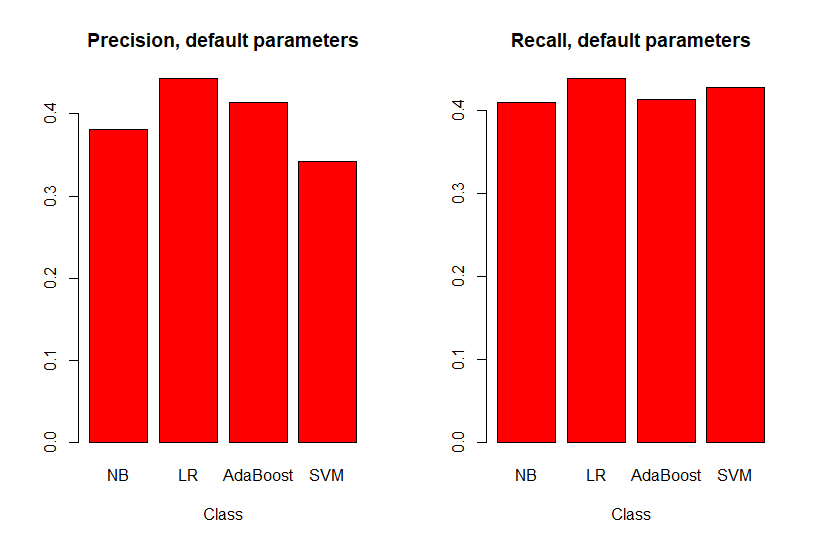
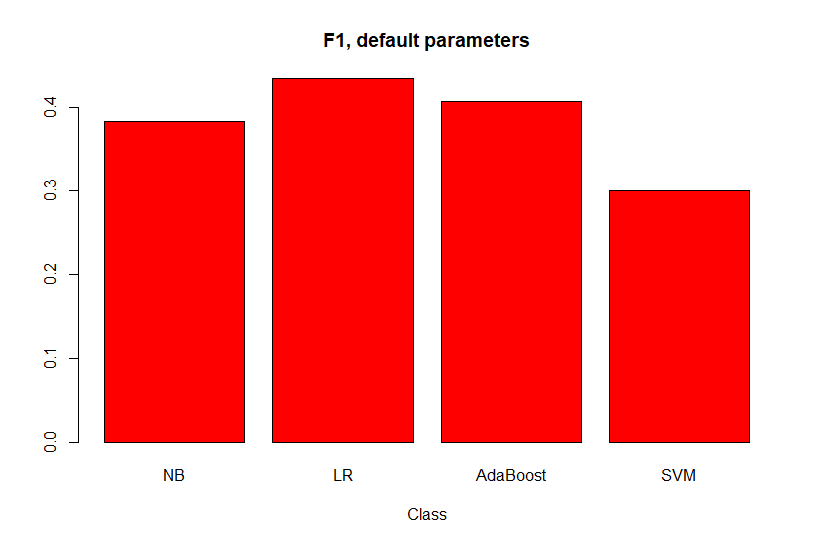
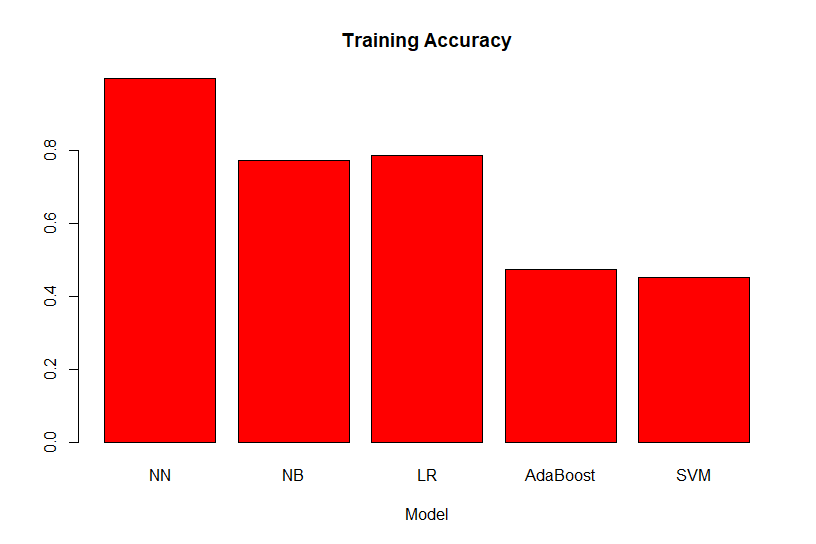
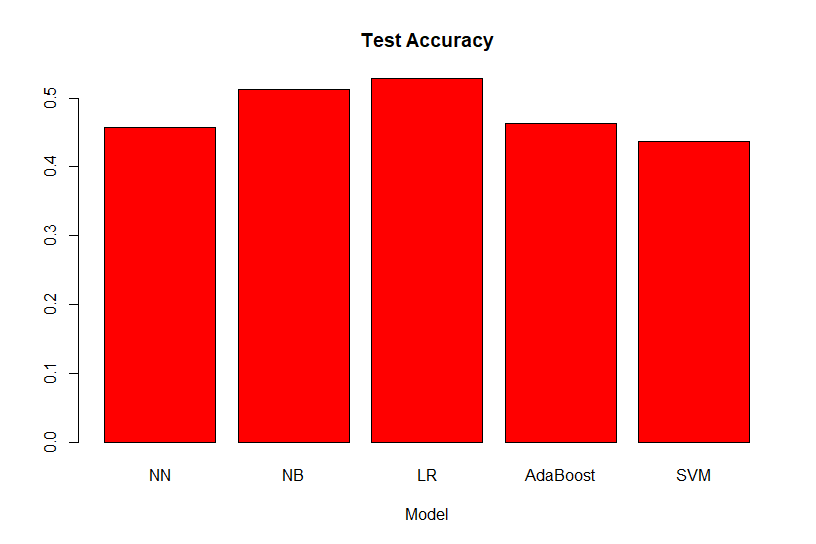
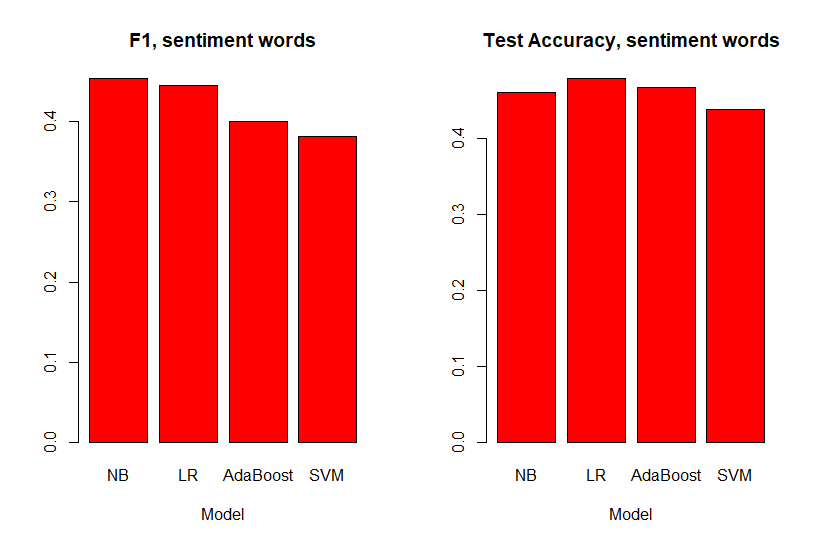
Casey Carr

