

FAKE NEWS DETECTION

Mini Project Report

Submitted by
SARAH C JOY

*Submitted in partial fulfillment of the requirements for the award of
the degree of*

*Master of Computer Applications Of
A P J Abdul Kalam Technological University*



**FEDERAL INSTITUTE OF SCIENCE AND TECHNOLOGY (FISAT)®
ANGAMALY-683577, ERNAKULAM(DIST)
MARCH 2022**

DECLARATION

I, **SARAH C JOJY**, hereby declare that the report of this project work, submitted to the Department of Computer Applications, Federal Institute of Science and Technology (**FISAT**), Angamaly in partial fulfillment of the award of the degree of Master of Computer Application is an authentic record of our original work.

The report has not been submitted for the award of any degree of this university or any other university.

Date : 04-03-2022

Place: Angamaly

**FEDERAL INSTITUTE OF SCIENCE AND
TECHNOLOGY (FISAT)®
ANGAMALY, ERNAKULAM-683577**

DEPARTMENT OF COMPUTER APPLICATIONS



CERTIFICATE

This is to certify that the project report titled "**FAKE NEWS DETECTION**" submitted by **SARAH C JOJY** towards partial fulfillment of the requirements for the award of the degree of Master of Computer Applications is a record of bonafide work carried out by them during the year 2022.

Project Guide

Head of the Department

Submitted for the viva-voice held on at

Examiner1 :

Examiner2 :

ACKNOWLEDGEMENT

Gratitude is a feeling which is more eloquent than words, more silent than silence. To complete this project work we needed the direction, assistance and co-operation of various individuals, which is received in abundance with the grace of God.

We hereby express our deep sense of gratitude to **Dr. Manoge George**, Principal of FISAT and **Dr. C Sheela**, Vice principal of FISAT, for allowing us to utilize all the facilities of the college. Our sincere thanks to **Dr. Deepa Mary Mathew**, Head of the department of MCA, FISAT , who had been a source of inspiration . During the period of our project work, we have received generous help from **Dr. Shahna K U** , our project guide , which we like to put on record here with deep gratitude and great pleasure.

We express our heartfelt thanks to all the faculty members in our department for their constant encouragement and never ending support throughout the project. We also express our boundless gratitude to all the lab faculty members for their guidance.

Finally we wish to express a whole heart-ed thanks to our parents, friends and well-wishers who extended their help in one way or other in preparation of our project. Besides all, We thank GOD for everything.

ABSTRACT

Fake news is any information fabricated to mislead readers to spread an idea for certain gains (usually political or economical). Now-a-days, sharing and accessing information is very fast and almost free. Online platforms and users are growing significantly day by day. Therefore, online platforms are perfect ground to spread information to a broader section of society. The information that could circulate between a relative few can now circulate globally overnight. This advantage also marked as the reason to increase in the number of fake news attacks by its users, which is unsuitable for a healthy society. Therefore, there is a need for good model to identify and take down fake information as soon as they appear. This project aims at solving the problem by automating the process of identifying fake news using its content. Evaluation metrics like the accuracy of correct classification, precision, recall and f1-score assess the performance of the model. In this model, we have used classification techniques like Logistic Regression, Random Forest Classifier and Decision Tree Classifier. Output of the model using feature extraction techniques as Term Frequency-Inverted Document Frequency (TF-IDF) and Logistic Regression as classifier, has accuracy of 91.58%.

Contents

1	INTRODUCTION	8
2	PROOF OF CONCEPT	9
3	SCRUM MEETINGS	11
4	IMPLEMENTATION	14
4.1	System Architecture	15
4.2	Dataset	15
4.3	Modules	15
4.3.1	DATA PRE-PROCESSING	15
4.3.2	MODEL EVALUATION	15
4.3.3	TRAINING THE MODEL	15
4.3.4	DEPLOYMENT	16
5	RESULT ANALYSIS	19
6	CONCLUSION AND FUTURE SCOPE	20
6.1	Conclusion	20
6.2	Future Scope	20
7	CODING	21
7.1	Fake News Detection.ipynb	22
7.2	app.py	23

7.3	index.html	24
8	SCREEN SHOTS	27
9	REFERENCES	30

Chapter 1

INTRODUCTION

Fake news spreads like a wildfire and this is a big issue in this era. A type of yellow journalism, fake news encapsulates pieces of news and is generally spread through social media and other online media. It influences decisions and opinions of viewers. It is a phenomenon which is having a strong impact on our social life, in particular in the political world. In the recent years, online content has been playing a significant role in fabricating fake news. Because of the wide spread of fake news in online platforms we had been lost our belief in these online platforms.

