National College of Ireland

# Configuration Manual

MSc Research Project
AI For Business

Ebubekir Ayhan
Student ID: x23141727

School of Computing
National College of Ireland

Supervisor: Victor Del Rosal

**National College of Ireland**

**MSc Project Submission Sheet**

**School of Computing**

**Student Name:** …Ebubekir Ayhan……

**Student ID:** …x23141727……..……

**Programme:** …AI For Business…… **Year:** …2024……..

**Module:** …MSc Research Project……..……

**Lecturer:** … Victor Del Rosal……..……

**Submission Due Date:** …16.09.2024……..……

**Project Title:** … Configuration Manual ……..……

**Word Count:** …1193…… **Page Count:** ……10……..……..……..……

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

**Signature:** ……Ebubekir Ayhan…..……

**Date:** ……16.09.2024……

**PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST**

| | |
|---|---|
| Attach a completed copy of this sheet to each project (including multiple copies) | □ |
| **Attach a Moodle submission receipt of the online project submission,** to each project (including multiple copies). | □ |
| **You must ensure that you retain a HARD COPY of the project**, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer. | □ |

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

| **Office Use Only** | |
|---|---|
| Signature: | |
| Date: | |
| Penalty Applied (if applicable): | |

# Configuration Manual

Ebubekir Ayhan
Student ID: x23141727

# 1 Introduction

Aim of this configuration manual document is to show the steps of the machine learning model that is built by Rapidminer.

# 2 System Requirements

To ensure optimal performance and stability of RapidMiner Server with an attached Job Agent, please make sure your system meets the following minimum requirements:

**Minimum System Requirements**

**Processor:**

- Dual-core 2GHz processor

**Memory:**

- 8GB RAM

**Disk Space:**

- More than 10GB of free disk space
    - Note: The filesystem needs to support UTF-8 encoding

**Operating System:**

- 64-bit recommended
- Windows Server 2008 R2, 2012, 2012 R2, 2016, 2019
- Linux

**Web Browser:**

- Internet Explorer 10 and above, or any modern web browser
    - Recommended browsers include Google Chrome, Mozilla Firefox, Microsoft Edge, and Safari

**Browser Resolution:**

- Minimum resolution of 1024x768 pixels

These requirements are essential to ensure that RapidMiner Server operates efficiently and reliably. Ensuring that your system meets these specifications will help you achieve the best performance and user experience.

# 3 Implementation

After installing rapidminer on your computer, the steps of machine learning model are shown below.

## 3.1. Data Collection

The dataset has been taken from the Kaggle. The dataset can be downloaded from the link https://www.kaggle.com/datasets/shilongzhuang/telecom-customer-churn-by-maven-analytics.

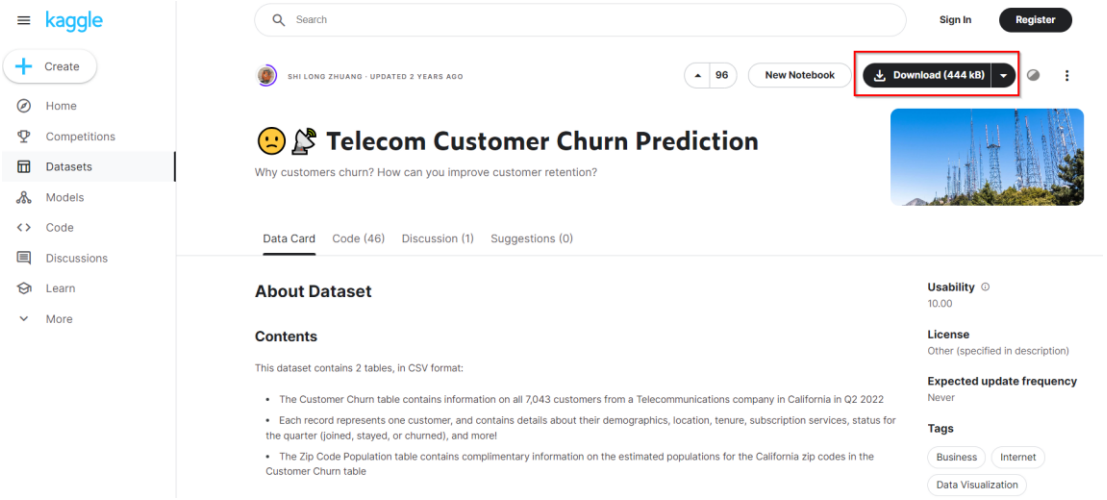


**Figure 1: Data Collection From Kaggle**

After this download, data can be uploaded to Rapidminer with “Import Data” process in Rapidminer. Then, the dataset should dragged into the process.

