THE FAKE NEWS DETECTION USING MACHINE LEARNING ALGORITHMS

Dr. C K GOMATHY[1], Ms. C.V.S.VASAVI[2], Mr. D.Y.V.RAJESH[3], Ms. A.SRIJA[4]

---***---

ABSTRACT: This project comes up with the applications of Natural Processing Language (NLP) to detect the true and fake news .These techniques are used for detecting the fake news happening from some trusted and untrusted platforms/non-trustable sources these days. These models are built using Various Machine Learning techniques to calculate the frequency and count of news from a set. We will use several news platforms like twitter and face book for analyzing the news to detect them whether true or fake.

Keywords: NLP, Fake News, True News, Logistic regression, Decision Tree.

I. INTRODUCTION

Now a day's fake news is creating different issues from sarcastic articles to fabricated news and government propaganda in some outlets. Even this fake news was destroying countries and one's individual peace and harmony and even some families are being destroyed through this.

The data which was collected might contain missing values that may lead to inconsistency .To gain better values the data must be in the media are growing problems with huge ramifications in our society .Obviously , a purposely misleading story is a "fake news" but lately blathering social media 's discourse is changing its definition. Some of them now use the term to dismiss the facts counter to their preferred view points.

II. LITERATURE SURVEY

There are several algorithms for detecting fake news. For that we analyze through different classifiers in different way. Some of the classifiers used here are Random Forest Classifier, Logic Regression; Decision Tree Classifiers are the classifiers which we used for obtaining accuracy.

Random Forest Classifier:

A random forest classifier is a machine learning technique that is used to solve regression and classification problems .It utilizes ensemble learning, which is a technique that combines many classifiers to provide solutions to complex problems. A random forest algorithm consists of many decision trees.

Logistic Regression:

It is a classification algorithm, used when the value of the target value is categorical in nature .Logistic Regression is most commonly used when the data in question has binary output .So when it belongs to one class or another is either 0 or 1.

Decision Tree Classifier:

A Decision Tree is a graphical representation of all possible solutions to a decision based on certain conditions .On each step or node of a decision tree, used for classification .We try to form a condition on the features to seperate all the labels or classes contained in a dataset to the fullest purity.

All these classifiers are used to obtain the best accuracy in the classification of results.

III. EXISTING SYSTEM

There exists a large body of research on the topic of machine learning methods for deception detection, most of it has been focusing on classifying online reviews and publicly available social media posts, Particularly now a days we can't even trust the news papers. So We used old classifiers with new modifications. The news taken can be listed as a small table below.

Table 1: Representing News Sources

Top Five Unreliable News Sources		Top Five Reliable News Sources	
Before It's News	2066	Reuters	3898
Zero Hedge	149	BBC	830
Raw Story	90	USA Today	824
Washington Examiner	79	Washington Post	820
Infowars	67	CNN	595

IV. PROPOSED SYSTEM

In the proposed model The existing data models have been used for training data with few modifications, Later the training models are being tested as multiple processes like passive aggression model, multinomial naive bayes and tested on holdout assets later the trained data is spitted and then generates count victimizer to test and train the data later the results are picked for models.

Here the diagram representing the training data models. In this way by using these models we can obtain accuracy in results.
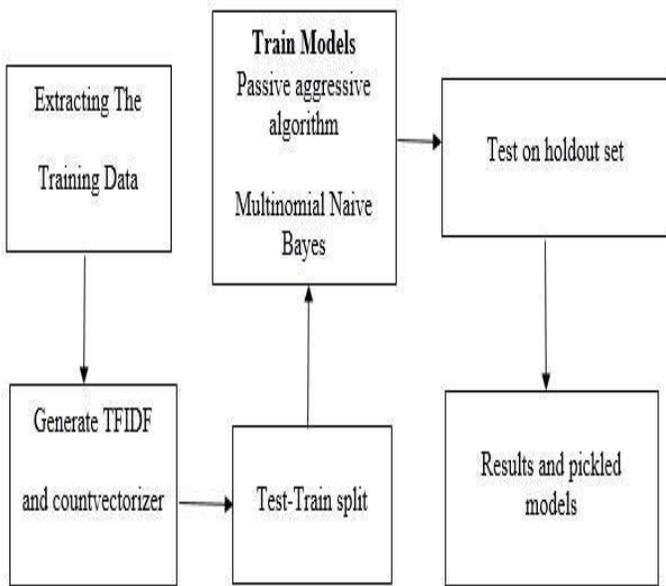

Fig 1: Representing Training Models

IV. METHODOLOGY

The methodology for it was how we are using machine learning algorithms in big data for extraction or classification and evaluation of data through various training data models and some random classifiers which are all listed in literature surveys.

