

数据挖掘-模型对比

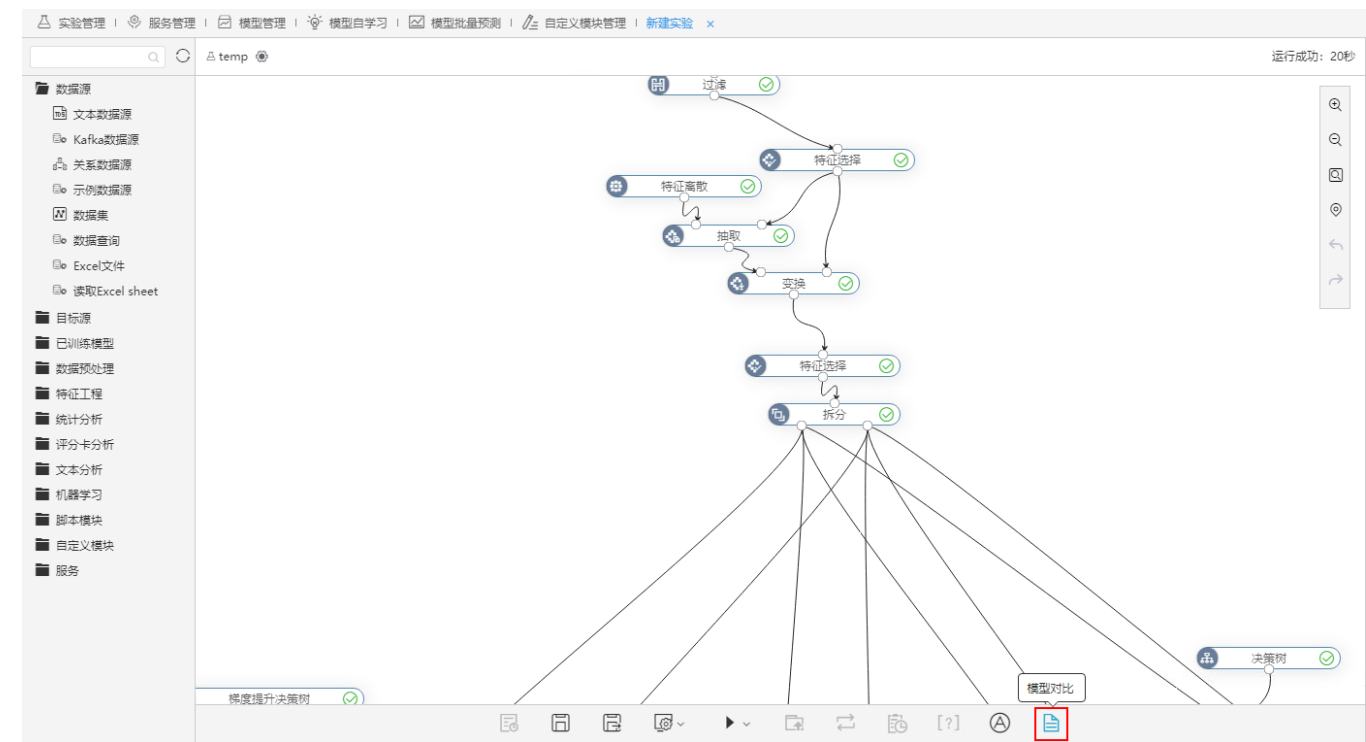
- 概述
- 模型对比
 - 模型报告详情
 - 分类评估报告
 - 回归评估报告
 - 聚类评估报告
 - 导出

