### DETECTION AND CATEGORIZATION OF FAKE NEWS WITH HYBRID APPROACH

#### A PROJECT REPORT

***Submitted by***

**HEMATHI.D [REGISTER NO: 211419104102] ISHANI.S [REGISTER NO:211419104107] MADHUMITHA.V [REGISTER NO:211419104157]**

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

**BACHELOR OF ENGINEERING**

**IN**

#### COMPUTER SCIENCE AND ENGINEERING



**PANIMALAR ENGINEERING COLLEGE**

**(An Autonomous Institution, Affiliated to Anna University, Chennai)**

**APRIL 2023**

### PANIMALAR ENGINEERING COLLEGE

**(An Autonomous Institution, Affiliated to Anna University, Chennai)**

**BONAFIDE CERTIFICATE**

Certified that this project report **“DETECTION AND CATEGORIZATION OF FAKE NEWS WITH HYBRID APPROACH”** is the bonafide work of “**HEMATHI.D(211419104102) , ISHANI.S(211419104107) , MADHUMITHA.V**

**(211419104157)”** who carried out the project work under my supervision.

#### SIGNATURE SIGNATURE

**Dr.L.JABASHEELA,M.E.,Ph.D., Dr. T.JACKULIN M.E,Ph.D., HEAD OF THE DEPARTMENT SUPERVISOR**

#### ASSOCIATE PROFESSOR

DEPARTMENT OF CSE, DEPARTMENT OF CSE,

PANIMALAR ENGINEERING COLLEGE, PANIMALAR ENGINEERING COLLEGE, NASARATHPETTAI, NASARATHPETTAI,

POONAMALLEE, POONAMALLEE,

CHENNAI-600 123. CHENNAI-600 123.

Certified that the above candidate(s) was examined in the End Semester Project Viva-Voce Examination held on...........................

**DECLARATION**

We **Hemathi.D(211419104102), Ishani.S(211419104107), Madhumitha.V (211419104157)** hereby declare that this project report titled “**Detection and categorization of Fake news with Hybrid Approach**”, under the guidance of **Dr. T.Jackulin M.E,Ph.D.,** is theoriginal work done by us and we have not plagiarized or submitted to any other degree in any university by us.

#### HEMATHI.D

**ISHANI.S**

**ACKNOWLEDGEMENT**

We would like to express our deep gratitude to our respected Secretary and Correspondent **Dr.P.CHINNADURAI, M.A., Ph.D.** for his kind words and enthusiastic motivation, which inspired us a lot in completing this project.

We express our sincere thanks to our beloved Directors

#### Tmt.C.VIJAYARAJESWARI, Dr.C.SAKTHI KUMAR,M.E.,Ph.D and

**Dr.SARANYASREE SAKTHI KUMAR B.E.,M.B.A.,Ph.D.,** for providing us with the necessary facilities to undertake this project.

We also express our gratitude to our Principal **Dr.K.MANI, M.E., Ph.D.** who facilitated us in completing the project.

We thank the Head of the CSE Department, **Dr. L.JABASHEELA , M.E.,Ph.D.,** for the support extended throughout the project.

We would like to thank our guide **Dr.T.JACKULIN M.E,Ph.D.,** and all the faculty members of the Department of CSE for their advice and encouragement for the successful completion of the project.

#### HEMATHI.D ISHANI.S MADHUMITHA.V

**TABLE OF CONTENTS**

|  |  |  |
| --- | --- | --- |
| **CHAPTER NO** | **TITLE** | **PAGE NO** |
|  | **ABSTRACT** | iii |
|  | **LIST OF FIGURES** | iv |
|  | **LIST OF TABLES** | v |
| **1** | **INTRODUCTION** |  |
|  | 1.1 Overview | 2 |
|  | 1.2 Problem Definition | 3 |
| **2** | **LITERATURE SURVEY** | 4 |
| **3** | **SYSTEM ANALYSIS** |  |
|  | 3.1 Existing System | 14 |
|  | 3.2 Proposed System | 14 |
|  | 3.3 Project Requirements | 15 |
| **4** | **SYSTEM DESIGN** |  |
|  | 4.1 UML Diagrams | 26 |
|  | 4.1.1 Use Case Diagram | 26 |
|  | 4.1.2 Activity Diagram | 27 |
|  | 4.1.3 Class Diagram | 28 |
|  | 4.2 Data Flow Diagram | 29 |

i

|  |  |  |
| --- | --- | --- |
| **5** | **SYSTEM ARCHITECTURE** |  |
| 5.1 System Architecture Overview | | 32 |
| **6** | **SYSTEM IMPLEMENTATION** |  |
| 6.1 Algorithm | | 34 |
| 6.1.1 Logistic Regression | | 34 |
| 6.1.2 Random Forest | | 35 |
| 6.1.3 Addition Tree Discriminant | | 36 |
| 6.1.4 XGBoost | | 36 |
| 6.2 Module Design Specification | | 37 |
| 6.3 Module Description | | 38 |
| **7** | **PERFORMANCE EVALUATION** |  |
| 7.1 Results and Discussion | | 41 |
| 7.2 Comparitive Analysis | | 42 |
| **8** | **CONCLUSION** |  |
| 8.1 Conclusion | | 46 |
| 8.2 Future Enhancements | | 46 |
| **APPENDICES** | |  |
| 1. Sample Dataset | | 48 |
| 2. Sample Coding | | 51 |
| 3. Sample Screens | | 58 |
| **REFERENCES** | | 60 |

#### ABSTRACT

Fake news is becoming a growing concern, with the potential to influence public opinion and cause social chaos. As a result, there is an urgent need to develop effective fake news detection systems. The three main types of fake news detection approaches are content-based, social context-based, and knowledge-based. Content-based approaches analyse the news's content, such as text, images, and videos, to identify characteristics that are indicative of fake news. Text analysis, sentiment analysis, and credibility analysis are examples of such techniques. Knowledge-based approaches detect fake news by utilising external knowledge sources such as databases, ontologies, and external facts.The social context-based approach focuses on analysing the social context in which the news is being shared. This can include analysing the behaviour of the users sharing the news, the network structure of the users, and the diffusion patterns of information. Existing fake news detection solutions take one of three approaches: linguistic, knowledge-based, or social context-based. Here, we suggest a hybrid methodology for identifying fake news that includes both knowledge-based and content-based methods. We also divide the fake news into different categories according to its content. In comparison to other methods, our experimental results show that the proposed approach detects fake news with high accuracy.

**LIST OF FIGURE**

|  |  |  |
| --- | --- | --- |
| **FIGURE NO.** | **FIGURE NAME** | **PAGE NO.** |
| Fig 4.1.1 | Use case Diagram | 26 |
| Fig 4.1.2 | Activity Diagram | 27 |
| Fig 4.1.3 | Class Diagram | 28 |
| Fig 4.2.1 | Level 0 DFD Diagram | 29 |
| Fig 4.2.2 | Level 1 DFD Diagram | 30 |
| Fig 5.1  Fig 7.1.1 | System Architecture  Screenshot of Web Page for fake news prediction | 32  41 |
| Fig 7.1.2  Fig 7.1.3  Fig 7.2.1  Fig 7.2.2  Fig 7.2.3  Fig 7.2.4  Fig 7.2.5 | Screenshot of News Entered in a text box for detection  Screenshot of Predicted result  Screenshot of Accuracy of Passive Aggressive classifier  Screenshot of Accuracy of Logistic Regression  Screenshot of Accuracy of Decision Tree  Screenshot of Accuracy of XGBoost  Comparison of Algorithm | 41  41  42  43  43  44  44 |

**LIST OF TABLES**

#### TABLE NO.

Table 1

#### TABLE NAME

Sample Dataset

#### PAGE NO.

48

# CHAPTER-1 INTRODUCTION

#### CHAPTER-1 INTRODUCTION

* 1. **OVERVIEW**

The spread of fake news has become a big challenge in the digital environment in recent years. With the rise of social media and online news sources, distinguishing between authentic news and fraudulent information has become increasingly challenging. As a result, there is an increasing demand for dependable and effective systems capable of detecting and identifying fake news.

To fulfil this demand, a new method that blends the capability of natural language processing with knowledge-based reasoning has evolved. This hybrid linguistic and knowledge-based approach leverages both linguistic features and background knowledge to identify and verify the accuracy of news articles.

This system is built around a sophisticated linguistic analysis that evaluates the content and structure of news articles. The system analyses the language used in an article using machine learning algorithms to identify patterns that are indicative of fake news. The tone, sentiment, and style of the writing, as well as the presence of certain words and phrases commonly used in fake news articles, are examples of linguistic features.

In addition to linguistic analysis, includes a knowledge-based component that uses external sources of information to verify the accuracy of news articles. This includes factual information databases like news archives and government reports, as well as expert opinions and analyses from credible sources. the use of linguistic and knowledge-based analysis enables it to provide a more comprehensive and accurate evaluation of news articles. The system can quickly identify potentially fake news articles and provide a high level of confidence in its assessments by combining both approaches.

This hybrid linguistic and knowledge-based approach is a highly effective tool to combat against fake news. As the spread of false information poses a significant threat to the integrity of our information systems, systems like this will be critical in ensuring that accurate information is widely disseminated and trusted.

