5.2 机器学习助力质量分类（下）

1 任务介绍

本次任务将介绍如何使用MLS工作流中的多个分类算法节点实现质量分类。

1）本次课程使用的MLS实例是MLS标准版实例，区域是华北区-北京一。假设创建时提示没有标准版配额，则说明您已经拥有了标准版实例。

2）MLS标准版实例的创建过程请参看https://support.huaweicloud.com/qs-mls/index.html中左侧“快速入门”——“【标准版】创建MLS实例【01】”。

3）释放资源，MLS工作流是按需收费，最好在课程结束前不要删除实例。如果要释放资源请首先在MLS实例页左侧的“设置”中，将NoteBookServer进行关闭。然后在MLS实例管理控制台，将标准版实例进行删除。

2 任务执行

2.1 数据上传

使用数据某生鲜渠道销售数据。

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

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，将数据文件上传至这个文件夹

数据已经做了打标签操作。

2.2 创建项目

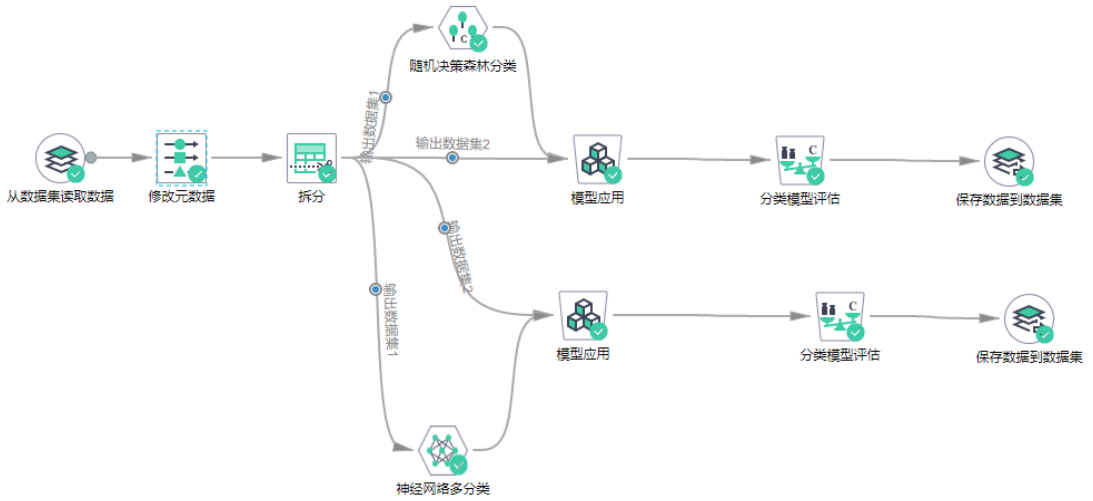
在MLS实例主页上单击“创建项目”，并写入项目名称“classification”，导入案例无需选择，完成后单击“确定”。

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。



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/chanpinzhiliang.csv（或者是自己上传的地址）;选择包括表头。

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

3）与“随机决策森林分类”相连“模型应用”：预测类型：分类

4）“随机决策森林”：改变参数

随机决策森林分类

* 树的数目:
300

* 最大树深度:
20

* 最大分箱数:
300

* 不纯度:
Gini

* 特征子集选取策略:
Auto

随机种子:
0

5） “随机决策森林”分支的“保存数据到数据集”：

保存数据到数据集

* 文件路径:
/classification/ ···

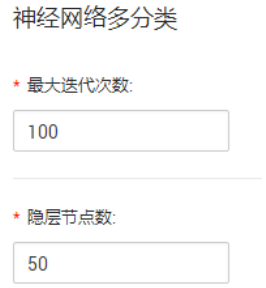
* 文件名:
evalue_rf

* 文件格式:
CSV

* 字段分隔符:
,

允许覆盖:

6）“神经网络多分类”：



7）与“神经网络多分类”相连“模型应用”：预测类型：分类

8）“神经网络多分类”分支的“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）在两个“分类模型评估”节点右键，单击“查看评估结果”，可以将混淆矩阵可视化。

查看评估结果

混淆矩阵

预测

真实	0	2	3	1	精确率	召回率	F1 测量
0	827	0	0	2	0.484	0.998	0.651
2	277	0	0	0	0.000	0.000	0.000
3	324	0	0	1	0.000	0.000	0.000
1	282	0	0	2	0.400	0.007	0.014

查看评估结果

混淆矩阵

预测

真实	2	1	3	0	精确率	召回率	F1 测量
2	0	0	0	277	0.000	0.000	0.000
1	0	1	1	282	0.200	0.004	0.007
3	0	1	0	324	0.000	0.000	0.000
0	0	3	4	822	0.482	0.992	0.649

可视化的结果显示:相比于课程4.2 对于1 2 3 三个类别的分类有些许提升，但是依然很差。这个时候就要考虑是否是特征设计的问题，本课程采用的是原始特征，尝试多种算法后依然效果不好，那么就必须在特征上进行改变，大多数时候特征设计大于算法选择。

大家可自行尝试。

一种特征设计后的数据：https://obs-class-mls2019.obs.cn-north-1.myhuaweicloud.com/chanpinzhiliang_new.csv

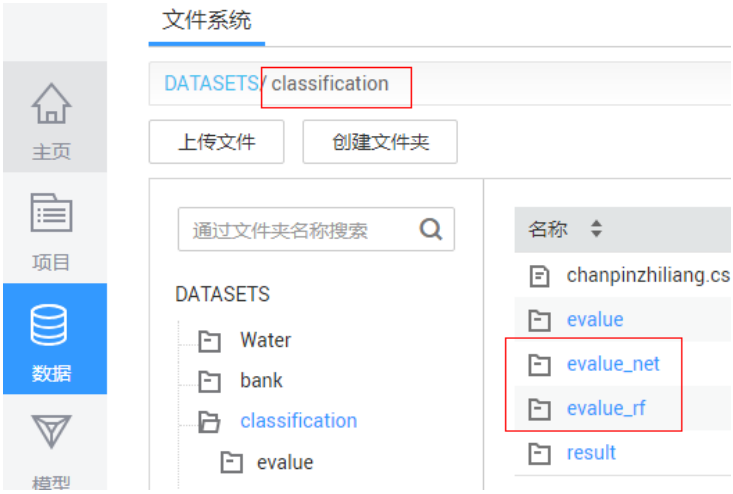
同样的工作流，效果很好，评估结果如下：

查看评估结果

混淆矩阵

真实 \ 预测	0	2	3	1	精确率	召回率	F1 测量
0	770	20	6	15	0.939	0.949	0.944
2	19	262	23	3	0.885	0.853	0.869
3	6	14	250	17	0.825	0.871	0.848
1	25	0	24	261	0.882	0.842	0.861

3）工作流运行完毕后，可以在“主页”—“数据”当中找到结果文件，进行查看



先单击meta.desc查看每一列的意义，再单击csv查看结果。最后一列是聚类结果列。

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