PROJECT REPORT

on

Fake News Detection Using Machine Learning

(CSE VI Semester project)

2021-2022

Submitted to:

Mr. UMANNG GARG

(CC-CSE-A-VI-Sem)

Guided by:

Submitted by:

Mr. HIMANSHU CHAND

Roll. No.: 1918375

CSE-A-VI-Sem

DEPARTMENT OF COMPUTER SCIENCE AND INFORMATION TECHNOLOGY

GRAPHIC ERA HILL UNVERSITY, DEHRADUN

Fake News Detection Using Machine Learning

ABSTRACT

Consuming news from social media is becoming increasingly popular nowadays. Social media brings benefits to users due to the inherent nature of fast dissemination, cheap cost, and easy access. However, the quality of news is considered lower than traditional news outlets, resulting in large amounts of fake news. Detecting fake news becomes very important and is attracting increasing attention due to the detrimental effects on individuals and the society. The performance of detecting fake news only from content is generally not satisfactory, and it is suggested to incorporate user social engagements as auxiliary information to improve fake news detection. Thus it necessitates an in- depth understanding of the correlation between user profiles on social media and fake news. I perform a comparative analysis over explicit and implicit profile features between these user groups, which reveals their potential to differentiate fake news.

INTRODUCTION

With the increasing popularity of social media, more and more people consume news from social media instead of traditional news media. Fake news is now viewed as one of the greatest threats to democracy, and freedom of expression. It has weakened public trust in government. The reach of fake news will be best highlighted all the time. The extensive spread of fake news has the potential for extremely negative impacts on individual and society. A type of yellow journalism, fake news encapsulates pieces of news that may be hoaxes and is generally spread through social media and other online media. This is often done to further or impose certain ideas and is often achieved with political agendas. Such news items may contain false and/or exaggerated claims, and may end up being verbalized by algorithms, and users may end up in a filter bubble.

MOTIVATION

In today's modern world we are so much connected to social media and this often helps us in gathering hot topics and current affairs on regular basis.

But sometimes the scenario is bit different when we encounter two contradicting news on different platforms. At this time we got confused in believing on particular side as we don't know which news is real and which one is fake. To overcome this situation one needs a platform where he/she can search for the news for its correctness.

So, the idea of creating such platform encouraged me to develop this project and help the society as well with this kind of project.

WORKING AND METHODOLOGY

Primarily, we take input in the form of a dataset through the user and then the dataset is handled using the model build based on the count vectorizer or a tf-idf matrix (word tallies how often they are used in other articles in our dataset).[1]

Now the next step is to extract the most optimal features for count vectorizer or tfidf-vectorizer, this is done by using a n-number of the most used words mainly removing the stop words which are common words such as “the”, “when”, and “there” and only using those words that appear at least a given number of times in a given text dataset.

Since this problem is a kind of text classification, Implementing **passive aggressive algorithm** will be best as this is standard for text-based processing. The actual goal is in developing a model which was the text transformation (count vectorizer vs tfidf vectorizer) and choosing which type of text to use.

