Assignment 2

COMP - 551

Student ID: 260880458

Mohd Safwan Ahmad

Prob.1

See data files on MyCourses.

Prob.2

YELP BBOW

a.

F1 Score of Random Classifier on Test: 0.17230888773618042

F1 Score of Majority Classifier on Test: 0.10392301998519615

b. See Q2 Yelp BBOW

c.

Model	*Parameters and values tried*
BernoulliNB (Naïve Bayes)	'alpha':[.01,.1,.5,1]})
Decision Tree	'random_state':[10] 'max_depth':[None,10,100,1000],' min_samples_split':[2,5,10]
SVM	'random_state':[10],'loss':['hinge','squared_hinge'],'C':[.5,1.0,2.0]}

d.

For '**sklearn.naive_bayes.BernoulliNB**'
Best params for Validation: {'alpha': 0.01}
Best F1 Score on Validation: 0.38051604534431227

Test score for best params: 0.36572174954585585

For '**sklearn.tree.tree.DecisionTreeClassifier**'

Best params for Validation: {'max_depth': 10, 'min_samples_split': 10, 'random_state': 10}
Best F1 Score on Validation: 0.28863171451423775
Test score for best params: 0.27117951516332034

For class '**sklearn.svm.classes.LinearSVC**'
Best params for Validation: {'C': 0.5, 'loss': 'squared_hinge', 'random_state': 10}
Best F1 Score on Validation: 0.43745568878627256

Test score for best params: 0.4026956035166604

e.

SVM performed best
Trying: {'C': 0.5, 'loss': 'squared_hinge', 'random_state': 10}
F1 Score Validation: 0.43745568878627256
Test score for SVM: 0.4026956035166604

Naïve Bayes classifier didn't perform as good because the data doesn't follow the NB Assumption.
The decision tree performed even worse because the data doesn't fit the decision tree model.

Prob.3

YELP FBOW

a. See Q3 Yelp FBOW

b.

Model	*Parameters and values tried*
GaussianNB (Naïve Bayes)	
Decision Tree	'random_state':[10] 'max_depth':[None,10,100,1000],' min_samples_split':[2,5,10]
SVM	'random_state':[10],'loss':['hinge','squared_hinge'],'C':[.5,1.0,2.0]}

c.

For class **'sklearn.naive_bayes.GaussianNB'**
Best params for Validation: {}
Best F1 Score on Validation: 0.2351740905717059
Test score for best params: 0.23929827431059195

For class 'sklearn.tree.tree.DecisionTreeClassifier'
Best params for Validation: {'max_depth': 10, 'min_samples_split': 5, 'random state': 10}
Best F1 Score on Validation: 0.3092932288750916
Test score for best params: 0.2628783432037355

For class 'sklearn.svm.classes.LinearSVC'
Best params for Validation: {'C': 1.0, 'loss': 'squared_hinge', 'random_state': 10}
Best F1 Score on Validation: 0.4773903905644473
Test score for best params: 0.4686722893287243

d.

SVM Performed better with Test score for best params: 0.4686722893287243

SVM Performed better than rest for both FBOW and BBOW but here the performance is better.

e.

SVM performs slightly better for the BBoW than the FBoW (0.4026956035166604 vs. 0.4686722893287243).

For **'sklearn.naive_bayes.BernoulliNB'**
BBOW
Test score: 0.36572174954585585
FBOW
Test score: 0.23929827431059195

BernoulliNB' performed better for BBOW than FBOW

For class 'sklearn.tree.tree.DecisionTreeClassifier'
BBOW
Test score: 0.27117951516332034
FBOW
Test score: 0.2628783432037355

DecisionTreeClassifier' performed almost the same for both BBOW and FBOW

For class **'sklearn.svm.classes.LinearSVC'**
BBOW
Test score: 0.4026956035166604
FBOW
Test score: 0.4686722893287243

SVM performed better for FBOW than BBOW

f. We cannot specifically say which is representation is better but looking at F1 measure for SVM it seems that FBOW is better than BBOW.

Prob.4

IMDB BBOW

a. F1 Score of Random Classifier on Test: 0.5035949168119481
b. See Q4 IMDB BBOW
c.

Model	*Parameters and values tried*
BernoulliNB (Naïve Bayes)	'alpha':[.01,.1,.5,1]})
Decision Tree	'random_state':[10] 'max_depth':[None,10,100,1000],' min_samples_split':[2,5,10]
SVM	'random_state':[10],'loss':['hinge','squared_hinge'],'C':[.5,1.0,2.0]}

d.

For '**sklearn.naive_bayes.BernoulliNB**'

Best params for Validation: {'alpha': 0.01}
Best F1 Score on Validation: 0.8445686588243655

Test score for best params: 0.8364324948115234

For '**sklearn.tree.tree.DecisionTreeClassifier**'

Best params for Validation: {'max_depth': 10, 'min_samples_split': 10, 'random_state': 10}
Best F1 Score on Validation: 0.7111417333906496

Test score for best params: 0.7062962555570182

For class '**sklearn.svm.classes.LinearSVC**'

Best params for Validation: {'C': 0.5, 'loss': 'squared_hinge', 'random_state': 10}
Best F1 Score on Validation: 0.8471989670650173

Test score for best params: 0.8367588275366029

e.
BernoulliNB(with alpha = 0.01) and SVM (with c = .5) performed almost same whereas DecisionTreeClassifier didn't perform as good due to the same reason – Data is not suited for the Decision Tree model.

Prob.5

IMDB FBOW

a. See Q5 IMDB FBOW

b.

Model	*Parameters and values tried*
GaussianNB (Naïve Bayes)	
Decision Tree	'random_state':[10] 'max_depth':[None,10,100,1000],' min_samples_split':[2,5,10]
SVM	'random_state':[10],'loss':['hinge','squared_hinge'],'C':[.5,1.0,2.0]}

c.