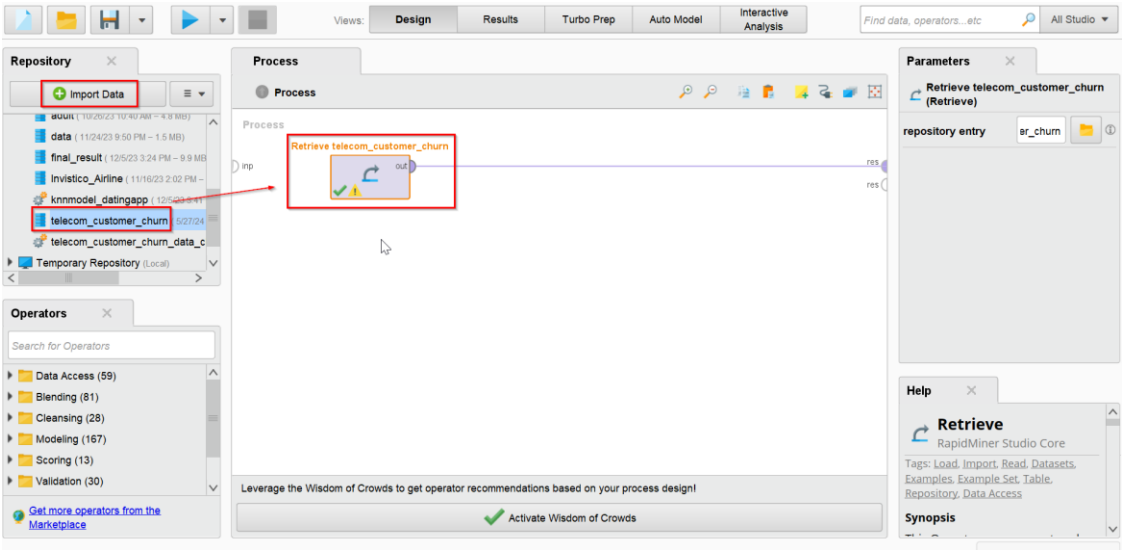


**Figure 2: Uploading Dataset to Rapidminer**

## 3.2 Data Cleaning

After uploading the data, remove the attributes that is not going to be used in the model which they do not affect the machine learning model. To do this, add “Select Attributes” process in the flow and choose “exclude attributes” and select which attributes to exclude.

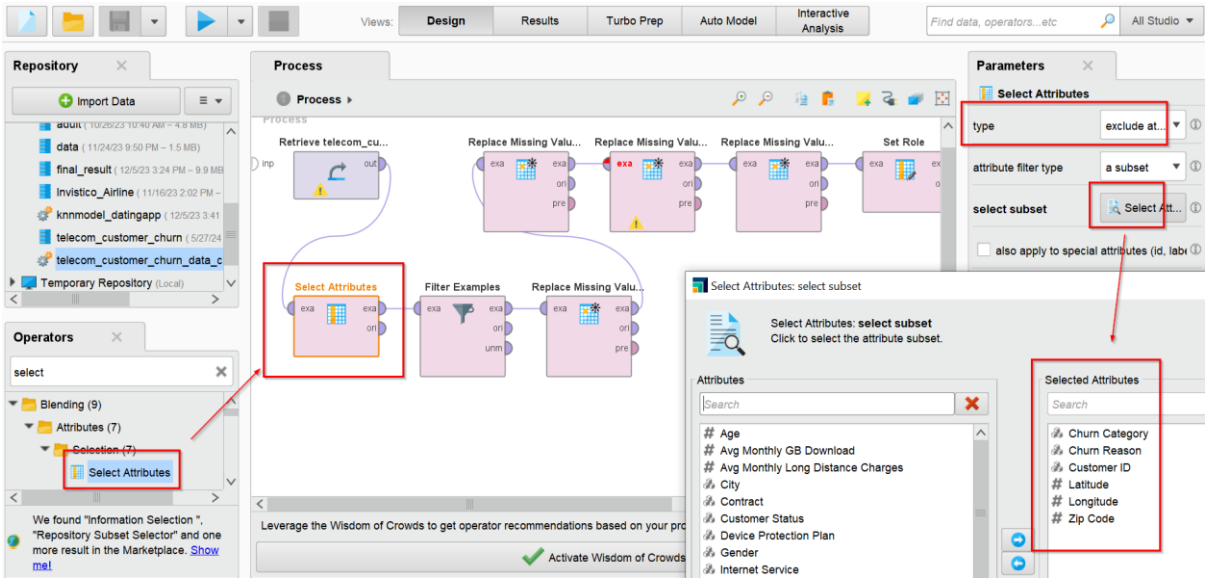


**Figure 3: Data Cleaning - Remove Attributes**

Customer status has 3 values; joined, stayed and churned. Since “joined” customers joined recently, they must not be used in the model. Therefore, we need to filter the dataset which only contains stayed and churned. To do this, add filter examples on Rapidminer and drag it into the flow. After dragging, choose customer status does not equal to “joined”.

