目录

[模型训练,预测过程使用说明 1](#_Toc424309771)

[数据输入 1](#_Toc424309772)

[模型训练 2](#_Toc424309773)

[模型存储 2](#_Toc424309774)

[预测数据输入 2](#_Toc424309775)

[预测 3](#_Toc424309776)

[结果可视化 3](#_Toc424309777)

[结果发布 3](#_Toc424309778)

# 模型训练,预测过程使用说明

Data Studio是一种结构化的数据挖掘流程设计器，它模拟数据挖掘的整个过程，一般的数据挖掘流程可分为，

模型训练过程：数据输入，数据探查，数据预处理，模型训练，模型校验，模型存储

预测过程：预测数据输入，预测，结果可视化，预测评估，结果发布

DataStudio会不断加入新的节点，覆盖挖掘的整个流程。

先针对现有节点介绍挖掘的整个流程，部分过程可能涉及不到，随着节点的不断加入，后续会不断修正此文档。

## 数据输入

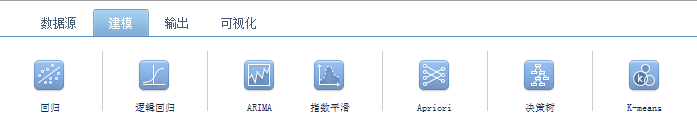
DataStudio现提供两种输入数据节点：文件，数据库读取。

前者可以读取数据文件，比如CSV，后者可以读取数据库中的数据。



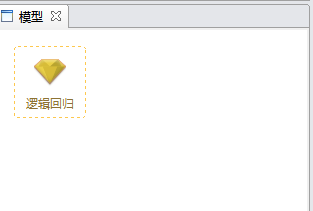
## 模型训练

建模部分提供多种算法节点，包括回归，分类，聚类，时间序列算法节点。



## 模型存储

生成的模型可以直接保存至服务器端，如下图所示，选中生成的模型，右键可将模型存储至服务器端。



## 预测数据输入

同数据输入，不过此时的数据为预测用数据。

## 预测

拖拽模型视图中的节点至工作区，引入预测数据可以执行预测过程。

## 结果可视化

提供多种图表，对数据结果集合的可视化，包含基础图表，也包含对挖掘模型可视化的针对性图表，比如平行坐标轴图表，是用来展现关联规则的。

图表会不断的丰富。



## 结果发布

预测结果数据可存储为本地文件，或者输出至数据库。

