Home Assignment 2

Naive Bayes Classifier Using Python

Code Description

Step 1:

- The train dataset file was organized to make it readable and simplify the calculation of probabilities.
- A **dictionary** function was created to count the occurrence of each word per row

Step 2:

- A **countWords** function was created to count the total number of words as well as the total occurrence of the targets (ham/spam)
- After counting, the targets were sorted to calculate their probabilities.

Could not continue past that.

Questions Answered

```
P(ham) =  0.42633333333333334
P(spam) =  0.5736666666666667
```

1. P(spam) = 0.5736666666666667
2. Undetermined due to unfinished code
3. Undetermined due to unfinished code
4. To beat the classifiers, a spammer would have to avoid word repetition in their emails as well as make each email slightly different than the other.

[Github Link]

References:

- https://www.geeksforgeeks.org/naive-bayes-classifiers/
- https://towardsdatascience.com/how-to-build-and-apply-naive-bayes-classification-for-spam-filtering-2b8d3308501
- https://www.kdnuggets.com/2020/07/spam-filter-python-naive-bayes-scratch.html
- https://www.analyticsvidhya.com/blog/2021/01/a-guide-to-the-naive-bayes-algorithm/
- https://machinelearningmastery.com/naive-bayes-classifier-scratch-python/