3.2 机器学习中的分类问题（上）

-机器学习服务操作指导

1 任务介绍

本次任务将介绍如何使用决策树和随机森林进行分类

2 任务执行

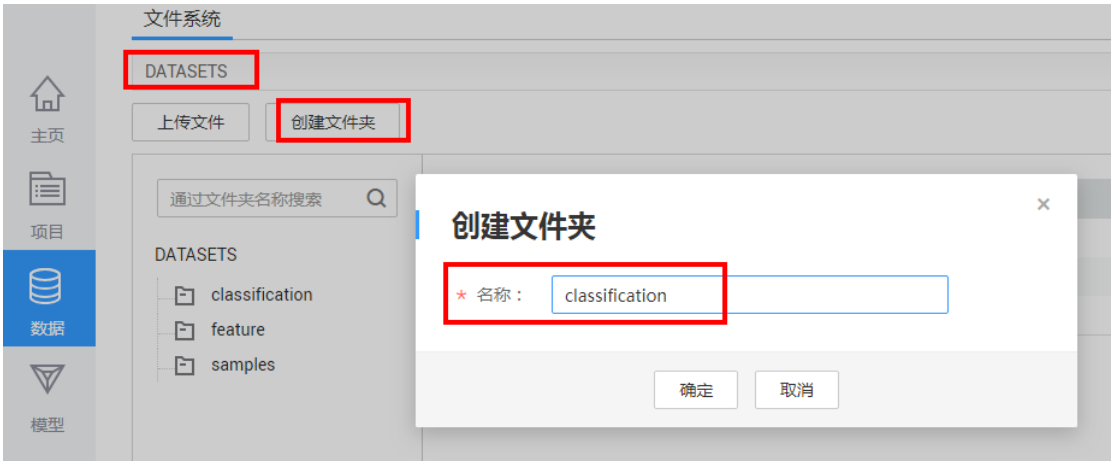
2.1 数据上传

使用数据某银行客户数据。

数据地址：https://obs-mlsclass12.obs.cn-north-1.myhwclouds.com/new_bank.csv

元数据地址：https://obs-mlsclass12.obs.cn-north-1.myhwclouds.com/bank_formal.desc

在MLS实例主页上单击“数据”---------单击“DATASETS”--------单击“创建文件夹”，文件夹名称为“classification”



点击“classification”进入文件夹，将数据文件和元数据文件上传至这个文件夹

2.2 创建项目

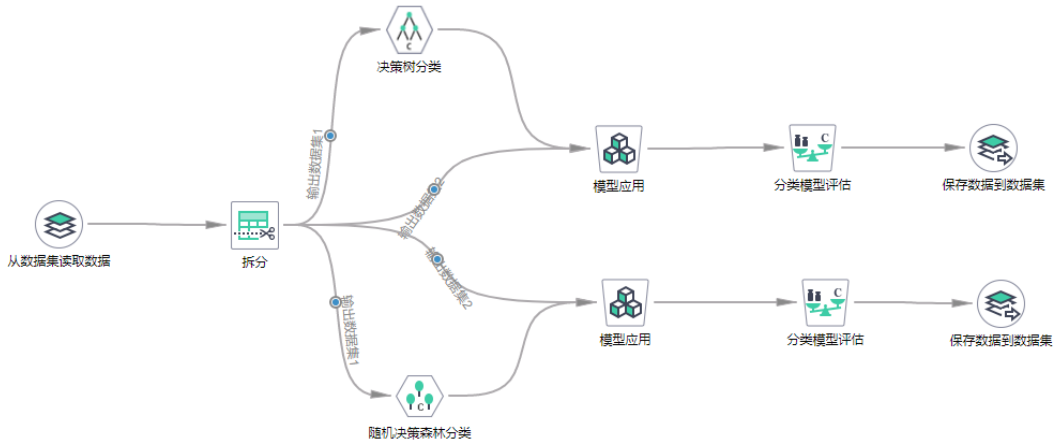
在MLS实例主页上单击“创建项目”，并写入项目名称，导入案例无需选择，完成后单击“确定”。

2.3 创建工作流

MLS实例主页单击“项目”—单击2.2中创建的项目名称----单击工作流-----单击“创建工作流”

2.4 编辑工作流

单击“工作流”—单击2.3中创建的工作流名称----打开一个空的工作流，然后按照下图的方式进行编辑，所有的算子在工作流页面的左侧“节点库”中都可以找到。



其中：“拆分”与“决策树分类”相连时选择“输出数据集1”；“拆分”与“随机决策森林分类”相连时选择“输出数据集1”；“拆分”与两个“模型应用”相连时选择“输出数据集2”；

