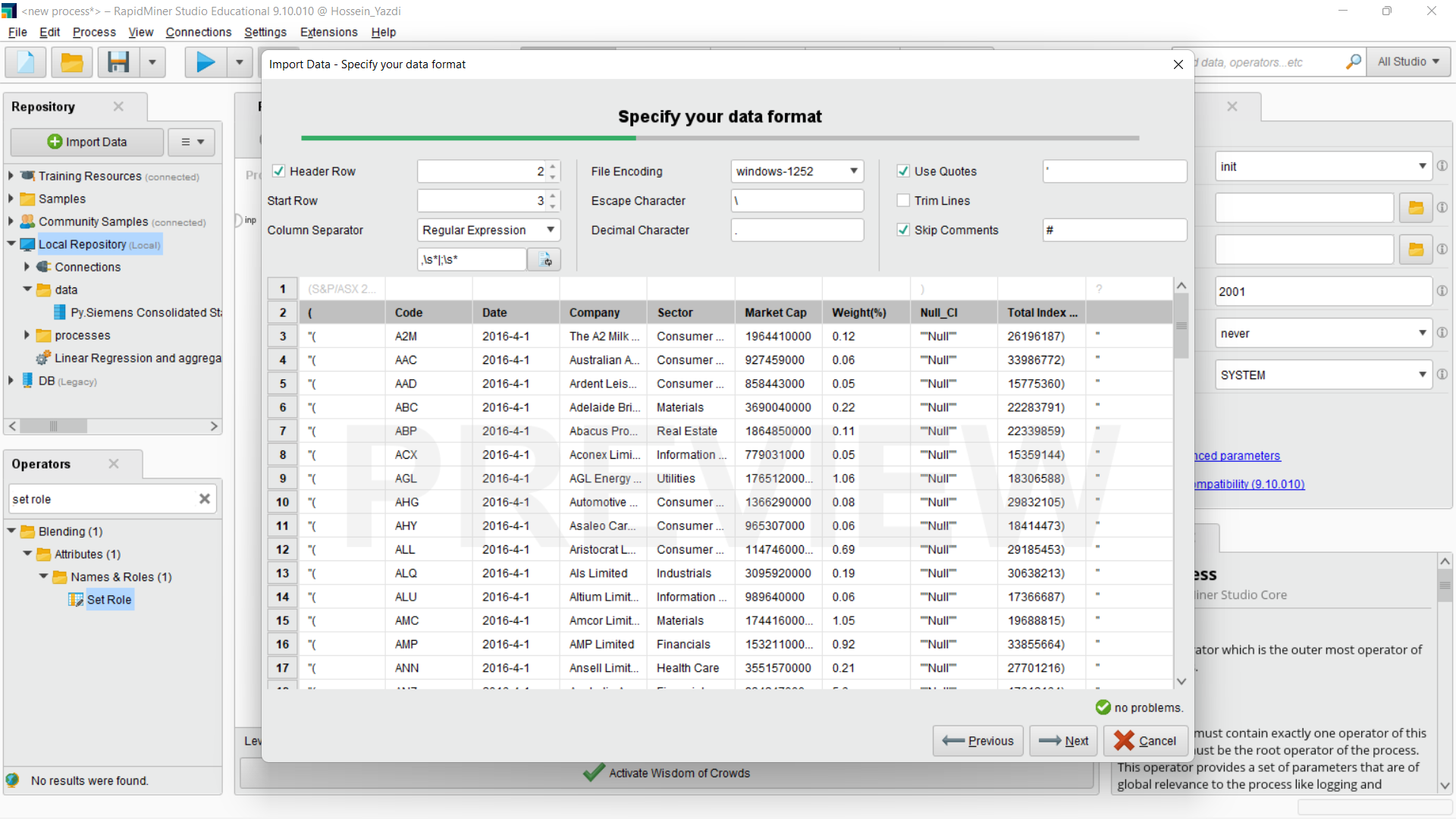
Importing the file “20170401-asx200”.

