

# The Full Report for RapidMiner with Screenshots and Description

The project on RapidMiner is based on the data from the Australian Market, ASX200 in five chapters.



## Contents

|  |    |
|--|----|
| Chapter 1 .....  | 4  |
| Section 1: Importing the file “20170401-asx200” .....      | 4  |
| Section 2: Setting Roles .....                             | 7  |
| Section 3: Removing with Attribute .....                   | 8  |
| Section 4: Decision Tree .....                             | 9  |
| Section 5: Apply Model .....                               | 10 |
| Section 6: Predicting in Unlabeled Dataset .....           | 11 |
| Chapter 2 .....  | 14 |
| Section 1: Data Loading .....                              | 14 |
| Section 2: Split Data .....                                | 15 |
| Section 3: Setting Role .....                              | 17 |
| Section 4: Other Operators .....                           | 17 |
| Section 5: Running the Model .....                         | 18 |
| Chapter 3 .....  | 19 |
| Section 1: Cross Validation .....                          | 19 |
| Section 2: Cross Validations’ Operators .....              | 19 |
| Section 3: Setting Role .....                              | 19 |
| Section 4: Results .....                                   | 20 |
| Chapter 4 .....  | 21 |
| Section 1: Dataset, Setting Role, and Split its Data ..... | 21 |
| Section 2: Adding “Apply Model” and “Naive Bayes” .....    | 22 |
| Section 3: “Performance” .....                             | 23 |
| Chapter 5 .....  | 24 |
| Section 1: Using the “Multiply” operator .....             | 24 |
| Section 2: Adding two “Cross Validation” .....             | 25 |
| Section 3: Providing the Cross Validations’ Inside .....   | 25 |
| Section 4: The Result .....                                | 26 |

## Chapter 1

In this chapter, two files from ASX200 from two different periods are going to be loaded to RapidMiner, one as the major file from sections 1 to 5 and the second in section 6. While the second one is called an unlabeled file, the prediction will run on that based on the first major dataset.

### Section 1: Importing the file “20170401-asx200”.

The header row must be two as the first row is some useless explanation.

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

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

Import Data - Specify your data format

Specify your data format

| 1  | (S&P/ASX 2... |      |          |                 |                 |              | )         |         | ?               |
|----|---------------|------|----------|-----------------|-----------------|--------------|-----------|---------|-----------------|
| 2  | (             | Code | Date     | Company         | Sector          | Market Cap   | Weight(%) | Null_CI | Total Index ... |
| 3  | "(            | A2M  | 2016-4-1 | The A2 Milk ... | Consumer ...    | 1964410000   | 0.12      | "Null"  | 26196187)       |
| 4  | "(            | AAC  | 2016-4-1 | Australian A... | Consumer ...    | 927459000    | 0.06      | "Null"  | 33986772)       |
| 5  | "(            | AAD  | 2016-4-1 | Ardent Leis...  | Consumer ...    | 858443000    | 0.05      | "Null"  | 15775360)       |
| 6  | "(            | ABC  | 2016-4-1 | Adelaide Bri... | Materials       | 3690040000   | 0.22      | "Null"  | 22283791)       |
| 7  | "(            | ABP  | 2016-4-1 | Abacus Pro...   | Real Estate     | 1864850000   | 0.11      | "Null"  | 22339859)       |
| 8  | "(            | ACX  | 2016-4-1 | Aconex Limi...  | Information ... | 779031000    | 0.05      | "Null"  | 15359144)       |
| 9  | "(            | AGL  | 2016-4-1 | AGL Energy ...  | Utilities       | 176512000... | 1.06      | "Null"  | 18306588)       |
| 10 | "(            | AHG  | 2016-4-1 | Automotive ...  | Consumer ...    | 1366290000   | 0.08      | "Null"  | 29832105)       |
| 11 | "(            | AHY  | 2016-4-1 | Asaleo Car...   | Consumer ...    | 965307000    | 0.06      | "Null"  | 18414473)       |
| 12 | "(            | ALL  | 2016-4-1 | Aristocrat L... | Consumer ...    | 114746000... | 0.69      | "Null"  | 29185453)       |
| 13 | "(            | ALQ  | 2016-4-1 | Als Limited     | Industrials     | 3095920000   | 0.19      | "Null"  | 30638213)       |
| 14 | "(            | ALU  | 2016-4-1 | Altium Limit... | Information ... | 989640000    | 0.06      | "Null"  | 17366687)       |
| 15 | "(            | AMC  | 2016-4-1 | Amcor Limit...  | Materials       | 174416000... | 1.05      | "Null"  | 19688815)       |
| 16 | "(            | AMP  | 2016-4-1 | AMP Limited     | Financials      | 153211000... | 0.92      | "Null"  | 33855664)       |
| 17 | "(            | ANN  | 2016-4-1 | Ansell Limit... | Health Care     | 3551570000   | 0.21      | "Null"  | 27701216)       |

no problems.

← Previous    → Next    X Cancel

The unnecessary columns are excluded from the table.

Import Data - Format your columns.

Format your columns.

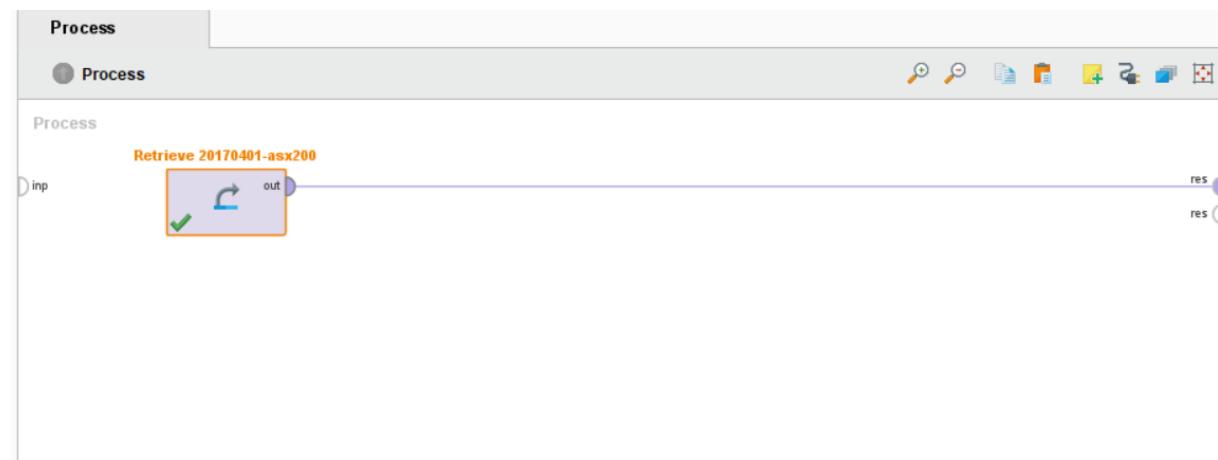
Date format yyyy-MM-dd 100% Replace errors with missing values

|    | (<br>polynomial) | Code<br>polynomial) | Date<br>date | Company<br>polynomial) | Sector<br>polynomial) | Market Cap<br>real | Weight(%)<br>real | Null_CI<br>polynomial) |
|----|------------------|---------------------|--------------|------------------------|-----------------------|--------------------|-------------------|------------------------|
| 1  | "(               | A2M                 | Apr 1, 2016  | The A2 Milk Com...     | Consumer Staples      | 1964410000.000     | 0.120             | "Null"                 |
| 2  | "(               | AAC                 | Apr 1, 2016  | Australian Agricul...  | Consumer Staples      | 927459000.000      | 0.060             | "Null"                 |
| 3  | "(               | AAD                 | Apr 1, 2016  | Ardent Leisure G...    | Consumer Discr...     | 858443000.000      | 0.050             | "Null"                 |
| 4  | "(               | ABC                 | Apr 1, 2016  | Adelaide Brighto...    | Materials             | 3690040000.000     | 0.220             | "Null"                 |
| 5  | "(               | ABP                 | Apr 1, 2016  | Abacus Property ...    | Real Estate           | 1864850000.000     | 0.110             | "Null"                 |
| 6  | "(               | ACX                 | Apr 1, 2016  | Aconex Limited         | Information Tech...   | 779031000.000      | 0.050             | "Null"                 |
| 7  | "(               | AGL                 | Apr 1, 2016  | AGL Energy Limit...    | Utilities             | 17651200000.000    | 1.060             | "Null"                 |
| 8  | "(               | AHG                 | Apr 1, 2016  | Automotive Holdi...    | Consumer Discr...     | 13662900000.000    | 0.080             | "Null"                 |
| 9  | "(               | AHY                 | Apr 1, 2016  | Asaleo Care Limi...    | Consumer Staples      | 965307000.000      | 0.060             | "Null"                 |
| 10 | "(               | ALL                 | Apr 1, 2016  | Aristocrat Leisur...   | Consumer Discr...     | 11474600000.000    | 0.690             | "Null"                 |
| 11 | "(               | ALQ                 | Apr 1, 2016  | Als Limited            | Industrials           | 30959200000.000    | 0.190             | "Null"                 |
| 12 | "(               | ALU                 | Apr 1, 2016  | Altium Limited         | Information Tech...   | 9896400000.000     | 0.060             | "Null"                 |
| 13 | "(               | AMC                 | Apr 1, 2016  | Amcor Limited          | Materials             | 17441600000.000    | 1.050             | "Null"                 |
| 14 | "(               | AMP                 | Apr 1, 2016  | AMP Limited            | Financials            | 15321100000.000    | 0.920             | "Null"                 |
| 15 | "(               | ANN                 | Apr 1, 2016  | Ansell Limited         | Health Care           | 3551570000.000     | 0.210             | "Null"                 |
| 16 | "(               | ANZ                 | Apr 1, 2016  | Australia And Ne...    | Financials            | 93424700000.000    | 5.600             | "Null"                 |
| 17 | "(               | AOG                 | Apr 1, 2016  | Aveo Group Stapl...    | Real Estate           | 18660900000.000    | 0.110             | "Null"                 |
|    | "(               | ADA                 | Apr 1, 2016  | ADA Group Shared       | Utilities             | 20044000000.000    | 0.000             | "Null"                 |

no problems.

Previous Next Cancel

The data which is saved is retrieved.

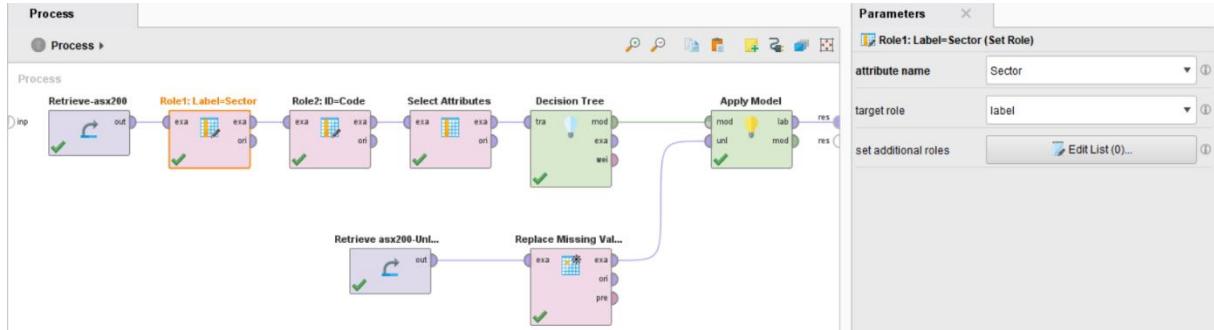


The result is as follows.