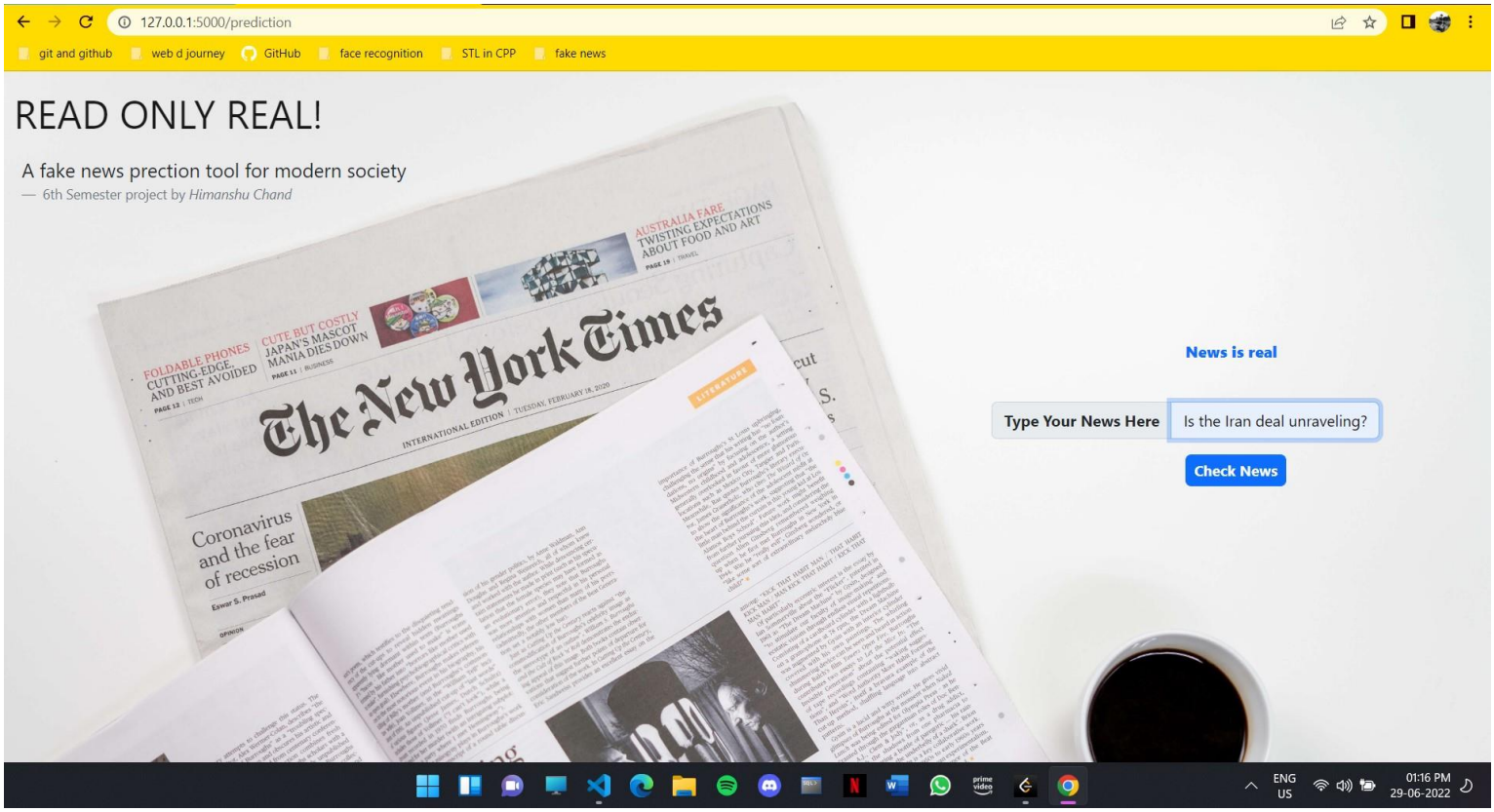
Finally, we are obtaining the accuracy and printing the true and false positives and negatives using confusion matrix. Confusion matrix is nothing but a table used to describe the performance of classification model or classifier on set of data set. We obtain the accuracy based on how well our classifiers are working and the data set is bet fit in it. Accuracy is calculated based on true and false positives and negatives.

The final accuracy achieved by the model was **94.71%**.

After successfully building the ML model it was then exported using pickle python library along with as ‘finalized_model.pkl’ and ‘vectorizer.pkl’ respectively.

Later on this model is then used with flask to create a web application to handle the new fake news search and provide a beautiful user interface with the help of HTML , CSS and Bootstrap.

