

# **Satyasutra -News Article Classification (Fake or Real)**

## **Introduction ::**

In an era where misinformation spreads rapidly, identifying fake news is crucial. This project, Satyasutra, leverages Natural Language Processing techniques to classify news articles as fake or real. The model is designed to assist users in distinguishing credible news sources from misleading information by analyzing textual patterns and linguistic cues.

## **Abstract ::**

Satyasutra utilizes advanced machine learning algorithms to perform text classification on news articles. The system is built using Python and various NLP libraries to preprocess, vectorize, and train models on labeled datasets. The classification task is executed using Logistic Regression or Naïve Bayes, ensuring efficient and accurate predictions. The final product incorporates a user-friendly Streamlit interface that allows users to input text and receive an instant classification result.

## **Tools Used ::**

**Python**

**Sklearn**

**Pandas**

**NLTK**

**TF-IDF Vectorization**

**Logistic Regression or Naïve Bayes**

**Streamlit**

## **Steps Involved in Building the Project ::**

1. Data Collection: Gather a labeled dataset from Kaggle containing both fake and real news articles.
2. Data Preprocessing: Use NLTK for text cleaning by removing stopwords, punctuation, and performing tokenization.
3. Vectorization: Convert text into numerical representations using TF-IDF for model training.

4. Model Training: Apply Logistic Regression or Naïve Bayes for classification and optimize performance using various tuning techniques.
5. Evaluation: Measure the model's effectiveness using F1-score and accuracy to ensure reliable predictions.
6. Interface Development: Build a Streamlit-based UI where users can enter news articles and obtain classification results seamlessly.

## **Conclusion ::**

Satyasutra provides an efficient and reliable solution for distinguishing fake news from genuine reports. By integrating advanced NLP techniques and ensuring accessibility, the model can contribute to reducing misinformation. Continuous improvements, such as expanding the dataset and incorporating deep learning methods, can further enhance its accuracy and effectiveness.

This project offers a practical approach to tackling misinformation by leveraging data-driven techniques, ensuring informed decision-making for users.