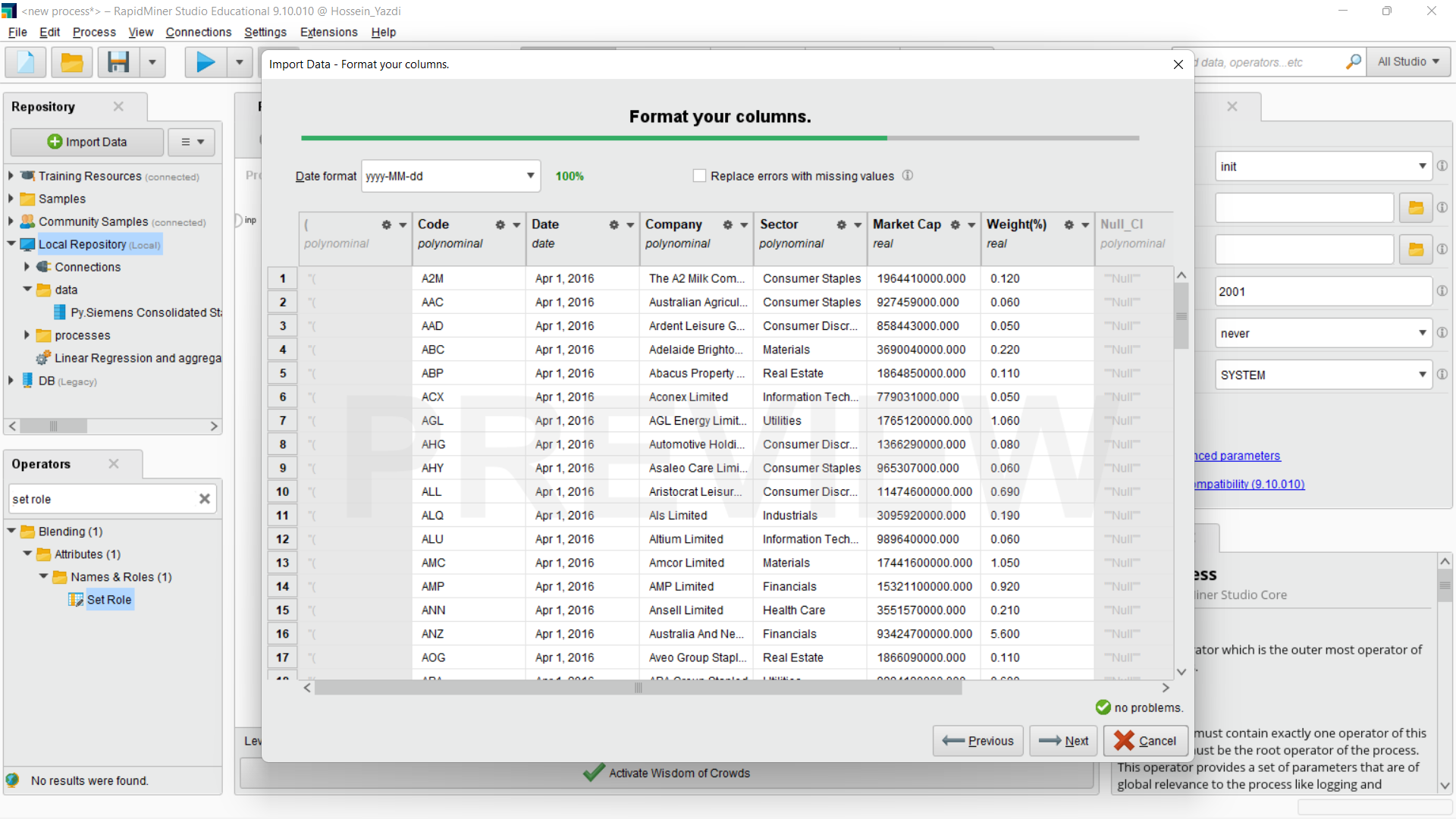
The header row must be 2, as the first row is some unuseful explanation.

The start row must be 3 because the first data begins there.

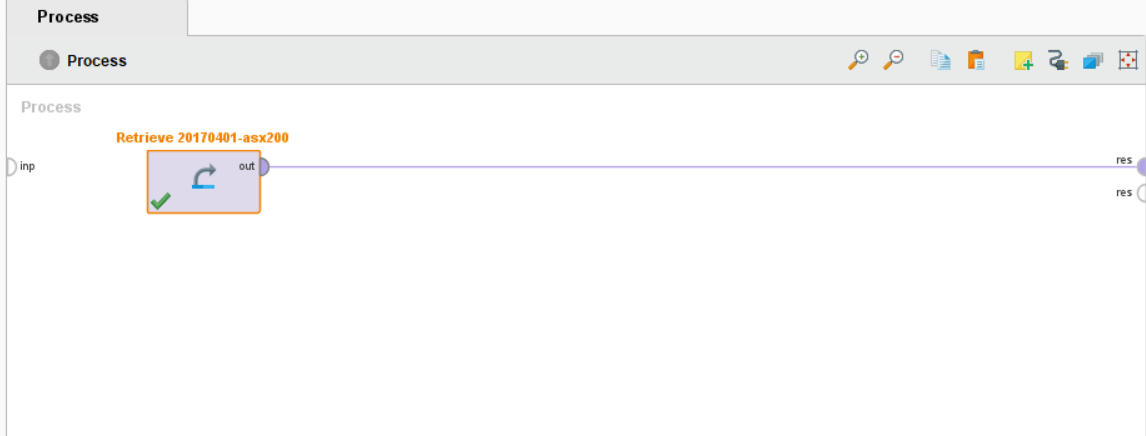
The column separator can not be a comma, because every row is surrounded by parentheses, so the “regular expression” is employed.



