

雪浪算盘 产品说明书





目 录

目录

1	产品概述.....	3
2	产品特点.....	4
3	功能架构.....	6
4	操作指导.....	7
	4.1 使用入口.....	7
	4.2 界面介绍.....	9
	4.3 创建项目.....	13
	5.4 搭建项目.....	14
	5.5 运行项目.....	18
	5.6 查看结果.....	20
	5.7 模型部署.....	22
	5.8 创建自定义组件.....	24
	5.9 部分算法集说明.....	28
	5.10 系统部署.....	错误！未定义书签。
6	附 录.....	错误！未定义书签。
	6.1 数据可视化.....	错误！未定义书签。
	6.2 模型可视化.....	错误！未定义书签。



1 产品概述

雪浪算盘是雪浪数制科技自主研发的一款面向工业领域的工业人工智能开发平台。内嵌了丰富的机器学习算法、深度学习算法、可视化套件、数据处理套件、工业行业机理组件，以及内置了涉及供、研、产、销过程的工业智能算法模板，开箱即用，组态化操作，可以实现机理模型与算法模型的快速融合，可以快速完成数据模型的建立、调优、部署应用，提高效率降低成本，让工业企业快速、便捷地用好人工智能。



2 产品特点

更快

高维特征规模训练
Intel MKL, GPU 软硬件加速
Spark, TensorFlow 等高效计算引擎

更强大

一键式任务部署和管理
通用 PMML 模型导出
自定义模型部署应用

更灵活

全 Docker 独立部署
多种异构数据集: ODPS/Hive
算法定制开发

更简单

工业面板自定义
图表拖拽显示
部署既应用



灵活的框架支持

支持 MaxCompute、Hadoop 等多计算引擎

数据源支持 HDFS, ODPS, OSS, MySql 等

支持开源 Spark 工具

支持 TensorFlow、PyTorch 等深度学习框架

可视化开发

拖拽式 workflow 开发

可视化 workflow 运行

运行状态实时显示

丰富组件类型

数据采集组件：数据采集，导入，验证，处理组件

机器学习算法：特征工程，机器学习，可视化分析，文本处理组件

可视化套件：数据可视化，模型可视化，时序数据分析

以上类型均支持自定义组件开发

在线预测

在建模完成后，可以一键部署为在线预测集群，提供模型在线预测服务

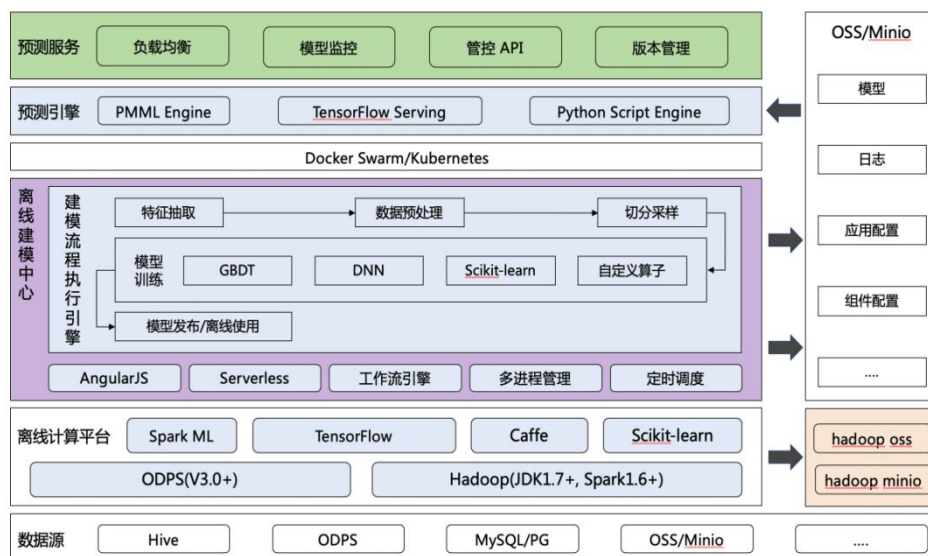
支持以可视化的方式进行 web 输入输出的配置

支持多个模型的串联预测

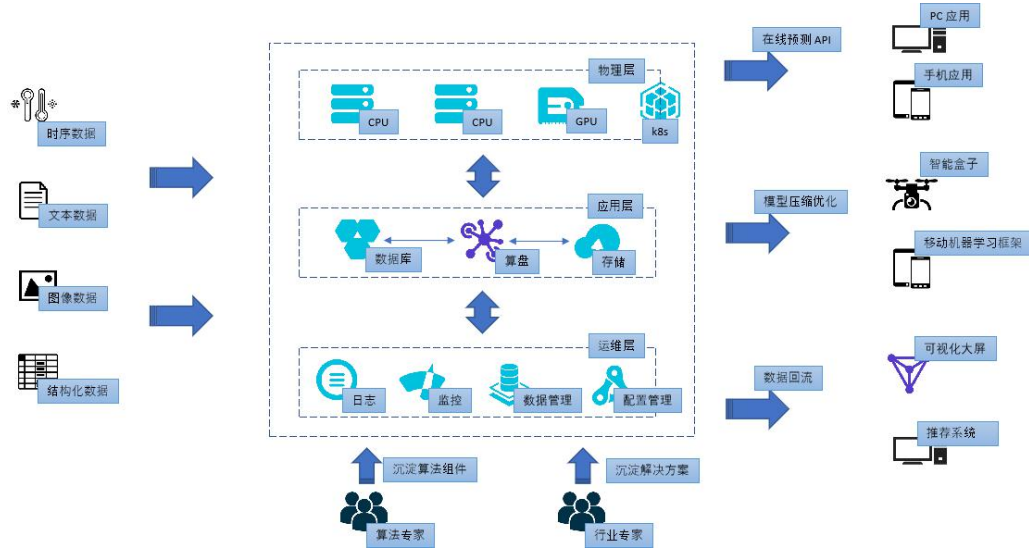


3 功能架构

系统至下而上分 3 层，数据接入层、离线数据计算服务层、实时计算服务层，共 8 大模块组成。其中数据接入层的功能在离线计算和实时计算中都有使用。



系统层次结构关系图



系统数据流程图

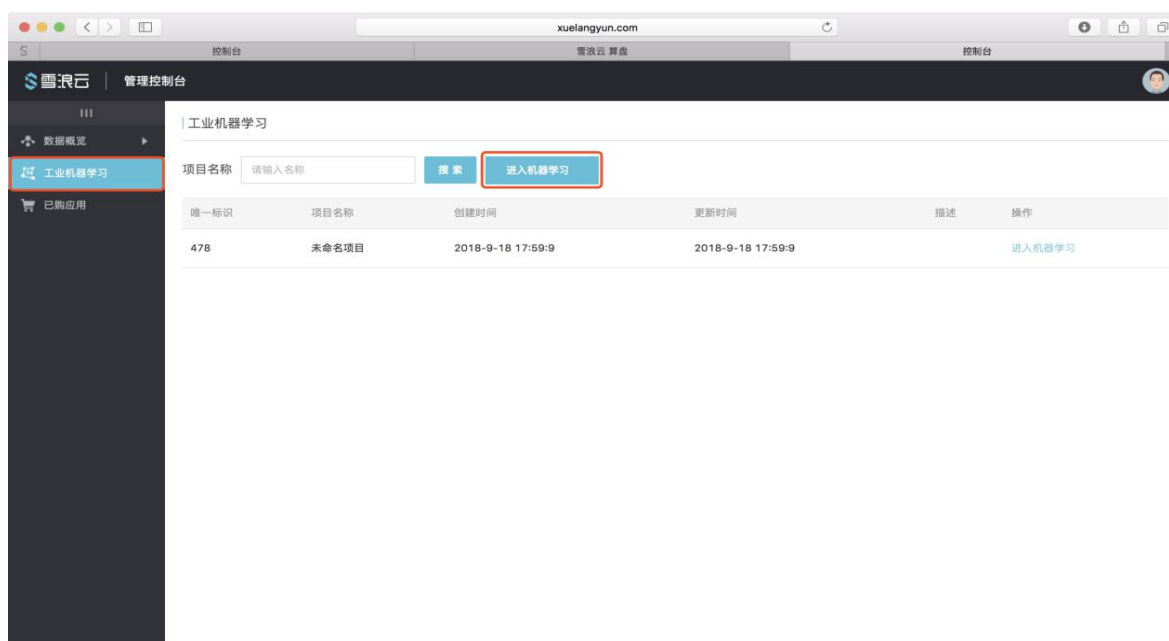
4 操作指导

4.1 使用入口

➤ 公有云版

登录雪浪云官网 (<http://www.xuelangyun.com>)，进入控制台，选择工业机器学习，进入工业机器学习。

首次使用需注册为雪浪云用户，申请使用工业机器学习。审核通过后即可使用工业机器学习（雪浪算盘）。



➤ 个人版

1. 准备安装包

从公有云算盘上首页右上角点击下载个人版安装包，支持 Windows、Mac、Linux 操作系统。



4.2 界面介绍

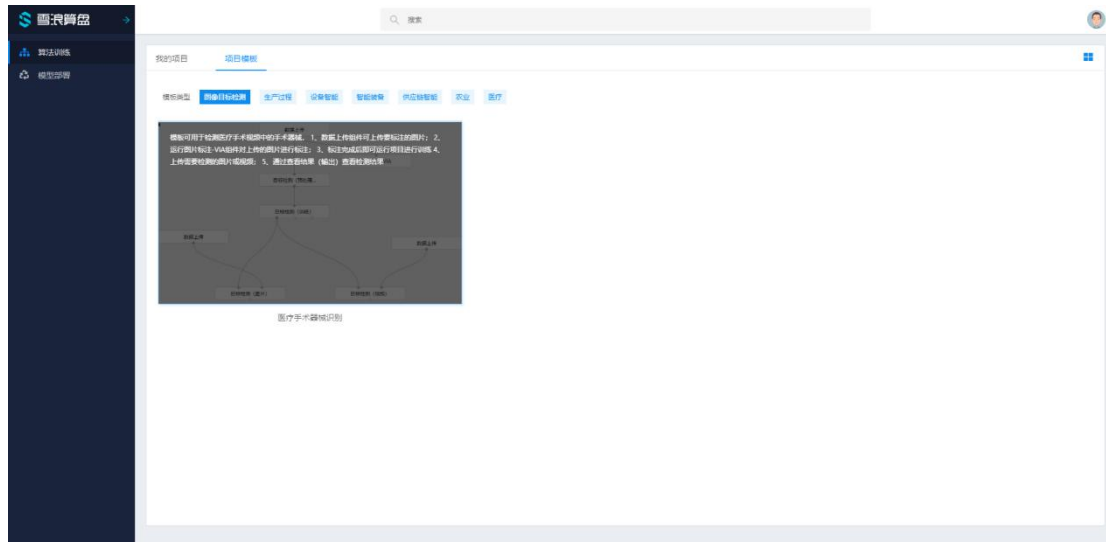
4.2.1 算盘功能分为：算法训练、模型部署、前面板设计

- 算法训练：为算法开发人员提供的图形化开发环境，通过组件连接搭建算法模型，输入数据进行训练，验证算法准确率。训练好的模型可发布成服务进行线上调用或边缘部署应用。
- 模型部署：通过服务组件和训练好的模型搭建部署服务，生成 API 调用，支持 Web 调用和边缘部署本地调用。
- 前面板设计（模型部署中）：在模型部署中为业务技术人员提供的图形化拖拽组件搭建前端显示页面，运行后可通过前面板显示模型部署中的数据并进行交互操作，可直接应用于生产现场。

