iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** XII **Month of publication:** December 2022

DOI: https://doi.org/10.22214/ijraset.2022.47843

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: 7.538
Volume 10 Issue XII Dec 2022- Available at www.ijraset.com

Fake News Detection Using Machine Learning

Akansha Jain[2], Dilpreet Kaur[2]
[1, 2]*Student, JIMS*

Abstract: ***Fake news is one of the most popular phenomena that effects on our social life[2]. Nowadays, growing faux information turns into very smooth due to user's considerable the usage of internet. It is a severe problem as its cap potential to motive a variety of social damage.In this paper, we can make an evaluation of studies associated with faux information detection. We will use system studying algorithms to pick the first-class version for detection of faux or actual information. The end result will display the accuracy of dataset. In this paper, we can use supervised studying like- Logistic Regression, svm, Naïve Bayes etc.***
Keywords: Fake news, Real news, Machine learning, Algorithm, Information, Positive, Negative

I. INTRODUCTION

Fake News is fake or deceptive facts provided as information. Fake information has additionally been known as junk information, pseudo-information, fake information. The term "Fake information" has turn out to be a common, mainly for describing the deceptive & fake articles to make cash from net page's[1]. Fake information intention is to control human beings opinions. Today we agree with in what we see on social media & we don't pursue to test if supplied information is actual or faux[1]. Sometimes, it's far tough to differentiate among faux and actual information due to the fact we want to spend long term to test the references of information.

This paper endorse a method to create a version on the way to come across if information is actual or faux primarily based totally on it's words, phrases, source & name via way of means of making use of supervised system studying algorithms at the dataset[6]. The algorithm will test the dataset, it will give accuracy of correctness of news means prediction of news whether it is real or fake. Then, we will choose best model on the basis of accuracy obtained from the algorithms.

II. LITERATURE REVIEW

Reham Jehad &Suhad A. Yousif, "Fake news Classification using Random Forest & Decision Tree (J48)", 2020 [2]

In this research paper, they utilize 2 different machine learning algorithms (Random forest, Decision Tree) to detect Fake news. They have taken dataset & pre-processed it by removing unnecessary special character, numbers, white spaces etc. & describe the workflow design of these 2 algorithms &at last experimental results. But in this paper, they have used only 2 algorithms of machine learning, there are many more algorithms that give more accuracy rate.

Uma Sharma, Sidarth Saran, Shankar M. Patil, "Fake News Detection using Machine Learning Algorithms", 2020 [5].In this research paper, they have used three algorithms of machine learning, main goal of this paper was to look at how these particular methods work for this particular problem given a manually labelled news dataset & to support the thought of using AI for fake news detection. In order to curb the phenomenon, it takes input from the user &classify it to be true or false. There is no dataset, which can be checked, user have to input the news manually.

Z Khanam1 , B N Alwasel1 , H Sirafi1 and M Rashid2, "Fake News Detection Using Machine Learning Approaches",2020[6]. In this paper, we analyze research related to fake news detection, examine traditional machine learning models and choose the best one, to create a model fora product equipped with supervised machine learning algorithm, which can classify fake news as true or false.

Using tools like Python Scikit-Learn, NLP for text analysis. This process leads to feature extraction and vectorization.This paper focuses on detecting the fake news by reviewing it in two stages: characterization and disclosure. The displayed fake news detection approach that is based on text analysis in the paper utilizes models based on speech characteristics and predictive models that do not fit with the other current models.

The researcher proposed a method for mining the productive information from web using classification algorithms particle swarm optimization and support vector machine[13].Author proposed a big data query optimization system for sentiment analysis of telecom customer tweets. The hybrid system suggested by researcher influenced the repeated neural network [12]. Researcher proposes a framework in which dataset of customers is analyzed using spearman method [13].

