Review Classification

Text classification is the process of categorizing the text into a group of words. By using NLP, text classification can automatically analyze text and then assign a set of predefined tags or categories based on its context. NLP is used for sentiment analysis, topic detection, and language detection.

In this mini-project, you need to develop an automated review classification system able to classify the customer reviews between one of the following classes (Mixed, Negative and Positive) using the attached dataset.

The dataset combines reviews from hotels, books, movies, products and a few airlines. It has three classes (Mixed, Negative and Positive). Most were mapped from reviewers' ratings with 3 being mixed, above 3 positive and below 3 negative. Each row has a label and text separated by a tab (tsv). Text (reviews) were cleaned by removing Arabic diacritics and non-Arabic characters. The dataset has no duplicate reviews.

Your Supervised learning must use 70% of the dataset as training and 30% as testing. You may use cross-folding for your classification.

The project team can be a max of 4 members.