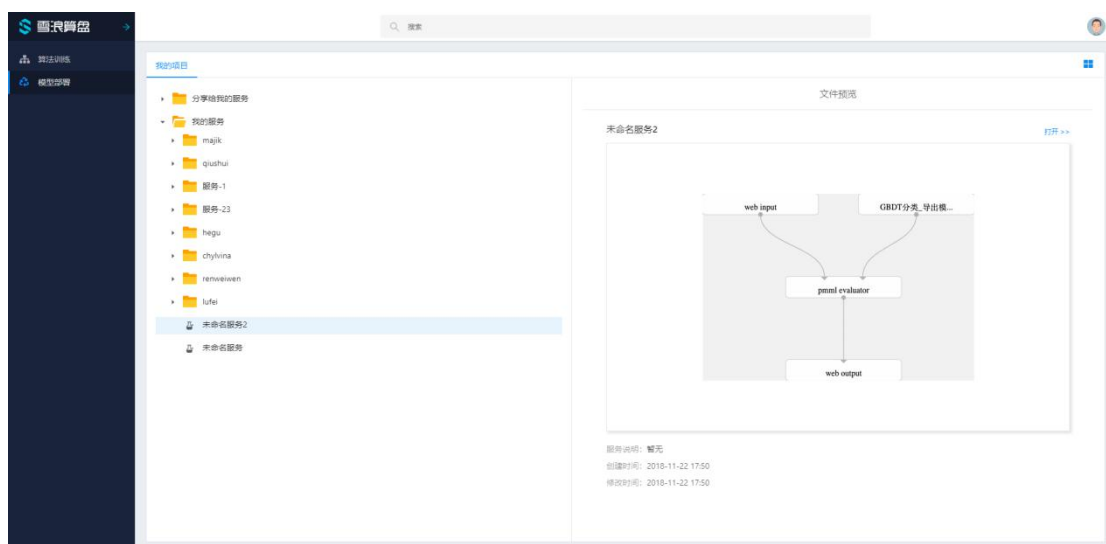


4.2.2 项目模板

雪浪算盘内置丰富模板，可直接订阅使用。模板包括工业领域的生产过程、智能设备、供应链智能以及农业和医疗等。

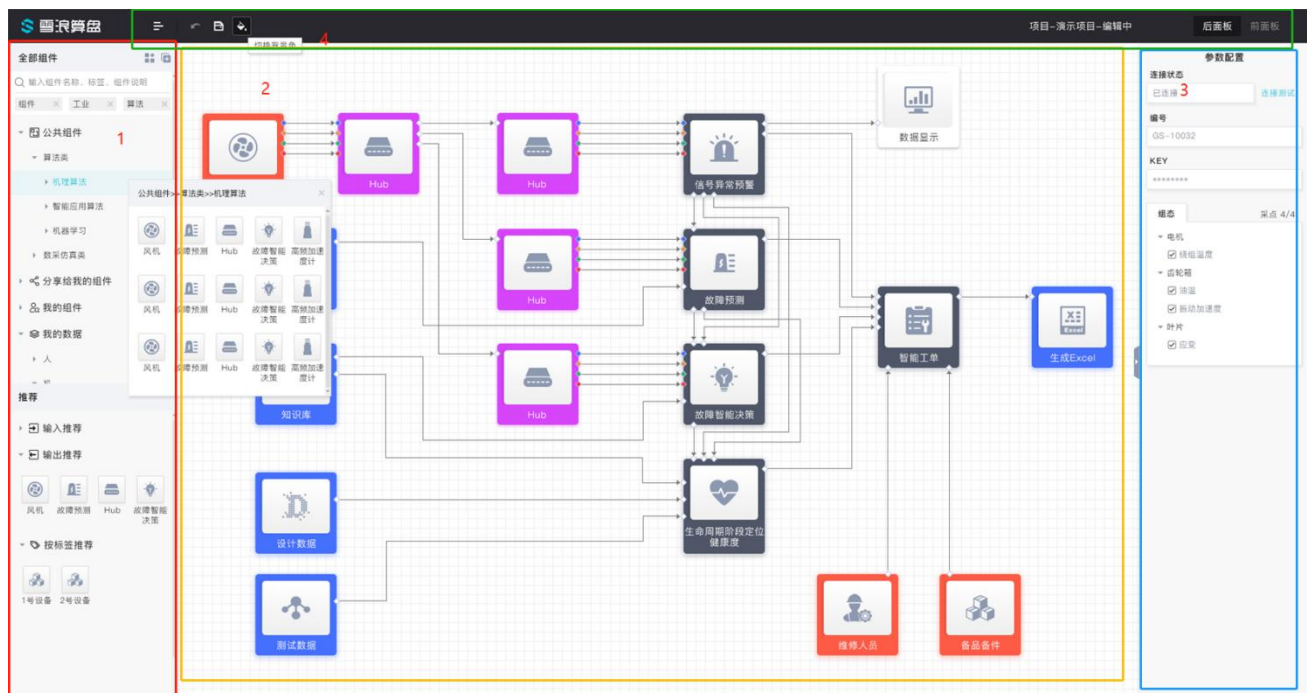


4.2.3 项目列表





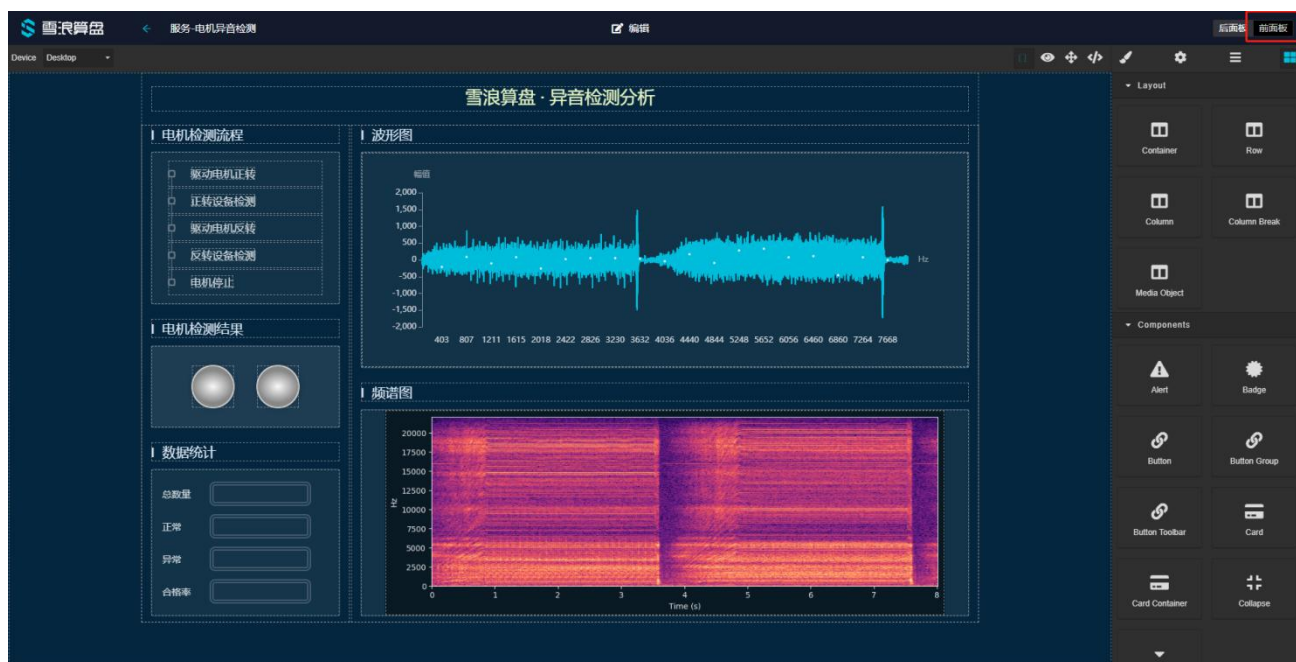