Initially, 5-fold cross validation was implemented on the classifiers to find precision, recall, and the F1 score. This was done using the default parameters for each learning algorithm. Cross Validation was not performed on the MPLClassifier because of the computational costs required. Logistic regression had the best performance with regards to the F1 score.

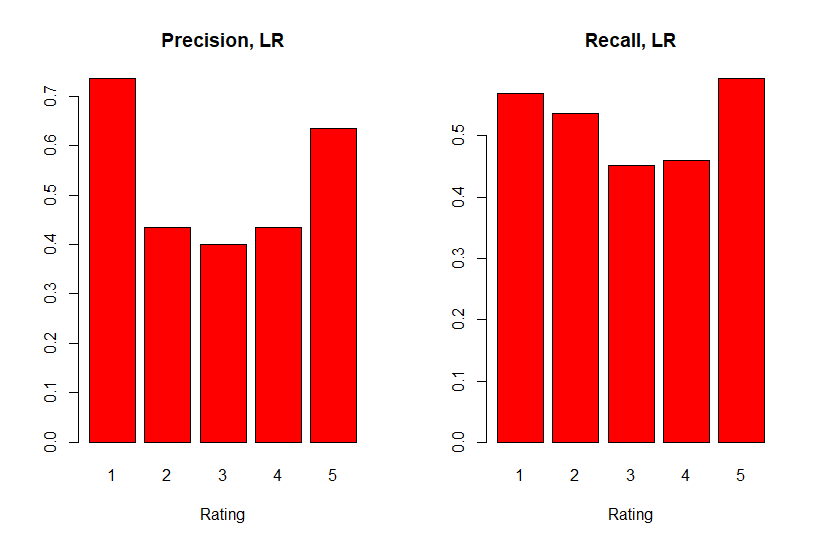


Next, the accuracy of the model was accessed on the training and test data. As expected, the neural network had the best performance for classifying the training set ratings (99% correct classification). However, the fit performed significantly worse on the test set. This was likely due to overfitting. Logistic regression remained the best fit when looking at model accuracy on the test set.



To improve performance, a list of words that express sentiment were used to create the vocabulary that the vectorizer would use. As a result, the F1 scores of every model increased. However, the test accuracy of the fits with the sentiment words was worse than the original fit. The use of sentiment words reduced the vocabulary of the vectorizer from roughly 22,000 to under 7,000. The reduction of predictors could have caused our models to underfit the data.

To better understand our data, we looked at metrics for the individual ratings. Precision and recall were generally higher for ratings of 1 and 5 than scores of 2-4. Intuitively, the outliers (ratings of 1 or 5) should be easier to differentiate than scores in the middle.



Lastly, the parameters of each model were tuned to improve the accuracy of the model. For neural networks, a slight improvement was found by adding a hidden layer and increasing the number of perceptrons in each layer ((100) → (1000, 100)). The adaboosting fit was improved by using logistic regression as the base estimator and increasing the number of estimators. The biggest improvement came from using a linear kernel for the SVM. The accuracy of logistic regression (relative to NN) suggests that simpler models can perform better on this data. The default parameters for naive Bayes and logistic regression produced the best fits.

