

目录

| | |
|---------------------------------|----|
| 目录 | 1 |
| 金融风控场景实例 | 3 |
| 示例简介 | 3 |
| 数据准备 | 3 |
| 新建实验 | 3 |
| 可视化建模 | 4 |
| 创建工作流示例 | 4 |
| 5.1 工作流中读入数据示例 | 4 |
| 5.2 数据预处理示例 | 5 |
| 5.3 特征工程示例 | 6 |
| 5.4 分类模型训练示例 | 8 |
| 5.5 模型预测示例 | 9 |
| 5.6 模型评估示例 | 9 |
| 5.7 模型对比优选示例 | 9 |
| 5.7.1 构建LightGBM模型 | 9 |
| 5.7.2 添加模型预测（LightGBM二分类） | 10 |
| 5.7.3 添加模型评估（LightGBM） | 10 |
| 5.7.4 模型比较（GBDT和LightGBM） | 10 |
| 5.8 模型注册示例 | 11 |
| 5.9 评分卡训练示例 | 11 |
| 5.10 将注册好的模型保存到推理pipeline示例 | 12 |
| 5.11 可视化建模完成，选择资源配置，单击画布中的运行按钮。 | 12 |
| 发布为推理服务 | 12 |
| 通过模型管理模块发布推理服务示例 | 12 |
| 通过推理pipeline模块发布推理服务示例 | 13 |
| 视觉—图像分类实例 | 14 |
| 示例简介 | 14 |
| 数据准备 | 15 |
| 训练集准备 | 15 |
| 数据添加 | 15 |
| 数据存储和文件选择 | 15 |
| 数据预览 | 15 |
| 确认详细信息 | 16 |
| 成功示例 | 16 |
| 测试集准备 | 16 |
| 数据添加 | 16 |
| 数据存储和文件选择 | 17 |
| 数据预览 | 17 |
| 确认详细信息 | 17 |
| 成功示例 | 18 |
| 新建实验 | 18 |
| 创建实验 | 18 |
| 创建成功 | 18 |
| 新建自动学习任务 | 18 |
| 新建图片分类训练任务 | 18 |
| 训练任务资源及参数设定 | 19 |
| 训练过程日志 | 20 |
| 模型评估 | 20 |
| 查看模型详情 | 20 |
| 模型评估 | 20 |

金融风控场景实例

示例简介

该场景满足金融信贷风控场景，主要涉及到数据管理、实验管理、工作流、模型管理、推理服务等模块。

该实例使用的数据集为某地信用卡客户贷款违约数据，数据集介绍如下：

| 数据字段 | 字段说明 |
|----------------------------|---|
| ID | 信用卡客户ID号 |
| LIMIT_BAL | 以新台币计算的信贷金额（包括个人和家庭/补充信贷） |
| SEX | 性别（1代表男性，2代表女性） |
| EDUCATION | 受教育程度（1=研究生，2=大学，3=高中，4=其他 5=未知，6=未知） |
| MARRIAGE | 婚姻状况（1=已婚，2=单身，3=其他） |
| AGE | 年龄 |
| PAY_0 | 2005年9月的还款记录（0=按时付款，1=延迟一个月付款，2=延迟两个月付款，8=延迟八个月付款，9=延迟九个月及以上付款） |
| PAY_2 | 2005年8月的还款情况（特征值含义如上） |
| PAY_3 | 2005年7月的还款情况（特征值含义如上） |
| PAY_4 | 2005年6月的还款情况（特征值含义如上） |
| PAY_5 | 2005年5月的还款情况（特征值含义如上） |
| PAY_6 | 2005年4月的还款情况（特征值含义如上） |
| BILL_AMT1 | 2005年9月账单金额（新台币） |
| BILL_AMT2 | 2005年8月账单金额（新台币） |
| BILL_AMT3 | 2005年7月账单金额（新台币） |
| BILL_AMT4 | 2005年6月账单金额（新台币） |
| BILL_AMT5 | 2005年5月账单金额（新台币） |
| BILL_AMT6 | 2005年4月账单金额（新台币） |
| PAY_AMT1 | 2005年9月支付金额（新台币） |
| PAY_AMT2 | 2005年8月支付金额（新台币） |
| PAY_AMT3 | 2005年7月支付金额（新台币） |
| PAY_AMT4 | 2005年6月支付金额（新台币） |
| PAY_AMT5 | 2005年5月支付金额（新台币） |
| PAY_AMT6 | 2005年4月支付金额（新台币） |
| default.payment.next.month | 是否违约还款（1代表是，0代表否） |

该场景涉及到以下几个环节：

- 数据准备**：在数据管理模块中，上传CSV格式的数据集。
- 新建实验**：在实验管理模块中，新建实验。
- 可视化建模**：在工作流画布中，拖拽相应的算子，创建完整的实验。
- 发布为服务**：在推理服务模块中，将生成的模型发布为服务。

注意事项：

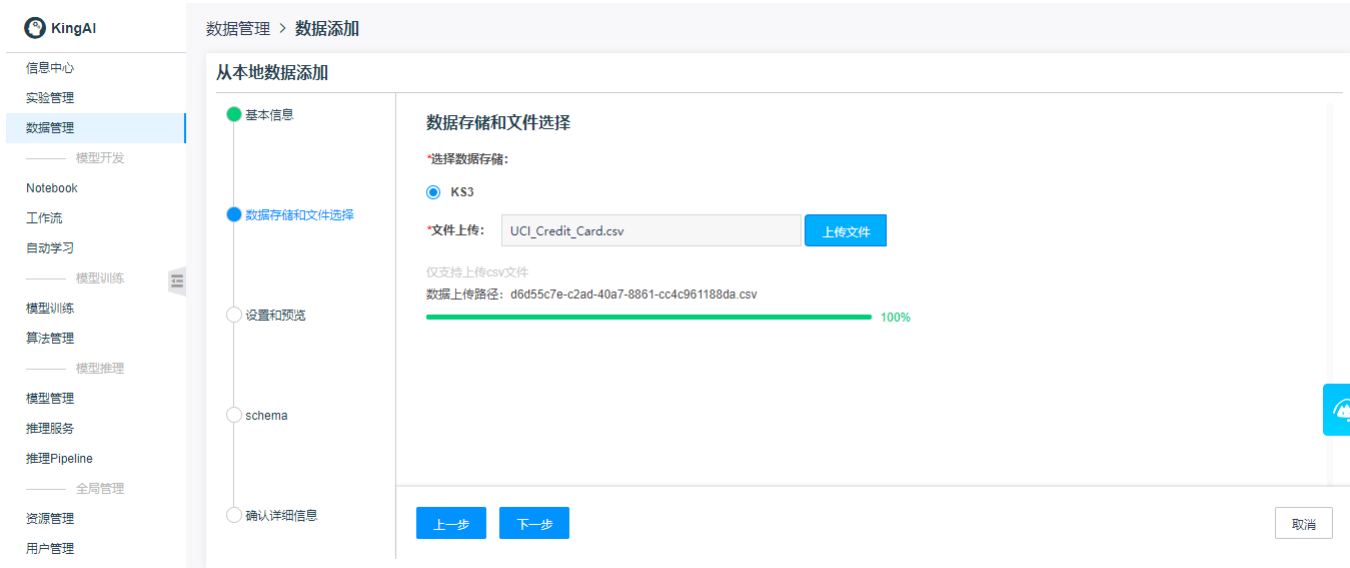
该场景只针对CSV（逗号分隔符）类型的数据文件。

数据准备

导入本地CSV类型的金融信贷数据到数据管理，预览并进行分析。

操作步骤：

- 登录[人工智能KingAI控制台](#)。
- 单击**立即开始**，进入KingAI平台工作区页面，在左侧导航栏中单击**数据管理**。
- 进入数据集管理页面，单击**创建>从本地文件**。
- 按平台引导将某地信用卡客户贷款违约数据上传在平台。



