

Results

Best achieved accuracy in Percent (including Second Maiden's Tragedy)

Results can be slightly different every time it runs due to the workings of the classifiers

Feature	Naïve Bayes	SVM	Fox et al.	N-Gram-Tracing (Grieve et al.)
Word Frequency (Baseline)	83.1	/	84.4	38,4 (7-char-ngrams) 60,8 (8-char-ngrams) 29,4 (2-word-ngrams)
-----	-----	-----	-----	-----
Syntactic	Number of stop	words in brackets		
POS tags + Stop Words + n-grams of POS & Stop Words	81.8 (305) 80.5/81.8 (710) 79.2 (747) 79.2 (507)	81.1 (305) 80.5/81.8 (710) 80.5 (747) 81.8 (507)	86	
All others are in combination with the Fox features				
Lexical				
Word frequency	81.8 (305) 80.5 (710)	81.8 (305) 81.8 (507)		
Stemming	83.1 (305) 84.4 (710)	81.8 (305) 81.8 (507)		
Lemmatizing	84.4 (305) 84.4 (710) 84.4 (747)	81.8 (305) 80.5 (507)		
Hapax Legomena Avg.	81.8 (305)			
Word Length Avg.	81.8 (305)			
Sentence Length	45.4 (305) 46.7 (710)	44.1 (507)		
Structural				
Line Length	71.4 (305) 70.1 (710)	68.8 (507)		

Content specific				
Keyword Extraction	85.7 (305) 84.4 (710)	81.8 (305) 80.5 (507)		