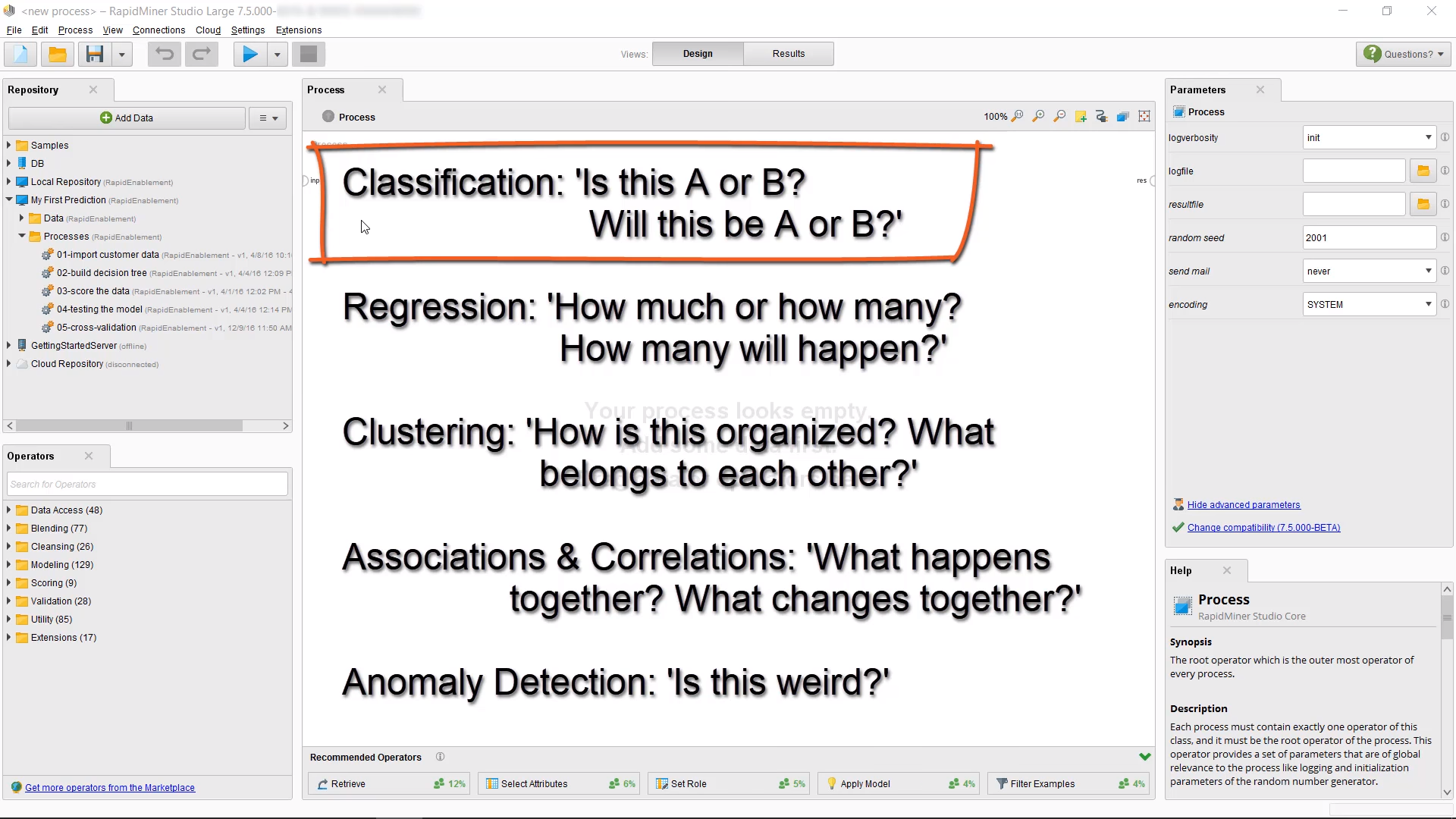
**Lab: Validations of Models**

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) To decide about the model selection, we need to analyse the results obtained from the various models.

2) First you should decide about the task bucket before the analysis.

