iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 7 **Issue:** VII **Month of publication:** July 2019

DOI: http://doi.org/10.22214/ijraset.2019.7067

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: 7.177
Volume 7 Issue VII, July 2019- Available at www.ijraset.com

Fake News Detection using Support Vector Machine learning Algorithm

Suchitra B. Deokate.
P.G. Student, Department of Computer Engineering, VPKBIET, Baramati, Maharashtra, India.

Abstract: ***Fake news has vast impact in our modern society. Detecting Fake news is an important step. This work purposes the use of machine learning techniques to detect Fake news. It is now used not only for social communication, but also as an important platform for exchanging information and news. Internet and social media made the access to the news information much easier and comfortable. Often Internet users can follow the events of their interest in online mode, and spread of the mobile devices makes this process even easier. This paper develops a method for fake news detection using machine learning on Twitter by learning to predict accuracy assessments in two credibility-focused Twitter datasets: CREDBANK, a crowdsourced dataset of accuracy assessments for events in Twitter, and PHEME, a dataset of potential rumors in Twitter and journalistic assessments of their accuracies. Apply this method to twitter content sourced from Buzz Feed's fake news dataset and show models trained against crowdsourced workers outperform models based on journalists. use of machine learning techniques to detect Fake news, using Support Vector Machine (SVM).***
Keywords: Social networking site, Twitter, reputation, credibility, fake news, machine learning.

I. INTRODUCTION

The vast amounts of information are produced on the social networking with various social media's formats. There have provided very big volumes of posts that explosive increasing of the social media data on the web. When some event has occurred, many people discuss it on the web through the social networking. They search or retrieve and discuss the news events as the routine of daily life. However, very large volume of news or posts made users face the problem of information overloading during searching and retrieving. Unreliable sources of information expose people to a dose of fake news, hoaxes, rumors, conspiracy theories and misleading news.

Some types of news such as bad events from nature phenomenal or climate are random. When the unexpected events happen there is also fake news that are broadcasted that create misunderstanding due to the nature of the events. Whom known the real fact from the event while the most people believe the forward news from their credible friends or relatives. The fake news comes from the misinformation, misunderstanding or the unbelievable contents which the creditability source. These are difficult to detect whether to believe or not when they receive the news information. Thailand is located in a tropical terrain. The rain is almost throughout the year therefore causes massive flooding in Thailand. Thai Meteorological department present the information of weather forecast, hydrological information and local climate.

They have broadcasted the forecasting information to notify the public beforehand and protect their properties. However, the unpredictable natural phenomena's news such as rain, floods, forest fires, earthquakes, storms, cold and hot weather which could be rapidly spread worldwide with misleading misunderstandings.

Hence, aimed to develop the method for fake news detection in twitter. This method uses a classification model to predict whether a thread of Twitter discussion will be labeled as accurate or inaccurate using features motivated by existing work on credibility of Twitter stories. We determine this approaches ability to identify fake news by evaluating it against the BuzzFeed dataset of 35 highly shared true and false political stories. This work is difficult by the limited availability of data on what is fake news online, however, so to train this system, we leverage two Twitter datasets that study credibility in social media: the PHEME journalist-labeled dataset and the CREDBANK crowd sourc

ed dataset. PHEME is a curated data set of discussion threads about rumors in Twitter replete with journalist annotations for truth, and CREDBANK is a large-scale set of Twitter discussions about events and corresponding crowd sourced accuracy assessments for each event. In particular use many features of twitter such as Structural feature, Content features and User features these features use for the user reputation and Credibility of content. Using these features to train our fake news detection model and improve its performance. Apply machine learning techniques to detect Fake news by using the support vector machine (SVM), algorithm used for classify the tweet with fake or real.

