

# Fake News Detection

Preethi Shankar Raman  
Aditya Sinha  
Ira Fulton School of Engineering  
Arizona State University  
Tempe, Arizona, USA  
pshanka8@asu.edu  
asinha80@asu.edu

## 1 Abstract

Large amounts of false information are created online for a variety of reasons, including financial and political benefit, due to how simple it is to propagate online. False information is widely circulated and can have a detrimental effect on people and society: People stop trusting the online forums and website. Manipulating the consumers to belief in false or wrong bias. Creating new behavior among the people to interpret and respond to news information.

Therefore, it is become extremely important to detect fake news and misinformation on social media and news website. The goal of this task is to classify the labels of given fake news dataset correctly with given metric (Accuracy). Using this metric, we are expected to get as high score as possible to minimize the error of classifier. We formally define the task as follows. Given a claim of misinformation related to COVID-19, participants are asked to classify the claim into one of the four categories (number indicates the label):

- FALSE(0): claim that is deceptive or untruthful.
- MISLEADING(1): claim that gives people an incorrect understanding of the situation.
- TRUE(2): claim that is verified to be correct.
- UNPROVEN(3): claim that is not able to be verified

## 2 Introduction

We have explored the possibility of using social media web mining to detect and predict the fake news being published around the world. The data set has a lot of unnecessary information, which data can remove without affecting the accuracy of the classifier. We have used many classifiers to classify the news into one category. Based on the analysis scores we classify the data into above mentioned 4 categories

---

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

© 2022 Association for Computing Machinery.  
ACM ISBN 978-x-xxxx-xxxx-x/YY/MM...\$15.00

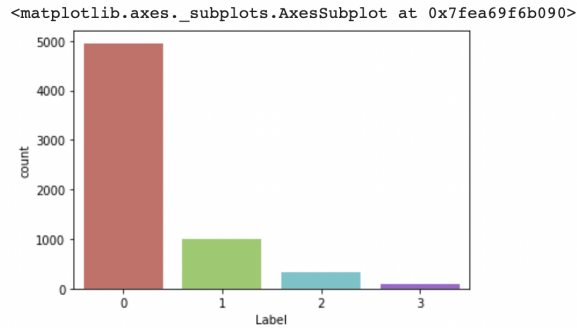
- False, Misleading, True, Unproven. The aim is to improve research efforts for companies that provide detection of fake news on social media websites.

## 3 Related Works

Websites and software devoted to forums, social networks, microblogging, social bookmarking, and wikis are considered to be part of social media. On the other hand, some academics believe that fake news is a result of unintentional events, such as educational shock or behaviors. In 2020, health-related fake news was pervasive and put the world's health at danger. Early in February 2020, the WHO issued a warning stating that the COVID-19 epidemic had sparked a significant "infodemic," or a burst of true and false news, which included a lot of disinformation. There are multiple groups and organizations which are involved in promoting false news on online social media sites. To prevent the further spread of fake news, we need a strong mechanism to counter it and stop it from spreading among the people. Machine Learning and Artificial Intelligence are one of the best tools which have been developed in recent times to counter fake news. Machine Learning model can be trained on data set, and they can be further used to detect fake news in a completely new dataset.

## 4 Model Description

Since we are dealing with text data, we will be performing text mining to analyze and draw meaningful patterns from the tweets. The given training dataset contains 6384 samples of news data which has been used to train our classification models. Pre-processing would involve removing stopwords and feature extraction in addition to removing the duplicates. In order to speed up learning algorithms while maintaining model performance, our model seeks to identify a strategy for extracting important terms and eliminating less relevant words based on TF-IDF information. 80 percent of the data would be used for training while 20 percent of the tweets would be used for testing the model. Our data preprocessing consists of the following:



**Figure 1.** In the graph we can see the no of false(0),misleading(1),true(2), unproven(3)

#### 4.1 Stopwords

Stopwords are the words that carry little or no significance such as the, is, are. We'll be removing stopwords by importing the stopwords list from nltk.corpus.

#### 4.2 Tokenization

Tokenization is used to split the tweets into respective words. The accuracy of tokenization is based on the vocabulary it is trained with. By examining the word order in our text data, tokenization aids in comprehending the text's meaning. It is performed by importing nltk.word\_tokenize.

#### 4.3 Feature Extraction

When dealing with textual data it is important to convert the texts into numericals. In order to speed up learning algorithms while maintaining model performance, our model seeks to identify a strategy for extracting important terms and eliminating less relevant words based on TF-IDF(Term Frequency - Inverses Document Frequency) information. It is equivalent to the number of times a word appears in a document by the total number of words in the document. Our model classifies the news into 4 categories: False(0), Misleading(1), True(2), Unproven(3).