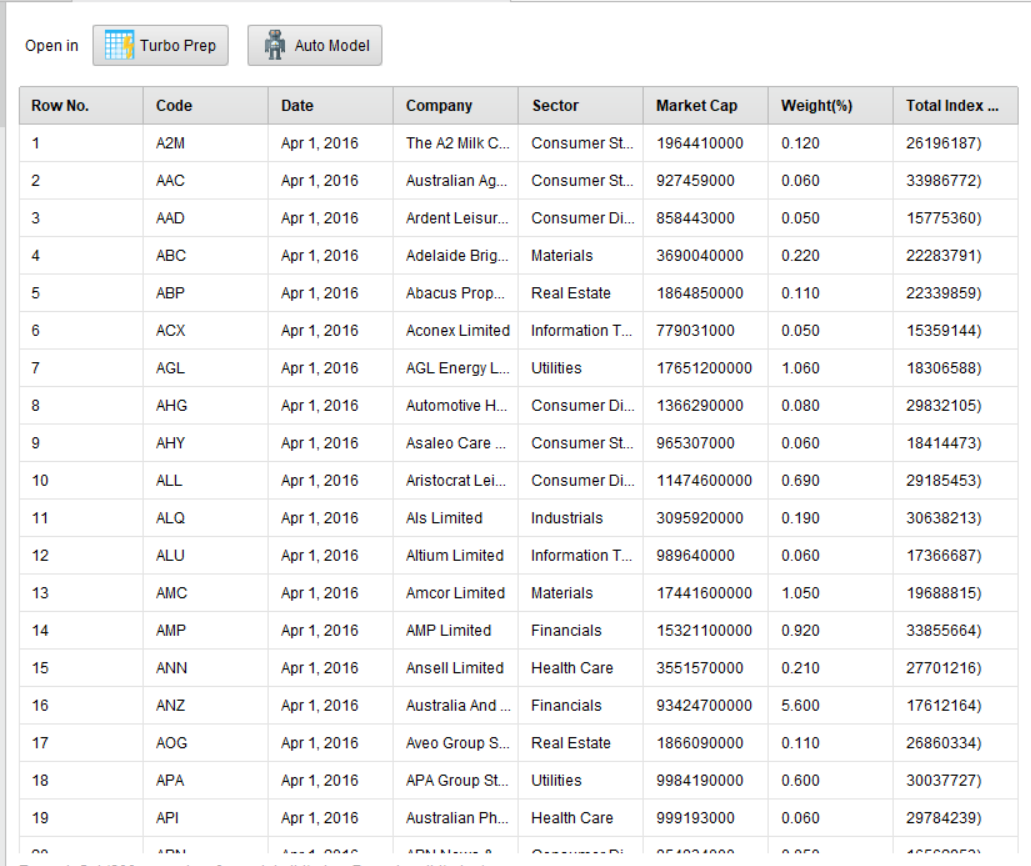
The unnecessary columns are excluded from the table.



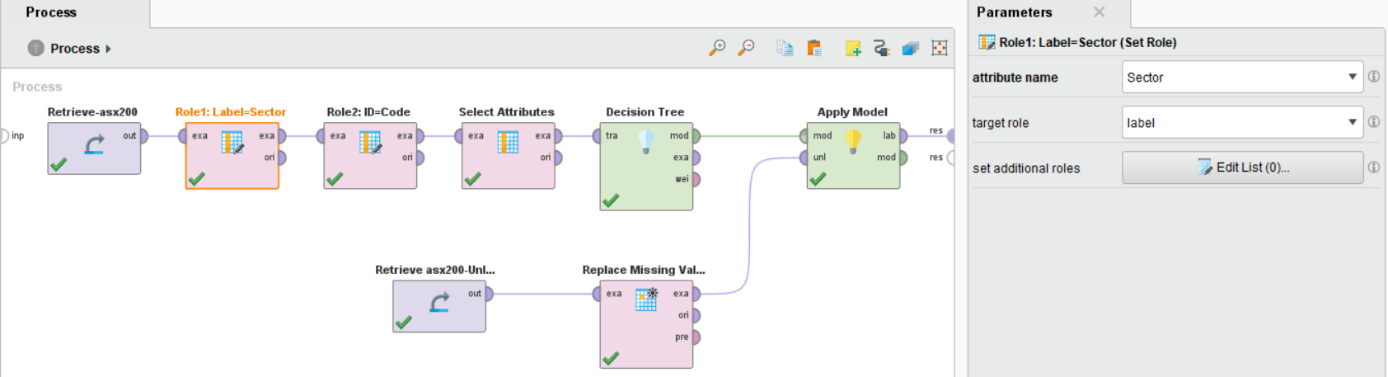
The data which is saved is retrieved.



