

# Latihan Tugas Materi 35 KNIME I

KNIME Analytics Platform

Home01. Importing Data - ExerciseHelpPreferencesMenu

Execute allCancel allReset all100%

Last update: Apr 26, 2024

Importing Data - Exercise

Exercise 1 for KNIME User Training

- Import data from three different file formats (table, csv, xls)
- Import data from SQLite database

External resources

- Data Access
- File Reader
- Table Reader
- Excel Reader
- The knime:// Protocol
- Slides (KNIME Analytics Platform Course)

Tags

ETLdata accessimportcsvtable+5

Activity: Importing Data

Import data from:

- Sentiment Analysis.table
- Sentiment Rating.csv
- Product Data2.xls

CSV ReaderExcel Reader

To show the node output, please select a configured or executed node.

KNIME Analytics Platform

Home01. Importing Data - Exercise02. Data\_Cleaning\_Exercise XHelpPreferencesMenu

ExecuteCancelResetCreate metanodeCreate component100%

Duplicate Row Filter

This node identifies duplicate rows. Duplicate rows have identical values in certain columns. The node chooses a single row for each set of duplicates ("chosen"). You can either remove all duplicate rows from the input table and keep only unique and chosen rows or mark the rows with additional information about their duplication status.

PortsOptionsViews

Input ports

Type:Input Data

The data table containing potential duplicates.

Output ports

Type:Filtered/Labeled Data

Either the input data without duplicates or the input data with additional columns identifying duplicates.

Exercise 3: Data Cleaning

In this exercise you will handle missing values.

1) Handle missing values

- 1.1 If the age of a customer is missing, replace the birthday with a missing value (Rule Engine node)
- Hint: Use the expression NOT MISSING \$Age\$>> \$Birthday\$
- 1.2 Impute the missing values in the age column with the column mean

2) Remove duplicates in the customer data

CSV ReaderRule EngineMissing ValueDuplicate Row Filter

Customer Information System 1Missing(Age)=meanRemove Duplicate Customers

1: Filtered/Labeled DataFlow Variables

Rows: 2398 | Columns: 9

TableStatistics

#	RowID	City	Country	CustomerID	FirstName	LastName	Birthday	Age	Email	Newsletter
	String		String	String	String	String	String	Number (double)	String	Number (integer)
1	Row0	Glasgow	United Kingdom	17-171-832-104	Alois	Berger	23.9.1972	47	Alois.Berger@...	0
2	Row1	Szczecin	Poland	37-370-580-177	Michaela	Schultz	9.6.1998	21	Michaela.Schu...	0
3	Row2	Sheffield	United Kingdom	27-270-743-182	Rotraut	GrÄunwald	20.4.1975	44	Rotraut.GrÄun...	0
4	Row3	Bochum-Hordel	Germany	64-647-953-993	Helga	Heindl	18.10.2000	19	Helga.Heindl@...	0
5	Row4	Dortmund	Germany	84-846-821-690	Mira	Gleich	18.3.1997	22	Mira.Gleich@...	0
6	Row5	Valencia	Spain	58-582-952-948	Joanna	Radke	13.12.1995	24	Joanna.Radke...	1