# 说明文档

[1. 项目概述及人员： 2](#_Toc298)

[1.1 项目概述 2](#_Toc30141)

[1.2 人员 2](#_Toc16157)

[2. 需求分析 2](#_Toc9188)

[3. 系统设计 3](#_Toc20549)

[3.1 总体设计 3](#_Toc8879)

[3.2 系统功能模块 3](#_Toc11705)

[4. 代码框架及函数设计 4](#_Toc10997)

[5. 测试 4](#_Toc10404)

[参考文献 4](#_Toc7708)

[结束语 4](#_Toc17945)

## 项目概述及人员：

### 1.1 项目概述

针对乳腺癌患者，可以依据中医阶段治疗法进行治疗，在辨症的基础上围绕各个病程的特殊症候先症而治型。依据医学指南，将乳腺癌辨证统一化，分为六种症型。即患者在围手术期、围化疗期、围放疗期和内分泌治疗期，基本都会出现特定的临床症状。可以通过一定的规则，挖掘出各中医症素与乳腺癌TNM分期之间的关系。探索不同分期阶段的乳腺癌患者的中医症素分布规律，以及阶段病变发展（71，72，73），指导乳腺癌的中医临床治疗。

（1）对数据进行预处理；

（2）利用数据挖掘的算法对数据进行分析，得到中医症素与乳腺癌TNM 分期之间的关系。根据实际的业务分析建议最小支持度设定为：6%，最小置信度为：75%；

（3）对所构建的模型进行分析，得出你的结论。

### 1.2 人员

|  |  |  |
| --- | --- | --- |
| 姓名 | 学号 | 分工 |
| 刘昕 |  | 贝叶斯算法，文档 |
| 李博伟 |  | svm，预处理 |
| 蒋浩丰 |  | 神经网络 |
| 苏莉娅 | 201628018629065 | 决策树，框架设计 |
| 昝文 |  | k近邻 |

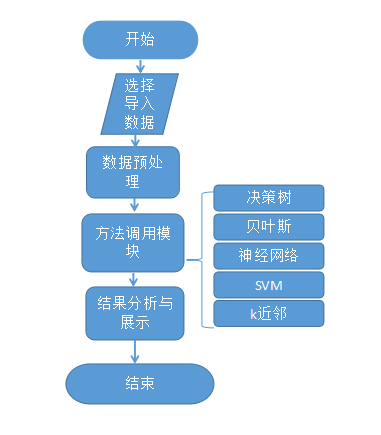
## 需求分析

（1）功能需求：利用数据挖掘的算法对数据进行分析，得到中医症素与乳腺癌TNM 分期之间的关系。根据实际的业务分析建议最小支持度设定为：6%，最小置信度为：75%；对所构建的模型进行分析，得出你的结论。

（2）性能需求：满足多算法优势比较。满足正确率在90%以上，错误率低于10%，运算速度可接受。

## 3. 系统设计

### 3.1 总体设计



### 3.2 系统功能模块

#### 3.2.1 预处理

对数据使用箱图方法去除噪声，再使用均值填充空值

#### 3.2.2 分类算法：

决策树：

贝叶斯：

svm：

k近邻：

神经网络：

#### 3.2.3 功能集成：

通过文件遍历的方法调用所有的方法，或者通过函数名直接调用具体某种方法。

#### 3.2.4 展示模块：

简单的可视化页面来调用具体方法，并给出参数，对于函数结果给出分析以及不同的结果之间的关系。

## 4. 代码框架及函数设计

## 5. 测试

交叉验证的方法

## 参考文献

## 结束语