Table A: Properties Experiment

	IDFTransform	TFTransform	lowerCase Tokens	minTerm Freq	stemmer	tokenizer	useStopList	comment	Correct%	Incorrect %	time to build the model	time to test the model
Default	FALSE	FALSE	FALSE	1	null	wordTokenizer	null		82.3217%	17.6783%	2m 58s 0ms	330ms
	TRUE								82.3217%	17.6783%	3m 1s 510ms	170ms
		TRUE							82.3217%	17.6783%	2m 53s 0ms	370ms
			TRUE						80.6718%	19.3282%	2m 18s 460ms	140ms
				10					81.9093%	18.0907%	2m 17s 960ms	350ms
				100					81.4968%	18.5032%	0m 14s 980ms	50ms
					IteratedLovins				80.5539%	19.4461%	1m 47s 490ms	130ms
					Lovins				81.32%	18.68%	1m 51s 330ms	130ms
					Snowball				82.3217%	17.6783%	3m 14s 860ms	240ms
						Alphabetic			82.0271%	17.9729%	2m 59s 740ms	150ms
						CharacterNGram		min:1; max:3	78.02%	21.98%	2m 26s 0ms	170ms
						NGram		min:1; max:3	82.2039%	17.7961%	4m 15s 10ms	400ms
							MultiStopwords		82.3217%	17.6783%	3m 12s 580ms	210ms
							Rainbow		83.5592%	16.4408%	3m 37s 960ms	400ms
							WordsFromFile	xpo6.com/download-stop-word-list/	82.4985%	17.5015%	3m 1s 390ms	470ms
Average									81.7246%		2m 40s 685ms	
	TRUE	TRUE							82.3217%	17.6783%	3m 14s 440ms	260ms
	TRUE	TRUE		100					81.4968%	18.5032%	0m 17s 160ms	40ms
	TRUE	TRUE	TRUE	50					81.0843%	18.9157%	0m 33s 90ms	70ms
			TRUE	100					81.0843%	18.9157%	0m 35s 590ms	90ms
					Snowball		Rainbow		83.5592%	16.4408%	3m 45s 990ms	480ms
					Snowball		MultiStopwords		82.3217%	17.6783%	3m 8s 120ms	5060ms
					Snowball		WordsFromFile	xpo6.com/download-stop-word-list/	82.4985%	17.5015%	3m 57s 780ms	660ms
					Snowball	Alphabetic	Rainbow		81.1432%	18.8568%	3m 36s 310ms	220ms
					Snowball	NGram	Rainbow		81.9682%	18.0318%	5m 12s 230ms	220ms
						Alphabetic	Rainbow		81.1432%	18.8568%	3m 51s 200ms	170ms
						NGram	Rainbow		81.9682%	18.0318%	5m 52s 790ms	310ms
Average									81.8718%		3m 5s 882ms	

	Naive-Bayes (naiveBayesMultinomial)	Support Vecto Machine (SMO)	Decision Trees (J48)
Training Set (50% resample)	Time taken to build model: 0.03 seconds === Evaluation on training set === Time taken to test model on training data: 0.6 seconds === Summary === Correctly Classified Instances 4544 91.0621 % Incorrectly Classified Instances 446 8.9379 % Kappa statistic 0.8843 Mean absolute error 0.0184 Root mean squared error 0.1265 Relative absolute error 11.9757 % Root relative squared error 45.6282 % Total Number of Instances 4990	Time taken to build model: 7.71 seconds === Evaluation on training set === Time taken to test model on training data: 2.32 seconds === Summary === Correctly Classified Instances 4821 96.6132 % Incorrectly Classified Instances 169 3.3868 % Kappa statistic 0.9559 Mean absolute error 0.1602 Root mean squared error 0.2722 Relative absolute error 104.2032 % Root relative squared error 98.2077 % Total Number of Instances 4990	
10 folds Cross-Validation (50% resample)	Time taken to build model: 0.02 seconds === Stratified cross-validation === === Summary === Correctly Classified Instances 4307 86.3126 % Incorrectly Classified Instances 683 13.6874 % Kappa statistic 0.8229 Mean absolute error 0.0281 Root mean squared error 0.1595 Relative absolute error 18.3076 % Root relative squared error 57.5491 % Total Number of Instances 4990	Time taken to build model: 7.87 seconds === Stratified cross-validation === === Summary === Correctly Classified Instances 4363 87.4349 % Incorrectly Classified Instances 627 12.5651 % Kappa statistic 0.8361 Mean absolute error 0.1608 Root mean squared error 0.2734 Relative absolute error 104.6106 % Root relative squared error 98.6254 % Total Number of Instances 4990 === Detailed Accuracy By Class ===	