**《数据挖掘实践任务教学平台》**

**用户手册**

|  |  |  |  |
| --- | --- | --- | --- |
| **作者：** | **马腾** | **日期：** | 2018.8.23 |
| **审核：** | **刘祥德** | **日期：** | 2018.9.9 |
| **批准：** |  | **日期：** |  |

**变更记录**

|  |  |  |  |
| --- | --- | --- | --- |
| **日期** | **版本** | **变更说明** | **作者** |
| 2018.08.23 | V1.0 | 文档创建 | 马腾 |
| 2018.09.02 | V1.1 |  | 马腾 |
|  |  |  |  |
|  |  |  |  |

目录

[1. 引言 5](#_Toc524299805)

[1.1 编写目的 5](#_Toc524299806)

[1.2 背景 5](#_Toc524299807)

[1.3 定义 5](#_Toc524299808)

[1.4 参考资料 5](#_Toc524299809)

[2. 用途 7](#_Toc524299810)

[2.1 功能 7](#_Toc524299811)

[2.2 性能 7](#_Toc524299812)

[2.2.1 精度 7](#_Toc524299813)

[2.2.2 时间特性 8](#_Toc524299814)

[2.2.3 灵活性 8](#_Toc524299815)

[2.3 安全保密 8](#_Toc524299816)

[3. 运行环境 9](#_Toc524299817)

[3.1 硬设备 9](#_Toc524299818)

[3.2 支持软件 9](#_Toc524299819)

[3.3 数据库 9](#_Toc524299820)

[4. 使用过程 9](#_Toc524299821)

[4.1 安装与初始化 9](#_Toc524299822)

[4.2 输入 10](#_Toc524299823)

[4.2.1 输入数据的现实背景 12](#_Toc524299824)

[4.2.2 输入格式 13](#_Toc524299825)

[4.2.3 输入举例 15](#_Toc524299826)

[4.3 输出 17](#_Toc524299827)

[4.3.1 输出数据的现实背景 19](#_Toc524299828)

[4.3.2 输出格式 20](#_Toc524299829)

[4.3.3 输出举例 22](#_Toc524299830)

[4.4 出错处理和恢复 24](#_Toc524299831)

[4.5 登录与注销 24](#_Toc524299832)

# 引言

## 编写目的

本文档是基于REST风格数据挖掘实践教学平台开发几近完成阶段针对用户所进行的用户手册，在本文档中我们通过对数据挖掘实践教学平台的两种不同使用者：教师与学生进行了详细而具体的需求分析，通过该文档读者可以了解该系统的所有功能及其运行环境，以便读者了解本系统的使用范围与使用方法。

## 背景

* 项目名称：基于REST风格的数据挖掘实践教学平台。
* 项目的委托单位：南京理工大学2017年立项的本科生科研训练项目。
* 项目指导教师：徐建
* 项目小组成员：刘祥德、马腾、刘鑫容。

## 定义

REST风格：REST即表述性状态传递（Representational State Transfer，简称REST）是一种软件架构风格，它是一种针对网络应用的设计和开发方式，可以降低开发的复杂性，提高系统的可伸缩性；目前在三种主流的Web服务实现方案中，因为REST模式的Web服务更加简洁，越来越多的web服务开始采用REST风格设计和实现；依照 Restful的设计原则，一切 Web 资源都以URI的方式进行统一的描述和访问，这也方便用户和开发者通过系统检索、获取资源数据。

* 教师：本系统的管理员，实践任务的发布者，管理学生信息，追踪任务进度。
* 学生：本系统的主要使用者，领取实践任务并上传数据挖掘阶段结果等。

## 参考资料

|  |  |  |  |
| --- | --- | --- | --- |
| **资料名称[标识符]** | **出 版 单 位** | **作 者** | **日 期** |
| 《计算机软件文档编制规范GB-T8567-2006》 |  |  | 2006 |
| 《南京理工大学本科生科研训练“百千万”计划开题报告》 |  | 刘祥德 马腾 刘鑫容 | 2017.11 |
| 《软件系统设计说明书》 |  | 刘祥德 马腾 刘鑫容 | 2018.8 |
| 《软件部署文档》 |  | 刘祥德 马腾 刘鑫容 | 2018.9 |
| 《软件需求规格说明书》 |  | 刘祥德 马腾 刘鑫容 | 2018.5 |
| 《系统设计说明书模板》 |  | 网友上传 | 2017 |
| 《系统需求规格说明书模板》 |  | 网友上传 | 2017 |