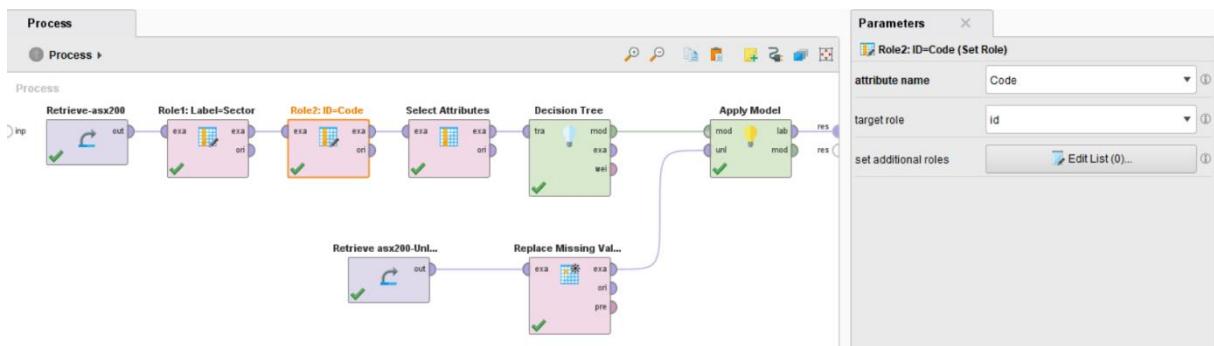
| Row No. | Code | Date        | Company           | Sector           | Market Cap  | Weight(%) | Total Index ... |
|---------|------|-------------|-------------------|------------------|-------------|-----------|-----------------|
| 1       | A2M  | Apr 1, 2016 | The A2 Milk C...  | Consumer St...   | 1964410000  | 0.120     | 26196187)       |
| 2       | AAC  | Apr 1, 2016 | Australian Ag...  | Consumer St...   | 927459000   | 0.060     | 33986772)       |
| 3       | AAD  | Apr 1, 2016 | Ardent Leisur...  | Consumer Di...   | 858443000   | 0.050     | 15775360)       |
| 4       | ABC  | Apr 1, 2016 | Adelaide Brig...  | Materials        | 3690040000  | 0.220     | 22283791)       |
| 5       | ABP  | Apr 1, 2016 | Abacus Prop...    | Real Estate      | 1864850000  | 0.110     | 22339859)       |
| 6       | ACX  | Apr 1, 2016 | Aconex Limited    | Information T... | 779031000   | 0.050     | 15359144)       |
| 7       | AGL  | Apr 1, 2016 | AGL Energy L...   | Utilities        | 17651200000 | 1.060     | 18306588)       |
| 8       | AHG  | Apr 1, 2016 | Automotive H...   | Consumer Di...   | 1366290000  | 0.080     | 29832105)       |
| 9       | AHY  | Apr 1, 2016 | Asaleo Care ...   | Consumer St...   | 965307000   | 0.060     | 18414473)       |
| 10      | ALL  | Apr 1, 2016 | Aristocrat Lei... | Consumer Di...   | 11474600000 | 0.690     | 29185453)       |
| 11      | ALQ  | Apr 1, 2016 | Als Limited       | Industrials      | 3095920000  | 0.190     | 30638213)       |
| 12      | ALU  | Apr 1, 2016 | Altium Limited    | Information T... | 989640000   | 0.060     | 17366687)       |
| 13      | AMC  | Apr 1, 2016 | Amcor Limited     | Materials        | 17441600000 | 1.050     | 19688815)       |
| 14      | AMP  | Apr 1, 2016 | AMP Limited       | Financials       | 15321100000 | 0.920     | 33855664)       |
| 15      | ANN  | Apr 1, 2016 | Ansell Limited    | Health Care      | 3551570000  | 0.210     | 27701216)       |
| 16      | ANZ  | Apr 1, 2016 | Australia And ... | Financials       | 93424700000 | 5.600     | 17612164)       |
| 17      | AOG  | Apr 1, 2016 | Aveo Group S...   | Real Estate      | 1866090000  | 0.110     | 26860334)       |
| 18      | APA  | Apr 1, 2016 | APA Group St...   | Utilities        | 9984190000  | 0.600     | 30037727)       |
| 19      | API  | Apr 1, 2016 | Australian Ph...  | Health Care      | 999193000   | 0.060     | 29784239)       |
| 20      | ADM  | Apr 1, 2016 | ADM Limited       | Consumer St...   | 854024000   | 0.050     | 46660060)       |

## Section 2: Setting Roles

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

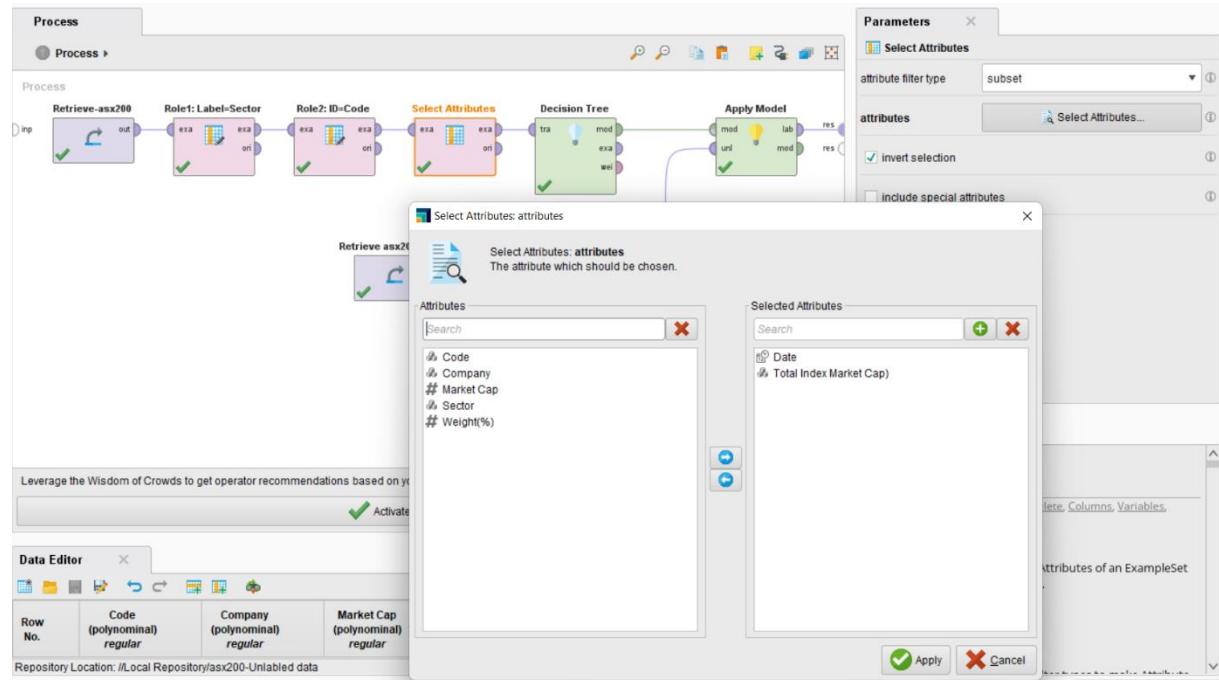


And the other is for the id which is code.



### Section 3: Removing with Attribute

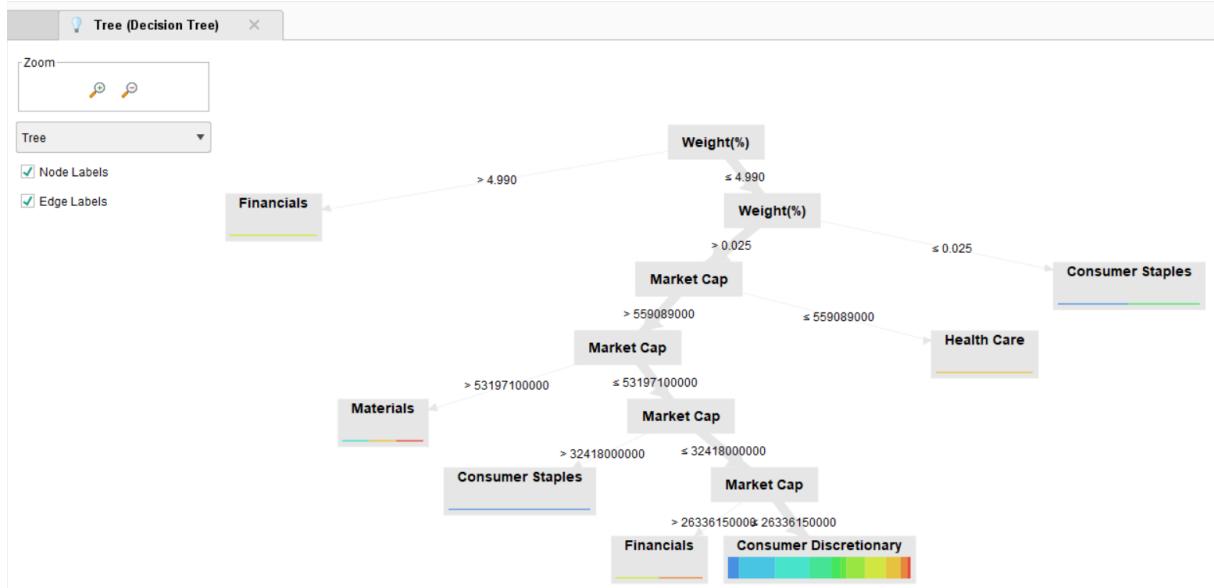
Adding the attribute to eliminate the wrong and unnecessary last column, total index market cap, and date, using the invert selection.



## Section 4: Decision Tree

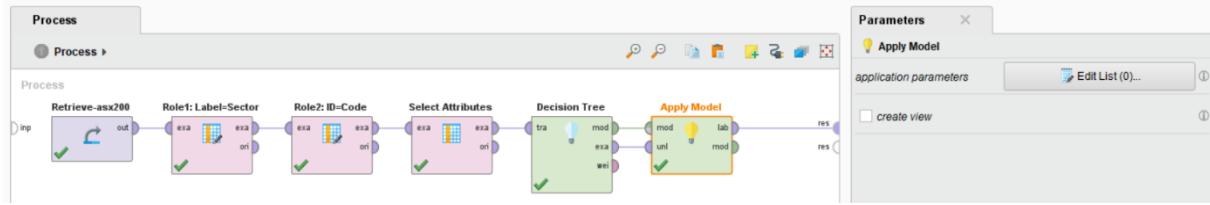
Adding “Decision Tree” to analyze the items of “Sector”:

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”.



## Section 5: Apply Model

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



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

| Row No. ↑ | Code | Sector                 | prediction(Sector)     | confidence(Consumer Staples) | confidence(Consumer Discretionary) | confidence(Materials) | confidence(Real Estate) | confidence(Information) |
|-----------|------|------------------------|------------------------|------------------------------|------------------------------------|-----------------------|-------------------------|-------------------------|
| 1         | AAD  | Consumer Discretionary | Consumer Discretionary | 0.059                        | 0.200                              | 0.189                 | 0.119                   | 0.049                   |
| 2         | AHG  | Consumer Discretionary | Consumer Discretionary | 0.059                        | 0.200                              | 0.189                 | 0.119                   | 0.049                   |
| 3         | ALL  | Consumer Discretionary | Consumer Discretionary | 0.059                        | 0.200                              | 0.189                 | 0.119                   | 0.049                   |
| 4         | ANZ  | Financials             | Financials             | 0                            | 0                                  | 0                     | 0                       | 0                       |
| 5         | APN  | Consumer Discretionary | Consumer Discretionary | 0.059                        | 0.200                              | 0.189                 | 0.119                   | 0.049                   |
| 6         | APO  | Consumer Discretionary | Consumer Discretionary | 0.059                        | 0.200                              | 0.189                 | 0.119                   | 0.049                   |
| 7         | ARB  | Consumer Discretionary | Consumer Discretionary | 0.059                        | 0.200                              | 0.189                 | 0.119                   | 0.049                   |
| 8         | BAL  | Consumer Staples       | Consumer Staples       | 0.500                        | 0                                  | 0                     | 0                       | 0.500                   |
| 9         | BAP  | Consumer Discretionary | Consumer Discretionary | 0.059                        | 0.200                              | 0.189                 | 0.119                   | 0.049                   |
| 10        | BHP  | Materials              | Materials              | 0                            | 0                                  | 0.333                 | 0                       | 0                       |
| 11        | BRG  | Consumer Discretionary | Consumer Discretionary | 0.059                        | 0.200                              | 0.189                 | 0.119                   | 0.049                   |
| 12        | CBA  | Financials             | Financials             | 0                            | 0                                  | 0                     | 0                       | 0                       |
| 13        | CTD  | Consumer Discretionary | Consumer Discretionary | 0.059                        | 0.200                              | 0.189                 | 0.119                   | 0.049                   |
| 14        | CWN  | Consumer Discretionary | Consumer Discretionary | 0.059                        | 0.200                              | 0.189                 | 0.119                   | 0.049                   |
| 15        | DMP  | Consumer Discretionary | Consumer Discretionary | 0.059                        | 0.200                              | 0.189                 | 0.119                   | 0.049                   |
| 16        | FLT  | Consumer Discretionary | Consumer Discretionary | 0.059                        | 0.200                              | 0.189                 | 0.119                   | 0.049                   |
| 17        | FXJ  | Consumer Discretionary | Consumer Discretionary | 0.059                        | 0.200                              | 0.189                 | 0.119                   | 0.049                   |
| 18        | GEM  | Consumer Discretionary | Consumer Discretionary | 0.059                        | 0.200                              | 0.189                 | 0.119                   | 0.049                   |
| 19        | GUD  | Consumer Discretionary | Consumer Discretionary | 0.059                        | 0.200                              | 0.189                 | 0.119                   | 0.049                   |

As another example, “Confidence(Financials)” is about 12 percent, showing a double probability compared to the latter.

