

**DEHRADUN CAMPUS**

**TERM WORK**

**MINI PROJECT**

**B. TECH (CSE) IV**

**2022-23**

**SUBMITTED TO: SUBMITTED BY:**

DEPARTMENT OF COMPUTER Name: Dikshant Negi

SCIENCE & ENGENIEERING University Roll No: 2118521

GRAPHIC ERA HILL UNIVERSITY Class Roll number: 27

Section: D



**DEHRADUN CAMPUS**

THIS IS TO CERTIFY THAT **DIKSHANT NEGI** HAS SATISFACTORILY COMPLETED ALL THE TERMWORK IN THE LABORATORY OF THIS COLLEGE. THE COURSE OF THE TERM WORK **MINI PROJECT** IN PARTIAL FULLFILLMENT OF THE REQUIREMENT IN **4th SEMESTER** OF B. TECH (CSE) DEGREE COURSE PRESCRIBED BY GRAPHIC ERA HILL UNIVERSITY, DEHRADUN DURING THE YEAR 2022 -23.

**CONCERNED FACULTY:** HEAD OF DEPARTMENT

MS. Rishika Yadav

NAME OF EXAMINER:

SIGNATURE OF EXAMINER:



**DEHRADUN CAMPUS**

ACKNOWLEDGEMENT

I WOULD LIKE TO TAKE THIS OPPORTUNITY TO EXPRESS MY SINCERE AND PROFOUND GRATITUDE TO MY MENTOR **MS. RISHIKA YADAV** FOR HIS CONSTANT ENCOURAGEMENT AND GUIDANCE THROUGH OUT THE COURSE OF MY PROJECT

AND FINALLY, THANKS TO ALL THOSE WHO INSPIRED ME DIRECTLY OR INDIRECTLY.

**Name:Dikshant Negi**

**University Roll No: 2118521**

**Student ID: 21011671**

**B.Tech. CSE, IVth SEM, Section ‘D’**

**Session: 2022-2023 GEHU, Dehradun**

WHAT IS FAKE NEWS ?

Fake news refers to false or misleading information presented as factual news. It encompasses deliberate disinformation, misinformation, or propaganda designed to deceive readers or viewers. Fake news often spreads rapidly through social media platforms and can have serious consequences, including shaping public opinion, influencing elections, or inciting fear and panic. It can be created by individuals or organizations with various motivations, such as political gain, financial profit, or simply to generate chaos. In an era of information overload, critical thinking, fact-checking, and media literacy are crucial in combating the proliferation of fake news and ensuring the dissemination of accurate and reliable information.

WHY DO WE NEED FAKE NEWS DETECTION

We need fake news detection for several reasons.

First, fake news has the potential to mislead and manipulate people, affecting their decisions and beliefs. It can undermine the integrity of democratic processes, public discourse, and trust in the media. Fake news detection helps in preserving the reliability and credibility of information sources.

Second, it promotes media literacy and critical thinking skills by encouraging individuals to question and verify the information they consume.

Third, it assists in preventing the spread of harmful or dangerous misinformation, particularly in sensitive areas such as health, science, and politics. Ultimately, fake news detection contributes to the maintenance of an informed and well-functioning society.

TECHNIQUES TO APPLY

FAKE NEWS DETECTION

1-Natural Language Processing (NLP): NLP techniques play a crucial role in fake news detection. Approaches like sentiment analysis, topic modeling, and named entity recognition can be used to analyze the content of news articles and identify suspicious patterns, biased language, or misleading information.

2-Machine Learning (ML) Algorithms: ML algorithms, such as decision trees, random forests, and support vector machines, can be trained on labeled datasets to classify news articles as either real or fake. These models learn patterns and features from the data to make predictions on new articles, enabling automated fake news detection.

3-Deep Learning: Deep learning models, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have shown promise in detecting fake news. These models can process textual and visual information, identifying patterns and hidden features that may indicate the presence of misinformation or fabricated content.

4-Social Network Analysis (SNA): SNA focuses on the structural analysis of social networks to identify patterns of information diffusion. By analyzing the spread of news articles and the credibility of the sources sharing them, SNA can help identify potential sources of fake news and uncover networks of disinformation.

Fact-Checking and External Sources: Fact-checking initiatives and external sources can be used as references to verify the accuracy of news articles. By comparing the claims made in an article with reliable sources, fact-checking can help identify inconsistencies or falsehoods, aiding in fake news detection.

5-Metadata and Source Analysis: Analyzing metadata, such as publication dates, author credibility, and website reputation, can provide valuable insights into the trustworthiness of news articles. Source analysis involves assessing the reliability and bias of the publishing platform or author, which can contribute to identifying fake news.

6-Cross-Referencing and Verification: Cross-referencing information across multiple sources and conducting thorough verification can help identify discrepancies or inconsistencies that may indicate the presence of fake news. This approach involves consulting diverse and reputable sources to ensure the accuracy of the news content.

7-User Feedback and Community Reporting: Incorporating user feedback and community reporting mechanisms can help in detecting fake news. Users can report suspicious articles or provide feedback on potentially misleading information, which can aid in identifying and flagging fake news content.

8-Hybrid Approaches: Combining multiple techniques and approaches, such as NLP, ML algorithms, fact-checking, and user feedback, can enhance the accuracy and reliability of fake news detection systems. Hybrid approaches leverage the strengths of different methods to provide more robust and comprehensive detection capabilities.