Information through conventional resources like newspapers, Television etc. follow some rules and regulations. But in the case of online websites and social media platforms, the sharing of information is free most of the time. So there may be a wide spread of fake news that affects the social life of the people. It may cause political and economical imbalance in the society. Fake News can affect the life of an individual or an entire world and causes a strong negative impact on various fields in our societies like health, political, economical .etc.

In this Project, news from different online platforms has taken. After training the model, the news from a website is tested. It will learn from the data that was already trained to predict whether a given news is fake or not.

Chapter 2

PROOF OF CONCEPT

It is based on a paper known as Detection of Online Fake News Using N-Gram Analysis and Machine Learning Techniques. by Hadeer Ahmed, Issa Traore and Sherif Saad.

The main objective of “Fake News Detection ” project is to predict whether a news is real or not. This will identify a given news is fake or not by learning the trained data. It is an effective way to check the genuineness of the news. System will help the human to find right information and to take the right decisions without any kind of misleads in the social life, in particular political and economical world.

- The proposed system is based on identifying the fakeness of a news. In this system the user can identify the real news, if it gets reported. Trained a combination of different machine learning algorithms using various machine learning methods that are not thoroughly explored in the current literature.

The objectives are:

- To predict whether a news is real or not.
- To create a machine learning algorithm with high accuracy and more efficiency.
- Comparison of accuracy results when different machine learning algorithms used.

- Detecting fake news based on text.

Chapter 3

SCRUM MEETINGS

On 24-11-2021

On this day I started searching the miniproject topic based on the new technology such as deep learning,IoT,machine learning,classification,prediction etc.

On 29-11-2021

The topic was selected and did the detail study of the topic,the required dataset was selected.The dataset was searched from the different site such as Kaggle,dataset etc.

On 06-12-2021

This day I submitted the synopsis and research paper to guide for the topic approval.

On 15-12-2021

After getting approval from the guide, the algorithm and model for the project were structured.Then the algorithm were choosen.

On 18-12-2021

On this day teacher took a detailed class on how to do the project, what IDEs to use, what paper are referred, what steps are followed to do the project and so on.

On 06-01-2022

According to the project the required IDE such as Jupyter Notebook, Colab are chosen. Even checked whether the system was efficient to train the model. Here colab to code the project, then started to deploying the model using the algorithm. Python language is used to code the project.

On 10-01-2022

After the project first review according to guide's opinion added two new algorithm to the project to find which algorithm is having more accuracy rate.

On 13-01-2022

Used different algorithm/data model then choose the maximum accuracy one. The algorithm used are:-

Random Forest Classifier
Decision Tree Classifier
Logistic Regression

On 19-01-2022

Started to do project coding. Firstly study the dataset and download the dataset from Kaggle. The dataset contains news from different news websites and online social media platforms.

On 25-01-2022

Testing the data application

On 28-01-2022

The training done in three different data model then choose the maximum accuracy algorithm for fake news detection. Logistic Regression model is used for prediction.

On 02-02-2022

Created the git repository.

On 07-02-2022

Used flask for connection.

Chapter 4

IMPLEMENTATION

Dataset contain news which is taken from different online platforms has taken from Kaggle.com.Dropped the label column from the data frame. Data pre-processing steps are applied and then split the dataset into train and test data.Feature extraction is done using tfidf vectorizer. I have created a decision tree classifier and random forest classifier to predict on tfidf train set. From sklearn, Logistic Regression is imported and response was predicted for test data set. Evaluated model performance is done with help of test accuracy, precision, recall and f1-score.

TOOLS OR PROGRAMMING LANGUAGE

- FRONT END :

- Html
 - Flask

- BACK END :

- Python

4.1 System Architecture

4.2 Dataset

The dataset have been used taken from Kaggle. Dataset consists of 6335 news in which 3164 are fake news and 3171 are real news. There were four attributes:-Id, title, text and label.

4.3 Modules

4.3.1 DATA PRE-PROCESSING

Data pre-processing is the process of importing and measuring information on variables of interest. Handles the missing values and encoding. Splitting of the dataset is also included in this process.

4.3.2 MODEL EVALUATION

Model Evaluation is the most important stage of the model development process. It helps to find the best model that represents our data and how well the chosen model will work in the future. Evaluating model performance with the data used for training is not acceptable in data science because it can easily generate overfit models. It helps to validate the performance of the model.