II. RELATED WORK

Shlok gilda.(2017) Proposed by This method demonstrates that term frequency is potentially predictive of fake news - an important first step toward using machine learning Classification for identification. The best performing models by overall ROC AUC are Stochastic Gradient Descent models trained on the TF-IDF feature set only compare the performance of models using three distinct feature sets to understand what factors are most predictive of fake news: TF-IDF using bi-gram frequency, syntactical structure frequency (probabilistic context free grammars, or PCFGs), and a combined feature union using machine classification for identification. This indicates that PCFGs are good for a Fake-News Filter type implementation versus, say, training fake news sites for review. [1] [4]. Mykhailo Granik, Volodymyr Mesyura (2017) they proposed by the Fake News Detection Using Naive Bayes Classifier was used for fake news detection method based on one of the artificial intelligence algorithms naive Bayes classifier. Naive Bayes classifiers are a general statistical technique of email filtering. Naïve Bayes typically use bag of words features to identify spam e-mail and method commonly used in text classification. And inspect how this particular method works for this specific problem given a physically labeled news dataset and to support the idea of using artificial intelligence for fake news detection.

Majed Alrubaian (2016) Proposed A Credibility Analysis System it is novel credibility assessment system that maintains complete entity-awareness in success a detailed information credibility judgment. This model comprises four integrated components, namely, a reputation based model, a feature ranking algorithm, a credibility assessment classifiers engine, and a user expertise model. All of these components operate in an algorithmic form to analyze and assess the credibility of the tweets on Twitter. The reputation-based technique helps to filter ignored information before starting the calculation process. The classifier engine component distinguishes between credible and non-credible content. [3] [6]. Supanya Aphiwongsophon and Prabhas Chongstitvatana Proposed the select dataset from Twitter are summarized with twenty two attributes. From this information, all the machine learning methods: Naive Bayes, Neural Network, Support vector machine, are very good at detecting Fake news There are classified to two classes with believable and unbelievable. The results from the classification are precision, recall, F-Measure, and accuracy [4]. Jacob Ross, Krishnaprasad Thirunarayan,they are create a robust and general feature set for learning to rank tweets based on credibility and newsworthiness. They are derived a set of features that are indicative of a tweets credibility regardless of the time period and topic of that tweet. Set of features were derived by combining popular and actual features from previous works, as well as deriving new features. Features can be broadly categorized as either user based features or tweet based features. They create sentiment based features that aim to capture when a tweets sentiment is irregular given the context of its topic. Juan Cao, Junbo Guo, Xirong Li, Zhiwei Jin, Han Guo, and Jintao Li [7] .Three different paradigms for rumor detection are elaborately labeled the feature-based classification approach, the credibility propagation approach and the neural networks approaches. For each method in these paradigm, it aims to two goals: one is to extract prominent features to representing the multimedia content comprise rumors and the social context generated by rumors on social network, the other goal is to build robust machine learning algorithms to separate rumors from common stories.

III.SYSTEM ARCHTECTURE

A. Text Pre-Processing

The tweets may contain a misspelling, acronyms, and even the interpretation is also ambiguous. For understanding the exact meaning of such data, we need to remove noise from tweets. Following are the text pre-processing steps [4].

1. Tweets may contain slang words, e.g. “omg”, “gn”. We replace slangs by their standard forms by using the slang word dictionary provided by http://noslang.com/dictionary/full
2. Tweets containing words with consecutive repeating letters, e.g. “yesssss”, “gooood”replace them by one so, original word remains as it is. We can recognize such words such words by using regular expressions.

B. Tweet Segmentation

Tweet segmentation is to split a tweet into a sequence of consecutive n-grams ($n \geq 1$) each of which is called a segment. A segment can be a named entity (e.g., a movie title “finding nemo"), a semantically meaningful information unit (e.g., “officially released "), or any other types of phrases which appear “more than by chance”. For Example -

1) *Before Tweet Segmentation:* Tweet - ’BRIDGEWATER Mass – Police in one Massachusetts town are warning residents about a menace on the streets Wild turkeys Bridgewater Police posted a video to Twitter on Sunday of four turkeys chasing after a police cruiser Some followers found the scene pretty amusing but police arent laughing CBS Boston reports Aggressive turkeys are a problem in townpolice wrote State law doesnt allow the police or the ACO to remove them Anyone having turkey trouble should call the MSPCA at 617-522-7400 police said’

