**A**

**MINI PROJECT REPORT** on

**CYBERCRIME DETECTION**

**USING MACHINE LEARNING**

**(FAKE NEWS DETECTION SYSTEM)**

**By**

AKANKSHA DUBEY

ROLL NO.: 2016600

B. Tech (CSE )4th SEM

SECTION -C



Under Supervision of

**Dr Priya Matta**

**Associate Professor, Department of CSE**

**Graphic Era Deemed to be University, Dehradun**



**CERTIFICATE**

This is Certified that Ms. Akanksha Dubey (Roll No.- 2016600) has developed mini project on “**Cybercrime Detection using Machine Learning**” for the fulfilment of B. Tech (CSE) 4th Semester Mini Project in Graphic Era Deemed to be University, Dehradun. The project carried out by student is bonafide work as best of my knowledge.

Date: 13-07-2022

**Project Guide**

**Dr Priya Matta**

**Associate Professor,**

**Department of CSE**

**GEU, Dehradun**

# 

# ACKNOWLEDGMENT

I Would like to thank my Project Guide **Dr Priya Matta, Associate Professor, Department of CSE** for her patience, support and encouragement throughout the completion of this mini project and for having faith in me.

I wish to thank my parents for their continuing support and encouragement. I also wish to thank them for providing me with the opportunity to reach this far in my studies.

At last, but not the least I am grateful indebted to all other persons who directly or indirectly helped me during this project.

**AKANKSHA DUBEY**

**ROLL NO. -2016600**

**B. TECH(CSE)**

**SECTION - C**

**SEMESTER – 4th**

**GEU, DEHRADUN**

# 

# TABLE OF CONTENT

|  |  |  |
| --- | --- | --- |
| SERIAL NO. | TITLE | PAGE NO. |
| 1. | INTRODUCTION  1.1 PURPOSE  1.2 FAKENEWS DETECTION SYSTEM OVERVIEW  1.3 MOTIVATION  1.4 PROBLEM STATEMENT | 1 |
| 2. | PROJECT  2.1 INTRODUCTION TO CYBERCRIMES  2.2 CYBERCRIME DETECTION TECHNIQUES  2.2.1 CYBERCRIME DETCETION UISNG MACHINE LEARNING  2.3 FAKE NEWS  2.3.1 FAKE NEWS DETECTION SYSTEM | 1-3  1  2  2  2  3  5 |
| 3. | METHODOLOGY FOLLOWED  3.1 LOAD THE DATA SET AND IMPORT LIBRARIES  3.2 DATA SPLITTING AND MODEL TRAINING  3.3CREATING FLASK APP AND WEB INTERFACE FOR THE MODEL | 3-4  3  4  4 |
| 4. | FUTURE SCOPE AND APPLICATION | 5 |
| 5. | LIST OF FIGURES  FIGURE 1. CATAGEGORIZATION OF CYBERCRIME DETECTION TECHNIQUES  FIGURE 2. WEB INTERFACE OF FAKE NEWS DETECTION SYSTEM  FIGURE 3. WORK FLOW OF PROJECT  CODE SNIPPTE  1. | 2  3  3 |
| 6. | REFERENECES | 5 |

## 1. INTRODUCTION

This section gives a scope description and overview of everything included in this Mini Project Report.

### 1.1. Purpose

The purpose of this document is to give a detailed description of Fake News Detection System to Detect the Cybercrime. It will illustrate the purpose and complete declaration for the building the machine learning model which will predict the news is fake or not. It will also explain about the Methodology Used to build this Cybercrime detection system. The process could be done automatically without having humans manually review thousands of news related articles

### 1.2. Overview

This Fake news Detection System which is the Sub Part of Cyber Crime Detection System is Based on Web based Machine Learning model based on Flask which is the web application framework written in **Python**. This Web Based system is easy to use and can detect the authenticity of the news headlines using the given dataset.

### 1.3. Motivation

The aim of this project is to create a responsive website using Machine Learning Model which will predict the authenticity of the news.