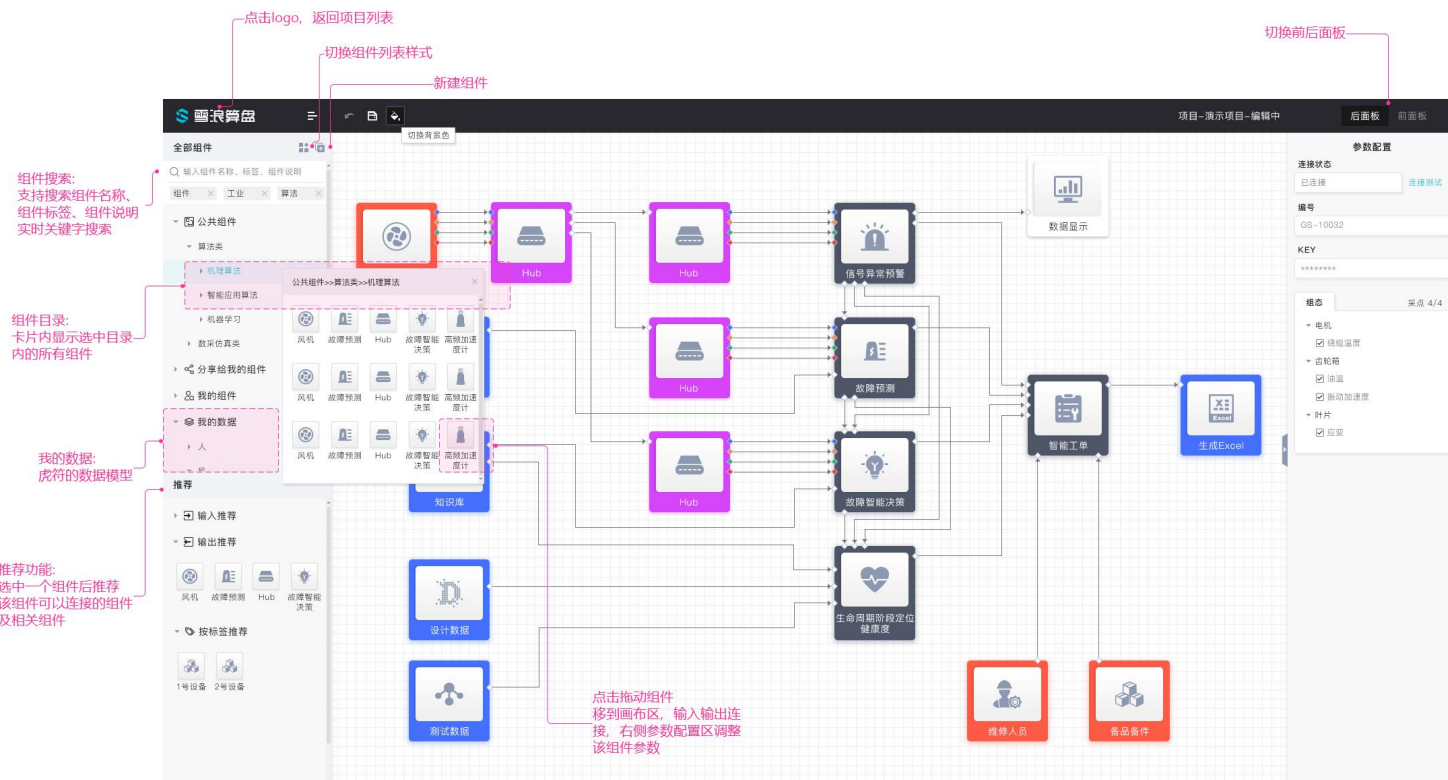
4.2.4 算法训练工作台



- 红色 1 组件栏：组件库，公共组件包括数据类、算法类、机理模型类，可搜索需要的组件拖进画布区进行使用。
- 黄色 2 项目编辑区：算法模型项目的开发区域，拖拽组件，连接不同组件之间的输入输出。
- 蓝色 3 属性去：配置项目及组件属性，调整优化组件参数等。
- 绿色 4 控制区：返回目录，编辑项目、运行&停止、运行历史等功能按键。