概述

用于展示各个模型的分析结果、不同算法模型的对比分析，并导出评估报告，报告内容包括模型数据、参数、特征变量、特征重要性、评估指标等。

功能入口

1、在工具栏中点击 **模型对比** 按钮，打开评估报告配置页面。



2、勾选需要生成评估报告的算法。

评估报告配置

☒ 分类算法

☒ 决策树

☒ 朴素贝叶斯

☒ 逻辑回归

☒ 梯度提升决策树

确定

取消

当一个实验中同时存在两种以上不同的算法时，算法会分组显示。同一个组内，算法个数较多的组会排在较前位置。

3、点击 **确定** 按钮，打开模型对比界面。

模型对比

模型对比

决策树

朴素贝叶斯

逻辑回归

梯度提升决策树

< 上一指标

准确率

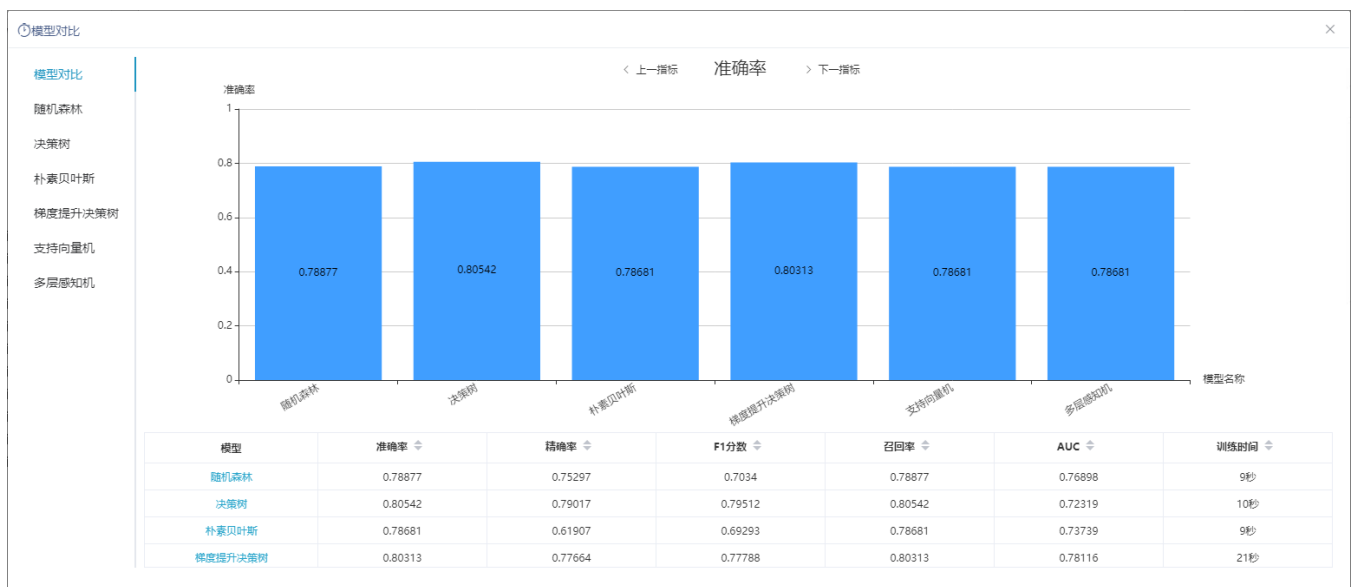
> 下一指标

模型	准确率	精确率	F1分数	召回率	AUC	训练时间
决策树	0.78681	0.72581	0.69355	0.78681	0.57802	4秒
朴素贝叶斯	0.72445	0.68318	0.69997	0.72445	0.58156	4秒
逻辑回归	0.78681	0.61907	0.69293	0.78681	0.58834	5秒
梯度提升决策树	0.78714	0.73281	0.6991	0.78714	0.59715	13秒