#### PROBLEM DEFINITION

The proliferation of fake news on social media and other online platforms is a growing concern that has significant implications for public opinion, public safety, and overall trust in information. Despite efforts to counter it, fake news continues to spread quickly and easily, often having a profound impact on public opinion and behavior. The aim of this project is to develop an effective and efficient solution for detecting and categorizing fake news using a combination of machine learning algorithms and knowledge-based approaches. The goal of the project is to provide a tool that can assist users in making informed decisions about the news they consume and to improve overall trust in information by reducing the spread of fake news on social media and other online platforms

# CHAPTER-2 LITERATURE SURVEY

#### CHAPTER-2 LITERATURE SURVEY

1. **TITLE**: A Hybrid Linguistic and Knowledge-Based Analysis Approach for Fake News Detection on Social Media

**YEAR:** 2022

**AUTHORS:** Noureddine Seddari , Abdelouahid Derhab, Mohamed Belaoued , Waleed Halboob , Jalal al-muhtadi , and Abdelghani Bouras

#### ABSTRACT:

The rapid development of different social media and content-sharing platforms has been largely exploited to spread misinformation and fake news that make people believing in harmful stories, which allow to influence public opinion, and could cause panic and chaos among population. Thus, fake news detection has become an important research topic, aiming at flagging a specific content as fake or legitimate. The fake news detection solutions can be divided into three main categories: content-based, social context-based, and knowledge-based approaches. In this paper, we propose a novel hybrid fake news detection system that combines linguistic and knowledge-based approaches and inherits their advantages, by employing two different sets of features: (1) linguistic features (i.e., title, number of words, reading ease, lexical diversity,and sentiment), and

(2) a novel set of knowledge-based features, called fact-verification features that comprise three types of information namely, (i) reputation of the website where the news is published, (ii) coverage, i.e., number of sources that published the news, and (iii) fact-check, i.e., opinion of well-known fact-checking websites about the news, i.e., true or false. The proposed system only employs eight features, which is less than most of the state-of-the-art approaches. Also, the evaluation results on a fake news dataset show that the proposed system employing both types of features can reach an accuracy of 94.4%, which is better compared to that obtained from separately employing linguistic features (i.e., accuracy=89.4% ) and fact-verification features (i.e., accuracy=81.2%).

1. **TITLE:** A Novel Hybrid Multi-Thread Metaheuristic Approach for Fake News Detection in social media.

**YEAR:** 2022

**AUTHORS:** Gungor Yildirim

#### ABSTRACT:

In fake news detection, intelligent optimization seems to be a more effective and explainable solution methodology than the black-box methods that have been extensively used in the literature. This study takes the optimization-based method one step further and proposes a novel, multi-thread hybrid metaheuristic approach for fake news detection in social media. The most innovative feature of the proposed method is that it uses a supervisor thread mechanism, which simultaneously monitors and improves the performance and search patterns of metaheuristic algorithms running parallel. With the supervisor thread mechanism, it is possible to analyse different key attribute combinations in the search space. In addition, this study develops a software framework that allows this model to be implemented easily. It tests the performance of the proposed model on three different data sets, respectively containing news about Covid-19, the Syrian War, and daily politics. The proposed method is evaluated in comparison to the results of fifteen different well-known deep models and classification algorithms. Experimental results prove the success of the proposed model and that it can produce competitive results.

1. **TITLE:** A Systematic Literature Review and Meta-Analysis of Studies on Online Fake News Detection.

**YEAR:** 2022

**AUTHORS:** Robyn C. Thompson, Seena Joseph and Timothy T. Adeliyi

#### ABSTRACT:

The ubiquitous access and exponential growth of information available on social media networks have facilitated the spread of fake news, complicating the task of distinguishing between this and real news. Fake news is a significant social barrier that has a profoundly negative impact on society. Despite the large number of studies on fake news detection, they have not yet been combined to offer coherent insight on trends and advancements in this domain. Hence, the primary objective of this study was to fill this knowledge gap. The method for selecting the pertinent articles for extraction was created using the preferred reporting items for systematic reviews and meta-analyses (PRISMA). This study reviewed deep learning, machine learning, and ensemble-based fake news detection methods by a meta-analysis of 125 studies to aggregate their results quantitatively. The meta-analysis primarily focused on statistics and the quantitative analysis of data from numerous separate primary investigations to identify overall trends. The results of the meta-analysis were reported by the spatial distribution, the approaches adopted, the sample size, and the performance of methods in terms of accuracy. According to the statistics of between-study variance high heterogeneity was found with τ 2 = 3.441; the ratio of true heterogeneity to total observed variation was I2 = 75.27% with the heterogeneity chi-square (Q) = 501.34, the degree of freedom = 124, and p ≤ 0.001. A p-value of 0.912 from the Egger statistical test confirmed the absence of a publication bias. The findings of the meta-analysis demonstrated satisfaction with the effectiveness of the recommended approaches from the primary studies on fake news detection that were included. Furthermore, the findings can inform researchers about various approaches they can use to detect online fake news.

1. **TITLE:** Evaluating the Effectiveness of Publishers’ Features in Fake News Detection on social media.

**YEAR:** 2022

**AUTHORS:** Ali Jarrahi, Leila Safari.

#### ABSTRACT:

With the expansion of the Internet and attractive social media infrastructures, people prefer to follow the news through these media. Despite the many advantages of these media in the news field, the lack of control and verification mechanism has led to the spread of fake news as one of the most critical threats to democracy, economy, journalism, health, and freedom of expression. So, designing and using efficient automated methods to detect fake news on social media has become a significant challenge. One of the most relevant entities in determining the authenticity of a news statement on social media is its publishers. This paper examines the publishers’ features in detecting fake news on social media, including Credibility, Influence, Sociality, Validity, and Lifetime. In this regard, we propose an algorithm, namely CreditRank, for evaluating publishers’ credibility on social networks. We also suggest a high accurate multi-modal framework, namely FR-Detect, for fake news detection using user-related and content-related features. Furthermore, a sentence-level convolutional neural network is provided to properly combine publishers’ features with latent textual content features. Experimental results show that the publishers’ features can improve the performance of content-based models by up to 16% and 31% in accuracy and F1, respectively. Also, the behavior of publishers in different news domains has been statistically studied and analyzed.

1. **TITLE:** Fake News Detection Techniques on social media: A Survey

**YEAR:** 2022

**AUTHORS:** Ihsan Ali, Mohamad Nizam Bin Ayub, Palaiahnakote Shivakumara, Nurul Fazmidar Binti Mohd Noor

#### ABSTRACT:

Social media platforms like Twitter have become common tools for disseminating and consuming news because of the ease with which users can get access to and consume it. This paper focuses on the identification of false news and the use of cutting-edge detection methods in the context of news, user, and social levels. Fake news detection taxonomy was proposed in this research. This study examines a variety of cutting-edge methods for spotting false news and discusses their drawbacks. It also explored how to detect and recognize false news, such as credibility-based, time-based, social context-based, and the substance of the news itself. Lastly, the paper examines various datasets used for detecting fake news and proposed an algorithm.

1. **TITLE:** Fake News Detection in Social Networks using Machine Learning: A Review

**YEAR:** 2021

**AUTHORS:** Sonali Raturi, Amit Kumar Mishra and Srabanti Maji.

#### ABSTRACT:

Fake News is spreading so rapidly these days. This is low-quality news that is generated to targeted someone. This could be created for financial gain or political gain. In no time, millions of tweets are generated and that could be false, people start believing in fake news when there is not enough information available to examine whether the information or the tweet that has been created

is true or false and also people start believing in the information that they hear frequently and that could be false. It has been continuing since traditional media but now it is easier in social media to share or comment on such false information. With the growth of this false news or information, it is impossible to manually filter such news. So, there is some computational approach to recognize fake news with different Machine Learning Algorithms like SVM, Naïve Bayes, etc. This review paper mentioned different types of techniques required to detect hoax news. Also discussed different methods used in existing models with their accuracy.

1. **TITLE:** Fake News Detection from Online media using Machine learning Classifiers

**YEAR:** 2021

**AUTHORS:** Shalini Pandey, San Keerthi Prabhakaran, N V Subba Reddy and Dinesh Acharya

#### ABSTRACT:

With the advancement in technology, the consumption of news has shifted from Print media to social media. The convenience and accessibility are major factors that have contributed to this shift in consumption of the news. However, this change has bought upon a new challenge in the form of "Fake news" being spread with not much supervision available on the net. In this paper, this challenge has been addressed through a Machine learning concept. The algorithms such as K- Nearest Neighbor, Support Vector Machine, Decision Tree, Naïve Bayes and Logistic regression Classifiers to identify the fake news from real ones in a given dataset and also have increased the efficiency of these algorithms by pre- processing the data to handle the imbalanced data more appropriately. Additionally, comparison of the working of these classifiers is presented along with the results. The model proposed has achieved an accuracy of 89.98% for

KNN, 90.46% for Logistic Regression, 86.89% for Naïve Bayes, 73.33% for Decision Tree and 89.33% for SVM in our experiment.