4.2.5 部署应用及前面板设计

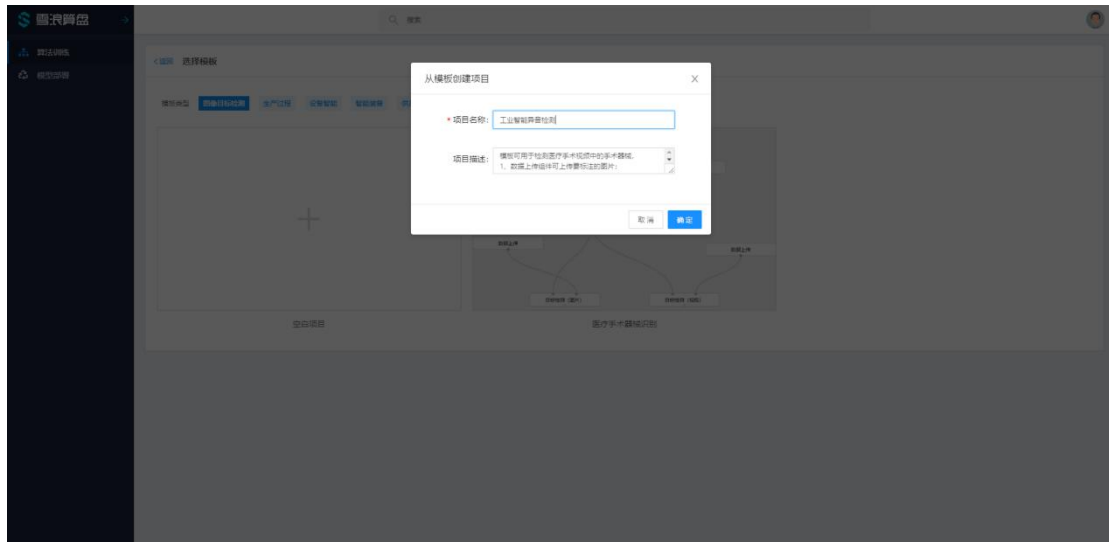




4.3 创建项目

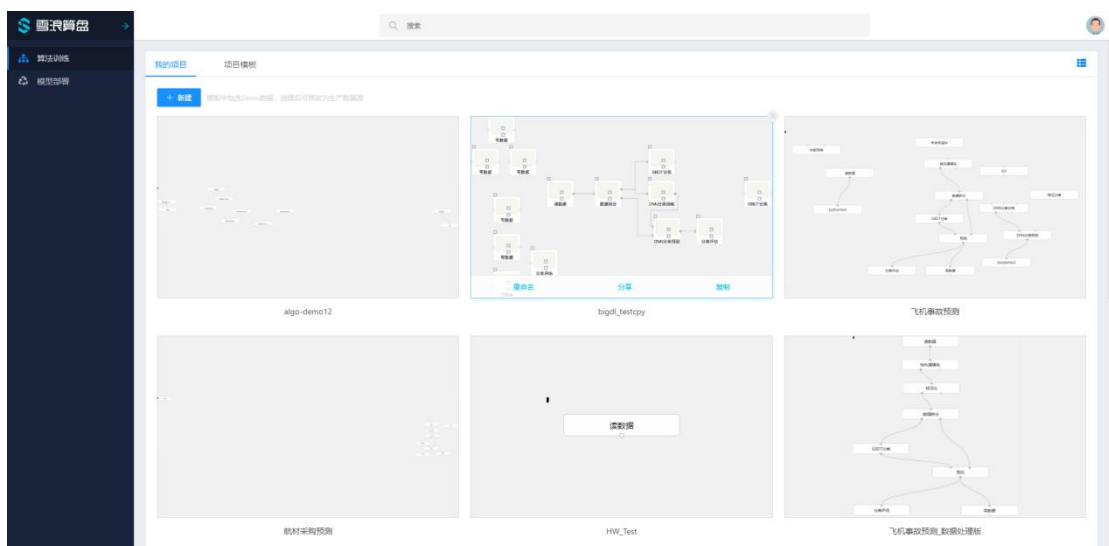
4.3.1 从模板中创建

模型训练 -> 项目模板，选择符合业务场景的项目模板，点击后弹出新建项目编辑框，输入项目名称、项目描述即可创建项目。



4.3.2 创建空白项目

模型训练 -> 我的项目，点击“新建”即可生成新的空白项目。

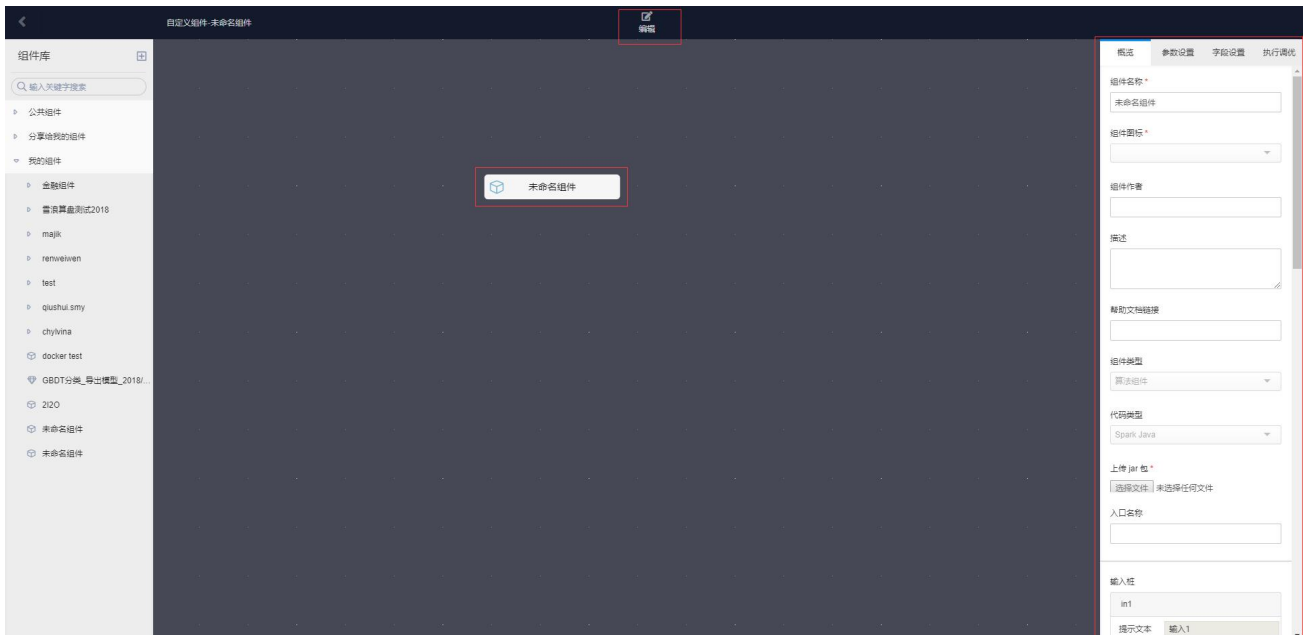




5.4 搭建项目

5.4.1 编辑项目

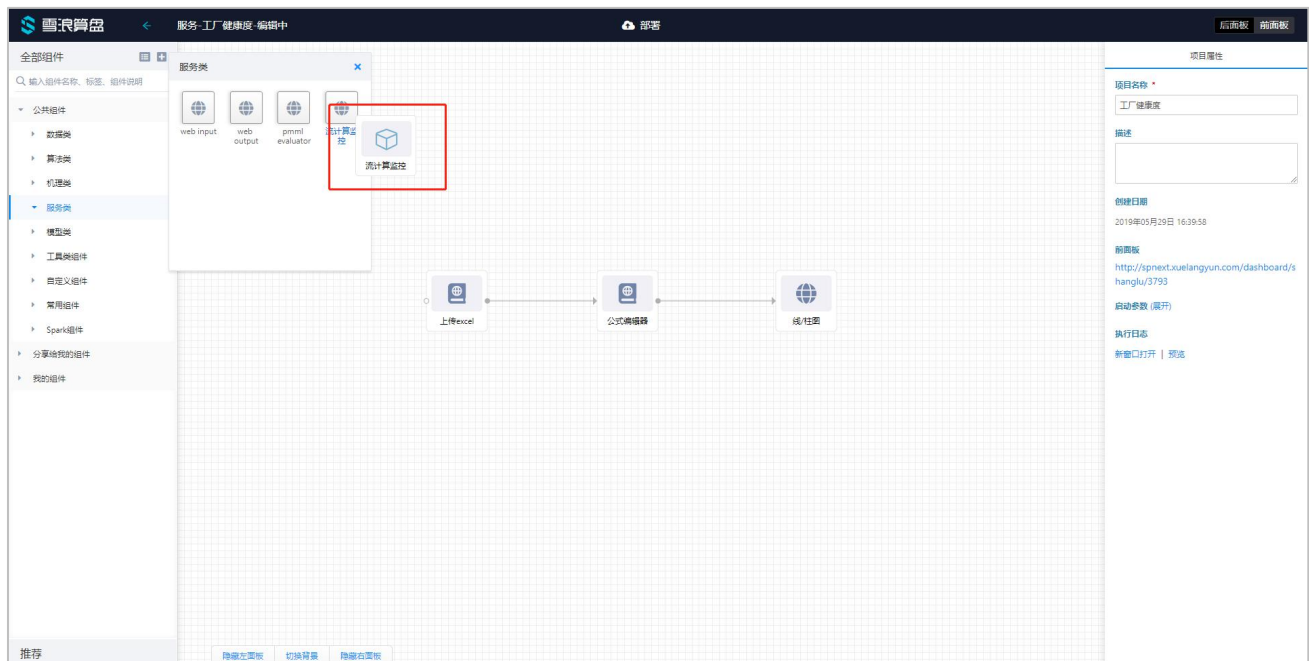
创建项目完成后，点击“编辑”获取编辑项目权限（同一时间只能进行唯一用户编辑），即可添加组件，搭建项目。（鼠标悬停在组件上 1s，可以查看组件使用说明）



5.4.2 添加组件

在“公共组件”和“我的组件”中拖拽任意组件到“项目建模编辑区”

