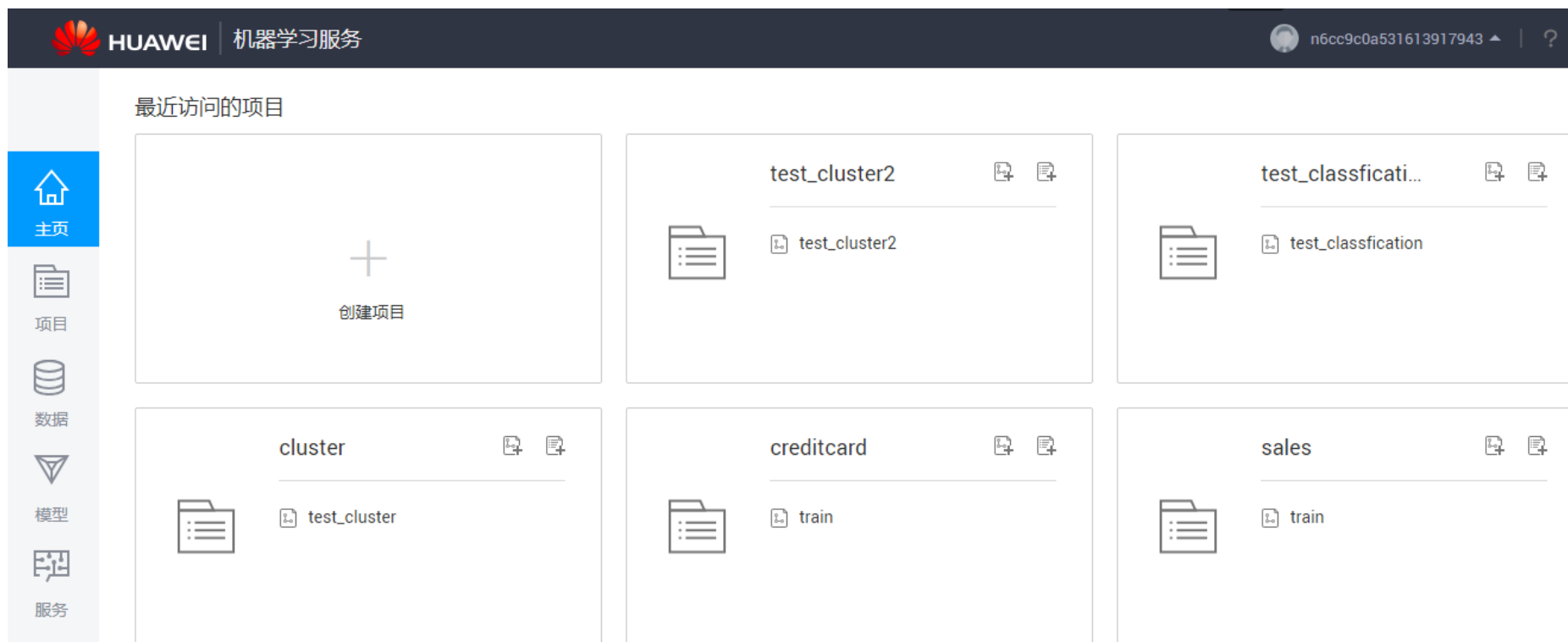


机器学习 试验

第一天

登录华为机器学习平台

<https://mls.cn-north-1.huaweicloud.com:22360/mls/index.html?inst=9a6465d2-c3e5-4006-b93d-1e0ef0ea5622®ion=cn-north-1&locale=zh-cn#/login>



这里 可以创建不同的项目，建立项目后可以在项目里创建 workflows


登录后数据 那里可以上传 项目的数据文件


 HUAWEI

机器学习服务


 n6cc9c0a531613917943

?

 主页

 项目

 数据

 模型

 服务

模型

名称

PMML1_result

PMML2_result

inner1

pmml1


salesFore

模型版本列表


通过版本号搜索

版本号	格式	修改时间	操作
20181229_225108_886	pmml	2018/12/29 22:51:08 GMT+08:...	打包并下载 删除 下载

点击 模型菜单，可以下载 生成的模型

 HUAWEI

机器学习服务

 n6cc9c0a531613917943 | ?