2) *After Tweet Segmentation:* Segments - bridgewater, mass, police, massachusetts, town, warning, residents, menace, streets, wild, turkeys, posted, video, twitter, sunday, chasing, cruiser, followers, found, scene, pretty, amusing, arent, laughing, cbs, boston, reports, aggressive, problem, townpolice,wrote, state, law, doesnt, aco, remove, turkey, trouble, call, mspca, 6175227400

C. Features Selection

The initial motivation for feature selection is that the social data often contain many different features that are difficult to deal with this feature, and most of the features are terminated except for specific tasks. to deal with this problem, Apply feature extraction methods. Feature selection is often preferred over extraction; [8] because the selected features have more understandable and useful they select the three main features first is Structural Features Structural features capture Twitter-specific properties of the tweet stream, including tweet volume and activity distributions. Second is User features capture properties of tweet authors, such as interactions, account ages, friend/follower counts, and Twitter verified status and third features Content features measure textual aspects of tweets, like polarity, subjectivity, no of comments and agreement.

1) *Structural Features:* Twitter, Facebook is common online social network services that provide the platform of communication. It enables users to read and send message of length 140 character. Structural features include the text features and sentiment features.

a) *Text Features:* Include some characteristics to the Twitter communication thread and are calculated across the entire thread. These features include the number of tweets, average of tweet length; thread lifetime is number of minutes between first and last tweet, and the depth of the communication tree. Include the frequency and ratio of tweets that contain media like images or video audio, mentions(@), re-tweets, and web links. Tweet meta-data Number of seconds since the tweet; Source of tweet (mobile / web/ etc); Tweet comprises geo-coordinates. Number of characters, Number of words, Number of URLs, Number of hash-tags, Number of unique characters, Occurrence of typical symbol, Occurrence of happy smiley, Occurrence of sad smiley, Tweet contains ‘via’; Occurrence of colon symbol.

b) *Sentiment Features:* calculating the sentiment for each tweet and number of positive and negative tweet, based on a predefined sentiment list. The credibility of the information was then based on the ratio of positive to negative tweet.

2) *User Features:* In this section focuses on activities and thread characteristics, the following Features are attributes of the users taking part in the conversations, their connectedness, and the density of interaction between these users. Each user had a personal record of information in his profile. Some of these features are latent and some of them explicitly revealed in user profiles. For example, age, gender, education, political orientation, and even any user preferences are considered as latent attributes. The number of followers, number of friends and the number of re-tweeted tweets as well as the replies of users tweets. Such as the number of likes and unlike on specific topic, and authored status counts, occurrence of verified authors, and whether the author of the first tweet in the thread is verified. This last user-centric feature, network density, is measured by first creating a graph representation of interactions between conversations constituent users.

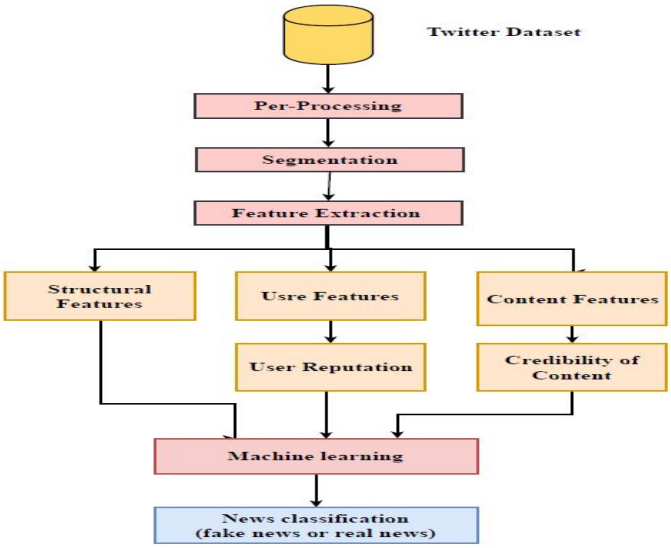


Figure 1 System architecture [9]

