

Assignment 6: Sentiment Analysis by Text Classification

Download “amazon_review_300.csv”. This dataset has three columns: label, title, and text. In this assignment, we build classifiers using label and text columns. The classifiers are used to detect review sentiment: 1 or 2.

1. Experiment 1: **Naïve Base classifier with cross validation**

Write a block of code to create a **MultinomialNB** classifier with 6-fold cross-validation using tf-idf matrix with stop words removed. Report macro precision/recall/f1 score for each fold.

2. Experiment 2: **Tune parameters using grid search**

Write a block of code to tune the classifier you created in (2) using grid search. The grid search is to find best values for the following parameters:

- **stop_words**: [None, "english"]
- **min_df**: [1,2,3,5]
- **alpha**: [0.5,1.0,1.5,2.0]

Compare the performance with the classifier in (2) and write your conclusion.

3. Experiment 3: **How many samples are enough? Show the impact of sample size on classifier performance**

Download “amazon_review_large.csv” which contains 20,000 reviews. Starting with 400 samples, in each round you build a classifier with 400 more samples. i.e. in round 1, you use samples from 0:400, and in round 2, you use samples from 0:800, ..., until you use all samples. In each round, do the following:

- a. create tf-idf matrix using **TfidfVectorizer** with stop words removed
- b. train the classifier using **linearSVC** model with 10-fold cross validation
- c. train another classifier using **multinomialNB** with 10-fold cross validation
- d. For each classifier, collect the average macro f1 score over 10 folds

Draw a line chart show the relationship between sample size and average macro f1 score.

Write your analysis on the following:

- How sample size affects each classifier’s performance?
 - How many samples do you think would be needed for each model for good performance?
 - How is performance of SVM classifier compared with Naïve Bayes classifier?
- e. (**Bonus**): Analyze model stability of these two models as the sample size increases. You can use f-score variance (or std) over folds to indicate model stability. The smaller the variance, the more stable a model.

Submission guidelines:

Submit a python file that contains script for all 3 experiments, but clearly separate your code for each experiment. Also, have a print statement in your code to print out the performance metrics of each experiment. Write all your analyses into a text file.