Assignment 1: 28, April.

1. Do NLP classification for any text doc , using one of ML rule supervised algorithms (SVM, KNN, NB)

I applied the SVM model with linear kernel and got the 0.8390% accuracy score on test samples.

Text

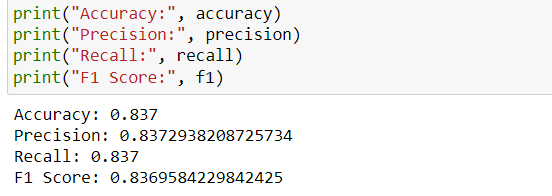
Description automatically generated

Text

Description automatically generated

1. Apply POS in classification and explain how POS will change the result.

After integrating the POS tag in classification model with text tokens I got the lower accuracy score.



1. Does using the morphological analyzer will change the result of classification.

Yes, the morphological analyzer negatively changed the result of classification.

It slightly decreased the accuracy of the classification model.

Text

Description automatically generated

1. Evaluate the classification using F1, Recall and precision.

Text, letter

Description automatically generated

1. What is TF-IDF and are you going to apply with the process of classification.

TF-IDF stands for "Term Frequency-Inverse Document Frequency". The idea behind TF-IDF is that if a word appears frequently in a particular document, but not in many other documents in the corpus, it is likely to be a key characteristic of that document. Conversely, if a word appears frequently across many documents in the corpus, it is less likely to be important in distinguishing one document from another.

I used this to extract the features from the review text for classification.