Fake News spread is also considered as the one of the cybercrimes which is the threat to the nation’s security and financial health. Fake news is a baseless information that is projected in the form of real news. The introduction of social media along with sharing of the huge volumes of unconfirmed information has led to creation of an environment where untruths, and lies spread rapidly. Violence is one of the real-world consequences of fake news.

So, there is a need to detect whether the news is fake or real before it gets spread to the community or people to stop this type of cybercrime.

**1.4. Problem Statement**

Cybercrime Detection using Machine Learning (Fake News Detection System).

## 2.PROJECT

### 2.1 Introduction to Cybercrimes

### The internet the world is growing rapidly. Which is the reason for growth to new opportunities in every field such as entertainment, business, sports, or education. If the internet has advantages, then there are also some disadvantages like cybercrimes.

### A Cybercrime is any form of criminal activity that involves the use of a computer, networked devices, or a network.

### Cybercrimes are increasing within the cyber world. There ought to be some advanced security measures taken to avoid the number of cyber-attacks.

### The term Cybercrime generally cover crimes such as credit card fraud, Bank Robbery, illegal downloading, fake news spreading, cyber terrorism, spam, and so on.

### 2.2 Cybercrime – Detection Techniques

The best detection is the strength of the implementation of security controls since attackers always try to attack the weakness and vulnerabilities therefore security controls and prevention measures offer detection of the potential attacks, prevention, and corrective capabilities to minimize the impact of cybercrime.

As there is an advancement in the ways of implementing cybercrime none of the traditional cybercrime detection systems implemented can completely stop or mitigate them. Cybercriminals improve their methods and utilize new technologies to commit crimes and achieve their goals.

Many studies and research have been conducted by the organization to develop methods for detecting cybercrimes. The main categories of these methods are.

**Cybercrime Detection Techniques**

**Statistical**

**Artificial Intelligence**

**Computer Vision Techniques**

**Biometric Techniques**

**Cryptography**

**Other Techniques**

**Data Mining**

**Machine Learning**

**Figure 1. Categorization of Cybercrime detection Techniques**

**2.2.1 Cybercrime Detection Using Machine Learning**

The Machine learning (ML) is a type of artificial intelligence ([AI](https://www.techtarget.com/searchenterpriseai/definition/AI-Artificial-Intelligence)) that allows software applications to become more accurate at predicting outcomes without being explicitly programmed to do so. Machine learning [algorithms](https://www.techtarget.com/whatis/definition/algorithm) use historical data as input to predict new output values.

Machine learning is the science of predicting outputs based on given input data, also called training data. The machine (i.e., computer) learns how to predict correct and appropriate outputs for specific inputs using the training data. This learning process can be supervised or unsupervised. In the supervised learning method, the training data contain pairs: an input and its corresponding output. The outputs are called labelled outputs because the correct output is already known. The machine tries to learn how pairs are built in order to make its own predictions later.

**2.3 Fake News**

Fake news means incorporating information that leads people or communities to the wrong path.

In today’s time, Fake news spreading is fast that people share the information without verifying it. This is often done with political agendas and disturbed the peace of the society. Spreading of fake news intentionally through network systems comes under Cybercrime.

For the media industry delivering true news is important for them to create trust among the people. Also, detection of fake news is important in time to avoid any conflicts between the communities or countries.**2**

**2.3.1 Fake News Detection System**

The Fake news Detection System is a web-based system designed with the aim to detect the authenticity of news to avoid any cybercrime. This system is developed using the Machine Learning algorithm and using the flask app, which helps to detect fake news from the given news articles. It is a web-based application so it is easy to use and predict fake news with accurate results

