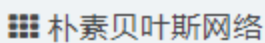


3.4.5.4 朴素贝叶斯

 朴素贝叶斯网络

图标:

描述: 朴素贝叶斯法是基于贝叶斯定理与特征条件独立假设的分类方法。

字段属性

特征列: 通过勾选的方式选择特征所在列。

标签列: 选择分类标签所在的列。

参数设置

条件概率分布满足何种分布: 包含高斯分布、多项式分布、二项分布, 默认高斯分布。

输出

表结果: 朴素贝叶斯算法分类结果。

报告: Confusion Matrix、Receiver Operating Characteristic(ROC)、Precision-Recall

示例

下列对某数据进行朴素贝叶斯算法分类:

- 选择特征数列, 标签列。如图 294 所示。
- 保留默认参数, 条件概率分布为高斯分布如图 295 所示。
- 朴素贝叶斯运行成功后, 可选择查看结果, 如图 296 所示。
- 朴素贝叶斯运行成功后, 可选择查看报告, 如图 297 所示。
- 模型评估配置如图 298 所示。
- 模型评估运行成功后, 选择查看数据, 如图 299 所示。
- 模型评估报告: Confusion Matrix、Receiver Operating Characteristic(ROC)、Precision-Recall
- 模型预测配置如图 301 所示。
- 模型预测结果如图 302 所示。

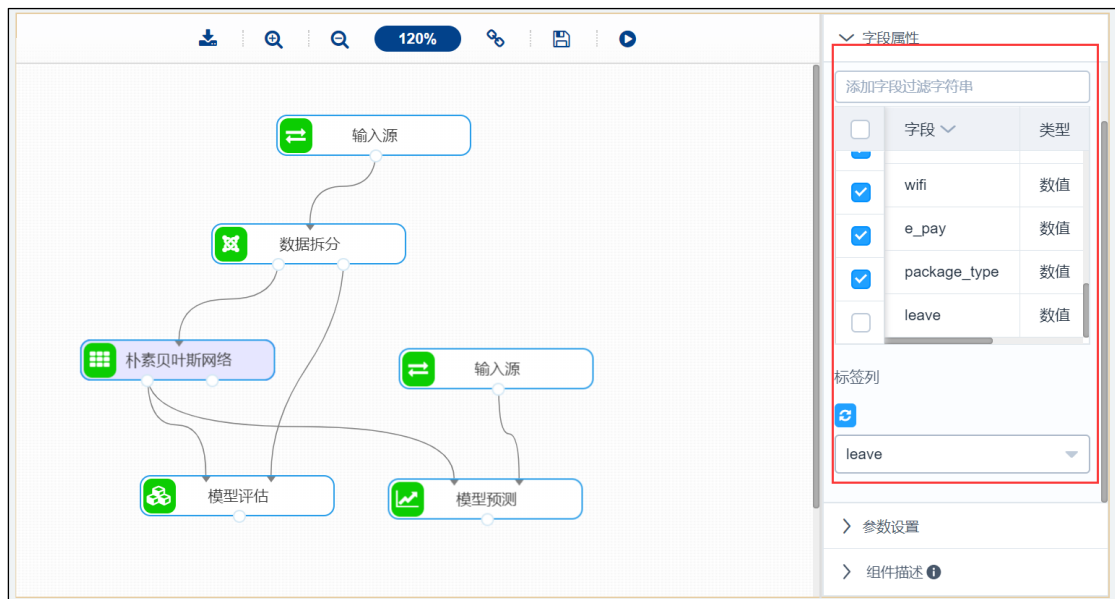


图 294

The screenshot shows the '参数设置' (Parameter Settings) panel for the '朴素贝叶斯网络' component. The '条件概率分布满足何种分布' (Which distribution does the conditional probability distribution satisfy?) parameter is currently set to '高斯分布' (Gaussian Distribution).

图 295

The screenshot shows the '预览数据' (Preview Data) window, displaying a table with 8 rows of data. The columns are 'wifi', 'e_pay', 'package_type', 'leave', and 'predict_label'. The 'predict_label' column represents the model's prediction for each row.

wifi	e_pay	package_type	leave	predict_label
0	1	2	1	1
28	1	4	1	1
0	0	3	1	1
0	0	3	0	0
0	0	3	0	0
0	1	1	0	1
0	1	2	1	1
0	0	2	0	1

At the bottom of the window, there is a pagination control showing '共 597 条' (Total 597 rows), '25 条/页' (25 rows per page), and a page number '1' out of '24'.

