

数据挖掘-自定义离散

概述

将连续的属性进行离散化操作，方便数据挖掘处理。用户可自定义规则进行离散化操作。

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

输入/输出

输入	一个输入端口，用于接收前置节点传下来的数据集。
输出	一个输出端口，用于输出增加了离散后的字段的数据集。

参数设置

参数名称	说明	备注
自定义规则	区间：以英文逗号分隔离散区间，负无穷为-INF，正无穷为INF。	必填。
	区间为空时，默认该字段不做离散处理。	区间填写示例：-INF, 30, 60, 90, INF
新增列后缀	离散后会生成新的字段，默认在原有字段名后追加Buckrizer后缀。该后缀支持修改。	后缀默认值为Buckerizer。

示例

使用“银行零售客户流失数据”，进行自定义离散化，自定义规则将年龄字段离散区间设置为-INF, 20, 30, 40, INF，其他字段区间为空，输出结果如下图：

1年内购买理财	下载手机银行	是否领取App权益	是否登录App	是否持有信用卡	是否关联还款	年龄Buckerizer
0	0	0	0	1	1	3.0
1	0	0	0	0	0	3.0
1	0	0	0	0	0	3.0
0	0	0	0	0	0	0.0
0	1	1	14	1	1	3.0
1	1	0	3	1	1	3.0
1	0	0	0	1	1	3.0
0	0	0	0	1	0	0.0
1	1	1	20	0	0	3.0
1	1	0	3	1	1	3.0
1	0	0	0	0	0	1.0