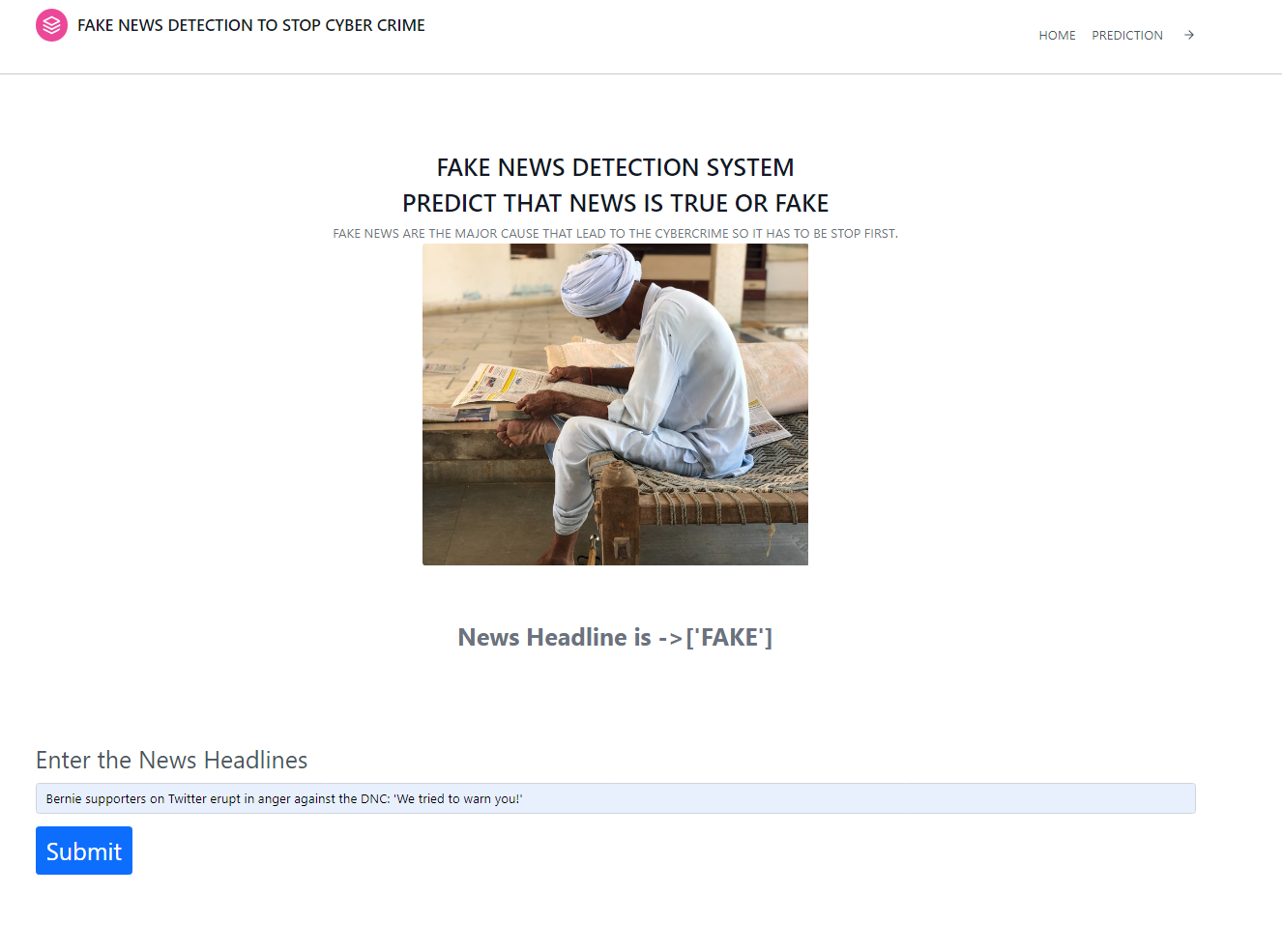


FIGURE 2. Web interface of the Fake News Detection System

### 3. Methodology Followed

#### FIGURE 3. WORK FLOW

#### 3.1 Load Data Set from Kaggle & import Libraries for Machine learning model

For any Machine Learning Project, the most fundamental part is the data. In this project I have used the fake and real news dataset from Kaggle to build my machine learning model. There is one CSV files in the folder, **news** corresponding to real and fake news.see (code 1)

df = pd.read\_csv("news.csv")