文件系统

DATASETS

上传文件

创建文件夹









通过名称搜索

通过名称搜索

通过文件夹名称搜索

DATASETS

- class
- classification
- cluster
- cluster1
- cluster2
- feature
- samples
- solution

名称	修改时间	大小	操作
 car_price2.csv	2018/12/28 19:34:13 GMT+08:00	35 KB	下载 删除
 class	-	0 KB	删除
 classification	-	0 KB	删除
 cluster	-	0 KB	删除
 cluster1	-	0 KB	删除
 cluster2	-	0 KB	删除
 feature	-	0 KB	删除
 samples	-	0 KB	删除

可以参见不同的 文件夹，有点像 OBS 的使用

节点库：

项目 > 工作流 > huiguitest

节点库


请输入关键字





[-] 数据转换


[-] 记录操作


 聚合

 去重


 过滤

 连接


 抽样

 排序


 拆分

 时序抽取

 追加


 执行SQL脚本

[+] 字段操作


 执行SQL脚本


■ 字段操作

 离散化


 二值化

 派生

 缺失值填充


 标准化

 重命名

 替换

 选择

 重建


 修改元数据


建模


- + 特征
- + 异常检测
- + 分类
- + 聚类
- + 回归
- + 推荐
- + 频繁模式挖掘
- + 关系分析
- + 时空分析
- + 文本分析
- + 时序分析


■ 建模

■ 特征


 卡方检验


 相关性分析

 N元语法

 主成分分析


■ 异常检测

 孤立森林


 基于PCA的异常检测


■ 分类


基于PCA的异常检测


 决策树分类


 梯度提升树分类


 K最近邻分类


 逻辑回归

 朴素贝叶斯


 神经网络多分类

 随机决策森林分类


 支持向量机


 XGBoost分类


■ 聚类


 K-均值


■ 回归


 决策树回归

 梯度提升树回归


 K最近邻回归


 线性回归

 随机决策森林回归


 XGBoost回归


■ 推荐

 域分解机

 交替最小二乘


■ 频繁模式挖掘


 Apriori

 FP-Growth




 PrefixSpan

■ 关系分析

 网页排名

 标签传播

■ 时空分析


-  轨迹点清洗
-  轨迹异常检测
-  轨迹切分


■ 文本分析


-  词频
-  文本断句
-  去除停用词
-  隐含狄利克雷分布
-  分词
-  繁简转换
-  关键词提取
-  命名实体识别
-  文本词向量
-  词频-逆文档频率
-  短文本相似度
-  文本规范化
-  文本推荐
-  自动摘要
-  HTML标签过滤