All these classifiers are used in obtaining the exact and nearly values to the problems. It particularly extracts the true and fake news lists in a graphical form and how they are impacting all as shown in fig5 in results. The methodology is very simple and quite interesting .All the classifiers implementation can be done by using Jupiter notebook tool to gain Results in a Way.

Datasets:

The datasets have been collected from various news journals and formed as datasets for two types as true and fake as shown in fig 3&4.All the collected is tested by using all the classifiers to obtain the accurate results from them.

Fig 2: True News Data Set

Fig 3: Fake News Data Set

V. REQUIREMENTS

- Python
- Numpy
- Pandas
- Itertools
- Matplotlib

With all these requirements the project is executed in jupyter notebook tool to obtain the accurate results.

VI. RESULTS

After testing the datasets mentioned above with all the classifiers the accurate and approximate results are obtained as follows.fig 5 represents the graphical representation of obtained results.fig 6,7 &8 presents a confusion matrix form of results without normalization. and also the count of real and fake news from datasets.
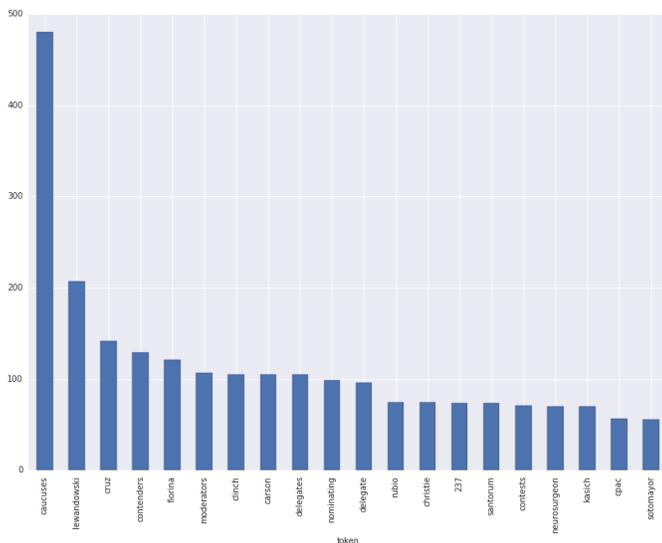
Fig 4: Graph Representing Twitter Fake News 1
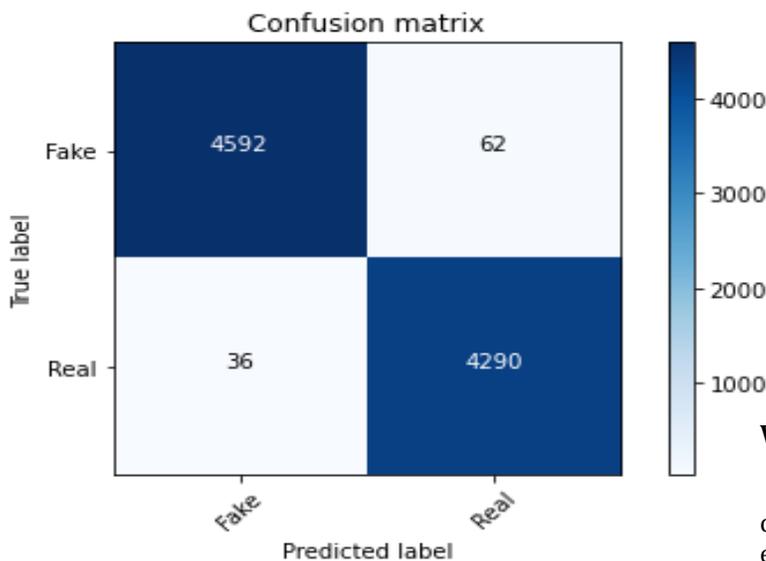
Fig 5: Confusion Matrix, Without Normalization
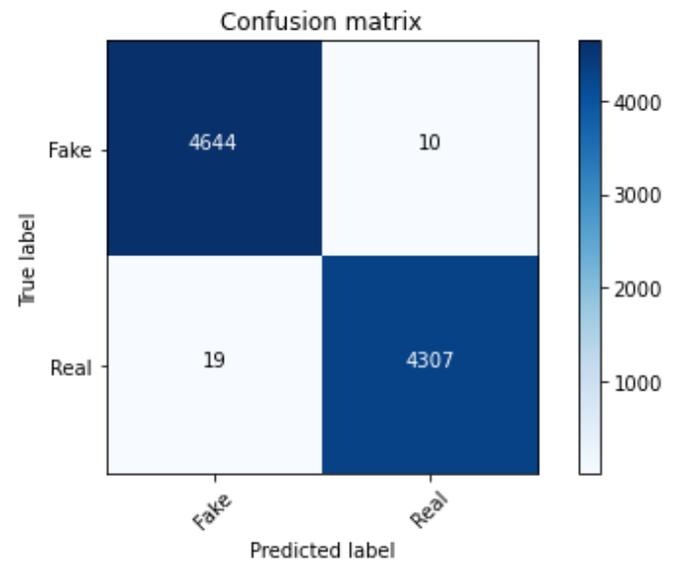
Fig 6: Confusion Matrix, Without Normalization
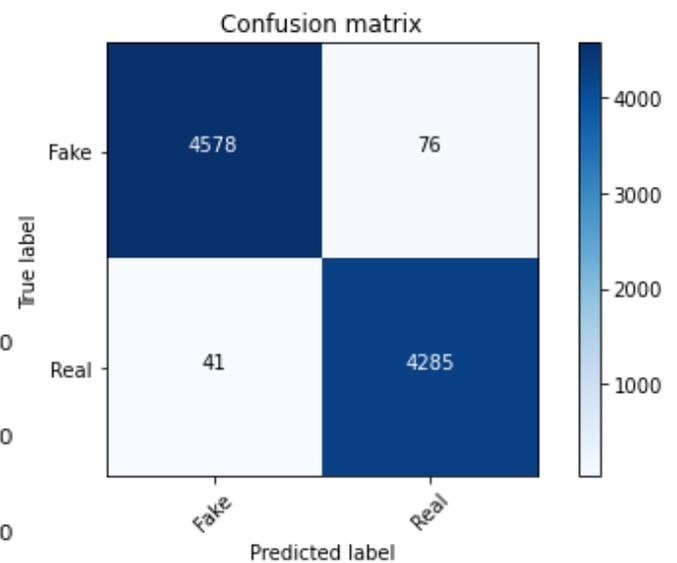
Fig 7: Confusion matrix, without normalization

