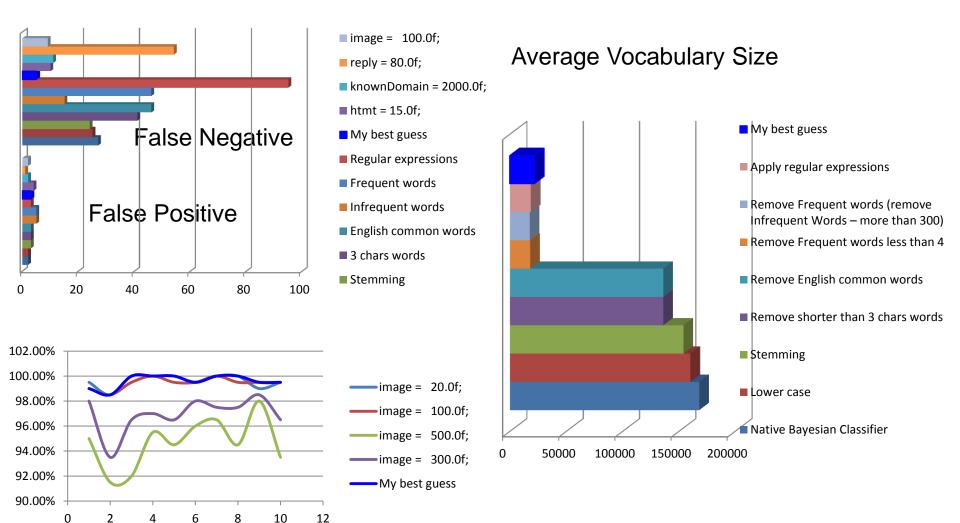
Results from Stratified cross-validation on different techniques 100.50% 100.00% knownDomain = 500.0f; 99.50% knownDomain = 1000.0f; 100.00% 99.00% knownDomain = 2000.0f; 98.50% knownDomain = 3000.0f; 98.00% My best guess 98.00% 97.50% 10 0 5 15 ■ Native Bayesian Classifier 100.50% 96.00% 100.00% html = 5.0f; Lower case 99.50% html = 8.0f; Stemming htmt = 10.0f; 99.00% 94.00% htmt = 15.0f; ■ 3 chars words 98.50% My best guess 98.00% ■ English common words 92.00% 5 10 15 105.00% Infrequent words reply = 100.0f; 100.00% 90.00% reply = 500.0f; Frequent words 95.00% - reply = 50.0f; ■ Regular expressions 90.00% reply = 10.0f; 88.00% 2 3 4 5 6 7 8 85.00% reply = 80.0f; My best guess My best guess 80.00% 10 15 0 5





T – test showing the difference between Naïve Bayes and the My Best Guess

0.005894292



