be set *a priori*. The corresponding classification results are summarized in Table 1. With the exception of the result achieved with logistic regression and the tf-idf representation, different combinations of base classifiers and representations performed nearly identically. Specifically, it can be seen that approximately 79% of tweets were correctly classified. This is a rather remarkable performance considering the brevity of tweets and the fact that some of the most informative (in the context of the classification task at hand) words were not used for classification. It is insightful to notice the consistently higher accuracy attained for control (84-85%) rather than ASD tweets (71-74%). We explored this finding in more depth by manually inspecting misclassified messages. Within the ASD data corpus, we found that the main source of classification error lied in the absence of any ADS-specific information in very short tweets, after the removal of our four keywords used to construct the data set in the first place (e.g. "it is the adhd, oops!"). This is a highly comforting finding since of course in any practical application these keywords would not be eliminated, thus improving classification performance dramatically.

Table 1: A summary of the classification results (ASD-related vs. non-ASD-related tweets). Each cell (corresponding to a combination of a classification method and a representation) shows the associated confusion matrix. The first row/column of a confusion matrix corresponds to the control class and second row/column to the ASD class (thus for e.g. using the term count representation, naïve Bayes correctly correctly classified 84% of control tweets and 71% of ASD tweets).

	Confusion matrix		
Representation	Binary	Term count	tf-idf
Naïve Bayes	0.84 0.16 0.29 0.71	0.84 0.16 0.29 0.71	0.84 0.16 0.38 0.62
Logistic regression	0.85 0.15 0.26 0.74	0.85 0.15 0.27 0.73	0.03 0.97 0.04 0.96