 ARIMA


[-] 评估

 模型应用


 分类模型评估比较


 分类模型评估


 回归模型评估比较

 回归模型评估

[-] 输出

 保存数据到数据集

 保存模型

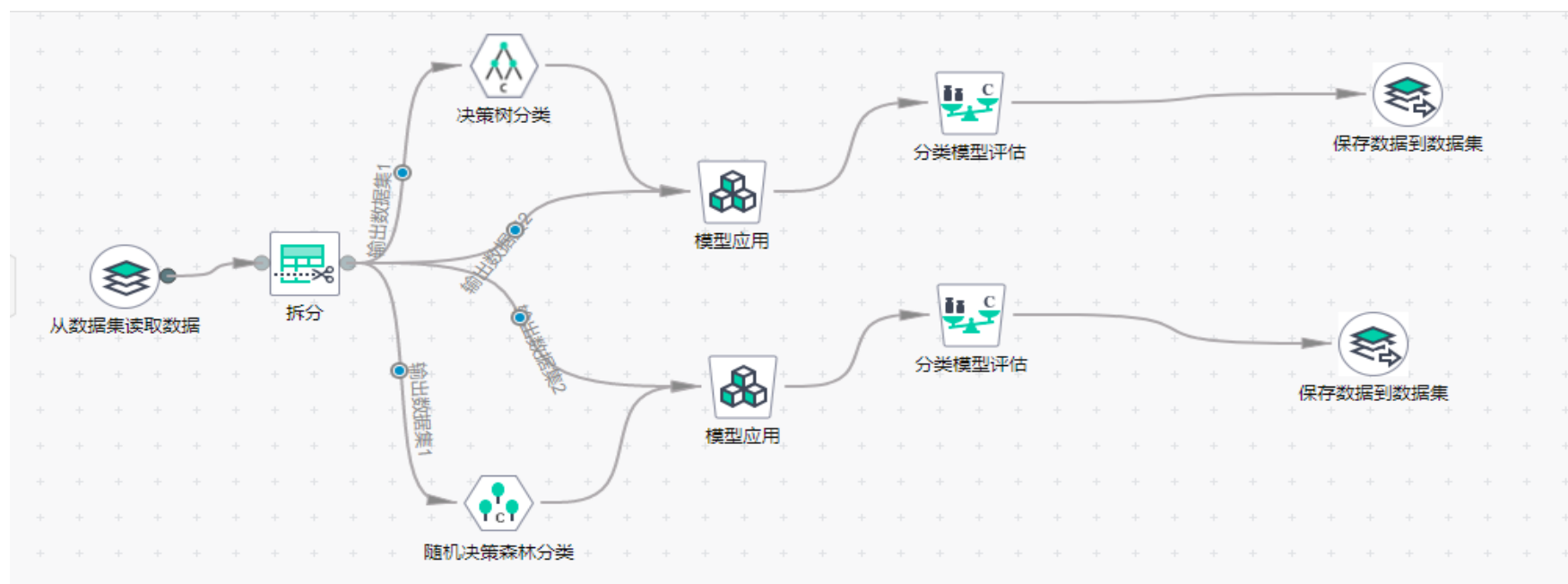
 保存数据到OBS

第一天的试验

使用数据某银行客户数据。

数据地址：https://obs-mlsclass7.obs.cn-north-1.myhuaweicloud.com/new_bank.csv

元数据地址：https://obs-mlsclass7.obs.cn-north-1.myhuaweicloud.com/bank_formal.desc



为了怕忘了具体设置,

每个节点的配置如下:

1) “从数据集读取数据”: 数据文件地址: /classification/new_bank.csv

元数据文件地址: /classification/bank_formal.desc

从数据集读取数据

* 数据格式:

* 数据文件:

导入元数据:

☒

* 元数据文件:

导入元数据 需要 钩上

2) “拆分”: 默认设置

3) “决策树分类”: 默认设置

4) 两个“模型应用”：预测类型：分类

5) “随机决策森林分类”：

随机决策森林分类	* 最大分箱数:
	<input type="text" value="200"/>
<hr/>	
* 树的数目:	* 不纯度:
<input type="text" value="100"/>	<input type="text" value="Entropy"/>
<hr/>	
* 最大树深度:	* 特征子集选取策略:
<input type="text" value="4"/>	<input type="text" value="Auto"/>
<hr/>	
* 最大分箱数:	随机种子:
<input type="text" value="200"/>	<input type="text" value="0"/>

6) 与“随机决策森林分类”分支相连的“保存数据到数据集”