模型对比页面主要包括模型对比，及各个算法的模型报告详情。

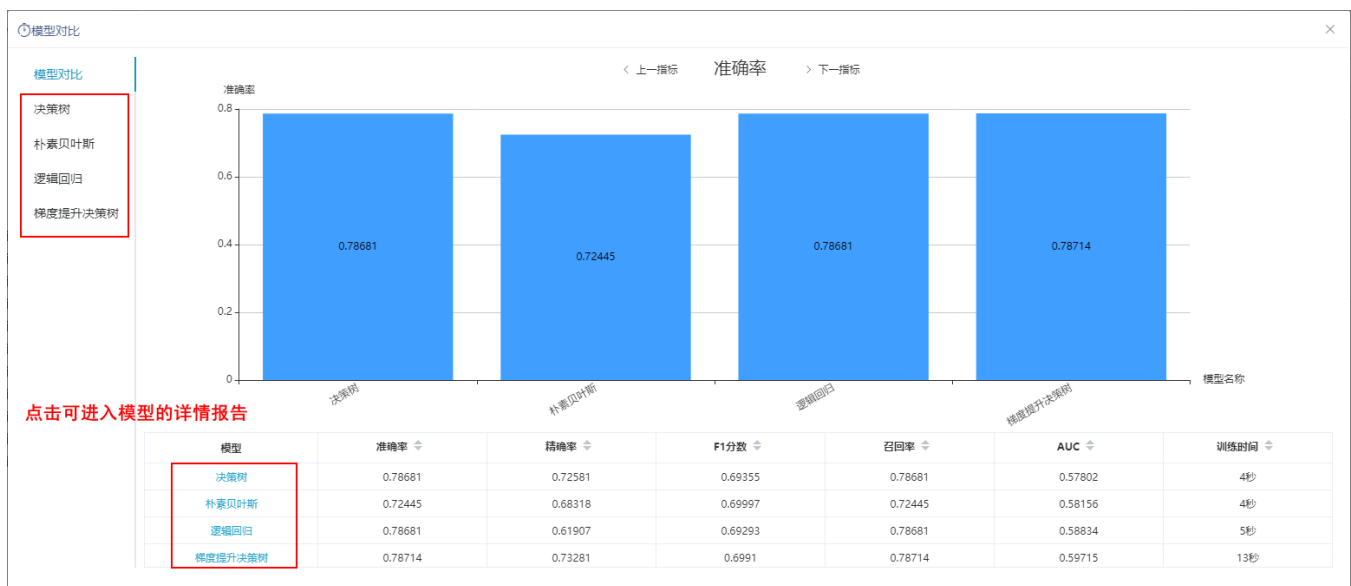
模型对比

用于不同算法模型之间的对比分析。



- 上一指标/下一指标：切换指标；
- 柱图：查看不同模型的各个指标对比情况。
- 模型对比表格：查看模型详细的指标数值。

点击左侧TAB页签或模型对比表格中的模型列，可跳转到指定模型的详情页面。



模型报告详情

各个分类算法所对应的评估报告不同，主要分为以下三种类型的报告：

分类评估报告

选择分类评估算法，这里以“随机森林”为例。

报告主要展示了以下内容：

- 分类算法可以显示的指标有：准确率、精确率、召回率、F1分数、特征重要程度、特征系数（线性回归和逻辑回归）、算法元数据、混淆矩阵、ROC曲线、KS曲线、树图（决策树算法）。
- 多分类模型不显示混淆矩阵、ROC曲线、KS曲线、树图。

模型对比

随机森林

决策树

朴素贝叶斯

梯度提升决策树

支持向量机

多层感知机

随机森林模型评估报告

导出



对于当前给定的测试数据集，模型预测正确的样本占总样本的比重为78.88%，加权精确率为75.30%，加权召回率为78.88%，结合加权精确率和加权召回率，模型的加权F1分数为70.34%。

特征

特征名
月均AUMBuckerizer
年龄Buckerizer
是否持有信用卡Buckerizer
购买理财产品数量Buckerizer
是否流失 [label]

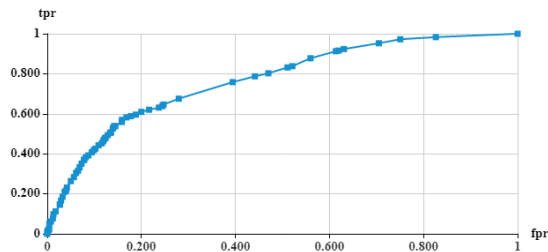
算法超参数

参数名称	说明	属性值
maxbins	分裂特征的数量	32
maxDepth	树的深度	4
numTrees	树的个数	20
impurity	衡量准则	gini