9-Continuous Learning and Adaptive Models: Fake news detection models can be designed to adapt and improve over time by continuously learning from new data. This allows the system to stay updated with evolving disinformation tactics and maintain effectiveness in identifying fake news.

LANGUAGE AND LIBRAIRES USED

Python - Python is a high-level programming language known for its simplicity, readability, and versatility. It was created by Guido van Rossum and released in 1991. Key features of Python include its easy-to-read syntax, extensive standard library, and vast ecosystem of third-party libraries and frameworks. Python is cross-platform compatible and widely used in web development, data science, machine learning, scripting, and automation. It offers powerful tools for data manipulation, analysis, visualization, and model development. Python's active community provides extensive support, resources, and documentation. Its simplicity and efficiency make it ideal for prototyping and rapid development. Overall, Python's popularity continues to grow due to its ease of use, flexibility, and wide range of applications.

Pandas: Pandas is a powerful data manipulation and analysis library in Python. It provides data structures and functions to efficiently handle and manipulate structured data, such as tables or spreadsheets. Pandas allows for tasks like data cleaning, filtering, grouping, merging, and aggregation. It is widely used in data analysis, data preprocessing, and data exploration tasks.

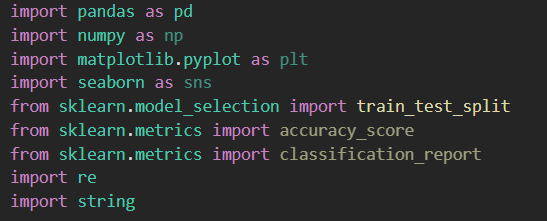
Scikit-learn (sk-learn): Scikit-learn is a popular machine learning library in Python. It provides a wide range of algorithms and tools for tasks like classification, regression, clustering, and dimensionality reduction. Sklearn offers consistent APIs, making it easy to use and integrate into machine learning workflows. It also includes utilities for data preprocessing, model evaluation, and feature selection.

re (Regular Expressions): The "re" library in Python provides support for regular expressions, which are powerful pattern matching and manipulation tools. Regular expressions allow you to search, match, and manipulate text based on specific patterns. The "re" library enables tasks such as text parsing, data extraction, string matching, and text manipulation.

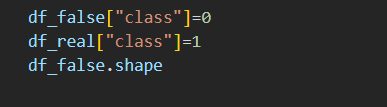
CODE SUMMARY

🡪Data Set- The datasets for fake and true news where taken from Kaggle which are already preprocessed.

🡪Libraries- First we import all these librarie’s



🡪Then we add a new column of class in it as shown –



The df\_false[“class”] =0 which shows that news is fake

And the df\_true[“class”] =1 which shows that news is real

🡪 In this step we take some news from real and some news from fake for manual testing and then dropping those news from the data set.

A computer screen shot of a code

Description automatically generated

🡪 In this step we merge real and fake news together



* Then we drop those column which are not required to make our modle more fast

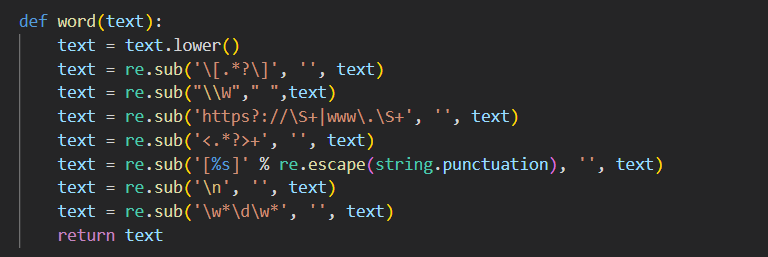


* In this step we suffle our data set so that real and fake news can be scrambled properly.

A black background with white text and numbers

Description automatically generated

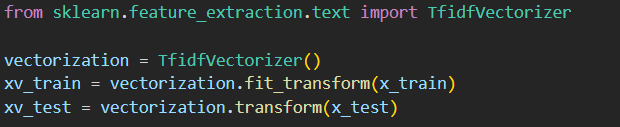
🡪Then write a function for clearing the text



* Now use sklearn library for train and test split

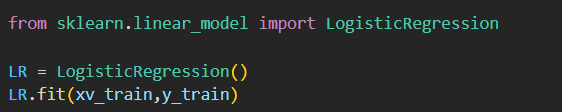


🡪Now use TF-IDF Vectorizer . It stands for Term Frequency-Inverse Document Frequency Vectorizer. TF-IDF Vectorizer converts text documents into numerical feature vectors based on the term frequency and inverse document frequency

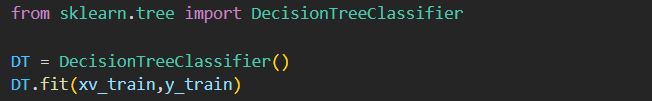


* I have used two models to train my model

1-Logistic regression-Logistic regression is a statistical model used for binary classification problems. It is a supervised learning algorithm that predicts the probability of an instance belonging to a certain class



2- Decision Tree-A decision tree is a supervised machine learning algorithm used for both classification and regression tasks. It creates a tree-like model of decisions and their possible consequences based on the input features.



🡪Lastly we check our prediction score

1-Logistic regression-

A screen shot of a computer

Description automatically generated

2-Decision Tree-

A computer screen shot of a computer

Description automatically generated

* In end created a function that can take news as an input and do all the step above mentioned and tell us wheater a news is fake or real.

A computer screen with text on it

Description automatically generated

***References***

🡪https://www.coursera.org/

🡪[https://www.geeksforgeeks.org](https://www.geeksforgeeks.org/)/