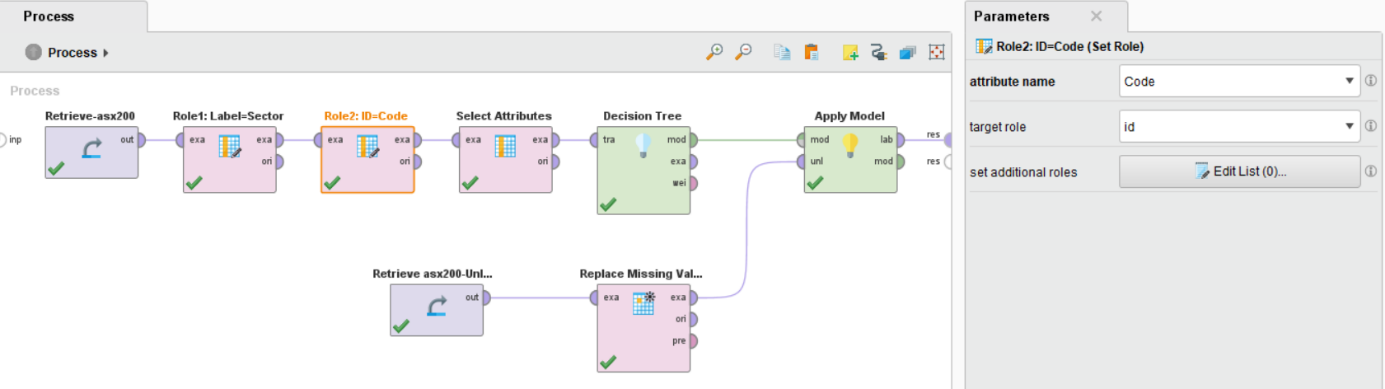
The result is as follows.



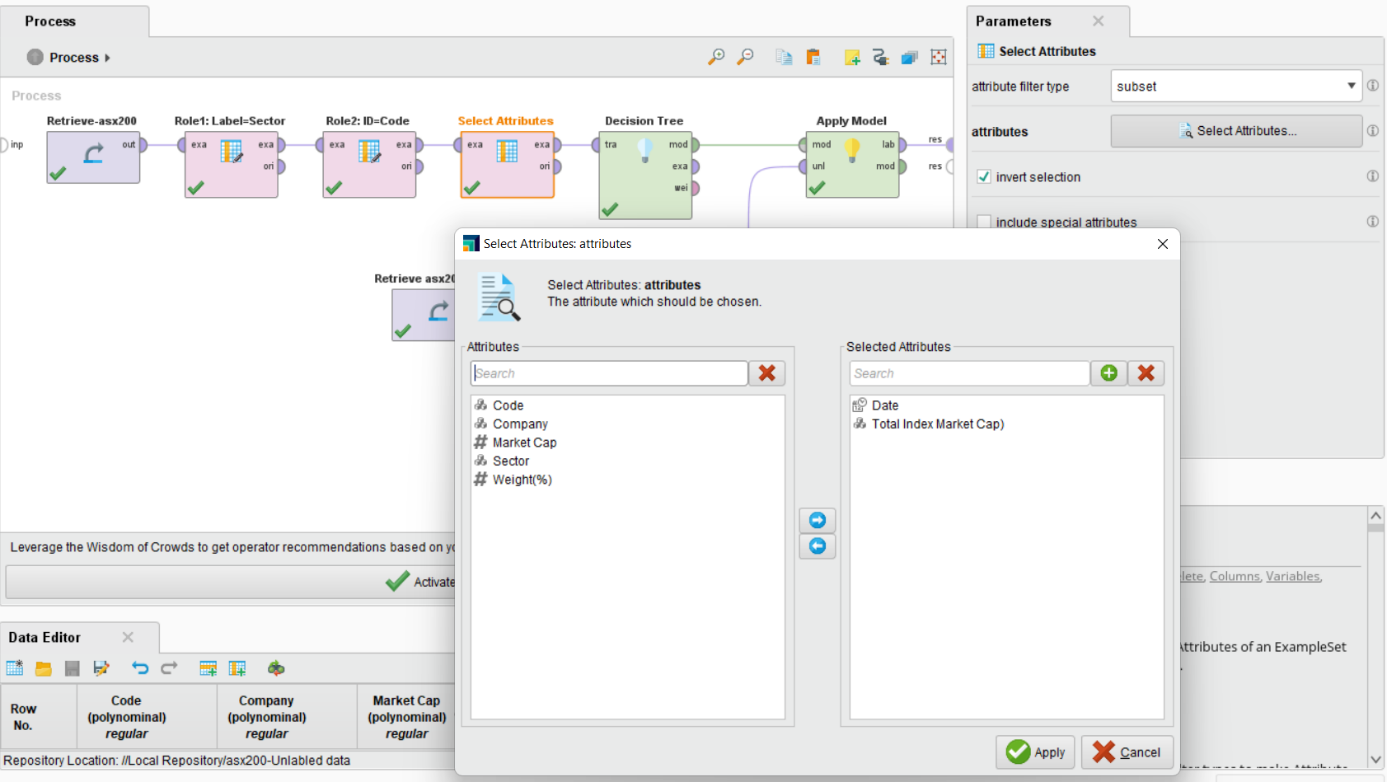
Two setting roles, one for the label which is the sector