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue XII Dec 2022- Available at www.ijraset.com

III. STUDY OF FAKE NEWS DETECTION

A. *Methodology*

This paper helps to identify fake news and real news. We identify news on the basis of machine learning algorithms. We have studied and trained the model with 3 algorithms.In this paper, we have used Python and its libraries. Python has various set of libraries, which can be easily used in machine learning. [5]. To identify the fake and real news following steps are used:-

Step 1: Choose appropriate fake news dataset

Step 2: Pre-Process the dataset

Step 3: Classify the dataset using algorithms

Step 4: Evaluate model performance using different metrics like- accuracy, correctness, recall, precision etc.

All these algorithms get as precise as possible.The dataset is applied to different algorithms in order to detect the fake news. The accuracy of the results obtained are then analyzed to conclude the final result.

B. *Algorithms*

1) *Logistic Regression:* Logistic regression is a supervised classification algorithm. It is used to predict a binary outcome based upon a set of independent variables. A binary outcome is one where there are simply two implicit scenarios—either the event happens (1) or it does not happen (0). Independent variables are those variables or factors which may impact the results (or dependent variable).

	Text	Label	Length
0	house dem aide' even see comey' letter jason...	REAL	3338
1	ever get feeling life circles roundabout rathe...	FAKE	2857
2	truth might get fired October 29 2016 tension	REAL	5328
3	videos 15 civilians killed single us airstrike...	REAL	2268
4	print iranian woman sentenced six years prison...	REAL	688

Table 1- After Preprocessing step[2]

a) Like in our case study we have dependent variable which are attribute label or outcome i.e. either the news is Real or fake.

b) The independent variable in our case study is the text attribute in dataset which influence our outcome that whether the give text is in the real class or in fake.

c) So: Logistic regression is the accurate type of algorithm to use when you're working with binary data. We are dealing with binary data when the output or dependent variable is bipartite or categorical in nature; in other words, if it fits into one of two categories (such as "yes" or "no", "true" or "false", "pass" or "fail", and so on).

d) Here the Label attribute is the dependent variable that we are predicting and Text is the independent variable.

e) So this is the equation of Sigmoid Function from which we predict that the Text is Fake or Real. [3]

f) Sigmoid function =y = 1/1+e^-x

g) Here eis Eulers Constant i.e . 2.718

h) We are trying to convert independent variable into expression of probability that ranges between 0 and 1 with respect to dependent variable.

i) In the Sigmoid function equation "x" is independent variable and "y" is the dependent variable i.e. Class Label

j) when the P(text=yes)≥0.5, then we say the text is real news.

k) When the P(text=yes)<0.4, then we say the text is fake news.

l) The probability will always range between 0 and 1. In the case of binary classification, the probability of real news and fake news will sum up to 1 as shown in Figure 1

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue XII Dec 2022- Available at www.ijraset.com

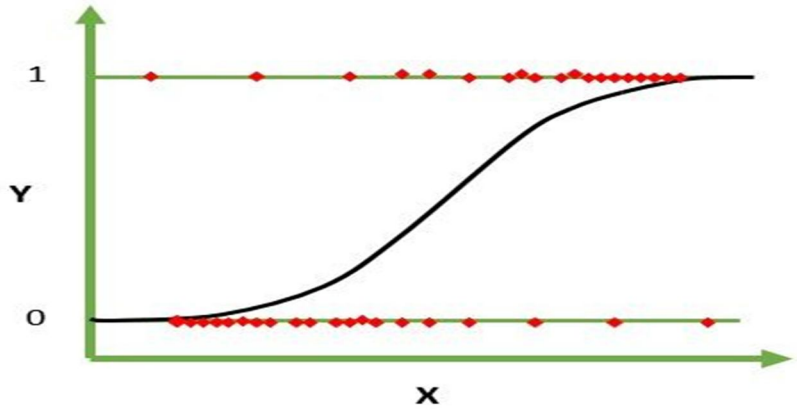


Figure 1: Result

2) *Support Vector Machines:* They're specifically effective at classification, numeral prediction, and pattern recognition tasks. SVMs find a line in between different classes of data similar that the distance on either side of that line or hyperplane to the next-closest data points is maximized.

In different words, help vector machines calculate a most-margin boundary that results in a homogeneous partition of all statistics factors. This classifies an SVM as a most margin classifier.

