Data Science Project – **Fake News Classifier using Python**

**OBJECTIVE :** To identify whether news is REAL or FAKE

**ABOUT :** We will build a TF-IDF Vectorizer on our dataset using ***sklearn***and fit the model into a ***PassiveAgressiveClassifier.*** Finally, we will find how our model fares using the ***confusion matrix*** and ***F1 scores.***

**DETAILS** :

* + **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 TF-IDFVectorizer converts a collection of raw documents into a matrix of TF-IDF features.

* + **PassiveAgressiveClassifier -** 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.