新建实验

操作步骤：

- 登录[人工智能KingAI控制台](#)。
- 单击**立即开始**，进入KingAI平台工作区页面，在左侧导航栏中单击**实验管理**。
- 进入实验集管理页面，单击**新建实验**。
- 填写实验名称，单击**确定**。



可视化建模

创建工作流示例

操作步骤：

1. 登录[人工智能KingAI控制台](#)。
2. 单击立即开始，进入KingAI平台工作区页面，在左侧导航栏中单击工作流。
3. 点击工作流，然后点击新建工作流：
4. 填写相关信息，工作流名需要合法，且不与已存在的工作流重名，选择对应选项，最后点击确认：



5. 在工作流中详细的操作步骤如下：

5.1 工作流中读入数据示例

工作流中多种导入数据方式，此处只展示从工作流自带开源数据集、数据管理、KS3三种：

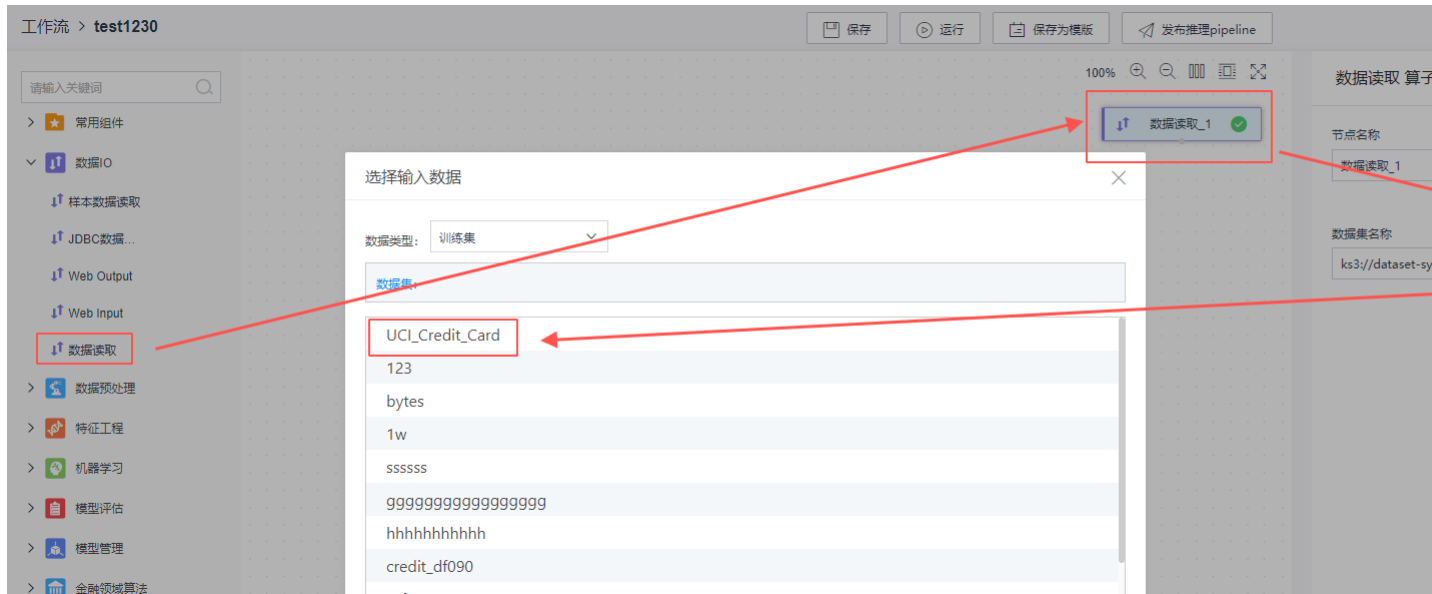
5.1.1. 工作流自带开源数据集：

- 1) 从左侧算子区展开-数据IO-拖动样本数据到空白处
- 2) 右键点击样本数据，点编辑，选择某地信用卡借贷数据



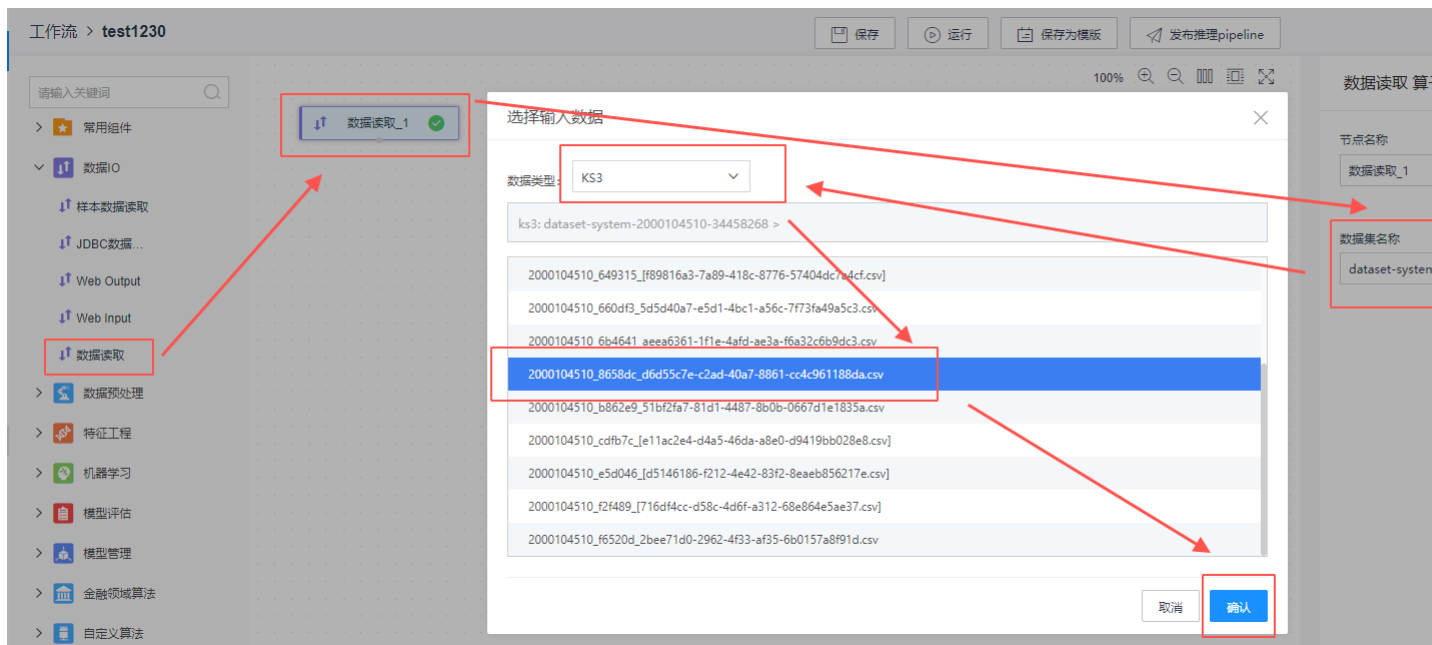
5.1.2. 导入数据管理中的表格数据：

- 1) 从左侧算子区展开-数据IO-拖动数据读取到空白处
- 2) 右键点击数据读取，点编辑，选择数据集名称，数据类型选训练集“UCI_Credit_Card”



5.1.3. 导入KS3中的表格数据：

- 1) 从左侧算子区展开-数据IO-拖动数据读取到空白处
- 2) 右键点击数据读取，点编辑，选择数据集名称，数据类型选KS3，然后选择对应的文件



5.2 数据预处理示例

5.2.1. 数据列筛选

- 1) 除id外其他列全选



5.2.2. 分层拆分（随机拆分类似）

1) 选default.t.payment.next.month为标签列，拆分比例0.8

工作流 > test1230

保存 运行 保存为模版 发布推理pipeline

请输入关键词

常用组件

数据IO

数据预处理

数据拆分

随机拆分数据

分层拆分数据

数据采样

单一值过滤

异常值检测

缺失值填充

数据类型转换

数据映射

异常字符处理

数据列筛选

数据读取_1

随机拆分数据_1

分层拆分数据_1

节点名称

分层拆分数据_1

选择数据的标签列

default.payme

分层拆分比例

0.8

数据预览

| 字段数 | 行数 |
|-----|-------|
| 25 | 23999 |

数据表

| ID | LIMIT_BAL | SEX | EDUCATION | MARRIAGE | AGE |
|-------|-----------|-----|-----------|----------|-----|
| 4557 | 150000 | 1 | 1 | 1 | 33 |
| 19808 | 60000 | 2 | 2 | 1 | 37 |
| 29827 | 100000 | 1 | 3 | 1 | 49 |
| 14839 | 120000 | 2 | 1 | 1 | 37 |

5.2.3. 分层采样

1) 选default.t.payment.next.month为标签列，采样比例0.7

2) 除分层采样外，工作流还支持随机采样等四种采样算子，使用方式类似

常用组件

数据IO

数据预处理

数据拆分

数据采样

按比例采样

分层采样

按条数采样

数据上采样

数据下采样

单一值过滤

异常值检测

缺失值填充

数据类型转换

数据映射

数据读取_1

随机拆分数据_1

分层拆分数据_1

分层采样_1

节点名称

分层采样_1

选择数据的标签列

default.payment

分层采样比例

0.7

数据预览

| 字段数 | 行数 |
|-----|-------|
| 25 | 20999 |

数据表

| ID | LIMIT_BAL | SEX | EDUCATION | MARRIAGE | AGE |
|------|-----------|-----|-----------|----------|-----|
| 3214 | 420000 | 1 | 2 | 2 | 28 |
| 5309 | 80000 | 1 | 1 | 2 | 35 |