RF_Reslut

7) 与“决策树分类”分支相连的“保存数据到数据集”：

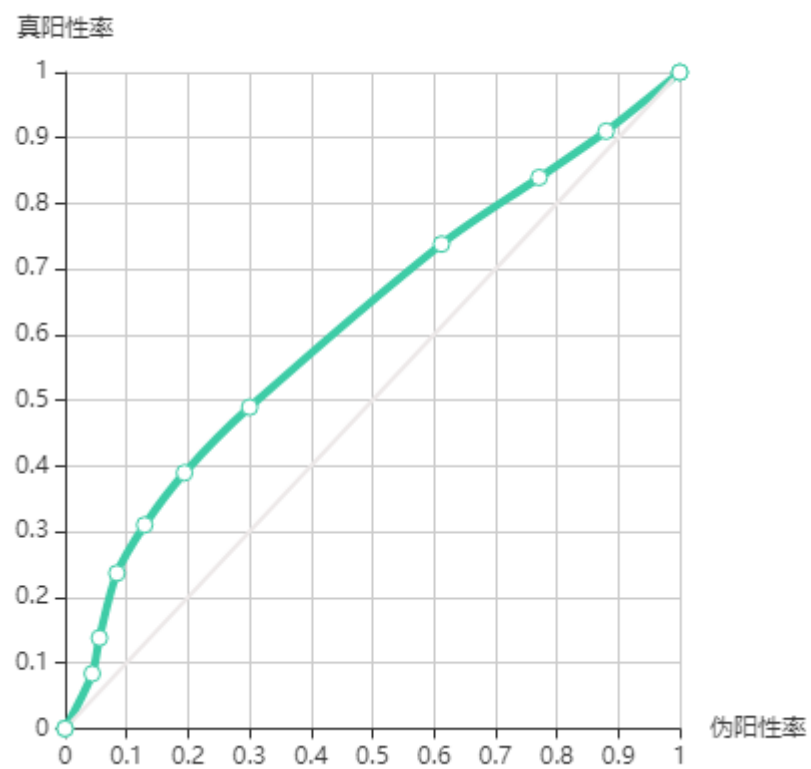
查看评估结果

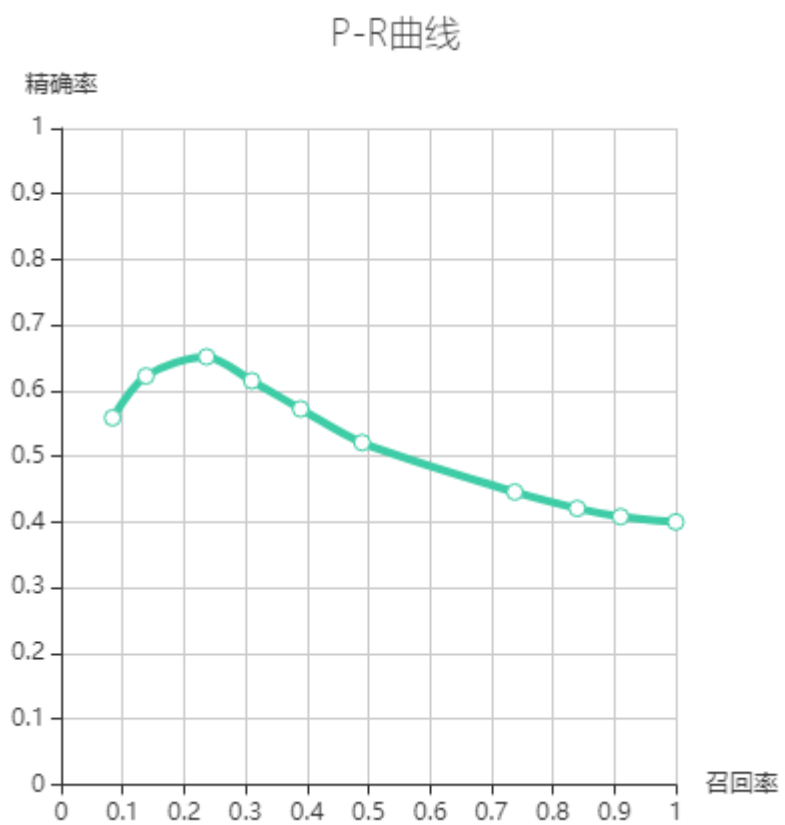
真阳性 265	伪阴性 415	准确率 0.639	精确率 0.572
伪阳性 198	真阴性 821	召回率 0.390	F1 测量 0.464
曲线下面积(AUC) 0.617			

阈值  0.5



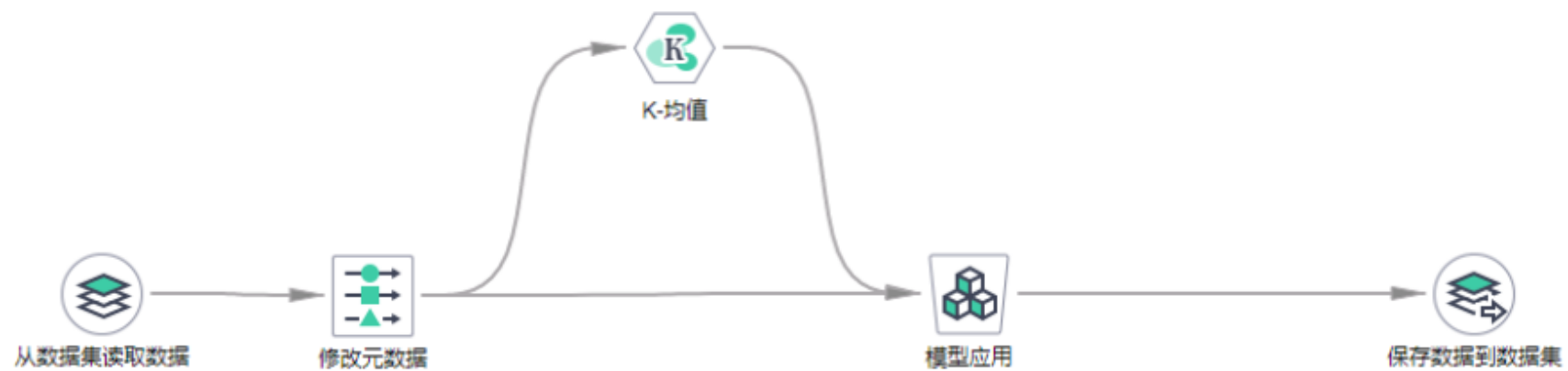
ROC曲线





第 2 天试验

数据地址: <https://obs-mlsclass7.obs.cn-north-1.myhuaweicloud.com/customer.csv>



2) “**修改元数据**”：进行配置时，点击  ，然后将特征“id”的角色改为“None”