VII .CONCLUSION

From all these we can conclude that all the results obtained here are accurate of our knowledge and are very easy to understand .As we all know that how people are addicted to social Media now-a-days even for small things they are very dependent on online sources. In that they are forgetting to analyze the thing they have seen all these happens in a large case when it comes to news, So this project is very much helpful in analyzing them to find out the truth.

VIII. REFERENCES

1) N. J. 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.

2) S. Feng, R. Banerjee, and Y. Choi, "Syntactic stylometry for deception detection," in Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers-Volume 2, Association for Computational Linguistics, 2012, pp. 171–175.

3)Shlok Gilda,Department of Computer Engineering, Evaluating Machine Learning Algorithms for Fake News Detection,2017 IEEE 15th Student Conference on Research and Development (SCOReD)

4. Dr.C K Gomathy, Article: A Semantic Quality of Web Service Information Retrieval Techniques Using Bin Rank A Cloud Monitoring Framework Perform in Web Services, International Journal of Scientific Research in Computer Science Engineering and Information Technology IJSRCSEIT | Volume 3 | Issue 5 | ISSN : 2456-3307,May-2018

5.Dr.C K Gomathy, Article: Supply chain-Impact of importance and Technology in Software Release Management, International Journal of Scientific Research in Computer Science Engineering and Information Technology (IJSRCSEIT) Volume 3 | Issue 6 | ISSN : 2456-3307, P.No:1-4, July-2018

6.Dr.C K Gomathy, Article: A Study on the recent Advancements in Online Surveying , International Journal of Emerging technologies and Innovative Research (JETIR) Volume 5 | Issue 11 | ISSN : 2349-5162, P.No:327-331, Nov-2018

Author's Profile:-

1.   **Ms.C.V.S.Vasavi**, Student, B.E. Computer Science and Engineering, Sri Chandrasekharendra SaraswathiViswa Mahavidyalaya deemed to be university, Enathur, Kanchipuram, India. Her Area of Big data analytics

2.   **Mr. D.Y.V.Rajesh**, Student, B.E. Computer Science and Engineering, Sri Chandrasekharendra SaraswathiViswa Mahavidyalaya deemed to be university, Enathur, Kanchipuram, India. His Area of Big data analytics

3.   **Ms.A.Srija** Student, B.E. Computer Science and Engineering, Sri Chandrasekharendra SaraswathiViswa Mahavidyalaya deemed to be university, Enathur, Kanchipuram, India. Her Area of Big data analytics

4.   **Dr.C.K.Gomathy** is Assistant Professor in Computer Science and Engineering at Sri Chandrasekharendra SaraswathiViswa Mahavidyalaya deemed to be university, Enathur, Kanchipuram, India. Her area of interest is Software Engineering, Web Services, Knowledge Management and IOT.