**Lab: Validating a Model**

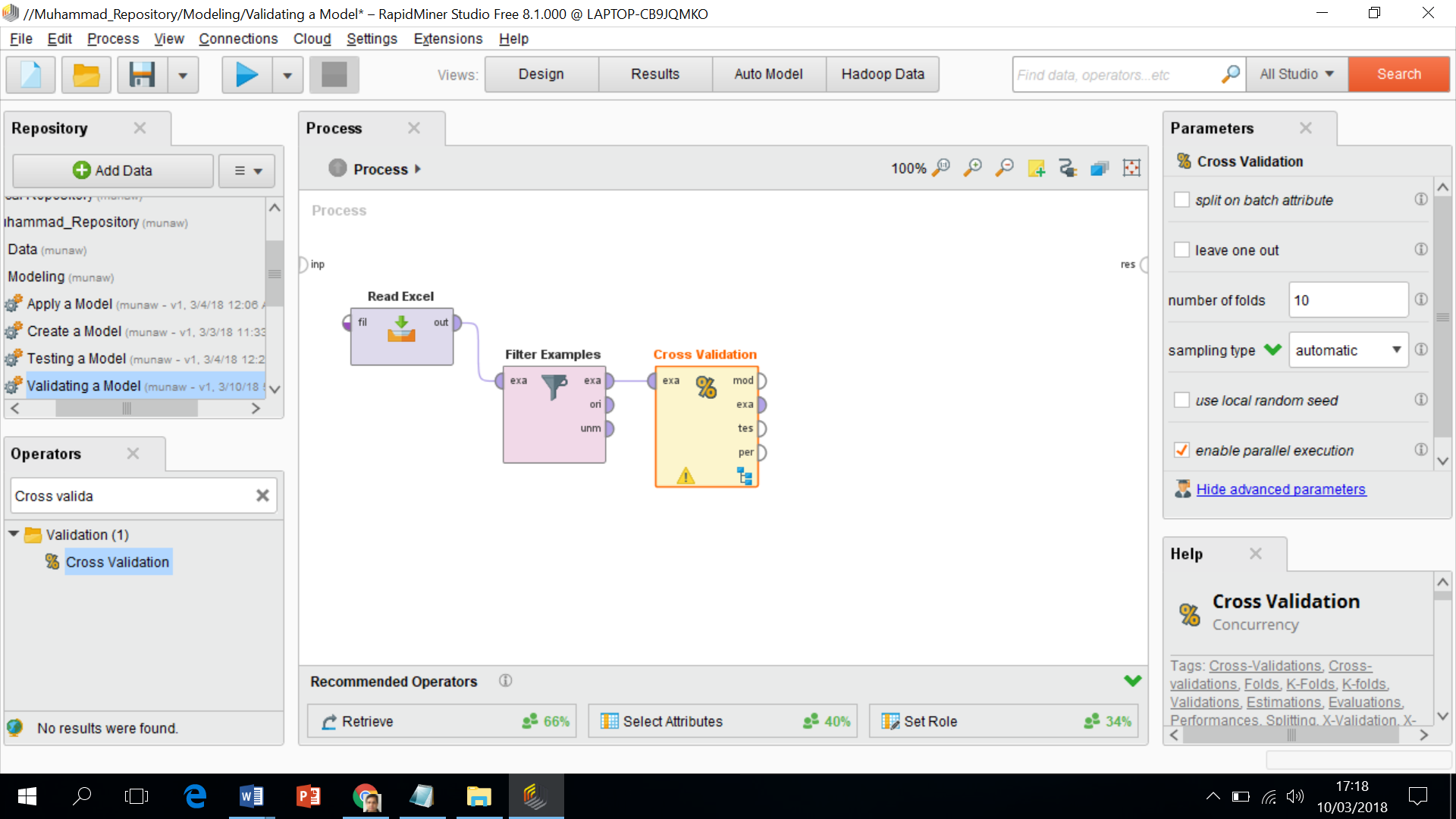
Download the excel sheet for the data set named as “customer-churn-data” from model. Follow the steps to prepare the data for the training set.