3) There are different types of problems as mentioned by the screen shot and this tutorial will focus on the classification problem.

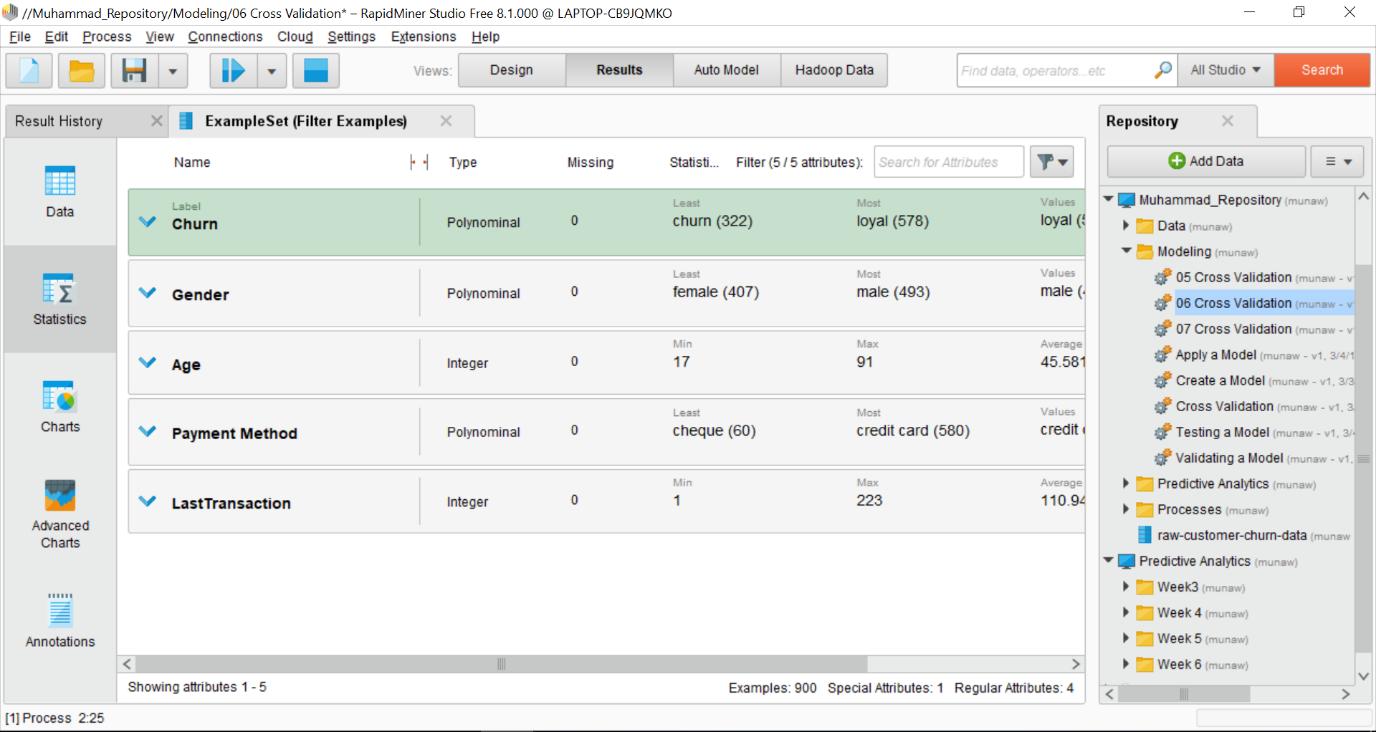


4) Open the previous cross validation model developed in the last week tutorial in rapid miner.

