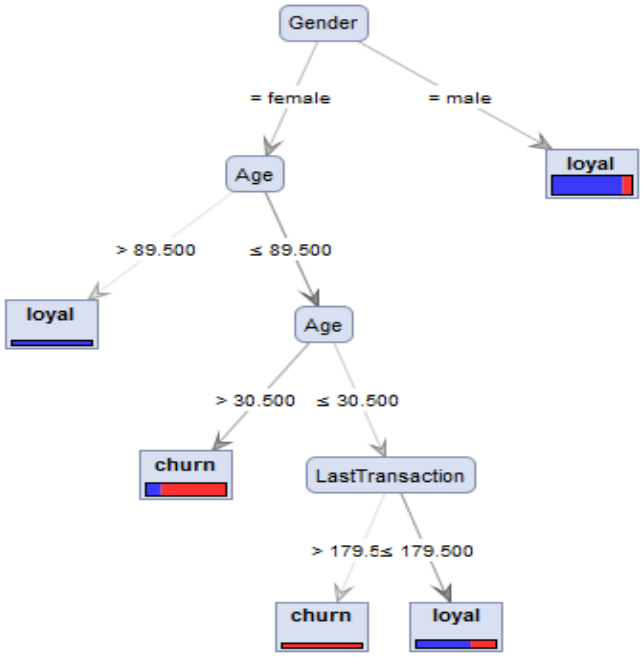
MIS 304: Using and Managing Information Systems

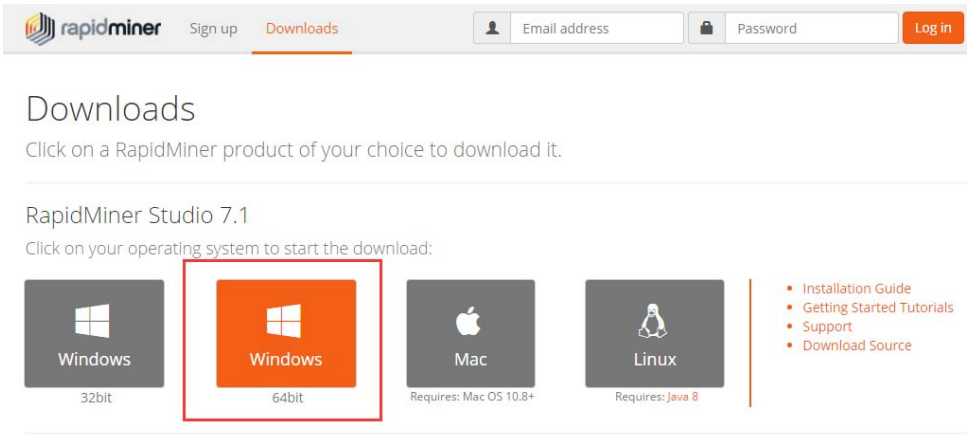
Lab Session 5: Business Intelligence Analysis – Decision Tree

The goal of this lab is to help you get started with RapidMiner for Business Intelligence (BI) Analysis. RapidMiner is a very popular predictive analytics platform which uses a graphical user interface. It brings the power of predictive analytics to the business user, with **NO** programming required. RapidMiner supports many different BI techniques, but we will focus only on decision trees here (Read: https://en.wikipedia.org/wiki/Decision_tree).

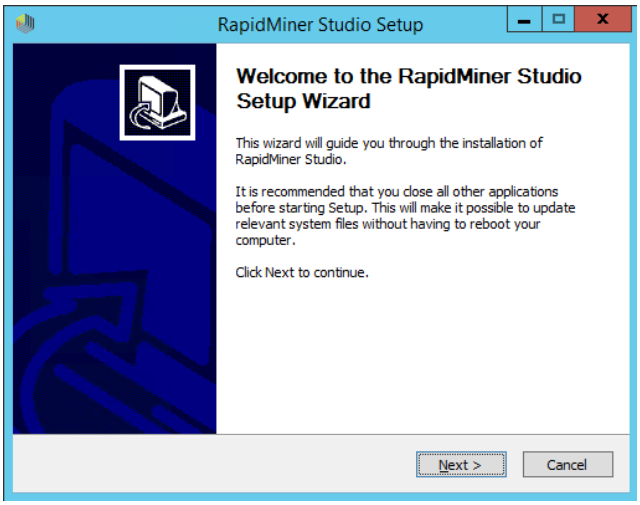


We will first install the RapidMiner Starter Edition. Then we will analyze a customer churn data with RapidMiner and build a decision tree model.

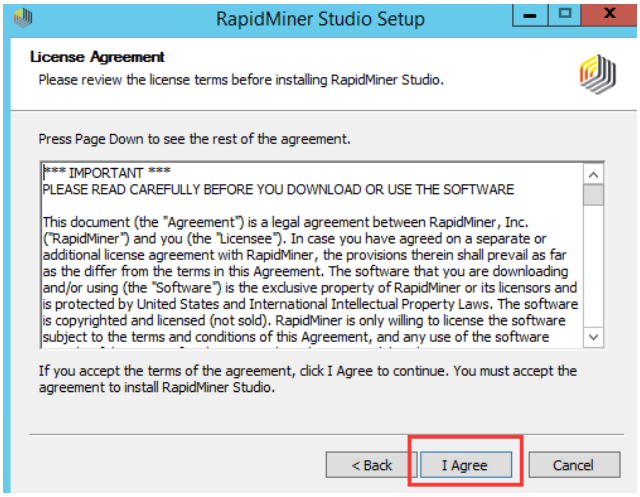
1. Install RapidMiner
   (1) Download RapidMiner Starter Edition (free):
   https://my.rapidminer.com/nexus/account/index.html#downloads



