

I Detecting Fake News

I.1 Introduction

In our days we do not trust all the news we hear or see in the social media. Most of the time a big number of news are not real. So the following question rises. how will you detect the fake news? In the following project we will give the answer using Python. By practicing this advanced python project of detecting fake news, we will easily make a difference between real and fake news. Before moving ahead in the algorithm, we must get aware with some terms related to it like fake news, tfidfvectorizer, PassiveAggressive Classifier.

I.2 TfidfVectorizer

TF (Term Frequency): The number of times a word appears in a document is its Term Frequency. A higher value means a term appears more often than others, and so, the document is a good match when the term is part of the search terms.

IDF (Inverse Document Frequency): Words that occur many times a document, but also occur many times in many others, may be irrelevant. IDF is a measure of how significant a term is in the entire corpus.

The TfidfVectorizer converts a collection of raw documents into a matrix of TF-IDF features.

I.3 PassiveAggressiveClassifier

Passive Aggressive algorithms are online learning algorithms. Such an algorithm remains passive for a correct classification outcome, and turns aggressive in the event of a miscalculation, updating and adjusting. Unlike most other algorithms, it does not converge. Its purpose is to make updates that correct the loss, causing very little change in the norm of the weight vector.

II How to Detect Fake News : building the model

Our perpos is to build a model to accurately classify a piece of news as REAL or FAKE.

II.1 The idea with sklearn

This algorithm of detecting fake news deals with fake and real news. Using sklearn, we build a TfidfVectorizer on our dataset. Then, we initialize a PassiveAggressive Classifier and fit the model. In the end, the accuracy score and the confusion matrix tell us how well our model fares.

II.2 The Dataset

the dataset we'll use for this project is called call it news.csv. This dataset has a shape of 7796×4 . The first column identifies the news, the second and third are the title and text, and the fourth column has labels denoting whether the news is REAL or FAKE. The dataset takes up 29.2MB of space and you can