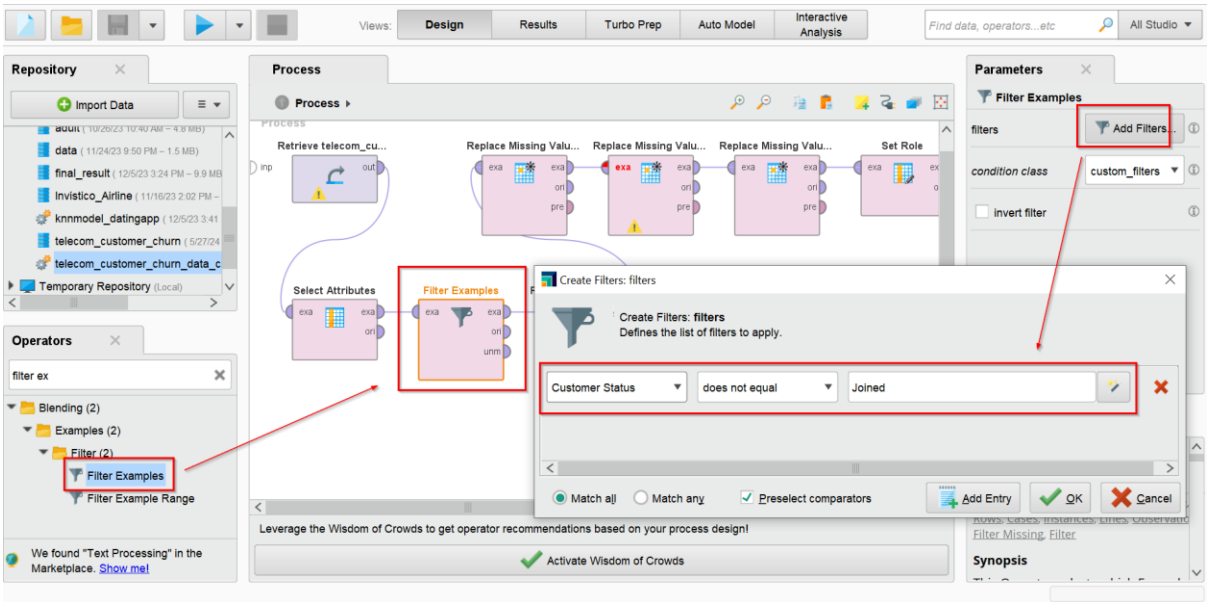


**Figure 4: Data Cleaning - Filtering Dataset**

After completing those, the missing values needs to be filled. “Avg Monthly GB Download” and “Avg Motnhly Long Distance Charges” had missing values and they are numerical values. Therefore, missing values are filled with the average of the dataset. To do this, drag “replace missing values" to the flow and select attributes to be replaced and select “average”.

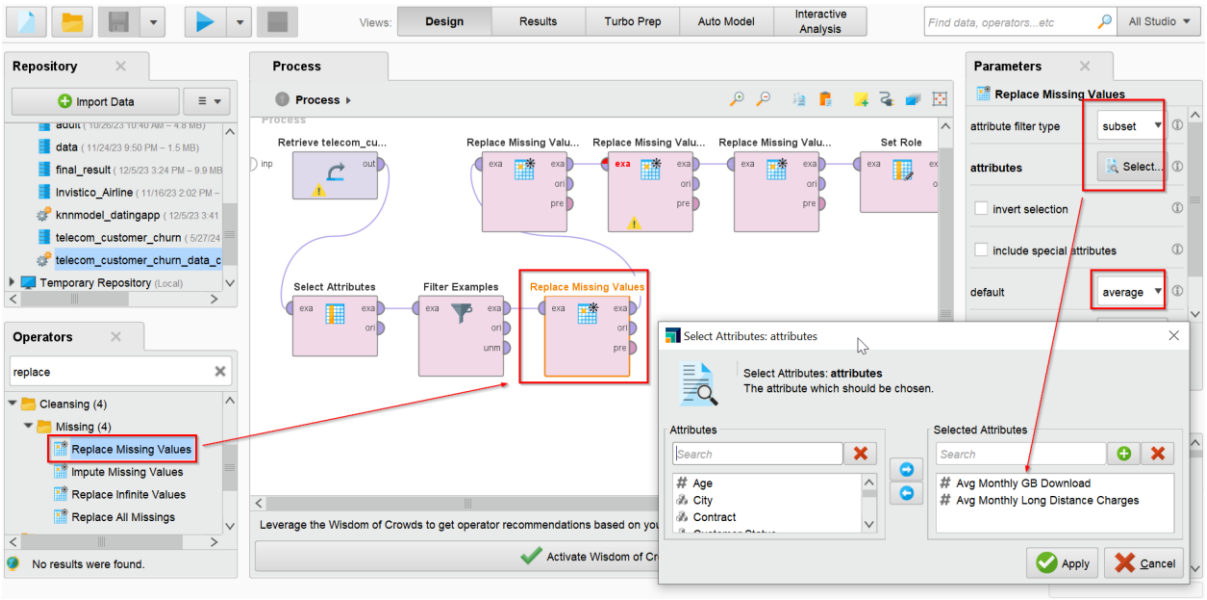


**Figure 5: Data Cleaning - Filling Missing Values**

Another type of the missing values are categorical attributes (yes/no). For the missing values of those attributes, “no” can be replaced which is probably the most correct answer. To do this, find the replace missing values process again and drag into flow. Then select the attributes that you want to fill missing values and write “No” as a replenishment value.