| Row... | Co... | Sector              | prediction(Sector)  | confidence(Energy) | confidence(Utilities) ↓ | confidence(Industrials) | confidence(Financials) | confidence(Health Care) | confidence(Telecommunication Servic... |
|--------|-------|---------------------|---------------------|--------------------|-------------------------|-------------------------|------------------------|-------------------------|--|
| 1      | AAD   | Consumer Discret... | Consumer Discret... | 0.038              | 0.032                   | 0.103                   | 0.114                  | 0.081                   | 0.016                                  |
| 2      | AHG   | Consumer Discret... | Consumer Discret... | 0.038              | 0.032                   | 0.103                   | 0.114                  | 0.081                   | 0.016                                  |
| 3      | ALL   | Consumer Discret... | Consumer Discret... | 0.038              | 0.032                   | 0.103                   | 0.114                  | 0.081                   | 0.016                                  |
| 5      | APN   | Consumer Discret... | Consumer Discret... | 0.038              | 0.032                   | 0.103                   | 0.114                  | 0.081                   | 0.016                                  |
| 6      | APO   | Consumer Discret... | Consumer Discret... | 0.038              | 0.032                   | 0.103                   | 0.114                  | 0.081                   | 0.016                                  |
| 7      | ARB   | Consumer Discret... | Consumer Discret... | 0.038              | 0.032                   | 0.103                   | 0.114                  | 0.081                   | 0.016                                  |
| 9      | BAP   | Consumer Discret... | Consumer Discret... | 0.038              | 0.032                   | 0.103                   | 0.114                  | 0.081                   | 0.016                                  |
| 11     | BRG   | Consumer Discret... | Consumer Discret... | 0.038              | 0.032                   | 0.103                   | 0.114                  | 0.081                   | 0.016                                  |
| 13     | CTD   | Consumer Discret... | Consumer Discret... | 0.038              | 0.032                   | 0.103                   | 0.114                  | 0.081                   | 0.016                                  |
| 14     | C...  | Consumer Discret... | Consumer Discret... | 0.038              | 0.032                   | 0.103                   | 0.114                  | 0.081                   | 0.016                                  |
| 15     | DMP   | Consumer Discret... | Consumer Discret... | 0.038              | 0.032                   | 0.103                   | 0.114                  | 0.081                   | 0.016                                  |
| 16     | FLT   | Consumer Discret... | Consumer Discret... | 0.038              | 0.032                   | 0.103                   | 0.114                  | 0.081                   | 0.016                                  |
| 17     | FXJ   | Consumer Discret... | Consumer Discret... | 0.038              | 0.032                   | 0.103                   | 0.114                  | 0.081                   | 0.016                                  |
| 18     | GEM   | Consumer Discret... | Consumer Discret... | 0.038              | 0.032                   | 0.103                   | 0.114                  | 0.081                   | 0.016                                  |
| 19     | GUD   | Consumer Discret... | Consumer Discret... | 0.038              | 0.032                   | 0.103                   | 0.114                  | 0.081                   | 0.016                                  |
| 20     | GXL   | Consumer Discret... | Consumer Discret... | 0.038              | 0.032                   | 0.103                   | 0.114                  | 0.081                   | 0.016                                  |
| 21     | HVN   | Consumer Discret... | Consumer Discret... | 0.038              | 0.032                   | 0.103                   | 0.114                  | 0.081                   | 0.016                                  |
| 22     | IVC   | Consumer Discret... | Consumer Discret... | 0.038              | 0.032                   | 0.103                   | 0.114                  | 0.081                   | 0.016                                  |
| 23     | JBH   | Consumer Discret... | Consumer Discret... | 0.038              | 0.032                   | 0.103                   | 0.114                  | 0.081                   | 0.016                                  |

## Section 6: Predicting in Unlabeled Dataset

To predict the sectors for another dataset, the “Unlabeled dataset”, is imported. Even though it contains the “Sector” column, we are going to make a new experience by predicting the sector and comparing the result of the prediction with the real information.

Firstly, as this dataset has some missing data in the “Market Cap” in rows number 12 and 14, they must be corrected.

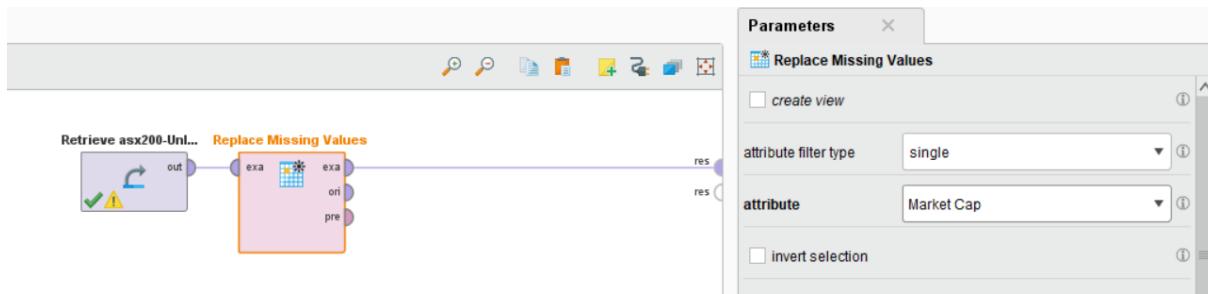
ExampleSet (Retrieve asx200-Unlabeled-data)

Open in [Turbo Prep](#) [Auto Model](#)

| Row No. | Code | Company           | Sector           | Market Cap  | Weight(%) |
|---------|------|-------------------|------------------|-------------|-----------|
| 4       | ABC  | Adelaide Brig...  | Materials        | 3488640000  | 0.210     |
| 5       | ABP  | Abacus Prop...    | Real Estate      | 1774210000  | 0.110     |
| 6       | ACX  | Aconex Limited    | Information T... | 675174000   | 0.040     |
| 7       | AGL  | AGL Energy L...   | Utilities        | 16209700000 | 1         |
| 8       | AHG  | Automotive H...   | Consumer Di...   | 1323180000  | 0.080     |
| 9       | AHY  | Asaleo Care ...   | Consumer St...   | 902233000   | 0.060     |
| 10      | ALL  | Aristocrat Lei... | Consumer Di...   | 10599800000 | 0.650     |
| 11      | ALQ  | Als Limited       | Industrials      | 3045500000  | 0.190     |
| 12      | ALU  | Altium Limited    | Information T... | ?           | 0.060     |
| 13      | AMC  | Amcor Limited     | Materials        | 16271900000 | 1         |
| 14      | AMP  | AMP Limited       | Financials       | ?           | 0.890     |
| 15      | ANN  | Ansell Limited    | Health Care      | 3139160000  | 0.190     |
| 16      | ANZ  | Australia And ... | Financials       | 90723500000 | 5.580     |
| 17      | AOG  | Aveo Group S...   | Real Estate      | 1848650000  | 0.110     |
| 18      | APA  | APA Group St...   | Utilities        | 9427040000  | 0.580     |
| 19      | API  | Australian Ph...  | Health Care      | 930621000   | 0.060     |
| 20      | APN  | APN News & ...    | Consumer Di...   | 817935000   | 0.050     |
| 21      | APO  | Apn Outdoor ...   | Consumer Di...   | 988024000   | 0.060     |
| 22      | ARB  | ARB Corporat...   | Consumer Di...   | 1143420000  | 0.070     |
| ...     | ACT  | Autoliv Corp...   | Automotive       | 5007440000  | 0.000     |

ExampleSet (1,048,572 examples, 0 special attributes, 5 regular attributes)

To manage 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 Dollars each.

**ExampleSet (Replace Missing Values)**

Open in [Turbo Prep](#) [Auto Model](#)

| Row No. | Market Cap     | Code | Company           | Sector           | Weight(%) |
|---------|----------------|------|-------------------|------------------|-----------|
| 5       | 1774210000     | ABP  | Abacus Prop...    | Real Estate      | 0.110     |
| 6       | 675174000      | ACX  | Aconex Limited    | Information T... | 0.040     |
| 7       | 16209700000    | AGL  | AGL Energy L...   | Utilities        | 1         |
| 8       | 1323180000     | AHG  | Automotive H...   | Consumer Di...   | 0.080     |
| 9       | 902233000      | AHY  | Asaleo Care ...   | Consumer St...   | 0.060     |
| 10      | 10599800000    | ALL  | Aristocrat Lei... | Consumer Di...   | 0.650     |
| 11      | 3045500000     | ALQ  | Als Limited       | Industrials      | 0.190     |
| 12      | 8193317489.796 | ALU  | Altium Limited    | Information T... | 0.060     |
| 13      | 16271900000    | AMC  | Amcor Limited     | Materials        | 1         |
| 14      | 8193317489.796 | AMP  | AMP Limited       | Financials       | 0.890     |
| 15      | 3139160000     | ANN  | Ansell Limited    | Health Care      | 0.190     |
| 16      | 90723500000    | ANZ  | Australia And ... | Financials       | 5.580     |
| 17      | 1848650000     | AOG  | Aveo Group S...   | Real Estate      | 0.110     |
| 18      | 9427040000     | APA  | APA Group St...   | Utilities        | 0.580     |
| 19      | 930621000      | API  | Australian Ph...  | Health Care      | 0.060     |
| 20      | 817935000      | APN  | APN News & ...    | Consumer Di...   | 0.050     |
| 21      | 988024000      | APO  | Apn Outdoor ...   | Consumer Di...   | 0.060     |
| 22      | 1143420000     | ARB  | ARB Corporat...   | Consumer Di...   | 0.070     |
| 23      | 5837110000     | AST  | Ausnet Servic...  | Utilities        | 0.360     |
| 24      | 6647000000     | AVM  | AVM Limited       | Financials       | 0.040     |

ExampleSet (1,048,572 examples, 0 special attributes, 5 regular attributes)

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

| ExampleSet (Apply Model) |                    |  |                                    |  |           |           |           |           |           |           |           |           |      |                   |  |
|--------------------------|--------------------|--|------------------------------------|--|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|------|-------------------|--|
| Open in                  |                    | Turbo Prep                             | Auto Model                         | Filter (1,048,572 / 1,048,572 examples): all |           |           |           |           |           |           |           |           |      |                   |  |
| Row No.                  | prediction(Sector) | confidence(Telecommunication Services) | confidence(Consumer Discretionary) | confid...                                    | confid... | confid... | confid... | confid... | confid... | confid... | confid... | confid... | Code | Company           |  |
| 1                        | Consum...          | 0.016                                  | 0.200                              | 0.189  | 0.119     | 0.049     | 0.032     | 0.103     | 0.114     | 0.081     | 0.038     | 0.059     | A2M  | The A2 Milk C...  |  |
| 2                        | Consum...          | 0.016                                  | 0.200                              | 0.189  | 0.119     | 0.049     | 0.032     | 0.103     | 0.114     | 0.081     | 0.038     | 0.059     | AAC  | Australian Ag...  |  |
| 3                        | Consum...          | 0.016                                  | 0.200                              | 0.189  | 0.119     | 0.049     | 0.032     | 0.103     | 0.114     | 0.081     | 0.038     | 0.059     | AAD  | Ardent Leisur...  |  |
| 4                        | Consum...          | 0.016                                  | 0.200                              | 0.189  | 0.119     | 0.049     | 0.032     | 0.103     | 0.114     | 0.081     | 0.038     | 0.059     | ABC  | Adelaide Brig...  |  |
| 5                        | Consum...          | 0.016                                  | 0.200                              | 0.189  | 0.119     | 0.049     | 0.032     | 0.103     | 0.114     | 0.081     | 0.038     | 0.059     | ABP  | Abacus Prop...    |  |
| 6                        | Consum...          | 0.016                                  | 0.200                              | 0.189  | 0.119     | 0.049     | 0.032     | 0.103     | 0.114     | 0.081     | 0.038     | 0.059     | ACX  | Aconex Limited    |  |
| 7                        | Consum...          | 0.016                                  | 0.200                              | 0.189  | 0.119     | 0.049     | 0.032     | 0.103     | 0.114     | 0.081     | 0.038     | 0.059     | AGL  | AGL Energy L...   |  |
| 8                        | Consum...          | 0.016                                  | 0.200                              | 0.189  | 0.119     | 0.049     | 0.032     | 0.103     | 0.114     | 0.081     | 0.038     | 0.059     | AHG  | Automotive H...   |  |
| 9                        | Consum...          | 0.016                                  | 0.200                              | 0.189  | 0.119     | 0.049     | 0.032     | 0.103     | 0.114     | 0.081     | 0.038     | 0.059     | AHY  | Asaleo Care ...   |  |
| 10                       | Consum...          | 0.016                                  | 0.200                              | 0.189  | 0.119     | 0.049     | 0.032     | 0.103     | 0.114     | 0.081     | 0.038     | 0.059     | ALL  | Aristocrat Le...  |  |
| 11                       | Consum...          | 0.016                                  | 0.200                              | 0.189  | 0.119     | 0.049     | 0.032     | 0.103     | 0.114     | 0.081     | 0.038     | 0.059     | ALQ  | Als Limited       |  |
| 12                       | Consum...          | 0.016                                  | 0.200                              | 0.189  | 0.119     | 0.049     | 0.032     | 0.103     | 0.114     | 0.081     | 0.038     | 0.059     | ALU  | Altium Limited    |  |
| 13                       | Consum...          | 0.016                                  | 0.200                              | 0.189  | 0.119     | 0.049     | 0.032     | 0.103     | 0.114     | 0.081     | 0.038     | 0.059     | AMC  | Amcor Limited     |  |
| 14                       | Consum...          | 0.016                                  | 0.200                              | 0.189  | 0.119     | 0.049     | 0.032     | 0.103     | 0.114     | 0.081     | 0.038     | 0.059     | AMP  | AMP Limited       |  |
| 15                       | Consum...          | 0.016                                  | 0.200                              | 0.189  | 0.119     | 0.049     | 0.032     | 0.103     | 0.114     | 0.081     | 0.038     | 0.059     | ANN  | Ansell Limited    |  |
| 16                       | Financials         | 0                                      | 0                                  | 0  | 0         | 0         | 0         | 1         | 0         | 0         | 0         | 0         | ANZ  | Australia And ... |  |
| 17                       | Consum...          | 0.016                                  | 0.200                              | 0.189  | 0.119     | 0.049     | 0.032     | 0.103     | 0.114     | 0.081     | 0.038     | 0.059     | AOG  | Aveo Group S...   |  |
| 18                       | Consum...          | 0.016                                  | 0.200                              | 0.189  | 0.119     | 0.049     | 0.032     | 0.103     | 0.114     | 0.081     | 0.038     | 0.059     | APA  | APA Group St...   |  |
| 19                       | Consum...          | 0.016                                  | 0.200                              | 0.189  | 0.119     | 0.049     | 0.032     | 0.103     | 0.114     | 0.081     | 0.038     | 0.059     | API  | Australian Ph...  |  |

