

|  |
| --- |
| FAKE NEWS DETECTION  2018  U. Lakshmi Prasanna |
|  |
|  |

# Detect Fake News

|  |
| --- |
| **Key Steps in the Process:**   1. Dataset Description**:**  * *Data has 5 columns: id, title, author, text, and label.* * *The dataset contains some missing values in the title and author columns, which are filled with empty strings.*  1. Preprocessing**:**  * *Combines author and title into a single column called content.* * *Applies stemming using the PorterStemmer to simplify words to their roots.* * *Removes stopwords and performs text cleaning by eliminating non-alphabetic characters.*  1. Feature Extraction:  * *Converts text data into numerical data using the TF-IDF Vectorizer.*    *Splits the data into training and testing sets in an 80:20 ratio.*   1. Model Training:  * *Uses Logistic Regression to train the model on the processed data.*  1. Evaluation:  * *Achieves a training accuracy of* ***98.63%*** *and testing accuracy of* ***97.91%****, demonstrating strong performance.*  1. Prediction:  * *Demonstrates how the model predicts whether a news article is real or fake using a test sample.* |
| Observations:   * *The use of TF-IDF ensures that the model captures the importance of words in relation to the entire dataset.* * *Logistic Regression is simple yet effective for binary classification problems like fake news detection.* * *The accuracy metrics suggest the model is well-suited for the dataset, though further evaluation (e.g., confusion matrix, precision, recall) is necessary for a complete assessment.*   Summary: "Detecting Fake News"  *The document details a machine-learning-based approach for detecting fake news using Python. It employs a dataset labeled as either real (0) or fake (1), consisting of attributes like title, author, text, and a label.* |