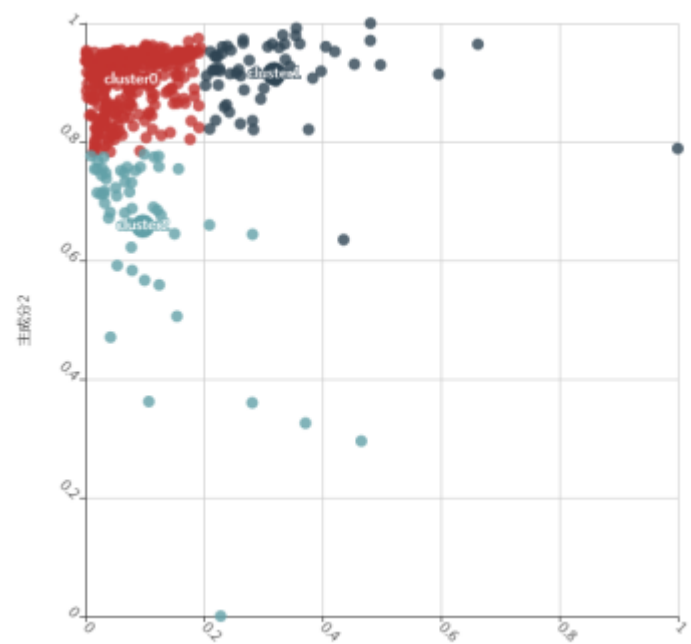
修改元数据

设置元数据



workflows run.

On the right-click menu for the “k-means” node, click “View Model”, to visualize the clustering model.