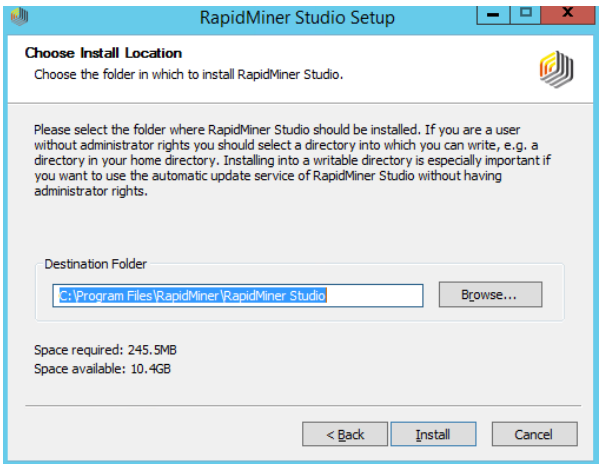
(2) Double-click on the file that was downloaded above. You may have a question asking you if you really want to run the file. Click "OK". Then click "Next".



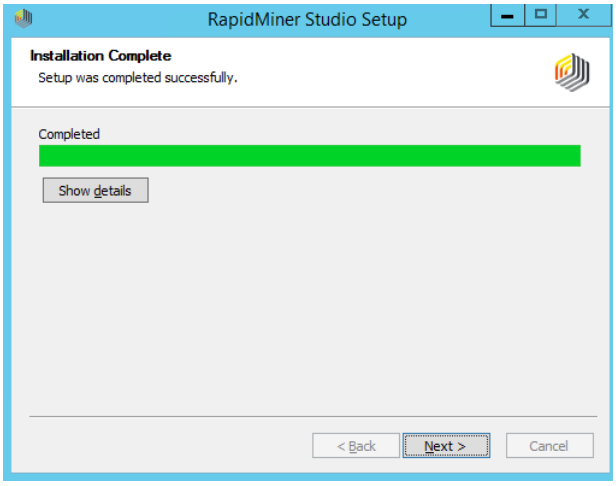
(3) Click "I Agree".

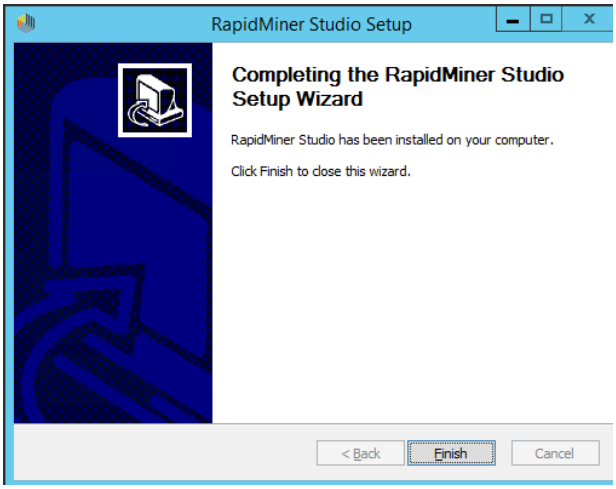


(4) Change where you install RapidMiner if you want. Click "Install" to continue.

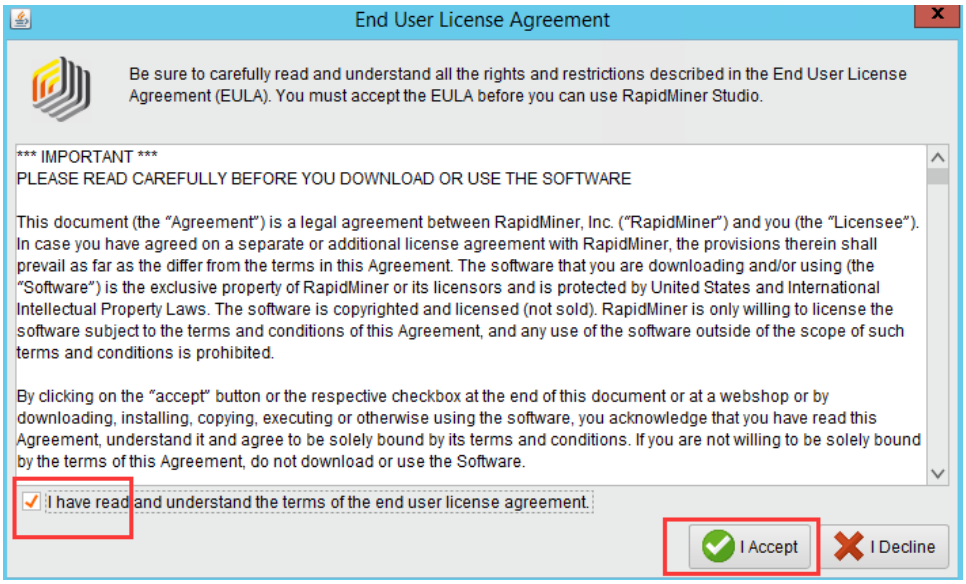


(5) Click “Next”. Then “Finish”.



