# Smartbi V10.5.12-数据挖掘

注意：（新特性列表中：+表示新增；^表示增强）



具体改进点如下：

| 新增 | 增强 |
| --- | --- |
| +【挖掘】通过产品帮助功能创建节点使用示例 | ^【挖掘】全表统计节点支持输出结果 |
| +【挖掘】新增移动平均算法节点 | ^【挖掘】评估节点支持输出结果 |
| +【挖掘】新增指数平滑算法节点 | ^【挖掘】模型的预测中把预测概率输出成字段 |
| +【挖掘】增加正则表达式处理节点 | ^【挖掘】数据清理节点合并 |
| +【挖掘】节点能够自动连接 | ^【挖掘】过滤和行选择节点合并 |
| +【挖掘】增加在线节点开发功能 | ^【挖掘】派生列增强函数帮助提示 |
| +【挖掘】增加日期时间节点 | ^【挖掘】SQL脚本输入表字段显示优化 |
| +【挖掘】监控建模增加停止功能 | ^【挖掘】提供节点iframe扩展配置组件 |
| +【挖掘】自定义帮助指引系统 | |
| +【挖掘】支持对源和目标节点运行结果进行数据透视 | |
| +【挖掘】节点支持自动布局 | |

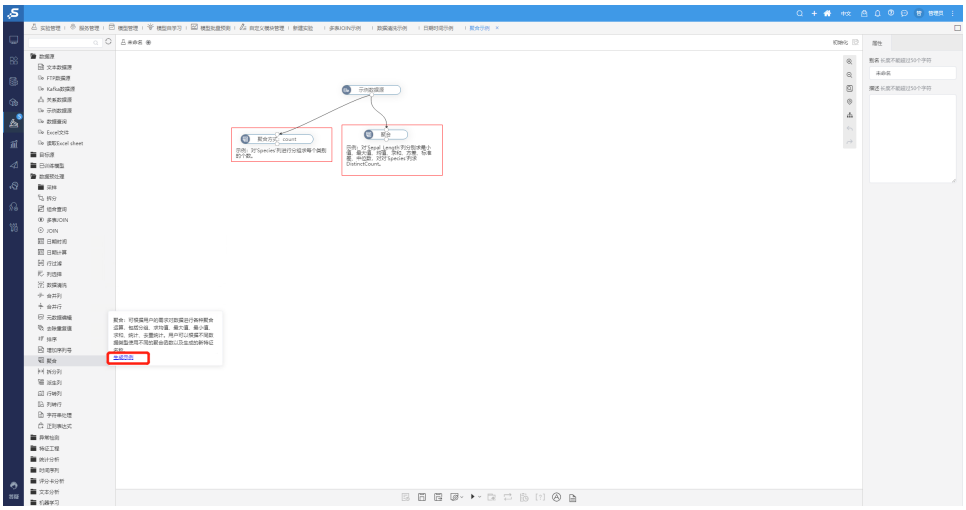
## +【挖掘】通过产品帮助功能创建节点使用示例

**背景介绍**

为了提高产品的易用性，帮助用户快速了解某一节点的功能。新版本在节点的帮助信息后面添加创建节点使用示例的入口，针对常用的数据预处理节点和统计分析节点支持一键生成示例，帮助用户快速了解节点的使用方法。

