ML model

Using the data topics and by doing some preproccesing tasks to separate tweets into two groups we made the tweets content ready for training a classification model.

- The text data is converted into numerical format using TF-IDF (Term Frequency-Inverse Document Frequency) vectorization. It considers the importance of each word in the document relative to its frequency across all documents.

- English stop words are removed.

- A Multinomial Naive Bayes classifier is chosen for text classification.

- The model is trained using the TF-IDF transformed training data.

- The trained model can be used to predict the labels for the text entered.

- Accuracy, precision, recall, and F1-score are as below to evaluate the performance of the model:

Accuracy: 0.9646246764452114

Classification Report:

precision recall f1-score

0 0.95 1.00 0.98

1 1.00 0.87 0.93

accuracy 0.96

macro avg 0.98 0.93 0.95

weighted avg 0.97 0.96 0.96

Some examples to try:

"The two presidents acknowledged the efforts of their respective teams" (Bot)

"I'm so happy to share that I trained this model after many efforts" (Real)