项目通常第一个组件为数据上传，算盘支持上传的数据类型包括：txt、csv、excel、图片、文件等，也支持 MySQL、ODPS 等数据库直接读取数据。



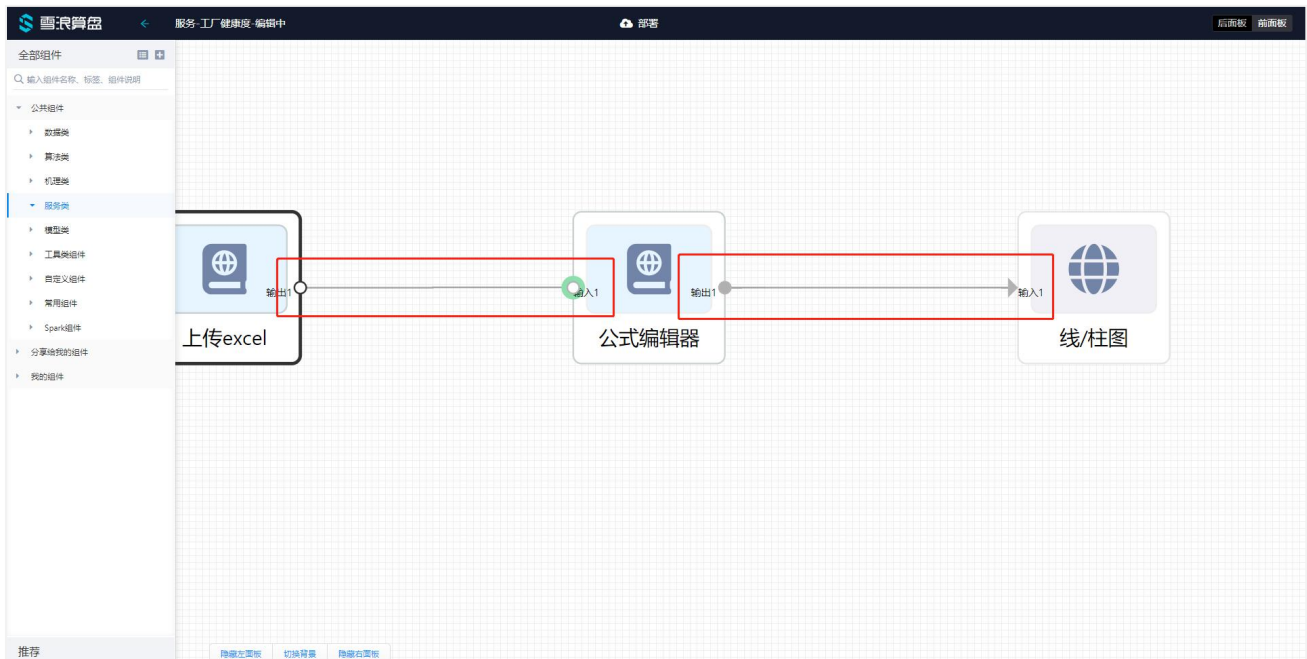
5.4.3 连接组件

点击并某个组件下面的输出点，会产出一条带箭头的线，再点击另一个组件上面的输入点，即可将两个组件连接在一起。

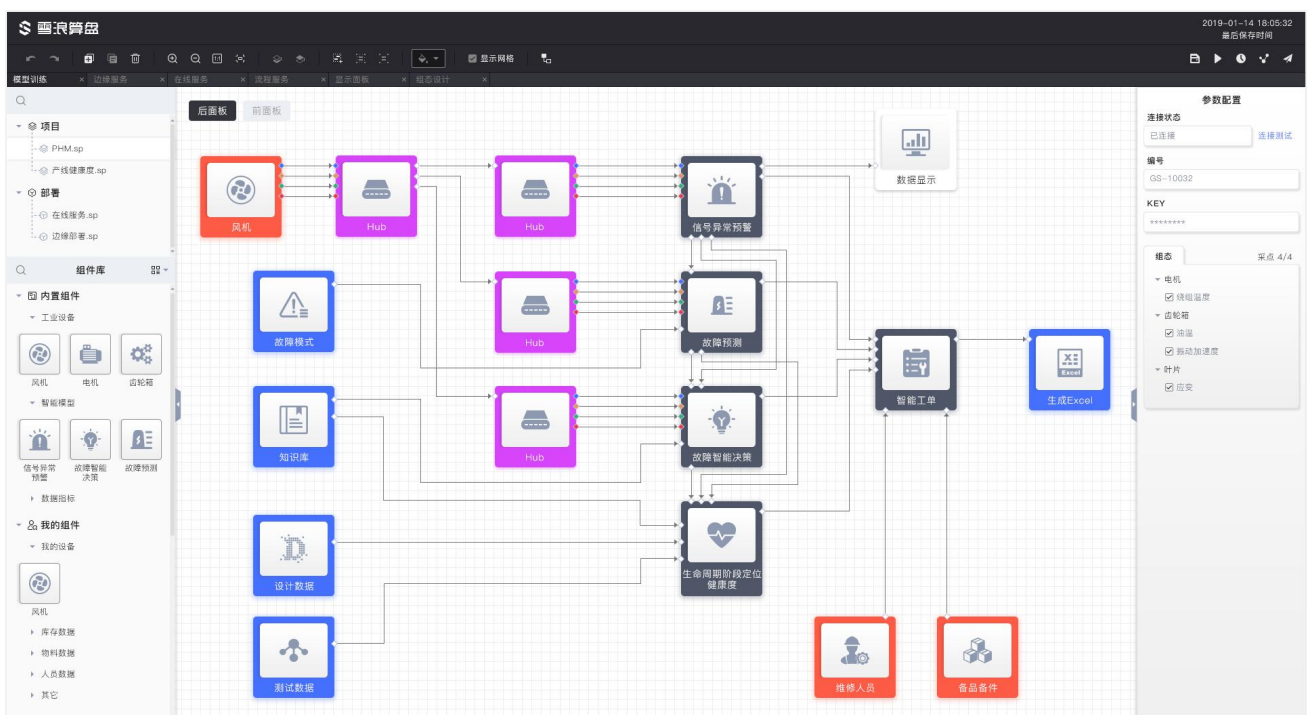
注意，一个组件的输出可能会有 4 种类型，即

- 数据表
- 模型
- 非结构化数据
- 可视化数据(不会作为输入)

对应的，一个组件的输入也是上面这 3 种类型。所以，如果一个组件的输入被定义为数据表，那么只能将输出为数据表的输出与其输入相连。如下图所示，读数据组件的输出为数据表，数据拆分组件的输入为数据表，因此可以相连。



以此类推，按照自己的想法将多个组件连接起来，形成一个完整的算法流程。



5.4.4 组件属性

组件概览:

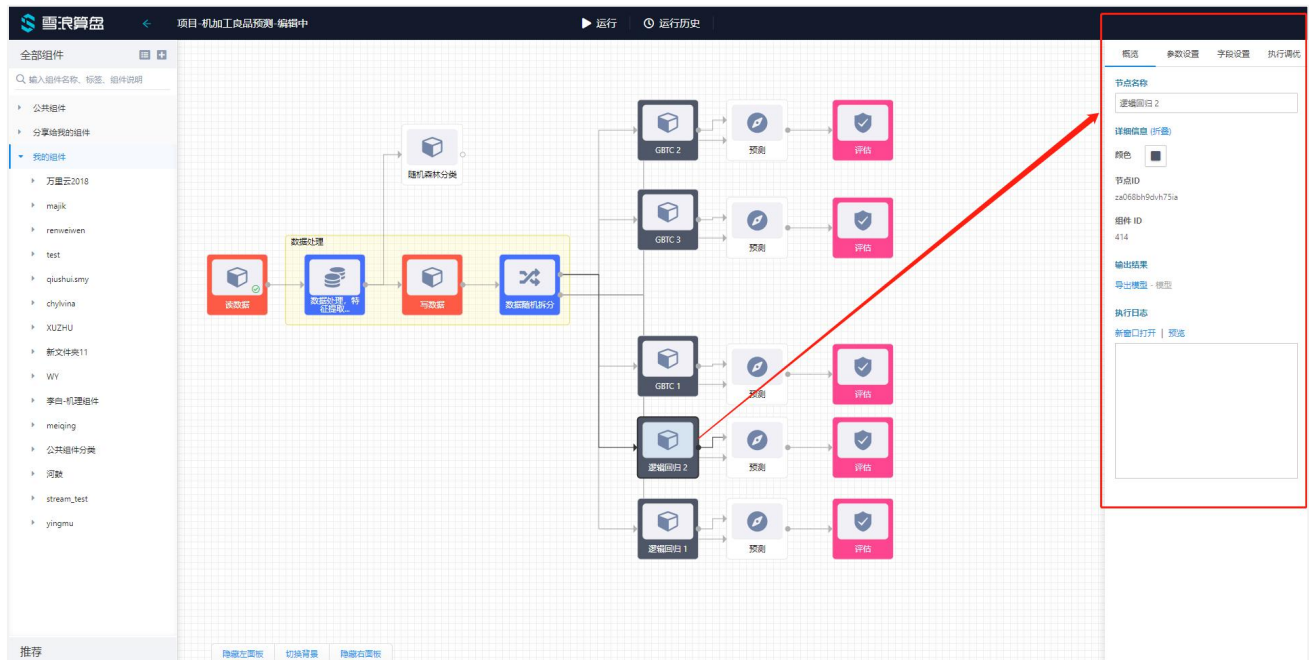
点击组件名称，右侧会出现组件的编辑区域，包括

- 组件名称: 说明组件的功能，可以修改
- 节点操作: 删除 (Backspace\Delete)
- 查看输出结果: 运行成功后，可以查看输出结果，不同组件的输出结果不一样，有数据表、模型、可视化数据、



非结构化数据

- 执行日志：显示该组件运行过程中的日志，若组件执行错误可以在此查看报错信息



参数设置&字段设置&执行调优

- 参数设置：设置组件的参数，如GBDT算法的迭代步长，树个数等参数
- 字段设置：主要设置输入表有关的参数，如训练特征列（下图中x/y就是输入表的字段）
- 执行调优：并行计算需要的CPU核数和每个CPU分配的内存



概览	参数设置	字段设置	执行调优
maxDepth:	5		
maxBins:	32		
minInstancesPerNode:	1		
minInfoGain:	0		
maxMemoryInMB:	256		
cacheNodeIds:	<input type="checkbox"/>		
checkpointInterval:	10		
lossType:	logistic		
maxIter:	20		
stepSize:	0.1		
seed:			

概览	参数设置	字段设置	执行调优
		标识字段: *	
		输入或者选择字段	
		特征字段: *	
		输入或者选择字段	

概览	参数设置	字段设置	执行调优
			executor数目:
			默认自动选择
			每个executor的core数目:
			默认自动选择
			每个executor的内存数(GB):
			默认自动选择
			队列:
			默认为 default

