International Journal of Psychosocial Rehabilitation, Vol. 24, Issue 02, 2020
ISSN: 1475-7192

Detection of Fake News and Probability

[1]Sireddy.Naveen Kumar Reddy, [2]Dr.M.Sujatha

Abstract--*The quality of information is an important issue in the present age. Preciseness of news or information through social media is very important and it is an increasing problem in our society. Fake news is the misleading or wrong information that is spread over the internet to damage the popularity of person or organization. To overcome the problem, research should be done to classify whether the news is 'Real' or 'Fake'. To classify the news, a machine learning algorithm is adopted. In the process of classifying news classifier algorithm like Support Vector Machine algorithm, Naive Bayes algorithm, Decision Tree algorithm, Random Forest algorithm, and Logistic regression classifier algorithm are required. Initially data sets are extracted and it should be processed by the machine learning algorithms. Buzzfeed, Credbank, Phema are some of the datasets which can extract the required information from social media. In our model the contents are processed through all the algorithms which will predict whether the news on social media is 'Fake' or 'Real' and the probability of truth.*

Keywords-- *Fake news, classifier, machine learning, data sets.*

I INTRODUCTION

In olden days the process of news is collected by professionals. Then the news is edited and published to the society. The whole process is supervised by police. The chance of publishing wrong information is impossible. The trusted mode for publishing the information is by newspapers and televisions.

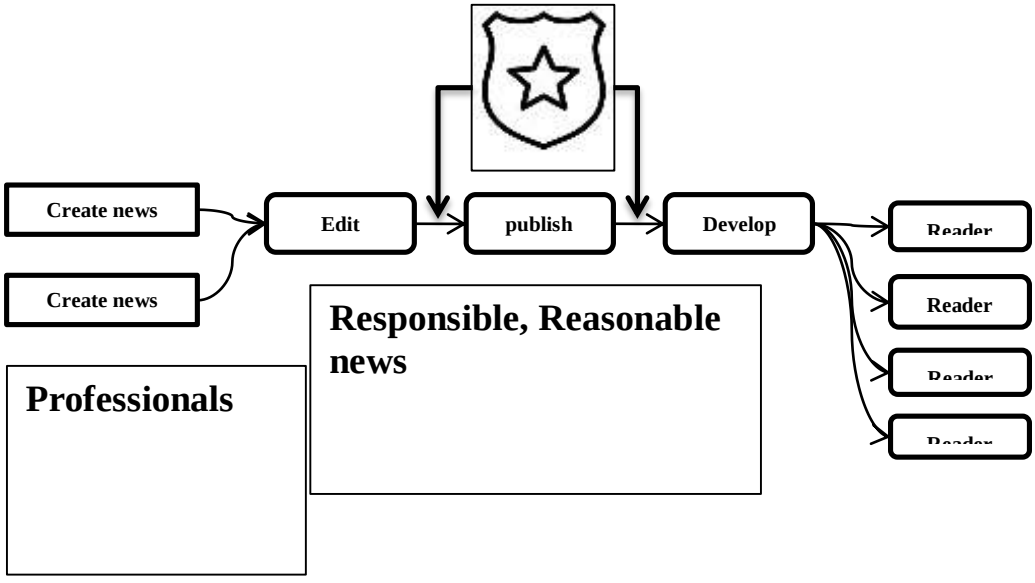


Figure 1: Traditional news publication pattern

[1]*Department of Electronics and Communication, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai-602105, Tamil Nadu, India, Email:sireddynaveen@gmail.com*

[2]*Department of Electronics and Communication, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai-602105, Tamil Nadu, India, Email:sujathamk.sse@saveetha.com*

Received: 23 Dec 2019 | Revised: 05 Jan 2020 | Accepted: 27 Feb 2020

Now a day's internet is getting popular in our society. There is a lot of change in the publication of news. In present days social media is the fastest mode for spreading the news in the world. But social media is a two sided knife [1]. It has advantages as well as disadvantages.

In social media there is cross verification of news by police and anyone can publish the news in society. Taking social media as a weapon so any people are publishing wrong information to spoil the popularity of some people and organizations. So there is a necessity to stop fake news publication in the social media. Due to absence of cross verification of publishing data through social media there is a threat in increase of fake news.

