Fib News Detection Using Machine Learning Algorithm

Pooja C.P, Mallamma .V. Reddy, Naveen Kumar .T. R

Department of Computer Science, Rani Channamma University, Belagavi, Karnataka 590001, India

Abstract: 21st generation is an Era of internet. Most of the people in this era depend on internet for their day to day activities and hence usage of Internet has increased. People now a days rely on the online services for the news instead of traditional news organization. Along with this, they use the social media platform like Facebook, Twitter, Websites etc. To know about what is happening around them. By the use of the social media news may spread rapidly among millions of people within a short span. Spammers who spread fake news among the people, affect the mindset of the people reading the news and make use of that. Aim of this paper is to show which news is true and which is false by implementing the system using machine learning algorithm for Classification. The proposed system provides the users to classify the news which is true and which is false.
Keywords: Classification algorithms, Internet, Machine Learning, Social media Websites.

1. Introduction:

Due to the advanced technology, facts and information is easily accessible to everyone. Internet provides a large amount of Information but the credibility of Information depends upon many factors. A big amount of Information is published daily online by social media like Facebook, Twitter and other sources. News is a report of current events as shown in figure 1 and 2

Fig 1: Online news

Fig 2: Social media websites

It is not easy to say that the published Information is true or false. It requires a deep study and analysis of stories which include checking of the facts by accessing the sources that spread the Information or by finding the original sources of the Information by checking credibility of authors.

The advantages of spreading fake news are just an attempt to damage an organization, an individual reputation, or it may be simply a moto to gain financially or politically. "FAKE NEWS" is used for the kind of information which misleads the people and their mindsets.

In the recent years, Many research has been conducted in the concept of detecting fake news with satisfactory results, with the help of AI, Neural networks, and the Machine Learning concepts. Technology has a relief to Humans from this dangerous efforts. Fake news detection with the help of Technology can save the society from the unnecessary chaos and social unrest.

The Main aim of this project is to build a system that helps the people to predict whether the news is tautology or contradiction.

The characteristics of the fake news are:

• Grammatical mistakes.
• Emotionally colored.
• Try to affect readers opinion.
• Content is not always true
• Their sources are not genuine most of the times.

2. LITERATURE REVIEW:

Uma Sharma [1] has aimed to perform binary classification on various news articles that are available online by using concepts like Artificial Intelligence, Machine Learning, and Natural Language Processing. Mykhalio Granik [2] has used a simple approach for fake news detection by using Bayes classifier. This is implemented as software system and tested against the set of data in Facebook news post, which has achieved the classification accuracy approximately about 74% Himmak Gupta [3] used a different framework based on machine learning that deals with various problems including accuracy shortage. Time lag and high processing time to handle thousands of tweets in 1 sec

.They are used to achieve the accuracy of 91.65% and surpassed the existing system by 18 percent.
Laxmisingh [4] created a tool for detecting the patterns of language that are characterized as fake and real through the use of machine learning.

Papadopoulou [5] has used two levels text-based classifier to notice click baits that need a big form of morphology, style, grammar, word-based options and sentiment analysis. Rubin [6] has used satriced cues to give a reason in between fake and real news. Their approaches trust the abrusity of the text, and grammatical options, achived a precision and recall of the 90 and 87 respectively.

Marco L. Della Vedova et. Al[11] Frist proposed a novel ML fake news news content and social media context features ,secondly they implement this method with in Facebook measures chots and validate it with real world application.

3. METHODOLOGY:

This paper explains the system which is developed in 3 phases.