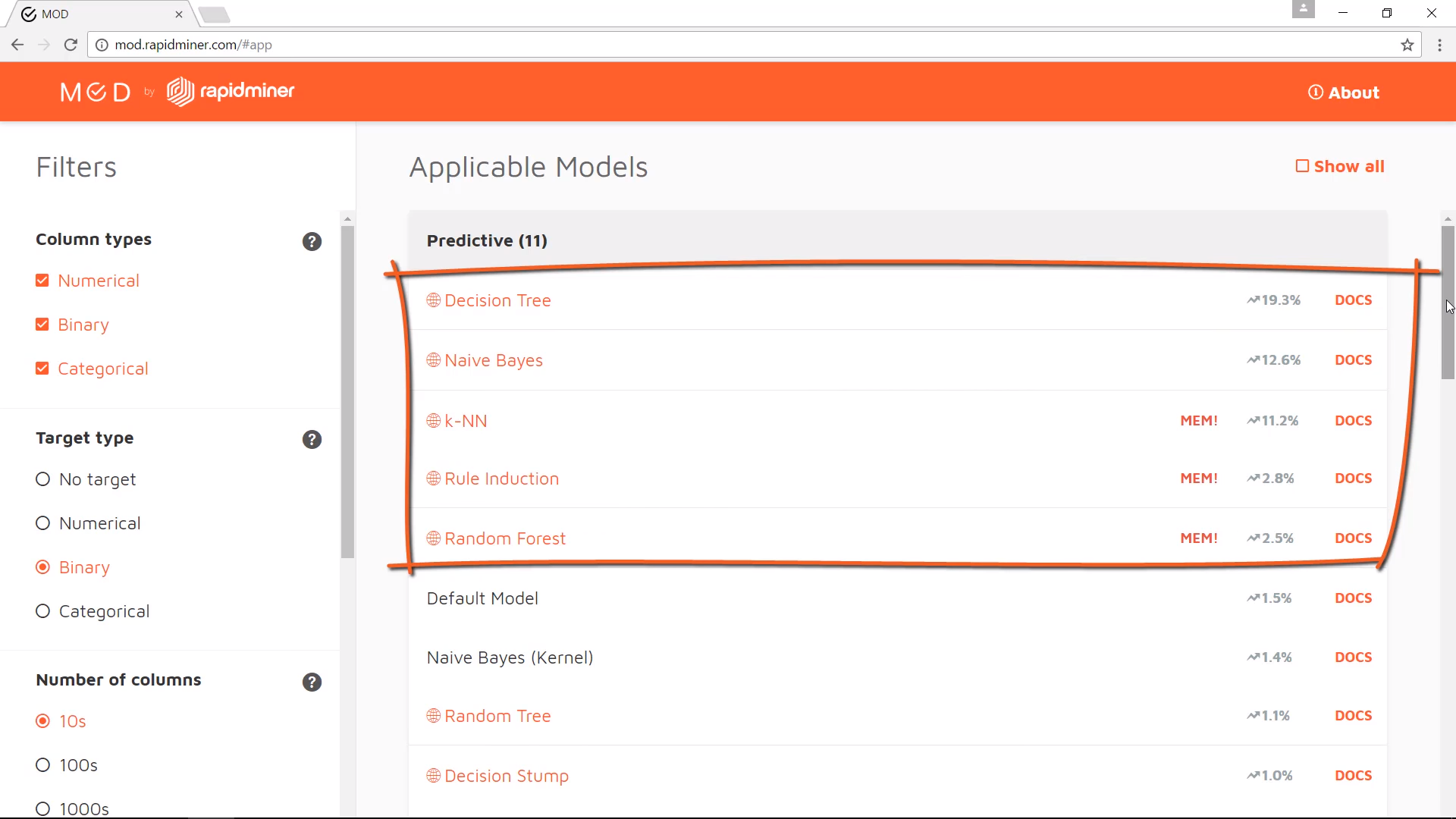
5) If you would like to see the data after two steps, you can use “Breakpoint After” that can be displayed by the right click on the process “Filter Examples”.

6) We can see the data after the execution after these steps and the data will be displayed after two steps. This facility is like “pause” rather than “stop”.

7) Let’s discuss the data types of the attributes used in the Customer churn data file as mentioned below. There are no missing values in the data sets as we applied the checks on the previous tutorial. The run process button is changed to pause and run after the check in option “Breakpoint After”. This method is used to check the behaviour of the various processes in the rapid miner. After completion of data review, remove the breakpoint after check in again.



8) Open this website as mentioned, “http://mod.rapidminer.com” and click on Use MOD button. You can see the list of applicable models on this website.