5.3 特征工程示例

5.3.1. 主成分分析

1) 降维的特征列全选

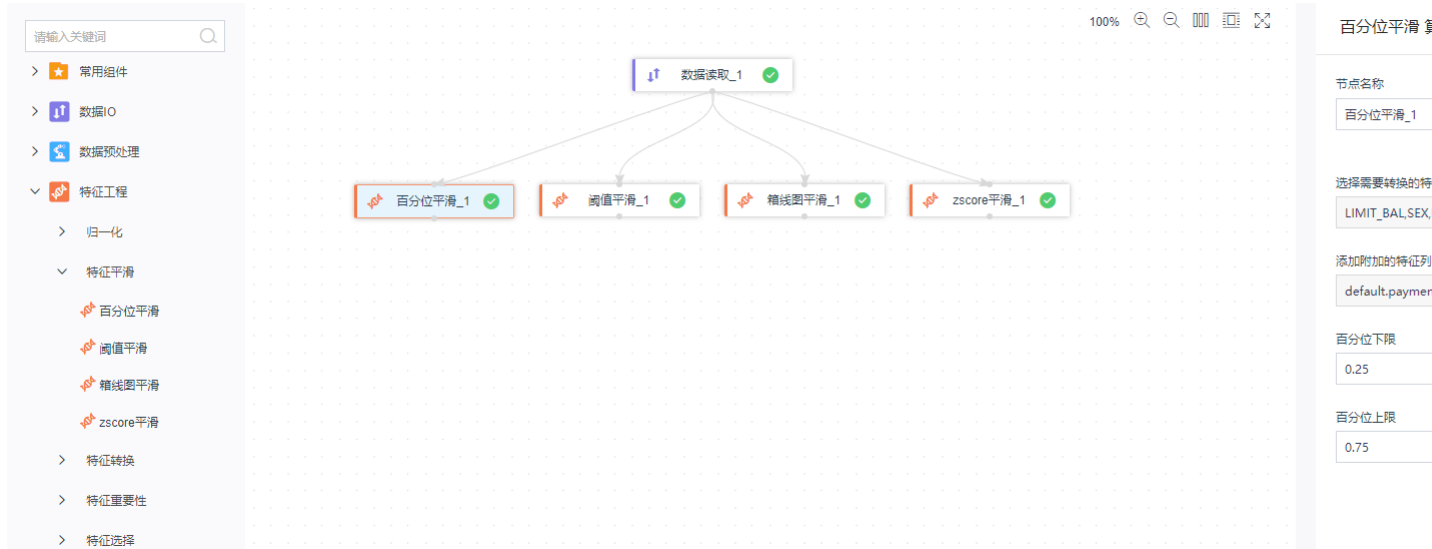
2) 附加的数据列选default.t.payment.next.month



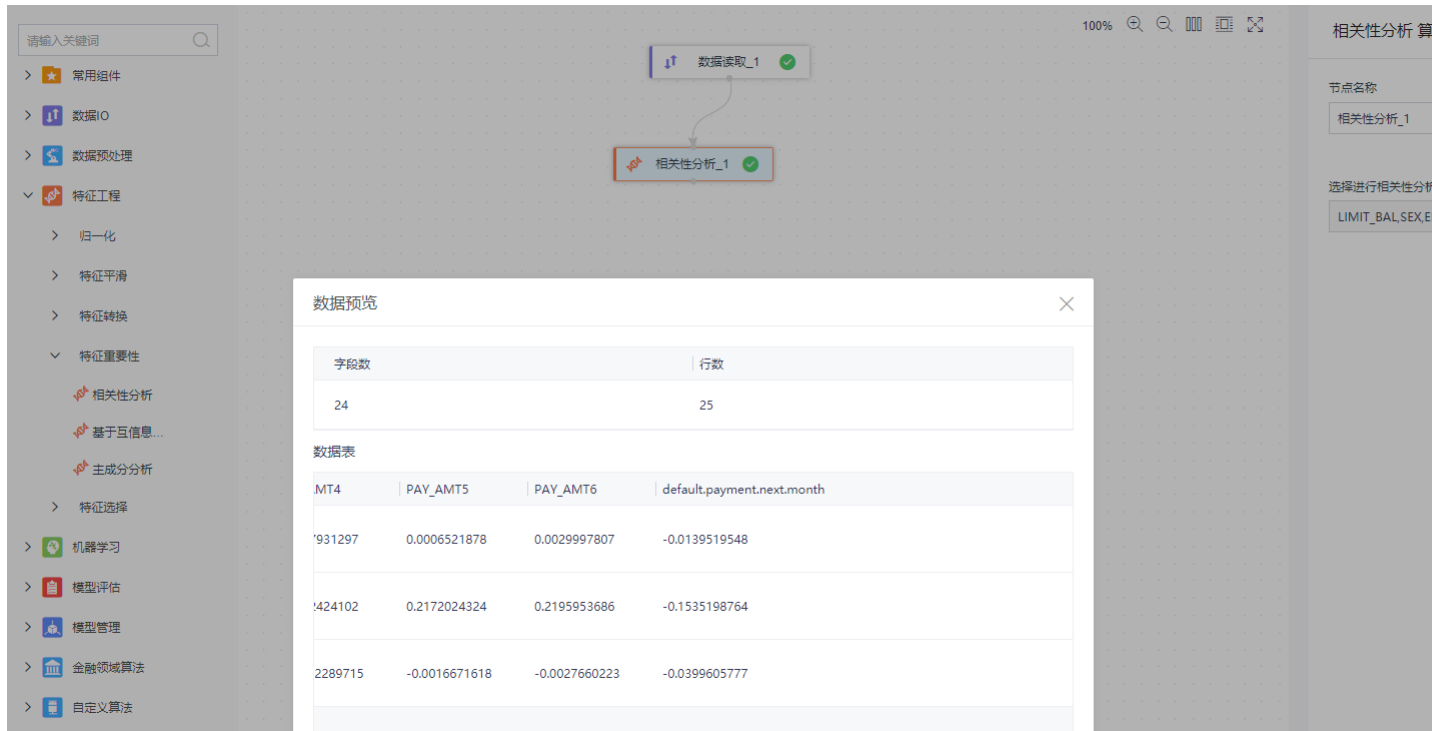
- 5.3.2. 归一化（最大最小归一化、标准归一化两者使用方法基本一致）
- 1) 归一化的特征列选择除ID和default.t.payment.next.month外的所有列
 - 2) 附加的数据列选default.t.payment.next.month



- 5.3.3. 百分位平滑
- 1) 百分位平滑的特征列选择除ID和default.t.payment.next.month外的所有列
 - 2) 附加的数据列选default.t.payment.next.month
 - 3) 另外工作流还提供阈值平滑、箱线图平滑、zscore平滑，使用方法类似

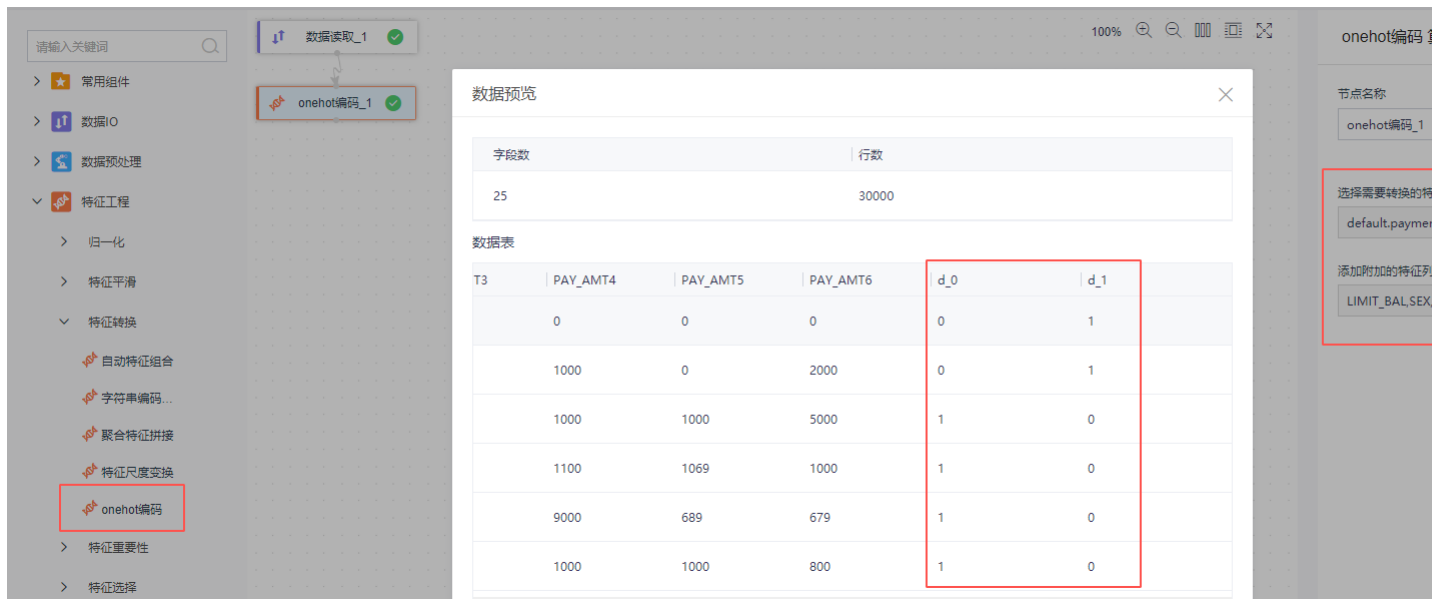


- 5.3.4. 相关性分析
- 1) 相关性分析的特征列选择除ID和default.t.payment.next.month外的所有列
 - 2) 附加的数据列选default.t.payment.next.month
 - 3) 另外工作流还提供阈值平滑、箱线图平滑、zscore平滑，使用方法类似



5.3.5. onehot编码

- 1) 对default t. payment. next. month列进行onehot编码
- 2) 附加的数据列选default t. payment. next. month外的其他列



5.4 分类模型训练示例

选择GBDT二分类算法算子



训练的特征列除default. payment. next. mont外全选

训练的标签选default.payment.next.month

5.5 模型预测示例

GBDT二分类算法模型预测

保存

运行

保存为模板

发布推理pipeline

数据IO

数据预处理

特征工程

机器学习

模型评估

聚类模型...

分类模型...

多分类模型...