**功能简介**

用户在鼠标移动到某一个节点之后，出现悬浮提示后会出现“生成示例”链接“，点击之后会在新窗口中生成对应的示例ETL。

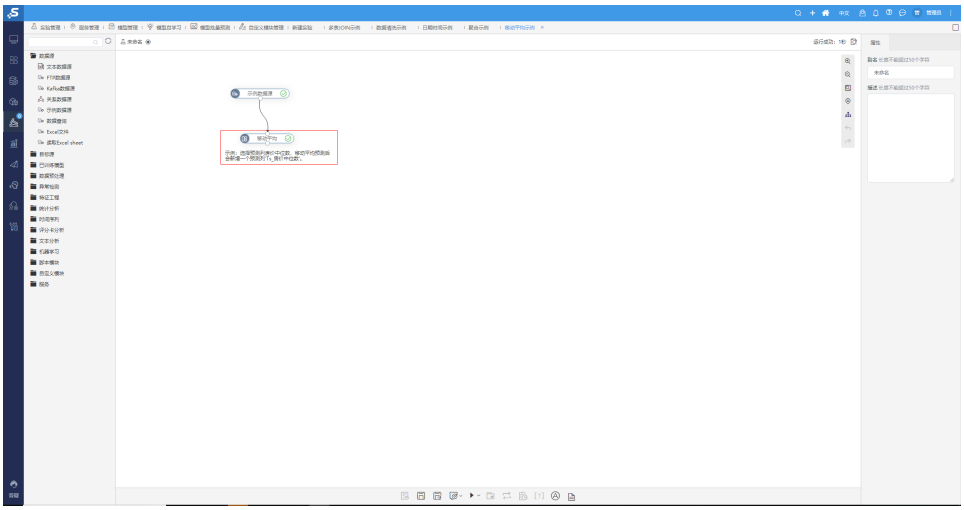


## +【挖掘】新增移动平均算法节点

**背景介绍**

移动平均法作为经典的基础时间序列算法，可以消除部分因素的影响，显示总体走势。

**功能简介**



**参考文档**

详情请参考 数据挖掘 - 移动平均。

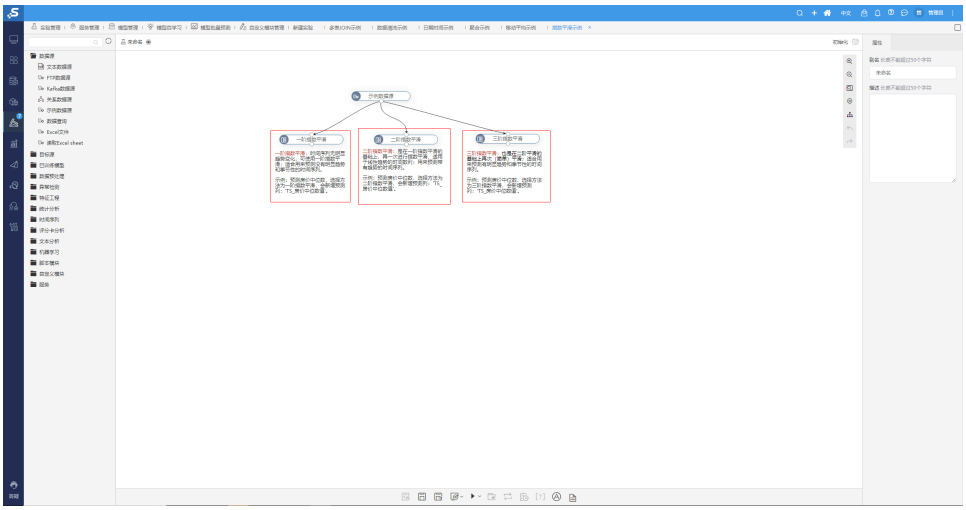
## +【挖掘】新增指数平滑算法节点

**背景介绍**

时间序列算法中常用的算法有指数平滑法，其中指数平滑常用的几种形式有（Brown）一次指数平滑、二次指数平滑、三次指数平滑、以及(Holt)双参数、（Holt-Winters）季节性指数平滑法。