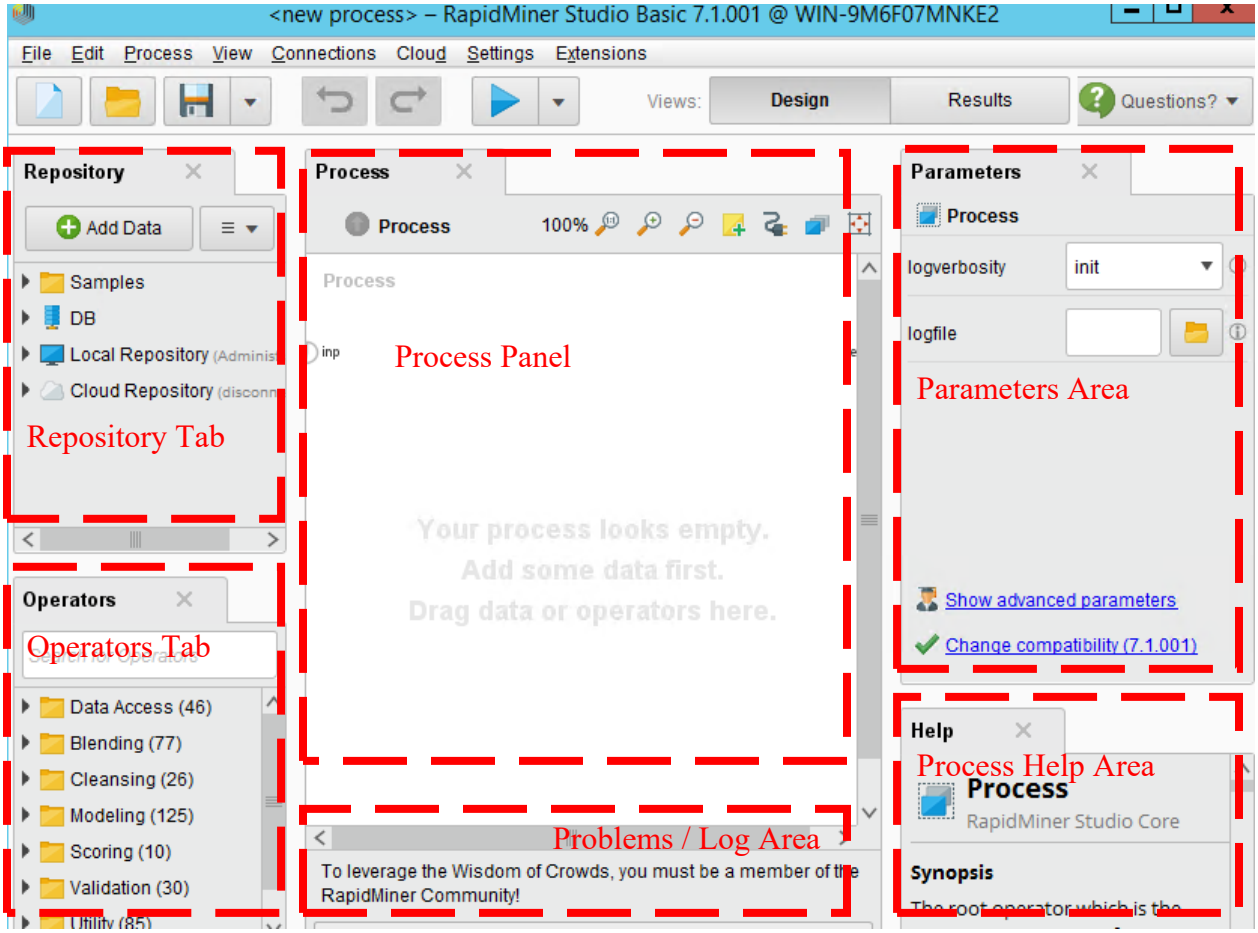


(6) Test if you can open RapidMiner.



(7) Click on "Continue using RapidMiner Studio Basic" (You can create your account later if you want to).

(8) The "new process page" should open. Below is a brief description of important areas in RapidMiner.



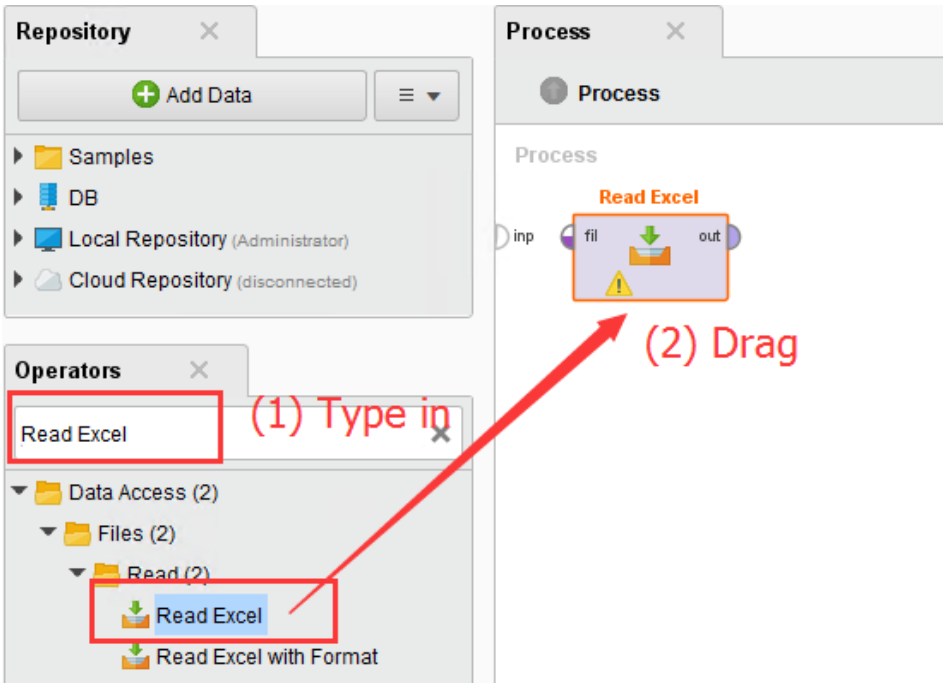
"Process Panel" is used to design any process you want to take on your data, for examples, data loading, classification and predication. Process is comprised of a set of connected operators, which can be understood as the building blocks of your BI analysis process.