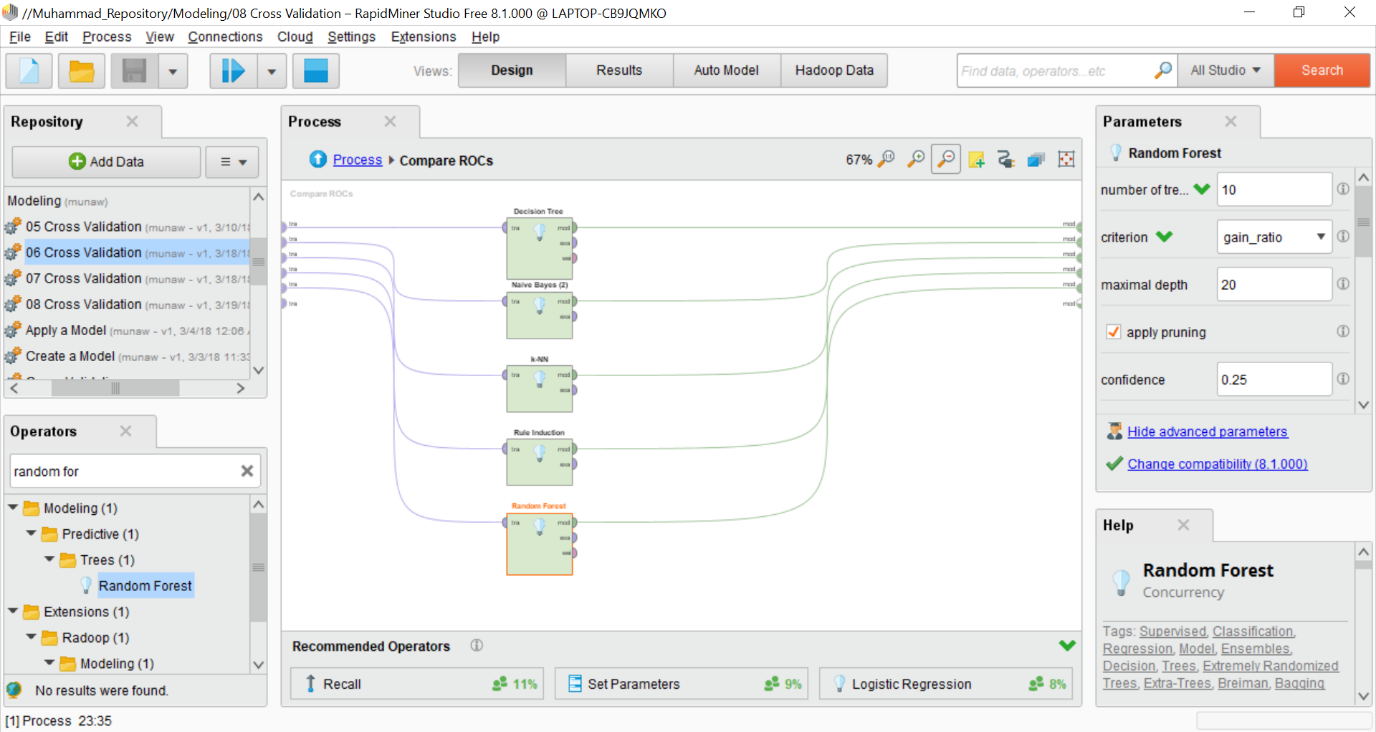


9) Disable the “Cross Validation” operator by right click and uncheck the “Enable” option and remove from the list of rapid miner active “Enabled” operators for a while.