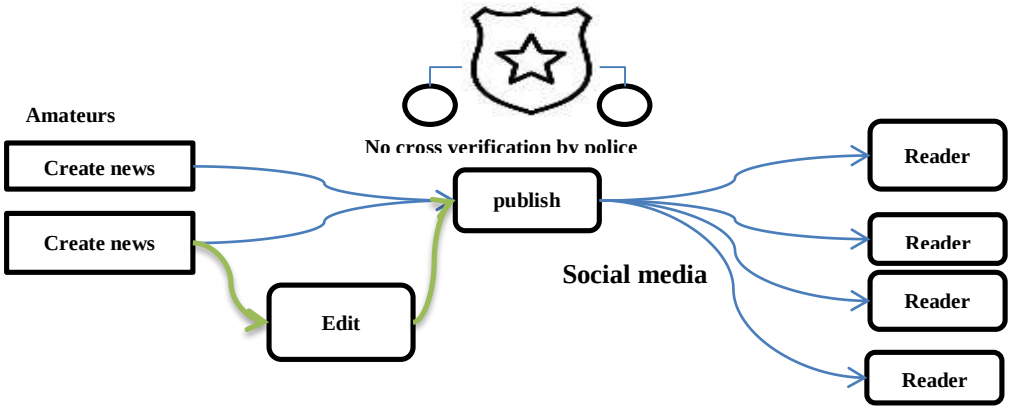


Figure 2: Internet news publication pattern

So there is a need to check the truth of the published information. To check it we need software which can classify which news is true. The required software should determine and classify the truth news. There are so many classifiers like Naive Bayes classifier, Random forest, support vector machine, stochastic gradient descent and Logistic regression classifier. [2]

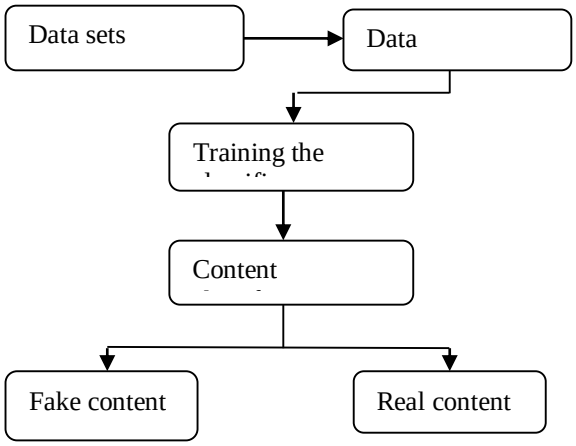


Figure 3: Process of classification

Received: 23 Dec 2019 | Revised: 05 Jan 2020 | Accepted: 27 Feb 2020

II CLASSIFIER

A. Data sets:

Some of the datasets are

- BUZZFEED
- CREDBANK
- PHEMA.

B. BUZZFEED:

Buzzfeed news makes datasets, libraries, analysis, guides and tools used in their articles available in github. The first data set was collected on September 2016 by journalists from its company over 2000 posts from some big verified FACEBOOK pages. The journalists from buzzfeed fact checked the posts noting as mostly true, group of both true and false, mostly false, no factual data.

C. CREDBANK::

This data set was a collection of posts from the social media platform, from its public feed and identifying the topics in the posts. In this dataset, contains more than 50,000,000 posts as false news. [5]

D. Phema Rumor Dataset:

University of warwick developed the phema rumor dataset in cooperation with swiss broadcasting company. It was developed in a set of 330 rumorous source posts across 140 tweets. In this posts 158 were classified as real, 69 was fake and 103 are unverified.