模型预测

回归模型...

台湾信用卡借贷...

数据列筛选

主成分分析

随机拆分数据

GBDT二分类

模型预测

模型预测 算子参数设置

节点名称

模型预测

选择特征列表

dimension0,dimension1,dimension2,dimension3,dim

选择模型的标签,添加该标签有利于后续评估模型

default,payment.next.month

特征列表除default.payment.next.month外全选

模型标签选择default.payment.next.month

5.6 模型评估示例

分类模型评估

保存

运行

保存为模板

发布推理pipeline

请输入关键词

常用组件

数据IO

数据预处理

特征工程

机器学习

模型评估

聚类模型...

分类模型...

多分类模型...

模型预测

回归模型...

台湾信用卡借贷...

数据列筛选

主成分分析

随机拆分数据

GBDT二分类

模型预测

分类模型评估

分类模型评估 算子参数设置

节点名称

分类模型评估

选择正标签值

1

5.7 模型对比优选示例

5.7.1 构建LightGBM模型

添加LightGBM二分类算法

保存

运行

保存为模板

发布推理pipeline

机器学习

回归算法

多分类算法

二分类算法

随机森林...

逻辑回归

lightgbm...

GBDT二...

xgboost二...

聚类算法

模型融合

模型评估

台湾信用卡借贷...

数据列筛选

主成分分析

随机拆分数据

模型预测

lightgbm二分类

模型预测

模型评估

分类模型评估

lightgbm二分类 算子参数设置

节点名称

lightgbm二分类

基学习器选择

gbdt

特征采样率

1

选择训练的特征列

dimension0,dimension1,dimension2,dimension3,dim

选择训练的标签

default,payment.next.month

特征列表除default.payment.next.month外全选

模型标签选择default.payment.next.month

5.7.2 添加模型预测（LightGBM二分类）

保存

运行

保存为模版

发布推理pipeline

数据IO

数据预处理

特征工程

机器学习

模型评估

聚类模型...

分类模型...

多分类模...

模型预测

回归模型...

台湾信用卡债...

数据列筛选

主成分分析

随机拆分数据

模型预测

lightgbm二分类

模型预测

模型预测 算子参数设置

节点名称

模型预测

选择特征列表

dimension0,dimension1,dimension2,dimension3,dim

选择模型的标签,添加该标签有利于后续评估模型

default.payment.next.month

特征列表除default.payment.next.month外全选
模型标签选择default.payment.next.month

5.7.3 添加模型评估（LightGBM）

保存

运行

保存为模版

发布推理pipeline

请输入关键词

常用组件

数据IO

数据预处理

特征工程

机器学习

模型评估

聚类模型...

分类模型...

多分类模...

模型预测

回归模型...

台湾信用卡债...

数据列筛选

主成分分析

随机拆分数据

模型预测

lightgbm二分类

模型预测

模型评估

分类模型评估

分类模型评估 算子参数设置

节点名称

分类模型评估

选择正标签值

1

5.7.4 模型比较（GBDT和LightGBM）

右键点击分类模型评估，查看模型报告

主成分分析

随机拆分数据

GBDT二分类

lightgbm二分类

模型预测

分类模型评估

模型注册

编辑

删除

查看模型报告

运行

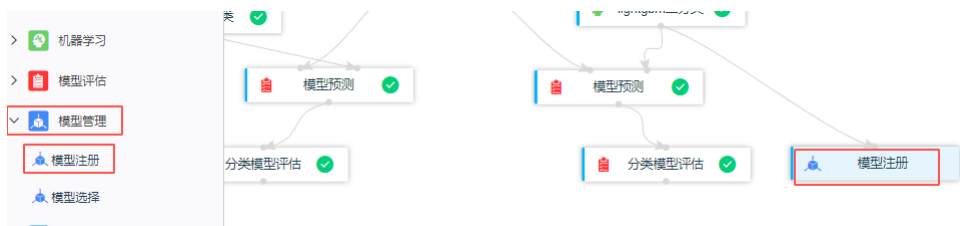
小数据量运行

分类模型评



5.8 模型注册示例

效果更好的模型2注册并准备发布及部署



5.9 评分卡训练示例

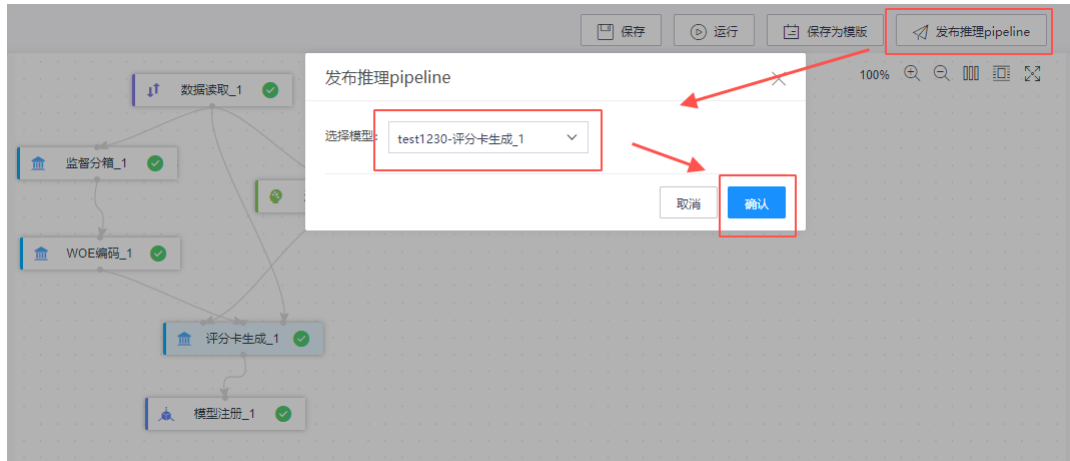
评分卡训练步骤:

1. 导入数据训练逻辑回归模型，特征列选择除ID和default.t.payment.next.month外的所有列；
2. 另外将导入的数据进行监督分箱然后将分箱结果进行woe编码；
3. 将woe编码结果、逻辑回归模型、原始数据输入到评分卡算子；
4. 进行模型注册