2. Download **customer-churn-data.xlsx** from Blackboard. Take a look at the data file. Information about customers' gender, age, payment methods and last transactions are included in the dataset. We are trying to know if they are loyal customers based on these information.

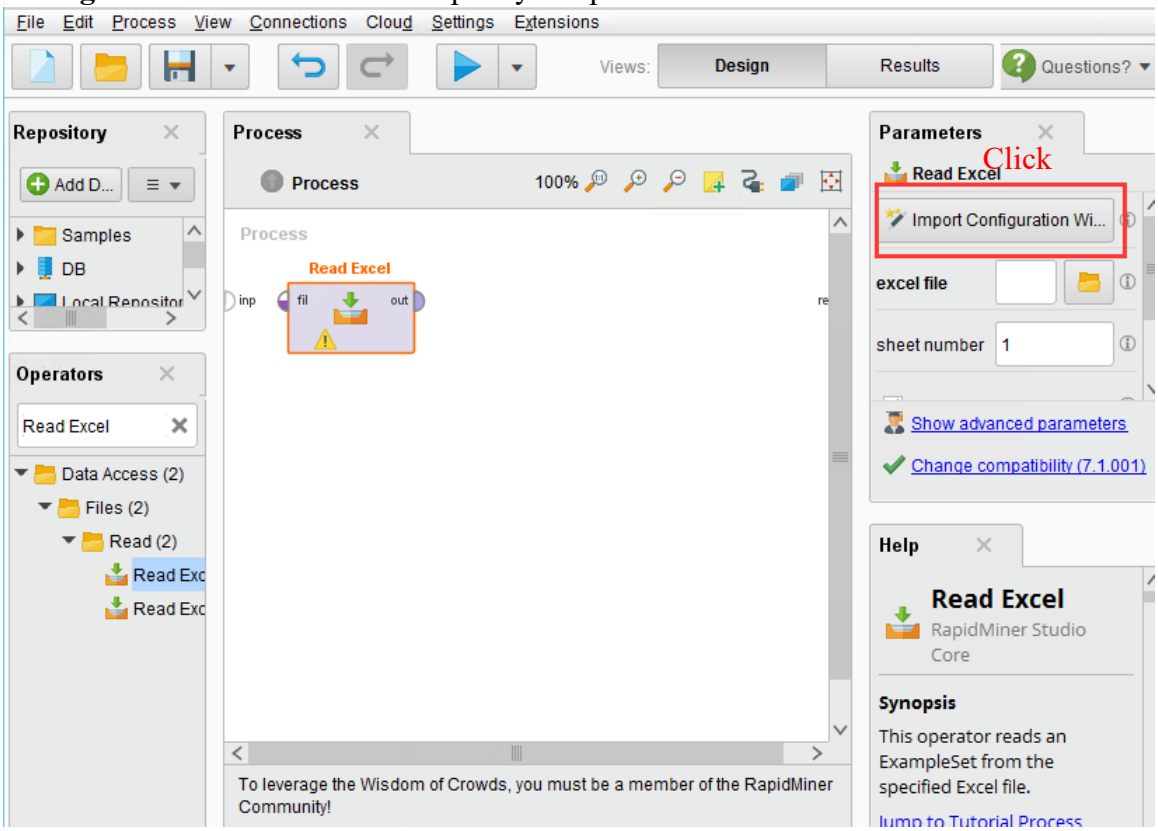
Gender	Age	Payment Meth	LastTransactic	Churn
male	64	credit card	98	loyal
male	35	cheque	118	churn
female	25	credit card	107	loyal
male	39	credit card	90	loyal
female	28	cheque	189	churn
female	21	credit card	102	loyal
male	48	credit card	141	loyal
female	70	credit card	153	churn
male	36	credit card	46	loyal
male	22	credit card	51	loyal
male	27	cash	137	loyal
male	22	cash	147	loyal
female	49	credit card	158	churn
female	24	cash	162	churn
male	45	credit card	55	loyal

3. Data analysis and decision tree building.
    (1) Import dataset into Rapidminer.

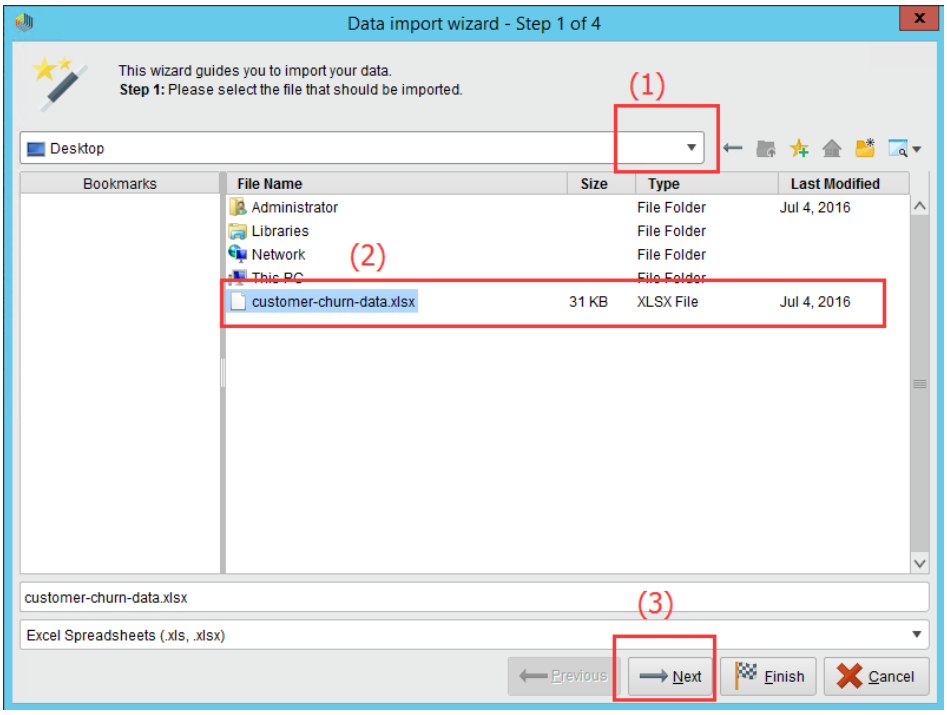
We need to choose the **operator** that imports from Excel files.
There are several ways to find the operator we are looking for: (i) Know which group the operator belongs to, and just expand the list until we find it. (ii) Start typing what we think the operator might be called into the filter field under the **Operators Tab**, so that the software will start showing suggestion as you type.
The operator we need here is called **Read Excel.**