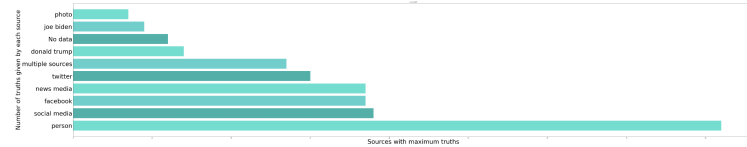
#### 4.4 Web Scrapping

A Python package called BeautifulSoup can extract data from XML and HTML files. It provides natural means of traversing, searching, and altering the parse tree in conjunction with our preferred parser. We are scrapping the title from the URLs. The data scraped from the URLs is stored in a new column called 'Headline'.

#### 4.5 Data Visualization

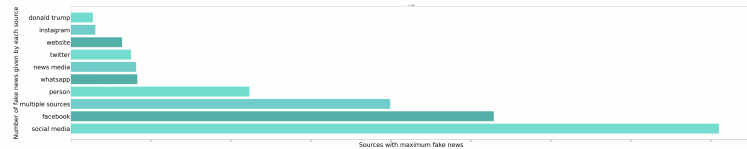
Data Visualization provides a pictorial representation of the trends, outliers and patterns in data. Matplotlib and Seaborn are the python libraries that have been used to plot our visualizations.

Plotting the horizontal bar plot using Matplotlib helps us understand that maximum number of truths is given by Person followed by Social Media. (fig 2)



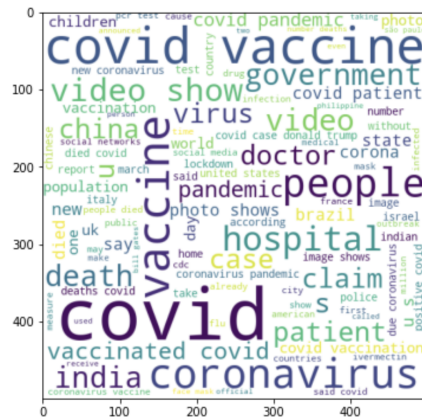
**Figure 2.** The figure represents the truths given by each source

Plot of maximum fake news helps us understand that maximum fake information is spread by Social Media.(fig 3)



**Figure 3.** The figure represents the truths given by each source

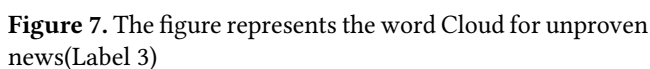
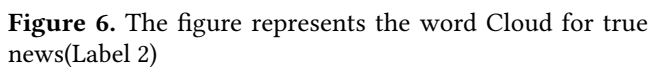
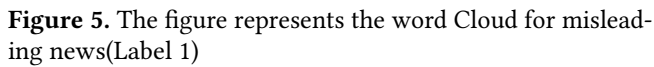
A wordcloud is essentially a visualization technique that shows the frequency of terms in a text by varying the size of each word. WordClouds have been plotted for different labels to understand the most frequently occurring words in True, fake, misleading and unproven news.



**Figure 4.** The figure represents the word Cloud for fake news(Label 0)

#### 4.6 Data Mining models

The difficulty of detecting fake news on social media makes it clear that a workable technique must include a number of components in order to effectively address the problem. We



**Logistic Regression** It is based on sigmoid function. The following is the definition of the sigmoid function with a weight parameter  $w$  and an input  $x(i)$ :- The equation  $h(x(i), w) = 1/(1 + e^{-(T \cdot x(i))})$  Because the sigmoid algorithm returns values between -1 and 1, we can categorize the predictions based on a specific cutoff.

Support Vector Machine works well with both classification and regression data. It classifies the data points into different classes using Hyperplane. Hyperplane equation:  $w \cdot x + c = 0$  where  $c$  is an offset and  $w$  is a vector's normal to the hyperplane.

The other model that we have used is the Decision Tree. It is a non parametric supervised learning algorithm. The objective is to learn straightforward decision rules derived from the data features in order to build a model that predicts the value of a target variable.

The performance of our model is measured by evaluating accuracy, precision, recall and F1 score on the 20

We have also plotted the Confusion matrix for our classifiers. The Confusion matrix helps us understand the True Positives (TP), True Negatives (TN), False Positives (FP) and False Negatives (FN) predicted by our model.

The given training dataset is split using `traintestsplit` function of `sklearn` with 80 percent of the data set aside for classification and 20 percent of the data set aside for testing our model.

As the first step we have pre-processed our data by removing stopwords from the list of stopwords defined by Corpus and have tokenized our data. We have then applied TF-IDF to our data to quantify the relevance of a word in our data amongst a collection of other words.

