Report:- ML for Cybersecurity Lab-1 --Vinay Bhapkar—vvb231@nyu.edu

top_10_features: [('language', 1) ('free', 1) ('remove', 1) ('linguistic', 1)
('university', 1) ('money', 1) ('our', 1) ('click', 1) ('business', 1)
('market', 1)]

Classifier	Spam precision	Spam Recall	Accuracy
Bernoulli NB with Binary Features N=10	0.87	0.82	0.948453608247
Multinomial NB with Binary Features N=10	0.89	0.82	0.951890034364
MultinomialNB with Term Frequency Feature N =10	0.85	0.94	0.962199312715
Bernoulli NB with Binary Features N=100	0.94	0.63	0.931271477663
Multinomial NB with Binary Features N=100	0.98	0.90	0.979381443299
MultinomialNB with Term Frequency Feature N =100	0.96	0.96	0.986254295533
Bernoulli NB with Binary Features N=1000	1.00	0.61	0.93470790378
Multinomial NB with Binary Features N=1000	1.00	0.94	0.989690721649
MultinomialNB with Term Frequency Feature N =1000	1.00	0.94	0.989690721649

Classifier	Spam precision	Spam Recall	Accuracy
SVM – N=1000 With binary features Hyperparameters: C = 1.5 gamma = 0.0074	1.00	0.90	0.9828

Output as obtained in Python:-

N = 10

Bernoulli NB with Binary Features

Accuracy: 0.948453608247

	precision	recall	f1-score	support
ham	0.96	0.98	0.97	242
spam	0.87	0.82	0.84	49
avg / total	0.95	0.95	0.95	291

Multinomial NB with Binary Features

Accuracy: 0.951890034364

	precision	recall	f1-score	support
ham	0.96	0.98	0.97	242
spam	0.89	0.82	0.85	49

avg / total 0.95 0.95 0.95 291

MultinomialNB with Term Frequency Feature

Accuracy: 0.962199312715

	precision	recall	f1-score	support
ham	0.99	0.97	0.98	242
spam	0.85	0.94	0.89	49
avg / total	0.96	0.96	0.96	291

N = 100

Bernoulli NB with Binary Features

Accuracy: 0.931271477663

	precision	recall	f1-score	support
ham	0.93	0.99	0.96	242
spam	0.94	0.63	0.76	49
avg / total	0.93	0.93	0.93	291

Multinomial NB with Binary Features

Accuracy: 0.979381443299

	precision	recall	f1-score	support
ham	0.98	1.00	0.99	242
spam	0.98	0.90	0.94	49
avg / total	0.98	0.98	0.98	291

MultinomialNB with Term Frequency Feature

Accuracy: 0.986254295533

	precision	recall	f1-score	support
ham	0.99	0.99	0.99	242
spam	0.96	0.96	0.96	49
avg / total	0.99	0.99	0.99	291

N = 1000

Bernoulli NB with Binary Features

Accuracy: 0.93470790378

	precision	recall	f1-score	support
ham	0.93	1.00	0.96	242
spam	1.00	0.61	0.76	49
avg / total	0.94	0.93	0.93	291

Multinomial NB with Binary Features

Accuracy: 0.989690721649

	precision	recall	f1-score	support
ham	0.99	1.00	0.99	242
spam	1.00	0.94	0.97	49
avg / total	0.99	0.99	0.99	291

MultinomialNB with Term Frequency Feature

Accuracy: 0.989690721649

	precision	recall	f1-score	support
ham	0.99	1.00	0.99	242
spam	1.00	0.94	0.97	49
avg / total	0.99	0.99	0.99	291

SVM Classifier for N = 1000

Hyperparameters:

C = 1.5

gamma = 0.0074

Accuracy: 0.982817869416

	precision	recall	f1-score	support
ham	0.98	1.00	0.99	242
spam	1.00	0.90	0.95	49
avg / total	0.98	0.98	0.98	291