III CLASSIFIERS

- Naïve bayes classifier
- Decision tree classifier
- Random forest classifier
- Support vector machine classifier
- Logistic regression classifier
- Stochastic gradient descent

A. Naïve bayes classifier

Naïve bayes classifier is machine learning algorithm among the easiest Bayesian community model. This classification algorithm is primarily based on assumption of non-dependant events. This algorithm is a kind of probabilistic classifier. The probability of event will be between 0 and1. The output of the classifier offers the in between it. Classifier estimates the chance relationship for every and each case such as the prevalence of that given event or case.[3]

$$P(P|Q) = P(Q|P) * \frac{P(P)}{P(Q)}$$

Received: 23 Dec 2019 | Revised: 05 Jan 2020 | Accepted: 27 Feb 2020

B. DECISION TREE CLASSIFIER:

Decision tree classifier is a well-known classifier. The structure is like a flow chart with tree structure. Every node indicates a test on attribute value and every branch gives the effect of the event or test. The leaves of tree characterize class of the distribution. [4]

For example in the classification of a set containing numbers, color and underline

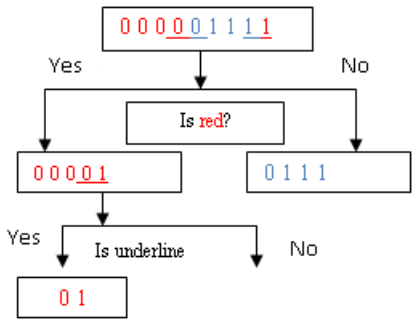
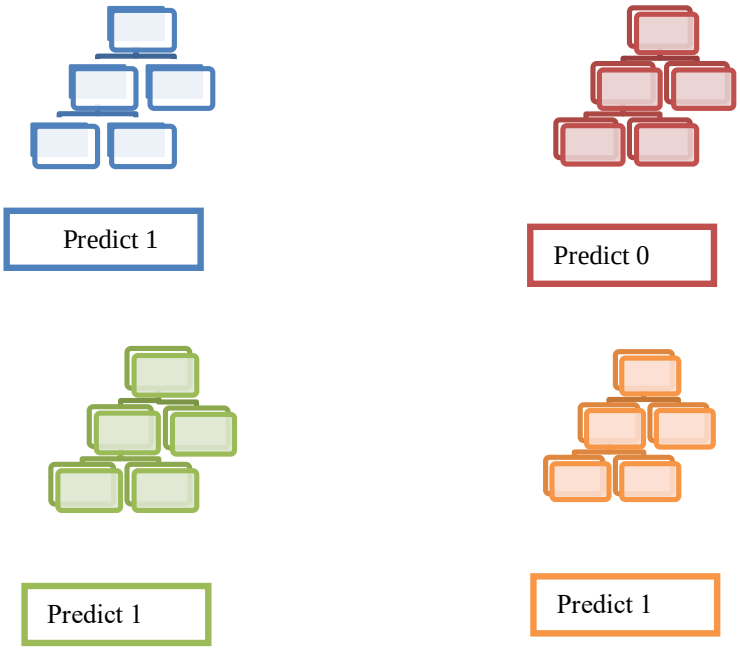


Figure 4: Decision tree classifier

C. The Random Forest Classifier:

Random forest classifier, by the name of the title refers, consists of a many independent decision trees that can function as an ensemble. Ensemble algorithms are the algorithms which combine algorithms of distinct kind or identical for classifying objects. Each individual decision tree in the random forest gives out a class prediction of the events or cases and the category with the most rated or voted will become our final prediction. [4]



Given three one's and one zero

Output: Prediction is 1

Received: 23 Dec 2019 | Revised: 05 Jan 2020 | Accepted: 27 Feb 2020

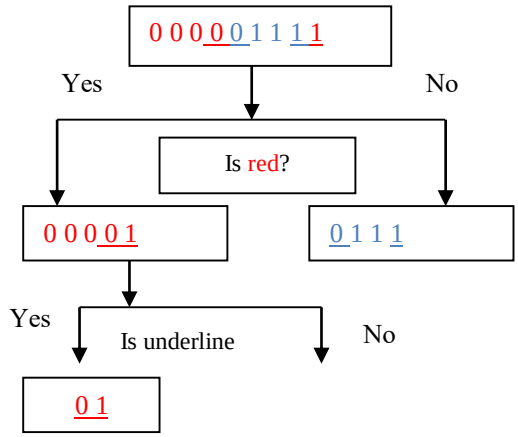


Figure 5: Random forest classifie

D. Support vector machine:

Support vector machine algorithm is abbreviated as SVM algorithm. It is mostly preferred classification algorithm as it can produce big accuracy. SVM algorithm may be used for classification tasks and regression.