1. **TITLE:** Fake News Detection Using Machine Learning Approaches

**YEAR:** 2020

**AUTHORS:** Z Khanam, B N Alwasel , H Sirafi and M Rashid

#### ABSTRACT:

The fake news on social media and various other media is wide spreading and is a matter of serious concern due to its ability to cause a lot of social and national damage with destructive impacts. A lot of research is already focused on detecting it. This paper makes an analysis of the research related to fake news detection and explores the traditional machine learning models to choose the best, in order to create a model of a product with supervised machine learning algorithm, that can classify fake news as true or false, by using tools like python scikit-learn, NLP for textual analysis. This process will result in feature extraction and vectorization; we propose using Python scikit-learn library to perform tokenization and feature extraction of text data, because this library contains useful tools like Count Vectorizer and Tiff Vectorizer. Then, we will perform feature selection methods, to experiment and choose the best fit features to obtain the highest precision, according to confusion matrix results.

1. **TITLE:** Fake News Detection Using Machine Learning Algorithms

**YEAR:** 2020

**AUTHORS:** Uma Sharma, Sidarth Saran, Shankar M. Patil

#### ABSTRACT:

In our modern era where the internet is ubiquitous, everyone relies on various online resources for news. Along with the increase in the use of social media platforms like Facebook, Twitter, etc. news spread rapidly among millions

of users within a very short span of time. The spread of fake news has far-reaching consequences like the creation of biased opinions to swaying election outcomes for the benefit of certain candidates. Moreover, spammers use appealing news headlines to generate revenue using advertisements via click-baits. In this paper, we aim to perform binary classification of various news articles available online with the help of concepts pertaining to Artificial Intelligence, Natural Language Processing and Machine Learning. We aim to provide the user with the ability to classify the news as fake or real and also check the authenticity of the website publishing the news.

1. **TITLE:** A Smart System for Fake News Detection Using Machine Learning

**YEAR:** 2019

**AUTHORS:** Anjali Jain, Harsh Khatter , AvinashShakya

#### ABSTRACT:

Most of the smart phone users prefer to read the news via social media over internet. The news websites are publishing the news and provide the source of authentication. The question is how to authenticate the news and articles which are circulated among social media like WhatsApp groups, Facebook Pages, Twitter and other micro blogs & social networking sites. It is harmful for the society to believe on the rumors and pretend to be a news. The need of an hour is to stop the rumors especially in the developing countries like India, and focus on the correct, authenticated news articles. This paper demonstrates a model and the methodology for fake news detection. With the help of Machine learning and natural language processing, author tried to aggregate the news and later determine whether the news is real or fake using Support Vector Machine. The results of the proposed model is compared with existing models. The proposed model is working well and defining the correctness of results upto 93.6% of accuracy.

# CHAPTER-3 SYSTEM ANALYSIS

#### CHAPTER-3 SYSTEM ANALYSIS

* 1. **EXISTING SYSTEM**

The major challenges that hinder the efficiency of the existing fake news detection solutions are related to the highly versatile nature of deceptive information. Indeed, it is very difficult to obtain a generalized dataset for fake news detection. Thus, it is very difficult to extract relevant features that can well represent and allow to detect fake news in various domains. Some existing solutions rely on ontologies in order to model fake news domain knowledge, which can be then used to distinguish fake from real news content. As previously discussed, the existing fake news detection solutions are linguistic-based, knowledge-based, or social context-based. Considering the limitations of the aforementioned categories, it would be a good idea to investigate combining two different categories in order to overcome their respective limitations.

#### DISADVANTAGES:

* + - Hosseinmoltagh and Paplexakis investigated the problem of identifying the different types of fake news with high accuracy.
    - Logistic regression is a linear algorithm used for binary classification problems.

#### PROPOSED WORK

We propose a hybrid fake news detection system that combines linguistic and knowledge-based approaches. It also discovers the category of fake news.The proposed fake news detection system is divided into two stages: training and testing.The feature extracting task extracts a set of relevant

features from the training dataset during the training phase, which are then fed

to several machine learning algorithms to build a fake news detection model.The detection model is applied on test data during the testing phase to determine whether the provided news articles are real or fake.

#### Advantages:

* + - A linguistic and knowledge-based hybrid fake news detection system that combines linguistic features with a novel set of knowledge-based features known as fact-verification features.
    - It is essential to detect a particular disease. In our country, many are not so educated to get correct information about all diseases.

#### PROJECT REQUIREMENTS General:

Requirements are the basic constrains that are required to develop a system. Requirements are collected while designing the system. The following are the requirements that are to be discussed.

1. Functional requirements
2. Non-Functional requirements
3. Environment requirements
   1. Hardware requirements
   2. software requirements

#### Functional requirements:

The software requirements specification is the first step in the requirements analysis process. It lists requirements of a particular software system. The following details to follow the special libraries like tensorflow, keras, matplotlib.

#### Non-Functional Requirements:

Process of functional steps,

* 1. Problem define
  2. Preparing data
  3. Evaluating algorithm
  4. Improving results
  5. Prediction the result

#### Environmental Requirements: A.Hardware system configuration:

Processor - Intel i3,i5,i7, AMD Processor RAM - Above 6 GB

Hard Disk - Above 500 GB

#### B.Software system configuration:

Operating System - Windows 7/8/10/11 Front End - HTML ,CSS

Scripts - Python language

Tool - Python

#### SOFTWARE DESCRIPTION

**INTRODUCTION TO PYTHON**

Python is a high-level object-oriented programming language that was created by Guido van Rossum. It is also called general-purpose programming language as it is used in almost every domain we can think of as mentioned below:

* Web Development
* Software Development
* Game Development
* AI & ML
* Data Analytics

This list can go on as we go but why python is so much popular let’s see it in the next topic.

**WHY PYTHON PROGRAMMING?**

You guys might have a question in mind that, why python? why not another programming language?

So let me explain:

Every Programming language serves some purpose or use-case according to a domain. for eg, Javascript is the most popular language amongst web developers as it gives the developer the power to handle applications via different frameworks like react, vue, angular which are used to build beautiful User Interfaces. Similarly, they have pros and cons at the same time. so if we consider python it is general-purpose which means it is widely used in every domain the reason is it’s very simple to understand, scalable because of which the speed of development is so fast. Now you get the idea why besides learning python it doesn’t require any programming background so that’s why it’s popular amongst developers as well. Python has simpler syntax similar to the English language andalso the syntax allows developers to write programs with fewer lines of code. Since it is open-source there are many libraries available that make developers’ jobs easy ultimately results in high productivity. They can easily focus on business logic and Its demanding skills in the digital era where information is available in large data sets.

#### INSTALLING PYTHON PACKAGES

To install Python packages, you can use pip, the package installer for Python. Here are the steps to install Python packages using pip:

* Open a command prompt (Windows) or terminal (Mac/Linux).
* Type pip install <package\_name> and press Enter. Replace <package\_name> with the name of the package you want to install.
* Wait for the installation to complete. pip will automatically download and install the package and its dependencies.

#### HOW TO USE A PYTHON PACKAGE

**Scikit-learn**

scikit-learn is a Python library for machine learning built on top of NumPy, SciPy, and matplotlib. It provides simple and efficient tools for data mining and data analysis, and is widely used for tasks such as classification, regression, and clustering. Here are some of the key features of scikit-learn:

* Support for a wide range of machine learning algorithms, including linear regression, logistic regression, decision trees, random forests, support vector machines, and k-nearest neighbors.
* Tools for feature selection, feature extraction, and data preprocessing, including scaling, normalization, and one-hot encoding.
* Evaluation metrics for machine learning models, including accuracy, precision, recall, and F1 score.
* Cross-validation and hyperparameter tuning tools for model selection and optimization.

Here are some basic commands to get started with scikit-learn:

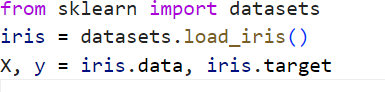
Install scikit-learn using pip:



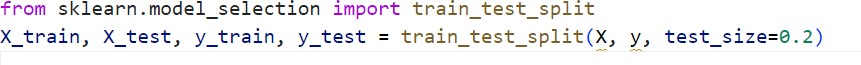
Import scikit-learn in your Python script:



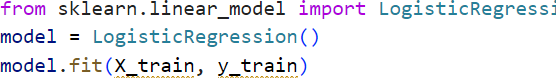
Load a dataset using scikit-learn's built-in datasets:



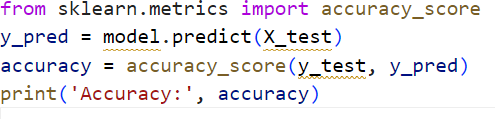
Split the data into training and testing sets:



Train a machine learning model:



Evaluate the Model:



#### PANDAS

Pandas is a popular open-source Python library for data manipulation and analysis. It provides powerful and flexible data structures for working with structured data, as well as tools for cleaning, merging, and reshaping data. Here are some of the key features of pandas:

* Data structures for working with tabular, time-series, and matrix data, including the DataFrame and Series classes.
* Tools for data cleaning and manipulation, including filtering, sorting, grouping, and reshaping data.
* Integration with other libraries for data analysis, such as NumPy, Matplotlib, and Scikit-learn.
* Support for reading and writing data in a variety of formats, including CSV, Excel, SQL databases, and JSON.

