

Dataset:

observation = 403465

spam = 146316

non-spam = 257149

	LRC-BOW	LRC-TFIDF	LRC-W2V	Naive Bayes	Decision Tree	Max Entropy	
Accry_trn	0.9634	0.9884	0.9913	0.9497	0.9292	0.9717	
Accry_tst	0.8493	0.9710	0.9671	0.9411	0.8908	0.9604	
Precision	0.9061	0.9417	0.9381	0.8438	0.9318	0.9145	
Recall	0.5413	0.9604	0.9505	0.9802	0.6766	0.9538	
F-score	0.6777	0.9510	0.9443	0.9069	0.7839	0.9338	
# Pred	23897	1000	300	ALL BOW	1000 -TFIDF BOW	1000 -TFIDF BOW	
Type II%	0.023	0.025	0.026	0.0751	0.0205	0.0369	
Comment	OF/ low recall low catch/ low T II conserv.	very good	very good	catch not good/ high precision given by imbalance/ T-II high aggressive	low recall low catch/ low T II conserv. sparse	Good Sparse	

LRC-BOW
[[715, 17],
[139, 164]]

LRC-TFIDF
[[714, 18],
[12, 291]]

LRC-W2V
[[713, 19],
[15, 288]]

NB
[[677, 55],
[6, 297]]

DT
[[717, 15],
[98, 205]]

MaxEnt
[[705, 27],
[14, 289]]