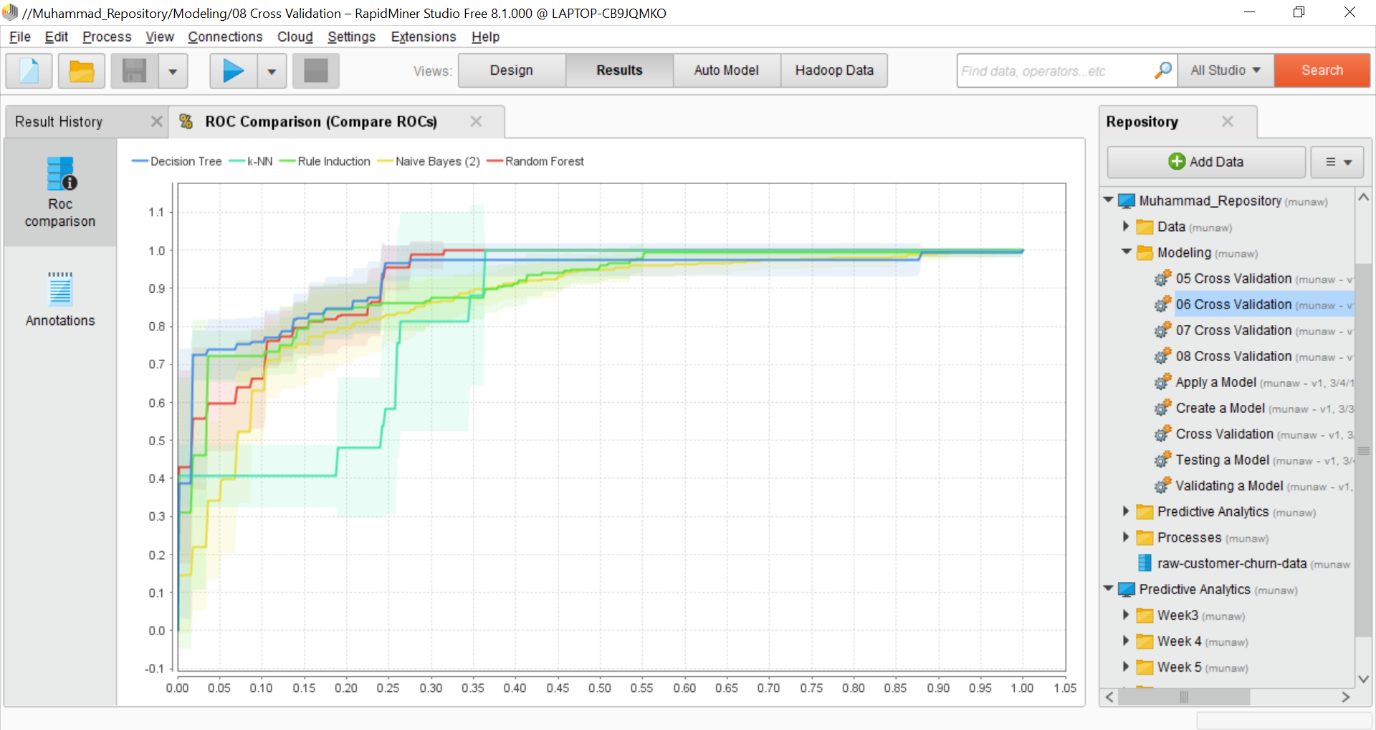
10) Drag the “Compare ROCs” operator into the process window and link with the active operators and link the roc port only. **ROC means Receiver Operator Characteristics.**

11) Rapidminer automatically perform cross validation on various models.

12) Double click on the relationship icon on the ROCs compare operator and a new sub process will open. Drag five operators in this sub process window, “Decision tree, Naïve Bayes, Knn, Rule Induction and Random Forest” and the it should like as mentioned below



13) Run the process and the resultant window will be displayed as mentioned below

