

北京邮电大学实践课程实验报告

实验一：熟悉大数据实验平台

实验目的

熟悉 BDAP 平台，工作流的基本操作和对应的组件

实验任务

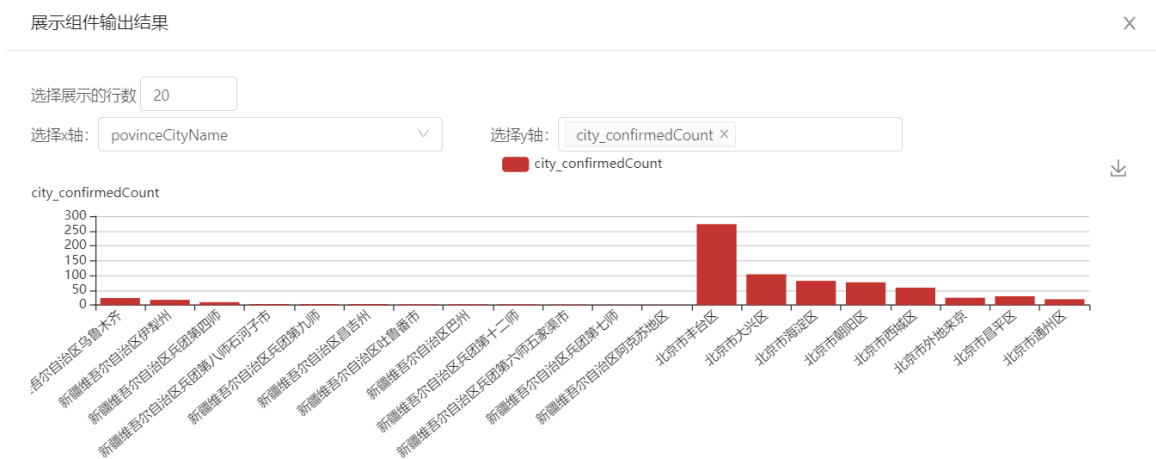
首先加载数据，然后合并两列数据（provinceName, cityName）；
投影展示所合并得到的数据列（provinceCityName）和 一 列 原 始 的 数 据 列（city_confirmedCount）；
最后根据所有数据中的某两列来绘制柱状图。

实验步骤和结果

加 载 数 据 集 → 计 算 生 成 列 → 统 计 → 列 投 影 → 统 计



一个由投影得到的数据可视化结果，一个柱状图



实验总结

实验二：常见的数据预处理实验

实验目的

熟悉常见的数据预处理操作

实验任务

以“大洲”为聚集字段，使用 `groupBy` 操作进行聚集，并统计确诊、治愈、死亡的人数；按照确诊人数，使用 `sort` 排序操作对整体数据进行排序。

实验步骤和结果

加载数据集→`groupby` 聚集→统计→`sort` 排序→统计



展示 sort 排序操作完成后的整体数据

展示组件输出结果

×

选择展示的行数

20

continentName	TotalConfirmed	TotalCuredcount	TotalDeadcount
北美洲	3094146	962203	167030
欧洲	2404525	1334691	191635
亚洲	2340860	1524827	57301
南美洲	2233316	1278731	85506
非洲	407651	187508	10179
大洋洲	9345	8410	131
其他	712	574	13