Now if both “Prediction(Sector)” and “Sector” are compared, the result is pointing out that this prediction is not for any kind of data applicable and true. The reason for it is that there is not a good and meaningful relationship between sectors of companies and their market cap or weight. In the next screenshots, the two sectors are beside each other to be clear about this issue.

| Row No. | prediction(Sector)     | Sector                 | confidence(Consumer Staples) | confidence(Consumer Discretionary) | confidence(Materials) | confidence(...) | confidence(L...) | confidence |
|---------|------------------------|------------------------|------------------------------|------------------------------------|-----------------------|-----------------|------------------|------------|
| 1       | Consumer Discretionary | Consumer Staples       | 0.059                        | 0.200                              | 0.189                 | 0.119           | 0.049            | 0.032      |
| 2       | Consumer Discretionary | Consumer Staples       | 0.059                        | 0.200                              | 0.189                 | 0.119           | 0.049            | 0.032      |
| 3       | Consumer Discretionary | Consumer Discretionary | 0.059                        | 0.200                              | 0.189                 | 0.119           | 0.049            | 0.032      |
| 4       | Consumer Discretionary | Materials              | 0.059                        | 0.200                              | 0.189                 | 0.119           | 0.049            | 0.032      |
| 5       | Consumer Discretionary | Real Estate            | 0.059                        | 0.200                              | 0.189                 | 0.119           | 0.049            | 0.032      |
| 6       | Consumer Discretionary | Information Technology | 0.059                        | 0.200                              | 0.189                 | 0.119           | 0.049            | 0.032      |
| 7       | Consumer Discretionary | Utilities              | 0.059                        | 0.200                              | 0.189                 | 0.119           | 0.049            | 0.032      |
| 8       | Consumer Discretionary | Consumer Discretionary | 0.059                        | 0.200                              | 0.189                 | 0.119           | 0.049            | 0.032      |
| 9       | Consumer Discretionary | Consumer Staples       | 0.059                        | 0.200                              | 0.189                 | 0.119           | 0.049            | 0.032      |
| 10      | Consumer Discretionary | Consumer Discretionary | 0.059                        | 0.200                              | 0.189                 | 0.119           | 0.049            | 0.032      |
| 11      | Consumer Discretionary | Industrials            | 0.059                        | 0.200                              | 0.189                 | 0.119           | 0.049            | 0.032      |
| 12      | Consumer Discretionary | Information Technology | 0.059                        | 0.200                              | 0.189                 | 0.119           | 0.049            | 0.032      |
| 13      | Consumer Discretionary | Materials              | 0.059                        | 0.200                              | 0.189                 | 0.119           | 0.049            | 0.032      |
| 14      | Consumer Discretionary | Financials             | 0.059                        | 0.200                              | 0.189                 | 0.119           | 0.049            | 0.032      |
| 15      | Consumer Discretionary | Health Care            | 0.059                        | 0.200                              | 0.189                 | 0.119           | 0.049            | 0.032      |
| 16      | Financials             | Financials             | 0                            | 0                                  | 0                     | 0               | 0                | 0          |
| 17      | Consumer Discretionary | Real Estate            | 0.059                        | 0.200                              | 0.189                 | 0.119           | 0.049            | 0.032      |
| 18      | Consumer Discretionary | Utilities              | 0.059                        | 0.200                              | 0.189                 | 0.119           | 0.049            | 0.032      |
| 19      | Consumer Discretionary | Health Care            | 0.059                        | 0.200                              | 0.189                 | 0.119           | 0.049            | 0.032      |

| Row No. | prediction(Sector)     | Sector                 | confidence(Consumer Staples) | confidence(Consumer Discretionary) | confidence(Materials) | confidence(...) | confidence(L...) | confidence |
|---------|------------------------|------------------------|------------------------------|------------------------------------|-----------------------|-----------------|------------------|------------|
| 20      | Consumer Discretionary | Consumer Discretionary | 0.059                        | 0.200                              | 0.189                 | 0.119           | 0.049            | 0.032      |
| 21      | Consumer Discretionary | Consumer Discretionary | 0.059                        | 0.200                              | 0.189                 | 0.119           | 0.049            | 0.032      |
| 22      | Consumer Discretionary | Consumer Discretionary | 0.059                        | 0.200                              | 0.189                 | 0.119           | 0.049            | 0.032      |
| 23      | Consumer Discretionary | Utilities              | 0.059                        | 0.200                              | 0.189                 | 0.119           | 0.049            | 0.032      |
| 24      | Consumer Discretionary | Financials             | 0.059                        | 0.200                              | 0.189                 | 0.119           | 0.049            | 0.032      |
| 25      | Consumer Discretionary | Materials              | 0.059                        | 0.200                              | 0.189                 | 0.119           | 0.049            | 0.032      |
| 26      | Consumer Discretionary | Industrials            | 0.059                        | 0.200                              | 0.189                 | 0.119           | 0.049            | 0.032      |
| 27      | Consumer Discretionary | Consumer Discretionary | 0.059                        | 0.200                              | 0.189                 | 0.119           | 0.049            | 0.032      |
| 28      | Consumer Discretionary | Financials             | 0.059                        | 0.200                              | 0.189                 | 0.119           | 0.049            | 0.032      |
| 29      | Consumer Discretionary | Consumer Staples       | 0.059                        | 0.200                              | 0.189                 | 0.119           | 0.049            | 0.032      |
| 30      | Materials              | Materials              | 0                            | 0                                  | 0.333                 | 0               | 0                | 0          |
| 31      | Consumer Discretionary | Consumer Staples       | 0.059                        | 0.200                              | 0.189                 | 0.119           | 0.049            | 0.032      |
| 32      | Consumer Discretionary | Materials              | 0.059                        | 0.200                              | 0.189                 | 0.119           | 0.049            | 0.032      |
| 33      | Consumer Discretionary | Materials              | 0.059                        | 0.200                              | 0.189                 | 0.119           | 0.049            | 0.032      |
| 34      | Consumer Discretionary | Financials             | 0.059                        | 0.200                              | 0.189                 | 0.119           | 0.049            | 0.032      |
| 35      | Consumer Discretionary | Energy                 | 0.059                        | 0.200                              | 0.189                 | 0.119           | 0.049            | 0.032      |
| 36      | Consumer Discretionary | Consumer Discretionary | 0.059                        | 0.200                              | 0.189                 | 0.119           | 0.049            | 0.032      |
| 37      | Consumer Discretionary | Materials              | 0.059                        | 0.200                              | 0.189                 | 0.119           | 0.049            | 0.032      |
| 38      | Consumer Discretionary | Financials             | 0.059                        | 0.200                              | 0.189                 | 0.119           | 0.049            | 0.032      |

## Chapter 2

In this chapter, the major file from the first chapter is loaded and then will be split into two datasets.

### Section 1: Data Loading

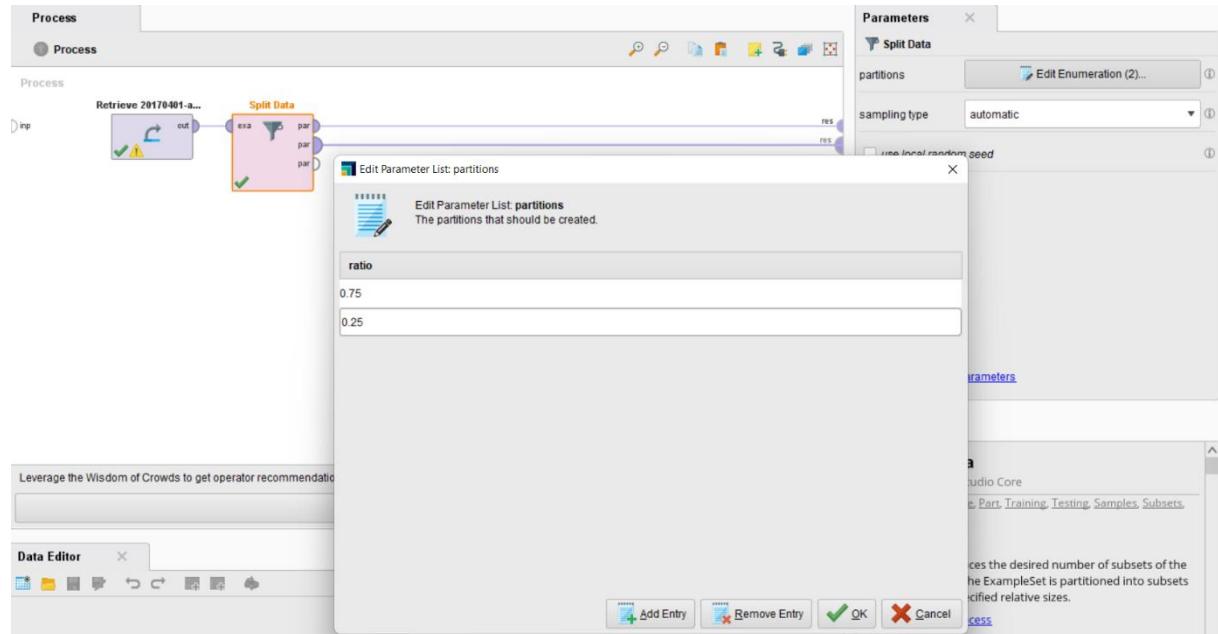
The first screen and the result after loading and running the major ASX200 file.

The screenshot shows the RapidMiner interface. At the top, there is a process titled "Process" with a "Process" icon. Below it, a "Process" window is open, showing a flow diagram with an "inp" port connected to a "Retrieve 20170401-asx200" operator, which has an "out" port connected to a "res" port. The "ExampleSet (Retrieve 20170401-asx200)" window is displayed below, containing a table of data. The table has columns: Row No., Code, Date, Company, Sector, Market Cap, Weight(%), and Total Index ... . The data consists of 20 rows of ASX200 company information.

| Row No. | Code | Date        | Company           | Sector           | Market Cap  | Weight(%) | Total Index ... |
|---------|------|-------------|-------------------|------------------|-------------|-----------|-----------------|
| 1       | A2M  | Apr 1, 2016 | The A2 Milk C...  | Consumer St...   | 1964410000  | 0.120     | 26196187)       |
| 2       | AAC  | Apr 1, 2016 | Australian Ag...  | Consumer St...   | 927459000   | 0.060     | 33986772)       |
| 3       | AAD  | Apr 1, 2016 | Ardent Leisur...  | Consumer Di...   | 858443000   | 0.050     | 15775360)       |
| 4       | ABC  | Apr 1, 2016 | Adelaide Brig...  | Materials        | 3690040000  | 0.220     | 22283791)       |
| 5       | ABP  | Apr 1, 2016 | Abacus Prop...    | Real Estate      | 1864850000  | 0.110     | 22339859)       |
| 6       | ACX  | Apr 1, 2016 | Aconex Limited    | Information T... | 779031000   | 0.050     | 15359144)       |
| 7       | AGL  | Apr 1, 2016 | AGL Energy L...   | Utilities        | 17651200000 | 1.060     | 18306588)       |
| 8       | AHG  | Apr 1, 2016 | Automotive H...   | Consumer Di...   | 1366290000  | 0.080     | 29832105)       |
| 9       | AHY  | Apr 1, 2016 | Asaleo Care ...   | Consumer St...   | 965307000   | 0.060     | 18414473)       |
| 10      | ALL  | Apr 1, 2016 | Aristocrat Lei... | Consumer Di...   | 11474600000 | 0.690     | 29185453)       |
| 11      | ALQ  | Apr 1, 2016 | Als Limited       | Industrials      | 3095920000  | 0.190     | 30638213)       |
| 12      | ALU  | Apr 1, 2016 | Altium Limited    | Information T... | 989640000   | 0.060     | 17366687)       |
| 13      | AMC  | Apr 1, 2016 | Amcor Limited     | Materials        | 17441600000 | 1.050     | 19688815)       |
| 14      | AMP  | Apr 1, 2016 | AMP Limited       | Financials       | 15321100000 | 0.920     | 33855664)       |
| 15      | ANN  | Apr 1, 2016 | Ansell Limited    | Health Care      | 3551570000  | 0.210     | 27701216)       |
| 16      | ANZ  | Apr 1, 2016 | Australia And ... | Financials       | 93424700000 | 5.600     | 17612164)       |
| 17      | AOG  | Apr 1, 2016 | Aveo Group S...   | Real Estate      | 1866090000  | 0.110     | 26860334)       |
| 18      | APA  | Apr 1, 2016 | APA Group St...   | Utilities        | 9984190000  | 0.600     | 30037727)       |
| 19      | API  | Apr 1, 2016 | Australian Ph...  | Health Care      | 999193000   | 0.060     | 29784239)       |
| 20      | APM  | Apr 1, 2016 | ADM Australia     | Consumer St...   | 954001000   | 0.050     | 40500050)       |

## Section 2: Split Data

By using the operator “Split Data”, the major data is separated into two groups of 0.75 and 0.25.