Phase 1 static search which works on the classifiers. Phase 2 dynamic search which takes the keywords from the dataset and matches with the articles to predict whether the news is true or false and phase 3 provides an authentic URL as input to the user. In this paper tools used are python and joblib library. Python has a huge set of library and extensions which can be easily used in machine learning to produce client side input GUI in Html, CSS, and Javascript. We have used Beautiful Soap(bs4) for requesting online scrapping.
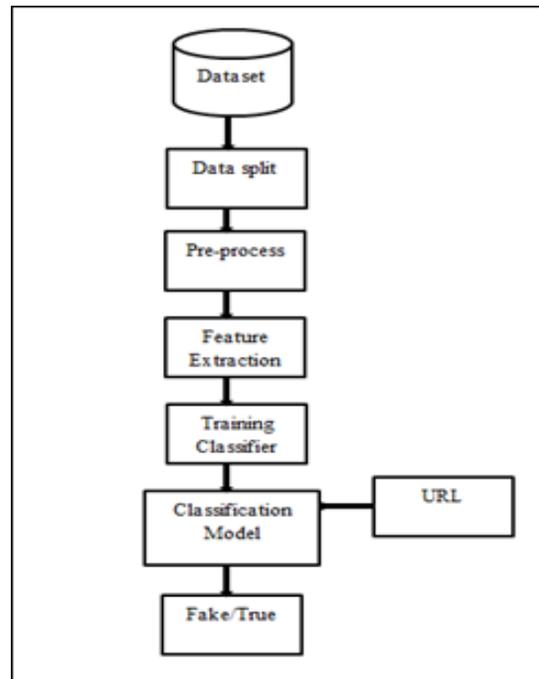
SYSTEM ARCHITECTURE:

Fig 3: Fib Detection

4. IMPLIMENTATION:
Data Collection and Data Analysis:
People get the information from the different online resources like social media, search engine, news agencies website, and others. There are only few publicly available datasets for fake news classification liker Buzzed news, LIAR, BS detector etc. These types of datasets are used in different research papers for determining the accuracy of news. In the following section of this paper we explained the source of dataset that is used in this work.
In this system we have used the Kaggle keyword extraction using HDF5 files as a dataset.

HDF5 is a hierarchical data format. A good abstraction to think about is a Python dictionary where a set of keys are mapped to a set of values. Just like a dictionary you can get the dataset associated with a key in an HDF5 using a file descriptor. This file descriptor can be generated with pandas HDFStore, and since HDF5 can be infamously prone to corruption issues when misused we recommend using it in simple operations inside context managers to ensure that the file descriptor is closed when not used.
Data splitting: A dataset used for machine learning should be partitioned into three subsets training, test and validation sets.
Pre-processing of data:
The data that is in the social media is unstructured. They are in the informal way like slangs bad-grammar to achive best insights, it needs to clean data which is used for modeling. For this purpose the basic thing that is the pre-processing of data must be done.

International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)
Web Site: www.ijettcs.org Email: editor@ijettcs.org, editorijettcs@gmail.com
Volume 11, Issue 4, July – August 2022   **ISSN 2278-6856**

Steps are:

Data cleaning: When we are reading data or Information data might be in the structured or unstructured format. This step needs to semi-structure data for better understanding. The cleaning up of data contains some steps:

a.Remove punctuation: Removing the punctuation which provides grammatical context to sentences.

b. Tokenization: This step separates the text into units like sentences and words.

c. Removing stop-words: The stop-words are common words that appear in any text. In these steps we are going to remove those words.

d. Stemming: It helps to reduce a word to its stem form. It removes the suffices like "ing",”ly”,”s”.

Classification Model(Random Forest):
Random Forest is one of the most popular machine learning algorithms. It does not require any data correction and modelling but it often leads to more accurate results. Random Forests are based on decision trees. Specifically, Random Forests are collections of deciduous trees, producing better predictable accuracy that is why it is called the 'forest'

- basically a group of trees to reduce to some minimum values. The basic idea is to grow multiple decision trees based on independent datasets. For each node, n-n variables without a feature set are randomly selected, and the best variation of these variables is available. Random Forests inherit many advantages of decision tree algorithms. They apply to both retrospective problems and categories; they are easy to count and quick to install. They often lead to better accuracy. However, unlike decision trees, it is not very easy to interpret the results. In decision-making trees, by examining the resulting tree, we can find important information about what changes are important and how they affect the outcome. This is not possible with random forests. It can also be described as a morestable algorithm than decision trees, if we change the data a bit the decision trees will change, almost reducing accuracy. This will not happen in random forest algorithms as it is a combination of many cutting trees, the random forest will remain stable.[10]

Decision tree:
As suggested from the name decision trees are data structures with a tree structure. The training database is used for tree construction, which is used to make predictions on experimental data. In this algorithm, the goal is to get the most accurate result with the least number of decisions to be made. Decision trees can be used for both partition and retreat problems. It is

based on the ideas of Entropy and Benefit of Knowledge. Entropy here refers to the level of uncertainty in data content. example, the entropy of toss money will be endless, because there is no way to guarantee the outcome. Conversely, throwing a two- headed coin on both sides that

can provide zero entropy as we can predict the outcome 100% before each throw.