3) *Content Features:* Content features are based on tweets textual and visual aspects and include polarity is the average positive or negative feelings expressed a tweet , subjectivity is the score of whether a tweet is objective or subjective, and difference, as measured by the amount of tweets expressing difference the conversation. [9]Also include the frequency and proportions of tweets that contain question marks, exclamation points, first/second/third person pronouns, and smiling emoji. Content Features are two main types Textual Features and Visual Features.

a) *Textual Features:* A text describing the news event. They provide details of the event and may contain certain opinions or thoughts towards the story. General textual Features are derived from the linguistics of a text; three categories of general textual features are commonly used: lexical features, syntactic features, and topic features. Lexical features are features extracted at the word-level of a rumor, which could be statistics of words, lexical rumor patterns or sentimental lexicons. Syntactic features represent rumors at the sentence level. The basic syntactic features are simple statistics of a rumor message, such as the number of keywords, the sentiment score or polarity of the sentence and part-of speech tagging. Topic features are extracted from the level at the message set, which aim to understand messages and their underlying relations within a corpus.

D. User Reputation and Credibility Assessment.

User reputation systems are commonly used in Ecommerce website and social networking sites, such as Twitter, Facebook etc. Most of the user reputation systems use the rule-based method or the voting systems to calculate user reputations. In present on-line social network has become the most popular communication tool of people. They can publish, transfer and rate different contents in a social network.

The boundary between content provider and customers becomes more and more partial, and each user has parts, content provider and content customer.[5] Once a user pastes a message, other users can read, comment, transfer, add to favorite, and degree it. These interactions between social network users are called as social activity of users. So in a social network are often to communicate with different user.

For security reasons, it is necessary to build a reputation system to evaluate his/her behaviors. In present most of the user reputation systems use the rule-based method or the voting systems to calculate user reputations but in this paper using machine learning techniques to calculate the user reputation. [4]The creditability of information was defined by many words such as trustworthiness, believability, reliability, accuracy, fairness, objectivity, and other with the same concepts and definitions. Credibility can be defined as the quality of being trusted and believed in, or the quality of being considerable or believable. A critical part of the system is the assessment of the credibility of tweets and the reputations of the users who posted them rescore to represent the level of trustworthiness of the posted content.

Use the term reputation score to represent the level of dependability of the user who posted the content. User's reputations are based on popularity measures. Describe a popular user as one who is recognized by other users on a similar network. The measures include the Follower-Rank and the Twitter Follower Followed ratio (TFF).

In addition, consider replies and rewets a measures of a user's popularity. In this viewpoint, sentiment defines the degree of antagonism of any user Tweet that affects social relationships, the emotional states of other users, and their orientation with respect to the given topic.

Propose a new reputation-based source credibility assessment method that introduces several new features into the existing models. Main approach users sentiments to identify and evaluate topically relevant and credible sources of information sentiment defines the degree of resentment of any user Tweet that affects social relationships, the psychological states of other users, and their orientation with respect to the given topic and also calculated the number of positive and negative words in a message, based on a predefined list of sentiment words. And also use the user popularity, the users social popularity score can be quantitatively evaluated using a simple algorithm that defines a user's popularity score based on certain features that are related to the user's reputation Based on some topics, re-tweets are considered to be one of the best indicators of user popularity from the calculable perspective. [7]This suggests that a tweet that has been re-tweeted many times is considered to be attractive to the reader.

But, the most critical indicators of the popularity of the person who posts the tweet (the tweeter) are qualitative, such as the relationship between the reader and the tweeter. In particular, measure the reputation or credibility of a Twitter user based on how popular he/she is, and how sentimental he/she is regarding a particular topic. In developing this approach, we have identified new features that can be used to find the most credible Twitter users.

Evaluated the proposed system using machine learning algorithms.

IV. METHODOLOGY

A. Dataset Alignment

While working with multiple datasets from different residents reduces bias in the final collection, to compare the resulting models, we must translate these datasets into a into a consistent format.