5.4.5 训练算法

可以根据查看可视化数据、模型及输出结果，验证模型的准确确定，调整模型组件的“参数设置”提高模型的准确性。

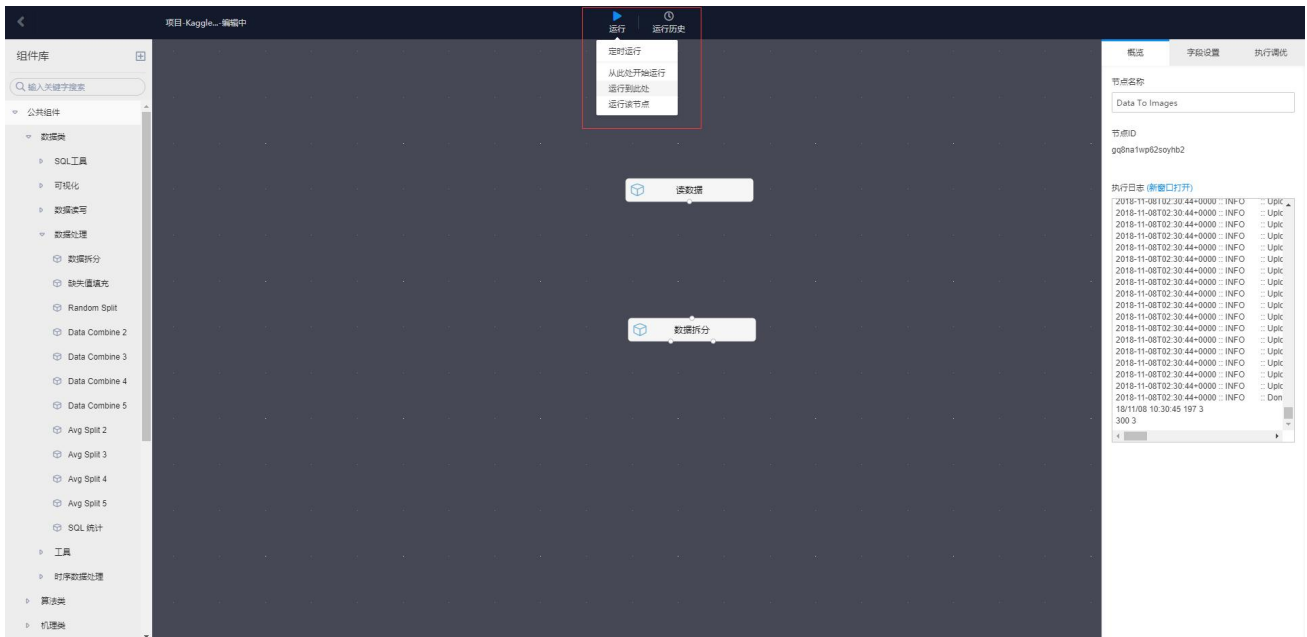
5.5 运行项目

5.5.1 执行项目

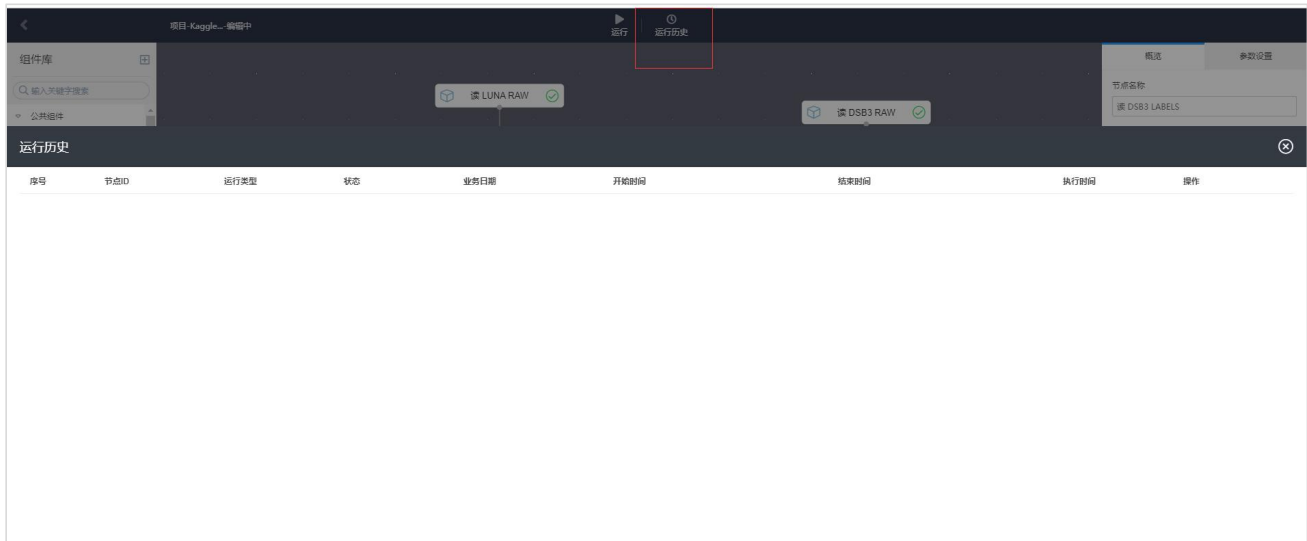
在运行中选择“从此开始执行”，就可以运行整个模型了，每个组件上会显示执行状态，绿色对勾表示运行成功，红色叉表示运行失败，可以点击运行失败的组件，在右侧组件操作区中的“执行日志”中查看失败原因。



在运行中也可以选择，定时运行、执行到此处、执行该节点

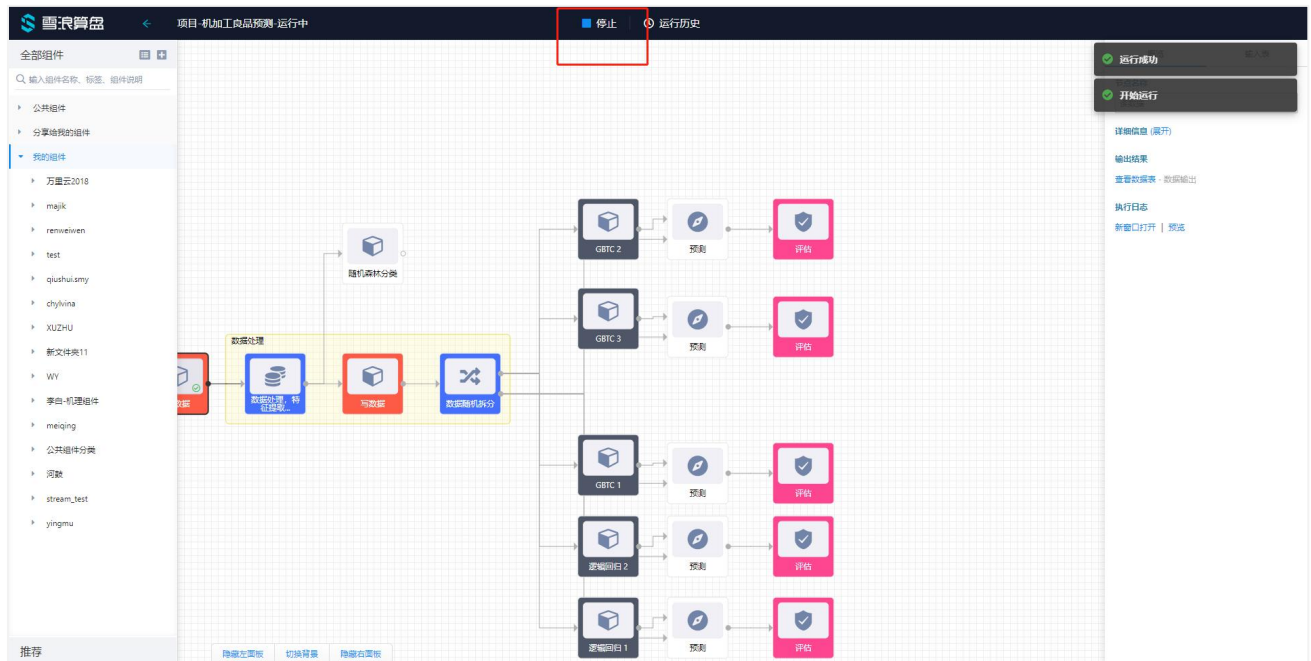


查看运行历史



5.5.2 停止运行项目

在项目运行过程中，如果想要停止运行项目，点击任意一个节点，在运行中点击“停止执行”即可停止项目。



5.6 查看结果

5.6.1 查看可视化数据

从左侧“公共组件”的“可视化套件”中拖拽出“散点图”，点击该组件右侧栏中的“执行到此处”，配置坐标参数，运行成功后，点击右侧栏中“查看可视化”，会出现以下弹窗。

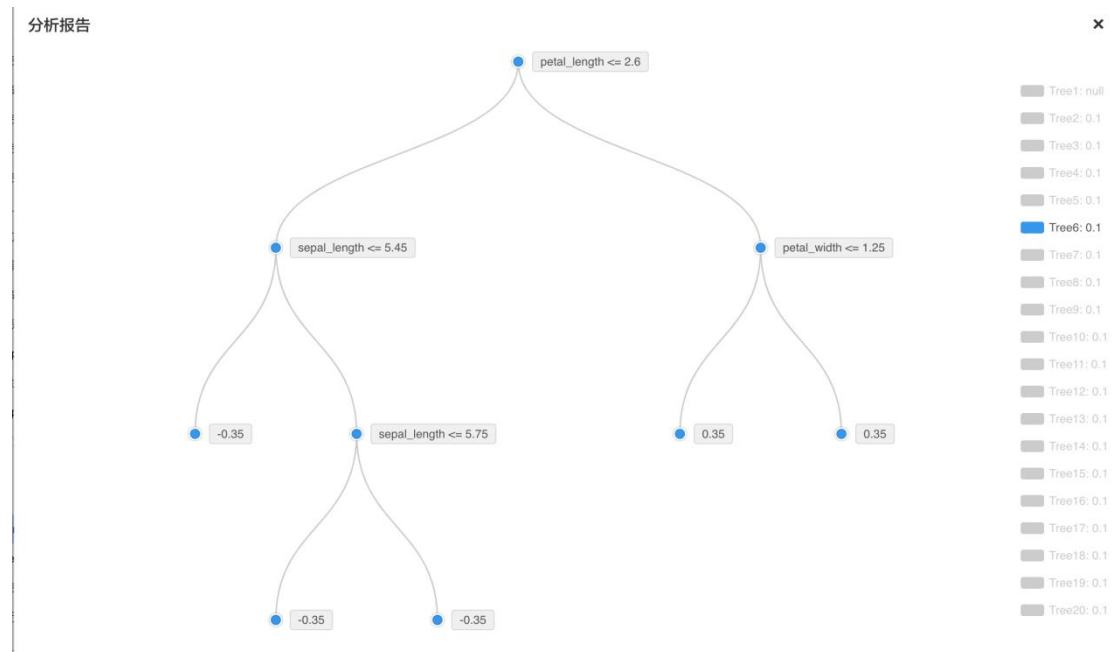




5.6.2 查看可视化模型

从左侧“公共组件”的“可视化套件”中拖拽出“GBTC 模型可视化”