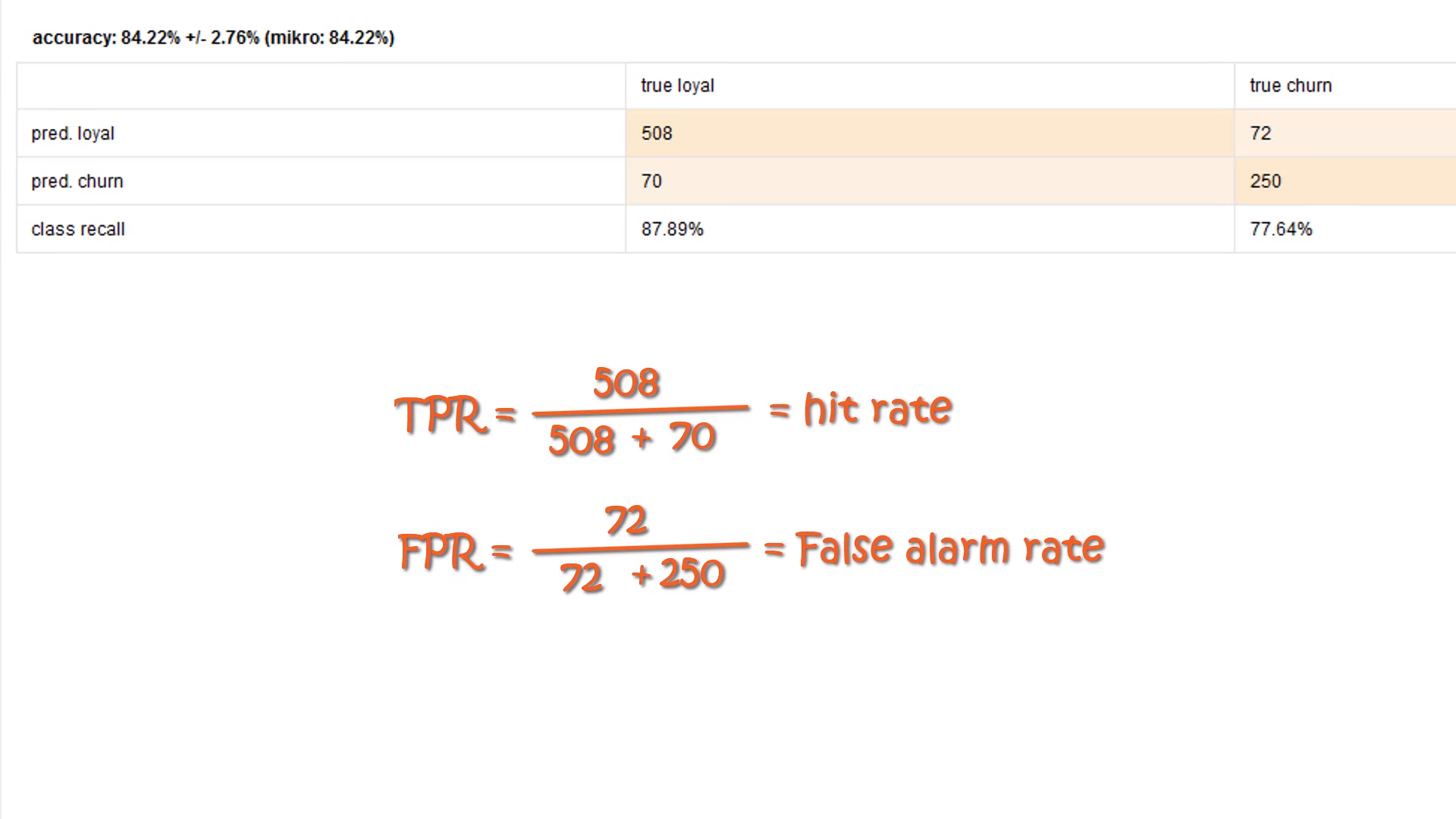


14) This is called as the ROC graph and these curves are obtained from the models. The true positive rate is plotted on the y-axis and false positive rate is plotted on the x-axis.