1) Extracting Twitter Threads from BuzzFeed's Facebook

a) Dataset: The most evident difference among our datasets is that BuzzFeed's data captures stories shared on Facebook, whereas CREDBANK and PHEME are Twitter-based. use the following awareness to extract Twitter threads that match the BuzzFeed dataset: Each element in the BuzzFeed data represents a story posted by an society to its Facebook page, and all of these societies have a presence on Twitter as well, so each story posted on Facebook is also shared on Twitter. To align this data with PHEME and CREDBANK.

2) Aligning Labels: While the PHEME and BuzzFeed datasets contain discrete class labels describing whether a story is true or false, CREDBANK instead contains a collection of annotator accuracy assessments on a Likert scale. Therefore convert CREDBANK's accuracy assessment vectors into discrete labels comparable to those in the other datasets. Given annotator bias towards "certainly accurate" assessments and the resulting negatively skewed distribution of average assessments, a labelling approach that addresses this bias is required.

3) Capturing Twitter's Threaded Structure: Another major difference between PHEME and CREDBANK/BuzzFeed is the form of tweet sets: in PHEME, topics are planned into threads, beginning with a popular tweet at the root and replies to this popular tweet as the children. This threaded structure is not present in CREDBANK or our BuzzFeed dataset CREDBANK contains all tweets that match the related event-topic's three-word topic query, and BuzzFeed contains popular tweeted headlines. To capture thread depth, which may be a proxy for controversy [21], we adapt CREDBANK's tweet sets and BuzzFeed's popular tweet headlines into threads using PHEME's thread-capture tool. For our BuzzFeed data, we use the popular headline tweets as the thread roots and capture replies to these roots to construct the thread structure representing PHEME's. In CREDBANK, we identify the most retweeted tweet in each event and use this tweet as the thread root.

B. Predicting BuzzFeed Fact-Checking

Applying the most performant CREDBANK and PHEME models to our BuzzFeed dataset shows both the pooled and CREDBANK-based models outperform the random baseline, but the PHEME-only model performs substantially worse. Fig 2 show the graph of fake news Vs real news count in Buzzfeed Dataset. Correctly Classified Instances is and accuracy of 94198.8445 % ,Incorrectly Classified Instance are11 accuracy of 1.1555 % Total Number of Instances are 952 ,Kappa statistic accuracy of 0.9769% ,Mean absolute error accuracy of 0.0116% ,Root mean squared error accuracy of 0.1075%, Relative absolute error accuracy of 21.3112 %, Root relative squared error accuracy of 21.4997 % ,Coverage of cases (0.95 level) accuracy of 98.8445 % ,Mean rel. region size (0.95 level) accuracy of 50 %.

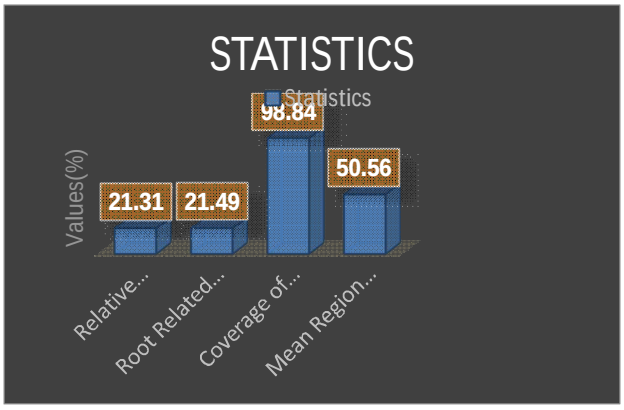


Fig. 2 Predicting BuzzFeed Fact-Checking accurancy result.

While predicting the fake news from dataset, these are the factors needs to be considered. Relative absolute error is the magnitude of the difference between the exact value and approximation. The relative error is represented in terms of per 100. Same result applicable for other parameter also.