图 296

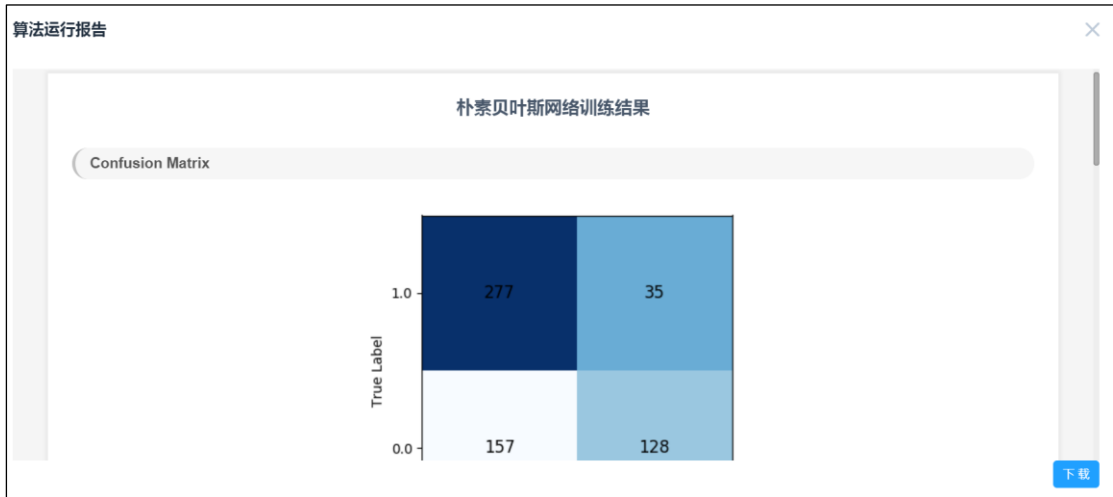


图 297

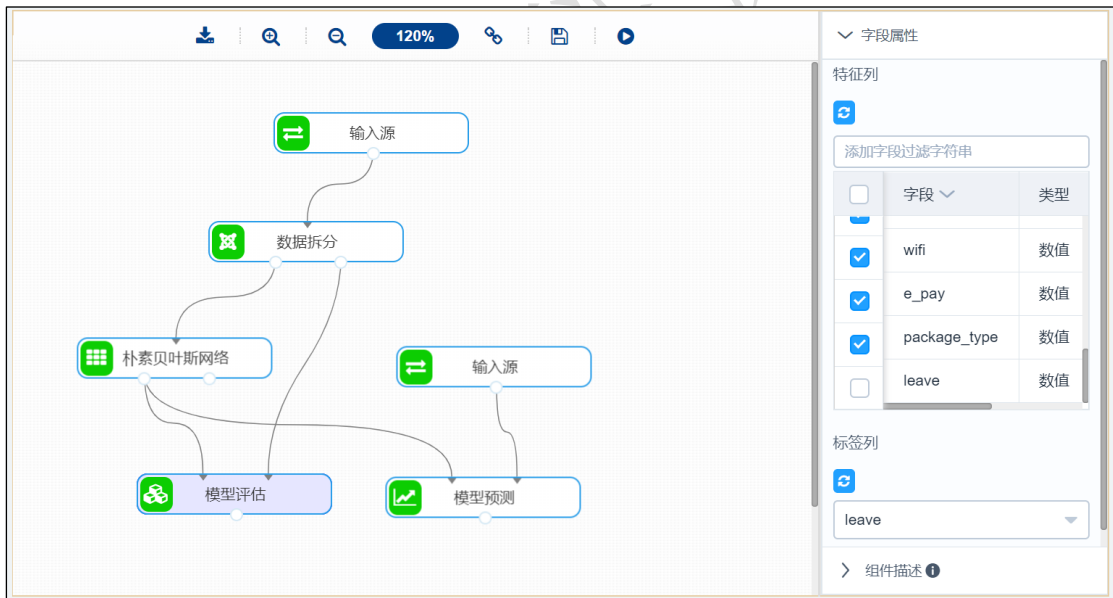


图 298

预览数据

wifi	e_pay	package_type	leave	predict_label
0	0	3	0	0
0	0	1	0	0
0	1	2	1	1
33	0	4	0	1
0	0	3	0	0
28	1	1	0	1
0	1	4	0	1
0	1	2	0	1

