Training/Testing Split:

75% / 25%

Text Vectorization Methods:

1. TfidfVectorizer

Classification methods:

1. Create new column for each vector classification possible, and put 1 if joke classifies as as the category, and 0 if it does not
2. Clean data, by taking content of joke and deleting stop words, unnecessary characters, and putting everything in lowercase letters
3. Convert a collection of raw documents to a matrix of TF-IDF features. With tfidfVectorizer (equivalent to countvectorizer)
4. Train data with Logistic Regression with cleaned joke contents, for each possible classification

Results:







