An 80/20 split is sufficient for our data. The purpose of splitting the dataset into a test set and a training set is to see if the model developed using the training set generalizes for other values. By adding a test set separate from the training set, the model is evaluated on rows it did not train on, i.e. rows it has never seen before. Splitting 80/20 is problematic in a few cases: first, if there are not enough data; second, if the data have time dependence; third, if the split that was chosen randomly happens to show performance that is better or worse than the actual performance of the model. None of the three cases apply to our data. The first case, that there is enough data, does not apply. Our data has 1,728 instances, which means that 346 instances are used in the test set. With 346 instances, lucky guesses have little effect; each instance represents less than 1 percent of the total result. The second case, that the data have time dependence, also does not apply. Time dependence exists in venues like finance, where stock price is a function of time. In our data set, new car models are not time dependent on old car models; it is not as if the number of doors that car models have increases over time. The third case, that the randomized 80/20 split happens to be a lucky one, is solved by running multiple randomizations. We run multiple randomizations to see how the model performs when trained and tested on different data to avoid the influence of one lucky guess.