**功能简介**

本节点实现的是布朗（Brown）的一次、二次、三次指数平滑。



**参考文档**

详情请参考　数据挖掘 - 指数平滑。

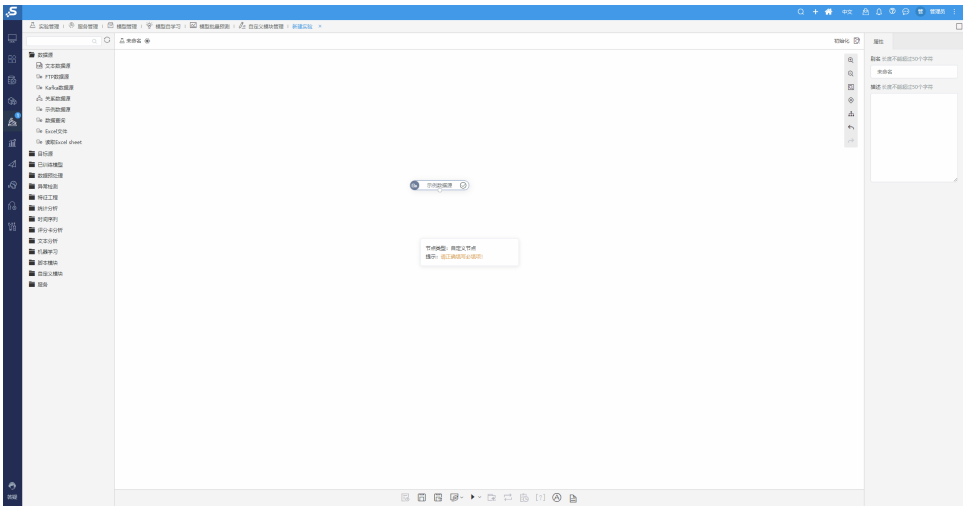
## +【挖掘】增加正则表达式处理节点

**背景介绍**

在处理文本过程中，经常会使用正则表达式去处理，目前ETL正则处理文本功能较弱，多个ETL项目中均有实施人员提到希望增强该功能。

**功能简介**

用正则表达式语法的强大模式匹配能力，对字符串数据进行解析、匹配或替换。



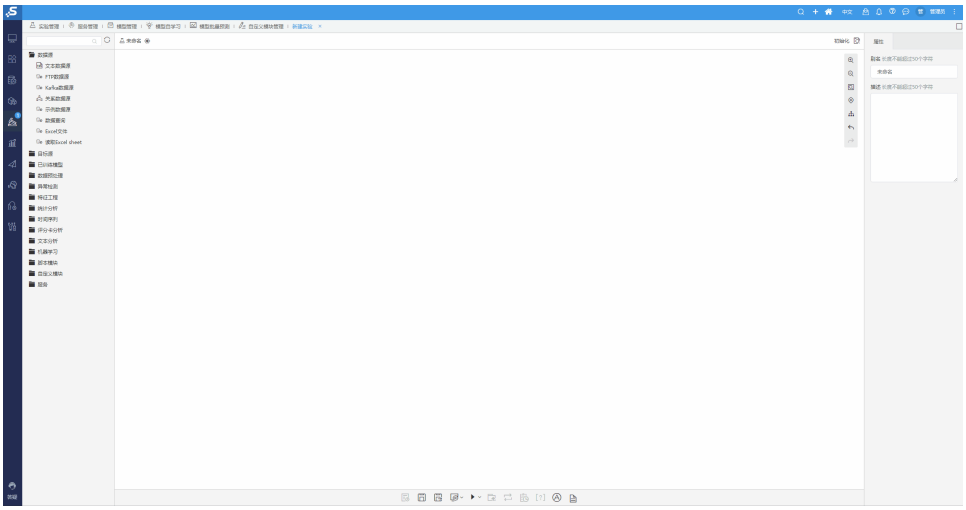
**参考文档**

详情请参考 数据挖掘-正则表达式。

## +【挖掘】节点能够自动连接

**功能简介**

新版本中添加了节点自动布局功能，能够让用户更加专注于建模。



当不需要自动连线功能时，可以在当前ETL中的如下入口禁用。或者可以配置系统选项：DISABLE_AUTO_CONNECTION=true

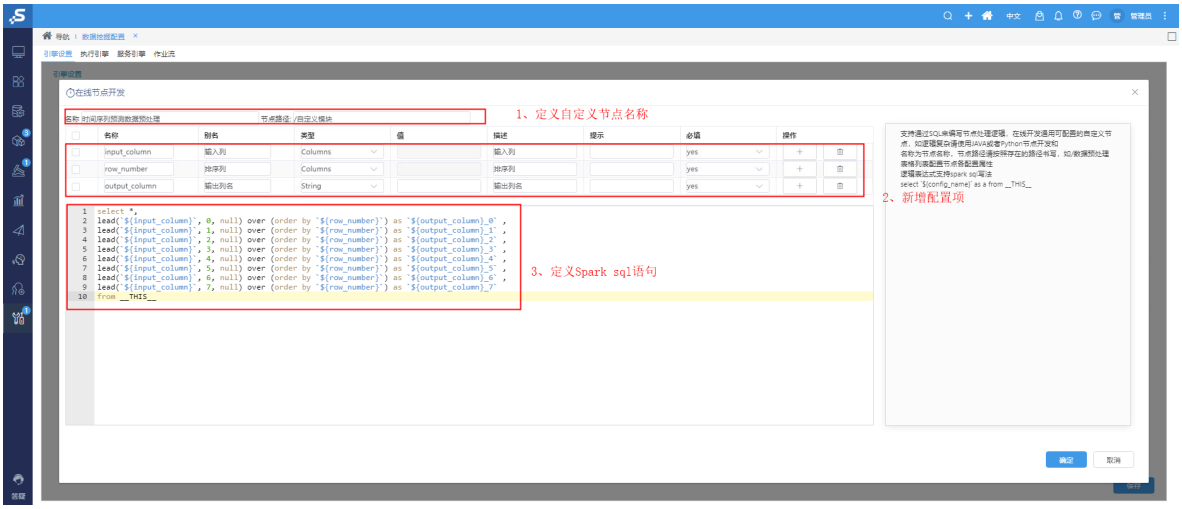

# +【挖掘】增加在线节点开发功能

**背景介绍**

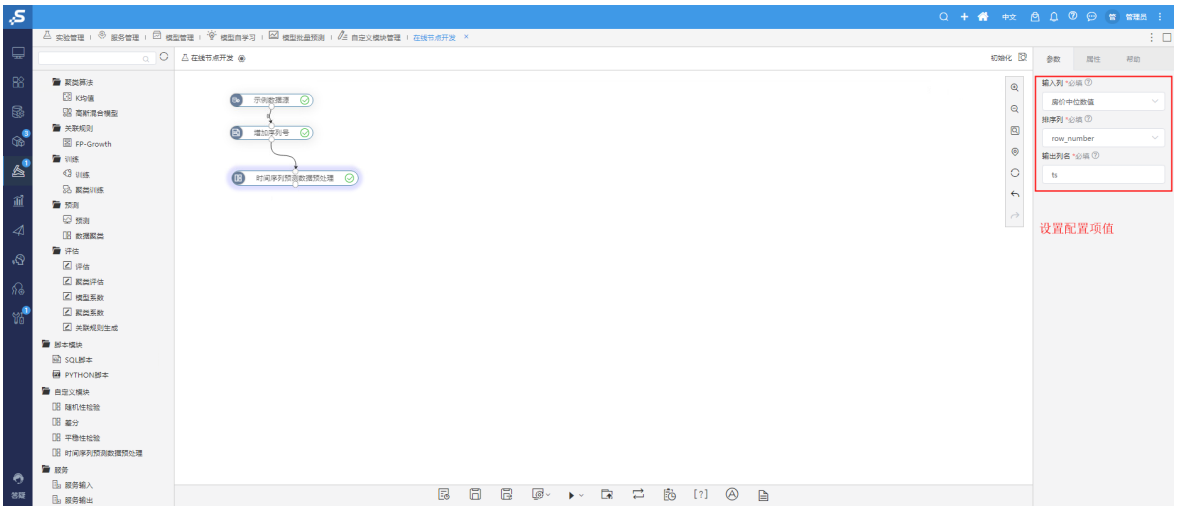
新版本中增加了在线节点开发功能，可以将部分可复用的SQL封装成预制节点进行复用，开发完成后和产品中预制节点使用方式一样。该功能入口在系统运维界面中，在/系统运维/数据挖掘配置/引擎配置/在线节点开发下，能够支持基础的配置选项，并且通过SQL来实现简单的数据处理逻辑。