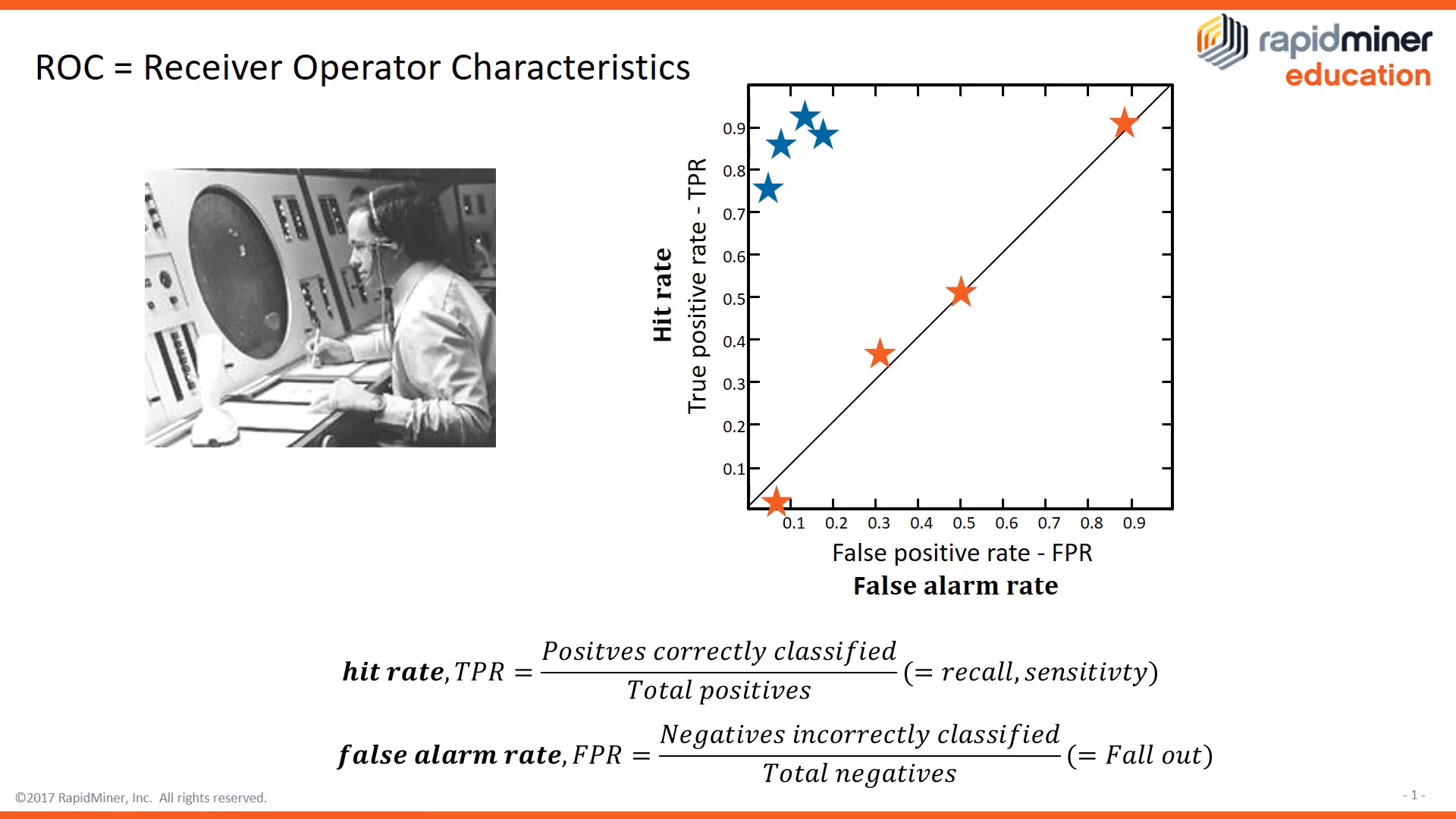
15) To understand these processes, reactive the “cross validation” operator again.

16) Use “Multiply” operator to link with the “cross validation” operator for the same data set to use in the cross validation. Link the performance port of the “Cross Validation” operator to result port.