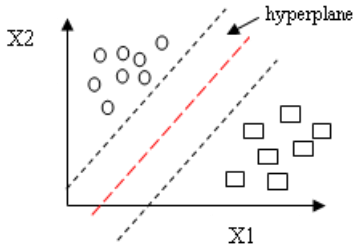


Figure 6: Support Vector Machine Graph

The goal of the SVM algorithm is to locate the hyperplane in an N-dimensional space. Thus the SVM can noticeably classify the fact points.[6]

E. Logistic regression classifier:

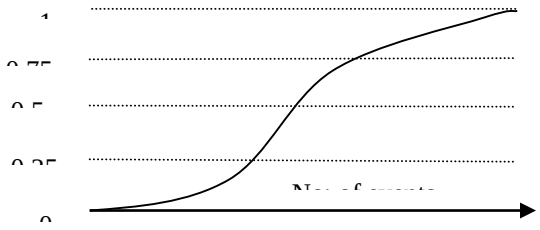


Figure 7: logic regression graph

Received: 23 Dec 2019 | Revised: 05 Jan 2020 | Accepted: 27 Feb 2020

Logistic regression classifier deals with the statistics and statistical learning method which is categorized under the supervised machine learning. The probability in this model is labeled between 0 and 1. [7]

IV Stimulation

In executing the project we can use two types of software. They are python and anaconda. When we are using this software we should have machine learning packages.

The packages are

- Sklearn
- Numpy
- Scipy

Using the required commands we should install it into our machine or PC. While using python, we need to setup the path variables. When the software is successfully installed we can execute the progr

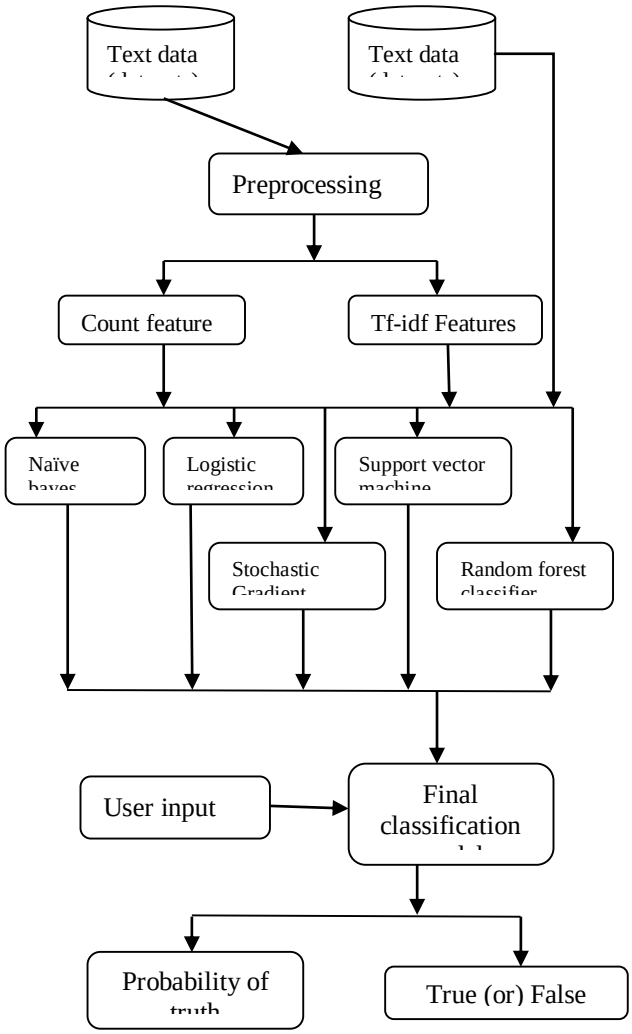


Figure 8: Stimulation module flow chart

For the fake news detection module datasets should trained. Preprocessing of the datasets is done and classification process is started. The input goes through all the classifiers and the final classification model will check the user input. The output the machine learning module will give the probability of truth of the news and gives whether the news is true or false.

