

Artificial Intelligence (CS571)

Assignment-7: Naive Bayes Classifier

(Read all the instructions carefully & adhere to them.)

Date: 03-09-2019

Deadline: 04-09-2019

Write a Python program that implements the Naive Bayes text classifier, as discussed in class. To avoid zero counts, make sure you also implement the add-one smoothing

Reading the dataset

Download and unpack [this zip file](#). The package contains a corpus (all_sentiment_shuffled.txt) of customer reviews from six of the review topics used in the paper by Blitzer et al., (2007). The data has been formatted so that there is one review per line, and the texts have been tokenized and normalized. Here is an example of a line.

```
music neg 544.txt i was misled and thought i was buying the entire cd and it contains one song
```

A line in the file is organized in columns:

- 0: topic category label (books, camera, dvd, health, music, or software)
- 1: sentiment category label (pos or neg)
- 2: document identifier
- 3 and on: the document tokens

Evaluating the classifier:

Report the 5-fold cross-validation results in terms of precision, recall, and F-score.

Instructions:

- Please submit your assignment here: <https://bit.ly/2oG4aDn>
- The submission file should be as follows:

Group-NUMBER Assignment-NUMBER.zip