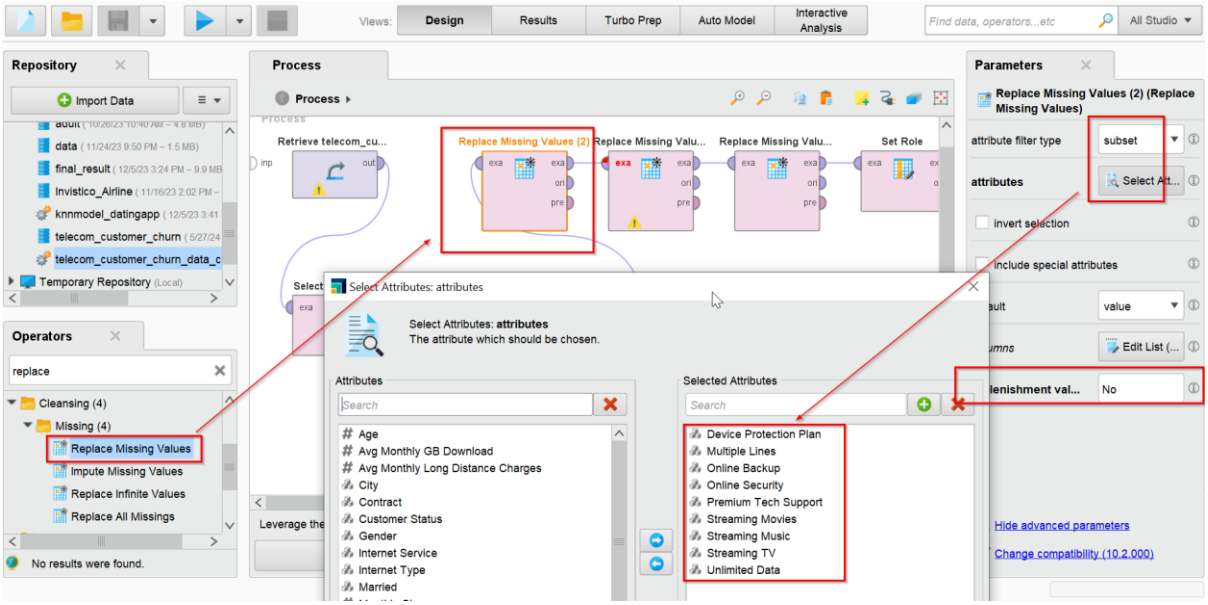


**Figure 6: Data Cleaning - Filling Missing Values (2)**

Last attribute that has missing values is “Internet Type” which includes cable, fiber optic, and DSL. This missing values can be filled as “none”. To do this in rapidminer, select “Replace Missing Values” process again and drag it into the flow. After that select “Internet Type” and wrote “None” as replenishment value.

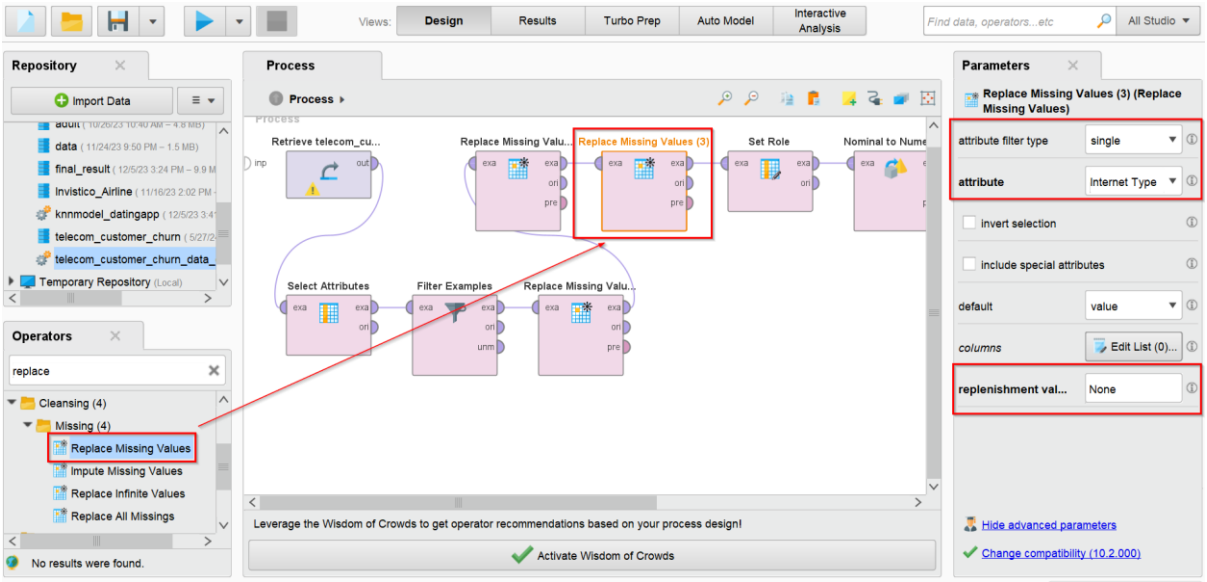


**Figure 7: Data Cleaning - Filling Missing Values (3)**

## 3.3 Data Transformation

Before applying machine learning model, all the variables except “Customer Status” need to be converted to numerical. Moreover, “customer status” needs to be selected as set so that rapidminer will understand customer status needed to be predicted. To do this, select “Set Role” process in rapidminer and select customer status as target label.

