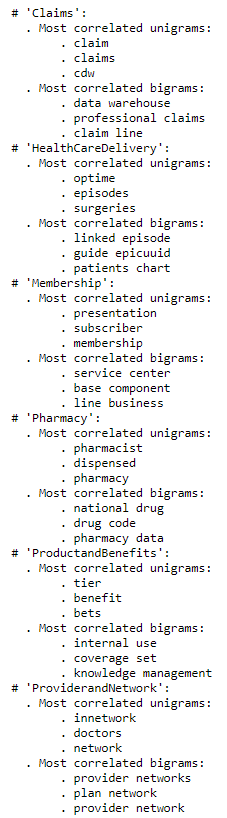
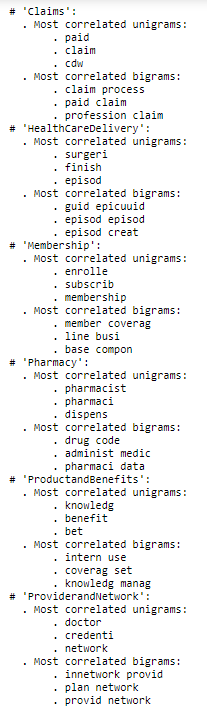


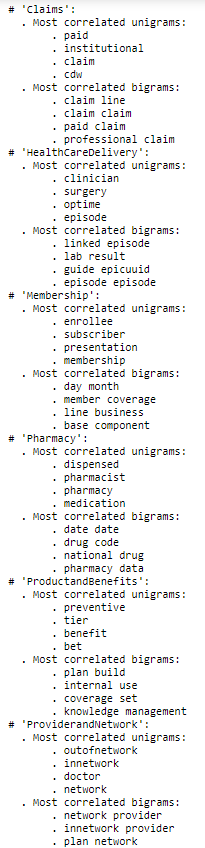
Snowball stemmer

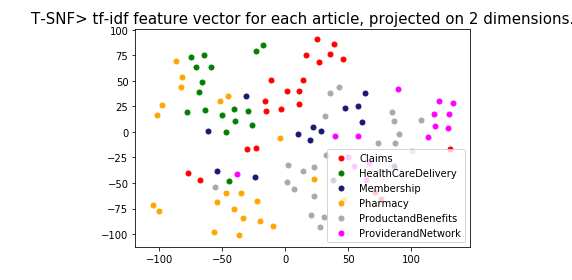


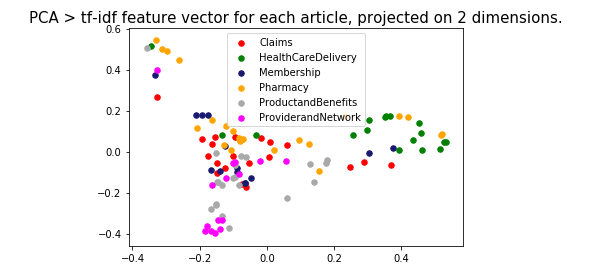
Porter Stemmer

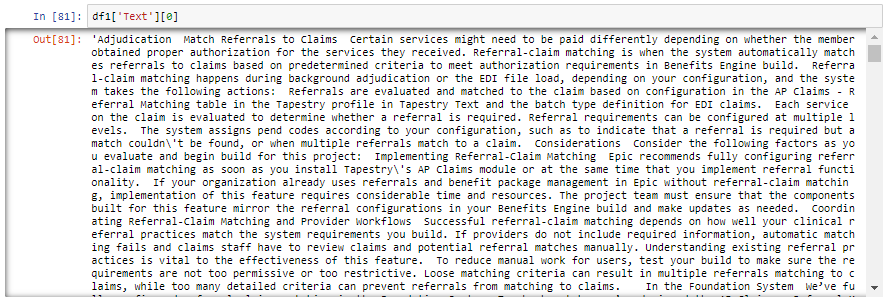


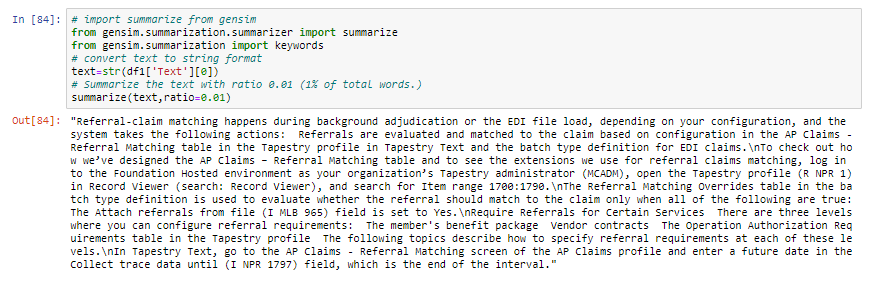
Lemmatize



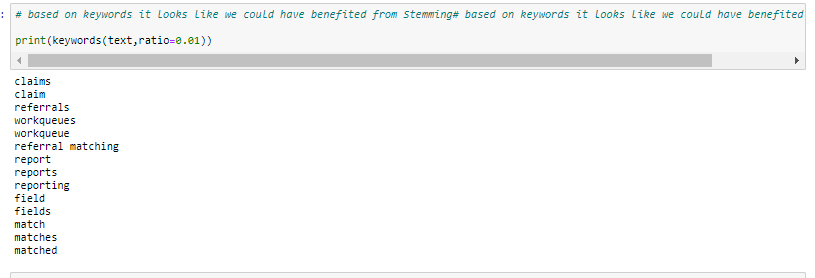








Keywords

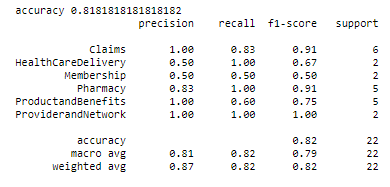


Model 1:

Logistics regression

Word2Vec

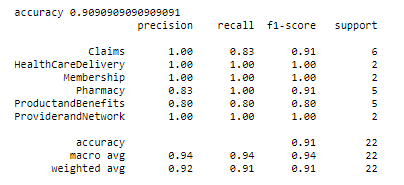
wv = gensim.models.KeyedVectors.load\_word2vec\_format("GoogleNews-vectors-negative300.bin.gz",



Model 2:

Sklearn CountVectorizer()),

('tfidf', TfidfTransformer()),

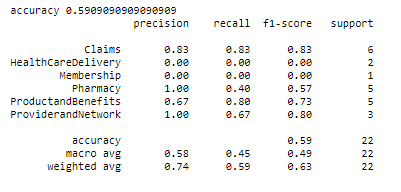


Model 3:

Gensim's Doc2Vec implementation requires each document/paragraph to have a label associated with it.

We do this by using the TaggedDocument method. The format will be "TRAIN\_i" or "TEST\_i" where "i" is

a dummy index of the post.



Reference: <https://github.com/susanli2016/NLP-with-Python/blob/master/Text%20Classification%20model%20selection.ipynb>