



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJIRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 11, Issue 1, January 2022

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 7.569

 9940 572 462

 6381 907 438

 ijirset@gmail.com

 www.ijirset.com



Fake News Detection Using Machine Learning

Vishal Shinde¹, Y Sai Varun¹, Himanshu Parmar¹, Pratima Patil²

Student, Department of Computer Engineering, Trinity Academy of Engineering, Pune, India¹

Professor, Department of Computer Engineering, Trinity Academy of Engineering, Pune, India²

ABSTRACT: Fake news nowadays is an important aspect in the life of social media, and in the political world. This paper makes an analysis of the research related to fake news detection and explores the traditional machine learning models to choose the best, in order to create a model of a product with supervised machine learning algorithm, that can classify fake news as true or false, by using tools like python, decision tree and other algorithms. We propose in this paper, a fake news detection using machine learning techniques. In other words if we're saying, the model is counted pretty much as good in accuracy if the model produces foreseen outcomes that is the prediction of fake and true news.

KEYWORDS: Machine Learning, Fake News, Fake News Detection.

I. INTRODUCTION

The advent of the World Wide Web and the rapid adoption of social media platforms (such as Facebook and Twitter) paved the way for information dissemination that has never been witnessed in human history before. Besides other use cases, news outlets benefitted from the widespread use of social media platforms by providing updated news in near real time to its subscribers. The news media evolved from newspapers, tabloids, and magazines to a digital form such as online news platforms, blogs, social media feeds, and other digital media formats[1].

Fortunately, there are a number of computational techniques that can be used to mark certain articles as fake on the basis of their textual content [2]. Majority of these techniques use fact checking websites such as "PolitiFact" and "Snopes". There are a number of repositories maintained by researchers that contain lists of websites that are identified as ambiguous and fake[3].

There are quite a few reasons why this problem is uniquely challenging. First, detecting fake news only on the basis of news content is not easy since it is written with the aim of misleading readers. Second, the fake news content differs a lot in terms of styles, topics and media platforms and distorts truth using diverse linguistic styles. For instance, fake news may cite true evidence within the incorrect context to support a claim that is not factual[4]. This paper proposes a methodology to create a model that will detect if an article is authentic or fake based on its words, phrases, sources and titles, by applying supervised machine learning algorithms on an annotated (labeled) dataset that are manually classified and guaranteed. Then, feature selection methods are applied to experiment and choose the best fit features to obtain the highest precision, according to confusion matrix results. We propose to create the model using different classification algorithms.

II. RELATED WORK

It pointed out various sources of media and made the suitable studies whether the submitted article is reliable or fake. The paper utilizes models based on speech characteristics and predictive models that do not fit with the other current models[6]. The paper used naive Bayes classifier to detect fake news by Naive Bayes. This method was performed as a software framework and experimented it with various records from the Facebook, etc., resulting in an accuracy of 74%. The paper neglected the punctuation errors, resulting in poor accuracy[7].

The estimated various ML algorithms and made the researches on the percentage of the prediction. The accuracy of various predictive patterns included bounded decision trees, gradient enhancement, and support vector machine were assorted. The patterns are estimated based on an unreliable probability threshold with 85-91% accuracy[8]. It proposes a methodology to create a model that will detect if an article is authentic or fake based on its words, phrases, sources and titles, by applying supervised machine learning algorithms on an annotated (labeled) dataset, that are manually classified and guaranteed[9].

Their square measures some tools that have been developed to spot fake news that spreads through examining lexical selection that seems in headlines and different intense language structures (Chen, Conroy, and Rubin 2015b). Another tool, developed to spot fake news on Twitter, includes an element known as the Twitter Crawler that collects and stores tweets in a very info (Atodiresei, Tan⁻ azalea, and Iftene 2018). Once a ~ Twitter user desire to ascertain the accuracy



of the news found they'll copy a link into this application, when that the link are going to be processed for fake news detection. This method is made on associate degree rule is a known as the NER (Named Entity Recognition) (Atodiresei, Tañ azalea, and Iftene 2018). ~ Their square measure several on the market approaches to assist the public to spot fake news and this paper aims to reinforce understanding of those by categorizing these approaches as found in existing literature.

