International Journal of Interdisciplinary Innovative Research &Development (IJIIRD)
ISSN: 2456-236X
Vol. **05** Issue **02** | **2021**

Fake News Detection Using Multiple Machine-Learning Algorithms.

Shashank H
Department of Computer Application, Jain University, Bangalore, India

ABSTRACT

With the ease of access to internet and also the various social media networks such as Facebook and Twitter. Which are almost used by half of the population on earth. The positive side of social media comes with sharing information to a large group of people and it is much faster compared to traditional news broadcast using television networks. Keeping the advantages aside let's discuss about the dark side of the social media the place for fake news and conspiracy. The amount of fake news that spreads through social media has become a threat to general public and differentiating the fake news from the real news has become a big problem. The reason behind this can be considered as there is no automatic engines within the social media network this allows users in the social network to post anything they like and this post gets shared between many users which can lead to confusion and unable to understand which message is real and which is fake. The project aims to classify the messages posted on social media and helps identifying the news or the post related to a specific topic is real or fake.

Keywords: Logistic Regression, Decision Tree Classifier, Gradient Boosting Classifier, Random Forest Classifier, Natural Language Processing.

1. INTRODUCTION

With the increase of spread of hoax or fake news on social media it's become difficult for general public to understand what's true and what's fake. The spread of fake news is faster than compared to the real or genuine news which is from a genuine source and which is verified before broadcasting the news to general public. The fake news poses a threat to modern day journalism sometimes the fake news is uploaded to the internet by using the name of a well-known news organisation this makes people that the news is coming from a genuine but the reality is that it's not real and everyone ends up believing and sharing the hoax news. The spread of is considered as a major tool in the presidential election of united states in 2016. It has also become a threat in healthcare domain also during the high time of COVID-19 the fake news cause panic among many individuals where they ended up believing the fake news was real and started spreading the same news. The ways to verify that the news is fake or not is by checking the source domain which is displaying the news and if the domain is genuine website the you might consider that the news is real. It is considered as a best practice to verify on a trusted domain about the news which is received from any social media.

The proposed approach used five machine learning techniques - Logistic regression, Decision tree classifier, Gradient boosting classifier, Random forest classifier and Natural language processing. These methods help us to get a better overall score of the content which will help us deciding if the news is real or fake. This combined with manual testing method which combines multiple machine learning algorithms which will help us to identify the genuineness of the news with multiple techniques at a single place.

2. ALGORITHMS AND METHODOLOGIES

2.1 Natural Language Processing

The natural language processing or NLP is a machine learning technique where you teach and train the computer to understand the human language. It can also be considered as communicating with the computer using human language. Its not easy to interact with the computer using natural language that humans speak the computer is designed to only communicate with the help of machine level language any operation computer does the instructions are converted to machine level language and are fed to the processing unit. In the proposed system by using NLP the processing of the data is done which helps the computer to decide if the set of data is well formatted and written by humans or it's written by a computer bot.

2.2 Logistic Regression:

Logistic regression is one of the most powerful technique used in machine learning, it is used combined with supervised learning. Supervised learning is the technique of labelling the data before giving it as an input to machine learning algorithm. The logistic regression returns only two values that is zero or one. It will also provide us with the probability value that is between zero and one, this helps us to find the probability of the data if the

result of the regression is close to one than its can be considered as true and if the result of logistic regression is close to zero then it can be considered as the fake. We use the above method of logistic regression in the manual testing mechanism in our model so that the test data given as input to the logistic regression and returns it returns us with the value either true or false.

	precision	recall	f1-score	support
0	0.99	0.99	0.99	5894
1	0.99	0.99	0.99	5326
accuracy			0.99	11220
macro avg	0.99	0.99	0.99	11220
weighted avg	0.99	0.99	0.99	11220

2.3 Decision Tree Classification:

The first thing with Decision Tree Classification is that they don't need a fixed set of parameters to work and to determine the probability using machine learning for the given set of data. It used non-parametric algorithm to work on a given set of training data. It also helps to solve the classification problem, it used tree-structure classifier which basically has two nodes that is Decision and Leaf node. The algorithm works on the given test dataset and helps us predicting the outcome of the data, the leaf node is used to produce output while taking the data from decision tree as an input.

```
DT.score(xv_test, y_test)
```

```
0.9967023172905526
```

	precision	recall	f1-score	support
0	1.00	1.00	1.00	5871
1	1.00	1.00	1.00	5349
accuracy			1.00	11220
macro avg	1.00	1.00	1.00	11220
weighted avg	1.00	1.00	1.00	11220