Here are some basic commands to get started with pandas:

* + Install pandas:
* *pip install pandas*
  + Import pandas in your Python script:
* *import pandas as pd*
  + Load data from a CSV file into a pandas DataFrame:
* *df = pd.read\_csv('data.csv')*
  + View the first few rows of the DataFrame:
* *print(df.head())*
  + Filter rows based on a condition:
* *filtered\_df = df[df['column\_name'] > 10]*
* Group the data by a column and compute summary statistics:
  + *grouped\_df = df.groupby('column\_name').mean()*
* Merge two DataFrames based on a common column:
  + *merged\_df = pd.merge(df1, df2, on='common\_column')*
* Write the DataFrame to a CSV file:
  + *df.to\_csv('output.csv', index=False)*

#### NUMPY

NumPy is a powerful open-source Python library for scientific computing. It provides tools for working with arrays, matrices, and numerical operations, making it a popular choice for data analysis, machine learning, and scientific research.

Here are some basic commands to get started with NumPy:

* Install NumPy using pip:
  + *pip install numpy*
* Import NumPy in your Python script:
  + *import numpy as np*
* Create a NumPy Array:
  + *arr = np.array([1, 2, 3, 4, 5])*
* Create a 2D NumPy Array:
  + *arr\_2d = np.array([[1, 2, 3], [4, 5, 6], [7, 8, 9]])*
* Perform element-wise operations on arrays:
  + *arr\_squared = arr \*\* 2*
* Perform Matrix Multiplication:
  + *mat\_product = np.dot(arr\_2d, arr)*
* Compute statistical measures on arrays:
  + *mean = np.mean(arr)*
  + *std\_dev = np.std(arr)*
* Reshape an array
  + *reshaped\_arr = arr.reshape((5, 1))*

#### FLASK

Flask is a micro web framework written in Python that allows developers to build web applications quickly and easily. It is lightweight and easy to use, making it a popular choice for small to medium-sized projects. Flask was designed to be simple and flexible, allowing developers to customize and extend it to suit their specific needs.

Some of the key features of Flask include:

* Routing: Flask provides a simple way to map URLs to Python functions, allowing developers to easily create a RESTful API.
* Templates: Flask supports a variety of templating engines, allowing developers to render dynamic HTML pages.
* HTTP request handling: Flask provides an easy way to handle HTTP requests and responses, including support for cookies, sessions, and file uploads**.**
* Extensions: Flask has a large number of third-party extensions available, allowing developers to add additional functionality to their applications.
* Debugging: Flask provides a built-in debugger that makes it easy to identify and fix errors in your code.

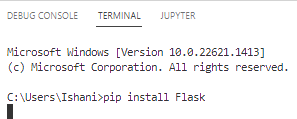
Overall, Flask is a powerful and versatile web framework that is easy to learn and use, making it a great choice for developers of all levels of experience.

**INSTALLING FLASK**

To install Flask, you can use pip, the package installer for Python. Open

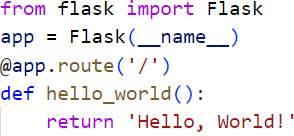
your

terminal or command prompt and type:



This will install Flask and any required dependencies.

Once Flask is installed, you can create a new Flask application using the following code:



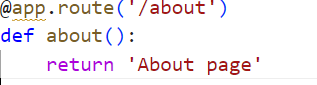
This code creates a new Flask application and defines a single route that responds with the string "Hello, World!" when accessed.

To run the application, save the code in a file (e.g., app.py) and execute the following command in your terminal or command prompt:

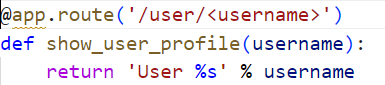


This will start a local web server at http://127.0.0.1:5000 (or another available port if 5000 is in use) and your application will be accessible at that URL.

Flask uses routes to map URLs to functions that handle requests. The @app.route decorator specifies the URL that the function should handle. For example, to handle requests to the URL /about, you can define a function like this:

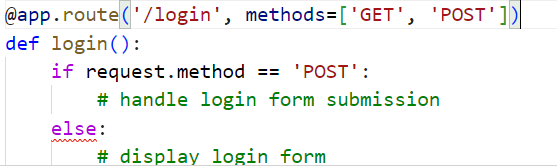


You can also specify URL parameters in routes using angle brackets (<parameter>). For example:



This route will match URLs like ‘/user/john’, ‘/user/sarah’, etc., and pass the value after ’user/’ to the ‘show\_user\_profile’ function as the ‘username‘ parameter.

Flask also supports HTTP methods like GET, POST, PUT, DELETE, and more. To specify the HTTP method that a function should handle, you can add the method name as an argument to the ‘@app.route’ decorator. For example:



This route will handle both GET and POST requests to the /login URL. If the request method is POST, the function will handle the form submission. Otherwise, it will display the login form. Flask has many more features and extensions available that allow you to create complex web applications.

**CHAPTER-4 SYSTEM DESIGN**

#### CHAPTER-4 SYSTEM DESIGN

* 1. **UML DIAGRAMS**

Unified Modeling Language (UML) is a general purpose modelling language. The main aim of UML is to define a standard way to visualize the way a system has been designed. It is quite similar to blueprints used in other fields of engineering.

#### USE CASE DIAGRAM:

Use case diagrams are considered for high level requirement analysis of a system.when the requirements of a system are analyzed the functionalities are captured in use cases. So, it can say that uses cases are nothing but thesystem functionalities written in an organized manner.

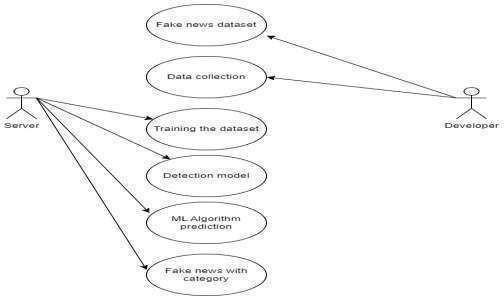


Figure 4.1.1 Use Case Diagram

#### ACTIVITY DIAGRAM:

A graphical representation of an executed set of procedural system activities and considered a state chart diagram variation. Activity diagrams describe parallel and conditional activities, use cases and system functions at a detailed level.

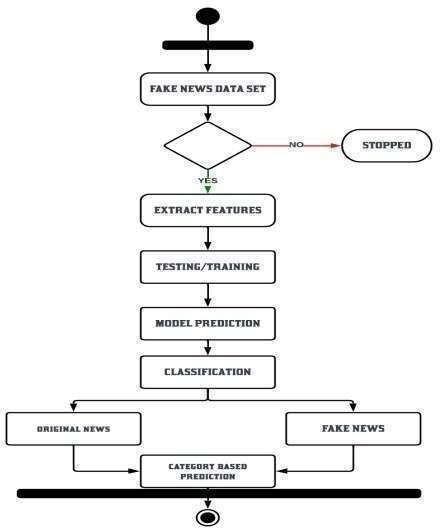


Figure 4.1.2 Activity Diagram

#### CLASS DIAGRAM:

Class diagram is basically a graphical representation of the static view of the system and represents different aspects of the application. The name of the class diagram should be meaningful to describe the aspect of the system. Each element and their relationships should be identified in advance.

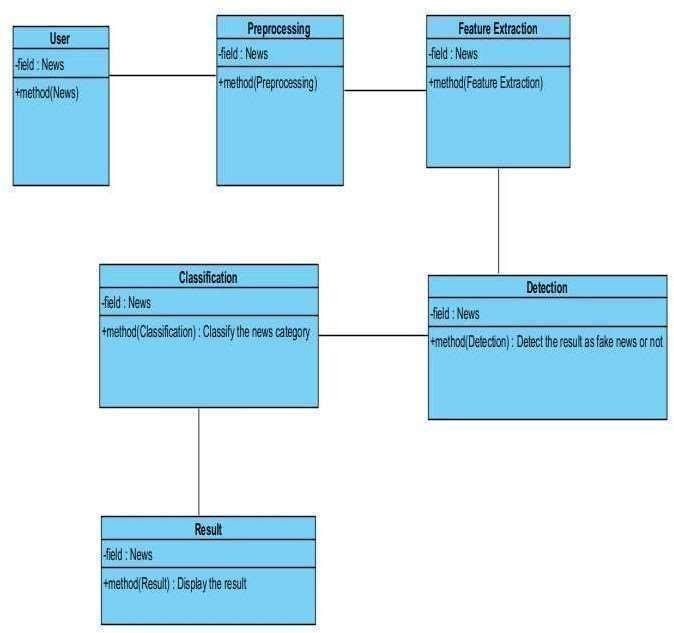


Figure 4.1.3 Class Diagram

#### Data Flow Diagram:

A data flow diagram (DFD) is a graphical representation of the "flow" of data through an information system, modeling its process aspects.It can be used for the visualization of data processing (structured design).Data flow diagrams are also known as bubble charts. DFD is a designing tool used in the top down approach to Systems Design. DFD levels are numbered 0, 1 or 2, and occasionally go to even Level 3 or beyond. DFD Level 0 is also called a Context Diagram.