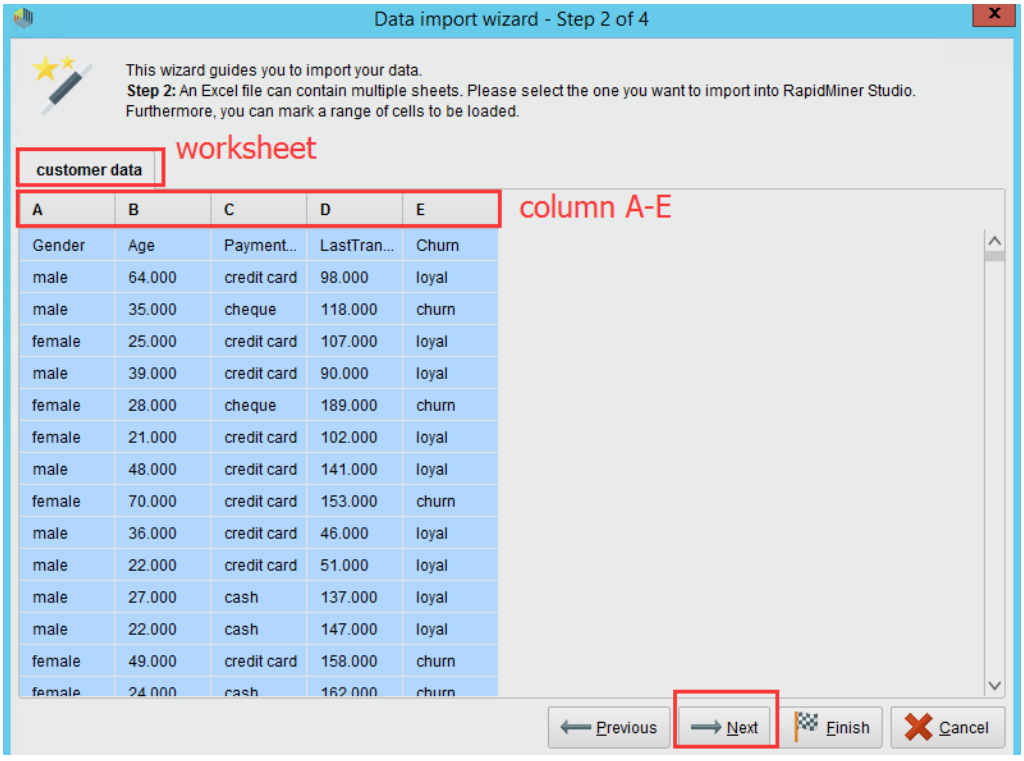
To import the data correctly we need to specify certain parameters, such as where the Excel file exists, which sheet(s) to extract from, columns, cell range etc. We will use the **Import Configuration Wizard** button to specify our parameters.



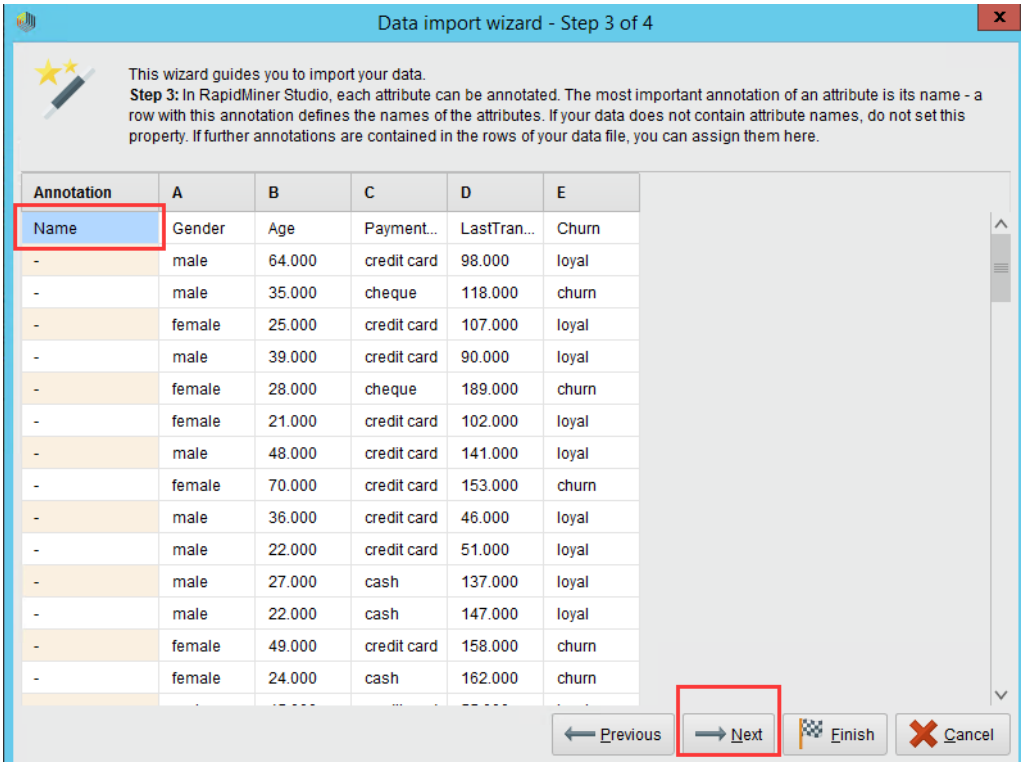
First, locate where the Excel file is. Once the required Excel file ***customer-churn-data.xlsx*** is located, click "Next".



