Report
Assignment – 2
Aditya Prakash axp171931

The accuracy improves by around 5-10% in general after removing the stop words. This is due to the case that the stopwords are low quality features which are usually unwanted by the programmer who is coding the classifier.
In most of the cases, including the stop words give negative results, example: Logistic Regression. In other cases, the stop words don't have as much of a negative effect on classification. This can be coincidental. Like in Naive Bayes classifier in this assignment.
But it does not change the fact that removing stopwords leads to more accurate classification.

(1) Naive Bayes

Accuracy without stop words: 92.22462203023758%

Accuracy with stop words : 95.46436285097192%

(2) Logistic Regression

Number of Iterations: 20

Lambda (λ)	Accuracy with Stop Words(%)	Accuracy without Stop Words(%)
0.25	80.5439330543933	89.5397489539749
0.50	80.75313807531381	87.86610878661088
0.75	76.56903765690377	87.86610878661088
1	80.75313807531381	85.77405857740585
0.80	81.17154811715481	88.70292887029288
0.90	80.9623430962343	86.19246861924687
0.70	81.3807531380753	87.65690376569037
0.60	77.61506276150628	89.3305439330544
0.40	82.84518828451883	88.70292887029288
0.30	79.49790794979079	84.72803347280335