AN OVERTURE   APPROACH TO DETECT THE FAKE NEWS USING DEEP LEARNING

Subramanian R[1], Ho Che Minh P[2]

**[1]*Computer Science and Engineering, RVS College Of Engineering, sweetestmani@gmail.com*
**[2]*Computer Science and Engineering, RVS College Of Engineering, hochiminh.p@gmail.com*

Abstract— This Paper Proposes an overture   approach to detect the fake news using deep learning. The first part of the work uses the conventional machine learning model to detect fake news. The second part of the work concentrates on designing a recurrent neural network model to detect fake news. In addition to this a model that is suitable for real-time processing would also be considered for comparison.

Keywords— Machine Learning, Deep Learning, Fake News Dedection, TF-IDF, Real-Time learning.

I.   INTRODUCTION

"Fake News" could be a term accustomed represent invented news or info comprising info communicated through ancient media channels like print, and tv in addition as non-traditional media channels like social media. the final motive to unfold such news is to mislead the readers, harm name of any entity, or to achieve from sensationalism. it's seen mutually of the best threats to Society. pretend news is more and more being shared via social media platforms like Twitter, WhatsApp and Facebook.

These platforms supply a setting for the overall population to share their opinions and views in a very raw and un-edited fashion. Some news articles hosted or shared on the social media platforms have additional views compared to direct views from the media outlets' platform. analysis that studied the rate of pretend news finished that tweets containing false data reach individuals on Twitter sixfold quicker than truthful tweets .

Technologies like AI (AI) and linguistic communication process (NLP) tools supply nice promise for researchers to make systems that might mechanically find faux news. However, sleuthing faux news could be a difficult task to accomplish because it needs models to summarize the news and compare it to the particular news so as to classify it as faux.

Moreover, the task of comparison planned news with the first news itself could be a intimidating task as its extremely subjective and self-opinionated pretend news will be are available in several forms, including: unintentional errors committed by news aggregators, outright false stories, or the stories that area unit developed to mislead and influence reader's opinion. whereas pretend news could have multiple forms, the result that it will wear individuals, government and organizations could usually be negative since it differs from the facts.

II.   FAKE NEWS DETECTION

Fake News is the process of using natural language processing, text analysis, and statistics to analyze the news in the social Medias such as Facebook, twitter etc. For finding the truthfulness of the news. As there is a paradigm shift from the Web 2.0 to Web 3.0 results in the increase of the content that are published in the social media . With the increase in the number of social media users, the speed at which particular news spread tremendously increases. So it becomes mandatory that the truthfulness of the news should be verified. The classification is the process of assigning a label to the input

Fig 1. Classification Problem

2.1 Problem Statement

Given a set of news, the model should be capable of classifying the news in to either true news or false news. It can be considered as a classification problem.

2.2 Models to be Employed

The work includes the designing of the following models

- ➢ Machine Learning Model
- ➢ Deep Learning Model
- ➢ Realistic Model

All these models required pre-processing. The pre-processing required to be done for sentimental analysis includes functionalities from natural language processing.

2.3 Natural Language Processing

Natural Language processing can be considered as a sub domain of computer science and artificial intelligence. It applies machine learning models to text and natural languages. It provides a platform that enables interaction between human and computers. Though the research related to natural language processing started from 1950 with the work of Alan turing, with the introduction of machine learning algorithms in late 1980s there is a drastic changes in the applications of natural language processing. Some of the applications of the natural language processing includes sentiment analysis, analyzing the opinions with huge set of texts,question answering which nowadays are being used as chat bots, translations, speech recognition and document summarization. The core processes involved in natural language processing are Text cleaning, Creating Bag of words and Applying machine learning model.
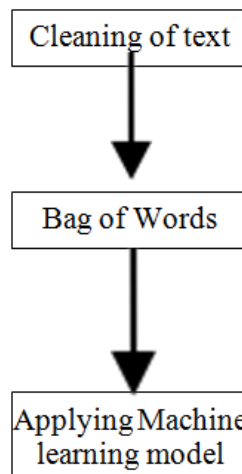

Fig 2. Natural Language Processing flow

III.   DETECTION METHOD

The proposed system includes the development of three models and comparing the results. The model includes machine learning models, Deep Learning Model and realistic model that could best suit to the increasing number of news continuously. This phase of the work concentrates on the implementation of the machine learning models and comparing the results.
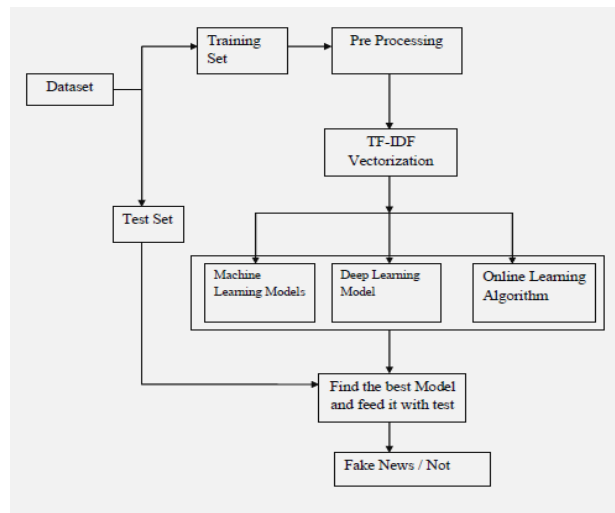
Fig 3.System Architechure

3.1Modules

This phase concentrates only on the following modules.