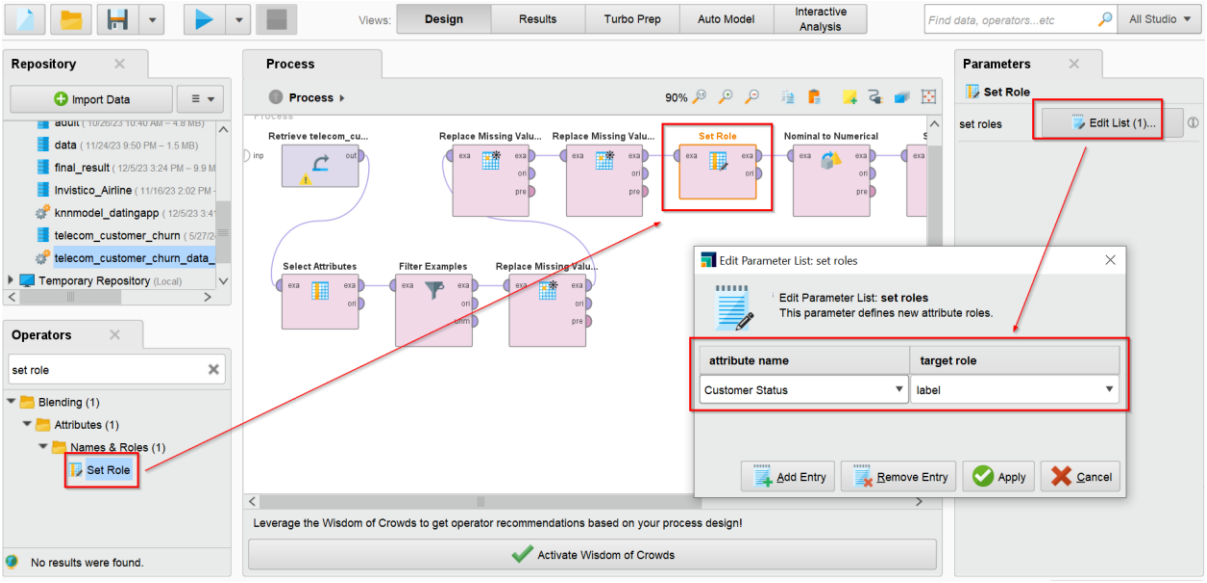


**Figure 8: Data Transformation - Set Role**

After that, select “nominal to numerical” to convert all data into numerical data. Select only customer status attribute and invert the selection. Model will convert all data into numerical except customer status.

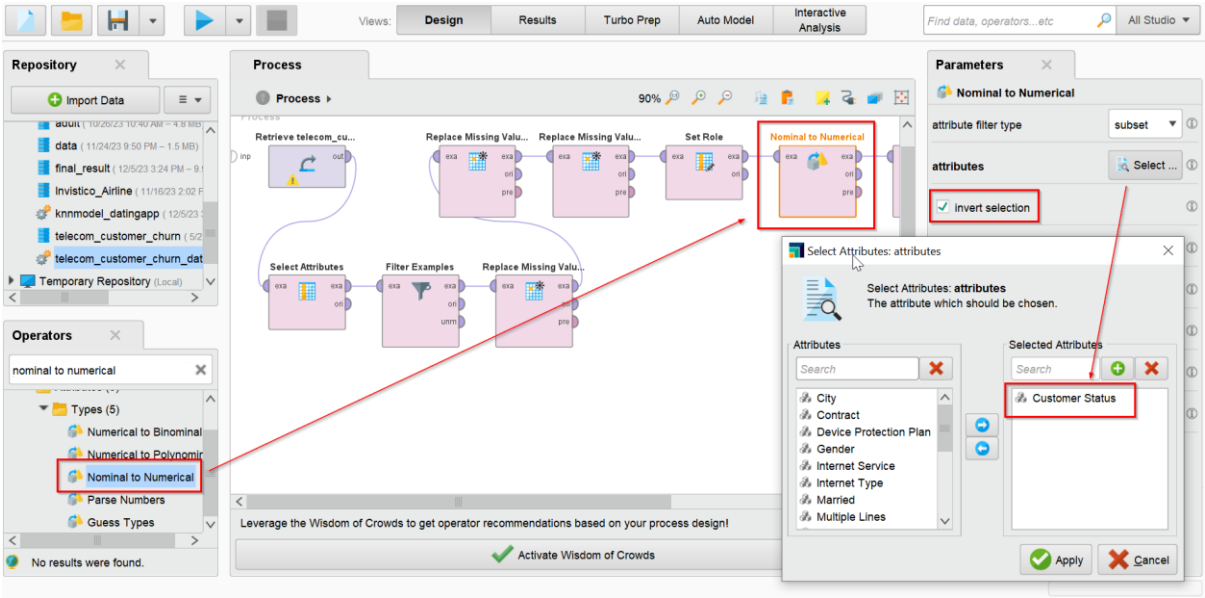


**Figure 9: Data Transformation - Nominal to Numerical**

After transforming all data into numerical, data needs to be splitted into two; one for training, and one for testing. %70 of the data is going to be splitted for training and %30 of the data for testing. To do this, select "Split Data" in Rapidminer and drag into the flow and select the splitting ratios.