4.3.3 TRAINING THE MODEL

A training data is a dataset that is used to train an machine learning algorithm. It consists of the sample output data and this sets of input data that have an influence on the output data. The training model is used to run the input data through the algorithm to correlate the processed output against the sample output. The result from this used for the modification of the model.

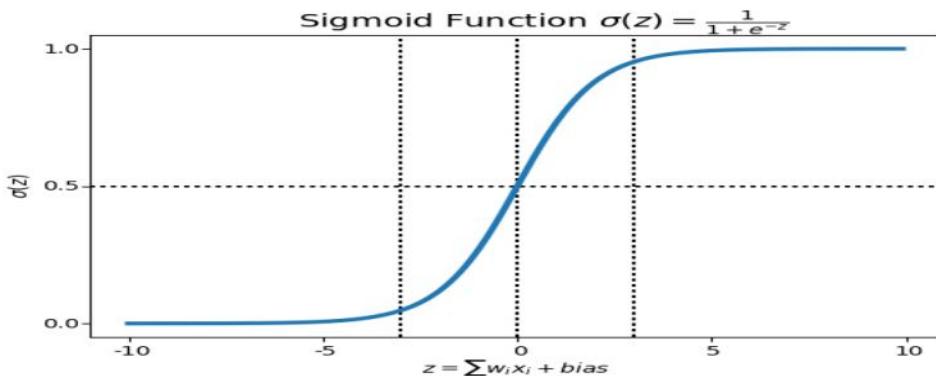
4.3.4 DEPLOYMENT

Model deployment is an engineering task of exposing an machine learning model to available to use.

Algorithm

Logistic Regression

Logistic regression is a classification algorithm used to predict the output of a categorical dependent variable. Logistic regression transforms its output using the logistic sigmoid function to return a probability value between 0 and 1. The outcome must be a categorical or discrete value. A Logistic Model Fitting Algorithm is a discriminative maximum entropy-based generalized linear classification algorithm that accepts a logistic model family. Context: It can range from (typically) being a Binomial Logistic Regression Algorithm to being a Multinomial Logistic Regression Algorithm.



Random Forest Classifier

This is an algorithm which ensembles the less predictive model to produce better predictive models. It aggregates the base model to create a large model. The features are sampled and passed to trees without replacement to build the highly uncorrelated decision trees. To select the best split it is required to have less correlation between the trees. The main concept that makes random forest different from the decision tree is aggregated uncorrelated trees.

Decision Tree Classifier

In a decision tree, the algorithm begins with the root node of the tree for predicting the class of the given data. It compares the values of root attribute with the real data attribute and based on the learning, follows the branch and jumps to the next node. For the next node, the algorithm again learns the attribute value with the other sub-nodes and move further. It continues the process until the leaf node of the tree reaches.

Chapter 5

RESULT ANALYSIS

Implemented an Fake News Detection System that would help us to detect the fake news based on the text. The system can be used to predict the genuineness of the news provided with the help of the news in the database.

Chapter 6

CONCLUSION AND FUTURE SCOPE

6.1 Conclusion

The goal of this project is to detect the fake news that are spreading through various news websites and online social media platforms. Also, to find a high accuracy machine learning model for the fake news detection. It uses three classification algorithm and good accuracy for the system. The accuracy obtained when Logistic Regression is used is 91.58%, 89.74% for Random Forest Classifier and 80.06% for Decision Tree Classifier.

6.2 Future Scope

Spread of fake news in areas of Politics, Health .etc affects the people. In future, run the model on some other publicly available datasets of different fields and large size dataset would give more accurate result and model

