****

**数据挖掘课程实验报告**

学 院： 信息管理学院

课程名称： 数据挖掘

教学班级：

姓 名：

学 号：

实 验 报 告

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 课程名称 | | 数据挖掘及财经应用 | | 教学班级 | |  | | 指导老师 | | 江腾蛟 |
| 学号 |  | | 姓名 | |  | | 行政班级 | |  | |
| 实验项目 | | 实验1：数据挖掘流程初探—基于Drug数据集 | | | | | | | | |
| 组员名单 | | 独立完成或姓名（学号）（操作性和验证性实验均独立完成，综合性实验3-5人一组） | | | | | | | | |
| 实验类型 | | ■操作性实验 □验证性实验 □综合性实验 | | | | | | | | |
| 实验地点 | |  | | | 实验日期 | | | 可写起止时间段 | | |

1. 实验目的和要求：

**目的：**

(1)初识数据挖掘与数据分析；

(2)发现数据挖掘与数据分析的趣味性。

**实验数据集介绍：**

数据文件(DRUG.TXT)以文本文件形式存放.

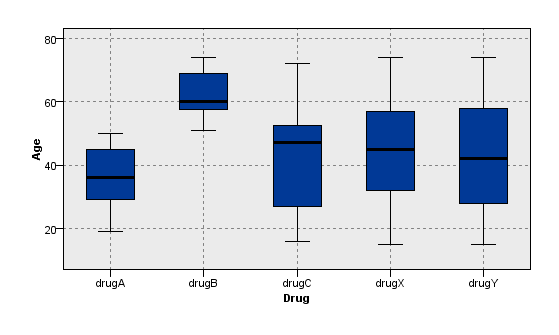
这是一份关于药物研究的数据。大批患有同种疾病的不同病人，服用五种药物中的一种（Drug A，Drug B，Drug C，Drug X和Drug Y）之后都取得了同样的治疗效果。现需要利用数据挖掘发现以往药物处方适用的规律，给出不同的临床特征病人更适合服用哪种药物的建议，为未来医生开具处方提供参考。

案例数据是随机挑选的部分病人服用药物前的基本临床检查数据，包括：病人的年龄（Age）、性别（Sex）、血压(BP）、胆固醇（Cholesterol）、唾液中钠元素（Na）、钾元素（K）含量和服用药物（Drug）。

**实验要求：**

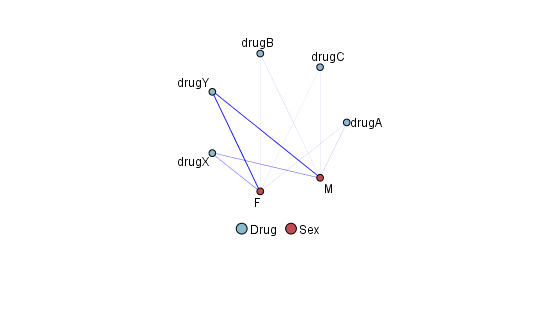
1. 读入数据源文件，输出表以查看具体数据内容并将表名重命名为Drug；
2. 添加类型节点，读取值并将Drug的角色设为目标，其余属性角色均为输入；
3. 添加数据审核节点，观察基本统计量与高级统计量下审核内容的区别，观察质量标签下各个选项的含义；
4. 分别观察各个输入变量与目标变量之间的相关性；

（4.1）Age与Drug



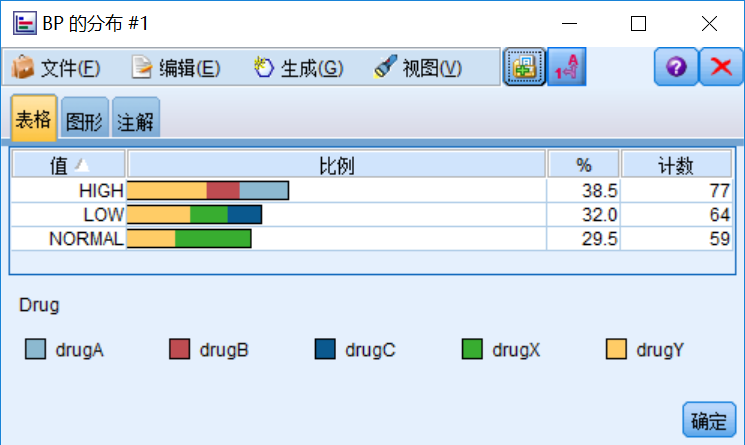
上图，你有什么发现？

（4.2）Sex与Drug



这个图呢？

（4.3）BP与Drug



这个图呢？

（4.4）……

1. 利用特征选择进行特征的重要性判定。
2. 建模，利用C5.0构建分类树并解读模型结果；比较特征选择与C5.0对特征重要性的判断是否一致。
3. 生成新属性Na/K
4. 分析Na/K与Drug的关系
5. 过滤节点去掉Na和K属性，因为已生成Na/K新属性。
6. 再次类型节点，读取值，设置输入与目标角色。
7. 利用特征选择进行特征的重要性判定。
8. 建模，利用C5.0构建分类树并解读模型结果；比较特征选择与C5.0对特征重要性的判断是否一致。
9. 利用分析节点，对比前后两个模型的正确率。

说明：提交一份尽你目前所能的详尽分析报告。

By the way,既然是分析报告，图文并茂肯定是最好的，对每张图都要有编号、有解读！

2．实验过程（记录实验步骤、分析实验结果）

2.1 XXXX

2.2 XXXX

……

3．问题反馈与收获（3.1实验过程中遇到的问题和解决办法; 3.2 通过此次实验操作有哪些收获，包括操作技巧、心得总结等均可记录下来。）

3.1 问题与解决办法

3.2 发现与收获

4．指导教师评语及成绩：

评语：

成绩： 指导教师签名：

批阅日期：月 日