5.10 将注册好的模型保存到推理pipeline示例

将上述步骤中的评分卡模型发布到推理pipeline



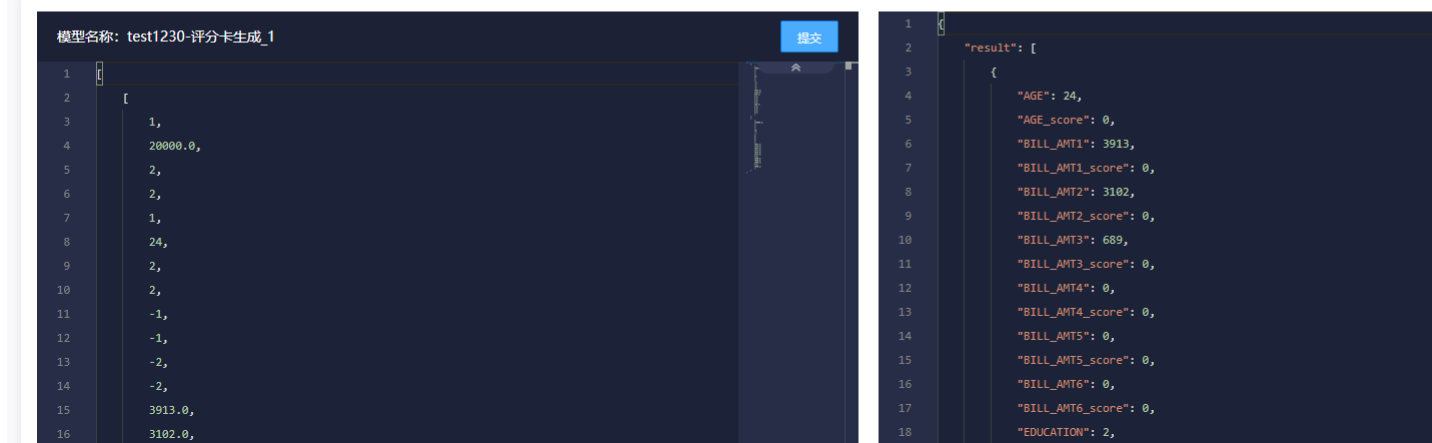
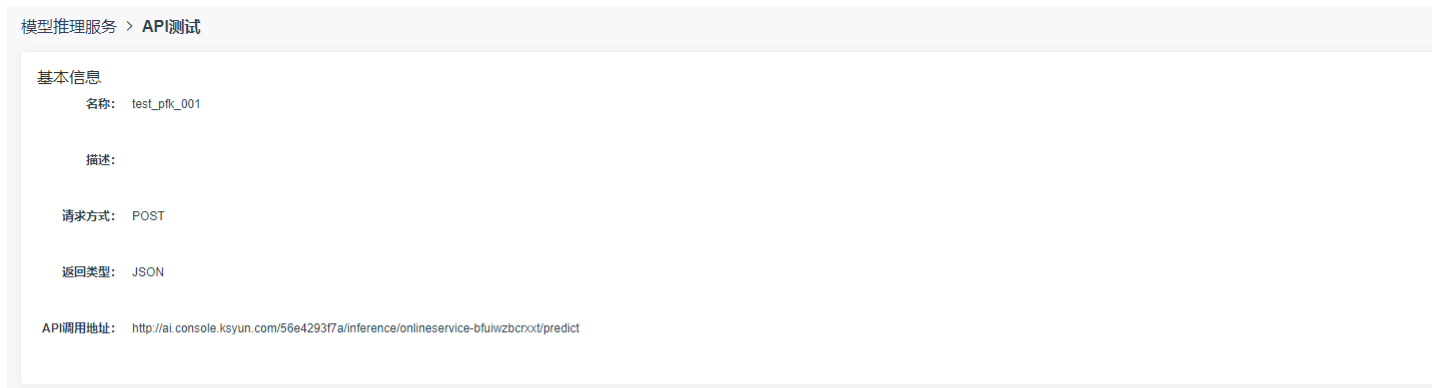
5.11 可视化建模完成，选择资源配置，单击画布中的运行按钮。

发布为推理服务

将金融信贷数据的工作流生成的评分卡模型发布为推理服务。

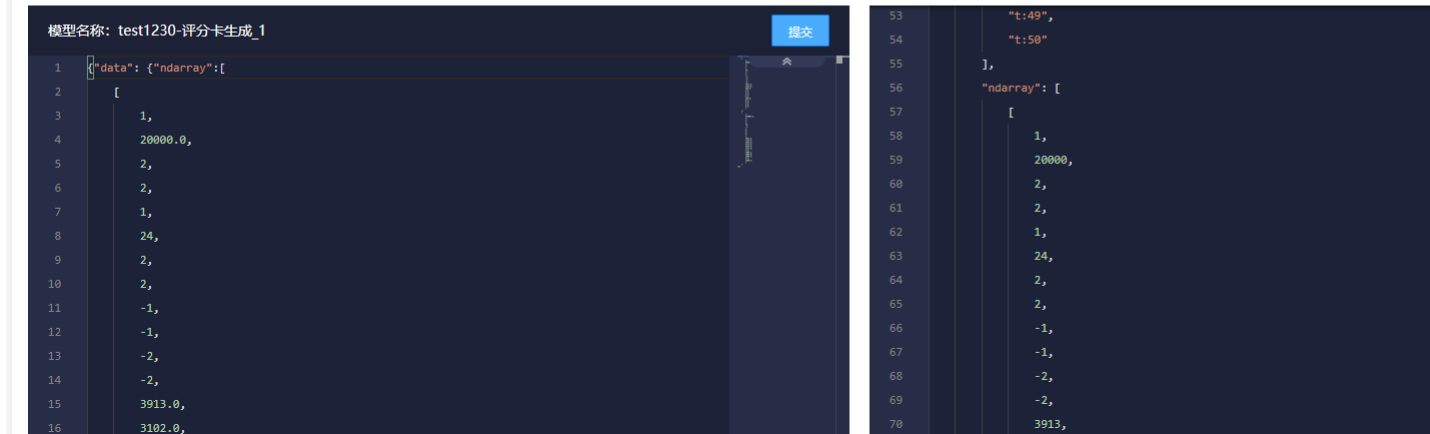
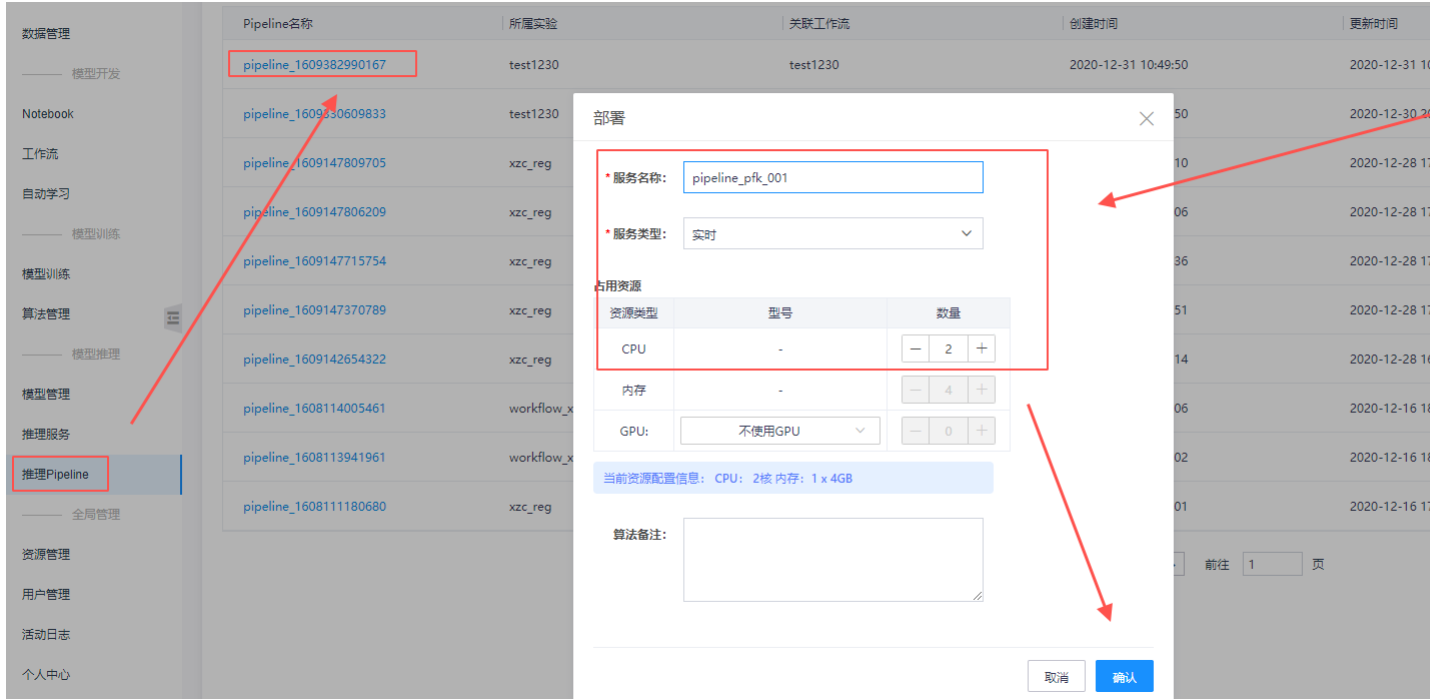
通过模型管理模块发布推理服务示例