, and the other for the id which is code.

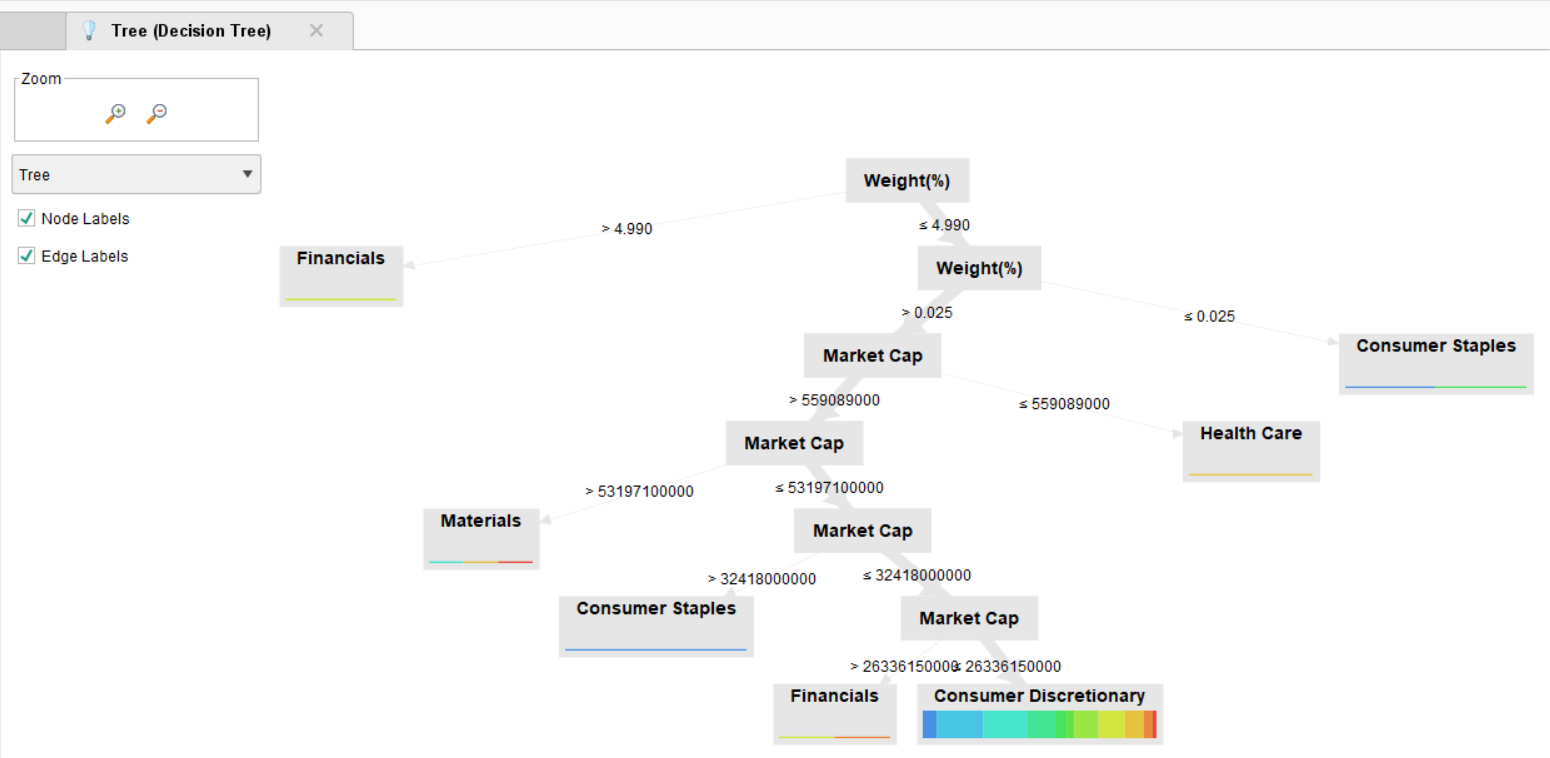


Adding the attribute to eliminate the wrong and unnecessary last column, total index market cab, using the invert selection.

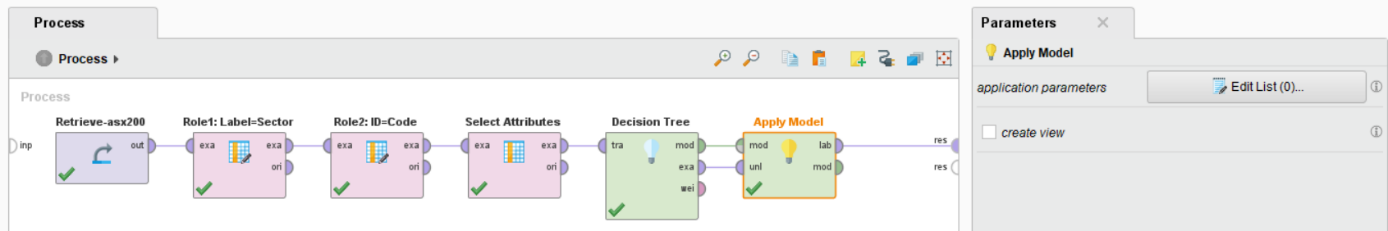


Adding “Decision Tree” to analyze the items of “Sector”:

As shown below, the “Financials” is the most important sector in terms of its weight in the