Level 0:



Figure 4.2.1 Level 0 DFD Diagram

Level 1:

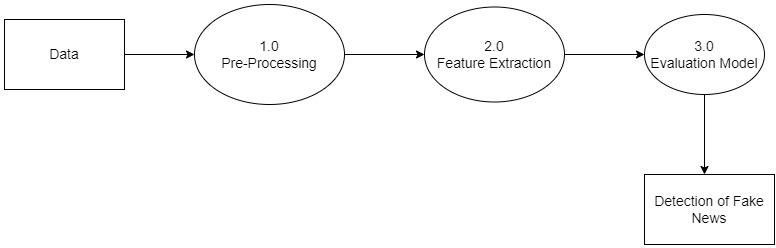


Figure 4.2.1 Level 1 DFD Diagram

## CHAPTER-5 SYSTEM ARCHITECTURE

#### CHAPTER-5 SYSTEM ARCHITECTURE

**5.1 SYSTEM ARCHITECTURE OVERVIEW**

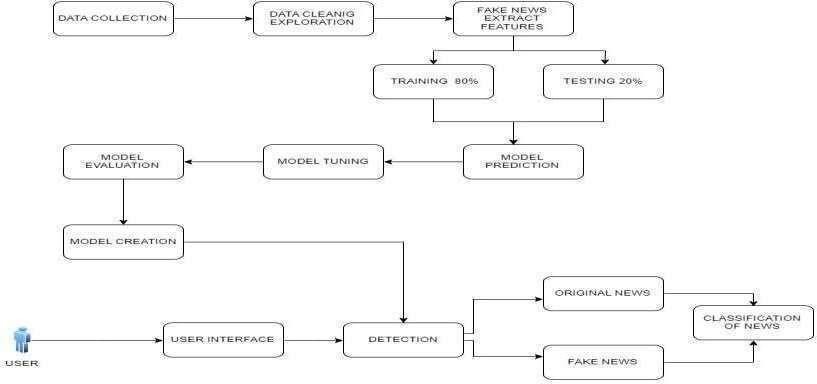


Figure 5.1 System Architecture

The architecture diagram shows the processes involved for building the project. It involves collecting dataset from website, the processing it to remove the noisy data, visualizing it and then implementing algorithms and finding the best model based on accuracy and then deploying it in the form of webpage.

## CHAPTER-6 SYSTEM IMPLEMENTATION

**CHAPTER-6 SYSTEM IMPLEMENTATION**

* 1. **ALGORITHMS:**

1.Logistic Regression 2.Random Forest

3.Additional Tree Discriminant 4.XG BOOST

#### LOGISTIC REGRESSION:

Logistic regression is a statistical technique used to model and analyze the relationship between a binary or categorical dependent variable and one or more independent variables. It is a type of regression analysis that is commonly used in fields such as economics, marketing, psychology, and medicine, among others.

In logistic regression, the dependent variable is modeled using a logistic function, which transforms the output of a linear equation into a value between 0 and 1, representing the probability of the dependent variable taking on a certain value. The independent variables are typically continuous, categorical, or binary.

The logistic regression model estimates the probability of the dependent variable given the values of the independent variables, and it can be used to predict the probability of a particular outcome or to identify the factors that influence the outcome.

Logistic regression is often used in machine learning for binary classification problems, where the goal is to predict whether a given input belongs to one of two classes. The model can also be extended to handle multi-class classification problems using techniques such as one-vs-all or softmax regression.

The formula for logistic regression can be expressed as follows:

\begin{equation}

P(Y=1|X) = \frac{1}{1 + e^{-\beta X}}

\end{equation} where:

P(Y=1|X) is the probability of the output variable Y being 1 given the input variable X

X is the input vector with n features

$\beta$ is the weight vector with n parameters

#### RANDOM FOREST:

Random forest is a popular ensemble machine learning algorithm that is used for both classification and regression tasks. It is an extension of the decision tree algorithm, and it is known for its high accuracy, robustness, and ability to handlelarge datasets.

Random forest works by creating multiple decision trees and combining their predictions to make the final prediction. Each decision tree in the forest is built using a random subset of the training data and a random subset of the input features. This helps to reduce overfitting and increase the diversity of the trees in the forest.

To make a prediction, the random forest algorithm passes the input data through each of the decision trees in the forest, and each tree provides a prediction. The final prediction is then made by combining the predictions of all the decision trees. For classification tasks, the final prediction is often the mode of the predictions of all the trees, while for regression tasks, it is the mean of the predictions.

Random forest is a powerful algorithm that can handle a wide range of input data types and is relatively easy to use. It can also provide information about the

importance of each input feature in the prediction, which can be useful for feature selection and understanding the underlying relationships in the data. However, it can be computationally expensive and may require tuning of the hyperparameters to achieve optimal performance.

The formula for the prediction of a random forest model can be represented as: For classification tasks:

y\_pred = mode(tree\_1.predict(X), tree\_2.predict(X), ..., tree\_n.predict(X)) where:

y\_pred is the predicted target variable

mode() returns the most frequent class among the predictions of the trees (for classification tasks)

#### ADDITIONAL TREE DISCRIMINANT:

Decision Trees are supervised machine learning algorithms where data is segmented according to a specific parameter. The objective of this algorithm is to build a training model that is used to predict the class of the target variable, which is used by the decision tree to solve the classification problem.Discriminant analysis creates a predictive model that determines to which group the class belongs. The model is composed of a discriminant function based on linear combinations of the variables used as predictors, i.e., offering the best discrimination between the groups.

#### XG BOOST:

XGBoost (short for eXtreme Gradient Boosting) is a popular machine learning algorithm that is used for supervised learning tasks such as classification, regression, and ranking. It is an ensemble learning algorithm that combines the predictions of multiple decision tree models to produce a final prediction.

XGBoost is a form of gradient boosting, which means that it uses an iterative approach to building decision trees. It starts with a simple decision tree model and then adds more trees to the model, with each tree attempting to correct the

errors of the previous trees. The algorithm uses a gradient descent optimization method to minimize a loss function that measures the difference between the predicted and actual values.

One of the key features of XGBoost is that it includes a regularization term in the loss function, which helps to prevent overfitting and improve the generalization of the model. It also includes other techniques such as weighted quantile sketch, sparsity-aware split finding, and cache-aware computation to improve performance and reduce memory usage.

XGBoost has become a popular algorithm for a variety of machine learning tasks because it is highly accurate, fast, and scalable. It has won several machine learning competitions and is used in a variety of industries such as finance, healthcare, and e-commerce. However, it can be computationally expensive and may require tuning of the hyperparameters to achieve optimal performance.

The XGBoost algorithm can be written as follows:

y\_hat = sum\_i^n f\_i(x)

where y\_hat is the predicted value, n is the total number of trees, and f\_i(x) is the prediction of the i-th tree.

The prediction of each tree is given by:

f\_i(x) = w\_q(x; \theta\_i)

where w\_q is the weighted sum of the leaf values of the i-th tree, and \theta\_i is the set of parameters that define the i-th tree.

#### MODULE DESIGN SPECIFICATION

1. Data collection
2. Data preprocessing
3. Feature extraction
4. Modeling Creation with Random Forest
5. Hyper parameter Tuning

#### MODULE DESCRIPTION

* + 1. **DATA COLLECTION:**

Data is the prime ingredient of this project, as these data features are extracted using Natural Language Processing. By using these features of the data, Machine Learning Algorithms are trained and models are created. In this proposal, we have news with equal proportionality of fake and real. Data is saved in Comma Separated Value format. This data set is divided in the training and testing of algorithms.

#### DATA PREPROCESSING:

The six-label classification problem was translated into a binary classification problem with True and False labels for the proposed scheme. In addition, only the news headline was used as an input for classification. Thus, in the preprocessing stage, the labels were first mapped using the above-mentioned mapping, after which only the labels and news statement columns were extracted from the dataset and saved in .csv format for future use. Following the preprocessing, we were able to obtain the following three cleaned files: • train.csv

#### FEATURE EXTRACTION:

The following feature extraction method is used to help machine learning models gain insights from news headlines: Count Vectorizer: First, the English stopwords were stripped from all of the news headlines using scikitlearn's Count Vectorizer, and then they were tokenized using spaces and punctuation marks as the delimiter. After all of the headlines had been tokenized, a sparse matrix of all of the news headlines as rows and the tokens as columns was restored. In addition to their morphological use, a number of n-grams were returned to make the tokens reflect the sense in which they were used.

#### MODEL CREATION WITH RANDOM FOREST:

Logistic Regression, Random Forest Classifier, Naive Bayes, SVM Classifier, and voting classifier were the models used for training. The features extracted from the Count Victories are used to train the models. After that, using Grid Search CV and a 5-hold out cross validation set, all of the models were hyper parameter tuned for all of the different possible parameters. The aim of this hyper parameter tuning was to boost the models' f1-score. After the models were fine- tuned, they were evaluated on a test range, and evaluation metrics for the models were determined.

#### HYPER PARAMETER TUNNING:

Parameters that define the model architecture are referred to as hyperparameters and thus the process of searching for the ideal parameter is referred to as hyper parameter tuning. We have used Grid Search CV to tune the parameter of each algorithm. The grid of values of each parameter is given as input and Grid Search CV will methodically build and evaluate a model for each combination of algorithm parameters specified in a grid. The model with the best parameter value is given as output.