将评分卡模型通过模型管理模块发布成推理服务



通过推理pipeline模块发布推理服务示例

将评分卡模型推理Pipeline模块发布成推理服务



视觉—图像分类实例

示例简介

在KingAI平台中，导入图片数据，完成图像分类算法，实现模型推理服务和测试。

数据集介绍

数据集为识别图片为跑步或者行走的数据集，训练集图片500张，测试集图片141张。

| 数据集分类 | 类别 | 数量 | 图片格式 |
|-------|------|-----|------|
| 训练集 | Run | 299 | Png |
| 训练集 | Walk | 301 | Png |
| 测试集 | Run | 82 | Png |
| 测试集 | Walk | 59 | Png |

涉及环节

该场景涉及到一下几个环节：

1. **数据准备**：在数据管理模块中，上传图片类数据集，包含训练集集和测试集。
2. **新建实验**：在实验管理模块中，新建实验。

3. 模型训练：在自动学习模块中，选择图片分类模板，创建模型训练任务。
4. 发布为服务：在推理服务模块中，将生成的模型发布为服务。

数据准备

训练集准备

数据添加

KingAI

信息中心

数据管理

实验列表

模型开发

模型开发

工作流

模型训练

模型训练

算法管理

模型推理

模型管理

推理服务

数据管理 > 数据添加

从对象存储添加

● 基本信息

○ 数据存储和文件选择

○ 数据预览

○ 确认详细信息

基本信息

*数据集名称:

Image_run_vs_walk_train_set

*数据集类型:

训练集

*数据类型:

图片

备注:

图像 (走/跑) 分类的训练集

数据存储和文件选择

KingAI

信息中心

数据管理

实验列表

模型开发

模型开发

工作流

模型训练

模型训练

算法管理

模型推理

模型管理

推理服务

数据管理 > 数据添加

从对象存储添加

● 基本信息

● 数据存储和文件选择

○ 数据预览

○ 确认详细信息

数据存储和文件选择

*选择数据:

xl-test/walk_or_run_train_for_ws/

标注文件:

xl-test/walk_or_run_train_for_ws/label_train.js

*标注类型:

图像分类

数据预览

KingAI

信息中心

数据管理

实验列表

模型开发

模型开发

工作流

模型训练

模型训练

算法管理

模型推理

模型管理

推理服务

推理Pipeline

全局管理

资源管理

活动日志

数据管理 > 数据添加

从对象存储添加

基本信息

数据存储和文件选择

数据预览

确认详细信息

数据预览

数据标签: 跑步 行人

标注情况: 已标注: 599 未标注: 0

确认详细信息

KingAI

信息中心

数据管理

实验列表

模型开发

模型开发

工作流

模型训练

模型训练

算法管理

模型推理

模型管理

数据管理 > 数据添加

从对象存储添加

基本信息

数据存储和文件选择

数据预览

确认详细信息

确认详细信息

基本信息

名称: Image_run_vs_walk_train_set

数据集类型: 训练集

数据类型: 图片

标注类型: 图像分类

备注: 图像(走/跑)分类训练集

数据存储和文件选择

数据存储: ks3

路径: xl-test/walk_or_run_train_for_ws/

成功示例

KingAI

信息中心

数据管理

实验列表

模型开发

模型开发

数据管理

数据集管理

数据源管理

创建 +

| 数据集名称 | 数据集类型 | 类型 | 创建人 | 创建时间 | 更新时间 |
|-----------------------------|-------|----|----------------|---------------------|---------------------|
| Image_run_vs_walk_train_set | 训练集 | 图片 | kdl_testuser01 | 2020-10-27 20:01:43 | 2020-10-27 20:01:43 |