way to be by far in the highest level compared to the others. Two sectors of “Consumer Staples” and “Health Care” are in the middle of weights but at the top level of “Market Cap”. The other sectors are located in the lower levels, both related to the “Weights” and “Market Cap”.



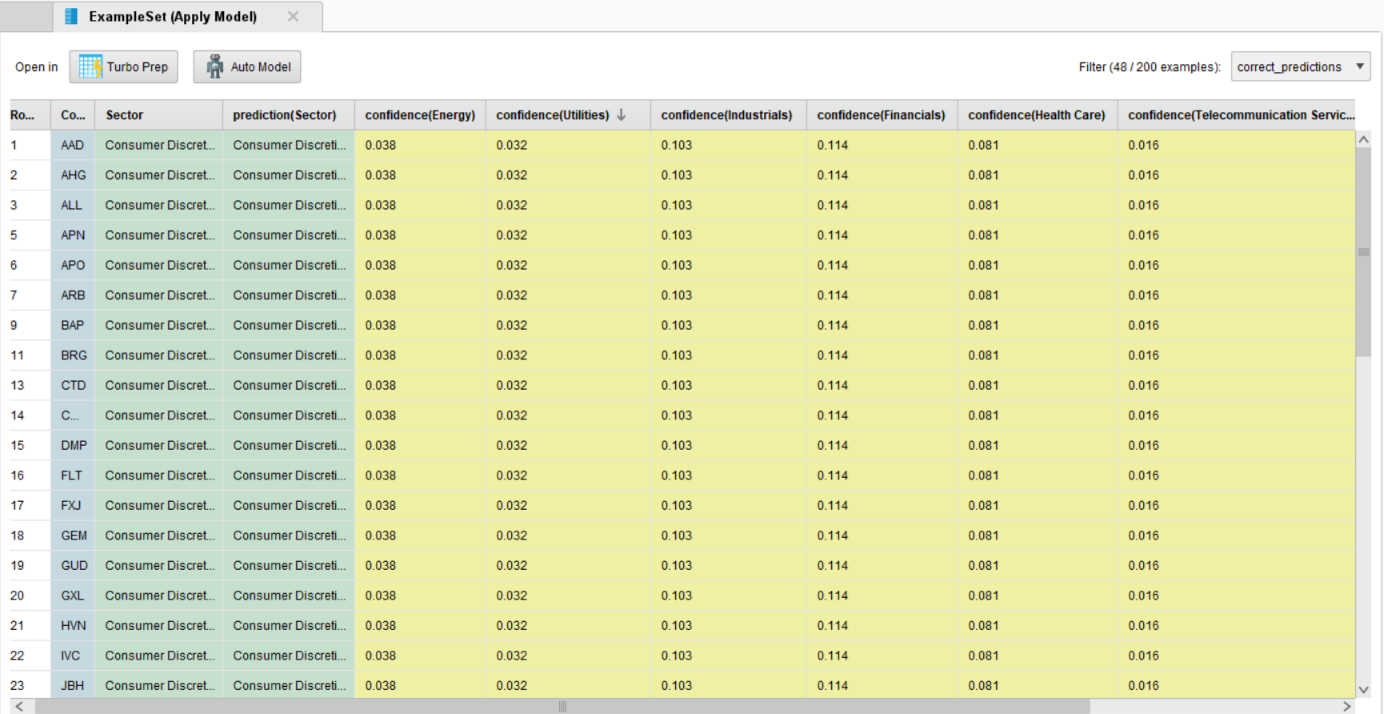
At this step, the “Apply Model” is applied to predict the different sectors.