**功能简介**

节点在线开发编辑界面如下：




使用在线自定义节点：



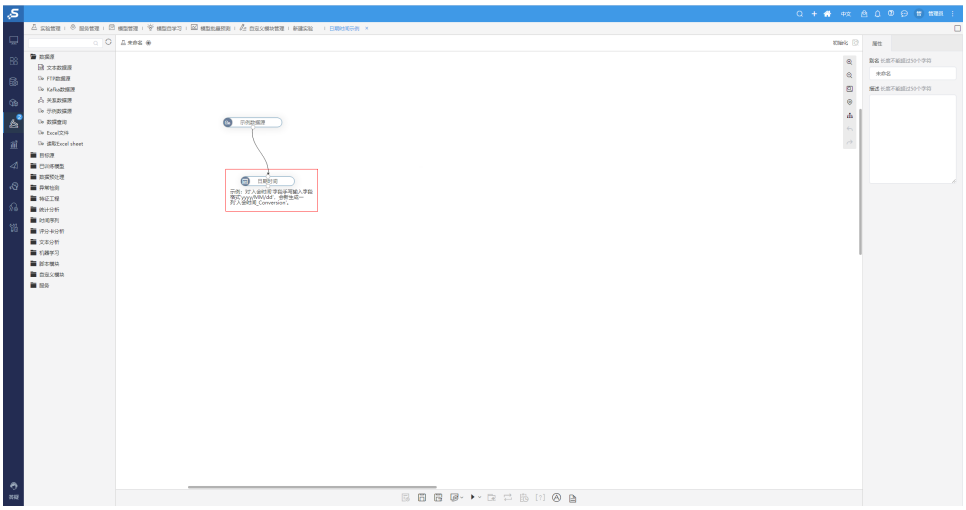

## +【挖掘】增加日期时间节点

**背景介绍**

用户在录入日期类型数据时，经常会遇到日期格式不统一的问题。

**功能简介**

新版本为了更方便用户统一日期时间格式，提供了日期类型字符转换功能。



详情请参考 数据挖掘-日期时间。

## +【挖掘】监控建模增加停止功能

**背景介绍**

异常情况下，引擎中运行的实验节点会存在无法停止运行的现象。因此，新版本在实验监控界面中提供了停止运行的功能。

**功能简介**

可以在系统监控下的实验监控下看到具体的引擎中运行的实验的记录，根据需要将其停止。


## +【挖掘】自定义帮助指引系统

**背景介绍**

产品的节点通常需要，帮助用户快速了解某一节点的功能。新版本在节点的帮助信息后面添加创建节点使用示例的入口，针对常用的数据预处理节点和统计分析节点支持一键生成示例，帮助用户快速了解节点的使用方法。

**功能简介**

用户在鼠标移动到某一个节点之后，出现悬浮提示后会出现对应的自定义帮助信息。

**功能简介**

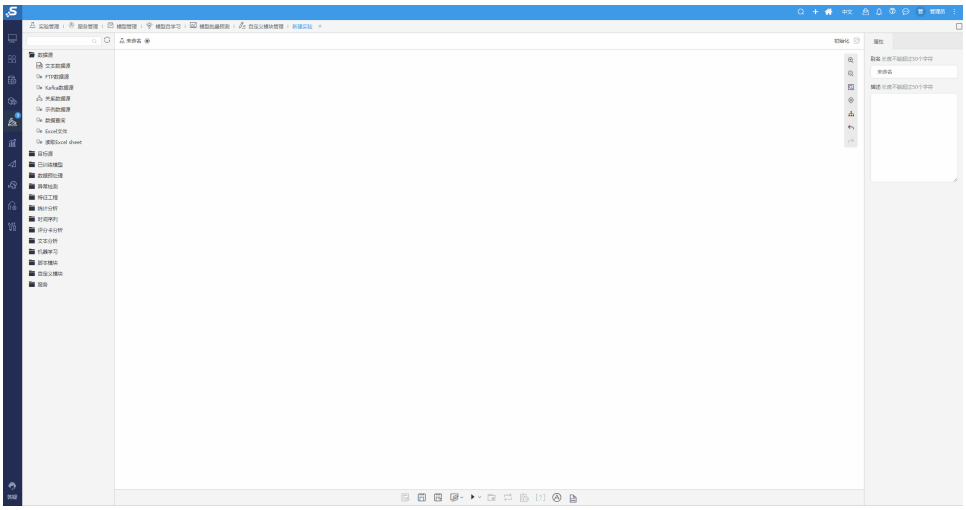

## +【挖掘】支持对源和目标节点运行结果进行数据透视

**背景介绍**

旧版本的数据挖掘实验的可视化实现方式不足，做数据探索需要导出数据到BI，或者用Python去做。所以增加数据透视功能，右键选择运行好的节点后能点击数据透视，进入一个透视分析界面，对该节点的运行结果进行透视分析。透视分析界面沿用BI分析展现模块的透视分析功能。