The output is in two tables, one with 50 members.

| Row No. | Code | Date        | Company           | Sector           | Market Cap  | Weight(%) | Total Index ... |
|---------|------|-------------|-------------------|------------------|-------------|-----------|-----------------|
| 1       | A2M  | Apr 1, 2016 | The A2 Milk C...  | Consumer St...   | 1964410000  | 0.120     | 26196187)       |
| 2       | AAC  | Apr 1, 2016 | Australian Ag...  | Consumer St...   | 927459000   | 0.060     | 33986772)       |
| 3       | ALU  | Apr 1, 2016 | Altium Limited    | Information T... | 989640000   | 0.060     | 17366687)       |
| 4       | APN  | Apr 1, 2016 | APN News & ...    | Consumer Di...   | 854834000   | 0.050     | 16562953)       |
| 5       | BAP  | Apr 2, 2016 | Bapcor Limited    | Consumer Di...   | 1619670000  | 0.100     | 22878183)       |
| 6       | BEN  | Apr 2, 2016 | Bendigo And ...   | Financials       | 5738350000  | 0.340     | 30901815)       |
| 7       | BXB  | Apr 2, 2016 | Brambles Li...    | Industrials      | 14853800000 | 0.890     | 14637229)       |
| 8       | CSL  | Apr 3, 2016 | CSL Limited       | Health Care      | 57047600000 | 3.420     | 15877428)       |
| 9       | CTX  | Apr 3, 2016 | Caltex Austral... | Energy           | 7691300000  | 0.460     | 15810313)       |
| 10      | CWY  | Apr 3, 2016 | Cleanaway W...    | Industrials      | 1909120000  | 0.110     | 17906023)       |
| 11      | DMP  | Apr 3, 2016 | Dominos Pizz...   | Consumer Di...   | 5165210000  | 0.310     | 26006390)       |
| 12      | DUE  | Apr 3, 2016 | Duet Group F...   | Utilities        | 6893450000  | 0.410     | 30516312)       |
| 13      | EVN  | Apr 3, 2016 | Evolution Mini... | Materials        | 3533880000  | 0.210     | 31473026)       |
| 14      | FLT  | Apr 3, 2016 | Flight Centre ... | Consumer Di...   | 2917940000  | 0.180     | 26727009)       |
| 15      | FPH  | Apr 3, 2016 | Fisher & Payk...  | Health Care      | 5018000000  | 0.300     | 22751676)       |
| 16      | GNC  | Apr 3, 2016 | Graincorp Li...   | Consumer St...   | 2080300000  | 0.120     | 28093765)       |
| 17      | GUD  | Apr 3, 2016 | G.u.d. Holdin...  | Consumer Di...   | 1015160000  | 0.060     | 29599469)       |
| 18      | GXL  | Apr 3, 2016 | Greencross L...   | Consumer Di...   | 836296000   | 0.050     | 22578274)       |
| 19      | IAG  | Apr 3, 2016 | Insurance Au...   | Financials       | 1432350000  | 0.860     | 32357731)       |
| 20      | ITD  | Apr 3, 2016 | Intertape Dr...   | Financials       | 662000000   | 0.150     | 22740400)       |

ExampleSet (50 examples, 0 special attributes, 7 regular attributes)

And the other with 150 members.

ExampleSet (Split Data) ExampleSet (Split Data)

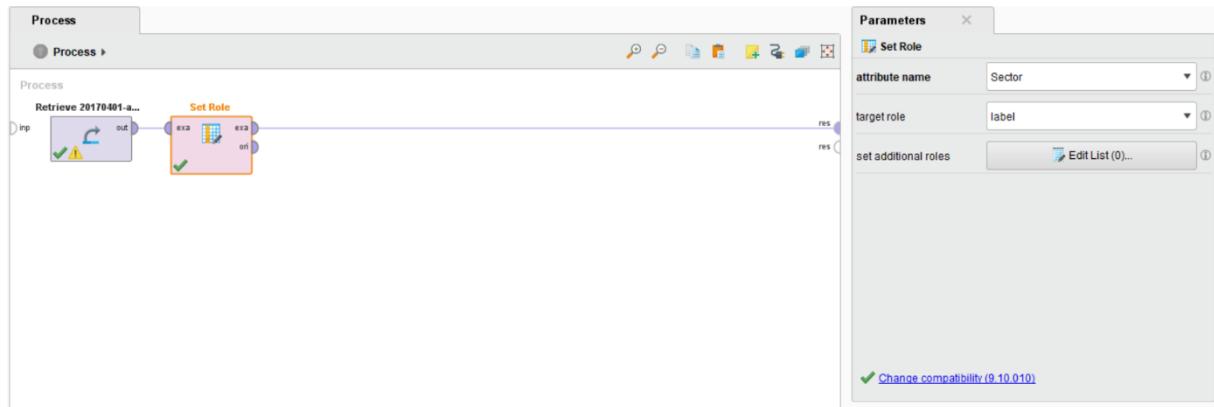
Open in [Turbo Prep](#) [Auto Model](#)

| Row No. | Code | Date        | Company           | Sector           | Market Cap  | Weight(%) | Total Index ... |
|---------|------|-------------|-------------------|------------------|-------------|-----------|-----------------|
| 1       | AAD  | Apr 1, 2016 | Ardent Leisur...  | Consumer Di...   | 858443000   | 0.050     | 15775360)       |
| 2       | ABC  | Apr 1, 2016 | Adelaide Brig...  | Materials        | 3690040000  | 0.220     | 22283791)       |
| 3       | ABP  | Apr 1, 2016 | Abacus Prop...    | Real Estate      | 1864850000  | 0.110     | 22339859)       |
| 4       | ACX  | Apr 1, 2016 | Aconex Limited    | Information T... | 779031000   | 0.050     | 15359144)       |
| 5       | AGL  | Apr 1, 2016 | AGL Energy L...   | Utilities        | 17651200000 | 1.060     | 18306588)       |
| 6       | AHG  | Apr 1, 2016 | Automotive H...   | Consumer Di...   | 1366290000  | 0.080     | 29832105)       |
| 7       | AHY  | Apr 1, 2016 | Asaleo Care ...   | Consumer St...   | 965307000   | 0.060     | 18414473)       |
| 8       | ALL  | Apr 1, 2016 | Aristocrat Lei... | Consumer Di...   | 11474600000 | 0.690     | 29185453)       |
| 9       | ALQ  | Apr 1, 2016 | Als Limited       | Industrials      | 3095920000  | 0.190     | 30638213)       |
| 10      | AMC  | Apr 1, 2016 | Amcor Limited     | Materials        | 17441600000 | 1.050     | 19688815)       |
| 11      | AMP  | Apr 1, 2016 | AMP Limited       | Financials       | 15321100000 | 0.920     | 33855664)       |
| 12      | ANN  | Apr 1, 2016 | Ansell Limited    | Health Care      | 3551570000  | 0.210     | 27701216)       |
| 13      | ANZ  | Apr 1, 2016 | Australia And ... | Financials       | 93424700000 | 5.600     | 17612164)       |
| 14      | AOG  | Apr 1, 2016 | Aveo Group S...   | Real Estate      | 1866090000  | 0.110     | 26860334)       |
| 15      | APA  | Apr 1, 2016 | APA Group St...   | Utilities        | 9984190000  | 0.600     | 30037727)       |
| 16      | API  | Apr 1, 2016 | Australian Ph...  | Health Care      | 999193000   | 0.060     | 29784239)       |
| 17      | APO  | Apr 1, 2016 | Apn Outdoor ...   | Consumer Di...   | 949703000   | 0.060     | 27315482)       |
| 18      | ARB  | Apr 1, 2016 | ARB Corporat...   | Consumer Di...   | 1165590000  | 0.070     | 16137397)       |
| 19      | AST  | Apr 1, 2016 | Ausnet Servic...  | Utilities        | 6071320000  | 0.360     | 19931265)       |
| 20      | AVM  | Apr 1, 2016 | AVM Limited       | Financials       | 8770000000  | 0.500     | 14054450)       |

ExampleSet (150 examples, 0 special attributes, 7 regular attributes)

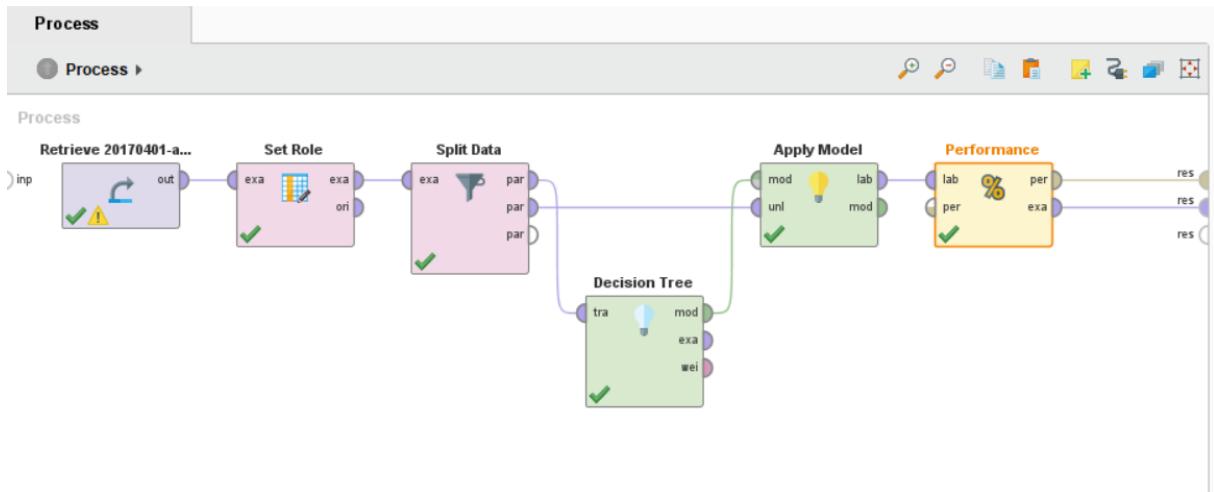
### Section 3: Setting Role

As the data is loaded not from samples, but from a repository, it must have a role to be set. So, the “Set Role” operator is employed by its “attribute name” as “sector” and “target role” as “label”.



### Section 4: Other Operators

We need “Split Data” after “Set Role”, then “Decision Tree”, “Apply Model”, and “Performance” to complete the process as the design is shown below.



## Section 5: Running the Model

It is noticeable that the prediction for the sector and its real sector are not the same.

| ExampleSet (Apply Model) |                    |                    | PerformanceVector (Performance) |                 |                 |                 |                 |                 |                 |                 |                 |                 |                 |                 |
|--------------------------|--------------------|--------------------|---------------------------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
|                          |                    |                    | Filter (47 / 47 examples): all  |                 |                 |                 |                 |                 |                 |                 |                 |                 |                 |                 |
| Row No.                  | Sector             | prediction(Sector) | confidence(...)                 | confidence(...) | confidence(...) | confidence(...) | confidence(...) | confidence(...) | confidence(...) | confidence(...) | confidence(...) | confidence(...) | confidence(...) | confidence(...) |
| 1                        | Materials          | Materials          | 0                               | 0               | 0.333           | 0               | 0               | 0               | 0.333           | 0               | 0.333           | 0               | 0               | 0               |
| 2                        | Information Tec... | Consumer Staples   | 0.500                           | 0.500           | 0               | 0               | 0               | 0               | 0               | 0               | 0               | 0               | 0               | 0               |
| 3                        | Materials          | Utilities          | 0                               | 0.167           | 0               | 0               | 0               | 0.333           | 0.167           | 0.333           | 0               | 0               | 0               | 0               |
| 4                        | Financials         | Utilities          | 0                               | 0.167           | 0               | 0               | 0               | 0.333           | 0.167           | 0.333           | 0               | 0               | 0               | 0               |
| 5                        | Utilities          | Utilities          | 0                               | 0.167           | 0               | 0               | 0               | 0.333           | 0.167           | 0.333           | 0               | 0               | 0               | 0               |
| 6                        | Consumer Disc...   | Consumer Staples   | 0.500                           | 0.500           | 0               | 0               | 0               | 0               | 0               | 0               | 0               | 0               | 0               | 0               |
| 7                        | Consumer Stap...   | Consumer Staples   | 0.500                           | 0.500           | 0               | 0               | 0               | 0               | 0               | 0               | 0               | 0               | 0               | 0               |
| 8                        | Materials          | Industrials        | 0                               | 0               | 0               | 0               | 0               | 0               | 0.667           | 0.333           | 0               | 0               | 0               | 0               |
| 9                        | Consumer Stap...   | Materials          | 0                               | 0               | 1               | 0               | 0               | 0               | 0               | 0               | 0               | 0               | 0               | 0               |
| 10                       | Financials         | Consumer Staples   | 0.667                           | 0.333           | 0               | 0               | 0               | 0               | 0               | 0               | 0               | 0               | 0               | 0               |
| 11                       | Financials         | Materials          | 0                               | 0               | 1               | 0               | 0               | 0               | 0               | 0               | 0               | 0               | 0               | 0               |
| 12                       | Real Estate        | Materials          | 0                               | 0               | 0.500           | 0               | 0.500           | 0               | 0               | 0               | 0               | 0               | 0               | 0               |
| 13                       | Health Care        | Materials          | 0.020                           | 0.192           | 0.222           | 0.101           | 0.051           | 0.030           | 0.101           | 0.131           | 0.121           | 0.020           |                 |                 |
| 14                       | Real Estate        | Materials          | 0.020                           | 0.192           | 0.222           | 0.101           | 0.051           | 0.030           | 0.101           | 0.131           | 0.121           | 0.020           |                 |                 |
| 15                       | Financials         | Materials          | 0.020                           | 0.192           | 0.222           | 0.101           | 0.051           | 0.030           | 0.101           | 0.131           | 0.121           | 0.020           |                 |                 |
| 16                       | Consumer Disc...   | Materials          | 0.020                           | 0.192           | 0.222           | 0.101           | 0.051           | 0.030           | 0.101           | 0.131           | 0.121           | 0.020           |                 |                 |
| 17                       | Consumer Disc...   | Materials          | 0.020                           | 0.192           | 0.222           | 0.101           | 0.051           | 0.030           | 0.101           | 0.131           | 0.121           | 0.020           |                 |                 |
| 18                       | Health Care        | Materials          | 0.020                           | 0.192           | 0.222           | 0.101           | 0.051           | 0.030           | 0.101           | 0.131           | 0.121           | 0.020           |                 |                 |
| 19                       | Real Estate        | Materials          | 0.020                           | 0.192           | 0.222           | 0.101           | 0.051           | 0.030           | 0.101           | 0.131           | 0.121           | 0.020           |                 |                 |

And the prediction which is prepared in the diagonal line shows very low correctness in prediction. It is obvious from the final result of the last chapter that the data is not highly predictable intrinsically. In total, the prediction has about 21 percent of accuracy.