➢ Pre-processing

➢ TF-IDF Vectorizer

➢ Implementation of Machine learning Model

3.1.1. Pre-Processing

It includes the process of splitting the dataset in to training set and test set. The split is made in the ratio of 80:20. 80 percent of the data is considered as the training set and remaining 20 percentages of the data is considered as the test set.

3.1.2. TF-IDF Vectorizer

It is the process which can be described as the combination of count vectorization and TF-IDF Transformation. Count vectorization can be explained as follows.

With the help of the Count Vectorizer, a bag of words is created. bag of words is created first by selecting the unique words from the set of words available and a table is formed as shown below.

There are n news

Set of reviews=$\{r_0, r_2, . \ r_n\}$

Set of unique words in all columns=$\{uw_1, uw_2, . \ uw_n\}$

A table is formed, where the rows denote the set of news and columns denote the set of unique words. The table is depicted below.

3.1.3 Implementation of the Machine Learning Modules

The following machine learning models are employed.

➢ Logistic regression

➢ Support Vector Classifier

➢ Random Forest

➢ Decision Tree

The TF-IDf vectors that are obtained are given as input to the model for training and testing

IV. RESULTS

The results obtained with respect to the test set are given below. The performance parameters that are measured and compared. Area under the curve is the region that is below the receiver operating characteristic curve which is drawn with the true positive rate and the false positive rate. Recall is known as the true positive rate and False positive rate can be defined as follows. The Following Figure represents the comparison of the test accuracy of the different machine learning models.
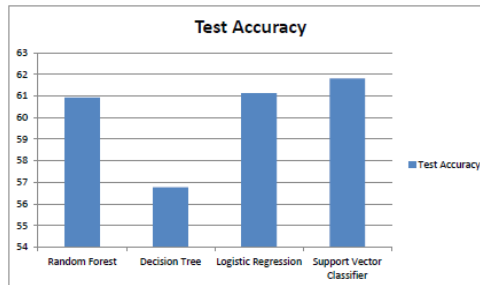

Fig 4. Comparison of Test Accuracy of different models

The Following Figure represents the comparison of the test recall of the different machine learning models.
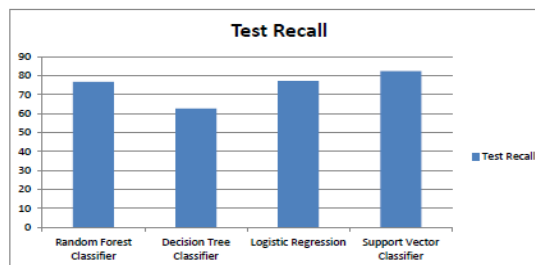

Fig 5. Comparison of Test Recall of different models

The Following represents the comparison of the test AUC of the different machine learning models.
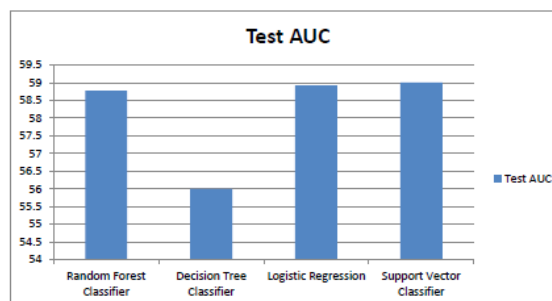

Fig 6. Comparison of Test AUC of different Machine Learning models

REFERENCES

[1] Deepak S, Bhadrachalam Chitturi, "Deep neural approach to Fake-News identification" , Procedia Computer Science, Volume 167, 2020, Pages 2236-2243

[2] Jozef Kapusta, Petr Hájek, Michal Munk, Ľubomír Benko, "Comparison of fake and real news based on morphological analysis", Procedia Computer Science, Volume 171, 2020, Pages 2285-2293.

[3] Yin-Fu Huang, Po-Hong Chen, "Fake news detection using an ensemble learning model based on Self-Adaptive Harmony Search algorithms" , Expert Systems with Applications, Volume 159, 2020.

[4] Feyza Altunbey Ozbay, Bilal Alatas, "Fake news detection within online social media using supervised artificial intelligence algorithms", Physica A: Statistical Mechanics and its Applications, Volume 540, 2020.

[5] V. Tiwari, R. G. Lennon and T. Dowling, "Not Everything You Read Is True! Fake News Detection using Machine

learning Algorithms," 2020 31st Irish Signals and Systems Conference (ISSC), Letterkenny, Ireland, 2020, pp. 1-4, doi: 10.1109/ISSC49989.2020.9180206.

[6] N. R. de Oliveira, D. S. V. Medeiros and D. M. F. Mattos, "A Sensitive Stylistic Approach to Identify Fake News on Social Networking," in IEEE Signal Processing Letters, vol. 27, pp. 1250-1254, 2020, doi: 10.1109/LSP.2020.3008087.

[7] Xinyi zhou et.al (2020) , "SAFE: Similarity-Aware Multi-Modal Fake News Detection" , arXiv:2003.04981v1

[8] Xinyi Zhou, Atishay Jain, Vir V. Phoha, and Reza Zafarani. 2020. Fake News Early Detection: A Theory-driven Model. Digit. Threat.: Res. Pract. 1, 2, Article 12 (June 2020), 25 pages.