**功能简介**

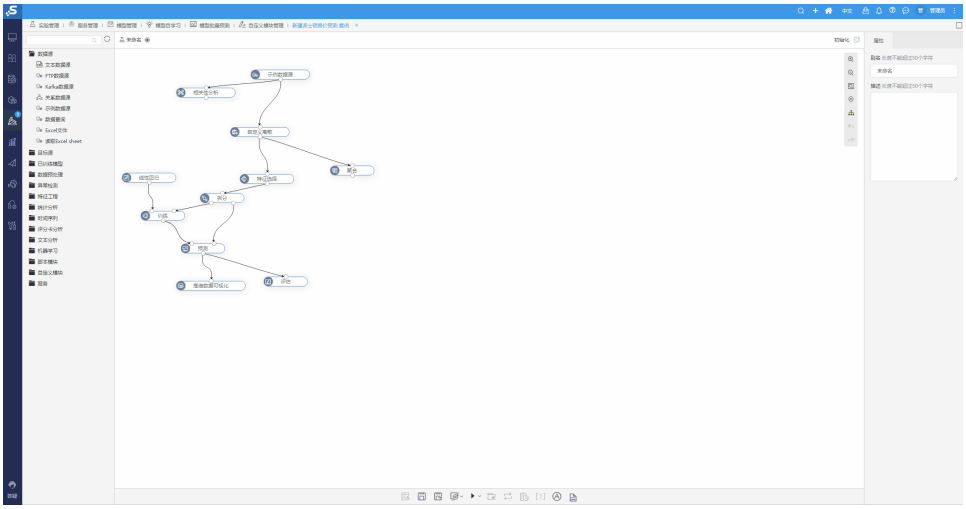
用户可以在关系数据源和关系目标源节点上右键，点击数据透视，打开一个透视分析界面，直接对数据源中的数据进行分析。



