iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** VIII **Month of publication:** August 2021

DOI: https://doi.org/10.22214/ijraset.2021.37799

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VIII Aug 2021- Available at www.ijraset.com

Rumor Detection on Twitter

Shilpa Singhal
C-DAC, Noida, India

Abstract: ***Social media interaction such as news spreading around the network is a great source of information nowadays. From one's perspective, its negligible exertion, straightforward access, and quick dispersing of information that lead people to look out and eat up news from internet-based life. Twitter is among the most well-known ongoing news sources that ends up a standout amongst the most dominant news spreading mediums. It is known to cause extensive harm by spreading bits of fake news among the people. Online clients are normally vulnerable and are reliable on web-based networking media as their source of information without checking the veracity of the information being spread.***
This research contributes to develops a system for detection of rumors about real- world events that propagate on Twitter and to design a prediction algorithm that will train the machine to predict whether the given data is information or a rumor. The work finds all the useful features of a Tweet. The dataset used is the pheme dataset of known Rumors and Non Rumors. Afterwards, we make a comparison between various known Machine learning algorithms such as Decision tree, SVM, Random Tree.

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 sociopragmatic 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 negative effects of rumors on our daily lives such as False rumors help Scammers to trick victims with serious scams and Rumors can be damaging, distressing and dangerous to individuals. Posting false rumors can stir up racial hatred which causes isolation towards ethnic minorities and it can induce more false rumors and untruths among people.

Our research in one way to reducing the finding the fake tweets and predicting whether it is rumor or information.

In our research the data consists of 350 tweets. These tweets are used for training and testing different machine algorithm with different features that are found in the research.

Four Different Classifiers such as Support Vector Machine (SVM), Decision Tree,Random Tree, Naïve Bayes are implemented for this task. A combination of these classification models are also tested to further enhance the accuracy of prediction

Paper is divided into different sections: section 1 describes precious research done in same domain, section 2 it describes basic machine learning algorithm, section 3 describes proposed steps, section 4 describes the data and research work, section 5 describes observation and results, section 6 contains references.

II. LITERATURE REVIEW

With the increase in the use of social platform a lot of research work has been done in this field.

In research [1] interaction of users and their belief on rumor propagation was studied. In this 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.

There are different machine learning algorithms that are implemented in many research works such as decision tree, naive bayes, support vector machine, neural network,Random Tree classifier

In research [2] the work was done to determine the veracity of the rumors on Twitter and implemented using Decision Tree and Random Tree classifier. Different studies has applied different machine learning algorithms: Decision tree, Regression, SVM, Naïve Bayes and Random tree with different set of data.

"Fake" news detection on twitter has been initially researched by many authors in the past. But during the presidential election of US, this issue became more popular and everyone was given their best to find out some better solutions for this classification. The previous history was briefed on "The Atlantic" news [8].

Different papers and journals explained the core of the problem and its sections [9] [10].

Some proposed data mining approaches [11], some uses the Stance Detection method by trained the machine using stances [12].

©IJRASET: All Rights are Reserved

Some authors also used a machine learning algorithm implemented as a software system for detection [13].
We mainly focused on a paper [14] which proposed a methodology for characterizing trending twitter threads. In the modern timeline, an outstanding science news site "SCIENCE" distributed an article where they ordered twitter news as evident or fictitious utilizing data from six free reality checking associations that showed 95 to 98% concession to the orders.
[15] Out of all the studies in this specific area, one study proposed a survey on the field of NLP for false news detection. This paper displays an overview of fake news recognition. Their review presents the difficulties of the methodology in the detection of news or twitters like this. They deliberately audit the data and NLP arrangements that have been produced for this whole study. They likewise examine the points of confinement of those twitters and issue definitions, our bits of knowledge, and suggested arrangements.
[16] One paper used the “Crowdsourcing a large corpus of clickbait” to address the urging task of clickbait detection, they constructed a new corpus of 38,517 annotated Twitter tweets, the Web is Clickbait Corpus 2017. [17]

Our approach works on Twitter data for which data is collected online and used for training set.

A. Approach

The objective of our research is to design an algorithm and train a machine to predict whether a given Tweet is a Rumor or Information.

The fig. 1 shows the diagram of the proposed approach of our research work. The approach consists of following steps:

1) Collecting the Tweet dataset which are labelled as Rumor/Non Rumor.
2) Feature Extraction for Rumor.
3) Feature Selection for Rumor.
4) Create machine algorithm to detect whether a given tweet is Rumor or Information.
5) Use the training dataset to optimize the function.
6) Use the function for Predicting the Tweets.

Our research approach consists of 6 steps which are listed above. In first step data is collected online which is the Pheme dataset or Rumor/Non Rumor that consists of collection of Twitter rumors and Non rumors that were posted during five breaking news in different locations such as German,Paris. Once the data is collected then the next step of project starts that is feature extraction which is done by subjective analysis of data and it comes out with the intermediate result that are direct and indirect features of a Tweet. After this hypothesis is made for building a essential list of features required for prediction. The list of final features contains 16 useful features out of 21. This hypothesis is base for our research for predicting Tweet as Rumor or Information Then using WEKA we will be implementing the machine learning algorithms(Decision Tree, Random Tree). By implementing we get 97.21% of accuracy for Random Tree and 95.71% with decision tree.

