# Fake News Classifier - NLP Project

# Project Overview

# Introduction

# This project serves as an in-depth exploration and implementation of a Fake News Classifier, a natural language processing (NLP) endeavor undertaken as part of an internship assignment. The primary goal is to develop a system capable of distinguishing between fake and real news articles.

# Project Implementation

# Dataset Acquisition

# The journey commenced with an exhaustive study of IEEE documentation related to fake news classification. Subsequently, a suitable dataset was identified and obtained from Kaggle. The dataset originally comprised three key features: author, text, and title, along with an outcome indicating whether the news is fake (0) or real (1).

# Code Development

# The coding process unfolded with the importation of essential libraries, including Pandas and NumPy, followed by the loading of the dataset in CSV format. The initial dataset contained 20,800 rows and 5 columns. However, to streamline the project and alleviate computational constraints, the 'text' column was excluded, resulting in a reduced dataset of 3 columns.

# Data Preprocessing

# The preprocessing phase involved meticulous steps to prepare the data for analysis. Initially, relevant columns (title and author) were selected, and a custom column was created by combining these features. String entries were converted to lowercase, punctuation was removed, and stopwords were eliminated. Stemming was then applied to further streamline the text data.

# Data Splitting

# The dataset was split into independent features (X) and the corresponding labels (Y). This marked the completion of the preprocessing stage, paving the way for the subsequent vectorization process.

# Vectorization

# The textual data was transformed into a machine-readable format using the TF-IDF (Term Frequency-Inverse Document Frequency) vectorizer from the scikit-learn library. This facilitated the conversion of text data into numerical vectors, enabling effective machine learning model training.

# Model Training

# The logistic regression algorithm was chosen for its simplicity and efficiency in this context. The dataset was divided into training and testing sets (80% and 20%, respectively), and the logistic regression classifier was applied to the training data.

# Model Evaluation

# Upon fitting the model, accuracy testing was performed using the testing dataset. The achieved accuracy score of 96% showcased the efficacy of the developed Fake News Classifier.

# Conclusion

# In conclusion, this internship project involved a comprehensive exploration of fake news classification using natural language processing techniques. The implemented model demonstrated remarkable accuracy, laying the groundwork for potential future enhancements and applications in the field of information verification and authenticity assessment.