III. METHODOLOGY

In this method supervised machine learning is used for classifying the dataset. The first step in this classification problem is the dataset collection phase, followed by pre-processing, implementing features selection, then performing the training and testing of the dataset and finally running the classifiers [10][11]. Figure [1] describes the proposed system methodology. The methodology is based on conducting various experiments on dataset using the algorithms named Random forest, KNN and Naïve Bayes, majority voting and other classifiers.

IV. EXPERIMENTAL RESULTS

The experiments will be conducted individually on each algorithm, and in combination among them for the purpose of best accuracy and precision. The LIAR-PLUS Master that has been used for cleaning and extracting the data set, and the algorithms are applied. This dataset has automatically extracted proof sentences from the full-text verdict article published in Politifact by journalists.

The data is divided into two parts: The first section, which consists of 75% of the data, is trained data, where the algorithm detects the real news and false news ,then the data is labeled in the form of 0 and 1 where 0 is for false news and 1 for true news. After that, the rest of the data, which is 25% of it, will do a test on it, so that it is sure whether the news is nature or forged, and then return it in case it was right or wrong, and according to the percentage of right and wrong, the algorithm percentage will be formed.

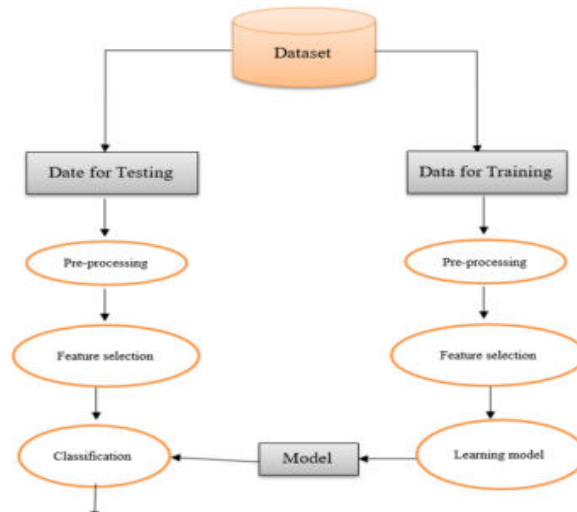


Figure [1] - Block Diagram

V. ALGORITHMS

Machine Learning (ML) is a class of algorithms that help software systems achieve more accurate results without having to reprogram them directly. When the training is completed, the algorithm splits the learned levels into new data [12].

A. Decision Tree

The decision tree is an important tool that works based on flow chart-like structure that is mainly used for classification problems. Each internal node of the decision tree specifies a condition or a “test” on an attribute



and the branching is done on the basis of the test conditions and result. Finally the leaf node bears a class label that is obtained after computing all attributes.

Decision Tree Pseudocode

GenerateDecisionTree(Sample s, features F)

1. If stop conditions(S,F) = true then
 2. root = create Node()
 3. root.test condition = find bestSplit(s,f)
 4. $v = \{v \mid v \text{ a possible outcome of root.test condition}\}$
 5. for each value v belongs V
 6. $sv = \{s \mid \text{root.test condition}(s) = v \text{ and } s \text{ belongs } S\}$;
 7. child = Tree Growth(sv ,F);
 8. Grow child as a descent of roof and label the edge (root → child) as v
- Return root

B. Random Forest

Random Forest is built on the concept of building many decision tree algorithms, after which the decision trees get a separate result. The results, which are predicted by a large number of decision trees, are taken up by the random forest.

Random Forest Pseudocode

To make n classifiers: For i = 1 to n do

Sample the training data T randomly with replacement for T_i output

Build a T_i -containing root node, N_i

Call BuildTree (N_i)

end For

BuildTree (N):

If N includes instances of only one class, then returns

else

Select z% of the possible splitting characteristics at random in N

Select the feature F with the highest information gain to split on

Create f child nodes of N, N_i, \dots, N_f , where F has f possible values (F_1, \dots, F_f)

For i = 1 to f do

Set the contents of N_i to T_i , where T_i is all instances in N that match F_i

Call Buildtree(N_i)

end for

end if

C. Naive Bayes

This algorithm works on Bayes theorem under the assumption that its free from predictors and is used in multiple machine learning problems. For example, the fruit will be classified as an apple when its of red color, swirls, and the diameter is close to 3 inches. Regardless of whether these functions depend on each other or on different functions, and even if these functions depend on each other or on other functions, Naive Bayes assumes that all these functions share a separate proof of the apples.

