1 Using NB:

1.1 Test Report Without Stop-words:

	ham	spam
Wrongly classified files /total files	11/348	21/130
Accuracy	96.83908045977012%	83.84615384615385%

1.2 Test Report With Stop-words

	ham	spam
Wrongly classified files /total files	13/348	19/130
Accuracy	96.26436781609196%	85.38461538461539%

Conclusion: accuracy will improve without stop-words. This is because NB produces real prob data, without stop-words the prob will get closer to real values;

2 Using Logistic Regression:

2.1 lambda: 0.001, eta: 0.01, iterations: 200

	With stop-words		Without stop-words	
	ham	spam	ham	spam
Wrongly classified files /total files	1/348	3/130	1/348	4/130
Accuracy	99.7%	97.6%	99.7%	96.9%

2.2 lambda: 0.035, eta: 0.01, iterations: 200

	With stop-words		Without stop-words	
	ham	spam	ham	spam
Wrongly classified files /total files	268/348	96/130	279/348	102/130
Accuracy	22.9%	26.1%	19.8%	21.5%

2.3 lambda: 0.075, eta: 0.01, iterations: 200

	With stop-words		Without stop-words	
	ham	spam	ham	spam
Wrongly classified files /total files	285/348	103/130	330/348	112/130
Accuracy	18.1%	20.7%	5.1%	13.8%

Conclusion: lambda=0.001 has good performance. accuracy will NOT improve without stop-words. This is because LR doesn’t produce real prob data, without stop-words the prob will NOT get closer to real values;