## +【挖掘】节点支持自动布局

**功能简介**

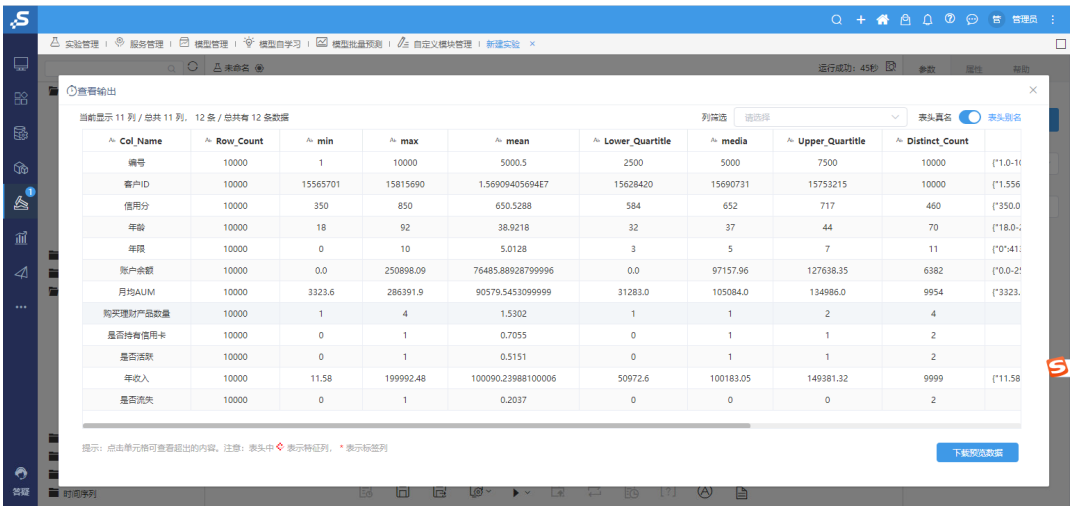
新版本中添加了节点自动布局功能，能够让用户更加专注于建模。



## ^【挖掘】全表统计节点支持输出结果

**功能简介**

旧版本的全表统计节点不支持输出统计结果，新版本中对该节点进行了改进，将各统计结果转换成Dataset输出，方便后续监控和图表展现分析等。

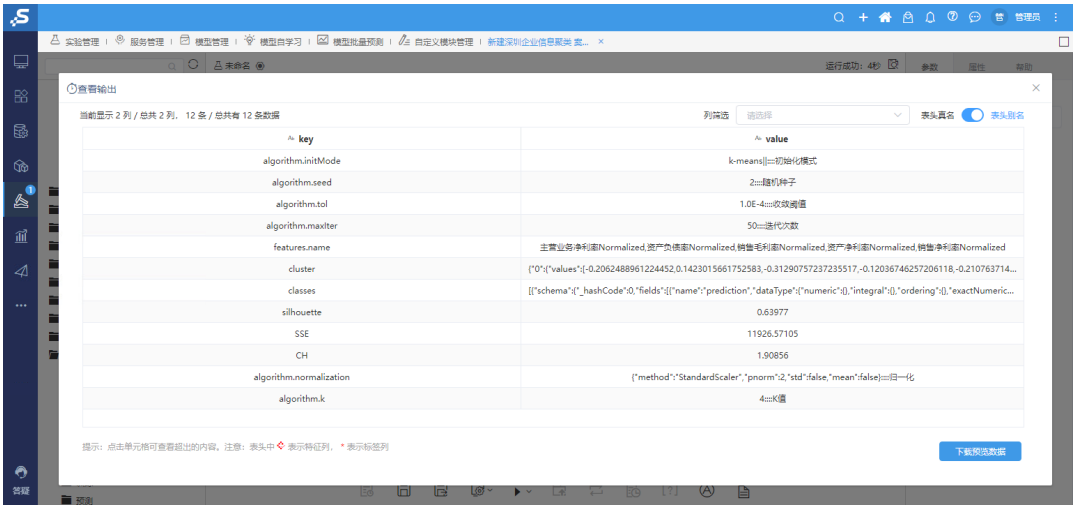



# ^【挖掘】评估节点支持输出结果

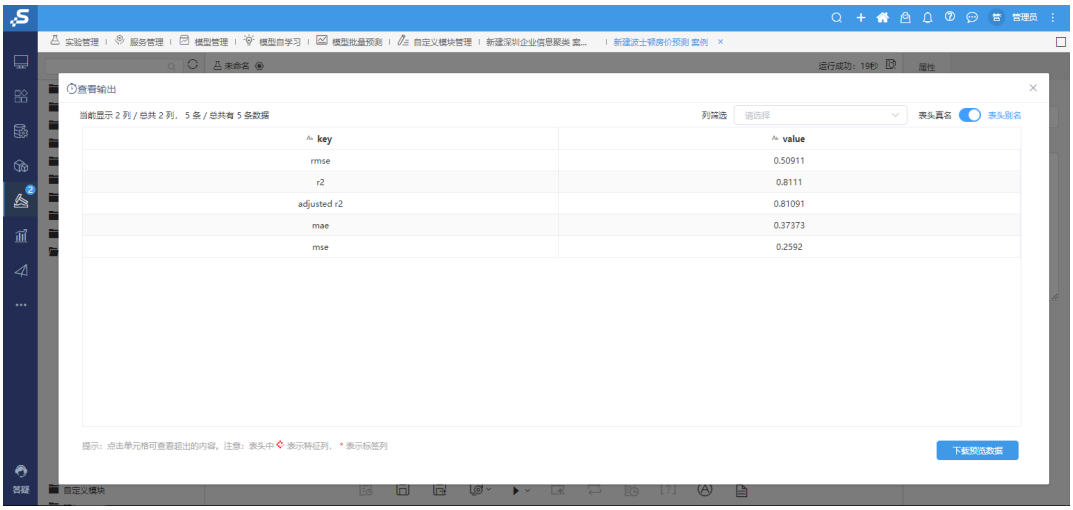
**功能简介**

在挖掘实验的建模流程中，模型训练以后进行验证时通常需要评估节点对模型进行评估。目前的评估节点只支持查看评估指标，不支持输出评估结果，在有大量模型训练的情况下难以把评估结果落地进行分析。所以需要在评估节点增加输出端口，输出评估指标结果。

