FAKE NEWS PROJECT

	Unnamed: 0	title	text	label
0	8476	You Can Smell Hillary's Fear	Daniel Greenfield, a Shillman Journalism Fello...	FAKE
1	10294	Watch The Exact Moment Paul Ryan Committed Pol...	Google Pinterest Digg Linkedin Reddit Stumbleu...	FAKE
2	3608	Kerry to go to Paris in gesture of sympathy	U.S. Secretary of State John F. Kerry said Mon...	REAL
3	10142	Bernie supporters on Twitter erupt in anger ag...	— Kaydee King (@KaydeeKing) November 9, 2016 T...	FAKE
4	875	The Battle of New York: Why This Primary Matters	It's primary day in New York and front-runners...	REAL
...	...	...	...	...
6330	4490	State Department says it can't find emails fro...	The State Department told the Republican Natio...	REAL
6331	8062	The 'P' in PBS Should Stand for 'Plutocratic' ...	The 'P' in PBS Should Stand for 'Plutocratic' ...	FAKE
6332	8622	Anti-Trump Protesters Are Tools of the Oligarc...	Anti-Trump Protesters Are Tools of the Oligar...	FAKE
6333	4021	In Ethiopia, Obama seeks progress on peace, se...	ADDIS ABABA, Ethiopia —President Obama convene...	REAL
6334	4330	Jeb Bush Is Suddenly Attacking Trump. Here's W...	Jeb Bush Is Suddenly Attacking Trump. Here's W...	REAL

The dataset

I worked with a dataset that contained target labels that stated whether a piece of news was real or fake. I used a TfIdfVectorizer in order to remove stop words and convert the natural language data into matrix form. I then split transformed the vectorizer on the dataset. After that I instantiated a Passive Aggressive Classifier, trained it, and tested it in order to find out its accuracy in classifying a piece of news. I then found the recall score, the precision score, the F1 score and the confusion matrix for the model as well as printed out its accuracy.

The formula for computing the recall score is given as follows - Number of True Positives / (Number of True Positives + Number of False Negatives)

The formula for computing the precision score is given as follows - Number of True Positives / (Number of True Positives + Number of False Negatives)

The formula for computing the F1 score is given as follows -   (2 * precision * recall) / (precision + recall). The F1 score is a measure of the model's accuracy.

Testing the model resulted in 571 true negatives, 44 false positives, 40 false negatives, and 612 true positives for an accuracy of 93.37%.