Sentiment Classification of Sentences using Naïve Bayes

You are given a dataset where every sentence is marked as 0 (negative) and 1 (positive). The tag indicates whether the sentence is carrying a positive or negative sentiment. Your task is to classify the sentences in these two categories using Naïve Bayes classifier.

The feature value of a word in a sentence is defined by $\log(1 + tf) * \log(N/df)$, where tf denotes the count of the word in the sentence, N is the total number of sentences in the training data and df denotes the number of sentences in the training data that contains the word.

Dataset link: https://archive.ics.uci.edu/ml/datasets/Sentiment+Labelled+Sentences and then go to data folder and download .zip file.

Note: There are three different datasets in the zip folder. You need to merge them to create a single dataset. Then use 80% of the data as training set and rest as test set.