V STIMULATION RESULTS

Table 1: Summary results of false news

S.No.	Input news	Probability of truth	True or false
1	Nearly half of African-American children under the age of 6 are living in abject poverty	0.7362282	True
2	When Obama was sworn into office, he DID NOT use the Holy Bible, but instead the Kuran (Their equivalency to our Bible, but very different beliefs)	0.4757452	False
3	Chemical weapons have been used probably 20 times since the Persian Gulf War	0.6322886	True
4	A flight from Atlanta to Houston was canceled due to a terrorist dry run	0.485419	False
5	Bill McCollumhas "recovered $200 million in Medicaid fraud."	0.5659615	True
6	Social Security is a Ponzi scheme.	0.2704136	False

VI CONCLUSION

In this paper we have learned the traditional and the present day news process of publication to the society. We came across through the different classifier and the datasets. The different classifier what we have discussed are like Naive Bayes classifier, Random forest, support vector machine, stochastic gradient descent and Logistic regression classifier. In this machine learning module we will get both the probability and truth of the news.

REFERENCES

1. Turk, Žiga. Technology as Enabler of Fake News and a Potential Tool to Combat It. European Parliament, 2018.

2. Ahmed, Hadeer, Issa Traore, and Sherif Saad. "Detecting opinion spams and fake news using text classification." Security and Privacy 1, no. 1 (2018): e9..
3. Jain, Akshay, and Amey Kasbe. "Fake news detection." In 2018 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS), pp. 1-5. IEEE, 2018.
4. https://towardsdatascience.com/understanding-random-forest-58381e0602d2.
5. Buntain, Cody, and Jennifer Golbeck. "Automatically identifying fake news in popular Twitter threads." In 2017 IEEE International Conference on Smart Cloud (SmartCloud), pp. 208-215. IEEE, 2017.
6. https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47.
7. https://machinelearningmastery.com/logistic-regression-for-machine-learning/
8. Parikh, Shivam B., and Pradeep K. Atrey. "Media-rich fake news detection: A survey." In 2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), pp. 436-441. IEEE, 2018..
9. Traylor, Terry, Jeremy Straub, and Nicholas Snell. "Classifying fake news articles using natural language processing to identify in-article attribution as a supervised learning estimator." In 2019 IEEE 13th International Conference on Semantic Computing (ICSC), pp. 445-449. IEEE, 2019.
10. Kim, Kyeong-Hwan, and Chang-Sung Jeong. "Fake News Detection System using Article Abstraction." In 2019 16th International Joint Conference on Computer Science and Software Engineering (JCSSE), pp. 209-212. IEEE, 2019.
11. Kaliyar, Rohit Kumar. "Fake news detection using a deep neural network." In 2018 4th International Conference on Computing Communication and Automation (ICCCA), pp. 1-7. IEEE, 2018.
12. Granik, Mykhailo, and Volodymyr Mesyura. "Fake news detection using naive Bayes classifier." In 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON), pp. 900-903. IEEE, 2017.
13. Kim, Namwon, Deokjin Seo, and Chang-Sung Jeong. "FAMOUS: Fake News Detection Model Based on Unified Key Sentence Information." In 2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS), pp. 617-620. IEEE, 2018.
14. Dey, Amitabha, Rafsan Zani Rafi, Shahriar Hasan Parash, Sauvik Kundu Arko, and Amitabha Chakrabarty. "Fake news pattern recognition using linguistic analysis." In 2018 Joint 7th International Conference on Informatics, Electronics & Vision (ICIEV) and 2018 2nd International Conference on Imaging, Vision & Pattern Recognition (icIVPR), pp. 305-309. IEEE, 2018.
15. Mahid, Zaitul Iradah, Selvakumar Manickam, and Shankar Karuppayah. "Fake News on Social Media: Brief Review on Detection Techniques." In 2018 Fourth International Conference on Advances in Computing, Communication & Automation (ICACCA), pp. 1-5. IEEE, 2018.
16. Maharaja, D., & Shaby, M. (2017). “Empirical Wavelet Transform and GLCM Features Based Glaucoma Classification from Fundus Image.” International Journal of MC Square Scientific Research, 9(1), 78-85.
17. Saravanan, N. (2013). “Hand Geometry Recognition based on optimized K-means Clustering and Segmentation Algorithm.” International Journal of MC Square Scientific R*esearch*, 5(1), 11-14.