

数据挖掘-去除重复值

概述

去除重复值是用于删除数据集中的重复行（假如有两行相同，保留其中一行）。



- [概述](#)
- [输入/输出](#)
- [参数设置](#)
- [示例](#)

输入/输出

输入	一个输入端口，用于接收数据集。
输出	一个输出端口，用于输出去除重复值的结果。

参数设置

设置去除重复值的参数：



设置说明如下：

参数	说明
选择列	用于选择进行去除重复值设置的列： 选择列 源数据列表0/0 请输入搜索内容 无数据 已选字段列表0/6 请输入搜索内容 ☐ #id ☐ #Sepal_Length ☐ #Sepal_Width ☐ #Petal_Length ☐ #Petal_Width ☐ A Species < 到左边 到右边 > 确定 取消

示例

1、原先关系数据源的输出结果有9条数据，对其进行去除重复值，在选择列弹框中选择所有的列。



2、输出结果有7条数据，如图：

当前显示 9 条 / 总共有 9 条数据 提示:点击单元格可查看超出的内容 关系数据源共有9条数据，有重复行

Species	A	B	C	D
versicolor	6.6	3	null	1.4
versicolor	6.8	2.8	4.8	1.4
versicolor	6.8	2.8	4.8	1.4
versicolor	6.6	3	4.4	1.4
versicolor	6.6	3	4.4	1.4
versicolor	5.5	2.4	null	1.1
versicolor	5.5	2.4	3.7	1
versicolor	5.8	2.7	null	1.2
versicolor	6	2.7	5.1	1.6

注意：表头中◆表示特征列，*表示标签列

去除重复值

表头真名 表头别名

当前显示 7 条 / 总共有 7 条数据 提示:点击单元格可查看超出的内容 输出结果有7条数据，无重复行

Species	A	B	C	D
versicolor	5.5	2.4	null	1.1
versicolor	6.8	2.8	4.8	1.4
versicolor	5.8	2.7	null	1.2
versicolor	5.5	2.4	3.7	1
versicolor	6	2.7	5.1	1.6
versicolor	6.6	3	4.4	1.4
versicolor	6.6	3	null	1.4

注意：表头中◆表示特征列，*表示标签列

表头真名 表头别名