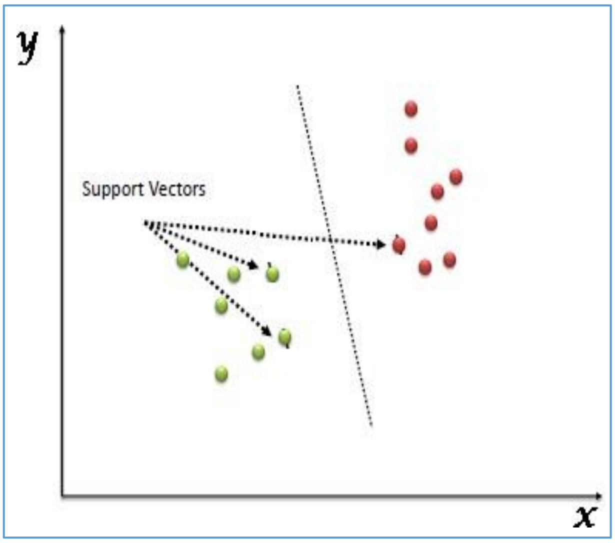

Fig. 11. Support Vector Machine

Naïve Bayes is classification algorithm technique which is related to Bayes' theorem with a theory or hypothesis of independence between predictors. In other words we can say that Naive Bayes assumes all the features are independent that is one feature don nit affect the other features that are present. For example, clock which is wall clock has feature round, 3 arms and about 5 cm radius. These features may depend on one another but classifier would consider then independent for classification of clock.

Weka is used for our implementation process. It is an open source java based platform that contains various machine learning algorithms that can be directly applied to our dataset or can be called from java code. It can perform various tasks on data such as Preprocessing, clustering, classification, regression, visualization and feature selection .All these can be performed on data contained in file where each data point has a fixed number of attributes that can be nominal or numeric.