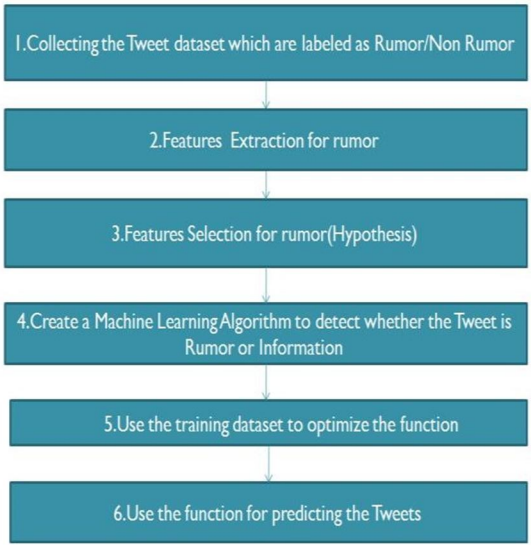


Fig. 4.1 Proposed Approach

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VIII Aug 2021- Available at www.ijraset.com

B. Features

With the subjective analysis and hypothesis we have come up with 16 features that are to be considered in our work to be useful.

S. No	Feature Name	Value
1.	User is verified or not	Yes/No
2.	No. of followers	Count
3.	No. of Retweets	Count
4.	No. of Comments	Count
5.	No. of Questioned Comments	Count
6.	Account has description	Yes/No
7.	Use of @	Yes/No
8.	Use of !	Yes/No
9.	Use of Highlighters	Yes/No
10.	Use of ?	Yes/No
11.	Use of ellipses	Yes/No
12.	Use of URLs	Yes/No
13.	Sentiment of Tweet	Post./Neg
14.	Has duplication	Yes/No
15.	Has Multimedia	Yes/No
16.	Time Span	Count

III. RESULTS

We have used the Weka Platform for our implementation process. The dataset used is the Pheme dataset that consists of collection of Twitter rumors and Non rumors that were posted during five breaking news in different countries. In our research the data consists of 350 tweets that are used as the training and testing dataset.

We have performed 4 machine learning algorithms 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.

We have come up with the different results and are comparing them as follows:

Sn o	Algorithm	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	92.14 %
4	Random Tree	68	2	95.21 %

©IJRASET: All Rights are Reserved

IV. CONCLUSION

On performing the classification on Random Tree it gave an accuracy of 95.21% but no pruning was done and all the features were considered to be equal.On performing decision tree with pruning accuracy measured was 95.71%. Some of features were further reduced in this such as Use of ellipses, Use of URLs, Use of Highlighters, No. of comments, Verified User etc.

If we compare all the results of Weka the r performance of Random Tree and decision Tree was better then other classifiers SVM,Naïve Bayes.

In this paper, we analysed a computerized model for checking the verification of news extracted from Twitter which gives general answers for information accumulation and expository demonstration towards fake news recognition. After having an idea from the supervised models, a deep learning-based model is proposed to identify fake news. The accuracy metric presumably would be altogether improved by methods for utilizing progressively complex model. It is worth noting, that even with the given dataset, only part of the information was used. The current project did not include domain knowledge related features, such as entity-relationships. Future studies could extract name entities from each pair of news headline and news body and analyse their relationships through a knowledge base.

The study demonstrated that even the very basic algorithms on fields like AI and Machine Learning may find a decent outcome on such a critical issue as the spread of fake news issues worldwide. Accordingly, the after effects of this examination propose much more, that systems like this might come very much handy and be effectively used to handle this critical issue.

This work exhibits a programmed model for identifying fake news in well-known Twitter strings. Such a model could be important to a huge number of social media users by expanding their own credibility decisions. The dataset in this examination is relied upon to be utilized for arrangements which utilized machine learning based statistical calculations, for example, Support Vector Machines (SVM), Naive Bayes (NB), Logistic Regression (LR). In this investigation, SVM performs best for characterization technique.

The studies already done in this field have done good work so far. Some of the difference that we found in our research and previous study are as follows:

1) No yet mechanism exist for proper identification of rumor on Twitter.

2) Difficult to maintain the rumor dictionary as new words are added each day . We try to maintain a relevant dictionary for our system that contains almost all possible words for rumor.

3) Further it can be used in Finding the source of the rumors once the rumor is identified.

In future we will work on the following features so that a reliable mechanism can be maintained to identify the authentification of the information on online medium.

Also we will try to improve the accuracy of the results to make it more accurate and ideal.

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 InformationSuggests 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] 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

[5] 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

[6] 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,2010

[7] 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,Nanjing, Jiang su ,210046, China 2012

[8] Meyer, "The Grim Conclusions of the Largest-Ever Study of Fake News."

