



Graded Questions

In this segment, you will use the [IMDB](#) movie reviews dataset to classify reviews as 'positive' or 'negative'. We have divided the data into training and test sets. The training set contains 800 positive and 800 negative movie reviews whereas the test set contains 200 positive and 200 negative movie reviews.

This was one of the first widely-available sentiment analysis datasets compiled by Pang and Lee's. The data was first collected in 2002, however, the text is similar to movies reviews you find on IMDB today. The dataset is in a CSV format. It has two categories: Pos (reviews that express a positive or favourable sentiment) and Neg (reviews that express a negative or unfavourable sentiment). For this exercise, we will assume that all reviews are either positive or negative; there are no neutral reviews.

You will need to build a Multinomial Naive Bayes classification model in Python for solving the questions.

Please find the imdb_train dataset [here](#) and imdb_test dataset [here](#).

Note:

- Tag negative(Neg) as 0 and positive(Pos) as 1.

Please answer the following questions based on the model you make using above data.