On the threshold of both facet of a margin lies pattern statistics labelled as help vectors, with at the least 1 help vector for every magnificence of statistics. These help vectors constitute the limits of the margin, and may be used to assemble the hyperplane bisecting that margin.

$\vec{W} \cdot \vec{X} + b = 0$ (1)

$Y = MX + b$ (2) [3]

Equations 2 and Equations1 represent the formulas for a line or hyperplane respectively. For all sample data *x*, an SVM should find weight such that the data points will be separated according to a decision rule. To elaborate, lets assume we have a set of 0 and 1 which is represented as negative and positive values in a two-dimensional Euclidean space, along with an original straight line (drawn in green) between the two classes of data points

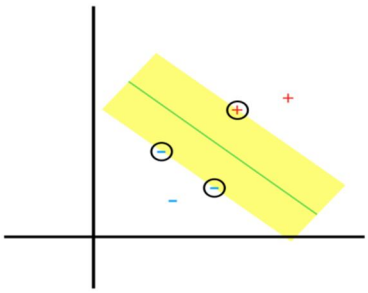


Figure 2: Euclidean space

Class 0 ,y=-1(for Fake News)

Class 1 , y=1(for Real News)

The yellow space in Figure.2 shows the periphery between the points of opposing classes that are closest to each other. The points that are encircled are the support vectors.

$$d(X^T) = \sum_{i=1}^{l} y_i \alpha_i X_i X^T + b_0$$

[3]

X is the input that is the text we are providing

Y is the output that is the class label that is in class 0 and 1

X^T is the new text input that we are providing

l = no. of support vectors

©IJRASET: All Rights are Reserved

IV. RESULTS

This tells us on which side of hyperplane are new input falls
If the sign is +ve then SVM predicts that new input belongs to class +1 that is the article is Real News.
If the sign is –ve the SVM predicts that new input belongs to class -1 that is the article is Fake News.
NAÏVE BAYES

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$$

P(A|B) is the posterior possibility, or the possibility of A given evidence B
P(B|A) is the liability, or the possibility of B given A
P(A) is the prior probability, or the probability of A, without taking any substantiation into account
P(B) is a normalization constant to gain a probability density function with a sum of 1

- Here C1 is for REAL Class and C2 is for Fake Class.
- Prior probability for C1= Number of C1 objects/ Total number of objects
- Prior probability for C2= Number of C2 objects/ Total number of objects

Having formulated our prior probability, we are now ready to classify a new object (TEXT circle). Since, all the objects are well clustered, it is reasonable to assume that the more GREEN (or RED) objects within the vicinity of X, the more likely that the new cases belong to that particular color. To measure this likelihood, we draw a circle around X which encompasses a number (to be chosen a prior) of points irrespective of their class labels.
Green Color is for class C1 and Red Color is for class C2.
Then, we will calculate the number of points in the circle belonging to each class label. From this we calculate the likelihood:
Likelihood of TEXT in class C1= number of C1 in the vicinity of that TEXT/Total number of C1 class
Likelihood of TEXT in class C2= number of C2 in the vicinity of that TEXT/Total number of C2 class.

V. EXPERIMENTAL RESULTS

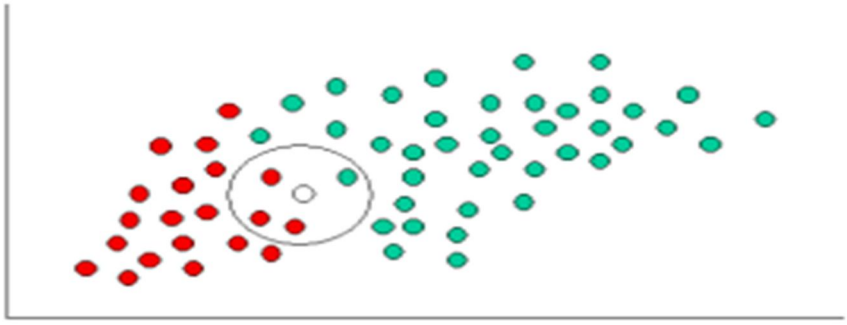


Figure 3: Result [4]

Posterior probability of Text being in C1 = p(C1|TEXT)= P(C1)*P(TEXT|C1)/P(TEXT)
Posterior probability of Text being in C2 = P(C2|TEXT)= P(C2)*P(TEXT|C2)/P(TEXT)
Finally, we classify TEXT as REAL OR FAKE WHOSE class membership achieves the largest posterior probability.

