FAKE NEWS DETECTION USING PYTHON AND MACHINE LEARNING

THE MAVERICKS
MRCET
DATE : July 15th

INTRODUCTION :

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.

This advanced python project of detecting fake news deals with fake and real news. Using sklearn, we build a Tfidf vectorizer 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.

REQUIREMENTS :

The Libraries that we imported for the execution of the code are Numpy for mathematical calculations ,Pandas for

Creation of dataframe .we also imported scikit learn which is one of the intel optimized libraries. Using sklearn, we build a ML library TfidfVectorizer on our dataset.Then,we initialize a PassiveAggressive Classifier and fit the model.

PLOT FOR FAKE NEWS:

Fake news is false or misleading information presented as news. Fake news often has the aim of damaging the reputation of a person or entity, or making money through advertising revenue.Although false news has always been spread throughout history. The prevalence of fake news has increased with the recent rise of social media, especially the Facebook news feed, and this misinformation is gradually seeping into the mainstream media. Several factors have been implicated in the spread of fake news, such as political polarization, post-truth politics, motivated reasoning, confirmation bias, and social media algorithms.

INTEL OPTIMIZED LIBRARIES :

We have used scikit learn which is one of the intel optimized libraries using which we build 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.We calculated the accuracy with accuracy_score() from sklearn.metrics.

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 in 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.

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.

GETTING ACCURACY:

This advanced python project of detecting fake news deals with fake and real news. We got an accuracy of 92.82% with this model. Finally, we printed the confusion matrix to gain insight into the number of false and true negatives and positives.

CONCLUSION :

So with this model, we have 589 true positives, 587 true negatives, 42 false positives, and 49 false negatives.

We learned to detect fake news with Python. We took a political dataset, implemented a TfidfVectorizer, initialized a PassiveAggressiveClassifier, and fit our model. We ended up obtaining an accuracy of 92.82% in magnitude