2.4 Gradient Boosting Classifier

The mechanism Gradient Boosting Classifier is that it uses multiple weak machine learning algorithms in order to create a robust model. It helps the algorithm to boost by combining multiple weak algorithm and the result produced by this is highly reliable the GBC score received in the current implementation is 0.9959893048128342 such high score helps us to predict the genuineness of the dataset. The Gradient Boosting Classifier has proved the effectiveness of it while dealing with the complex dataset or big dataset.

```
GBC.score(xv_test, y_test)
```

```
0.9959893048128342
```

	precision	recall	f1-score	support
0	1.00	0.99	1.00	5871
1	0.99	1.00	1.00	5349
accuracy			1.00	11220
macro avg	1.00	1.00	1.00	11220
weighted avg	1.00	1.00	1.00	11220

2.5 Random Forest Classifier

It also falls under the group of supervised learning, the advantage with RFC is that it can be used for both classification and regression. The way random tree works in order to predict the news is it will randomly select for a set of data from the big group of data sample and then it will create a decision tree from the sample and starts prediction once it gets predictions from all the trees then it uses the voting mechanism in order to predict if the news real or fake. This works well while you're dealing with a large set of data it also has the ability to operate on multiple inputs and it also provides a highly efficient score of 0.988680926916221

```
RFC.score(xv_test, y_test)
```

```
0.988680926916221
```

	precision	recall	f1-score	support
0	0.99	0.99	0.99	5871
1	0.99	0.99	0.99	5349
accuracy			0.99	11220
macro avg	0.99	0.99	0.99	11220
weighted avg	0.99	0.99	0.99	11220

3. SCOPE

With the development of an effective machine learning algorithm which is reliable and agile it becomes easy for social media networks as well as websites to get the genuineness of the news before posting it to a group of subscribers or sharing it among multiple users. Since the model uses four different machine learning algorithms the output which comes out of each of this model is highly accurate and it also provides accurate results while compared to the models which use single algorithm.

4. CONCLUSION

By integrating the model with a social media networks, it is possible to get the fake new detected and removed even before the news gets into circulation. This helps social media become a safe place for general public and will also allow only real news to circulate in the network. This will help build a sense of trust among the online news which is received by the people.

5. REFERENCES

[1] A. S. S. S. A. N. B. R. A. M. Gaurav Bhatt, "On the Benefit of Combining Neural, Statistical and External Features for Fake News Identification," arXiv, vol. 1712, no. 03935v1, 2017.
[2] B. M. P. C. LUÍS BORGES, "Combining Similarity Features and Deep Representation Learning for Stance Detection in the Context of Checking Fake News," arXiv, vol. 1811, no. 00706v1, 2018.
[3] I. A. G. P. S. S. R. Benjamin Riedel, "A simple but tough-to-beat baseline for the Fake News Challenge stance detection task," arXiv, vol. 1707, no. 03264v2, 2018.
[4] M. M. R. B. J. G. M. Alessandro Moschitti, "Automatic Stance Detection Using End-to-End Memory Networks," arXiv, vol. 1804, no. 07581v1, 2018.
[5] A. B. D. Sourya Dipta Das, "A Heuristic-driven Ensemble Framework for COVID-19 Fake News Detection," arXiv, vol. 2101, no. 03545v1, 2021.
[6] J. D. R. R. Kellin Pelrine, "The Surprising Performance of Simple Baselines for Misinformation Detection," arXiv, vol. 2104, no. 06952v1, 2021.
[7] L. Singh, "Fake News Detection: a comparison between available Deep Learning techniques in vector space," 2018.
[8] I. B. a. Z. Boukhers, "ECOL: Early Detection of COVID Lies Using Content, Prior Knowledge and Source Information," arXiv, vol. 2101, no. 05499v1, 2021.
[9] G. B. M. L. D. V. M. a. L. d. A. Eugenio Tacchini, "Some Like it Hoax: Automated Fake News Detection in Social Networks," arXiv, vol. 1704, no. 07506v1, 2017.
[10] S. S. Y. L. Natali Ruchansky, "CSI: A Hybrid Deep Model for Fake News Detection," arXiv, vol. 1703, no. 06959v4, 2017.