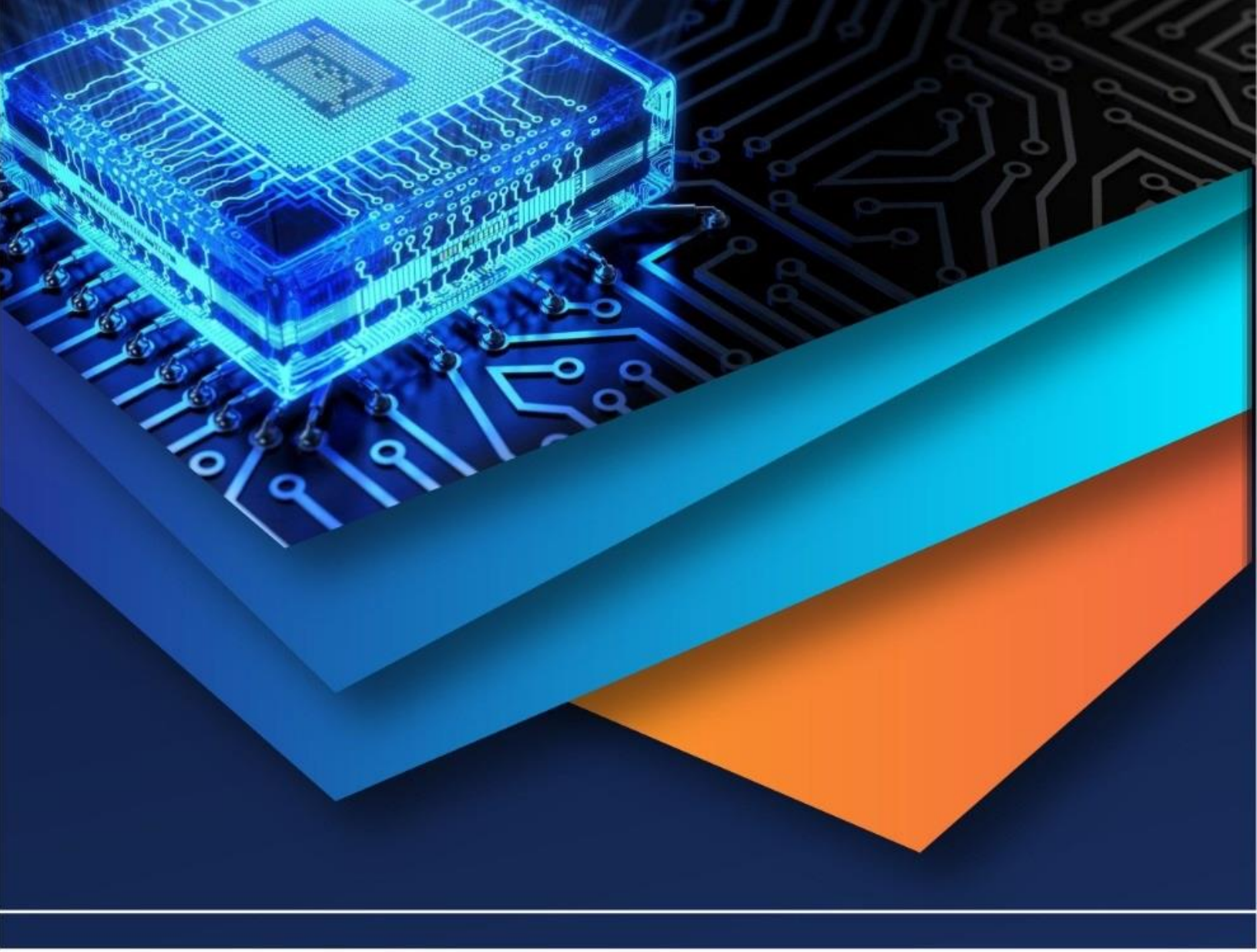
VI. CONCLUSION

Although there is evident success in detection of fake news and posts using various Data Mining approaches. However ever changing characteristics and features of fake news in social media networks is posing a challenge in categorization of fake news.Due to increasing use of internet, it's now easy to spread fake news. A Large number of persons are regularly connected with internet and social media platforms. There's no any restriction while posting any news on these platforms. So some of the people take the advantage of these platforms and start spreading fake news against the individuals or associations. This can destroy the reputation of an individual or can affect a business. Through fake news, the view of the people can also be fluctuated for a political party. There's a necessity for the way to detect these fake news. Data mining classifiers are using for different purposes and these can also be used for detecting the fake news. These classifiers are first trained with a data set called training data set. After that, these classifiers will automatically detect fake news.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

REFERENCES

[1] Gilda S., "Evaluating machine learning algorithms for fake news detection", 2017.
[2] RehamJehad&Suhad A. Yousif, "Fake news Classification using Random Forest & Decision Tree (J48)", 2020
[3] Ahmed, H., Traore, I., &Saad, S. (2017). Detection of online fake news using n-gram analysis and machine learning techniques. Proceedings of the International Conference on Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments, 127–138, Springer, Vancouver, Canada, 2017. https://doi.org/10.1007/978-3-319-69155-8_9
[4] Abdullah-All-Tanvir, Mahir, E. M., Akhter S., &Huq, M. R. (2019). Detecting Fake News using Machine Learning and Deep Learning Algorithms. 7th International Conference on Smart Computing & Communications (ICSCC), Sarawak, Malaysia, Malaysia, 2019, pp.1-5, https://doi.org/10.1109/ICSCC.2019.8843612