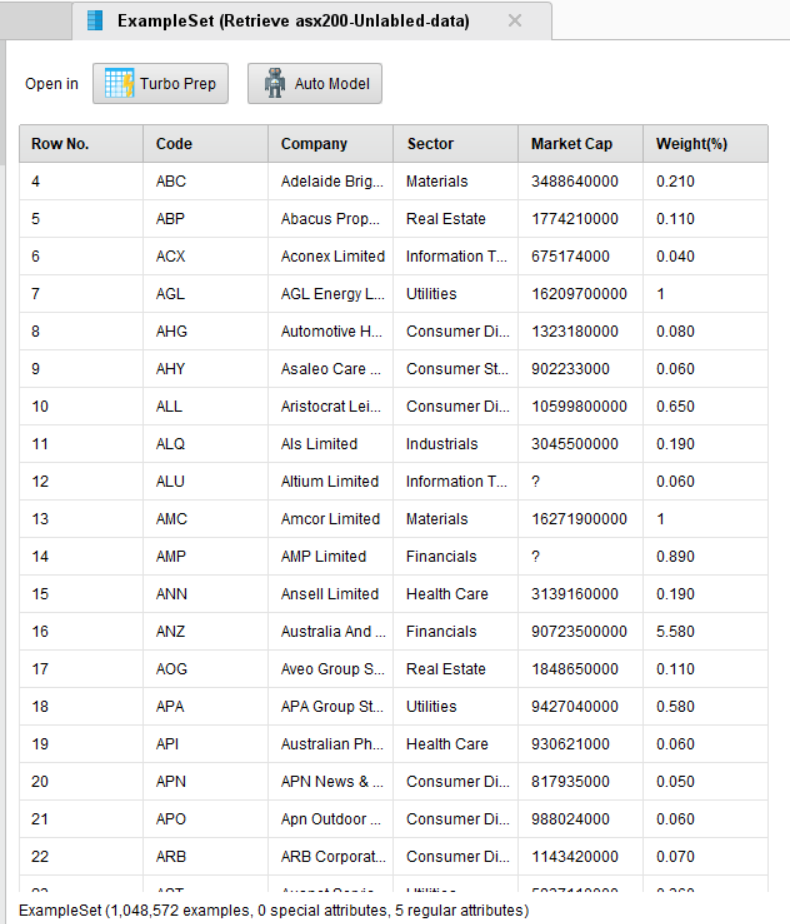
For example, “Confidence(Consumer Staples)” is about 6 percent, showing a lower probability in comparison with others.