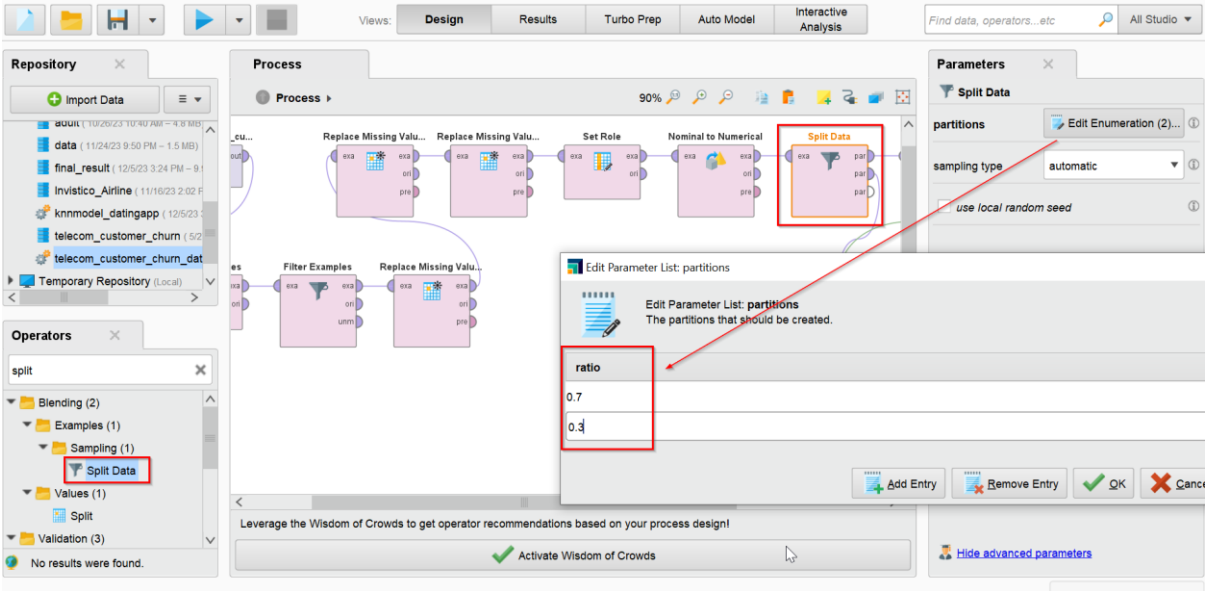


**Figure 10: Data Transformation - Splitting Data**

## 3.4 Data Modelling

After finishing all the preprocessing, machine learning model can be implemented. After splitting the data, one of the parts (%70) needs to be connected to the model which is "Neural Network". The other part (%30) and neural network process need to be connected apply model for testing. After all, "Performance" operator should be added to see the accuracy and the other performance parameters. After modelling, the model is ready to run.

