

COMP 551: Assignment 3

Anand Kamat

ID# 260773313

- Both the datasets, IMDB and yelp were converted to Binary Bag Of Words and Frequency Bag Of Words.
- Yelp:** Hyperparameter Tuning:
 - For Decision Trees the hyperparameters tuned were: max_depth, max_samples, min_samples_leaf all of which set to 100 different values
 - For SVM the hyperparameter tuned was 'C' which was set to values ranging from 0 to 1 for Binary Bag of Words and 1 – 100 for Frequency Bag of Words.
 - For Bernoulli Naïve Bayes the hyperparameter tuned was 'alpha' which ranged from 0 to 1 as well.

The following were the best hyperparameters calculated based on the F1-Score.

Yelp	Hyperparameters on Binary Bag Of Words	Hyperparameters on Frequency Bag of Words
Decision Trees	Max_depth=530, max_leaf_node=140, min_samples_split=4, min_samples_leaf=2	Max_depth=6, max_leaf_node=1, min_samples_split=4, min_samples_leaf=2
SVM	C = 0.1	C = 120
Naïve Bayes(Bernoulli)	Alpha=0.275	Alpha=0.267
Naïve Bayes(Gaussian)	No hyperparameters	No Hyperparameters

Performance:

Decision Trees (Binary Bag of words):

Training F1-Score: 0.410

Validation F1-Score: 0.397

Test F1-Score: 0.3001

Decision Trees (Frequency Bag of words):

Training F1-Score: 0.4341

Validation F1-Score: 0.389

Test F1-Score: 0.3510

SVM (Binary Bag of words):

Training F1-Score: 0.9289

Validation F1-Score: 0.460

Test F1-Score: 0.4488

SVM (Frequency Bag of words):

Training F1-Score: 0.904

Validation F1-Score: 0.4682

Test F1-Score: 0.4517

Bernoulli (Binary Bag of words):

Training F1-Score: 0.78

Validation F1-Score: 0.3999

Test F1-Score: 0.3843

Gaussian (Frequency Bag of words):

Training F1-Score: 0.783

Validation F1-Score: 0.2437

Test F1-Score: 0.2570

Random Classifier:

Training F1-Score: 0.154212

Validation F1-Score: 0.1577181

Test F1-Score: 0.144524

Majority Classifier:

Training F1-Score: 0.3525714

Validation F1-Score: 0.356

Test F1-Score: 0.351

Report: There is no significant improvement between the F1-Scores of binary bag of words and frequency bag of words. All the classifiers perform better than the baseline classifiers (random and majority class).

3. IMDB Dataset

Similar process as yelp was used to tune the hyperparameters for the IMDB dataset.

IMDB:

	<i>Hyperparameters on Binary Bag Of Words</i>	<i>Hyperparameters on Frequency Bag of Words</i>
<i>Decision Trees</i>	Max_depth=530, max_leaf_node=75, min_samples_split=4, min_samples_leaf=2	Max_depth=320, max_leaf_node=75, min_samples_split=4, min_samples_leaf=2
<i>SVM</i>	C = 0.01	C = 90
<i>Naïve Bayes(Bernoulli)</i>	Alpha=0.68	Alpha=0.70
<i>Naïve Bayes(Gaussian)</i>	No hyperparameters	No Hyperparameters

Performance:

Decision Trees (Binary Bag of words):

Training F1-Score: 0.7655

Validation F1-Score: 0.73412

Test F1-Score: 0.7390

Decision Trees (Frequency Bag of words):

Training F1-Score: 0.7612

Validation F1-Score: 0.7250

Test F1-Score: 0.72812

SVM (Binary Bag of words):

Training F1-Score: 0.95533

Validation F1-Score: 0.8766

Test F1-Score: 0.8723

SVM (Frequency Bag of words):

Training F1-Score: 0.9433

Validation F1-Score: 0.8789

Test F1-Score: 0.87433

Bernoulli (Binary Bag of words):

Training F1-Score: 0.873

Validation F1-Score: 0.8390

Test F1-Score: 0.83432

Gaussian (Frequency Bag of words):

Training F1-Score: 0.862214

Validation F1-Score: 0.763

Test F1-Score: 0.6977

Random Classifier:

Training F1-Score: 0.49593

Validation F1-Score: 0.50855

Test F1-Score: 0.502444

Report: There is no significant improvement between the F1-Scores of binary bag of words and frequency bag of words. All the classifiers perform better than the baseline classifiers (random and majority class). The binary classification problem for the IMDB dataset seems to have shown improvement of the performance of classifiers as compared to the 5 class yelp dataset. The relative performance of classifiers remain the same. All the classifiers here also perform better than the baseline classifiers.

Datasets:

- Vocabulary: For the yelp and IMDB datasets, the 'Vocab' folder contains the text files. IMDB-vocab.txt and yelp-vocab.txt represent the vocabulary list for the combined dataset(training+validation+ test)
- The id review datasets are stored in 'Generated datasets' folder.