Name: Krishna Chaitanya Sripada
CU Identikey: krsr8608

Explanation of Features:

The process of creating additional features is:

1. I've used the tokenizer parameter that can be passed to the CountVectorizer. This tokenizer parameter is a dictionary which is returned by the features I've used.

2. English stopwords are removed by passing the "stop_words" parameter to the CountVectorizer function.

3. Feature-1: Bag of words is used where the count of each word present in the data is maintained.

4. Feature-2: A custom function is written to generate the bigrams where the input for this feature is a tokenized list of the "Text" column of the training data.

5. The tokenizer used here is the "WordPunctTokenizer" which takes the data and uses regular expression to tokenize the data.