

# NAAN MUDHALVAN IBM: AI101

ARTIFICIAL INTELLIGENCE PHASE 3

Fake News Detection Using NLP

**PROJECT NO: 08**

**TEAM MEMBERS:**

1. SIVA JEGADEESH C B
2. RAJKUMAR M
3. BABITH SARISH S
4. SHARVESH

**MENTOR**

* Dr.SUDHAKAR T

**MADRAS INSTITUTE OF TECHNOLOGY, ANNA UNIVERSITY, CHENNAI**

# Fake News Detection Using NLP PHASE 3

**Problem Statement**: **The fake news dataset is one of the classic text analytics datasets available on Kaggle. It consists of genuine and fake articles’ titles and text from different authors. Our job is to create a model which predicts whether a given news is real or fake.**

**In this crucial phase of our project, we embark on the foundational steps of creating a fake news detection model. Our primary objective is to load and preprocess the dataset, setting the stage for the subsequent stages of model development and training.**

**Detecting fake news is a pertinent challenge in the age of information**

**overload, where misinformation can have far-reaching consequences. The effectiveness of our model hinges on the quality and appropriateness of the data it is trained on, making dataset loading and preprocessing pivotal.**

**In the following sections, we will take a closer look at how we load the fake news dataset, examine its structure, and implement preprocessing**

**techniques to ensure that the textual data is in a format conducive to model training. These initial steps lay the groundwork for the development of a robust and accurate fake news detection model, which has the**

**potential to contribute significantly to the critical task of distinguishing fact from fiction in the digital age.**

**Data Source: We use a** [***fake news dataset***](https://www.kaggle.com/datasets/clmentbisaillon/fake-and-real-news-dataset) **available on Kaggle. This dataset contains articles' titles and text, along with their corresponding labels indicating whether the news is genuine or fake.**

**TOOLS:**

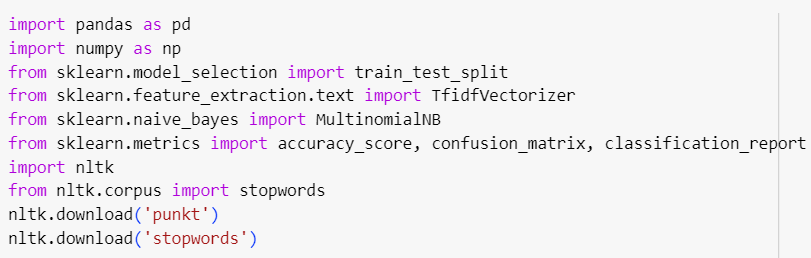
**Google Colab: Google Colab, a cloud-based Jupyter notebook environment, serves as our primary coding platform**

**Python and other libraries for natural language processing (NLP) and machine learning.**

IMPLEMENTATION STEPS:

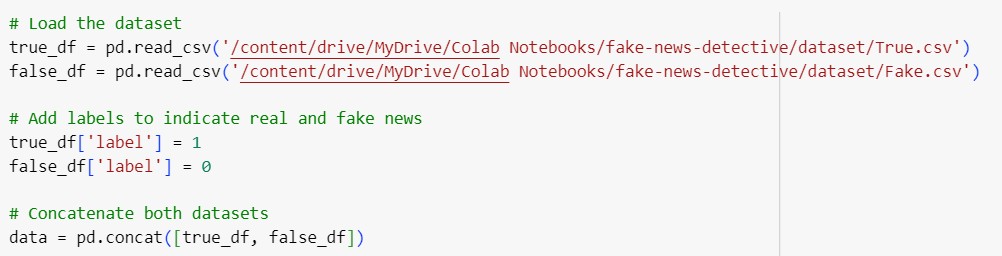
## Import Necessary Libraries:

**Start by importing the required Python libraries, such as pandas, numpy, scikit-learn, and natural language processing libraries like NLTK or spaCy.**

*CODE:*

## Load and Explore the Dataset:

**Load the CSV files 'true.csv' and 'false.csv' using pandas and explore the dataset to understand its structure.**

*CODE:*

## Data Preprocessing:

**Data preprocessing is essential for text data. Perform the following preprocessing steps:**

**Lowercasing: Convert text to lowercase. Tokenization: Split text into words or tokens.**

**Stopword Removal: Remove common words like 'and', 'the', etc.**

**Text Vectorization: Convert text into numerical format (e.g., using TF-IDF or Count Vectorization).**

*CODE:*



## Feature Extraction (TF-IDF):

**Choose a text vectorization method. You can use either Count Vectorization or TF-IDF Vectorization.**

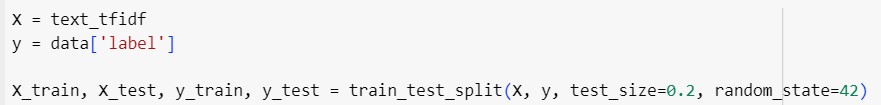
***CODE:***



## Split the Data into Training and Testing Sets:

**Split the dataset into training and testing sets to evaluate the model's performance.**

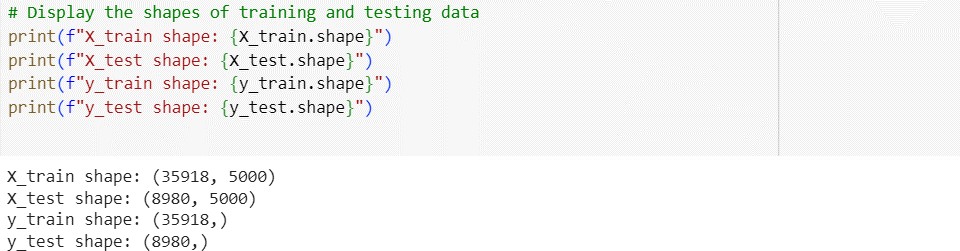
***CODE:***

****

The code for the dataset preprocessing and data splitting stages can be found in our GitHub repository, which streamlines the implementation process and ensures reproducibility. For access to this code, please refer to the following GitHub link: [Fake New Detective](https://github.com/vijaisuria/Fake-News-Detective).

https://github.com/vijaisuria/Fake-News-Detective

## SAMPLE OUTPUT (Data Preprocessing and splitting):



Conclusion

**In this section, we loaded the dataset, performed data preprocessing, and split the data into training and testing sets. The dataset is now ready for model building and evaluation**.