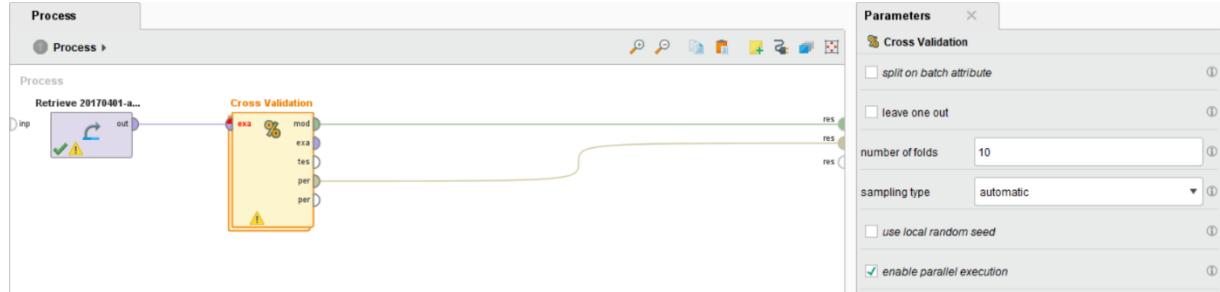
| Criterion               |        | accuracy      |               |                |                |                 |                |                  |                 |                 |             |                |                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|-------------------------|--------|---------------|---------------|----------------|----------------|-----------------|----------------|------------------|-----------------|-----------------|-------------|----------------|-----------------|--|--|--|--|--|--|--|--|--|--|--|--|--|--|
|                         |        | Table View    |               |                |                |                 |                |                  |                 |                 |             |                |                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| <b>accuracy: 21.28%</b> |        |               |               |                |                |                 |                |                  |                 |                 |             |                |                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                         |        |               |               |                |                |                 |                |                  |                 |                 |             |                |                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                         |        | true Consu... | true Consu... | true Materials | true Real E... | true Informa... | true Utilities | true Industri... | true Financi... | true Health ... | true Energy | true Teleco... | class precis... |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| pred. Consu...          | 1      | 1             | 0             | 0              | 1              | 0               | 0              | 1                | 0               | 0               | 0           | 0              | 25.00%          |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| pred. Consu...          | 0      | 0             | 0             | 0              | 0              | 0               | 0              | 0                | 0               | 0               | 0           | 0              | 0.00%           |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| pred. Materi...         | 2      | 7             | 7             | 5              | 1              | 0               | 5              | 4                | 3               | 2               | 0           | 0              | 19.44%          |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| pred. Real E...         | 0      | 0             | 0             | 0              | 0              | 0               | 0              | 0                | 0               | 0               | 0           | 0              | 0.00%           |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| pred. Inform...         | 0      | 0             | 0             | 0              | 0              | 0               | 0              | 0                | 0               | 0               | 0           | 0              | 0.00%           |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| pred. Utilities         | 0      | 0             | 1             | 0              | 0              | 1               | 0              | 1                | 0               | 0               | 0           | 0              | 33.33%          |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| pred. Indust...         | 0      | 0             | 1             | 0              | 0              | 0               | 0              | 0                | 0               | 0               | 0           | 0              | 0.00%           |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| pred. Financ...         | 0      | 0             | 0             | 0              | 0              | 0               | 0              | 0                | 0               | 0               | 0           | 0              | 0.00%           |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| pred. Health...         | 0      | 0             | 0             | 0              | 0              | 0               | 0              | 0                | 0               | 0               | 0           | 0              | 0.00%           |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| pred. Energy            | 0      | 0             | 0             | 0              | 0              | 0               | 0              | 0                | 1               | 0               | 0           | 0              | 0.00%           |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| pred. Teleco...         | 0      | 1             | 0             | 0              | 0              | 0               | 0              | 0                | 0               | 0               | 0           | 1              | 50.00%          |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| class recall            | 33.33% | 0.00%         | 77.78%        | 0.00%          | 0.00%          | 100.00%         | 0.00%          | 0.00%            | 0.00%           | 0.00%           | 0.00%       | 100.00%        |                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

## Chapter 3

In this chapter, the same data is used and analyzed by cross-validation.

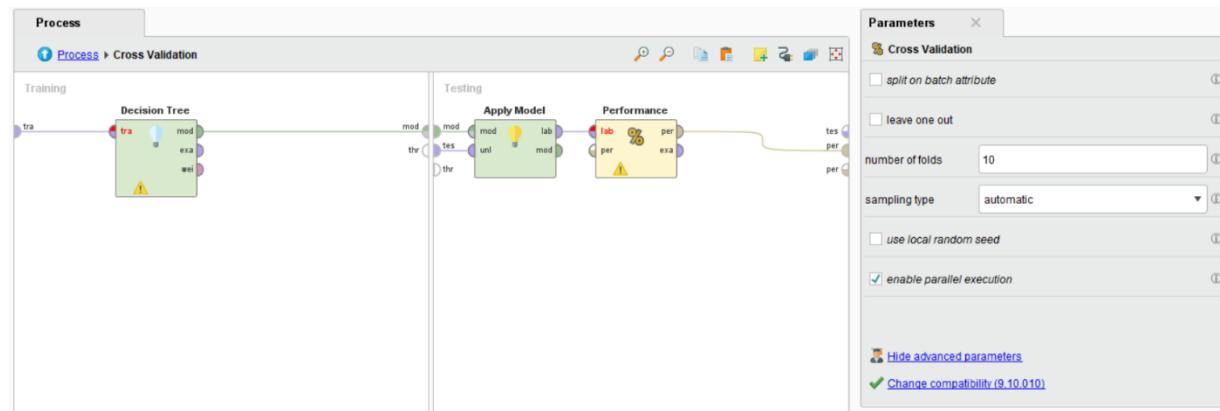
### Section 1: Cross Validation

Connecting the data by the operator called “Cross Validation” by two final connections, “mod” and “Per”.



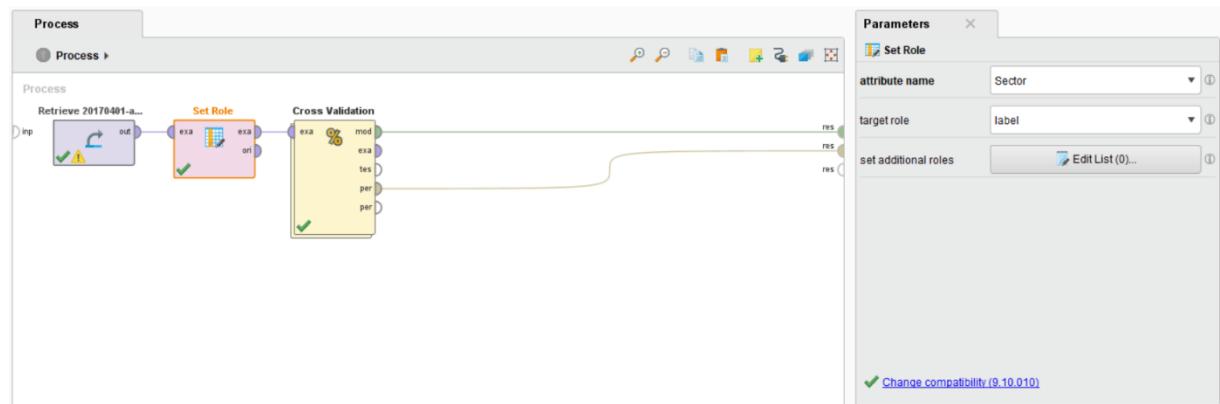
### Section 2: Cross Validations' Operators

Inside the “Cross Validation”, the operators of “Decision Tree”, “Apply Model”, and “Performance” are being used. Pay attention that “per” must be connected to “per” in “Performance” and also to the point of having the “number of folds” of 10.



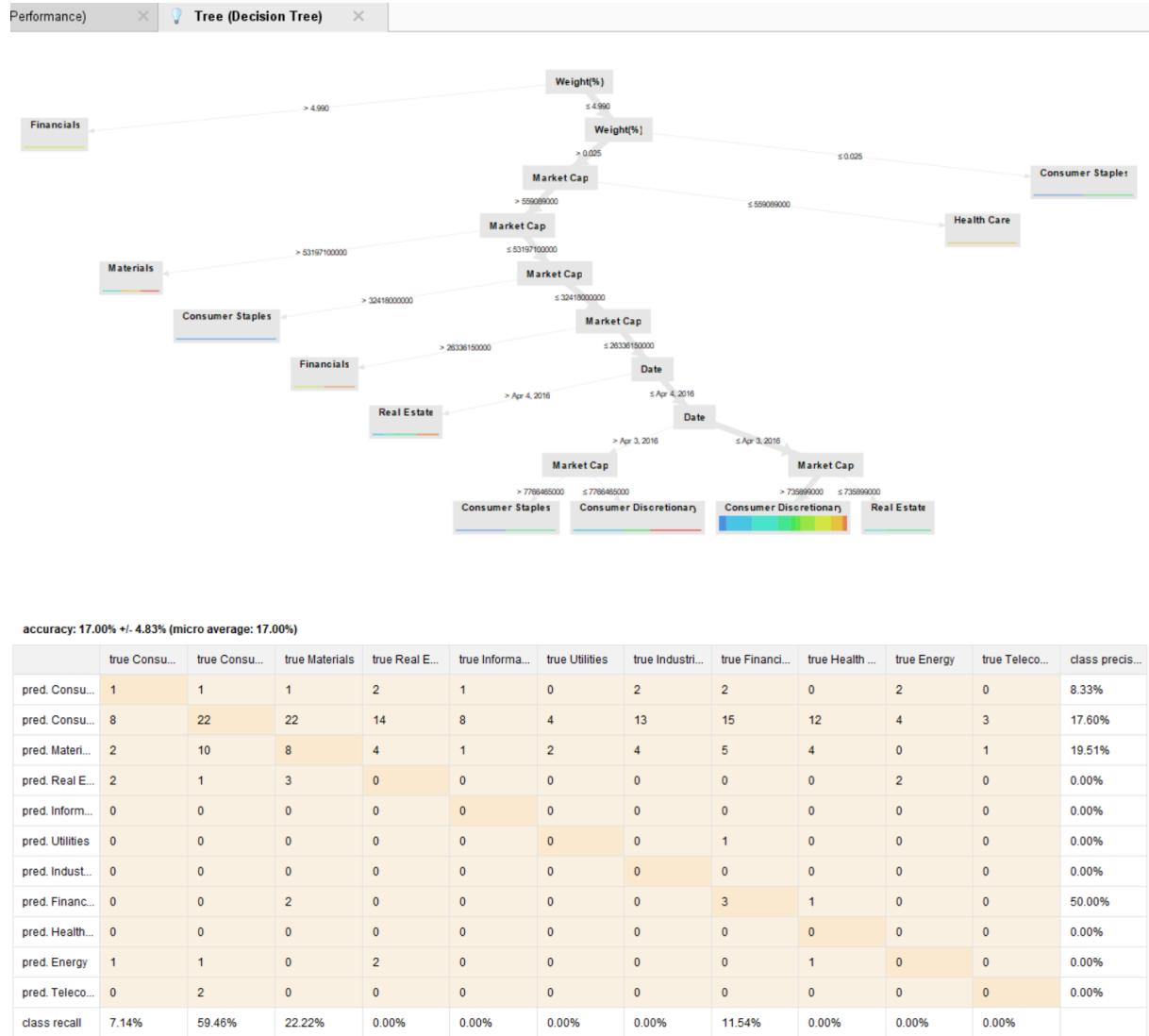
### Section 3: Setting Role

But there was an error, so the “Set Role” to provide the label has been added. The “Sector” as “label” is set.



## Section 4: Results

The results are as follows. Regardless of the “Decision Tree”, the ““Performance”” shows about 17 percent of accuracy. Compared to the last chapter’s outcome, it is less valuable.



## Chapter 4

Using the “Naive Bayes” to evaluate the same data

### Section 1: Dataset, Setting Role, and Split its Data

In setting the role, the label is named for “Sector”.

In split data, it must have the percentages to be able to operate. So, in this time like in chapter 2, the weights of 75 and 25 percent are applied.

The result is the same as before.