[5] Uma Sharma, Sidarth Saran, Shankar M. Patil, "Fake News Detection using Machine Learning Algorithms", 2020
[6] Z Khanam1 , B N Alwasel1 , H Sirafi1 and M Rashid2, "Fake News Detection Using Machine Learning Approaches",2020
[7] Al Asaad, B., &Erascu, M. (2018). A Tool for Fake News Detection. 2018 20th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC), Timisoara, Romania, 2018, pp.379-386. https://doi.org/10.1109/SYNASC.2018.00064
[8] Donepudi, P. K., Ahmed, A. A. A., Saha, S. (2020a). Emerging Market Economy (EME) and Artificial Intelligence (AI): Consequences for the Future of Jobs. Palarch's Journal of Archaeology of Egypt/Egyptology, 17(6), 5562- 5574. https://archives.palarch.nl/index.php/jae/article/view/1829
[9] AhlemDrif, ZinebFerhatHamida, "Fake News Detection Method Based on Text-Features",2019
[10] Uma Sharma, Sidarth Saran, Shankar M. Patil, "Fake News Detection Using Machine Learning Algorithms", 2020
[11] Chugh A., Sharma V.K., Bhatia M.K., Jain. C (2021) A Big Data Query Optimization Framework for Telecom Customer Churn Analysis. In: 4TH International Conference on Innovative Computing and Communication, Advances in Intelligent Systems and Computing.Springer, Singapore.
[12] Kaushik N., Bhatia M.K. ,Rastogi S. (2020) SVM and cross validation using RStudio. International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249-8958, Volume-10 Issue-1, October 2020.
[13] Kaushik N., Bhatia M.K. (2020) Information Retrieval from Search Engine Using Particle Swarm Optimization. In: Sharma H.,Govindan K., Poonia R., Kumar S., El-Medany W. (eds) Advances in Computing and Intelligent Systems. Algorithms for Intelligent Systems. Springer, Singapore. https://doi.org/10.1007/978-981-15-0222-4_11



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)