

Naive Bayes Classifier from the Scratch

Download the Email dataset from [here](#). This dataset is from Lingspam dataset. There are two types of emails, "spam" and "non-spam". You need to implement your own Naive Bayes classifier to categorize emails as "spam" or "non-spam". To convert the text data into vectors, you can use [TfidfVectorizer](#).

Train your Naive Bayes on data given in folder "nonspam-train" and "spam-train". Test it on emails given in folder "nonspam-test" and "spam-test". Report accuracy, F1-score, confusion matrix and area under ROC curve (AUC). Also, report your observations from the result.

Use sklearn's inbuilt Naive Bayes classifier to train on the same dataset. Compare the performance on test dataset with your implementation.