Chapter 7

CODING

7.1 Fake News Detection.ipynb

```

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import pickle
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.linear_model import LogisticRegression
from sklearn.ensemble import RandomForestClassifier
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import accuracy_score, confusion_matrix, classification_report
df = pd.read_csv('news.csv')
df.columns = ['id', 'title', 'text', 'label']
print(df.shape)
print(df.head())
print(df.info())
print(df.isnull().sum())

print(df.groupby(['label'])['text'].count())
df.groupby(['label'])['text'].count().plot(kind="bar")
plt.show()

x = df["text"]
y = df["label"]
tfidf_vectorizer = TfidfVectorizer(stop_words='english', max_df=0.7)

X_train, X_test, y_train, y_test = train_test_split(x,y,test_size=0.3)

# Fit and transform train set, transform test set
tfidf_train = tfidf_vectorizer.fit_transform(X_train)
tfidf_test = tfidf_vectorizer.transform(X_test)

# Initialize classifiers
print("LOGISTIC REGRESSION")
LR = LogisticRegression()
LR.fit(tfidf_train,y_train)
pred_lr=LR.predict(tfidf_test)
LR.score(tfidf_test, y_test)
print(classification_report(y_test, pred_lr))
print(confusion_matrix(y_test, pred_lr,labels=['FAKE','REAL']))

print("RANDOM FOREST CLASSIFIER")
RFC = RandomForestClassifier(random_state=0)
RFC.fit(tfidf_train, y_train)
pred_rfc=RFC.predict(tfidf_test)
RFC.score(tfidf_test, y_test)
print(classification_report(y_test, pred_rfc))
print(confusion_matrix(y_test, pred_rfc,labels=['FAKE','REAL']))

print("DECISION TREE CLASSIFIER")
DT = DecisionTreeClassifier()
DT.fit(tfidf_train, y_train)
pred_dt=DT.predict(tfidf_test)
DT.score(tfidf_test, y_test)
print(classification_report(y_test, pred_dt))
print(confusion_matrix(y_test, pred_dt,labels=['FAKE','REAL']))
# saving vectorizer
with open('tfid.pickle', 'wb') as f:
    pickle.dump(tfidf_vectorizer,f)

# saving model
with open('model_fakenews.pickle','wb') as f:
    pickle.dump(LR,f)

```

7.2 app.py

```
1 from flask import Flask, render_template, request, url_for, Markup, jsonify
2 import pickle
3
4 app = Flask(__name__)
5 pickle_in = open('model_fakenews.pickle', 'rb')
6 LR = pickle.load(pickle_in)
7 tfid = open('tfid.pickle', 'rb')
8 tfidf_vectorizer = pickle.load(tfid)
9
10 @app.route('/')
11 def home():
12     return render_template("index.html")
13
14 @app.route('/newscheck')
15 def newscheck():
16     abc = request.args.get('news')
17     input_data = [abc.rstrip()]
18     # transforming input
19     tfidf_test = tfidf_vectorizer.transform(input_data)
20     # predicting the input
21     y_pred = LR.predict(tfidf_test)
22     return jsonify(result = y_pred[0])
23
24
25 if __name__ == '__main__':
26     app.run(debug=True)
27
```

7.3 index.html

```
<!DOCTYPE html>
<html>
<head>
    <title>FAKE NEWS DETECTION</title>
    <meta charset="utf-8">
    <meta name="viewport" content="width=device-width, initial-scale=1">
    <link rel="stylesheet" href="https://maxcdn.bootstrapcdn.com/bootstrap/4.4.1/css/bootstrap.min.css">
    <script src="https://ajax.googleapis.com/ajax/libs/jquery/3.4.1/jquery.min.js"></script>
    <script src="https://cdn.jsdelivr.net/npm/popper.js@1.16.0/dist/umd/popper.min.js"></script>
    <script src="https://maxcdn.bootstrapcdn.com/bootstrap/4.4.1/js/bootstrap.min.js"></script>
    <script src="https://cdn.jsdelivr.net/npm/bootbox.js@5.4.0/bootbox.min.js"></script>
    <style>
        body{
            background: black;
            color: white;
        }
        .modal-body,.modal-footer,.modal-header{
            background:#fff;
            border: none;
            color: black;
        }
        .modal-header .close{
            color: red;
            text-shadow: none;
            size: 1.9rem;
        }
        .close{
            opacity: 1;
        }
    </style>
</head>
```

```
<body>
    <div align="center" style="margin-top: 50px;" class="container">
        <div class="card bg-transparent">
            <div class="card-header bg-transparent">
                <p id="source" data-text="FAKE NEWS DETECTION !"></p>
                <h1 id="dest"></h1>
            </div><br>
            <div class="card-body bg-transparent">
                <form><br>
                    <textarea type="text" class="form-control bg-dark text-white" name="news" id="news" placeholder="Paste your news here!" rows="5"></textarea><br>
                    <button type="submit" class="btn btn-lg btn-success" name="submit" id="submit">PREDICT</button>
                </form>
            </div>
        </div>
    </div>
</body>
</html>
<script>
    var source = $('#source').attr('data-text');
    var dest = $('#dest');
    function typeWriter(text, n) {
        if (n < (text.length)) {
            dest.html(text.substring(0, n+1));
            n++;
            setTimeout(function() {
                typeWriter(text, n)
            }, 150);
        }
    }
    typeWriter(source, 0);

    $SCRIPT_ROOT = "{{ url_for('newscheck') }}";
    $(function() {
        $('#submit').bind('click', function() {