# CHAPTER-7 PERFORMANCE EVALUATION

#### CHAPTER-7 PERFORMANCE EVALUATION

* 1. **RESULTS AND DISCUSSIONS:**



Figure 7.1.1 Screenshot of webpage for Fake news Prediction This above screenshot shows the main page where the news as a text is given as input and the type of news is detected



Figure 7.1.2 Screenshot of News entered in a text box for detection The above screenshot shows the way of entering news into the textbox

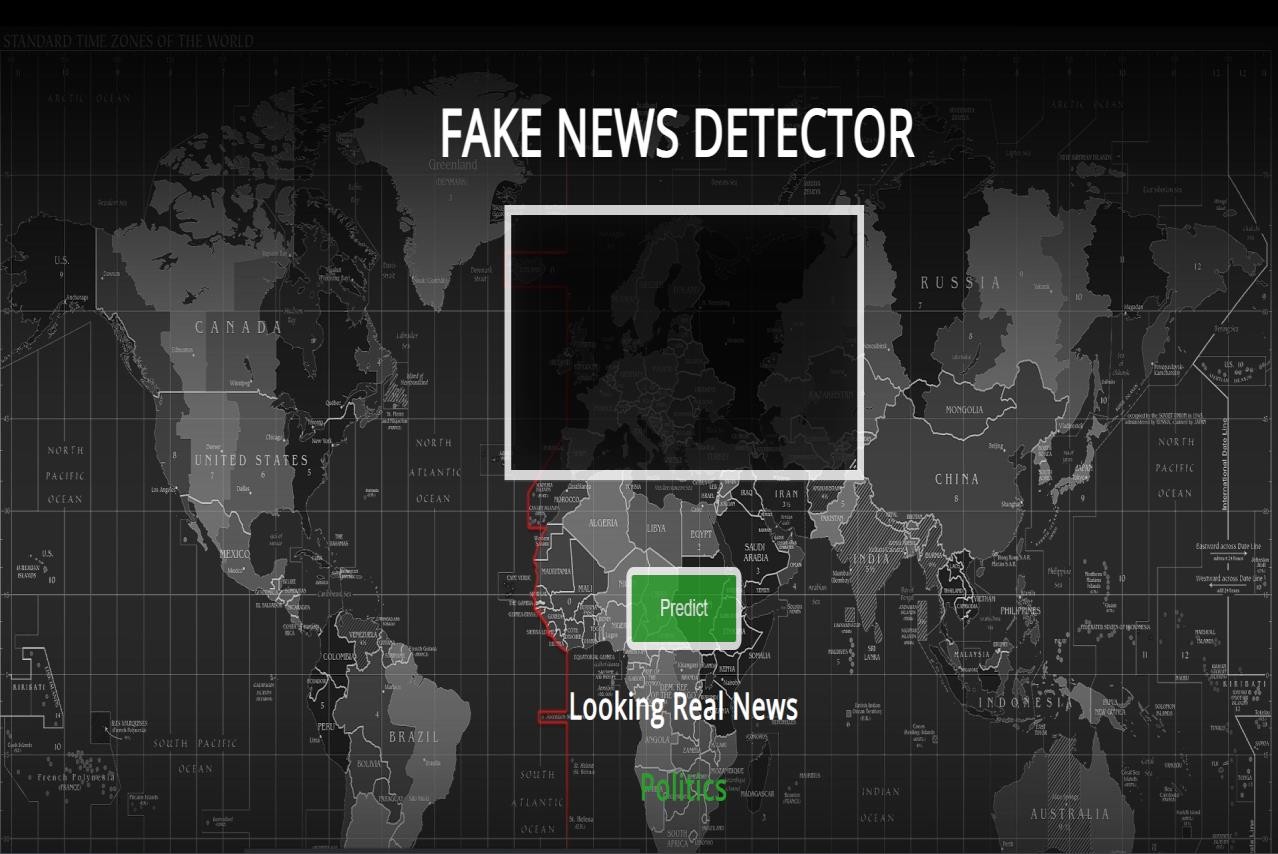


Figure 7.1.3 Screenshot of Predicted result

The above screenshot shows the predicted output of the given news in the text box

#### COMPARATIVE ANALYSIS:

Figure 7.2.1 Screenshot of accuracy of Passive Aggressive classifier The above diagram shows the accuracy , precision , recall, f1-score obtained

by training the model using Passive aggressive classifier

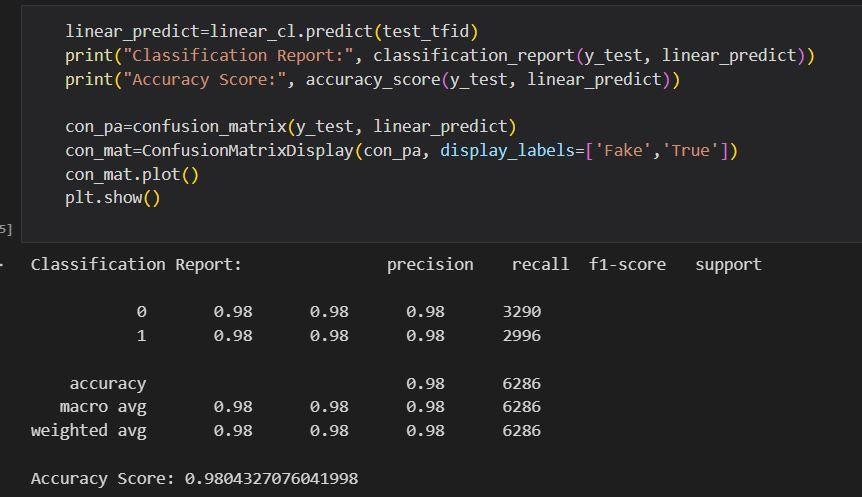


Figure 7.2.2 Screenshot of accuracy of Logistic Regression.

The above diagram shows the accuracy , precision , recall, f1-score obtained by training the model using Logistic regression

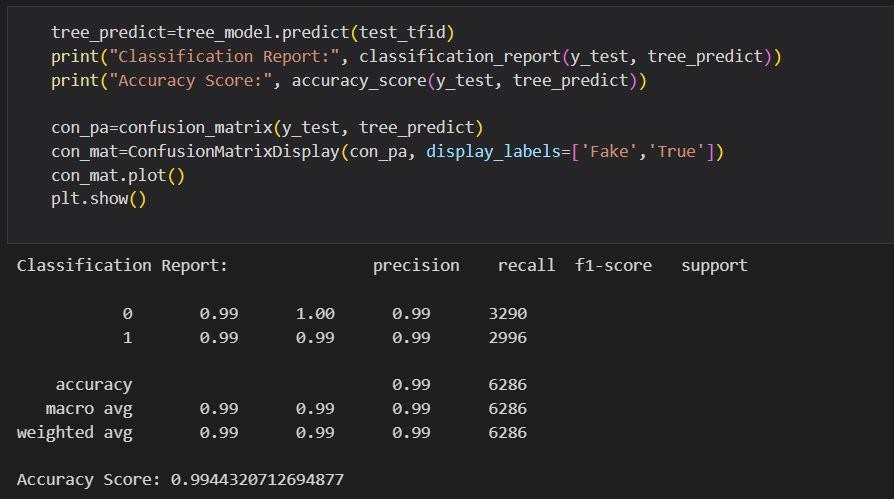


Figure 7.2.3 Screenshot of accuracy of Decision Tree

The above diagram shows the accuracy , precision , recall, f1-score obtained by training the model using Decision Tree

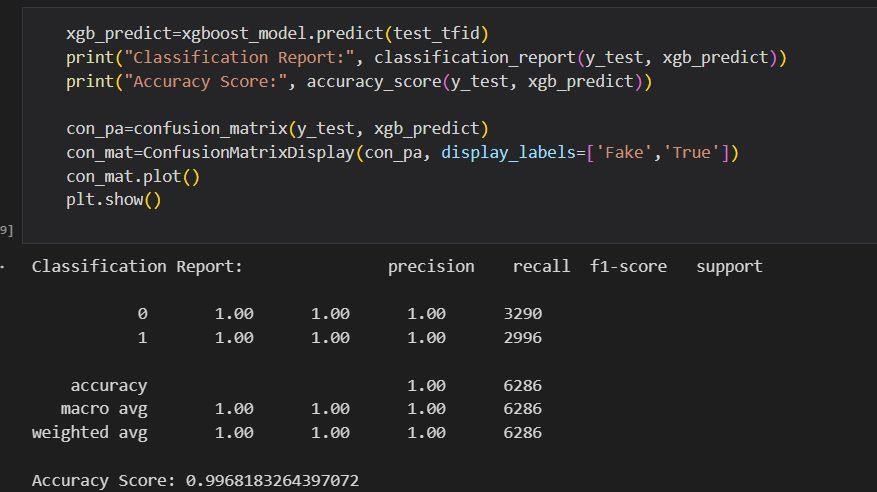


Figure 7.2.4 Screenshot of accuracy of XGBoost

The above diagram shows the accuracy , precision , recall, f1-score obtained by training the model using XGBoost