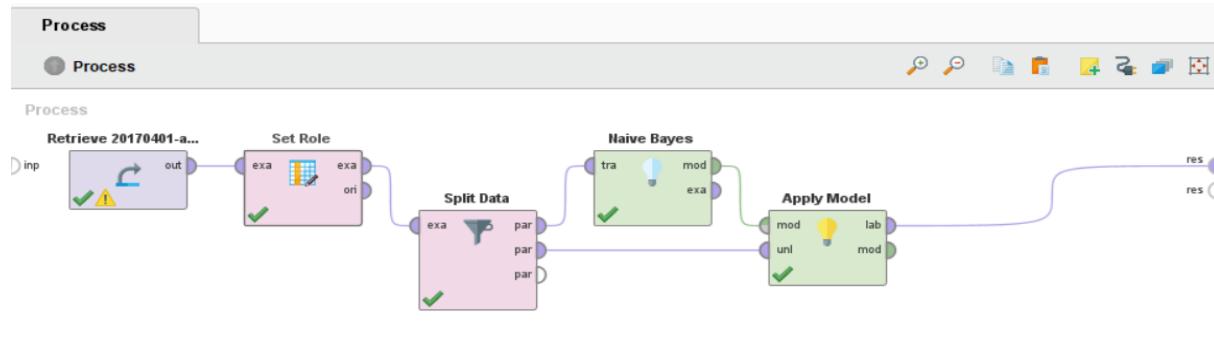
---

Open in [!\[\]\(497ef498b0f335171f109e4718220ab7\_img.jpg\) Turbo Prep](#) [!\[\]\(fe9ce875776950d006aea88a8ca4569b\_img.jpg\) Auto Model](#)

| Row No. | Sector                 | Code | Date        | Company           | Market Cap  | Weight(%) | Total Index ... |
|---------|------------------------|------|-------------|-------------------|-------------|-----------|-----------------|
| 1       | Consumer Staples       | A2M  | Apr 1, 2016 | The A2 Milk C...  | 1964410000  | 0.120     | 26196187)       |
| 2       | Consumer Staples       | AAC  | Apr 1, 2016 | Australian Ag...  | 927459000   | 0.060     | 33986772)       |
| 3       | Consumer Discretionary | AAD  | Apr 1, 2016 | Ardent Leisur...  | 858443000   | 0.050     | 15775360)       |
| 4       | Real Estate            | ABP  | Apr 1, 2016 | Abacus Prop...    | 1864850000  | 0.110     | 22339859)       |
| 5       | Utilities              | AGL  | Apr 1, 2016 | AGL Energy L...   | 17651200000 | 1.060     | 18306588)       |
| 6       | Consumer Discretionary | AHG  | Apr 1, 2016 | Automotive H...   | 1366290000  | 0.080     | 29832105)       |
| 7       | Consumer Staples       | AHY  | Apr 1, 2016 | Asaleo Care ...   | 965307000   | 0.060     | 18414473)       |
| 8       | Consumer Discretionary | ALL  | Apr 1, 2016 | Aristocrat Lei... | 11474600000 | 0.690     | 29185453)       |
| 9       | Industrials            | ALQ  | Apr 1, 2016 | Als Limited       | 3095920000  | 0.190     | 30638213)       |
| 10      | Information Technology | ALU  | Apr 1, 2016 | Altium Limited    | 989640000   | 0.060     | 17366687)       |
| 11      | Financials             | AMP  | Apr 1, 2016 | AMP Limited       | 15321100000 | 0.920     | 33855664)       |
| 12      | Health Care            | ANN  | Apr 1, 2016 | Ansell Limited    | 3551570000  | 0.210     | 27701216)       |
| 13      | Real Estate            | AOG  | Apr 1, 2016 | Aveo Group S...   | 1866090000  | 0.110     | 26860334)       |
| 14      | Health Care            | API  | Apr 1, 2016 | Australian Ph...  | 999193000   | 0.060     | 29784239)       |
| 15      | Consumer Discretionary | APO  | Apr 1, 2016 | Apn Outdoor ...   | 949703000   | 0.060     | 27315482)       |
| 16      | Consumer Discretionary | ARB  | Apr 1, 2016 | ARB Corporat...   | 1165590000  | 0.070     | 16137397)       |
| 17      | Utilities              | AST  | Apr 1, 2016 | Ausnet Servic...  | 6071320000  | 0.360     | 19931265)       |
| 18      | Financials             | ASX  | Apr 1, 2016 | ASX Limited       | 9772680000  | 0.590     | 14354150)       |
| 19      | Materials              | AWC  | Apr 1, 2016 | Alumina Limit...  | 5154920000  | 0.310     | 24803538)       |
| 20      | Industrials            | ATI  | Apr 1, 2016 | Autoliv Limit...  | 40774700000 | 0.650     | 65104670)       |

## Section 2: Adding “Apply Model” and “Naive Bayes”

One connection is from “split data” which is the main one to the “Naive Bayes” and the other to the “Apply Model”.



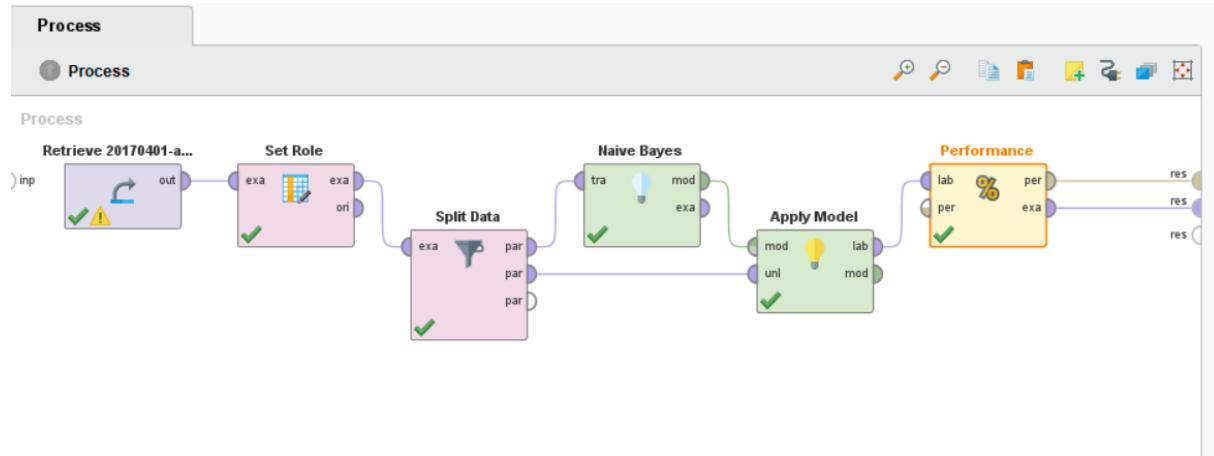
So far, the result is as below. As we see, the prediction is not very accurate because of the reason which is mentioned in the last chapters, as data is not intrinsically predictable.

Open in [Turbo Prep](#) [Auto Model](#) Filter (47 / 47 examples): all

| Row No. | Sector                 | prediction(Sector)     | confidence(...) |
|---------|------------------------|------------------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| 1       | Materials              | Information Technology | 0.014           | 0.058           | 0.001           | 0.031           | 0.448           | 0.371           | 0.054           | 0.000           | 0.003           |
| 2       | Information Technology | Information Technology | 0.017           | 0.059           | 0.001           | 0.025           | 0.556           | 0.279           | 0.039           | 0.000           | 0.004           |
| 3       | Materials              | Utilities              | 0.204           | 0.000           | 0.000           | 0.004           | 0               | 0.562           | 0.000           | 0.006           | 0.033           |
| 4       | Financials             | Financials             | 0.000           | 0               | 0               | 0               | 0               | 0               | 1.000           | 0.000           | 0.000           |
| 5       | Utilities              | Utilities              | 0.032           | 0.000           | 0.001           | 0.036           | 0.000           | 0.836           | 0.033           | 0.001           | 0.006           |
| 6       | Consumer Discretionary | Information Technology | 0.017           | 0.059           | 0.001           | 0.025           | 0.559           | 0.277           | 0.039           | 0.000           | 0.004           |
| 7       | Consumer Staples       | Information Technology | 0.021           | 0.058           | 0.001           | 0.028           | 0.517           | 0.305           | 0.042           | 0.001           | 0.004           |
| 8       | Materials              | Financials             | 0.000           | 0               | 0               | 0               | 0               | 0               | 0               | 0.859           | 0.000           |
| 9       | Consumer Staples       | Utilities              | 0.019           | 0.003           | 0.023           | 0.067           | 0.003           | 0.634           | 0.155           | 0.001           | 0.023           |
| 10      | Financials             | Information Technology | 0.007           | 0.056           | 0.009           | 0.021           | 0.689           | 0.125           | 0.065           | 0.001           | 0.009           |
| 11      | Financials             | Utilities              | 0.016           | 0.011           | 0.023           | 0.070           | 0.026           | 0.581           | 0.188           | 0.001           | 0.020           |
| 12      | Real Estate            | Information Technology | 0.005           | 0.056           | 0.008           | 0.020           | 0.705           | 0.119           | 0.065           | 0.000           | 0.006           |
| 13      | Health Care            | Information Technology | 0.006           | 0.052           | 0.017           | 0.024           | 0.719           | 0.074           | 0.056           | 0.001           | 0.010           |
| 14      | Real Estate            | Utilities              | 0.021           | 0.001           | 0.049           | 0.081           | 0.000           | 0.474           | 0.125           | 0.002           | 0.036           |
| 15      | Financials             | Information Technology | 0.004           | 0.054           | 0.016           | 0.024           | 0.718           | 0.079           | 0.063           | 0.000           | 0.007           |
| 16      | Consumer Discretionary | Information Technology | 0.009           | 0.051           | 0.036           | 0.059           | 0.387           | 0.215           | 0.146           | 0.001           | 0.015           |
| 17      | Consumer Discretionary | Information Technology | 0.004           | 0.051           | 0.015           | 0.021           | 0.744           | 0.068           | 0.053           | 0.000           | 0.008           |
| 18      | Health Care            | Information Technology | 0.007           | 0.052           | 0.019           | 0.025           | 0.698           | 0.080           | 0.059           | 0.001           | 0.012           |
| 19      | Real Estate            | Utilities              | 0.020           | 0.001           | 0.051           | 0.083           | 0.001           | 0.470           | 0.135           | 0.002           | 0.035           |

### Section 3: "Performance"

By adding the operator of ““Performance””, the accuracy of prediction with about 8 percent is prepared, very lower than the prediction accuracies in the prior chapters.



accuracy: 8.51%

|                 | true Consu... | true Consu... | true Materials | true Real E... | true Informa... | true Utilities | true Industri... | true Financi... | true Health ... | true Energy | true Teleco... | class precis... |
|-----------------|---------------|---------------|----------------|----------------|-----------------|----------------|------------------|-----------------|-----------------|-------------|----------------|-----------------|
| pred. Consu...  | 0             | 0             | 0              | 0              | 0               | 0              | 0                | 0               | 0               | 0           | 0              | 0.00%           |
| pred. Consu...  | 0             | 0             | 0              | 0              | 0               | 0              | 0                | 0               | 0               | 0           | 0              | 0.00%           |
| pred. Materi... | 0             | 0             | 0              | 0              | 0               | 0              | 0                | 0               | 0               | 0           | 0              | 0.00%           |
| pred. Real E... | 0             | 0             | 0              | 0              | 0               | 0              | 0                | 0               | 0               | 0           | 0              | 0.00%           |
| pred. Inform... | 2             | 9             | 6              | 2              | 2               | 0              | 3                | 2               | 3               | 0           | 1              | 6.67%           |
| pred. Utilities | 1             | 0             | 1              | 2              | 0               | 1              | 1                | 2               | 1               | 2           | 0              | 9.09%           |
| pred. Indust... | 0             | 0             | 0              | 0              | 0               | 0              | 0                | 0               | 0               | 0           | 0              | 0.00%           |
| pred. Financ... | 0             | 0             | 1              | 0              | 0               | 0              | 0                | 1               | 0               | 0           | 0              | 50.00%          |
| pred. Health... | 0             | 0             | 0              | 0              | 0               | 0              | 0                | 0               | 0               | 0           | 0              | 0.00%           |
| pred. Energy    | 0             | 0             | 1              | 0              | 0               | 0              | 0                | 0               | 0               | 0           | 0              | 0.00%           |
| pred. Teleco... | 0             | 0             | 0              | 1              | 0               | 0              | 1                | 1               | 0               | 0           | 0              | 0.00%           |
| class recall    | 0.00%         | 0.00%         | 0.00%          | 0.00%          | 100.00%         | 100.00%        | 0.00%            | 16.67%          | 0.00%           | 0.00%       | 0.00%          |                 |

## Chapter 5

In this chapter, the data will be multiplied and then by using cross-validation, it will be assessed.

### Section 1: Using the “Multiply” operator.

By adding this operator, the data is doubled, meaning we have a copy of the dataset to perform further analysis.

The screenshot shows the KNIME interface with a process flow and a data table.