[9] Tandoc Jr. Lim. Ling, (2017, August 30). Defining "Fake News" A typology of scholarly definitions.

[10] D Horne. Sibel Adali, Fake News Packs a Lot in Title, Uses Simpler, Repetitive Content in Text Body, More Similar to Satire than Real News, Benjamin.

[11] Shuy. Silva. Wangy.Tang. Huan, "Fake News Detection on Social Media: A Data Mining Perspective."

[12] Machine Learning (CS229), "Fake News Stance Detection."

[13] Sharf. Zareen. Saeed," Twitter News Credibility Meter."

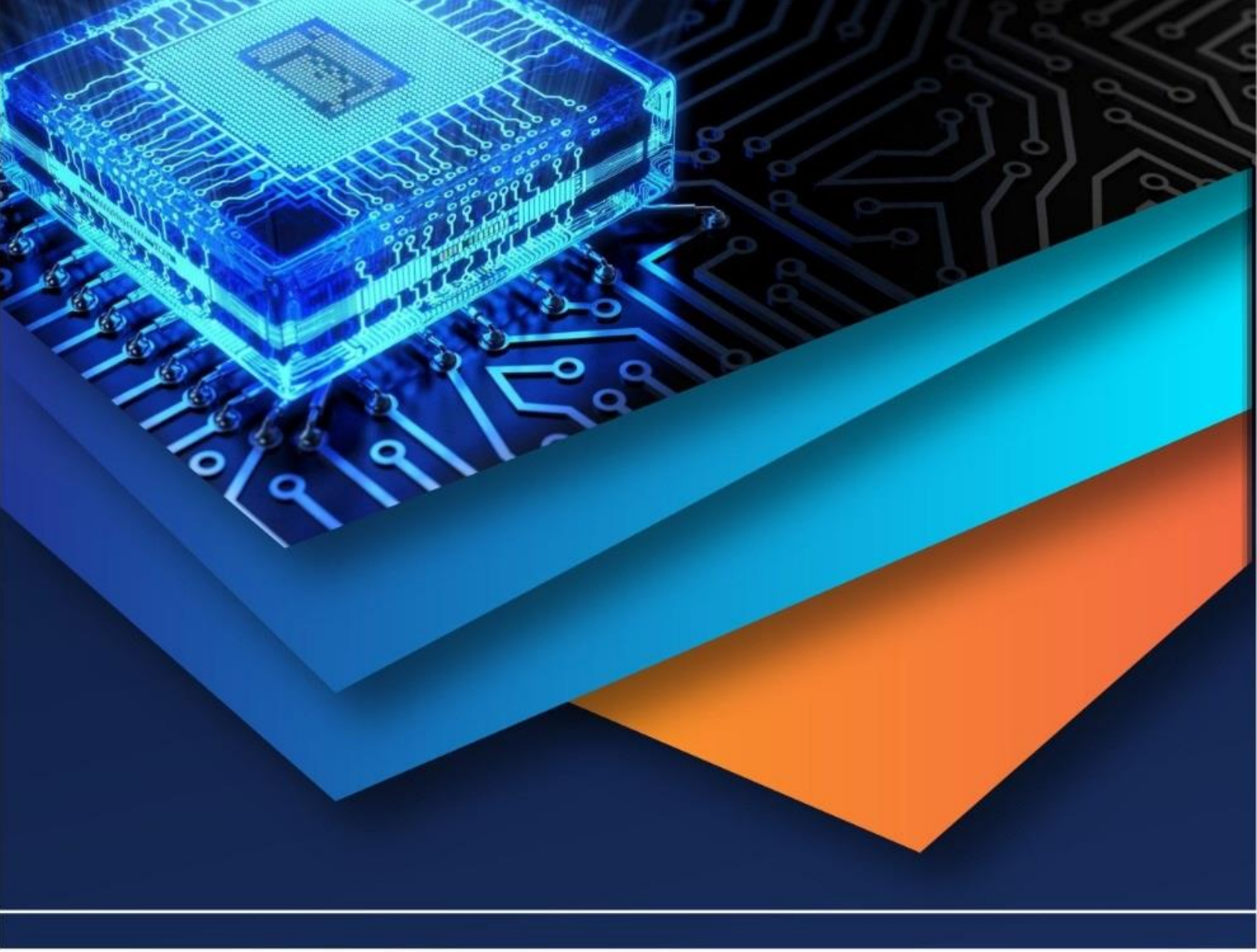
[14] Cody Buntain,."Automatically Identifying Fake News in Popular Twitter Threads." IEEE International Conference on Smart Cloud,2017

[15] Soroush Vosoughi, Roy, Aral, (March 2018), "The spread of true and false news online."

[16] Oshikawa.Qian.Wang, (November 2018), "A Survey on Natural Language Processing for Fake News Detection."

[17] Golbeck, Jennifer, et al. "A Large Labeled Corpus for Online Harassment Research" Proceeding of the 2017 ACM on Web Science Conference. ACM,2017

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)