#### COMPARISON OF ALGORITHM

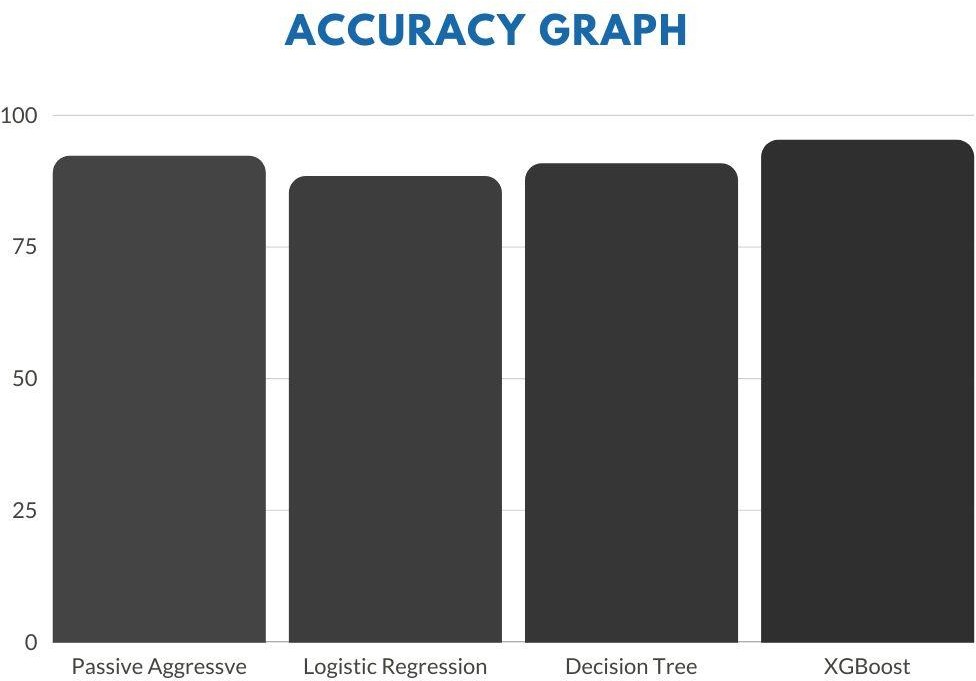


Figure 7.2.5 Comparison of Algorithms

# CHAPTER-8 CONCLUSION

#### CONCLUSION

In this project, we have proposed a novel hybrid fake news detection system that employs two types of features: linguistic and fact-verification features. The proposed detection system employs only eight features, which less compared to the stat-of-the-art approaches. It operates in two phases: training and testing. In the training phase, the detection system runs four machine learning algorithms, i.e., Logistic Regression (LR), Random Forest (RF), Additional Trees Discriminant, and XGBoost, in order to select the best classifier for the testing phase. Evaluation results on the News data set show that the proposed detection system reaches an accuracy of 99% under XGBoost. As future work, we aim at improving the accuracy of our detection system by investigating other discriminating features such as visual-based and style-based features.Moreover, we plan to further detect other types of false information such as biased/inaccurate news and misleading/ambiguous news.

#### FUTURE ENHANCEMENT

The project's future scope is that fake news detectors can help to filter different websites that contain fake news, and the goal is to help users avoid being drawn in by misleading headlines. With some modifications, the project can also be used on many social media platforms where there is a massive amount of fake data that can cause harm to society. To avoid detection, fake account creators are constantly changing their tactics. Future research could concentrate on developing machine learning models that can adapt to these changing tactics while still detecting fake accounts.

# APPENDICES

#### APPENDICES

1. **SAMPLE DATASET:**

|  |  |
| --- | --- |
| **TITLE** | **LABEL** |
| Donald Trump Sends Out Embarrassing  New Year Message; This is Disturbing | Fake |
| Drunk Bragging Trump Staffer Started  Russian Collusion Investigation | Fake |
| Sheriff David Clarke Becomes An Internet Joke For Threatening To Poke People In The  Eye | Fake |
| Trump Is So Obsessed He Even Has Obama  Name Coded Into His Website (IMAGES) | Fake |
| Pope Francis Just Called Out Donald Trump  During His Christmas Speech | Fake |
| Racist Alabama Cops Brutalize Black Boy  While He Is In Handcuffs (GRAPHIC IMAGES) | Fake |
| Fresh Off The Golf Course, Trump Lashes  Out At FBI Deputy Director And James | Fake |
| Trump Said Some INSANELY Racist Stuff Inside The Oval Office, And Witnesses Back  It Up | Fake |
| Former CIA Director Slams Trump Over UN  Bullying, Openly Suggests He Acting Like A Dictator (TWEET) | Fake |
| WATCH: Brand-New Pro-Trump Ad  Features So Much Kissing It Will Make You Sick | Fake |
| Papa John Founder Retires, Figures Out  Racism Is Bad For Business | Fake |
| WATCH: Paul Ryan Just Told Us He Does not Care About Struggling Families Living In  Blue States | Fake |
| Bad News For Trump Mitch McConnell  Says No To Repealing Obamacare In 2018 | Fake |
| WATCH: Lindsey Graham Trashes Media  For Portraying Trump As he Forgets His | Fake |

|  |  |
| --- | --- |
| Own Words |  |
| Heiress To Disney Empire Knows GOP  Scammed Us SHREDS Them For Tax Bill | Fake |
| Tone Deaf Trump: Congrats Rep. Scalise On  Losing Weight After You Almost Died | Fake |
| The Internet Brutally Mocks Disney New  Trump Robot At Hall Of Presidents | Fake |
| Mueller Spokesman Just F-cked Up Donald  Trumps Christmas | Fake |
| SNL Hilariously Mocks Accused Child  Molester Roy Moore For Losing AL Senate Race (VIDEO) | Fake |
| Republican Senator Gets Dragged For Going  After Robert Mueller | Fake |
| In A Heartless Rebuke To Victims, Trump  Invites NRA To Xmas Party On Sandy Hook Anniversary | Fake |
| KY GOP State Rep. Commits Suicide Over  Allegations He Molested A Teen Girl (DETAILS) | Fake |
| Meghan McCain Tweets The Most  AMAZING Response To Doug Jones Win In Deep-Red Alabama | Fake |
| CNN CALLS IT: A Democrat Will  Represent Alabama In The Senate For The First Time In 25 Years | Fake |
| As U.S. budget fight looms, Republicans flip  their fiscal script | True |
| U.S. military to accept transgender recruits  on Monday: Pentagon | True |
| Senior U.S. Republican senator: 'Let Mr.  Mueller do his job' | True |
| FBI Russia probe helped by Australian  diplomat tip-off: NYT | True |
| Trump wants Postal Service to charge 'much  more' for Amazon shipments | True |
| White House, Congress prepare for talks on  spending, immigration | True |

|  |  |
| --- | --- |
| Trump says Russia probe will be fair, but  timeline unclear: NYT | True |
| Factbox: Trump on Twitter (Dec 29) -  Approval rating, Amazon | True |
| Trump on Twitter (Dec 28) - Global  Warming | True |
| Alabama official to certify Senator-elect  Jones today despite challenge: CNN | True |
| Jones certified U.S. Senate winner despite  Moore challenge | True |
| New York governor questions the  constitutionality of federal tax overhaul | True |
| Factbox: Trump on Twitter (Dec 28) - Vanity  Fair, Hillary Clinton | True |
| Trump on Twitter (Dec 27) - Trump, Iraq,  Syria | True |
| Man says he delivered manure to Mnuchin to  protest new U.S. tax law | True |
| Virginia officials postpone lottery drawing to  decide tied statehouse election | True |
| U.S. lawmakers question businessman at  2016 Trump Tower meeting: sources | True |
| Trump on Twitter (Dec 26) - Hillary Clinton,  Tax Cut Bill | True |
| U.S. appeals court rejects challenge to Trump  voter fraud panel | True |
| Treasury Secretary Mnuchin was sent gift-  wrapped box of horse manure: reports | True |
| Federal judge partially lifts Trump's latest  refugee restrictions | True |
| Exclusive: U.S. memo weakens guidelines  for protecting immigrant children in court | True |
| Trump travel ban should not apply to people  with strong U.S. ties: court | True |
| Second court rejects Trump bid to stop  transgender military recruits | True |

Table 1. Sample Dataset

1. **SAMPLE CODING:**

**SERVER SIDE CODING:**

from flask import Flask, render\_template, request

from sklearn.feature\_extraction.text import TfidfVectorizer from sklearn.linear\_model import PassiveAggressiveClassifier import pickle

import pandas as pd

from sklearn.model\_selection import train\_test\_split app = Flask( name )

tfvect = TfidfVectorizer(stop\_words='english', max\_df=0.7)loaded\_model = pickle.load(open('model.pkl', 'rb')) dataframe = pd.read\_csv('news.csv',low\_memory=False) dataframe = dataframe.fillna(' ')

x = dataframe['text'] y = dataframe['label']

x\_train, x\_test, y\_train, y\_test = train\_test\_split(x, y, test\_size=0.2, random\_state=0)

def fake\_news\_det(news):

tfid\_x\_train = tfvect.fit\_transform(x\_train) tfid\_x\_test = tfvect.transform(x\_test) input\_data = [news]

vectorized\_input\_data = tfvect.transform(input\_data) prediction = loaded\_model.predict(vectorized\_input\_data) return prediction