B. Implementation

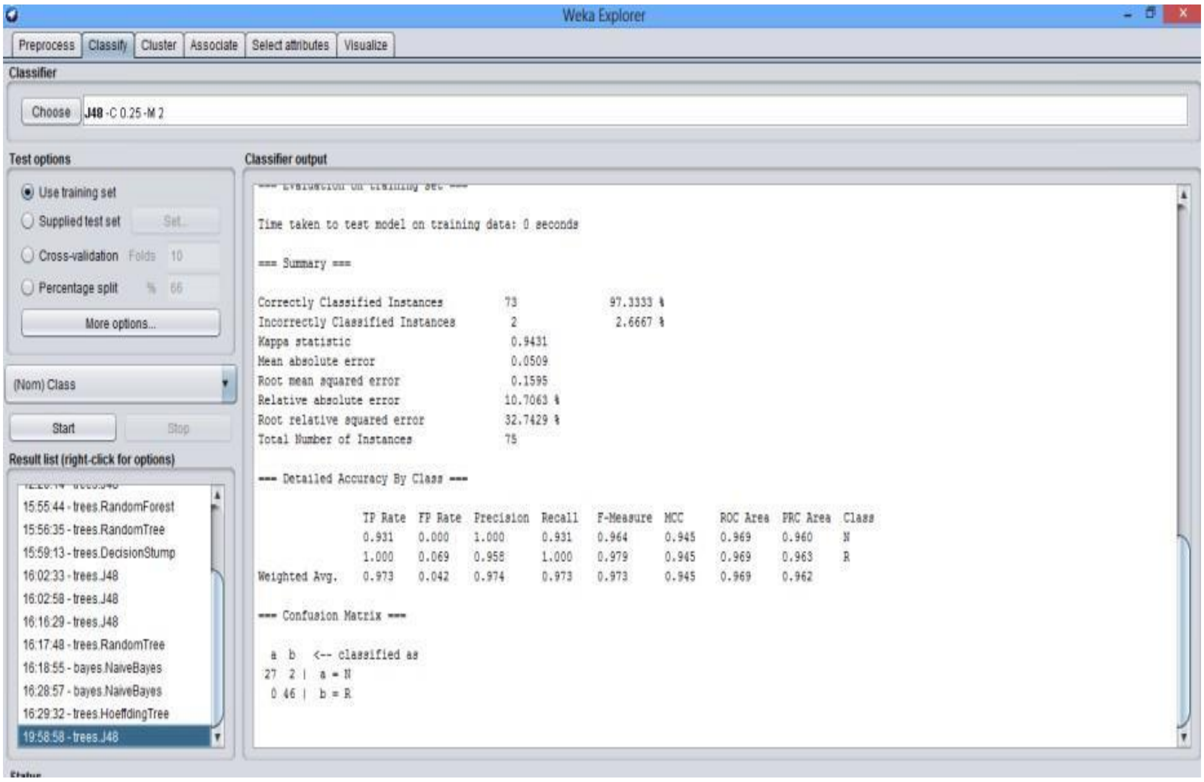

Fig. 12. Result of Random Tree

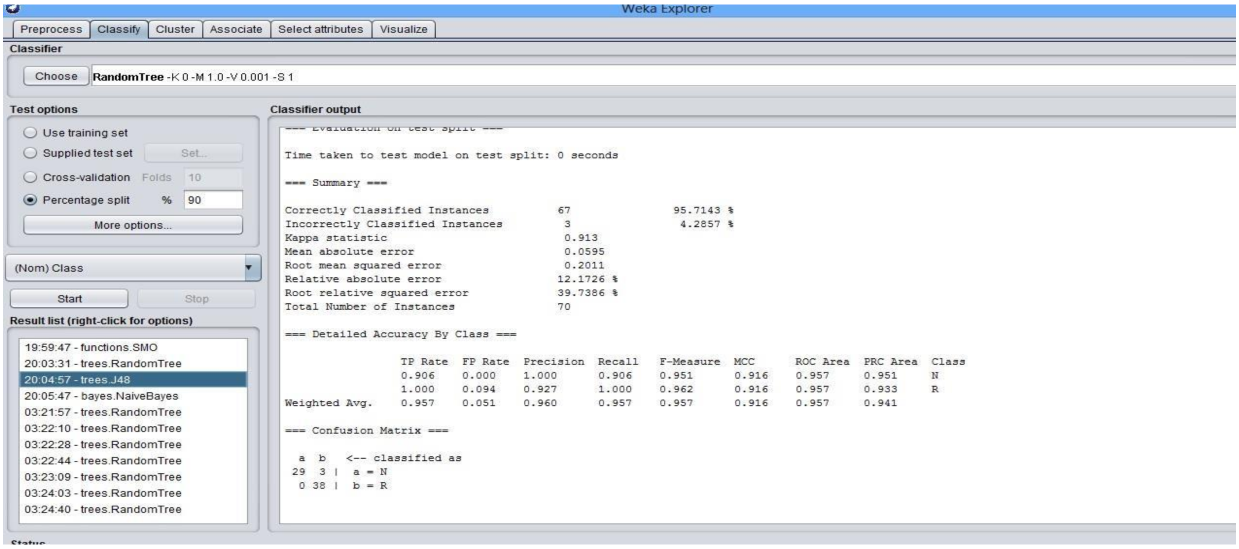

Fig. 13. Result of Decision Tree

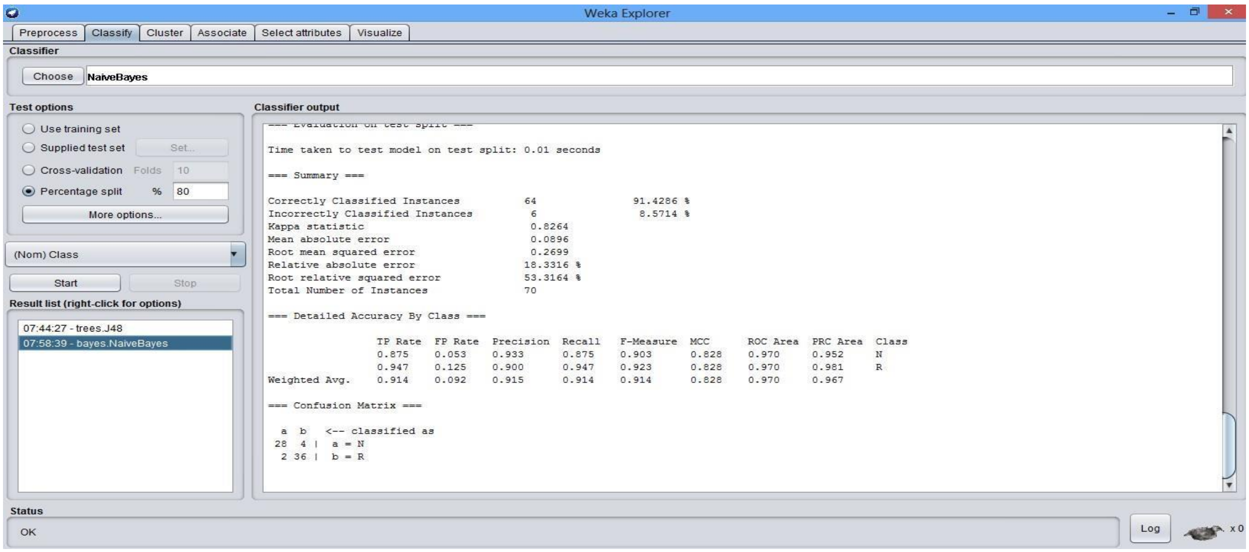

Fig. 14. Result of Naive Bayes

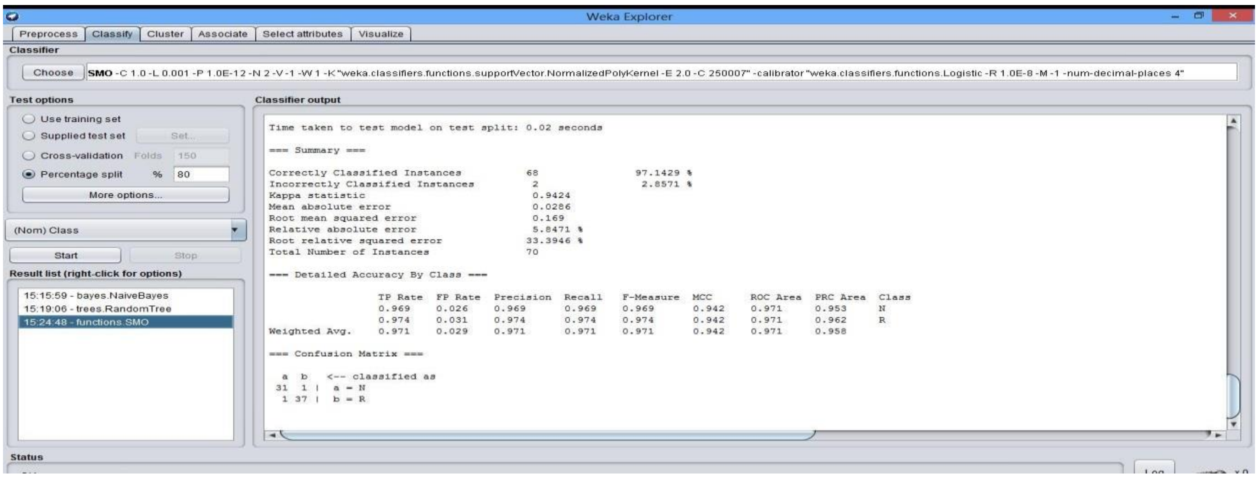

Fig. 15. Result of SVM

C. Comparative results and analysis

We have performed 4 machine learning algorithms using WEKA platform with different combinations of splits and set of features and obtained different results . We split the data into training (80%) and testing (20%) sets. These two sets also preserve the overall ratio of true to false observations.