3

In order to perform the classifications, we need collection of modules or codes that can be used repeatedly in different programs to make the program efficient. In this project the libraries used are in( code 2)

import pandas as pd #**working with relational or labelled data**

import numpy as np #for working with arrays

import itertools # **to iterate over data structures that can be stepped over using a for-loop**

! pip install sklearn #**use for data analysis, classification, clustering**

import TfidfVectorizer #**compute the word occurrence frequency matrix**

import PassiveAggressiveClassifier #**machine learning algorithm for classification**

import pickle **#used for serializing and de-serializing**

#### 3.2 Data Splitting and Model Training and Validation

For Data splitting means dividing the data into two subparts one is used for train the model and other is used to compare or test the data. In this project train\_test\_split function from Sklearn model selection id used for splitting the arrays into two parts: for testing data and for training data.

x\_train, x\_test, y\_train, y\_test = train\_test\_split(df["text"],labels,test\_size = 0.2, random\_state = 20)

**Passive Aggressive Classifier:** In this project I have used this Machine learning algorithm. It works by responding as passive for real/correct prediction and aggressive for any fake or miscalculations. I have used this machine learning algorithm to train my fake news detection model and to check the accuracy of the model. A confusion matrix is used to summaries the prediction result of the classification into fake and real.

#Initilise a PassiveAggressiveClassifier

pac = PassiveAggressiveClassifier(max\_iter=50)

pac.fit(tf\_train,y\_train)

#To predict the datasets

from sklearn. metrics import accuracy\_score, confusion\_matrix

y\_pred = pac.predict(tf\_test)

score = accuracy\_score (y\_test, y\_pred)

print (f"Accuracy of the model: {round(score\*100,2)}%")

Accuracy of the model: 94.95%

#Confusion matrics

confusion\_matrix (y\_test, y\_pred,labels=['FAKE','REAL'])

array([[625, 23], [ 41, 578]], dtype=int64)

**Pickle and load Model:** After training the dataset it requires to save the model so that it can be used elsewhere when required.

#save model

import pickle

filename = 'finalized\_model.pkl'

pickle.dump(pac, open(filename, 'wb'))

#### 3.3 Creating Flask App and Web Interface

Flask is the web application framework in Python which helps user to interacts with our python code or ML model. It allows us to create a web application easily. The setup requires two packages first virtual environment to avoid any compatibility problem with different versions on the libraries and second flask Also to style our application I have used Tailwind blocks so that it is visually appealing also.

def prediction():

if request.method == "POST":

news = str(request.form['news'])

print(news)

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

print(predict)

return render\_template("prediction.html",prediction\_text="News Headline is ->{}".format(predict))

else:

return render\_template("prediction.html")

### 4.Future Scope and Application

Advancement in the technology is also leading the chances of increasing cybercrime and on parallel hands it also leads in the enhancement of better and improved measured to stop these cybercrimes. This project on Fake News Detection System can be used by individuals and news channels to verify the news before delivering it to the audience. This practice of verifying the news before sharing it will lead to reduce the chances of cybercrime in one another form.

### 

### 5.References

[Machine Learning: Algorithms, Real-World Applications and Research Directions | SpringerLink](https://link.springer.com/article/10.1007/s42979-021-00592-x)

[Tailwind Blocks](https://tailwindblocks.com/)

[Welcome to Flask — Flask Documentation (2.1.x) (palletsprojects.com)](https://flask.palletsprojects.com/en/2.1.x/)

[Comprehensive Review of Cybercrime Detection Techniques | IEEE Journals & Magazine | IEEE Xplore](https://ieeexplore.ieee.org/abstract/document/9146148)

[Hacking and cybercrime | Proceedings of the 1st annual conference on Information security curriculum development (acm.org)](https://dl.acm.org/doi/abs/10.1145/1059524.1059553?casa_token=Pa2BUqxHLbUAAAAA:RKWFep_ROzMvzbTMGAZ8EYI_DGSaomkMzio3rDCj28Th1o4NhFF8NkXqDIAi_zS2dBeoiTHudBJKeA)