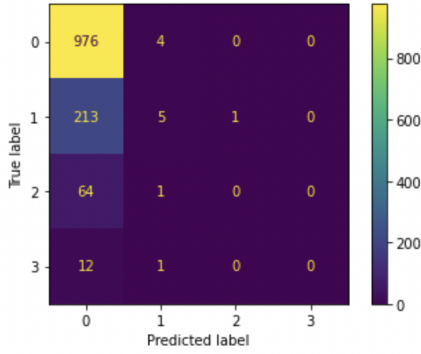
We have then applied 5 supervised learning models to train our data and measure its accuracy against the test data set aside from the training dataset.

We have trained the following models and its evaluation metrics is as follows:

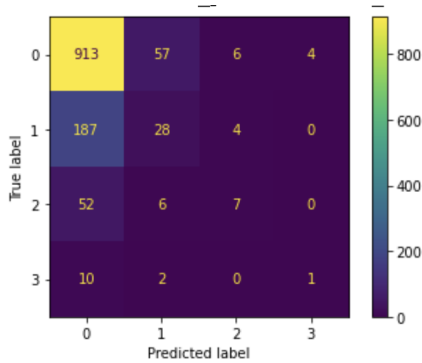
Model	Accuracy	Precision	Recall	F1Score
Logistic	0.768	0.768	0.771	0.768
KNN	0.743	0.743	0.743	0.743
SVM	0.771	0.771	0.771	0.771
ADABOOST	0.765	0.765	0.771	0.765
DecisionTree	0.767	0.767	0.771	0.767

The confusion matrix has been plotted for the various data mining models.

The Confusion matrix for SVM works very well in predicting the performance our dataset. It works very well in predicting false news articles.



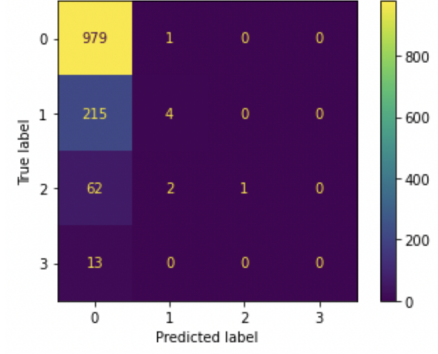
**Figure 8.** Confusion matrix of Logistic Regression



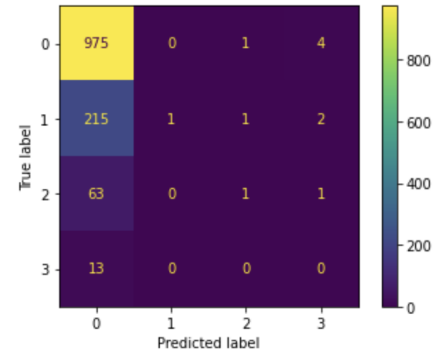
**Figure 9.** Confusion matrix of KNN model

From the results we understand that Support Vector Machine gives the best accuracy of about 0.771. This is because SVM uses the most optimal hyperplanes. Hence we use SVM to predict the labels for our test data. It also gives high precision(0.771), recall(0.771) and F1 score(0.771).

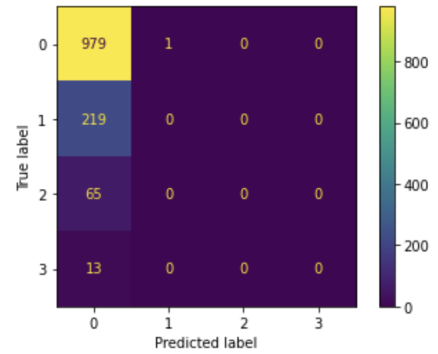
By examining the data using various visualizations we find that the data is wildly skewed, with the bulk of news falling into the false category and endorsing the theory of fake news propagation.



**Figure 10.** Confusion matrix of SVM model



**Figure 11.** Confusion matrix of ADABOOST model



**Figure 12.** Confusion matrix of DecisionTree model

## 6 Future Works

Currently, we trained our machine learning model on data set that has data related to social media websites. In the future, it can be further extended to take data from other sources as well. We have gained enough knowledge on this topic that we are confident that we want to implement it in real-world scenarios. Currently, the input file is in text format, and we can further generalize the model to take the image, video, or binary as input. We have trained the model on a particular set of machine learning algorithms, and the

model can further test it on many different machine learning algorithms. We can then compare the accuracy of all the algorithms and pick the best which suits our case.

## References

1. [Data Mining Techniques in Social Media a survey.](#)
2. [Social Media Mining: An Introduction.](#)
3. [A Study on Social Network Analysis through Data Mining Techniques](#)
4. [Social Media in Data Mining Review Paper](#)
5. [Social media mining under the COVID-19 context](#)
6. [Fake news detection using machine learning approaches](#)
7. [Numpy Framework](#)
8. [Pandas Framework](#)
9. [Seaborn Framework](#)
10. [Matplotlib Framework](#)
11. [BeautifulSoup Framework](#)
12. [Wordcloud Framework](#)
13. [Sklearn Framework](#)