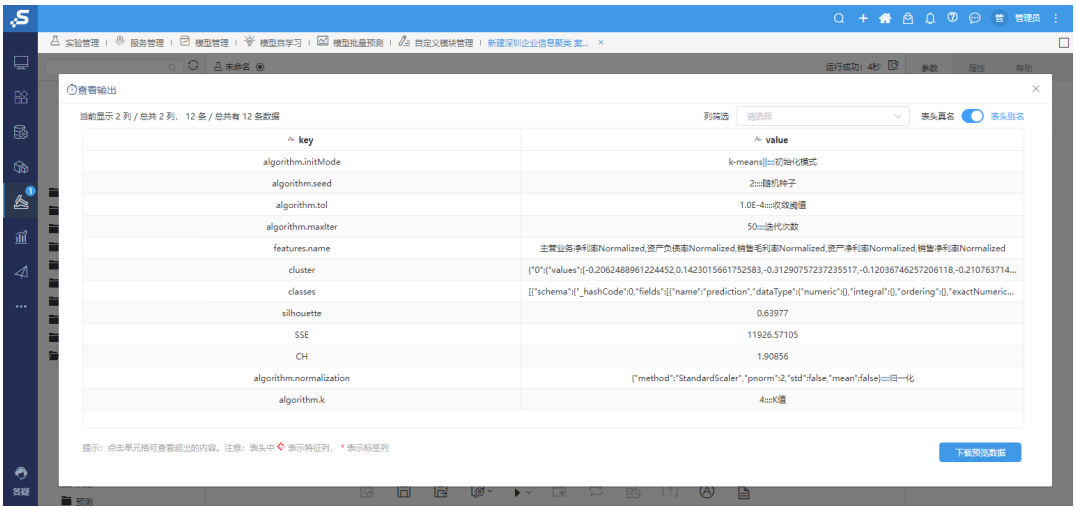
分类算法评估指标：




回归算法评价指标：



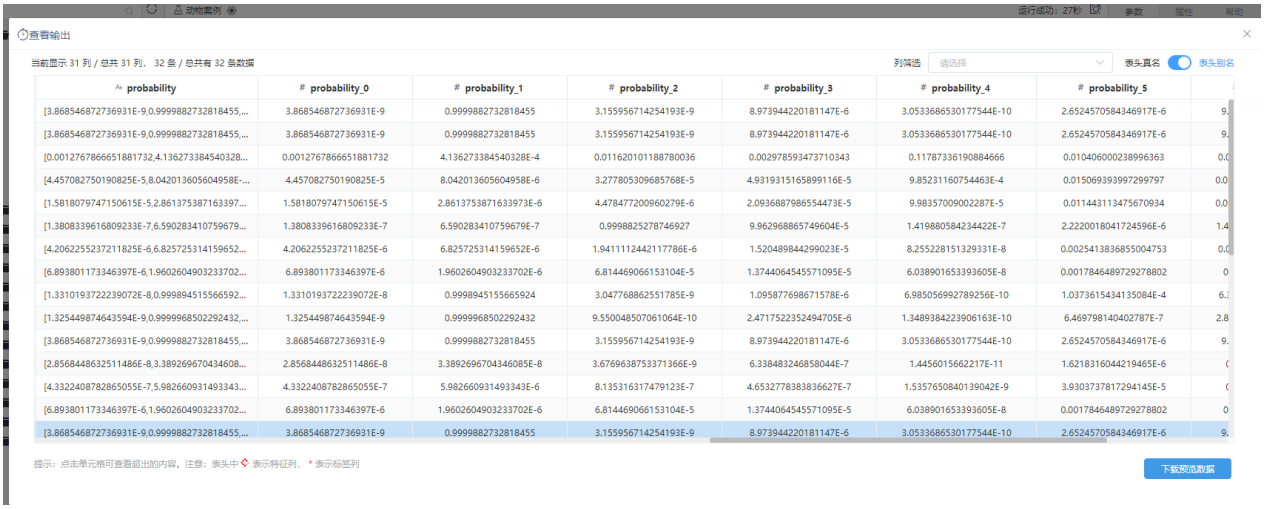

聚类算法评价指标：




# ^【挖掘】模型的预测中把预测概率输出成字段

**功能简介**

数据挖掘中的分类模型完成预测时，预测输出结果中包含了预测的概率，但目前这个概率字段格式不好，无法直接用于后续节点中进行处理，导致在项目中需要另写脚本提取相关预测概率的数据。所以需要支持把概率输出成如数值型的字段，更方便对预测结果进行分析处理。



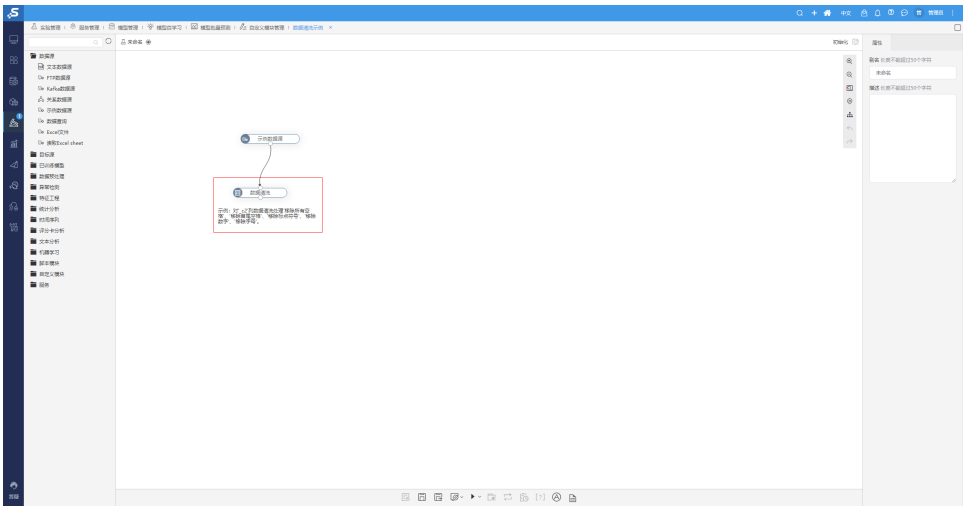
## ^【挖掘】数据清理节点合并

**功能简介**

新版本的数据清洗合并了旧版本中的空值处理、值替换、数据清理节点。通过该节点，可以实现以下几点功能：

（1）空值替换为均值、最大频数或者用户自定义的值等，实现空值的填充或者过滤；

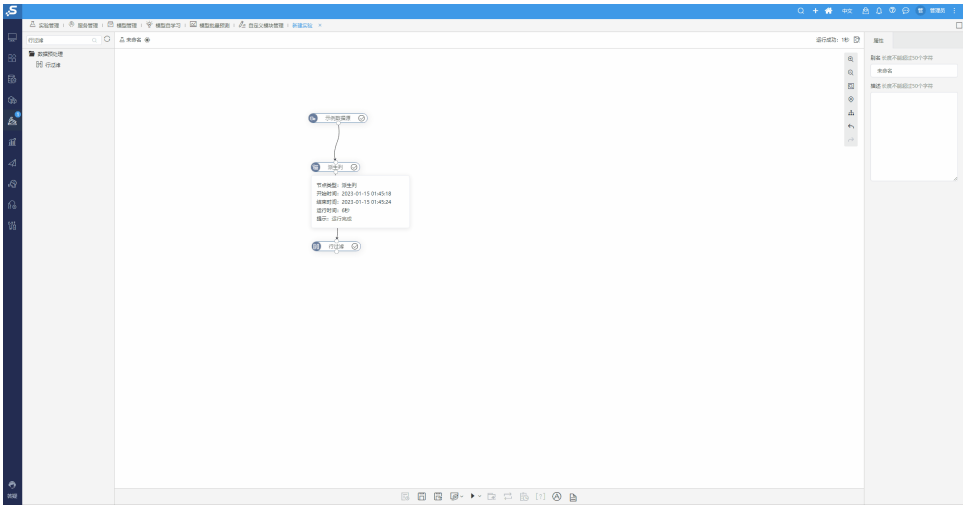
（2）移除字符串中空格、标点符号、字母、数字等不必要的字符，或设置大小写方式。



## ^【挖掘】过滤和行选择节点合并

**功能简介**

新版本的过滤节点，整合了旧的过滤节点和旧的行选择节点。提供了两种类型的筛选器。基本筛选器可以根据用户需求设置不同的筛选或者删除条件，选择不同数量的行；自定义筛选器通过写SQL语句(片段)，对数据按照过滤表达式进行筛选。