实验总结

实验三：决策树模型

实验目的

了解决策树的组成和构造算法，熟悉 BDAP 中机器学习相关的工作流程和组件。

实验任务

决策树模型的训练、验证和预测

实验步骤和结果

加载数据集→训练模型

加载数据集→决策树→表格→多分类评估器→表格



训练出来的决策树模型，在预测集上的评估结果

展示组件输出结果

X

选择展示的行数

20

metric
0.639344262295082

1

实验总结

实验四：聚类实验 1

实验目的

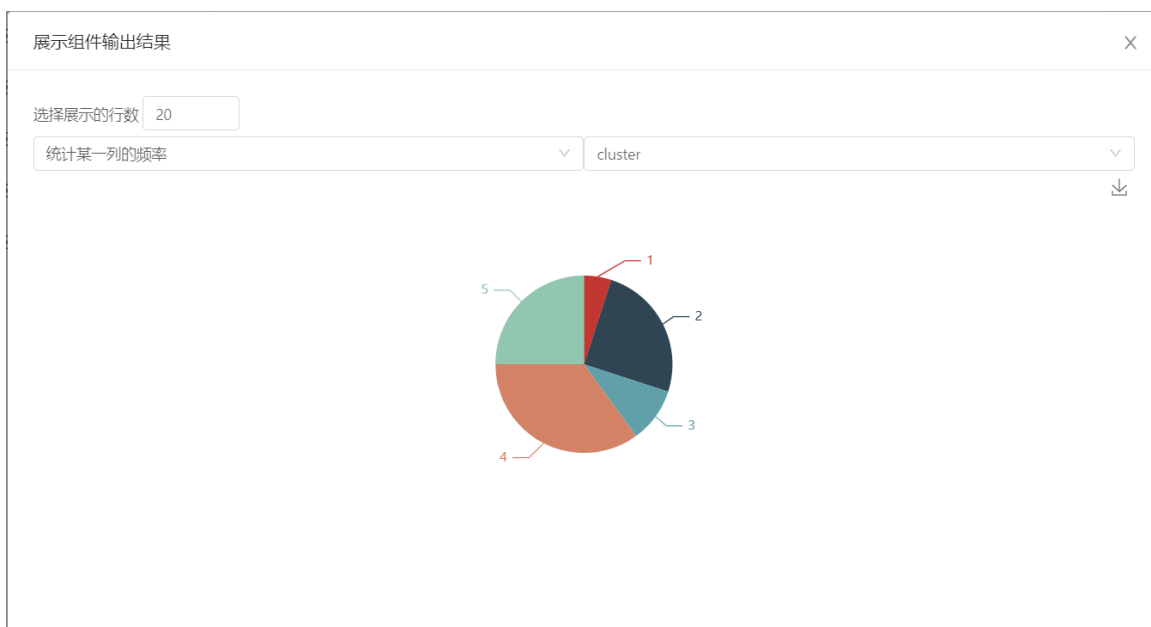
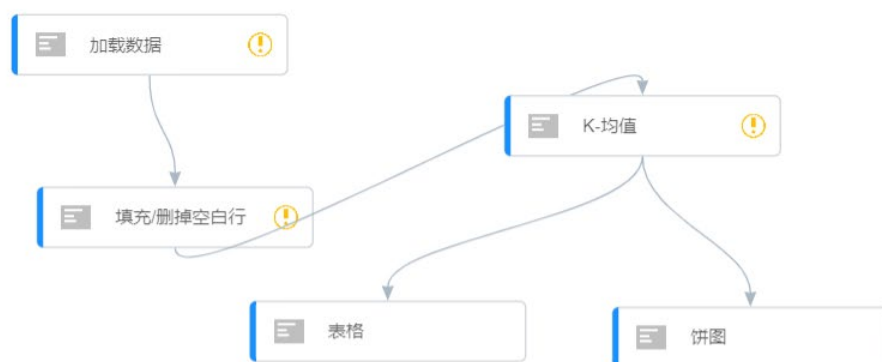
了解无监督学习

实验任务

删除空白行，通过聚类实验分析王者荣耀英雄类别

实验步骤和结果

加载数据集→删除空白行→选取聚类中心→表格



实验总结

实验五：关联规则实验

实验目的

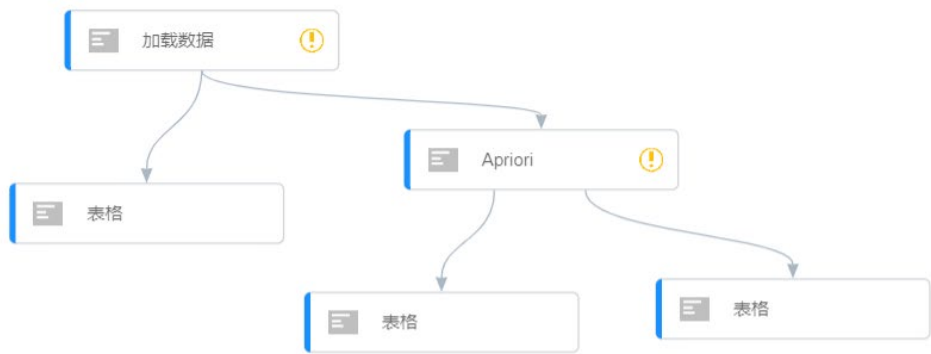
了解关联规则算法

实验任务

某个杂货店一个月真实的交易记录，列属性为交易商品名称，记录

实验步骤和结果

加载数据集→Apriori 算法→表格



展示组件输出结果

选择展示的行数 20

KnownItem	RelatedItem	Confidence
tropical fruit,other vegetables	root vegetables	0.34277620396600567
tropical fruit,root vegetables	other vegetables	0.5845410628019324
other vegetables	whole milk	0.38675775091960063
tropical fruit,rolls/buns	whole milk	0.4462809917355372
curd	yogurt	0.3244274809160305
citrus fruit	other vegetables	0.34889434889434895
whole milk,soda	other vegetables	0.3477157360406091
soda,other vegetables	whole milk	0.42546583850931674

展示组件输出结果

X

选择展示的行数

20

Item	Support
bottled beer,soda	0.016980172852058974
tropical fruit,other vegetables,root vegetables	0.012302999491611592
sausage,yogurt	0.019623792577529234
whole milk,other vegetables	0.07483477376715811
tropical fruit,whole milk,rolls/buns	0.010981189628876462
tropical fruit,root vegetables	0.021047280122013217
soda,whipped/sour cream	0.011591255719369599
citrus fruit,whipped/sour cream	0.010879511947127605

实验总结

实验六：SVM 拓展实验

实验目的

了解 SVM 基本原理

实验任务

下载数据集，转换数据集格式，数据集划分，模型训练，模型预测

实验步骤和结果

加载数据

多分类SVM

加载数据

S[多分类SVM]

表格

多分类评估器

表格

展示组件输出结果

选择展示的行数20

metric

0.5777777777777777

实验总结