Fake News Detection to Eradicate Yellow Journalism

[1] A.Akshay Deep, [2] T.Akshay Goud, [3] P.Akshaya Reddy, [4] Akshaya sadrollu,

[5] S. Akshaya , [6] L.Akhil Goud ,[7] Dr.D.Thiyagarajan

[12345] Students, [6] Assistant Professor

Artificial Intelligence and Machine Learning.

Department Of Computer Science And Engineering.

Malla Reddy University, Hyderabad, Telangana, India.

ABSTRACT :

In the modern era where the internet is ubiquitous, everyone relies on various online resources for news. Together with the increase in the use of social media platforms like Whatsapp, Facebook, Twitter, etc.. The need for an hour is to stop the spread of articles in developing countries like India.This Project comes up with the three applications of Natural Language Processing techniques for detecting 'fake news, that is, misleading news stories that come from unknown sources. Only by building a model based on a count vectorizer or a TF-IDF matrix, can only get you so far. But these models don't consider the important qualities like word ordering and context. This process will result in feature extraction; we propose using Python sci-kit-learn library to perform extraction of text data because this library contains useful tools like Count Vectorizer, Tf-IDF Vectorizer, and Hashing Vectorizer.Then, we will perform the training model and then classify with the user input then it will obtain the highest accuracy and precision according to confusion matrix results.

Keywords : Fake News, Passive Aggressive Classifier, Machine Learning, Natural Language Processing

INTRODUCTION :

A great deal of fake news is roaring through the various social media platforms. During this case classification of any news, post, story, journal into fake or real one has become a crucial them as fake and true and it's conjointly attracted a good interest from researchers round the world. In line with several analysis studies that are administered to hunt out the impact of any false and fictional news on of us upon returning through such fake news details. Falsified news or news is used in such how that individual begin basic mental process in one issue that may not true. The best example for fake news is that the pandemic situation occurring within the entire world. There are variant of news articles till presently that are falsified and used merely to create confusion and disturbance inside the minds of individual and to misguide their minds to believe that false news. However, can anyone perceive if it's fake or real? False information on Indian social media caused form voters to drink cow wewe or eat dung, thus on stop infection, whereas in Country, artiodactyl weewee with lime was hailed as a protection against the coronavirus. The scientists put together looked into completely different rumors, like uptake garlic, sporting heat socks and spreading goose fat on one's chest, as treatment for the likely fatal virus. Conspiracy

theories were put together monitored, just like the notion that it's a bioweapon funded by enterpriser to further antigen sales. Our project aims to develop a machine learning program to identify when a news source may be producing fake news. We use a corpus of labelled real and fake articles to build a classifier that can make decisions about information based on the content from the corpus. Our model focuses on identifying sources of fake news, based on multiple articles originating from a source. Once a source is labelled as a producer of fake news, we predict that all future articles from the same source are also a producer of fake news. The intended application of our project is to assist in applying visibility weights in social media. Social networks can make use of the weights produced by the model to obscure stories that are highly likely to be fake news

I.   IMPLEMENTATION

ysis of claims and additional evidence, context, and reports from authoritative sources. Generally, news data with annotations can be gathered in the following ways: Expert journalists, Fact-checking websites, Industry detectors, and Crowd sourced workers. However, We can get online news from different sources like social media websites, search engine, homepage of news agency websites or the factchecking websites. On the Internet, there are a few publicly available datasets for Fake news classification like Buzzfeed News, LIAR [15], BS Detector etc. These datasets have been widely used in different research papers for determining the veracity of news. In the following sections, I have discussed in brief about the sources of the dataset used in this work. Online news can be collected from different sources, such as news agency homepages, search engines, and social media websites. However, manually determining the veracity of news is a challenging task, usually requiring annotators with domain expertise who performs careful analthere are no agreed upon benchmark datasets for the fake news detection problem. Data gathered must be preprocessed- that is, cleaned, transformed and integrated before it can undergo training process [16]. The dataset that we used is explained below: LIAR: This dataset is collected from fact-checking website PolitiFact through its API [15]. It includes 12,836 human labelled short statements, which are sampled from various contexts, such as news releases, TV or radio interviews, campaign speeches, etc. The labels for news truthfulness are fine-grained multiple classes: pants-fire, false, barely-true, half-true, mostly true, and true. The data source used for this project is LIAR dataset which contains 3 files with .csv format for test, train and validation. Below is some description about the data files used for this project. 1. LIAR: A Benchmark Dataset for Fake News Detection William Yang Wang, "Liar, Liar Pants on Fire": A New Benchmark Dataset for Fake News Detection, to appear in Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL 2017), short paper, Vancouver, BC, Canada, July 30-August 4, ACL. Below are the columns used to create 3 datasets that have been in used in this project- ● Column1: Statement (News headline or text). www.ijcrt.org © 2020 IJCRT | Volume 8, Issue 6 June 2020 | ISSN: 2320-2882 IJCRT2006189 International Journal of Creative Research Thoughts (IJCRT) www.ijcrt.org 1397 ● Column2: Label (Label class contains: True, False) The dataset used for this project were in csv format named train.csv, test.csv and valid.csv. 2. REAL_OR_FAKE.CSV we used this dataset for passive aggressive classifier. It contains 3 columns viz 1- Text/keyword, 2-Statement, 3-Label (Fake/True)

