

Fake News Detection Using Machine Learning

What is Fake News?

A type of yellow journalism, fake news encapsulates pieces of news that may be hoaxes and is generally spread through social media and other online media. This is often done to further or impose certain ideas and is often achieved with political agendas. Such news items may contain false and/or exaggerated claims, and may end up being viralized by algorithms, and users may end up in a filter bubble.

What is a Tf/idf Vectorizer?

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.

What is a Passive Aggressive Classifier?

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.

The fake news Dataset

The dataset used for this project is '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.

Download The Dataset: [dataset](#)

Steps for detecting fake news

1. Make necessary libraries import
2. Read the dataset into a DataFrame, and get the shape of the data and the first 5 records.
3. Get the labels from the DataFrame.
4. Split the dataset into training and testing sets.
5. Initialize a Tf/idf Vectorizer
6. Initialize a Passive Aggressive Classifier.
7. Got an accuracy of 92.82% with this model..

Summary

This is a project to detect fake news. Took a political dataset, implemented a Tf/idf Vectorizer, initialized a Passive Aggressive Classifier, and fit the model. After that, ended up obtaining an accuracy of 92.82% in magnitude.