Here we can choose the appropriate sheet and cell range. Our data resides in the "customer data" worksheet, and columns A through E are needed. The cells should be highlighted as shown in below. Then click "Next".

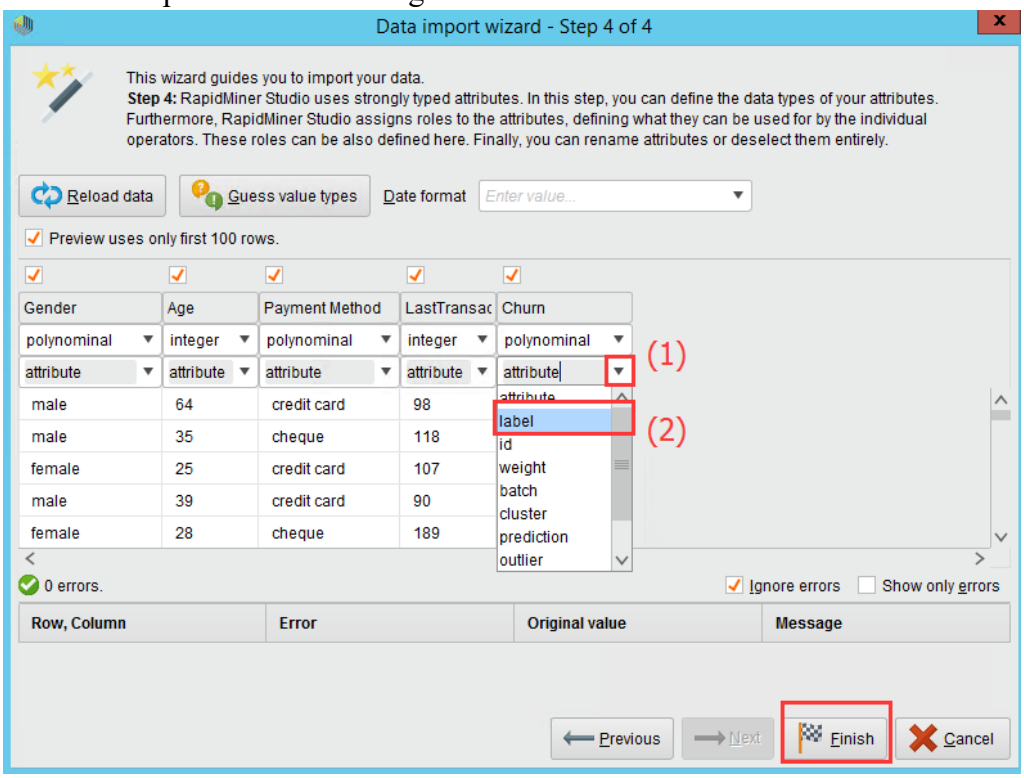


Then, make sure each attribute is annotated. Under "Annotation" choose "Name" for the column titles row as shown in below. Click "Next".

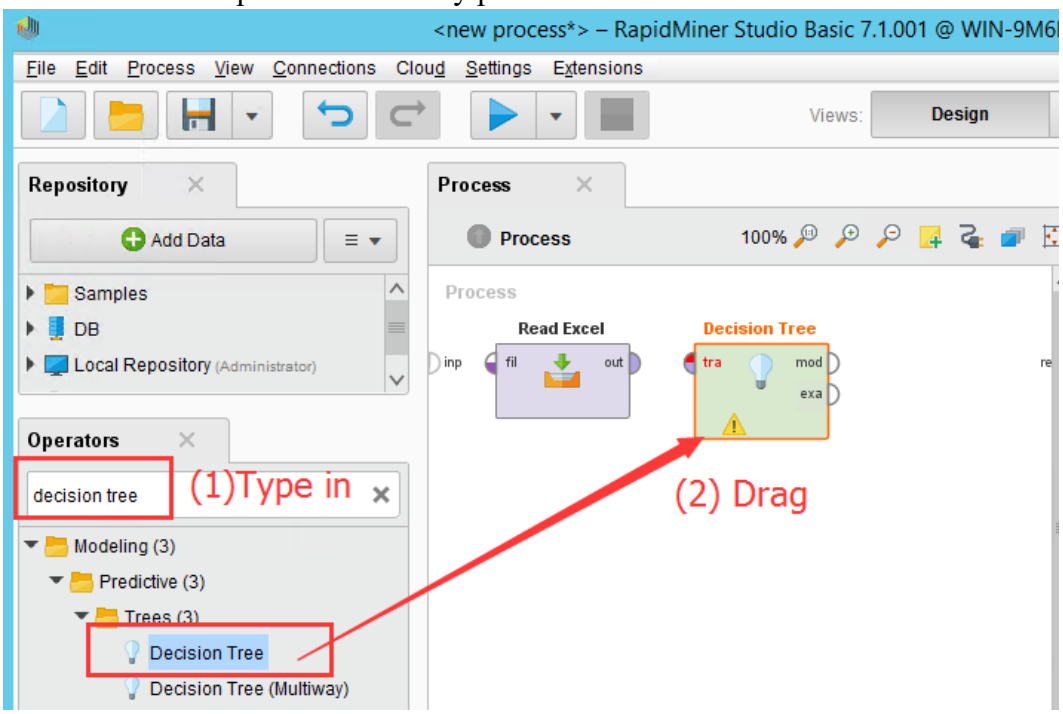


Now we choose which attribute (column) we are trying to forecast based on the Data we have. In our case that is "Churn".