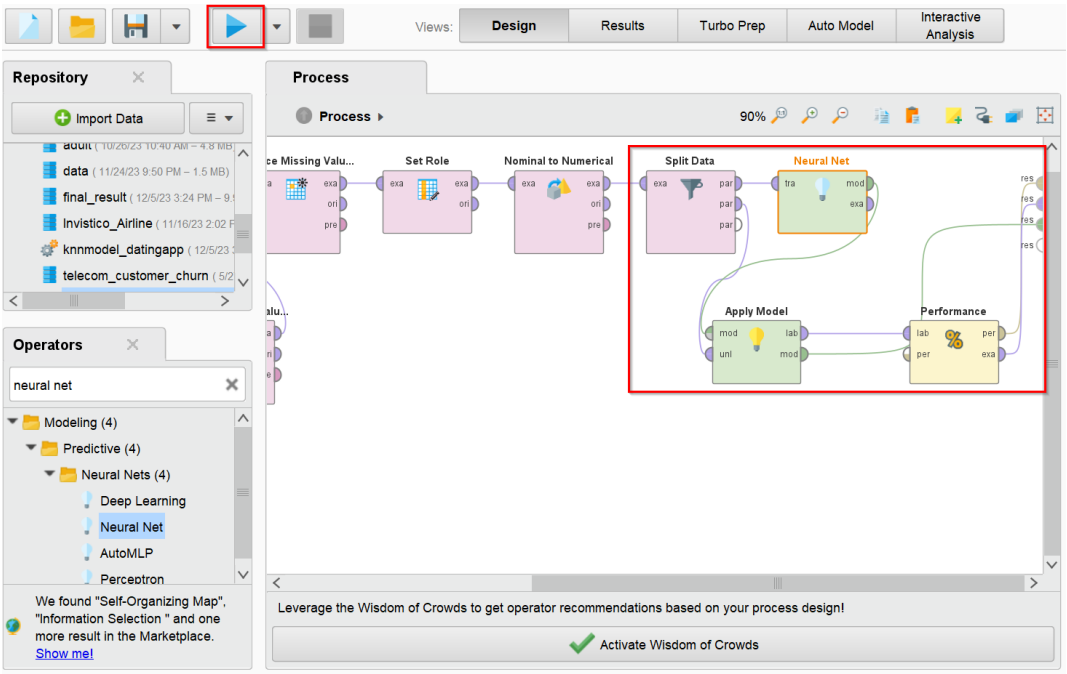


**Figure 11: Data Modelling for Neural Network**

To apply other machine learning algorithms, only “neural network” process needs to be changed with the new one and model needs to be runned to get results.