V. ALGORITHM

There are some research uses the machine learning approach to determine the creditability of tweets message Using Support Vector Machines (SVM) to classify the fake or real news form twitter. Support Vector Machines (SVM) are an arrangement of related supervised learning techniques operated for grouping and classification SVM is a pair-wise ranking technique that uses SVM.[1]The stop words are been removed from the text data and the features are extracted successfully. After text feature extraction, SVM Classier performs classification on the data; and defines the fake news or real news.

Algorithm 1 for detecting News is fake or not:

1) *Input:* D dataset, on-demand features, aggregation-based features
2) *Output:* Classification of News
a) for each application *news_id* in D do
b) Get on-demand features and stored on vector x for *news_id*
c) x.add (Get_Features(*news_id*));
d) end for
e) for each application in x vector do
f) Fetch first feature and stored in b, and other features in w.
g) $h_{w,b}(x) = g(z)$ here $z = (w^T x + b)$
h) if ($z \geq 0$)
i) assign g(z)=1;
j) **else** g(z)=-1;
k) **end if**
l) end for

A .Mathematical formulation:

We will consider a linear classifier for a binary classification problem with labels **y** and features **x**.
Here, y €{-1,1} (instead of {0, 1}) to denote the class labels.
We will use parameters w, b, and write our classifier as,
$h_{w,b}(x) = g(w^T x + b)$
Here, g(z)= 1 if $z \geq 0$, and g(z)= -1 otherwise;
This "w, b" notation allows us to explicitly treat the intercept term b separately from the other parameters.
Thus, b takes the role of what was previously θ0, and w takes the role of $[\theta 1 \ldots \theta n]^T$.
Note also that, from our definition of g above, our classifier will directly predict either 1 or −1.