测试集准备

数据添加

金山云

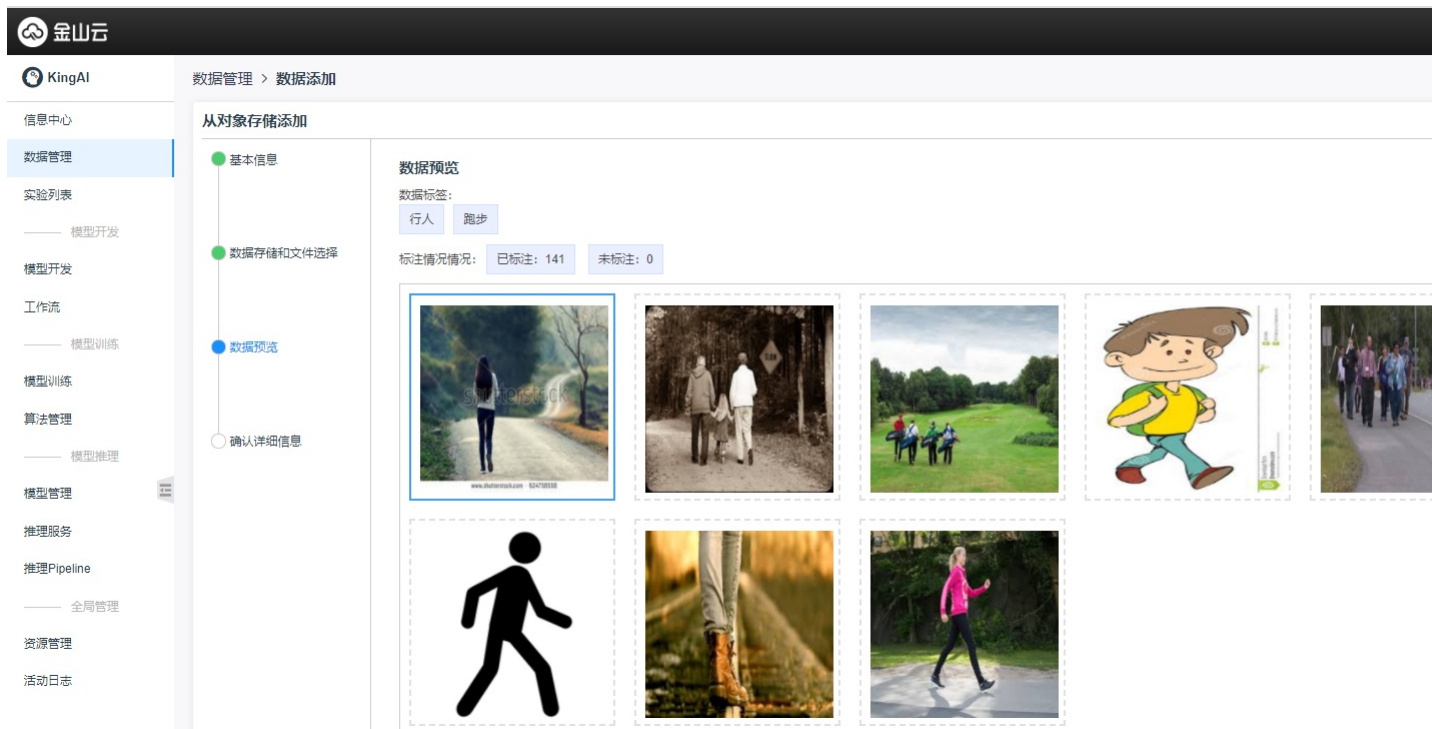
16/20



数据储存和文件选择



数据预览



确认详细信息

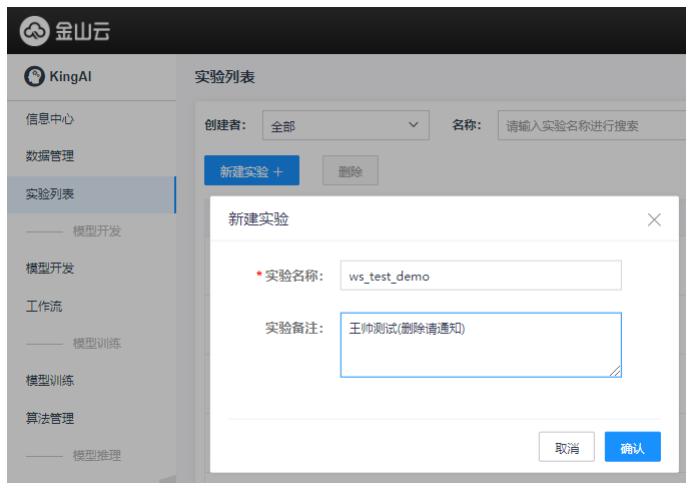


成功示例

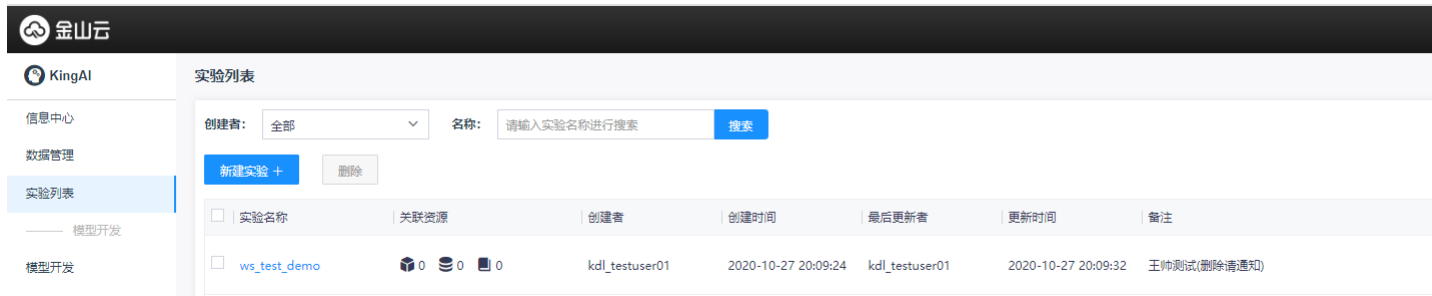


新建实验

创建实验



创建成功



新建自动学习任务

新建图片分类训练任务

KingAI

信息中心

实验管理

数据管理

模型开发

容器开发

可视化建模

自动学习

模型训练

模型训练

算法管理

模型推理

模型管理

推理服务

推理Pipeline

全局管理

资源管理

用户管理

活动日志

个人中心

自动学习

图片分类

帮助您更快的识别一张图片中是否包含某种物体

创建模板

目标检测

帮助您更快的识别出图片中每个物体的位置及类别

创建模板

| 任务名称 | 任务类型 | 状态 | 创建时间 | 更新时间 | 备注 |
|----------------------------|------|------|---------------------|---------------------|------|
| ffffrfrfr | 图像分类 | 训练失败 | 2021-01-15 19:08:35 | 2021-01-15 19:09:50 | rrrr |
| zidongxuexi | 图像分类 | 训练失败 | 2021-01-15 17:13:46 | 2021-01-15 17:14:58 | |
| jzk_detect_monitor_gpu_p40 | 目标检测 | 训练失败 | 2021-01-12 19:40:19 | 2021-01-12 19:43:19 | |
| jzk_detect_gpu_titan_again | 目标检测 | 训练完成 | 2021-01-12 14:15:52 | 2021-01-12 15:34:08 | |
| jzk_detect_gpu_titan | 目标检测 | 训练完成 | 2021-01-12 12:23:21 | 2021-01-12 13:37:59 | |
| jzk_detect_monitor_gpu | 目标检测 | 训练完成 | 2021-01-12 11:19:32 | 2021-01-12 12:24:25 | |
| jzk_test_gpu_monitor | 目标检测 | 训练完成 | 2021-01-11 19:55:26 | 2021-01-11 21:53:47 | |
| ddd | 图像分类 | 训练完成 | 2021-01-11 18:50:45 | 2021-01-11 19:08:47 | ffff |
| test_log | 图像分类 | 训练完成 | 2021-01-11 16:17:01 | 2021-01-11 16:33:18 | |
| dm auto face | 目标检测 | 训练完成 | 2021-01-11 15:29:09 | 2021-01-11 16:18:37 | fff |

KingAI

信息中心

实验管理

数据管理

模型开发

容器开发

可视化建模

自动学习

模型训练

模型训练

算法管理

模型推理

模型管理

推理服务

推理Pipeline

全局管理

资源管理

用户管理

活动日志

个人中心

训练任务资源及参数设定

训练过程日志

模型评估

[查看模型详情](#)

在模型管理页面，单击生成的模型名称，查看模型详情

模型评估

在模型管理页面，选择该条模型，单击操作列中的更多>模型评估

金山云