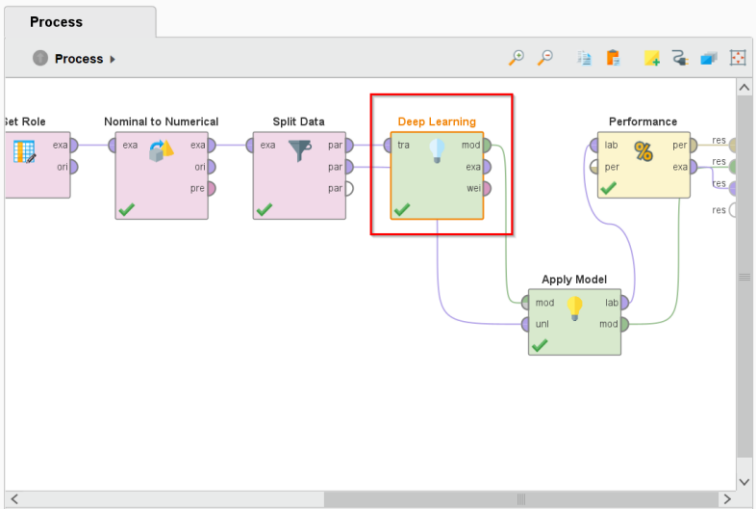


**Figure 12 : Data Modelling for Deep Learning**

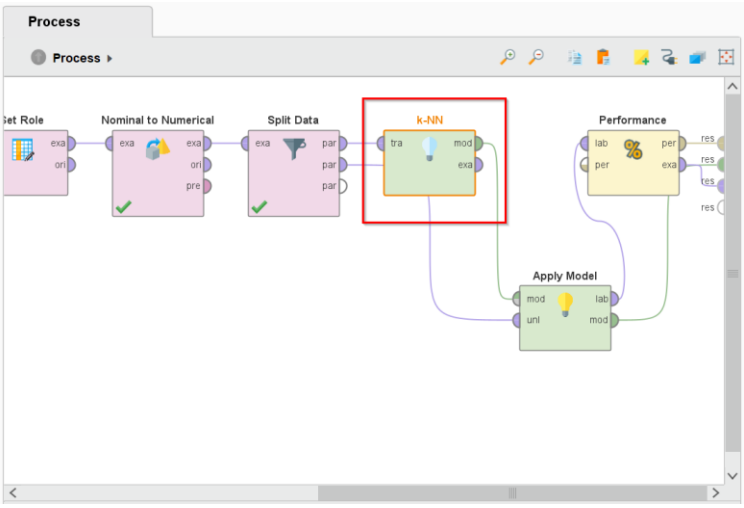


**Figure 13: Data Modelling for k-NN**

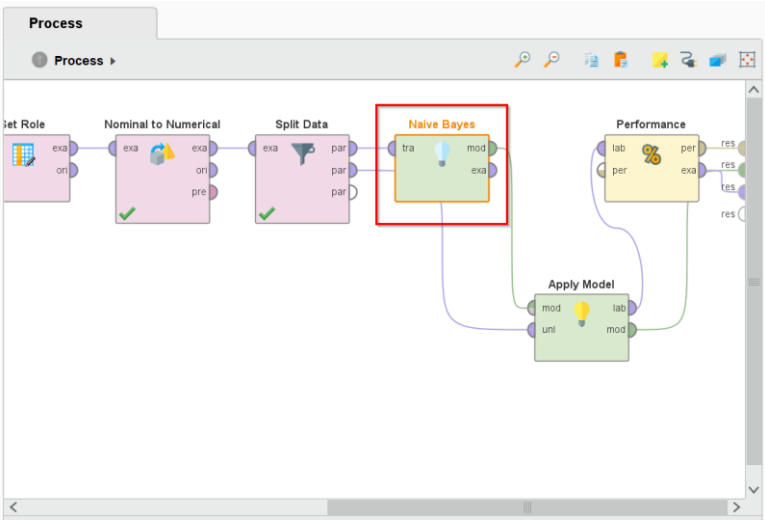


**Figure 14: Data Modelling for Naïve Bayes**

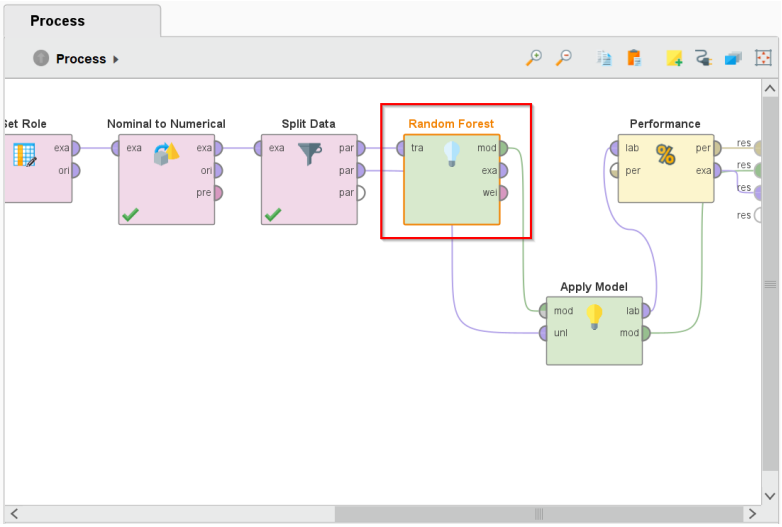


**Figure 15: Data Modelling for Random Forest**

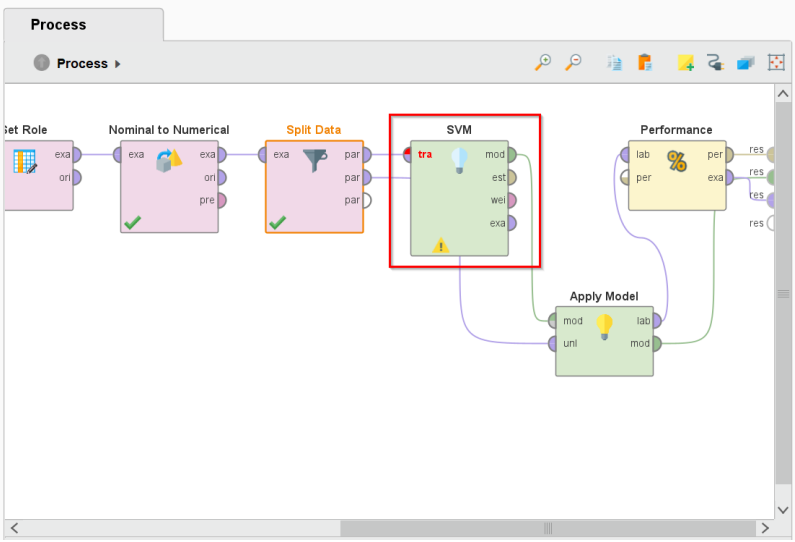


**Figure 16: Data Modelling for Support Vector Machine (SVM)**

# 4 Evaluation

After running the model, performance tab of Rapidminer shows the results of your model. Since "joined" customer status removed from the model, it shows 0 for joined. There is a total %80 accuracy in the model.

**accuracy: 80.02%**

| | true Stayed | true Churned | class precision |
|---|---|---|---|
| pred. Stayed | 1411 | 390 | 78.35% |
| pred. Churned | 5 | 171 | 97.16% |
| class recall | 99.65% | 30.48% | |

**Figure 17: Neural Network Results**

**accuracy: 77.04%**

| | true Stayed | true Churned | class precision |
|---|---|---|---|
| pred. Stayed | 1153 | 191 | 85.79% |
| pred. Churned | 263 | 370 | 58.45% |
| class recall | 81.43% | 65.95% | |

**Figure 18: Deep Learning Results**

**accuracy: 78.81%**

| | true Stayed | true Churned | class precision |
|---|---|---|---|
| pred. Stayed | 1294 | 297 | 81.33% |
| pred. Churned | 122 | 264 | 68.39% |
| class recall | 91.38% | 47.06% | |

**Figure 19: k-NN Results**

**accuracy: 44.31%**

| | true Stayed | true Churned | class precision |
|---|---|---|---|
| pred. Stayed | 468 | 153 | 75.36% |
| pred. Churned | 948 | 408 | 30.09% |
| class recall | 33.05% | 72.73% | |

**Figure 20: Naïve Bayes Results**

**accuracy: 71.83%**

| | true Stayed | true Churned | class precision |
|---|---|---|---|
| pred. Stayed | 1416 | 557 | 71.77% |
| pred. Churned | 0 | 4 | 100.00% |
| class recall | 100.00% | 0.71% | |

**Figure 21: Random Forest Results**

**accuracy: 78.76%**

| | true Stayed | true Churned | class precision |
|---|---|---|---|
| pred. Stayed | 1117 | 121 | 90.23% |
| pred. Churned | 299 | 440 | 59.54% |
| class recall | 78.88% | 78.43% | |

**Figure 22: SVM Results**