共 199 条 25 条/页 1 2 3 4 5 6 ... 8 前往 1 页

图 299

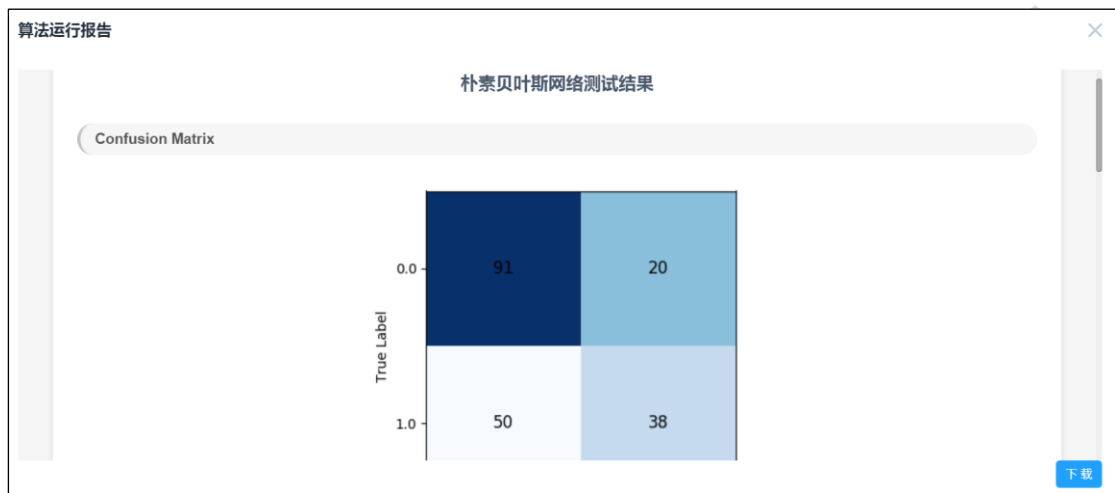


图 300

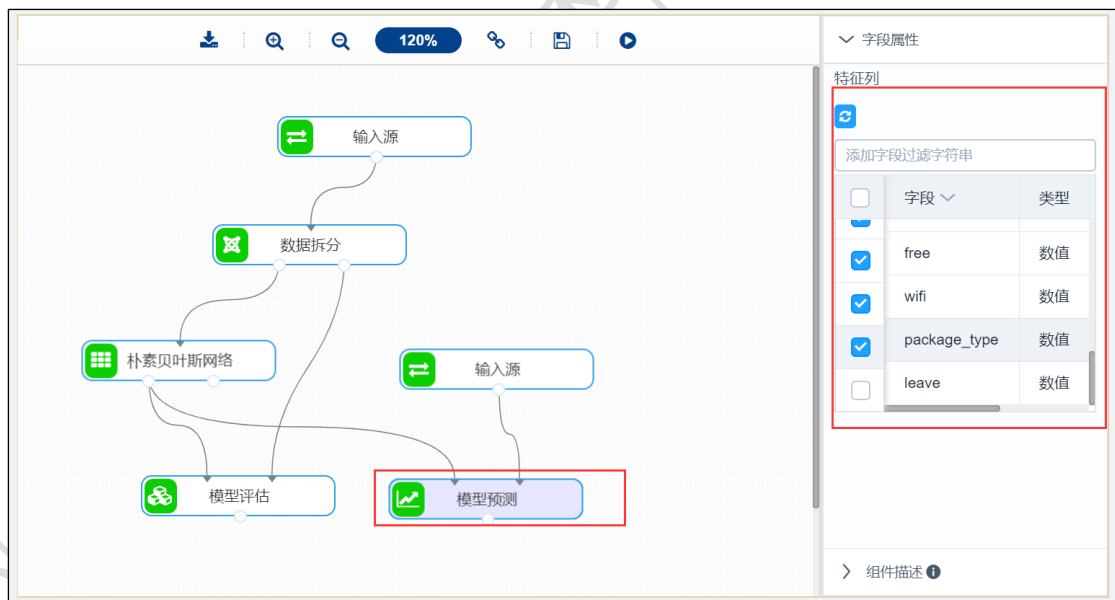


图 301

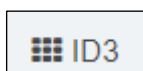
onths	capital_cost	free	wifi	package_type	predict_label
	4	0	0	1	1
	6	0	0	3	1
	12	19	0	3	1
	10	29	0	3	1
	24	0	0	2	1
	5	0	0	1	1
	7	22	0	3	1
	9	0	0	1	1

共 796 条 25 条/页 < 1 2 3 4 5 6 ... 32 > 前往 1 页

图 302

3.4.5.5 ID3

图标:



描述: ID3 算法是一种贪心算法，用来构造决策树。以信息熵的下降速度为选取测试属性的标准，即在每个节点选取还尚未被用来划分的具有最高信息增益的属性作为划分标准，然后继续这个过程，直到生成的决策树能完美分类训练样例。ID3 算法可用于划分标称型数据集，不能处理连续分布的数据特征。

字段属性

特征列：请选择标称类数据。

标签列：选择分类标签所在的列。

输出

表结果：ID3 算法分类结果。

报告：Confusion Matrix、Receiver Operating Characteristic(ROC)、Precision-Recall

示例

下列对某数据进行 ID3 算法分类：

- 选择特征数列，标签列。如图 303 所示。
- ID3 运行成功后，可选择查看结果，如图 304 所示。
- ID3 运行成功后，可选择查看报告，如图 305 所示。
- 模型评估配置如图 306 所示。