II.   ARCHITECTURE
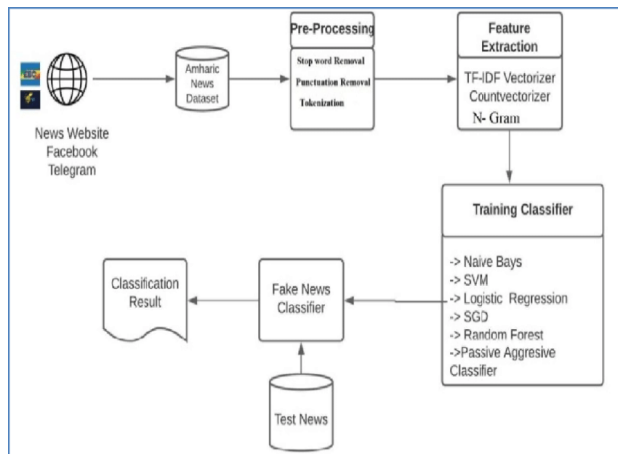
III.   ACTIVITY FLOW DIAGRAM
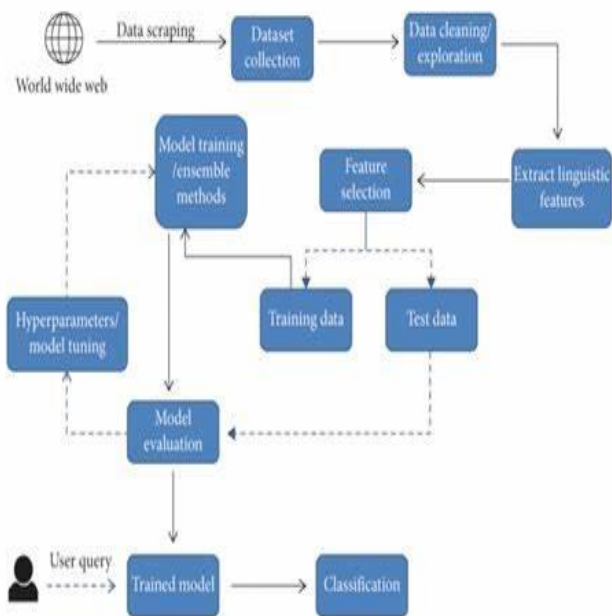
IV. RESULTS

In the fake news detection technology, there have been multiple instances where both unsupervised learning and supervised learning algorithms are used to classify text. Most of the literature survey focus on specific domains, most important the domain of politics. Therefore, the algorithm trained best works on a particular type of article's domain and does not gives optimal results when presented to articles from different areas. Since articles from various areas have a special literary construction, it is hard to train a generic algorithm that works best on all specific news spaces
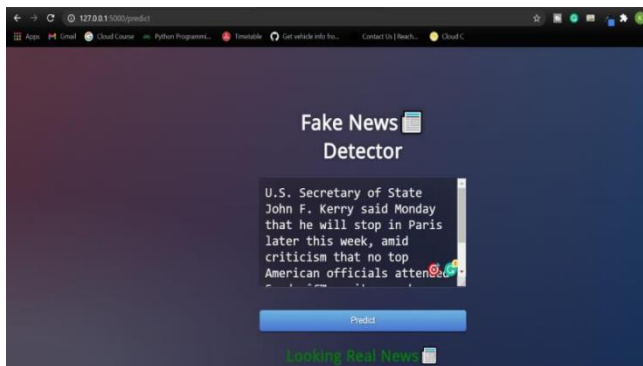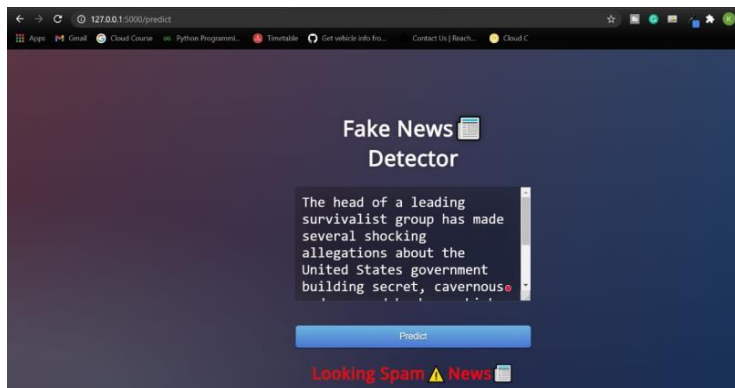
V. CONCLUSION

After analyzing the outcomes of the five models, a mix of machine learning techniques and natural language processing methods is chosen. These five models were created by combining all Machine learning algorithms with a variety of NLP techniques. A user interface is built and linked to a machine learning model that has been trained. A Passive Aggressive Classifier with TF-IDF vectorizer is trained and used to predict user-inputted news. U sers may submit news into this platform, and it will determine if the article is true or not.

VI.   REFERENCES

[1] Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu, "Fake News Detection on Social Media: A Data Mining Perspective"

arXiv:1708.01967v3 [cs.SI], 3 Sep 2017

[2] Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu, "Fake News Detection on Social Media: A Data Mining Perspective"

arXiv:1708.01967v3 [cs.SI], 3 Sep 2017

[3] M. Granik and V. Mesyura, "Fake news detection using naive Bayes classifier," 2017 IEEE First Ukraine Conference on Electrical

and Computer Engineering (UKRCON), Kiev, 2017, pp. 900-903.

[4] Fake news websites. (n.d.) Wikipedia. [Online]. Available: https://en.wikipedia.org/wiki/Fake_news_website. Accessed Feb. 6, 2017

[5] Cade Metz. (2016, Dec. 16). The bittersweet sweepstakes to build an AI that destroys fake news.