1) We are going to test the already developed model based on unknown data by using two methods called as Split validation or Cross Validation

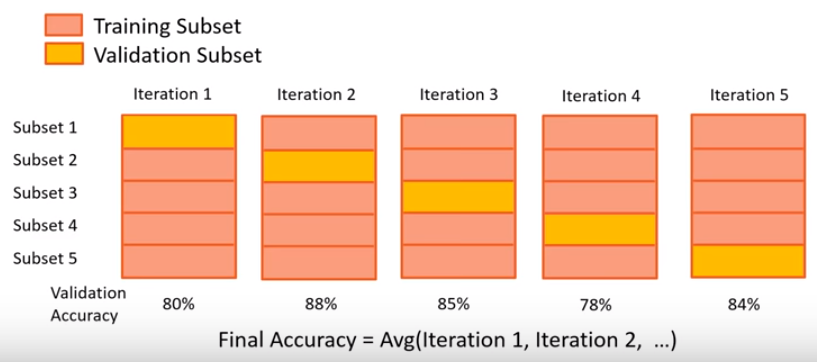
2) We discussing “Cross Validation” in this exercise.

3) Delete all the operators except the Read Excel and Filter Example operators and remove the output port also.

4) Find the “Cross Validation” Operator and drag into the process window and it looks like as mentioned below



5) What is meant by number “10” in the parameter pane? This means that It will break the data set into 10 small subsets of data. It will build a model on 9 sets of data and one will be used for the testing. This process Iterate to cover all ten parts one by one.



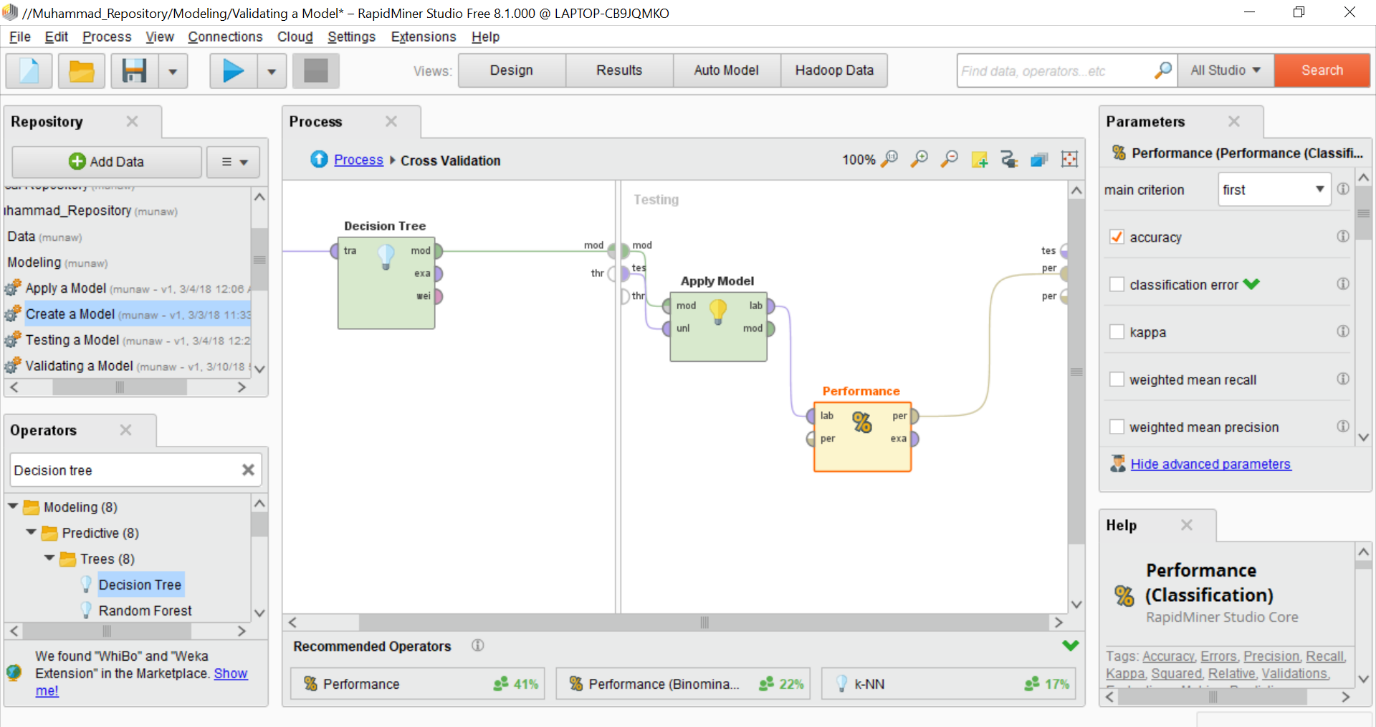
6) The blue hierarchy symbol pointed by arrow in the operator “Cross Validation” means that there is sub process going on under it. Double click on that and we will see two windows. There are two sections in this window, one called as the **training** and other called as **testing**.

7) Drag the “decision tree” operator into the training data window and build a model based on that. Connect the input and model ports.