17) Run the process again and you will see a confusion matrix as mentioned below



18) True positive rate and false positive rate are mentioned below

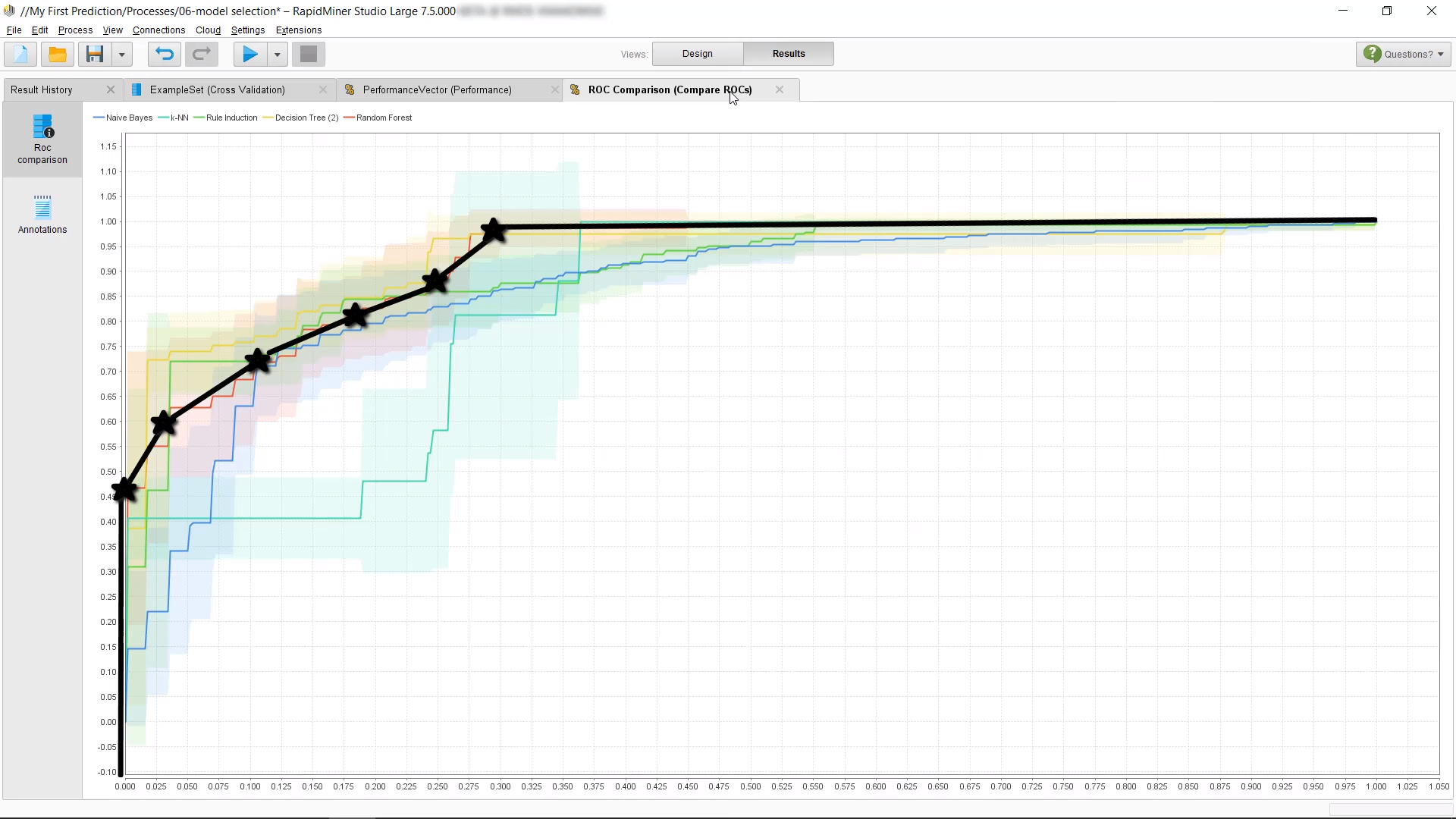


19) To retrieve the confidence values, we will click on the “Cross Validation” operator and link the link the “exa” port of the Performance operator to the “tes” resultant port.

20) In the main process window for the “Cross Validation” operator, join the “tes” and “per” ports with the resultant port.

21) After completion of these steps, Run the process and the confidence of the values changed and you can analyse the data generated by the rapidminer.

22) The ROC graph should like this as mentioned below and black star liked is used to explain the graph and It will not appear on your system.



23) Regarding the understanding of the ROC graph, I recommend to see the last two minutes of the video to understand the insight about these operators.

24) This tutorial provides you an understanding to evaluate the performance or confidence on the results based on various models.

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

* https://www.youtube.com/watch?v=C8Ko3-2f-pA&index=12&list=PLssWC2d9JhOZLbQNZ80uOxLypglgWqbJA