点击该组件右侧栏中的“执行到此处”，运行成功后，点击右侧栏中“查看可视化”，会出现以下弹窗。



5.6.3 查看输出结果

从左侧“公共组件”的“算法类”中拖拽出“预测”

点击该组件右侧栏中的“执行到此处”，运行成功后，点击右侧栏中“查看数据表（预测结果）”，会出现以下弹窗。



数据探查

序号	htime	val	predic...	point_id	equip_id	residual
0	2018-...	-0.199...	3.074...	effect	14950...	-3.274...
1	2018-...	-0.199...	3.074...	effect	14950...	-3.274...
2	2018-...	-0.199...	3.074...	effect	14950...	-3.274...
3	2018-...	-0.199...	3.074...	effect	14950...	-3.274...
4	2018-...	-0.199...	3.101...	effect	14950...	-3.301...
5	2018-...	-0.199...	3.101...	effect	14950...	-3.301...
6	2018-...	-0.199...	3.101...	effect	14950...	-3.301...
7	2018-...	-0.199...	3.101...	effect	14950...	-3.301...
8	2018-...	-0.199...	3.101...	effect	14950...	-3.301...
9	2018-...	-0.199...	3.095...	effect	14950...	-3.295...
10	2018-...	-0.199...	3.095...	effect	14950...	-3.295...
11	2018-...	-0.199...	3.095...	effect	14950...	-3.295...
12	2018-...	-0.199...	3.095...	effect	14950...	-3.295...
13	2018-...	-0.199...	3.095...	effect	14950...	-3.295...
14	2018-...	-0.199...	2.971...	effect	14950...	-3.171...
15	2018-...	-0.199...	2.971...	effect	14950...	-3.171...
16	2018-...	-0.199...	2.971...	effect	14950...	-3.171...

5.7 模型部署

5.7.1 新建服务

项目运行成功后，模型就保存在文件服务器中了，这时可以通过部署服务来实现模型的在线调用。

模型部署 -> 我的项目，点击新建即可创建服务。

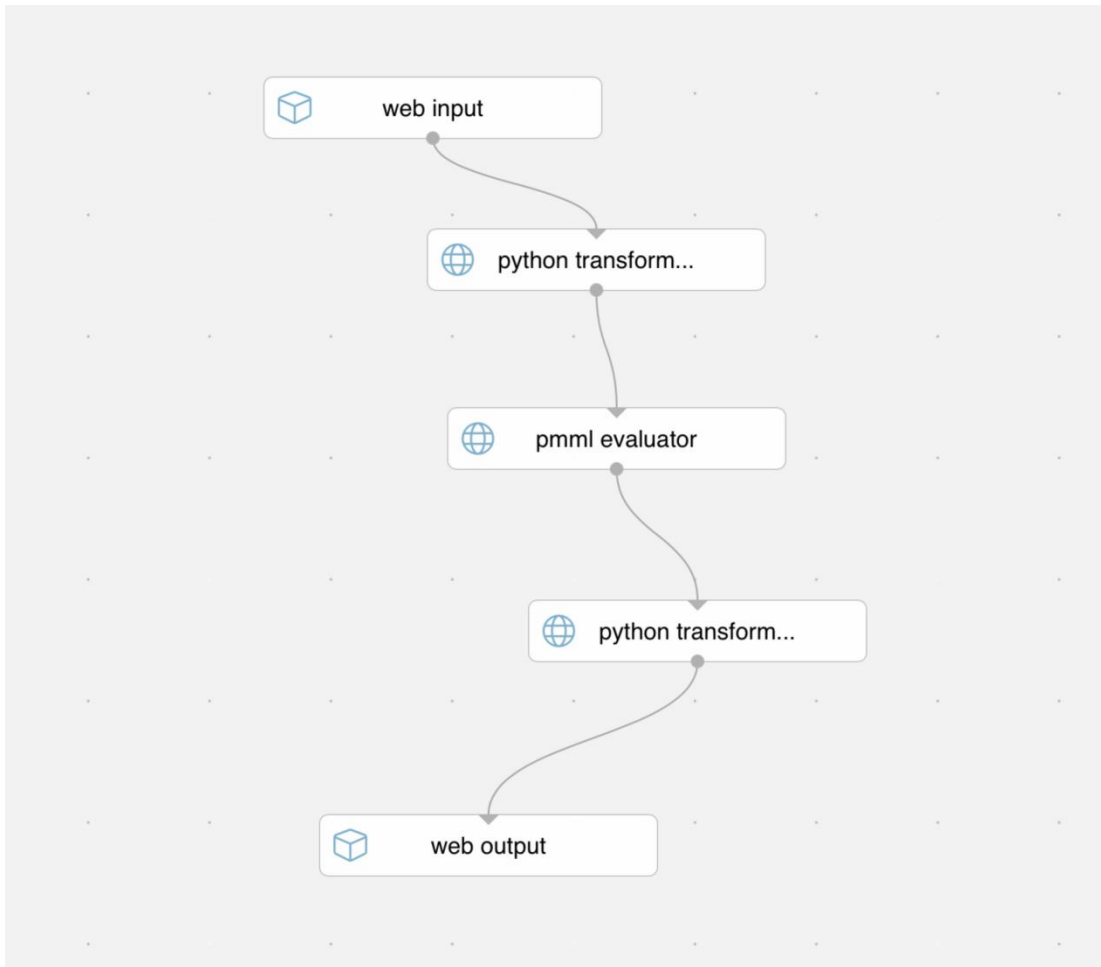
在组件列表中选择服务类，拖拽进工作区即可进行使用（操作和搭建项目相同）。

5.7.2 配置 web 输入输出

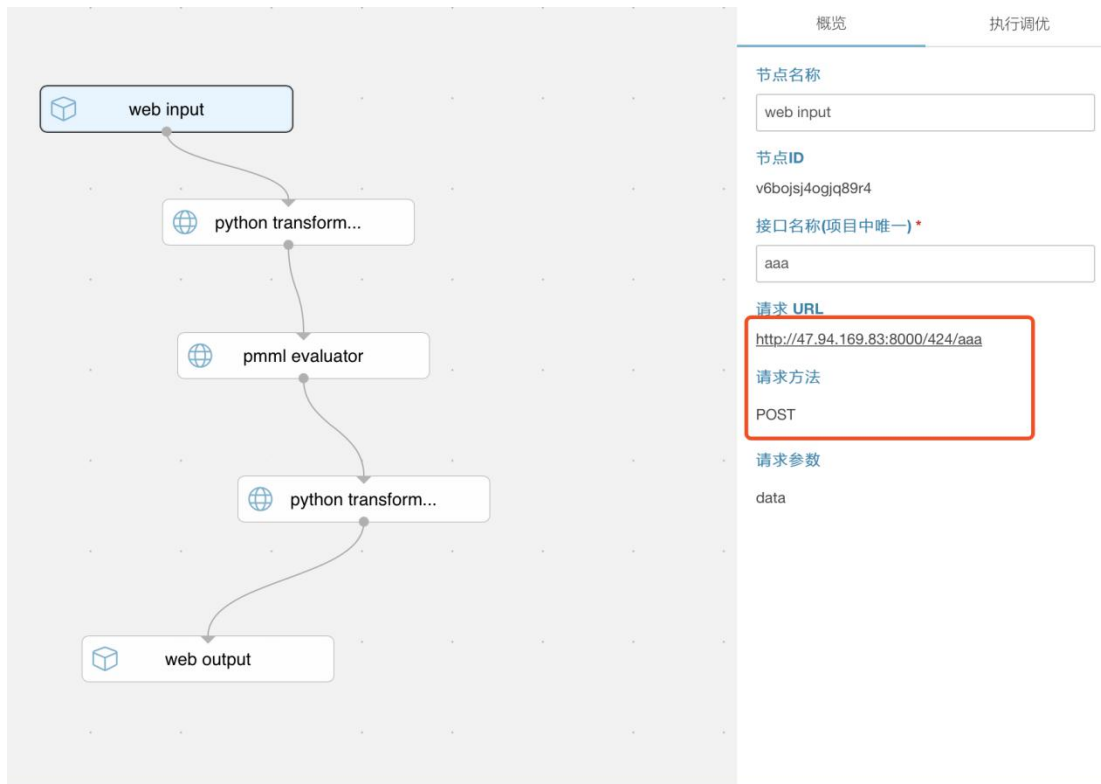
web input 节点负责接收 web 输入，然后通过其输出端口将数据传到与其连接的节点。Web output 节



点表示要将传到其输入端口的数据返回给 web 请求。



点击 web input 节点，在右侧的配置栏中可以看到如下界面：



请求 URL 由三部分组成，即 `http://[host:port]/[web 服务名称]/[接口名称]`。其中 `[host:port]` 是系统配置，`[web 服务名称]` 在项目属性中配置，全局唯一，`[接口名称]` 是该 `web input/web output` 节点对应的接口的名称。