VI. RESULTS

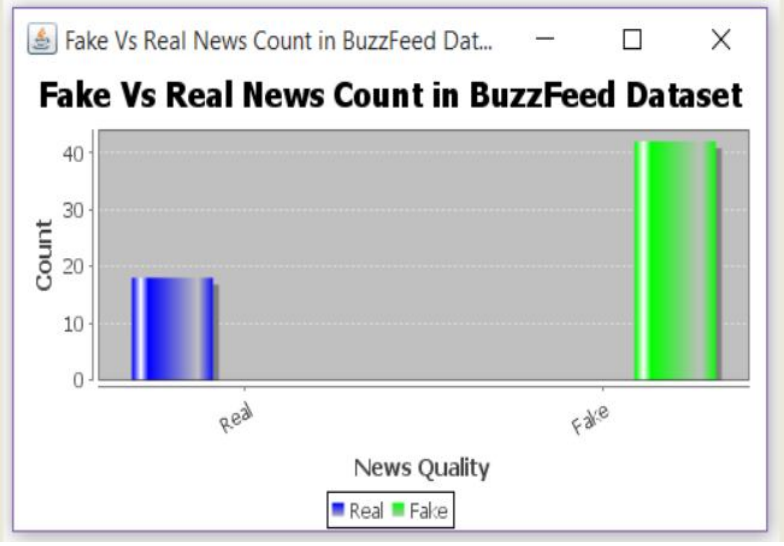


Fig 3 Fake Vs Real News Count BuzzFeed Dataset

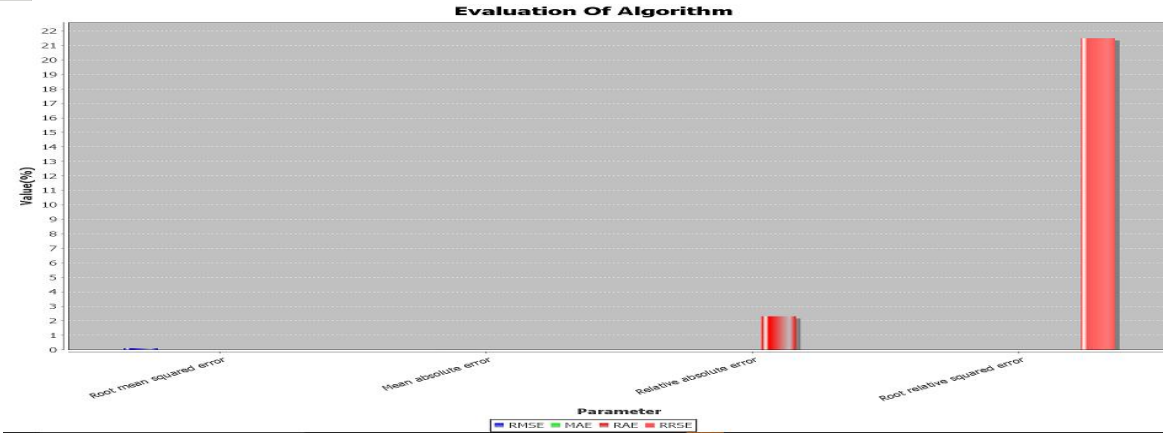


Fig .4 Evaluation of Algorithm

Applying the most perform CREDBANK and PHEME models to our BuzzFeed dataset shows both the pooled and CREDBANK-based models outperform the random baseline, but the PHEME-only model performs substantially worse, compare to this dataset better accuracy result obtained by BuzzFeed dataset show in fig 3.and fig 4 shows the accuracy results Mean absolute error accuracy of 0.0116% ,Root mean squared error accuracy of 0.1075%, Relative absolute error accuracy of 21.3112 %, compare to this algorithm better accuracy result obtained by Root relative squared error accuracy of 21.4997 %.

VII. CONCLUSIONS

This work demonstrates an automated system for detecting fake news in popular Twitter threads. Identifying misinformation is authoritative in online social media platforms, because information is circulated easily across the social media by unsupported sources. Automatically detect fake news using machine learning algorithm.

Using the reputation-based technique to each users profile and calculating sentiment score established based on the users history. Their tweets also solve the problem of assessing information credibility on Twit-ter. The issue of information credibility has comes under scrutiny.

VIII. ACKNOWLEDGMENT

This paper would not have been written without the valuable advices and encouragement of Mrs. S. S. Nandgaonkar, supervisor of ME Dissertation work. Author's special thanks to all the professors of Computer Engineering department of VPKBIET, Baramati, for their support and for giving an opportunity to work on fake news detection using Support vector machine learning Algorithm.

REFERENCE

[1] Shlok Gilda, "Evaluating Machine Learning Algorithms for Fake News Detection," 2017 IEEE 15th Student Conference on Research and Development (SCOReD).

[2] Mykhailo Granik, Volodymyr Mesyura, "Fake News Detection Using Naive Bayes Classifier", 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON).

[3] Mohammad Mehedi Hassan, Member, IEEE and Atif Alamri, "A Credibility Analysis System for Assessing Information on Twitter Member," IEEE TRANSACTIONS ON DEPENDABLE AND SECURE COMPUTING .

[4] Supanya Aphiwongsophon, Prabhas Chongstitvatana, "Detecting Fake News with Machine Learning Method," IEEE.

