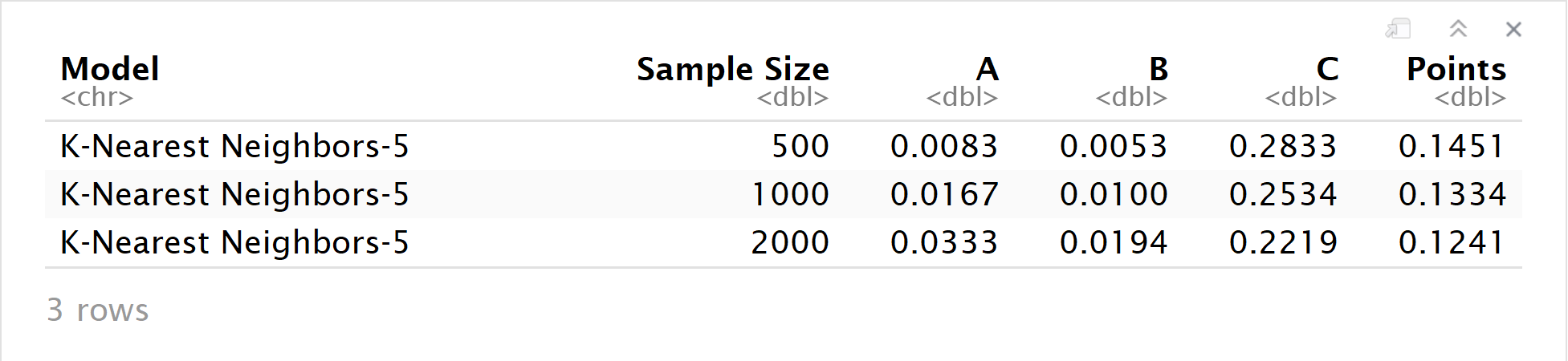
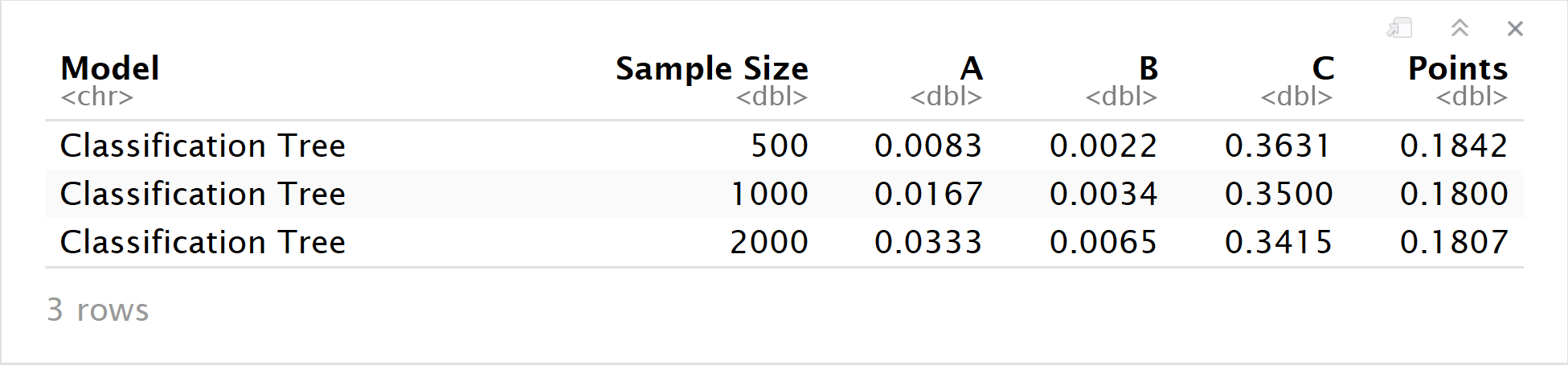


First we discuss the Part A scoring result, since we use the same n.values of 500-1000-2000, so we must get the same A scoring of 0.0083, 0.0167, 0.0333. The model score will be puninshed for using more data to learn and then we can see which one a better model is. For the B scoring part, the time to run the model increases as we try larger samples from 500 to 2000. This make sense since the computer need more time to process more data. More than that, we see the higher accuarcy as long as we use more data to predict the test data set. The Points decreases with the larger sample size so maybe we should try to use a larger sample size to decrease our score and to find the lowest points level.



Compared with the KNN – 10 tests above, we didn’t see much difference in the running time for KNN – 5 Model, this maybe we still use a relatively small sample size and the running time might increse if we use more group numbers. For the accuracy part, we see a increase in the accuarcy compared with KNN – 10. In general, the KNN – 5 have a better performance compared with KNN- 10 under the sample size of 500 – 1000 and 2000.



The Classification Tree is a relatively “quick” model, all the B scores are very samll compared with other model. However, compared the accruarcy status with other models, the Classification Tree did’t perform so well. And we see that there are still a lot of room left to trade off between the running time and the accuracy. Maybe for the next step we should just increase the sample sizes. Allowing the running time to a acceptable status and we get the better prediciton result.