Naive Bayes Equation

$$P(c/x) = P(x/c)P(c)/P(x) \quad P(c/X) = P(x_1/c) \times P(x_2/c) / \dots / P(x_2/c) \times P(c)$$

Where: $P(c/X)$ is the posterior Probability.

$P(x/c)$ is the Likelihood.

$P(c)$ is the Class Prior Probability.

$P(x)$ is the Predictor Prior Probability.

Naive Bayes Pseudo-code

Training dataset T, $F = (f_1, f_2, f_3, \dots, f_n)$ // value of the predictor variable in testing dataset.



Output: A class of testing dataset.

D. KNN Algorithm

KNN classifies new positions based on most of the sounds from the neighboring k with respect to them. The position assigned in the class is highly mutually exclusive between the nearest neighbors K , as measured by the role of the distance.

KNN Pseudo-code

Classify (X, Y, x) // X : training data, Y : class labels of X , x : unidentified sample

For $i = 1$ to m do

Calculate distance $d(X_i, x)$

end for

Calculate set I containing indices for k smallest distances $d(X_i, x)$

return majority label for Y_i where i to I

VI. CONCLUSION

This paper focuses on detecting the fake news by reviewing it in two stages: characterization and disclosure. In the first stage, the basic concepts and principles of fake news are highlighted in social media. During the discovery stage, the current methods are reviewed for detection of fake news using different supervised learning algorithms. Fake news detection has many open issues that require the attention of researchers.

For instance, in order to reduce the spread of fake news, identifying key elements involved in the spread of news is an important step. Graph theory and machine learning techniques can be employed to identify the key sources involved in the spread of fake news. Likewise, real time fake news identification in videos can be another possible future direction. By considering our project's ideology people can at least be able to check whether the news they have got in front of their eyes are legit or not and the people will become more aware of the fake news circulation.

REFERENCES

- [1] A. Douglas, "News consumption and the new electronic media," 7e International Journal of Press/Politics, vol. 11, no. 1, pp. 29–52, 2006.
- [2] N. K. Conroy, V. L. Rubin, and Y. Chen, "Automatic deception detection: methods for finding fake news," Proceedings of the Association for Information Science and Technology, vol. 52, no. 1, pp. 1–4, 2015.
- [3] F. T. Asr and M. Taboada, "Misinfotext: a collection of news articles, with false and true labels," 2019.
- [4] Shu, K., Sliva, A., Wang, S., Tang, J. and Liu, H. (2018). Fake News Detection on Social Media: A Data Mining Perspective. [online] Arxiv.org. Available at: <https://arxiv.org/abs/1708.01967> [Accessed 16 Mar. 2018].
- [5] Economic and Social Research Council. Using Social Mmedia.
- [6] MykhailoGranik and VolodymyrMesyura. "Fake news detection using naive Bayes classifier". First Ukraine Conference on Electrical and Computer Engineering (UKRCON). Ukraine : IEEE.2017.
- [7] Gilda, S. "Evaluating machine learning algorithms for fake news detection." 15th Student Conference on Research and Development (SCORed) (pp. 110-115). IEEE. 2017
- [8] Akshay Jain and AmeyKasbe. "Fake News Detection." 2018 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS). Bhopal, India: IEEE.2018
- [9] Z Khanam et al 2021 IOP Conf. Ser. "Fake News Detection Using Machine Learning Approaches" Mater. Sci. Eng. 1099 012040.
- [10] Z Khanam et al 2021 IOP Conf. Ser. "Fake News Detection Using Machine Learning Approaches" Mater. Sci. Eng. 1099 012040.
- [11] Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake news detection on social media: A data mining perspective. ACM SIGKDD explorations newsletter, 19(1), 22-36.
- [12] SumeetDua, Xian Du. "Data Mining and Machine Learning in Cybersecurity". New York: Auerbach Publications.19 April 2016.



**Impact Factor:
7.569**



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details