## OPEARTION

# VIDURA

## 1. Introduction

**Fake news detection is a critical task in today's world, where social media and digital platforms enable rapid information spread. This project utilizes Natural Language Processing (NLP) and machine learning to classify news articles as either FAKE or REAL.**

## 2. Objective

**To build a fake news detector that uses NLP techniques to identify the authenticity of news articles. The model is trained using labeled real and fake news datasets and can also handle text extracted from images.**

## 3. Technologies Used

**- Python**
**- Pandas**
**- NumPy**
**- scikit-learn**
**- NLTK**
**- Tesseract OCR**
**- OpenCV**
**- Google Colab**

## 4. Workflow

**1. Load and preprocess real and fake news datasets.**
**2. Combine, clean, and vectorize the text using TF-IDF.**
**3. Train a PassiveAggressiveClassifier model.**
**4. Evaluate model accuracy.**
**5. Predict using text or OCR from an image.**

## 5. Dataset Details

**The project uses two CSV files — 'Fake.csv' and 'True.csv'. Each file contains headlines and text. Labels are added as 0 (Fake) and 1 (Real) before merging the datasets.**

## 6. Model Accuracy

**The model achieved an accuracy of approximately 92-94% on the test data.**

## 7. Code Overview

**Key parts of the code include data preprocessing, vectorization, model training, and prediction functions for both text and images.**

## 8. Image-based Detection

**Using Tesseract OCR and OpenCV, the system can extract text from an uploaded image and classify the extracted news as fake or real.**

## 9. Conclusion

**This project successfully demonstrates the use of NLP and machine learning in detecting fake news. The hybrid input method (text and image) increases its usability across real-world scenarios.**

## 10. Sample Prediction Functions

**Text Prediction: predict_text('sample news content')**
**Image Prediction: extract_text_from_image('image_path')**