Chapter 4
Data Pre-Processing and Example of Data Classification With RapidMiner

ABSTRACT

In this book, the focus is on data mining with RapidMiner. However, it's important to note that there are other essential steps to consider when delving into the realm of data mining. This chapter serves as an introduction to the process of data pre-processing using RapidMiner, allowing readers to practice with a data set example available on the platform. With RapidMiner, data pre-processing begins with exploring the data visually and then selecting the features that will be analysed with each data mining technique. Managing missing values in a feature is also a crucial step in this process, which can be achieved by either eliminating or replacing them with appropriate values. In addition, RapidMiner allows data scientists to detect outliers and normalize features easily using diagram design, without requiring any computer programming skills. To help readers become familiar with the tools offered by RapidMiner, a classification technique will be demonstrated step-by-step in the book.

INTRODUCTION

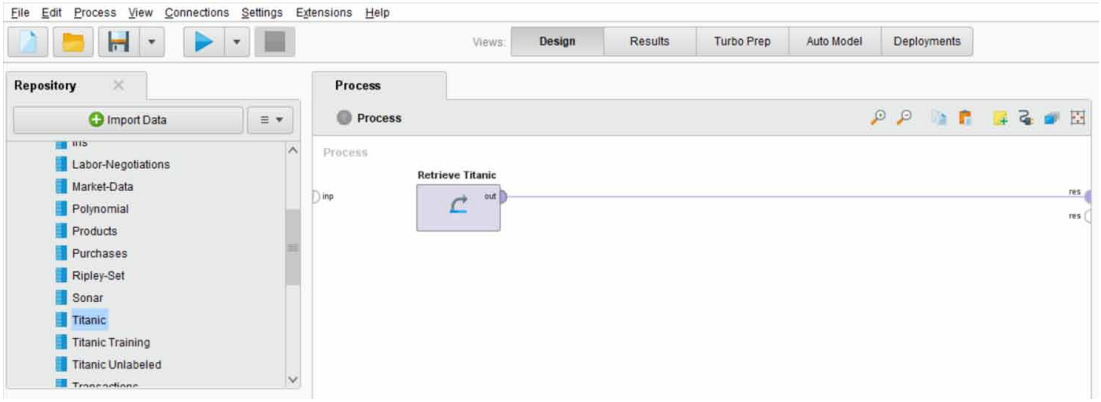
RapidMiner is a tool supporting throughout the process of data science work (Mat, Lajis & Nasir, 2018). Data scientists can import data sets into RapidMiner software in order to prepare the data ready for processing with various data mining techniques (Dai et al., 2016: Samsani, 2016: Phan, Wu & Phan, 2021). In Missing Value Management and Outlier Management, when the data is available, data scientists can use various data mining models, either Supervised Learning or Unsupervised Learning (Cai et al., 2016: Mandhare & Idate, 2017: Susanti & Azizah, 2017: Abu-Soud, 2019). Processing algorithms are designed through model connections, and then the result of data processing is executed. To connect such models, the need for computer programming is unnecessary. Therefore, the scientists can quickly modify models at each step and perform reprocessing to improve the model's processing accuracy.

DOI: 10.4018/978-1-6684-4730-7.ch004

DATA PRE-PROCESSING USING RAPIDMINER SOFTWARE

To import datasets into a data mining model with RapidMiner software, data scientists can manipulate the data with Microsoft Excel. For example, Attribute Transformation is done to change the shape of the data before it is imported into RapidMiner software. They can also choose the operator within Rapid-Miner as well. An experiment can be performed to prepare the data as follows:

Figure 1. Using the Titanic sample dataset



Data scientists can choose to use the Titanic sample dataset for experimentation. After the dataset is imported into the process, data scientists connect the dataset's output to the Result section, and execute the model. The result is as follows.

Figure 2. Titanic dataset details

Row No.	Passenger ...	Name	Sex	Age	No of Sibling...	No of Parent...	Ticket Numb...	Passenger F...	Cabin	Port of Emb
1	First	Allen, Miss. E...	Female	29	0	0	24160	211.338	B5	Southampt
2	First	Allison, Mast...	Male	0.917	1	2	113781	151.550	C22 C26	Southampt
3	First	Allison, Miss. ...	Female	2	1	2	113781	151.550	C22 C26	Southampt
4	First	Allison, Mr. H...	Male	30	1	2	113781	151.550	C22 C26	Southampt
5	First	Allison, Mrs. ...	Female	25	1	2	113781	151.550	C22 C26	Southampt
6	First	Anderson, Mr....	Male	48	0	0	19952	26.550	E12	Southampt
7	First	Andrews, Mis...	Female	63	1	0	13502	77.958	D7	Southampt
8	First	Andrews, Mr. ...	Male	39	0	0	112050	0	A36	Southampt
9	First	Appleton, Mrs...	Female	53	2	0	11769	51.479	C101	Southampt
10	First	Artagaveytia, ...	Male	71	0	0	PC 17609	49.504	?	Cherbourg
11	First	Astor, Col. Jo...	Male	47	1	0	PC 17757	227.525	C62 C64	Cherbourg
12	First	Astor, Mrs. Jo...	Female	18	1	0	PC 17757	227.525	C62 C64	Cherbourg
13	First	Aubart, Mme. ...	Female	24	0	0	PC 17477	69.300	B35	Cherbourg
14	First	Barber, Miss. ...	Female	26	0	0	19877	78.850	?	Southampt
15	First	Barkworth, Mr...	Male	80	0	0	27042	30	A23	Southampt
16	First	Baumann, Mr...	Male	?	0	0	PC 17318	25.925	?	Southampt
17	First	Baxter, Mr. Qu...	Male	24	0	1	PC 17558	247.521	B58 B60	Cherbourg
18	First	Baxter, Mrs. J...	Female	50	0	1	PC 17558	247.521	B58 B60	Cherbourg