

Figure 1
Top Terms by TF-IDF

Term	80th Percentile TF-IDF
boeing	0.410
plane	0.348
safety	0.272
flight	0.268
airline	0.233
pilot	0.183
issue	0.182
would	0.154
aircraft	0.153
software	0.145

Figure 2

Pearson Correlation of TF-IDF Features

Term	boeing	plane	safety	flight	airline	pilot	issue	would	aircraft
boeing									
plane	0.08								
safety	-0.56	0.10							
flight	-0.37	0.26	-0.04						
airline	-0.57	0.23	0.15	0.77					
pilot	-0.30	-0.25	-0.23	0.08	-0.19				
issue	0.33	-0.66	-0.21	-0.47	-0.46	-0.19			
would	-0.23	0.04	0.49	0.21	0.52	-0.41	-0.30		
aircraft	-0.35	0.14	0.54	0.22	0.33	-0.15	-0.23	0.38	
software	0.92	-0.15	-0.52	-0.31	-0.58	-0.26	0.47	-0.27	-0.29

Figure 3
Logistic Model Variables

Term	Coefficient	Relative Importance
boeing	0.954	1.000
safety	-0.457	0.479
software	0.393	0.412
pilot	-0.338	0.354
plane	0.213	0.223
airline	-0.129	0.135
would	0.090	0.095
issue	-0.087	0.092
flight	-0.069	0.072
aircraft	-0.055	0.057

Figure 4
Model Score Output

Document	y-value	Predicted P(y=1)	% > Pop
Doc1	1	0.424	27%
Doc2	1	0.359	8%
Doc3	1	0.474	42%
Doc4	1	0.486	46%
Doc5	0	0.277	-17%
Doc6	0	0.286	-14%
Doc7	0	0.259	-22%
Doc8	0	0.294	-12%
Doc9	0	0.269	-19%
Doc10	0	0.287	-14%
Doc11	0	0.297	-11%
Doc12	0	0.289	-13%
Test 1	0	0.281	-16%
Test 2	1	0.414	24%