FAKE NEWS DETECTION USING NLP

Name: Muthamil.S

NM\_ID: au813821106071

Phase\_3

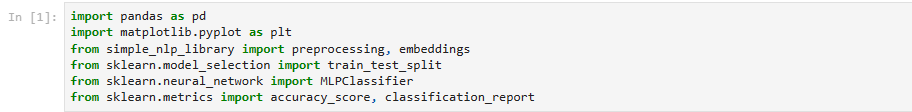


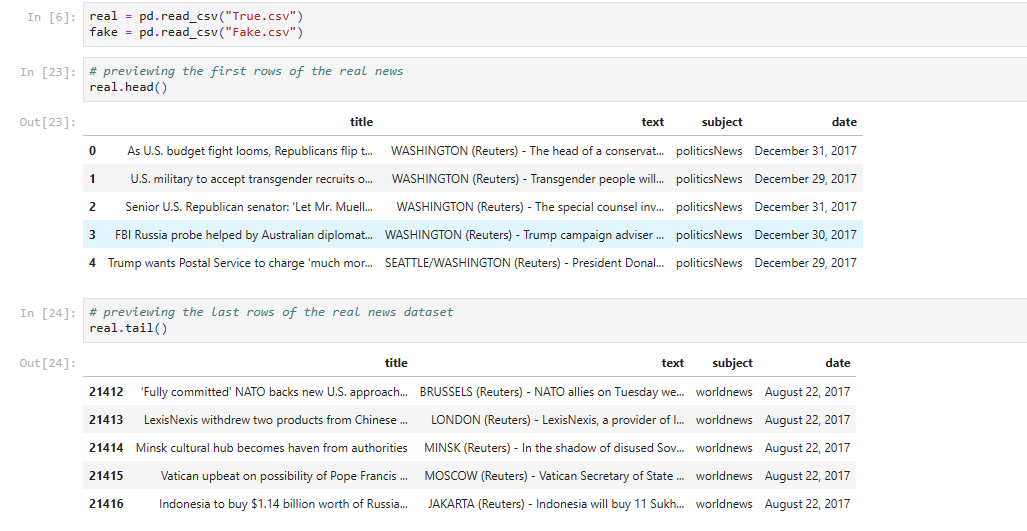
Fake news detection using Natural Language Processing (NLP) involves developing algorithms and techniques to automatically identify the misinformation and deceptive content from the vast amount of information available online. NLP techniques analyze the textual content of news articles, social media posts, and other sources.

Building the Project

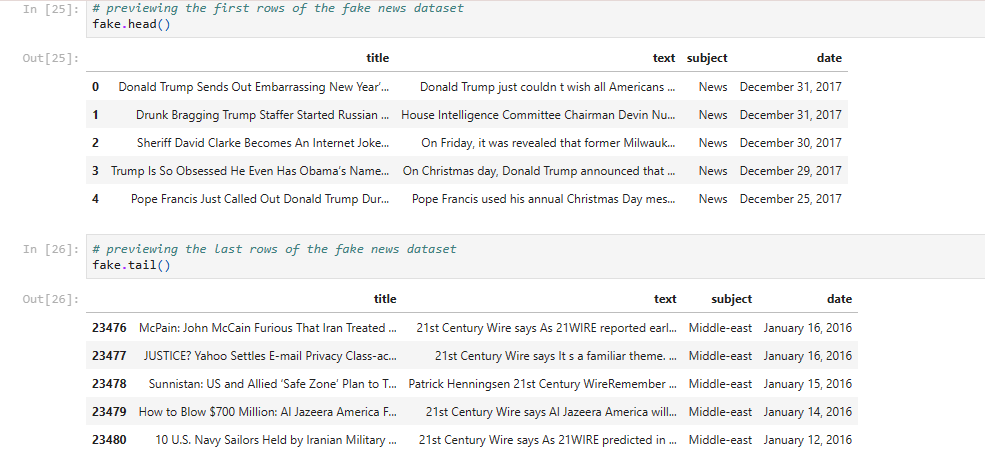
Libraries

These lines import necessary libraries and modules. It includes Pandas for data manipulation, Matplotlib for plotting, scikit-learn (sklearn) for machine learning tools, and custom libraries for NLP-related tasks like text preprocessing and word embeddings.

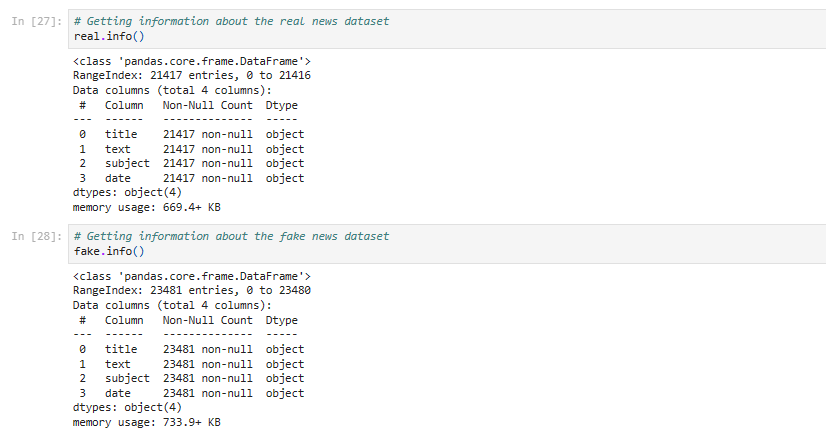
Loading Data

Here, two CSV files, "True.csv" and "Fake.csv," are read into Pandas DataFrames, presumably containing real and fake news headlines.

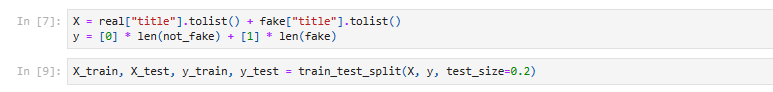
* Printing the first and last rows of the dataset.



* Printing the lables in the dataset.



Combining Data and Labels and Splitting Data into Training and Testing Sets

Loading Data:

Two CSV files, presumably containing real and fake news data, are loaded into Pandas DataFrames: real and fake.

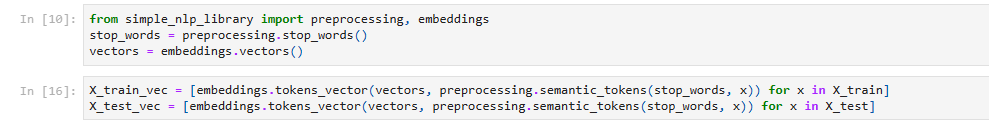
Combining Data:

The 'title' column of both real and fake datasets are combined into a list X.

A binary label is assigned: 0 for real news and 1 for fake news, forming the list y.

Splitting Data:

The data is split into training and testing sets using train\_test\_split from scikit-learn. 80% of the data is used for training (X\_train and y\_train) and 20% for testing (X\_test and y\_test).

Text Preprocessing and Tokenization

The code uses functions from the "simple\_nlp\_library" to preprocess the text data. It removes stop words and converts text into semantic tokens. It then generates word embeddings (vectors) for each headline using an unspecified embedding model.

Conclusion

These codes Loads and preprocesses the data, trains, and evaluates its performance on both training and testing datasets.

THANK YOU