CS838 Project
Dan Howe
Keith Funkhouser
1 March 2017

Our documents were course reviews posted on RateMyProfessor.com. From these we attempted to extract adjectives.

> *I found Eric Bach's class to be extremely difficult!*
> *The TAs' office hours were useful in learning the course material.*
> *Exams in CS367 tested only for a basic comprehension of Data Structures.*

We found 2553 adjectives in our collection of documents, and added 2665 negative instances, which were words other than adjectives. Set I contained 386 documents, in which there were 1660 adjectives. Set J contained 192 documents, in which there were 893 adjectives.

After the first stage of cross-validation, we found that a decision tree performed the best. The cross-validation on set I resulted in a precision of .901, a recall of .929, and thus an F1 score of .915.

After working to improve the performance of the classifiers, we elected to continue using the decision tree. On set J, we found the final version of this classifier (classifier X) to have a precision of .918, a recall of .935, and thus an F1 score of .926.

Since the classifier already met the minimum requirements for precision and recall, we did not add any post-processing rules. As a result, there was no difference in performance between the classifier X above and the final version of our classifier (classifier Y).