SCREENSHOTS OF THE PROJECT



flask file:-

Prediction.html app.py

```python
from flask import Flask, escape, request, render_template
import pickle

vector = pickle.load(open("vectorizer.pkl", 'rb'))
model = pickle.load(open("finalized_model.pkl", 'rb'))
app = Flask(__name__)

@app.route("/")
def home():
    return render_template("prediction.html")

@app.route("/prediction",methods=['GET','POST'])
def prediction():
    if request.method == "POST":
        news = str(request.form['news'])

        predict = model.predict(vector.transform([news]))
        print(predict)

        if(predict=='REAL'):
            output="News is real"
        elif(predict=='FAKE'):
            output="News is fake"
        return render_template("prediction.html", result=output)
    else:
        return render_template("prediction.html")



if __name__ == '__main__':
    app.run()
```

Collab notebook :-

Fake News Detection.ipynb

File Edit View Insert Runtime Tools Help All changes saved

+ Code + Text

```
4    REAL
Name: label, dtype: object
```

```
from sklearn.model_selection import train_test_split
```

```
x_train, x_test, y_train, y_test= train_test_split(df["text"],labels, test_size=0.2, random_state =20)
```

```
x_train.head()
```

```
4741    NAIROBI, Kenya – President Obama spoke out Sun...
2089    Killing Obama administration rules, dismantlin...
4074    Dean Obeidallah, a former attorney, is the hos...
5376      WashingtonsBlog \nCNN's Jake Tapper hit the ...
6028    Some of the biggest issues facing America this...
Name: text, dtype: object
```

```
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.linear_model import PassiveAggressiveClassifier
```

```
vector =TfidfVectorizer(stop_words='english', max_df=0.7)
```

```
tf_train = vector.fit_transform(x_train)
tf_test =vector.transform(x_test)
```

```
pac= PassiveAggressiveClassifier(max_iter=50)
pac.fit(tf_train,y_train)
```

```
PassiveAggressiveClassifier(max_iter=50)
```

```
from sklearn.metrics import accuracy_score, confusion_matrix
y_pred= pac.predict(tf_test)
```

```
score =accuracy_score(y_test, y_pred)
```

```
print(f"Accuracy : {round(score*100,2)}%")
```

```
Accuracy : 94.71%
```

HTML File:-

Prediction.html app.py

```
<!doctype html>
<html lang="en">
  <head>
    <meta charset="utf-8">
    <meta name="viewport" content="width=device-width, initial-scale=1">
    <title>READ ONLY REAL</title>
    <style>
        body,nav{
            background-image:url('../static/news.jpg');
            background-size: 100%;
        }
    </style>
    <link href="https://cdn.jsdelivr.net/npm/bootstrap@5.2.0-beta1/dist/css/bootstrap.min.css" rel="stylesheet" integrity="sha384-0evHe/X
    +R7YkIZDRvuzKMRqM+OrBnVFBL6DOitfPri4tjfHxaWutUpFmBp4vmVor" crossorigin="anonymous">
  </head>
  <body >
    <!-- <h1>Hello, world!</h1> -->
    <nav class="navbar navbar-expand-lg bg-light">
        <div class="container-fluid">
          <a class="navbar-brand" style="font-size:40px ;">READ ONLY REAL!</a>
          </button>
          <div class="collapse navbar-collapse" id="navbarNavAltMarkup">
            <div class="navbar-nav">
            </div>
          </div>
        </div>

      </nav>
      <div style="padding-left:20px;">
        <figure>
          <blockquote class="blockquote">
            <p>A fake news prection tool for modern society</p>
          </blockquote>
          <figcaption class="blockquote-footer">
            6th Semester project by <cite title="Source Title">Himanshu Chand</cite>
          </figcaption>
        </figure>
    </div>
```

REFERENCES

1.-This project is inspired by Adam Geitgey

Data science playlist by codejay

2.GeeksForGeeks

Passive Aggressive Classifiers - GeeksforGeeks

3.Udemy Machine learning courses

Machine Learning A-Z (Python & R in Data Science Course) | Udemy