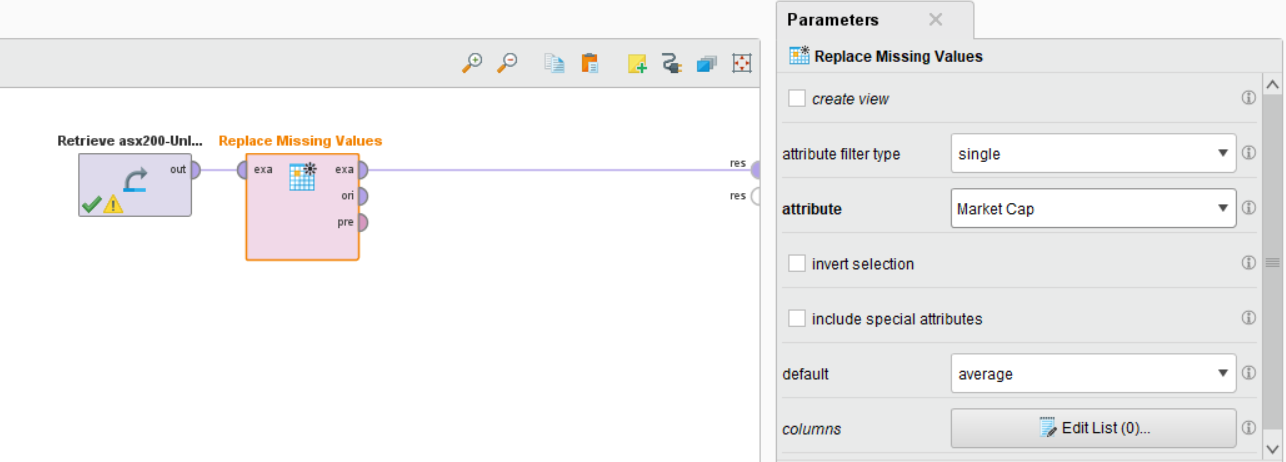
As another example, “Confidence(Financials)” is about 12 percent, showing a double probability in comparison with latter one.



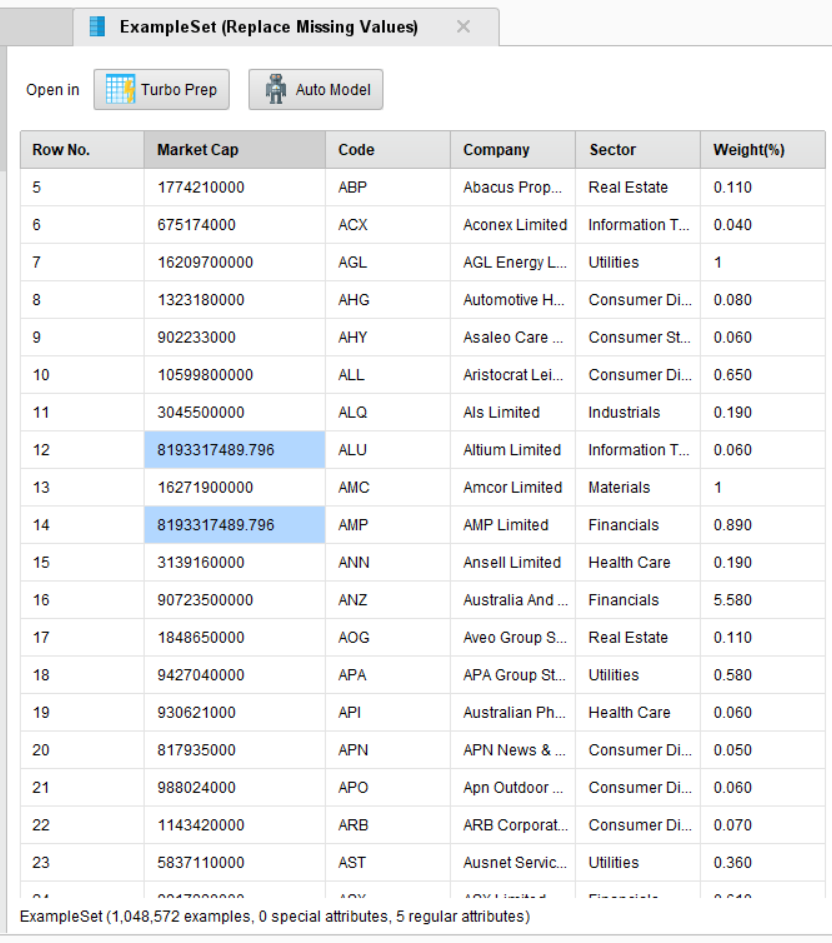
To predict the sectors for a dataset without this item, the “Unlabeled dataset” is imported. But this dataset has some missing data in the “Market Cap” in rows number 12 and 14.



To handle this problem, we add “Replace Missing Values” with “Attribute filter type” as single, the attribute as “Market Cap”, and default as “average”.



The result is about 8.2 billion Dollar each.



Now by adding these last two steps to the first dataset analysis, the confidences for prediction are available. The greatest and least predictions are “Consumer Discretionary” with 20 and “Telecommunication Services” with approximately 2 percent respectively.