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

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”：

6）与“随机决策森林分类”分支相连的“保存数据到数据集”：

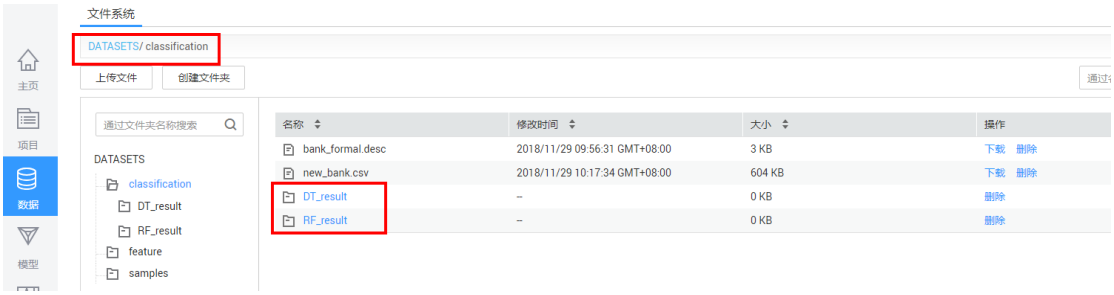
7）与“决策树分类”分支相连的“保存数据到数据集”：：

2.5 运行工作流

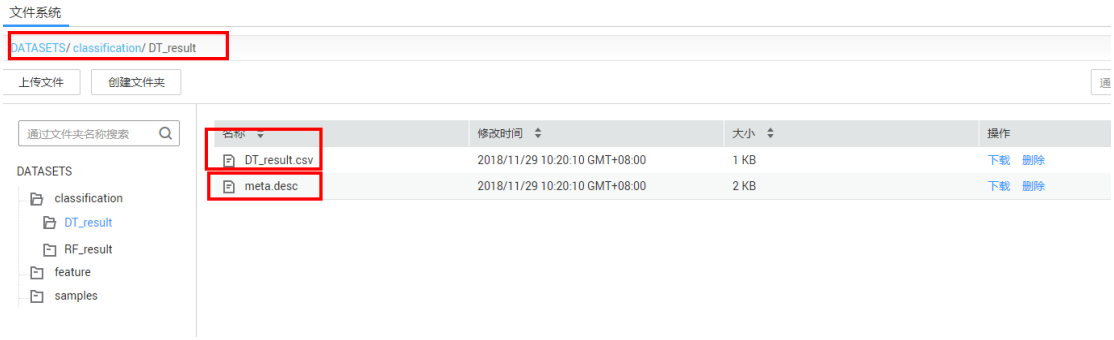
1）单击 运行工作流。

在下方的运行日志查看运行结果。

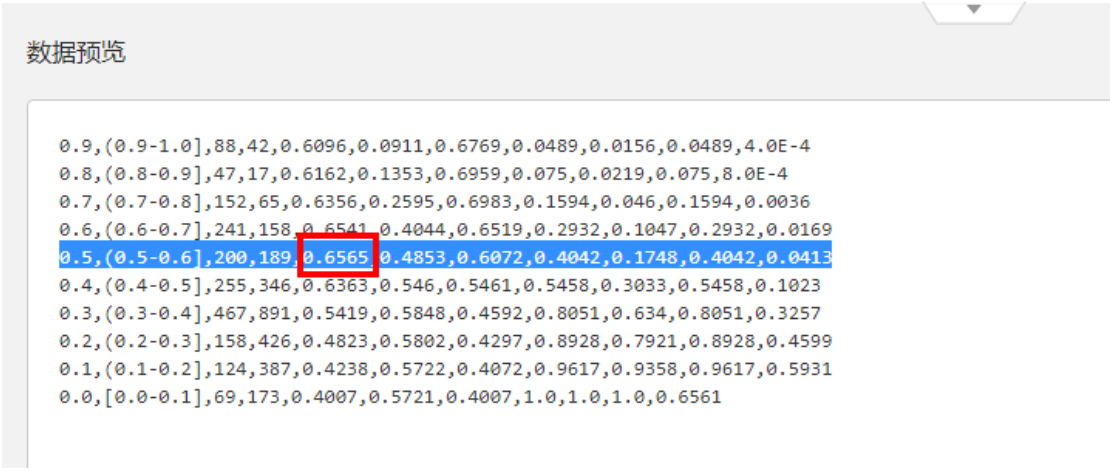
2）工作流运行完毕后，可以在“主页”—“数据”当中找到两个结果文件，进行查看



3）比如查看“决策树分类”的模型评估结果



先单击meta.desc查看每一列的意义，再单击DT_result.csv查看结果。随机森林分类的结果同理。分类问题一般查看的是第一列“阈值”=0.5的那一行的结果。



在分类概率阈值=0.5时，决策树分类的准确率为0.6565

3 打卡任务

3.1 完成单元测试

3.2 任务截图

1、 在2.4工作流界面进行截图：

1）右上角为用户名、下方为“工作流运行成功”

2）工作流与图示相同