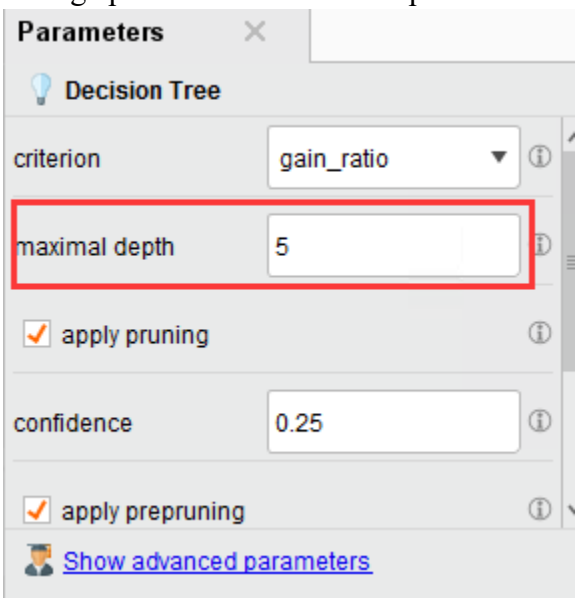
To do so, we go to "Churn", to the cell that says attribute, which is above our actual data, and from the drop-down menu change it from "attribute" to "label". Then click "Finish".



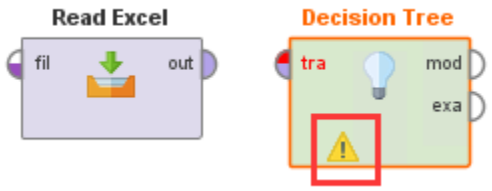
(2) Now that the "Read Excel" preferences are set up, we need to select the Decision Tree Operator to actually produce the decision tree we desire.



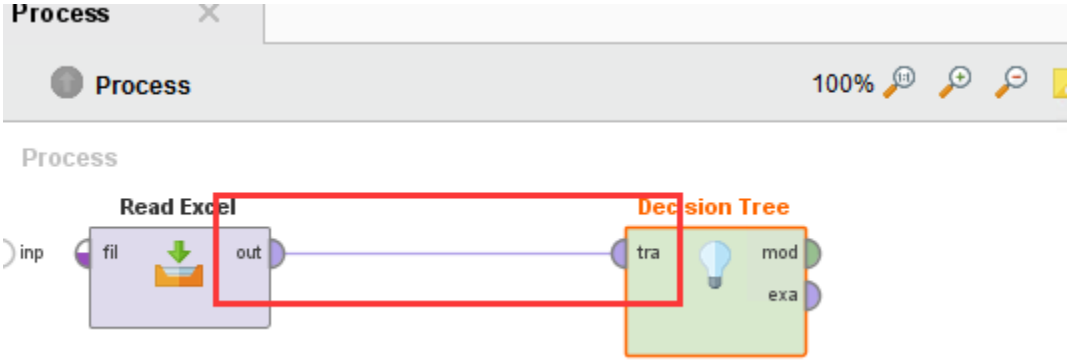
Change parameter "maximal depth" to 5 for the Decision Tree Operator.



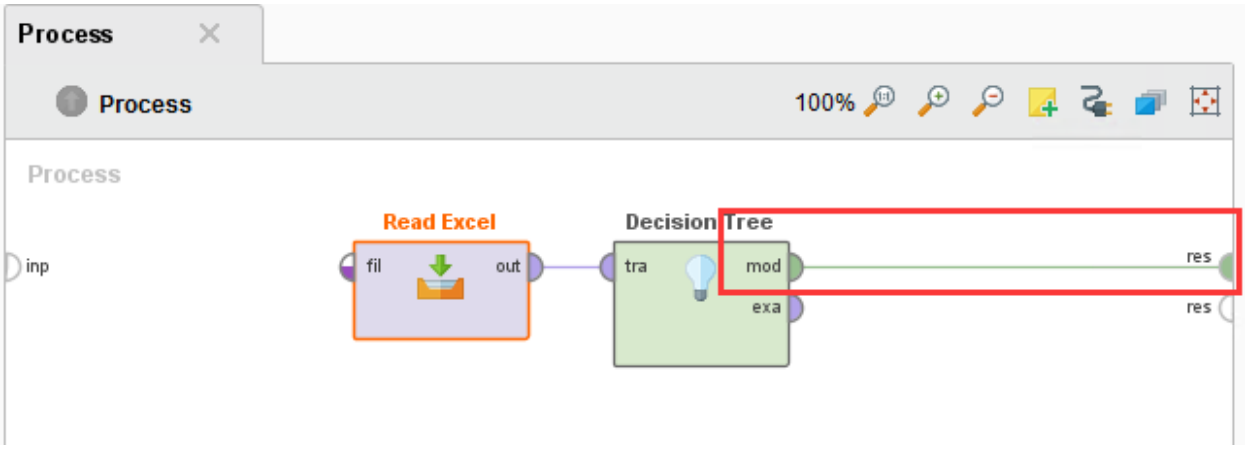
On each operator if you have a "yellow" triangle with an exclamation mark, that usually means something is wrong with your setup and needs to be rectified before proceeding further, otherwise your model will probably not run. For example:



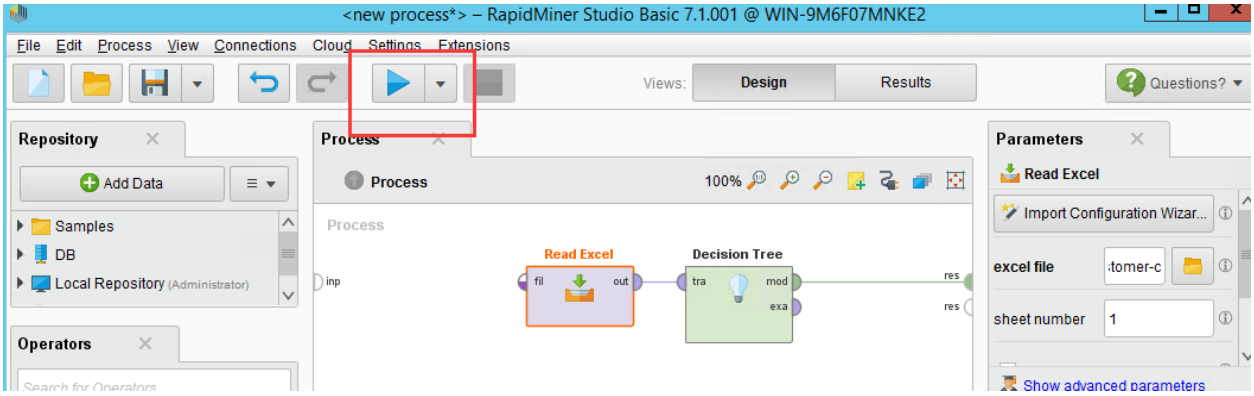
Here we need to connect “Read Excel” and “Decision Tree” on “out” and “tra” (tra = training).



Next, connect “Decision Tree” and “Process Panel” on their “mod” (mod = model) and “res” (res = results)



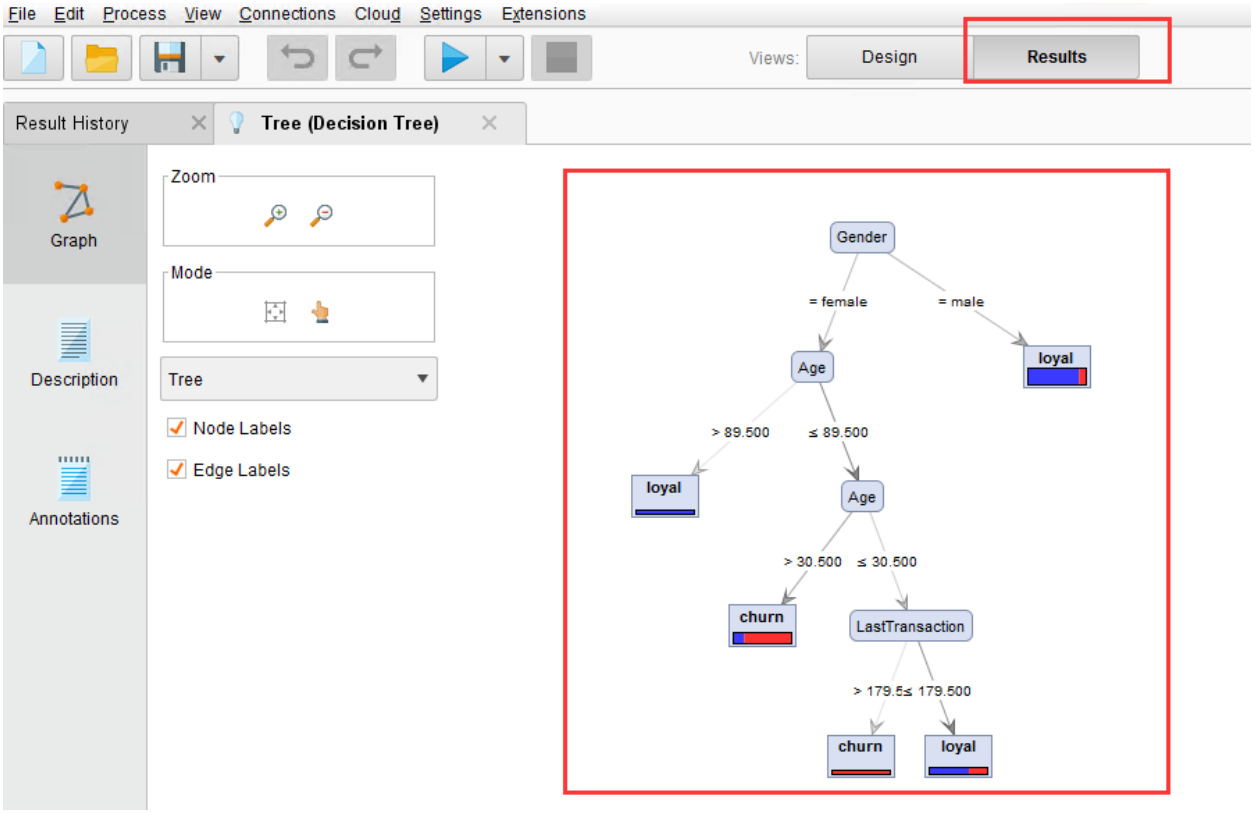
Then click the “Run” button.



You should get your results – the decision tree.

We can find that men are more loyal customers than women for this company. Women whose ages are greater than 89.5 are loyal customers.

If you are the manager of this company, what would you do in the future?



Deliverables

Take a screen shot of your decision tree and paste it on a word file **yourNetID_L5.docx**, submit in TurnItIn.

*** Copyright: Originally created by Wenli Zhang for MIS 304.*