@app.route('/') def home():

return render\_template('index.html')

@app.route('/predict', methods=['POST']) def predict():

if request.method == 'POST': message = request.form['message']

pred = fake\_news\_det(message) print(pred)

politics="obama trump donald political politician law government governance diplomatic polity public administration diplomatical political opportunism partisanship election republic politics aristotle politically politicalscience localgovernment partisan politics economics state debate ideologies monarchy politic democratic religion politicking policy nationalism smooth wars political system activism politicians suave politicos morality politicizing regionalism governor judiciary policymaking political expediency expedient geopolitics parliament history society police republicanism policies sagacious demagoguery elections voting electoral populist ban conspiracy polis principle politics sociopolitical social mudslinging politicdrama divisive divisiveness factionalism ideological statecraft journalism parochialism cynicism gubernatorial ethics worldpopulism regime economy power partisan bickering pandering rhetoric bureaucracy intrigue federal government personal democracy identity tribes controversial civil legal unitednations lawmaking allegiance justice celibacy governmental civilised sinfulness stuffy meddles unclothed crystallizing psyche policy-making anarchy engrained latinisation pols humorless clubby metabolizes political parties politicize multinational realpolitik crist demagogue supremacy nations politick statesmanship conservatism pragmatism electioneering electorate idealism dukes sovereign state slinging tribalism partisan curfew government indiangovernmrnt donald trump obama narendra modi Modi minister exminister parliment council barack department"

weather="Weather weather rain cloud rainbow temperature pressure overcast shower sunrise dry tornado sunset humidity cold heat wind cloudy heat wave fog breeze humid lightning blustery humidity thunder snow heat index thunderstorm downpour drought tropical water cycle temperate moisture drizzle warm hail icicle climate storm flood muggy gale flash flood atmosphere cold front mist isobar cold snap condensation forecast ice storm freeze barometric gust snowfall rainfall raining sunny monsoon"

sports="aerobics archer archery arena arrow athlete athletics axel badminton ball base baseball basketball bat baton batter batting biathlon bicycle bicycling bike biking billiards bobsleigh bocce boomerang boules bow bowler bowling boxer boxing bronze medal bunt cricket shuttle football goal billiards wicket bowler boldout catchout losethematch cricket"

politics=politics.split() message=message.split() weather=weather.split() sports=sports.split() politics=tuple(politics) message=tuple(message) weather=tuple(weather) sports=tuple(sports)

a=" "

print(len(message))

for x in range(len(message)): if(message[x] in politics):

a="Politics" elif(message[x] in weather):

a="weather" elif(message[x] in sports):

a="sports" #else:

#a="Others" if a ==" ":

a="Others"

return render\_template('index.html', prediction=pred,newstype=a) else:

return render\_template('index.html', prediction="Something went wrong") if name == ' main ':

app.run(host='0.0.0.0',port='400',debug=True)

**CLIENT SIDE CODING:**

<!DOCTYPE html>

<html >

<head>

<meta charset="UTF-8">

<title>Fake News📰 Detection System</title>

<link href='https://fonts.googleapis.com/css?family=Pacifico' rel='stylesheet' type='text/css'>

<link href='https://fonts.googleapis.com/css?family=Arimo' rel='stylesheet' type='text/css'>

<link href='https://fonts.googleapis.com/css?family=Hind:300' rel='stylesheet' type='text/css'>

<link href='https://fonts.googleapis.com/css?family=Open+Sans+Condensed:300' rel='stylesheet' type='text/css'>

<link rel="stylesheet" href="{{ url\_for('static', filename='style.css') }}">

</head>

<body style="color:black">

<br>

<h1 style="color:rgb(255, 255, 255);font-size:48px;margin-left:42px;";>&nbsp FAKE NEWS DETECTOR</h1><br><br>

<div class="login" style="color:black">

<br>

<form action="{{ url\_for('predict')}}" method="POST">

<br>

<center>

<textarea name="message" rows="6" cols="50" required="required" style="font-size: 18pt"></textarea>

</center>

<br> </br>

<button type="submit" class="btn">Predict</button>

<div class="results">

{% if prediction == ['FAKE'] %}

<h2 style="color:rgb(255, 36, 36);">Looking Spam⚠️News </h2>

{% elif prediction == ['REAL'] %}

<h2 style="color:rgb(255, 255, 255);"><b>Looking Real News</b></h2>

{% endif %}

h2 style="color:#29a02b;"><b></b>{% print(newstype) %}</b></h2>

<br><br>

</div>

</form>

</div>

@import url(https://fonts.googleapis.com/css?family=Open+Sans);

.btn{

background-color: #29a02b; /\* Green \*/ border: 6px solid #ffffff;

color: rgb(255, 255, 255); padding: 15px 32px;

text-align: center; text-decoration: none; display: inline-block; font-size: 17px; border-radius: 8px; opacity: 0.8;

}

\* { -webkit-box-sizing:border-box; -moz-box-sizing:border-box; -ms-box- sizing:border-box; -o-box-sizing:border-box; box-sizing:border-box; }

html { width: 100%; height:100%; overflow:hidden; }

body {

background: url("\static\1.jpg"); width: 100%;

height:100%;

font-family: 'Open Sans', sans-serif; background-image: url("/static/1.jpg"); color: #fff;

font-size: 18px; text-align:center;

letter-spacing:1.2px;

background-image: url("/static/1.jpg"); background-size: cover;

}

.login {

position: absolute; top: 40%;

left: 50%;

margin: -150px 0 0 -150px; width:400px; height:400px;

}

.login h1 { color: #fff; text-shadow: 0 0 10px rgba(0,0,0,0.3); letter-spacing:1px; text-align:center; }

textarea {

width: 100%;

margin-bottom: 10px; background: rgba(0, 0, 0, 0.942); border: 8px solid #ffffff; padding: 10px;

font-size: 25px; color: #38da3b;

box-shadow: inset 0 -5px 45px rgba(100,100,100,0.2), 0 1px 1px

rgba(255,255,255,0.2);

-webkit-transition: box-shadow .5s ease;

-moz-transition: box-shadow .5s ease;

-o-transition: box-shadow .5s ease;

-ms-transition: box-shadow .5s ease; transition: box-shadow .5s ease; opacity: 0.8;

}

input:focus { box-shadow: inset 0 -5px 45px rgba(100,100,100,0.4), 0 1px 1px rgba(255,255,255,0.2); }

**Output Screenshot:**









**REFERENCES**

1. N. Seddari, A. Derhab, M. Belaoued, W. Halboob, J. Al-Muhtadi and A. Bouras, "A Hybrid Linguistic and Knowledge-Based Analysis Approach for Fake News Detection on Social Media," in *IEEE Access*, vol. 10, pp. 62097-62109, 2022, doi: 10.1109/ACCESS.2022.3181184.
2. Yildirim, G. A novel hybrid multi-thread metaheuristic approach for fake news detection in social media. *Appl Intell* (2022). <https://doi.org/10.1007/s10489-022-03972-9>
3. Thompson, R.C.; Joseph, S.; Adeliyi, T.T. A Systematic Literature Review and Meta-Analysis of Studies on Online Fake News Detection. *Information* 2022, *13*, 527. https://doi.org/10.3390/info13110527
4. Jarrahi, A., Safari, L. Evaluating the effectiveness of publishers’ features in fake news detection on social media. *Multimed Tools Appl* 82, 2913–2939 (2023). <https://doi.org/10.1007/s11042-022-12668-8>
5. Ihsan Ali, Mohamad Nizam Bin Ayub, Palaiahnakote Shivakumara, Nurul Fazmidar Binti Mohd Noor, "Fake News Detection Techniques on Social Media: A Survey", *Wireless Communications and Mobile Computing*, vol. 2022, Article ID 6072084, 17 pages, 2022. https://doi.org/10.1155/2022/6072084
6. Raturi S., Mishra A. and Maji S. (2021). Fake News Detection in Social Networks using Machine Learning: A Review. In *Proceedings of the 3rd International Conference on Advanced Computing and Software Engineering - Volume 1: ICACSE,* ISBN 978-989-758-544-9, pages 177-181. DOI: 10.5220/0010564800003161
7. Pandey, S., Prabhakaran, S., Reddy, N. V. S., & Acharya, D. (2022). Fake News Detection from Online media using Machine learning Classifiers. *Journal of Physics: Conference Series*, *2161*(1), [012027]. https://doi.org/10.1088/1742- 6596/2161/1/012027
8. Khanam, Z & Alwasel, B & Sirafi, H & Rashid, Mamoon. (2021). Fake

News Detection Using Machine Learning Approaches. IOP Conference Series: Materials Science and Engineering. 1099. 012040. 10.1088/1757- 899X/1099/1/012040.

1. Sharma, Uma & Saran, Sidarth & Patil, Shankar. (2020). Fake News Detection Using Machine Learning Algorithms. 2320-2882.
2. A. Jain, A. Shakya, H. Khatter and A. K. Gupta, "A smart System for Fake News Detection Using Machine Learning," *2019 International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT)*, Ghaziabad, India, 2019, pp. 1-4, doi: 10.1109/ICICT46931.2019.8977659.