```

```
var news = $('#news').val();
if(news === "" || news === " " || news === "\n" || news === null){
    bootbox.alert({
        size: "big",
        title: "EMPTY FIELD",
        message: "Please enter some news!",
        backdrop: true
    });
}
else{
    $.getJSON($SCRIPT_ROOT, {
        news: news,
    }, function(data) {
        if(data.result == "REAL"){
            var src = "static/success.gif";
        }
        else{
            var src = "static/fail.gif";
        }
        bootbox.alert({
            size: "big",
            title: "PREDICTION",
            message: "<div align='center'><h2>NEWS IS "+data.result+"</h2><img style='width:240px;height:232px;' src='"+src+"' /></div>",
            backdrop: true,
            callback: function(){
                setTimeout(function(){
                    //do what you need here
                    location.reload();
                }, 100);
            }
        });
    });
}
return false;
});
});
</script>
```

Chapter 8

SCREEN SHOTS

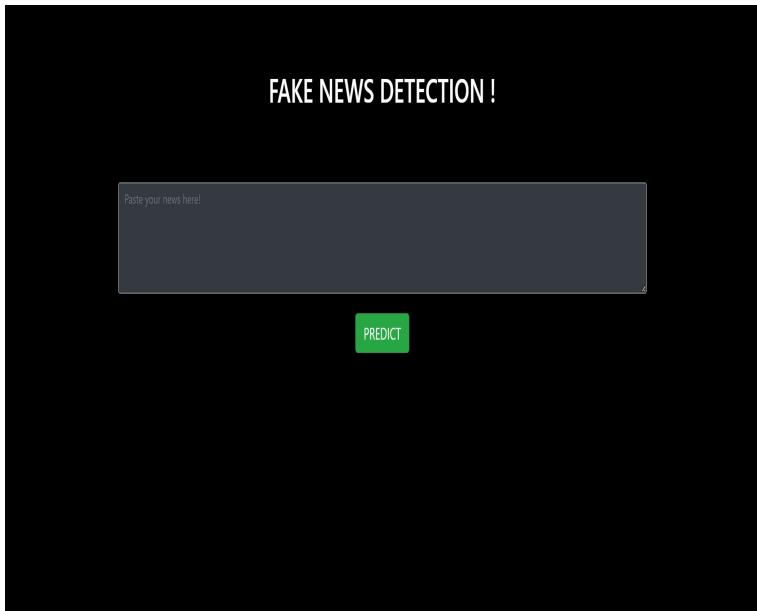
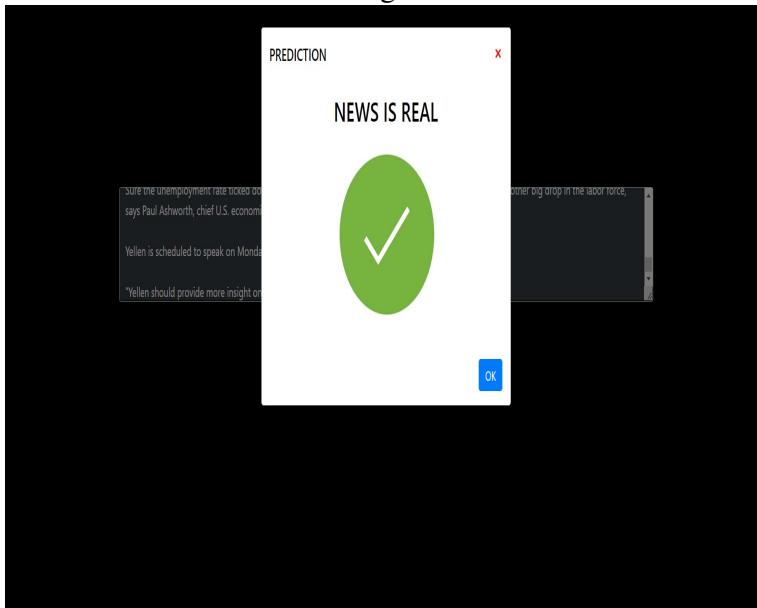


Figure 8.1: PREDICTION





Chapter 9

REFERENCES

- www.youtube.com
 - www.wikipedia.com
- 1 Detection of Online Fake News Using N-Gram Analysis and Machine Learning Techniques by Hadeer Ahmed, Issa Traore and Sherif Saad.
- [2] Inventive Communication and Computational Technologies pp 155-164: Fake News Detection Using Passive-Aggressive Classifier by Saloni Gupta and Priyanka Meel
- [3] <https://www.kaggle.com>