主页



项目



DATASETS/ cluster/ kmeans_result

上传文件

创建文件夹

通过文件夹名称搜索



DATASETS

classification

名称

修改时间

kmeans_result.csv

2018/11/29 1

meta.desc

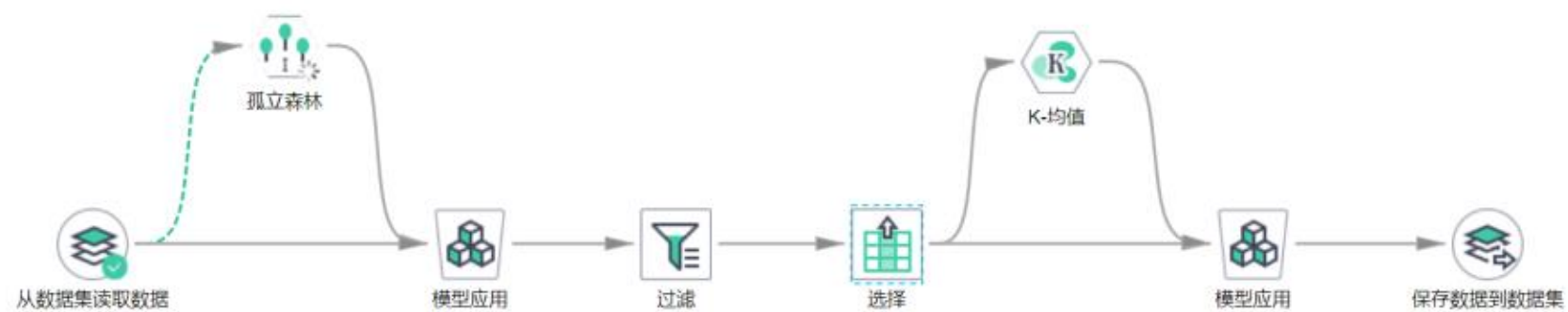
2018/11/29 1

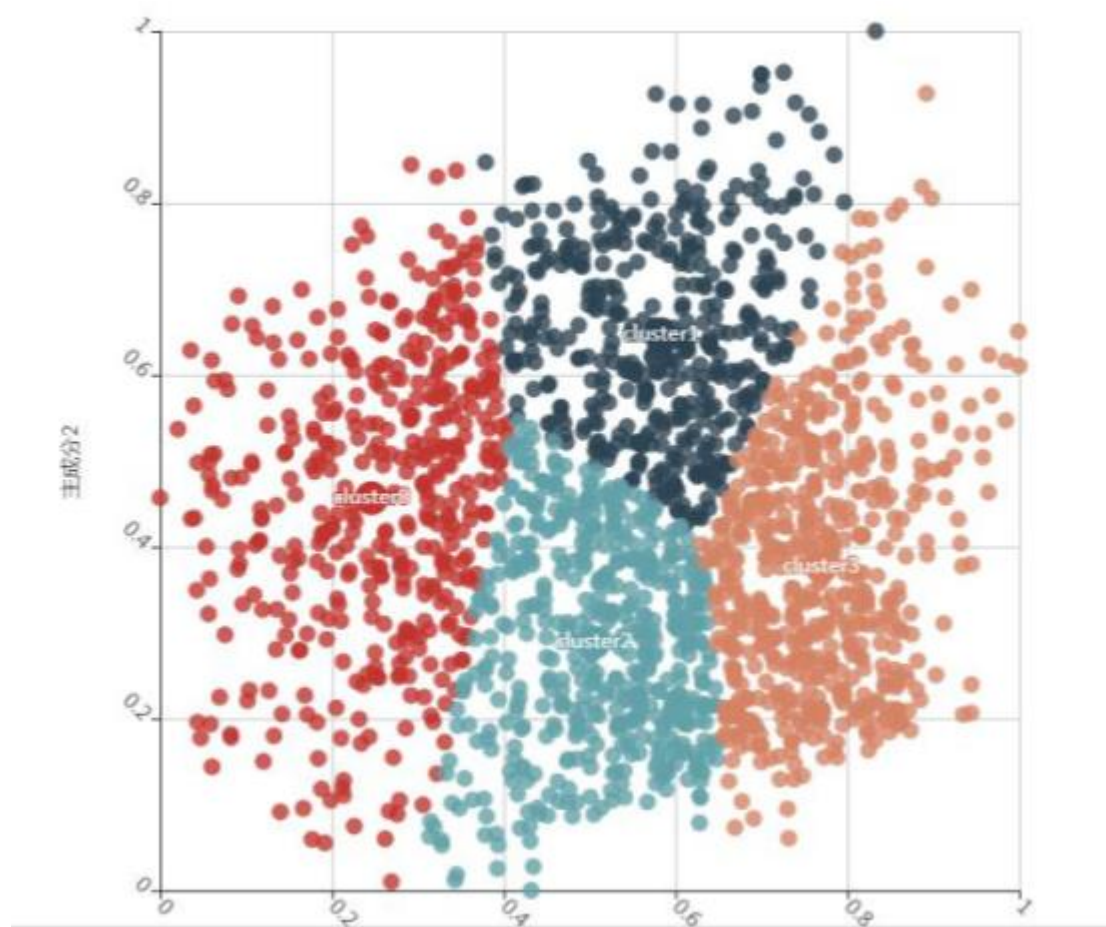
数据预览

```
16,1,3,10253,1114,3821,397,964,412,2  
231,2,1,11072,5989,5615,8321,955,2137,2  
396,1,3,9602,1316,1263,2921,841,290,2  
226,1,1,12680,3243,4157,660,761,786,2  
383,1,3,34454,7435,8469,2540,1711,2893,1  
272,1,1,2083,5007,1563,1120,147,1550,2
```

这两个样本被聚集到了不同的簇当中。

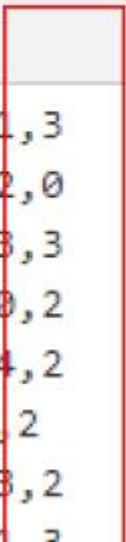
第 3 天 试验





类结果列。

数据预览



```
115.79,138.0,265,61.7,547.1,3
124.74,140.6,285,67.5,527.2,0
121.45,137.6,278,64.3,548.3,3
110.22,139.1,252,59.1,541.0,2
106.63,136.5,245,56.0,543.4,2
106.5,138.4,243,56.9,554.6,2
106.73,138.4,245,56.9,553.3,2
112.57,138.4,258,60.1,530.1,3
```

自己折腾 的这里是什么问题？



运行日志

2019/03/03 23:18:22 GMT+08:00 节点[拆分]开始运行...

2019/03/03 23:18:27 GMT+08:00 节点[拆分]运行成功。

2019/03/03 23:18:27 GMT+08:00 节点[随机决策森林回归]开始运行...

2019/03/03 23:18:32 GMT+08:00 节点[随机决策森林回归]运行失败。

2019/03/03 23:18:32 GMT+08:00 Label measurement type must be continuous for regression model.

2019/03/03 23:18:32 GMT+08:00 工作流运行失败。

- 2019/03/03 23:18:32 GMT+08:00 Label measurement type must be continuous for regression model.
- 2019/03/03 23:18:32 GMT+08:00 工作流运行失败。