8) In testing section, Drag the “Apply Model” and “Performance” operators in the second section. Model ports connect with the “Apply Model” Operator and Testing port connect with “unl” port of “Apply Model”.

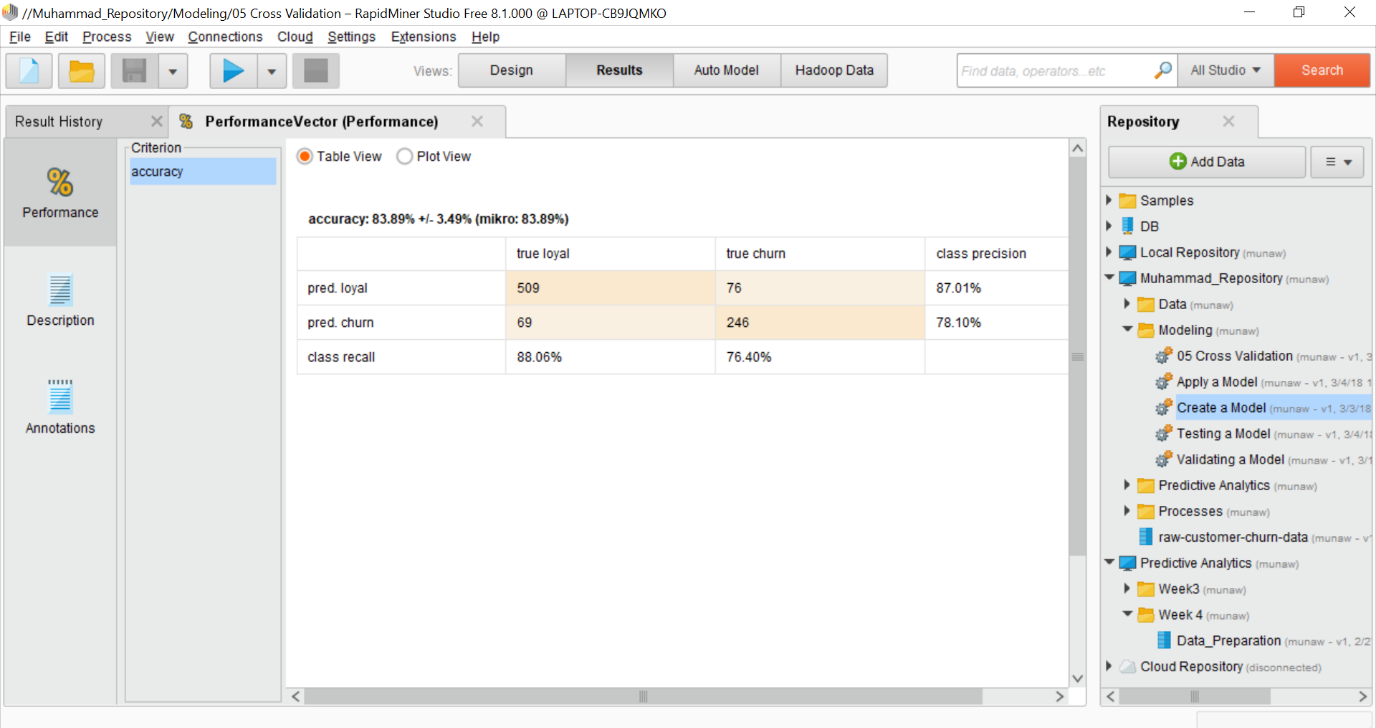
9) “Per” port of the Performance operator will connect with the output port “par”.

10) This finish work in this subprocess and it looks like as mentioned in the window



11) Now move outside the sub process, In main process window, connect the “per” port to the main output result port.

12) Save this first before running of this process and give a different name if you like. The result will be like as mentioned below



13) This shows a confusion matrix. The accuracy is reduced than the previous testing case and +- signs shows the standard deviation.

14) If the standard deviation value is smaller, the model will be more stable. The stable model high accuracy than the unstable model.

15) Pred. loyal means predicted loyal and similarly for other label.

16) We have 88.06% loyal customers and 76.40% churn customers. A better way to look at the diagonal way and it shows the true loyal and true churn. (509 + 246) / 900 = 0.8389 = 83.89 % correct predictions. For churn customers, 246 / (69 + 246) = 0.78 = 78%

17) We can predict 76.40 correct prediction out 100 churners.

If you like to explore further, you can watch this youtube video as mentioned below

* <https://www.youtube.com/watch?v=Hw3tVDZ4Pmo&list=PLssWC2d9JhOZLbQNZ80uOxLypglgWqbJA&index=11>