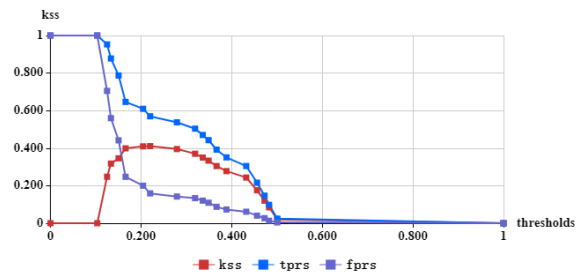
混淆矩阵

	真实 No	真实 Yes
预测 No	2400	637
预测 Yes	10	16

ROC曲线 (AUC = 0.7690)



KS曲线



回归评估报告

选择回归算法，这里以“梯度提升回归树”为例。

报告主要展示了以下内容：

- 回归报告中显示的指标包括：特征重要程度、特征系数、算法超参数、回归算法指标。
- 梯度提升回归树没有特征系数指标。

梯度提升回归树模型评估报告

导出

特征

特征名
gearbox
power
kilometer
v_14
v_13
v_12
v_11
v_10
v_9
v_8
v_7
v_6
v_5
v_4

算法超参数

参数名称	说明	属性值	
normalization	归一化	method	StandardScaler
		pnorm	2
		std	false
		mean	false
maxbins	分裂特征的数量	150	
maxDepth	树的深度	10	
impurity	计算信息增益的方式	gini	

回归指标

指标名	指标值
mae(平均绝对误差)	1184.15926
mse(均方误差)	8501202.775
rmse(均方根误差)	2915.68221
r2(决定系数)	0.85317
adjusted r2(校正决定系数)	0.8523

聚类评估报告

选择聚类算法，这里以“K均值”算法为例。

报告主要展示了以下内容：

- 样本的各个聚类的个数、特征（不展示标签列）。
- 存在模型的超参数信息和聚类的指标。
- 各个聚类中心的坐标。

K均值模型评估报告

导出

样本分布



特征

特征名
Alcohol
Malic_acid
Ash
Alcalinity_of_ash
Magnesium
Total_phenols
Flavanoids
Nonflavanoid_phenols
Proanthocyanins
Color_intensity
Hue
OD315_of_diluted_wines
Proline

算法超参数

参数名称	说明	属性值
silhouette	0.81935	
SSE	4545746.292	
CH	23.51379	

聚类指标

指标名	指标值
silhouette	0.81935
SSE	4545746.292
CH	23.51379

聚类中心

类名	维度	数值
cluster-0	Alcohol	12.70655738
	Malic_acid	2.533606557
	Ash	2.340245902
	Alcalinity_of_ash	20.42377049
	Magnesium	96.68852459
	Total_phenols	2.067704918
	Flavanoids	1.64852459
	Nonflavanoid_phenols	0.39352459
	Proanthocyanins	1.456639344
	Color_intensity	4.823852451
	Hue	0.910737705
	OD315_of_diluted_wines	3.047678571
cluster-1	Malic_acid	1.906607143
	Ash	2.42375
	Alcalinity_of_ash	17.47142857
	Magnesium	106.3928571
	Total_phenols	2.790535714
	Flavanoids	2.85875
	Nonflavanoid_phenols	0.292857143
	Proanthocyanins	1.883392857
	Color_intensity	5.568392857
	Hue	1.059214286
	OD315_of_diluted_wines	3.047678571
	Proline	1146.428571

导出

在模型详情页面，点击右上角的 导出 按钮，可将报告保存为PDF文件。

模型对比

决策树

朴素贝叶斯

逻辑回归

梯度提升决策树

决策树模型评估报告

导出

 导出报告



对于当前给定的测试数据集，模型预测正确的样本占总样本的比重为78.68%，加权精确率为72.58%，加权召回率为78.68%，结合加权精确率和加权召回率，模型的加权F1分数为69.36%。

特征

特征名
月均AUMBuckerizer
年限
账户余额
是否流失 [label]

算法超参数

参数名称	说明	属性值
------	----	-----