Table- V: Comparison of Classifiers

S.No	Classifier	Correctly Classified Instances	Incorrectly Classified Instances	Accuracy
1.	Naïve Bayes	64	6	91.42%
2.	Decision Tree	67	3	95.71%
3.	Support Vector Machine	65	5	97.14%
4.	Random Tree	68	2	97.21%

V. CONCLUSION AND FUTURE SCOPE

In this section we will conclude the overall performance of our work assumption taken and the scope of project in future. On comparing the results of different classifiers we find the best accuracy of 97.21% for Random Tree and 95.71% for SVM with 97.14%.For Decision Tree the accuracy achieved was 95.71% and for Naïve Bayes its 91.42%.We also come up with other metrics such as TP rate, FP rate, F-measure ,Precision and confusion matrix for each algorithm. Our key findings suggest that the Decision Tree, Random Forest and SVM are the best classifiers and also the F1-score for Random Tree was highest than other classifiers.

REFERENCES

1. Georgios Giasemidis, Colin Singleton, Ioannis Agrafiotis, Jason R. C.Nurse, "Determining the veracity of rumours on Twitter", in the proceedings of 8th International Conference on Social Informatics (SocInfo2016), 2016.
2. Quanzhi Li, Xiaomo Liu, Rui Fang ," User Behavior in Newsworthy Rumors:A Case Study of Twitter" 10th International AAAI Conference on Web and Social Media (ICWSM-16), At Cologne, Germany ,May 2016.
3. Pascal Boyer, Nora Parren,"Threat-Related Information Suggests Competence: A Possible Factor in the Spread of Rumors" Department of Psychology, Washington University in St. Louis, St. Louis, Missouri, United States of America, June 10, 2015.
4. Sardar Hamidian and Mona Diab,"RumorDetectionandClassificationforTwitterData" The Fifth International Conference on Social Media Technologies, Communication, and Informatics, 2015.
5. Yongli Zana,b, Jianliang Wua, Ping Lic, Qinglin Yua," SICR rumor spreading model in complex networks: Counterattack and self-resistance" Physica A: Statistical Mechanics and its Applications Volume 405, PP: 159-170, 1 July 2014.
6. Laijun Zhao, Jiajia Wang, Yucheng Chen, Qin Wang, Jingjing Cheng, Hongxin Cui," SIHR rumor spreading model in social networks" Physica A: Statistical Mechanics and its Applications Volume 391, Issue 7, PP:2444-2453 ,1 April 2013.
7. Laijun Zhao, Hongxin Cui, Xiaoyan Qiu, Xiaoli Wang, Jiajia Wang," SIR rumor spreading model in the new media age " Physica A: Statistical Mechanics and its Applications Volume 392, Issue 4, PP:995-1003, 15 February,2012.
8. Nanjing, Jiang su , Third International Conference on Digital Manufacturing & Automation Rumor spreading in the online social network: A case of a Renren account Gu Yiran, Meng Fanrong Nanjing University of Posts and Telecommunications, 210046, China 2012.
9. Rudra M. Tripathy, Sameep Mehta," A study of Rumor Control Strategies on online social networks" CIKM '10 Proceedings of the 19th ACM international conference on Information and knowledge management Pages 1817-1820 126 October,2010Ali Dorri, "An EDRI-based approach for detecting and eliminating cooperative black hole nodes in MANET", Wireless Networks (2016), Springer.

AUTHORS PROFILE

Deepak Sharma is currently pursuing a Ph.D. in Computer Science at M. D. University, Rohtak. He has completed his M.tech from C-DAC: Centre for Development of Advanced Computing, Ministry of Communications and Information Technology, Government of India affiliated from Guru Gobind Singh Indraprastha University, Delhi. His main research areas include Data mining, Mobile Adhoc Network (MANET), wireless sensor network (WSN) and Internet of things (IoT).

Shilpa Singhal has successfully completed her master's degree in Computer Science and Engineering from C-DAC, Noida affiliated to Guru Gobind Singh Indraprastha University, New Delhi. Her research dimension covers computer networking and machine learning.