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

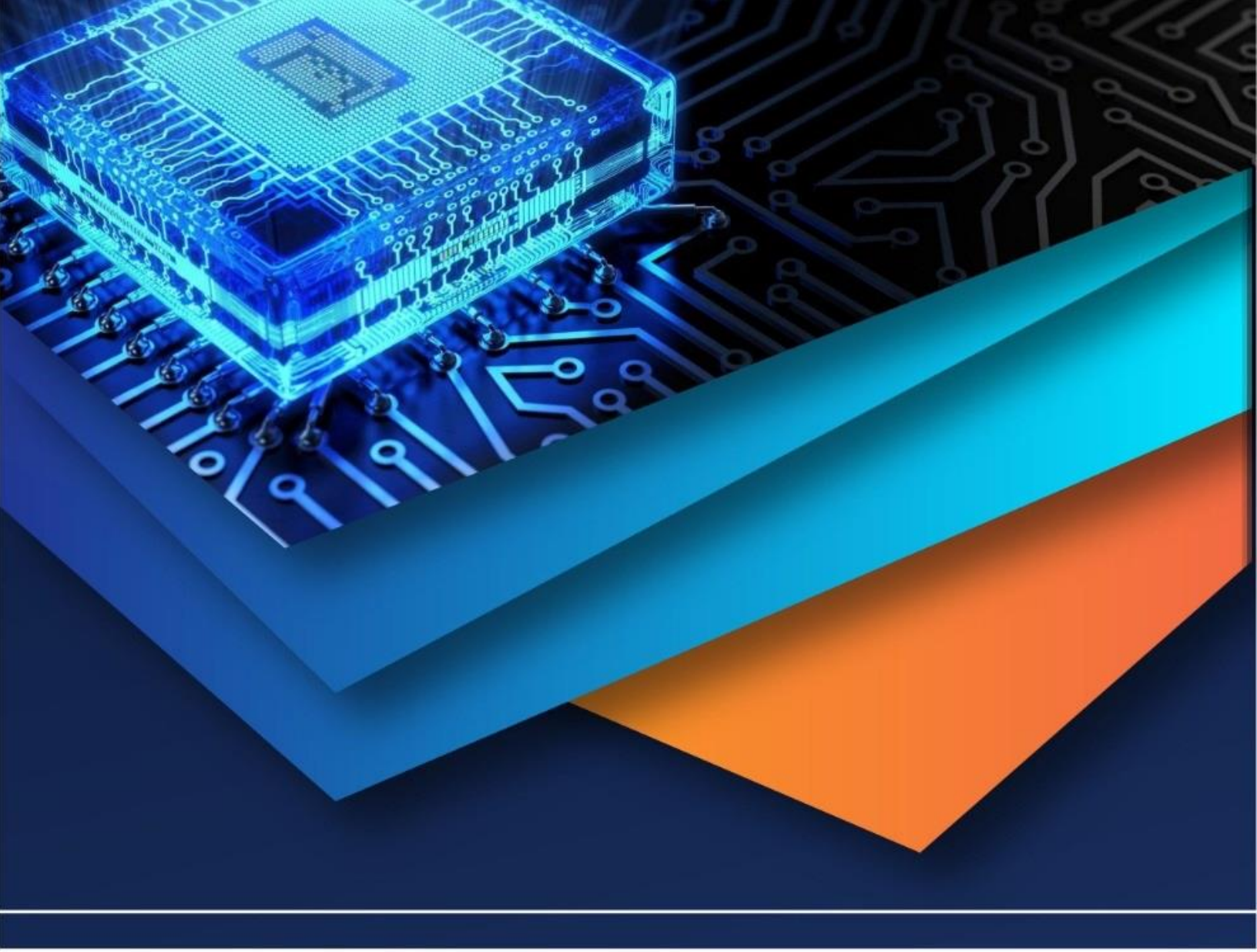
[6] Juan Cao, Junbo Guo, Xirong Li, Zhiwei Jin, Han Guo, and Jintao Li, "Automatic Rumor Detection on Microblogs: A Survey", arXiv:1807.03505v1 [cs.SI] 10 Jul 2018.

[7] Kai Shu,Suhang Wang, Huan Liu, "Understanding User Profiles on Social Media for Fake News Detection, "2018 IEEE Conference on Multimedia Information Processing and Retrieval

[8] Tanushree Mitra and Eric Gilbert, "CREDBANK: A Large-Scale Social Media Corpus with Associated Credibility Annotations , "School of Interactive Computing GVU Center Georgia Institute of Technology tmitra3, gilbert@cc.gatech.edu,Proceedings of the Ninth International AAAI Conference on Web and Social Media.

[9] Jacob Ross, Krishnaprasad Thirunarayan, "Features for Ranking Tweets Based on Credibility Newsworthiness",Kno.e.sis: Ohio Center of Excellence in Knowledge-enabled Computing Department of Computer Science and Engineering Wright State University Dayton, Ohio 45435 ross.138, t.k.prasad@wright.edu,2016 .

[10] Granik, m., & mesyura, V. 2017. "Fake news detection using naive Bayes classifier, " 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON), 900-903.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY