FAKE NEWS DETECTION USING NLP

SUBMITTED BY:

NAME:A.SUSHMA

REG NO:723921104001

E-MAIL:yadururusushma@gmail.com

# Fake News Detection Project Documentation

## Problem Statement

The objective of this project is to develop a system that can accurately differentiate between real and fake news articles. Fake news poses a significant challenge in today's information landscape, influencing opinions and creating societal discord. The goal is to create a machine learning model that can identify and classify fake news accurately.

## Design Thinking Process and Development Phases

### 1. Understanding the Problem:

- Defined the scope and importance of fake news detection.

- Researched existing methodologies and challenges in the field.

### 2. Data Collection and Preparation:

- Gathered a dataset of news articles labeled as real or fake.

- Cleaned and preprocessed the data, removing irrelevant information, handling missing values, and ensuring consistency.

### 3. Feature Extraction:

- Extracted features from the text, including word frequencies, TF-IDF scores, and other relevant linguistic attributes.

### 4. Model Selection and Training:

- Chose a classification algorithm suitable for text-based classification.

- Trained the model using the processed data and optimized its parameters for improved accuracy.

## Dataset and Data Processing

### Dataset Used:

The project used a dataset containing news articles labeled as real or fake. The dataset comprises text content and corresponding labels for classification.

### Data Preprocessing Steps:

- Removed special characters, stop words, and performed tokenization.

- Applied stemming/lemmatization to reduce words to their root form.

- Conducted vectorization or feature engineering to prepare the data for model input.

### Feature Extraction Techniques:

- Utilized TF-IDF (Term Frequency-Inverse Document Frequency) for feature representation.

- Extracted n-grams, word embeddings, or other linguistic features to enhance model performance.

## Classification Algorithm and Model Training

### Choice of Algorithm:

- Selected algorithms such as Naive Bayes, Logistic Regression, Support Vector Machines, or neural network-based approaches suitable for text classification tasks.

### Model Training Process:

- Split the dataset into training and validation sets.

- Trained the selected model using the training set.

- Tuned hyperparameters and evaluated model performance on the validation set.

## Submission Details

### Code Compilation:

- Compile all code files related to data preprocessing, model training, and evaluation steps in a structured format.

### README File:

- Create a comprehensive README file detailing how to execute the code, required dependencies, and an overview of the project.

### Dataset Source and Description:

- Include information about the dataset source, its size, the format of data, and a brief description.

### Platform for Sharing:

- Upload the project to a platform like GitHub or a personal portfolio for access and review by others.

By following these outlined documentation and submission guidelines, the fake news detection project can be properly organized, well-documented, and easily accessible for review and utilization.