**Process Flow:**

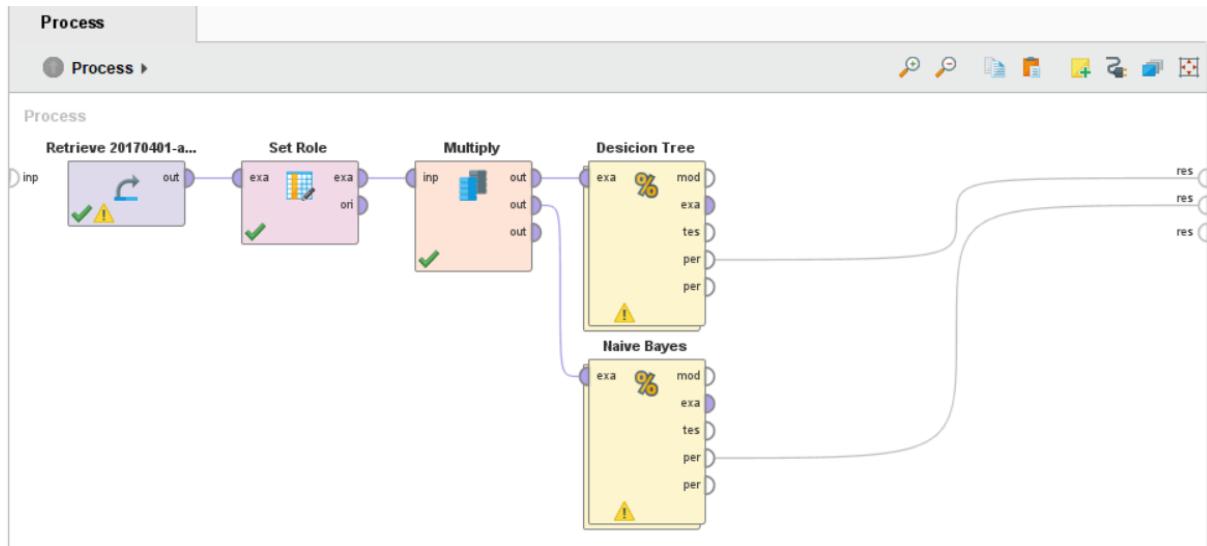
```
graph LR; A[Retrieve 20170401-a... ] --> B[Set Role]; B --> C[Multiply]; C --> D[ExampleSet (Multiply)];
```

**Data Table:**

| Row No. | Sector           | Code | Date        | Company           | Market Cap  | Weight(%) | Total Index. |
|---------|------------------|------|-------------|-------------------|-------------|-----------|--------------|
| 1       | Consumer St...   | A2M  | Apr 1, 2016 | The A2 Milk C...  | 1964410000  | 0.120     | 26196187)    |
| 2       | Consumer St...   | AAC  | Apr 1, 2016 | Australian Ag...  | 927459000   | 0.060     | 33986772)    |
| 3       | Consumer Di...   | AAD  | Apr 1, 2016 | Ardent Leisur...  | 858443000   | 0.050     | 15775360)    |
| 4       | Materials        | ABC  | Apr 1, 2016 | Adelaide Brig...  | 3690040000  | 0.220     | 22283791)    |
| 5       | Real Estate      | ABP  | Apr 1, 2016 | Abacus Prop...    | 1864850000  | 0.110     | 22339859)    |
| 6       | Information T... | ACX  | Apr 1, 2016 | Aconex Limited    | 779031000   | 0.050     | 15359144)    |
| 7       | Utilities        | AGL  | Apr 1, 2016 | AGL Energy L...   | 17651200000 | 1.060     | 18306588)    |
| 8       | Consumer Di...   | AHG  | Apr 1, 2016 | Automotive H...   | 1366290000  | 0.080     | 29832105)    |
| 9       | Consumer St...   | AHY  | Apr 1, 2016 | Asaleo Care ...   | 965307000   | 0.060     | 18414473)    |
| 10      | Consumer Di...   | ALL  | Apr 1, 2016 | Aristocrat Lei... | 11474600000 | 0.690     | 29185453)    |
| 11      | Industrials      | ALQ  | Apr 1, 2016 | Als Limited       | 3095920000  | 0.190     | 30638213)    |
| 12      | Information T... | ALU  | Apr 1, 2016 | Altium Limited    | 989640000   | 0.060     | 17366687)    |
| 13      | Materials        | AMC  | Apr 1, 2016 | Amcor Limited     | 17441600000 | 1.050     | 19688815)    |
| 14      | Financials       | AMP  | Apr 1, 2016 | AMP Limited       | 15321100000 | 0.920     | 33855664)    |
| 15      | Health Care      | ANN  | Apr 1, 2016 | Ansell Limited    | 3551570000  | 0.210     | 27701216)    |
| 16      | Financials       | ANZ  | Apr 1, 2016 | Australia And ... | 93424700000 | 5.600     | 17612164)    |
| 17      | Real Estate      | AOG  | Apr 1, 2016 | Aveo Group S...   | 1866090000  | 0.110     | 26860334)    |
| 18      | Utilities        | APA  | Apr 1, 2016 | APA Group St...   | 9984190000  | 0.600     | 30037727)    |
| 19      | Health Care      | API  | Apr 1, 2016 | Australian Ph...  | 999193000   | 0.060     | 29784239)    |
| 20      | Consumer St...   | APU  | Apr 1, 2016 | Autoliv Inc.      | 254001000   | 0.050     | 40000000)    |

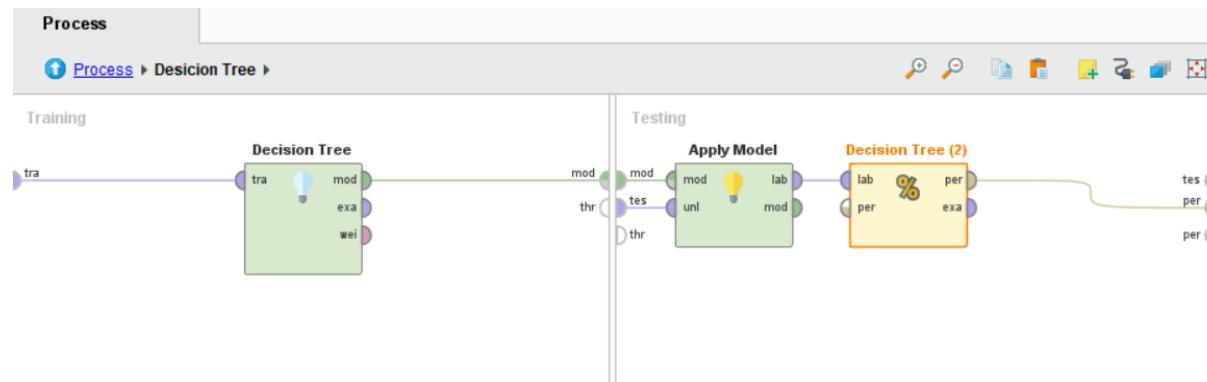
## Section 2: Adding two “Cross Validation”

One of them is for “Decision Tree” and the other is for “Naïve Bayes”. So after adding, they are renamed.

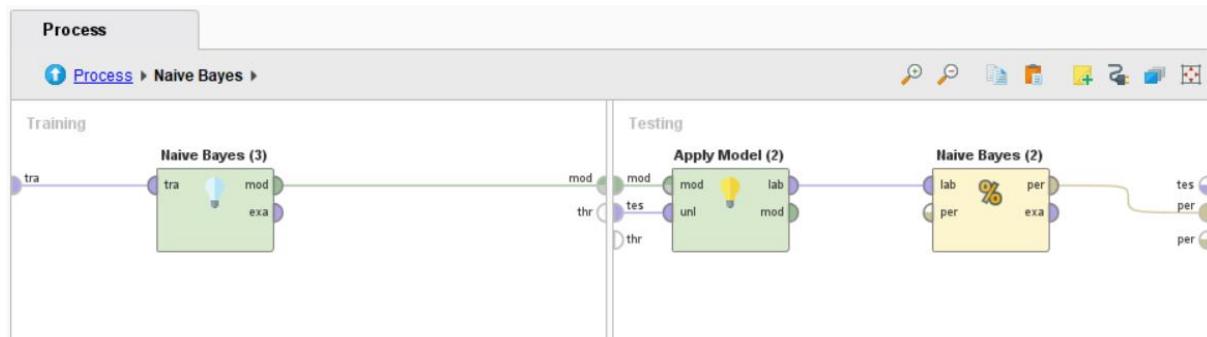


## Section 3: Providing the Cross Validations’ Inside

The first cross-validation includes the decision tree. Additionally, the “Apply Model” and “Performance” operators are added. To be more exact for the final results, the “Performance” operators are renamed to their basic operators accordingly.



The second cross\_validation includes the Naïve Bayes. Additionally, the “Apply Model” and “Performance” operators are added. To be more exact for the final results, the “Performance” operators are renamed in this step to “Naïve Bayes”.



## Section 4: The Result

The accuracy of the result of “Naïve Bayes” is about 7%, while it is about 17% in the “Decision Tree”. It is very interesting that “Decision Tree” can predict far better than “Naïve Bayes” in this case.

Coming back to section 5 of chapter 2, the result for accuracy is more than 21 percent which is again based on the “Decision Tree”.

PerformanceVector (Naïve Bayes (2))
PerformanceVector (Decision Tree (2))

Criterion
accuracy

Table View
 Plot View

accuracy: 7.00% +/- 3.50% (micro average: 7.00%)

|                 | true Consu... | true Consu... | true Materials | true Real E... | true Informa... | true Utilities | true Industri... | true Financi... | true Health ... | true Energy | true Teleco... | class precis... |
|-----------------|---------------|---------------|----------------|----------------|-----------------|----------------|------------------|-----------------|-----------------|-------------|----------------|-----------------|
| pred. Consu...  | 0             | 0             | 0              | 0              | 0               | 1              | 0                | 2               | 0               | 0           | 0              | 0.00%           |
| pred. Consu...  | 0             | 0             | 0              | 0              | 0               | 0              | 0                | 0               | 0               | 0           | 0              | 0.00%           |
| pred. Materi... | 0             | 0             | 0              | 0              | 0               | 0              | 0                | 0               | 0               | 0           | 0              | 0.00%           |
| pred. Real E... | 0             | 0             | 0              | 0              | 0               | 0              | 0                | 0               | 0               | 0           | 0              | 0.00%           |
| pred. Inform... | 10            | 31            | 23             | 12             | 8               | 2              | 12               | 11              | 12              | 3           | 1              | 6.40%           |
| pred. Utilities | 1             | 4             | 8              | 5              | 2               | 3              | 5                | 7               | 4               | 4           | 1              | 6.82%           |
| pred. Indust... | 0             | 0             | 0              | 0              | 0               | 0              | 0                | 0               | 0               | 0           | 0              | 0.00%           |
| pred. Financ... | 1             | 0             | 1              | 0              | 0               | 0              | 0                | 3               | 1               | 0           | 1              | 42.86%          |
| pred. Health... | 0             | 0             | 0              | 0              | 0               | 0              | 0                | 0               | 0               | 1           | 0              | 0.00%           |
| pred. Energy    | 1             | 1             | 2              | 3              | 0               | 0              | 1                | 2               | 1               | 0           | 1              | 0.00%           |
| pred. Teleco... | 1             | 1             | 2              | 2              | 0               | 0              | 1                | 1               | 0               | 0           | 0              | 0.00%           |
| class recall    | 0.00%         | 0.00%         | 0.00%          | 0.00%          | 80.00%          | 50.00%         | 0.00%            | 11.54%          | 0.00%           | 0.00%       | 0.00%          |                 |

|                 | true Consu... | true Consu... | true Materials | true Real E... | true Informa... | true Utilities | true Industri... | true Financi... | true Health ... | true Energy | true Teleco... | class precis... |
|-----------------|---------------|---------------|----------------|----------------|-----------------|----------------|------------------|-----------------|-----------------|-------------|----------------|-----------------|
| pred. Consu...  | 1             | 2             | 0              | 1              | 0               | 0              | 0                | 0               | 2               | 0           | 1              | 14.29%          |
| pred. Consu...  | 6             | 13            | 14             | 8              | 3               | 3              | 8                | 6               | 6               | 3           | 1              | 18.31%          |
| pred. Materi... | 2             | 17            | 13             | 10             | 4               | 3              | 10               | 13              | 7               | 2           | 1              | 15.85%          |
| pred. Real E... | 1             | 1             | 1              | 0              | 1               | 0              | 0                | 0               | 1               | 3           | 0              | 0.00%           |
| pred. Inform... | 0             | 0             | 0              | 0              | 0               | 0              | 0                | 0               | 0               | 0           | 0              | 0.00%           |
| pred. Utilities | 0             | 0             | 2              | 0              | 1               | 0              | 0                | 2               | 1               | 0           | 0              | 0.00%           |
| pred. Indust... | 0             | 0             | 1              | 0              | 0               | 0              | 0                | 0               | 0               | 0           | 0              | 0.00%           |
| pred. Financ... | 2             | 1             | 4              | 1              | 0               | 0              | 1                | 5               | 0               | 0           | 0              | 35.71%          |
| pred. Health... | 0             | 0             | 0              | 0              | 0               | 0              | 0                | 0               | 0               | 0           | 0              | 0.00%           |
| pred. Energy    | 1             | 1             | 1              | 2              | 0               | 0              | 0                | 0               | 1               | 0           | 0              | 0.00%           |
| pred. Teleco... | 1             | 2             | 0              | 0              | 1               | 0              | 0                | 0               | 0               | 0           | 1              | 20.00%          |
| class recall    | 7.14%         | 35.14%        | 36.11%         | 0.00%          | 0.00%           | 0.00%          | 0.00%            | 19.23%          | 0.00%           | 0.00%       | 25.00%         |                 |