5.8 使用自定义组件

雪浪算盘支持创建组件（同时支持 `vscode` 在线编辑代码，搜索 `vscode` 组件使用），在功能栏中选择“组件”，点击组件列表上方“新建图标”即可创建组件。

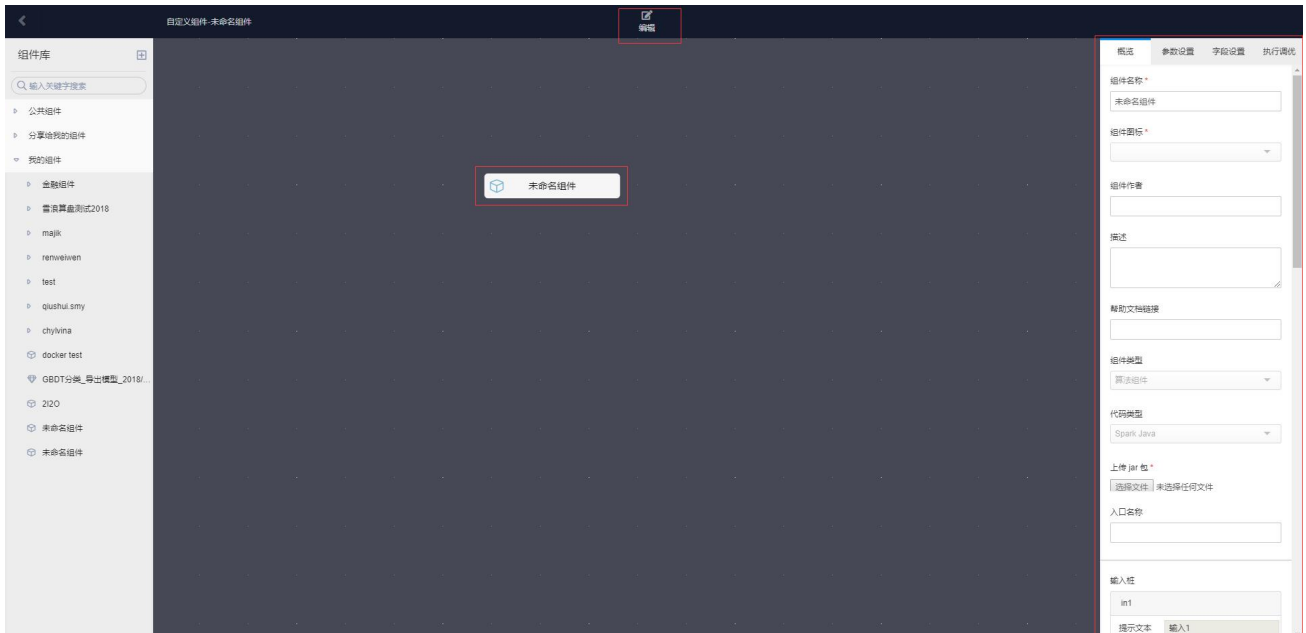
自定义组件类型包括：算法组件、服务组件、复合组件





5.7.3 编辑组件属性

新建组件完成后，会在右侧出现组件编辑界面，点击“编辑”后即可修改组件参数。



5.7.4 组件概览配置

配置组件名称、组件图标、组件作者、描述、帮助文档链接后，选择“组件类型”、“代码类型”（目前支持 java、python）中的一种，选择文件上传。

上传后，配置输入输出，最多配置 5 个输入和 5 个输出

提示文本是该输入输出的名称

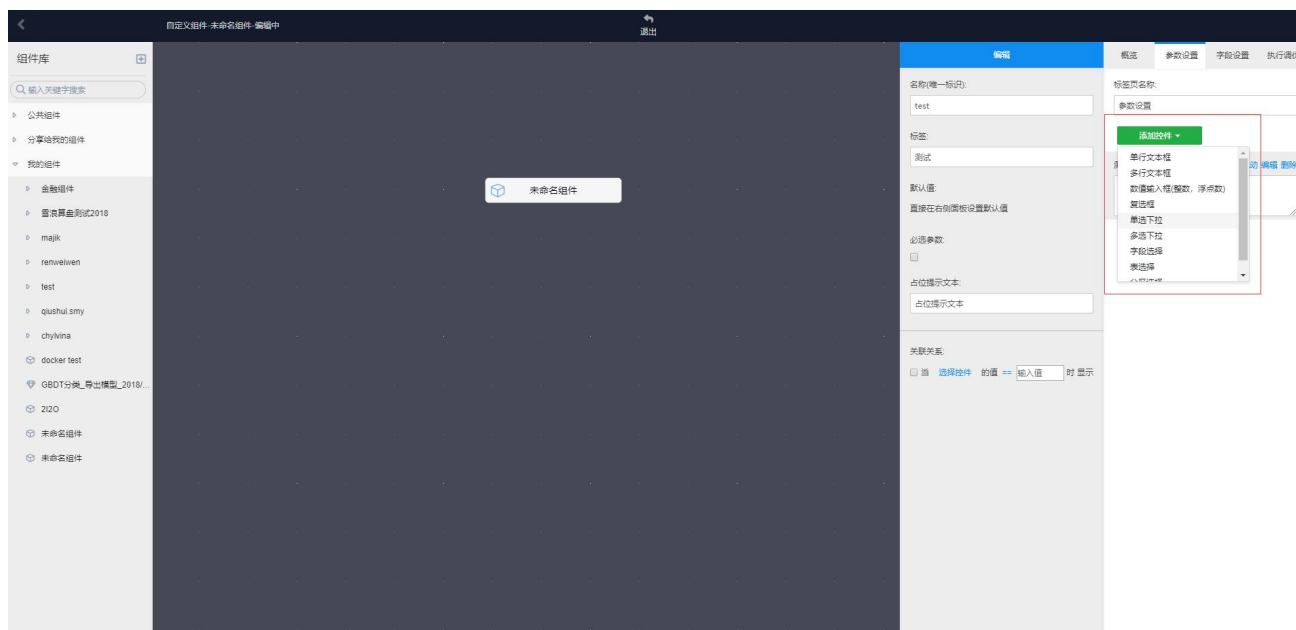
输入类型包括数据表、模型、非结构化数据

输出模型包括数据表、模型、可视化数据、非结构化数据

是否显示，勾选则在项目模型编辑时可以连接其他组件的输入输出

组件参数配置

可以选择多种控件配置参数的输入方式



5.7.5 组件字段配置

可以选择多种控件配置字段的输入方式





5.7.6 执行调优配置

默认了 CPU 数量、核数、内存数的设置，可以删除调整位置

名称(唯一标识): coreNum

标签: 核心数

默认值: 默认自动选择

直接右侧面板设置默认值

必选参数: ☐

占位提示文本: 默认自动选择

关联关系: ☐ 当 选择控件 的值 == 输入值 时 显示

标签页名称: 执行调优

添加控件

核心数: 默认自动选择

内存数: 默认自动选择

也可以选择多种控件配置调优的输入方式

名称(唯一标识): coreNum

标签: 核心数

默认值: 默认自动选择

直接右侧面板设置默认值

必选参数: ☐

占位提示文本: 默认自动选择

关联关系: ☐ 当 选择控件 的值 == 输入值 时 显示

标签页名称: 执行调优

添加控件

单行文本框

多行文本框

数值输入框(整数, 浮点数)

复选框

单选下拉

多选下拉

字段选择

表选择

分区选择



5.9 部分算法集说明

算法分类	算法集	解释
分类	逻辑回归分类、随机森林 (Random Forest) 、 梯度提升树 (Gradient-Boosted Trees) 、 贝叶斯方法 (NavieBayes) 等	分类算法通过对已知类别训练集的分析,从中发现分类规则,以此预测新数据的类别。
聚类	Kmeans 算法、混合高斯聚类等	聚类分析是研究分类问题的一种统计分析方法,以相似性为基础,在一个聚类中的模式之间比不在同一聚类中的模式之间具有更多的相似性。
回归	逻辑回归 (LogisticRegression) 、 随机森林回归、GBDT 回归等	回归分析法是在掌握大量观察数据基础上,利用数理统计方法建立因变量与自变量之间的回归关系函数表达式,这种技术通常用于预测分析。
文本分析	LDA、IDF、Word2Vec 等	文本分析是指对文本的表示及其特征项的选取,可以用来识别大规模文档集或语料库中潜藏的主题信息
统计	卡方相关性检验、单变量分析等	根据统计学计算给定数据集的数理统计,是机器学习的基础算法。
深度机器学习	深度神经网络 (Deep Neural Network)、卷积网络 (Convolutional Network) 等	深度机器学习是模仿人脑的机制来解释图像,声音和文本数据,有监督学习与无监督学习之分。
可视化套件	关系网格、箱线图、线状图、散点图等	可视化套件是将输入进行可视化的组件,此外用户可以自定义可视化的组件,支持 各类前端插件。
特征工程	One-hot 编码、规范化、离散化、LabelEncode 等	特征工程的目的是最大限度地从原始数据中提取特征以供算法和模型使用

普惠机器智能 释放人类智慧



从制造业中来，到制造业中去