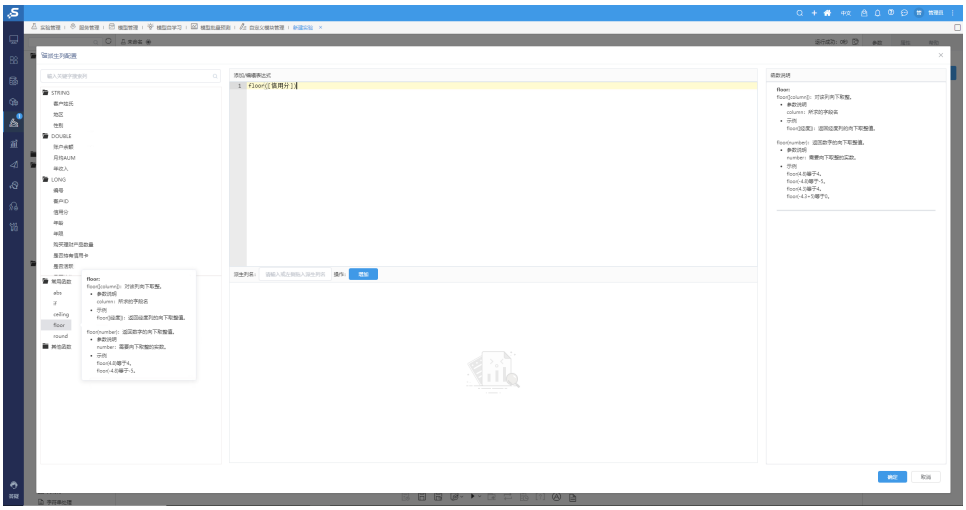
## ^【挖掘】派生列增强函数帮助提示

**背景介绍**

旧版本的派生列节点编辑界面中没有任何的指引，业务用户必须知道有这个函数，才能够写表达式，而且还没有语法规则和示例。

**功能简介**

新版本中对左侧函数资源树进行了优化，将用户常用的函数提前，并在鼠标悬浮时，以及左侧面板中显示对应的函数说明信息。



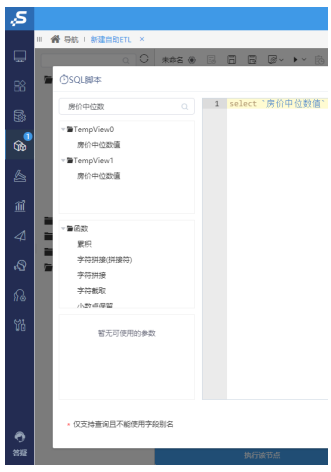
## ^【挖掘】SQL脚本输入表字段显示优化

**背景介绍**

SQL脚本节点主要通过执行Spark SQL语句，实现数据的查询。旧版本的SQL脚本节点中，没有显示节点相关的表及其字段，导致SQL书写困难。

**功能简介**

新版本中显示节点相关的表及其字段，用户可以直接拖拽表和字段，方便用户书写Spark SQL相关语句。

| 旧版本 | 新版本 |
| --- | --- |
|  |  |



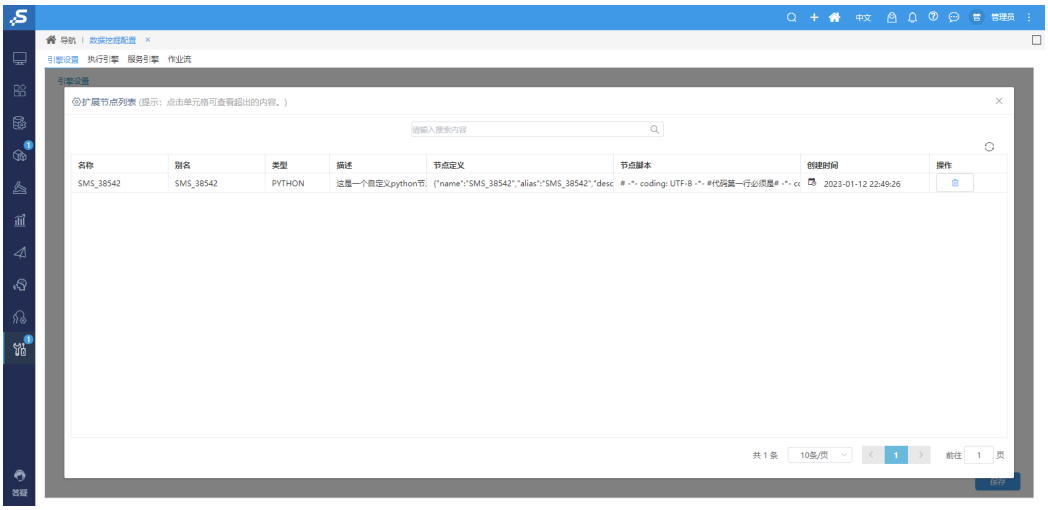
## ^【挖掘】提供节点iframe扩展配置组件

**背景介绍**

项目上有时需要一些特殊的配置项来进行自定义节点（自定义Java节点和自定义Python节点）开发，目前产品内置的配置项控件无法满足。

**功能简介**

因此，新版本中需要提供iframe扩展配置项，可以通过V的机制进行扩展开发，满足个性化需求。




**参考文档**

详情请参见：如何自定义java节点。