

Name: Thiago Santos

Class: Bio NLP

Homework 7

Problems:

- Due to this dataset being a multilabel classification, some modifications had to be made for simplicity. From the sentence label(1 to 25) I rearrange to be only between 1 and 5, and transform this problem into a multi-class classification, from labels 1 to 5.
- In order to use MetaMap, we must register with UMLS. However, the registration seems like it will take over 4+ days. I have requested for the registration this past Monday, Oct 5th, but haven't got the acceptance/rejection yet.
- Therefore, I solve the problem in a different way, but with the same goal. I then choose the features as follow:
 - Count of Words
 - Bag of Words
 - TF-IDF
- Features selection:
 - I use different combinations of my word space in order to create some feature selection, as:
 - Count of Words + Bag of Words
 - Bag of Words + TF-IDF
 - Count of Words + TF-IDF
 - Count of Words + Bag of Words + TF-IDF

F1 Micro and F1 Macro is reported using Random Forest algorithm