5. IMPLIMENTATION STEPS AND RESULTS:

Step 1: Take input as URL of the website or news from the user to predict.

Step 2: Pre-process of Information is done and it extract the features.

Step 3: There is a classifier called Random forest to predict the news as fake and true. Each of the extracted features are fed to the classifier.

Step 4: The classifier predicts the news as fake or true and gives the results as true news or fake news.

Implementation is done using the Machine learning Algorithms like Random Forest, Which is used to predict the news as Fake and truth. We found the results by system after the implementation.
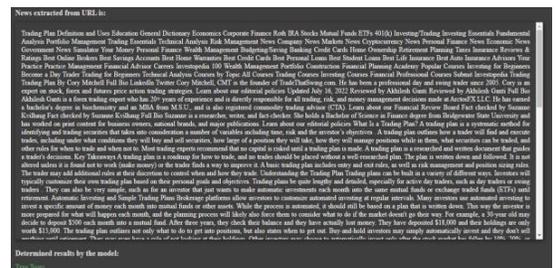
Snapshots of working system:
1. Input from the User as URL:

2. After Entering the URL: a.TrueNews:
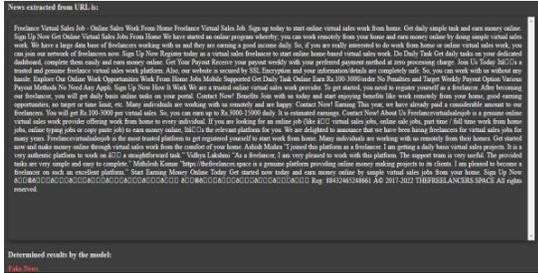

Before Detect

After detect

b.Fake News:

Before Detect

After Detect

6. CONCLUSION AND FUTURE ENANCEMENT:

The majority of works are carried out in online, in early days the newspaper are used to grab the news that are happening around us, but now a days this is substituted by the social media like Facebook, Twitter, Online news agencies etc. The news that are published online are unpredictable. It may be true or false. The main objective is to build this system to help people differentiate between news between true and fake news. This may help to save the people and society from unnecessary chaos and social unrest.

As we have implemented the system to detect fake news in website, social media and other sources in future it may implement real time fake news identification in videos.

REFERENCES:

1. UmaSharmaet.al.https://www.ijert.org/fake-news-detection-using-machine-learning-algorithms.
2. 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.
3. H. Gupta, M. S. Jamal, S. Madisetty and M. S. Desarkar, "A framework for real-time spam detection in Twitter," 2018 10th International Conference on Communication Systems & Networks (COMSNETS), Bengaluru, 2018, pp. 380-383.
4. https://www.semanticscholar.org/paper/Fake- News-Detection-using-Machine-Learning-Singh/07f76273e0fd8b962142f9126b1a2654b8df2 1c
5. C.Budak, "What happened? The Spread of Fake News PublisherContent During the 2016 U.S. Presidential Election," 2019, pp. 139–150.
.6. A. Mitchell, J. Gottfried, and K. E. Matsa, "Facebook Top Source for Political News Among Millennials —

Pew Research Center,"2015.[Online]. Available:http://www.journalism.org/2015/06/01/facebook-top-source-for-political-news-among-millennials
7 .https://en.wikipedia.org/wiki/Fakenews
8. https://en.wikipedia.org/wiki/Machine_learning
9.https://www.kaggle.com/code/diegovicente/a-short-introduction-to-hdf5-files/notebook
10.https//medium.com/anirudhpalapathi/understanding-random-forest-algorithm-79b87bdde79a.
11.M. L. Della Vedova, E. Tacchini, S. Moret, G. Ballarin, M. DiPierro and L. de Alfaro, "Automatic Online Fake News Detection Combining Content and Social Signals," 2018 22nd Conference of Open Innovations Association (FRUCT), Jyvaskyla, 2018, pp. 272- 279.