# 用途

## 功能

本系统意在提供一个为教师、学生提供一个实践教学平台，管理员可以在平台给学生分配指定类型的数据发掘任务，并同步追踪学生的数据发掘进度，进行有效的教学评估。学生可以根据任务在平台上传数据集，进行数据挖掘流程的分类、估计、预测、分析、总结工作，并上传期望进度，汇报发掘结果。系统会提供诸如图形、表格的数据呈现手段，帮助教学任务较好的完成。

对于每个数据挖掘任务、每位教师：

* 可以上传一个或者多个数据集以供使用
* 可以管理维护上传的数据集
* 可以批量上传学生信息
* 可以管理更新学生信息
* 可以对学生进行一键分组
* 可以发布与数据集和分组关联的数据挖掘任务
* 可以对任务分阶段设置目标
* 可以追踪学生任务的进度
* 可以查看获取阶段任务的结果

对于每个数据挖掘任务、每位学生：

* 可以查看下载数据集
* 可以查看自己接到的数据挖掘任务
* 可以查看自己所在的分组信息
* 可以分阶段上传数据挖掘结果
* 可以通过系统以图表呈现数据集

## 性能

### 精度

|  |  |
| --- | --- |
| **项目类型[标识符类型]** | **输入精度** |
| 账号信息 | 整数、整字 |
| 数据集描述 | 整数、整字 |
| 学生信息 | 整数、整字 |
| 分组信息 | 整数、整字 |
| 数据挖掘结果 | 一位小数 |

### 时间特性

系统响应的时间特性：用户对系统请求的反应时间控制在5秒内，如果超过5秒考虑设计交互友好的方式显示等待信息。

系统更新处理时间：系统对刷新请求的反应时间控制在5秒内。

系统数据传输时间：系统对上传与下载请求的反应时间一般可以控制在5秒内，如果文件过大或用户网络状态较差依据情况耗时不等。

系统计算时间：系统对一般大小的数据集图表呈现计算可以控制在10秒内，如果数据集实例较多或进行复杂表格的呈现，耗时不等。

### 灵活性

本系统基于REST风格的微服务架构，支持用户在web端、移动端做无缝场景切换。

当系统的输入不符合规范时，例如注册输入的用户名不符合命名规则，系统会给出提示信息用户操作不规范，其他操作也会有类似提示机制。

## 安全保密

本系统在Shiro安全维护的框架基础上，进一步使用了Token客户端令牌方案，针对每个请求帧，在客户端生成令牌，严格执行用户名的合法性验证机制，计算时间戳的差值，防止Token被截取下的重放攻击。令牌由身份验证服务进行签名，并且必须包含足够的信息，以便可以在所有微服务中建立用户身份。令牌会附加到每一个请求上，为微服务提供身份验证的信息。

# 运行环境

## 硬设备

#### 显示器尺寸：笔记本15.4寸，台式显示屏27寸；

#### 处理器：四核及以上；

#### 内存：4G及以上；

#### 硬盘容量：100G及其以上；

## 支持软件

Google浏览器

## 数据库

数据库系统：MySQL 5.6及以上

# 使用过程

## 安装与初始化

系统已经打包部署到了腾讯云，用户在任何网络连接顺畅的浏览器或移动端浏览器即可访问主机名或者域名http://118.24.62.94:9527 /data\_mining\_platform来使用系统。

其中<http://118.24.62.94:9527>是可变的，应用部署的主机不同、端口不同，对应的地址也不相同。

## 输入

注册要求：

|  |  |
| --- | --- |
| **学生字段[标识符]** | **输入要求** |
| 用户名 | 字母、数字、下划线 |
| 密码 | 字母、数字、下划线 |
| 学号/工号 | 数字、字母 |
| 姓名 | 汉字、字母 |
| 年级 | 年份 |
| 班级 | 数字、字母 |
| 专业 | 计算机科学与技术、软件工程、人工智能、  网络工程 |

数据集上传要求：

|  |  |
| --- | --- |
| **数据集字段[标识符]** | **输入要求** |
| 数据集名称 | 数字、字母、下划线 |
| 描述摘要 | 数字、字母、下划线 |
| 数据特征 | Multivariate-多变、Univariate-单变量、Sequential-按序、Time-Series-时序、Text-文本、Domain-Theory-畴理论 |
| 属性类型 | Categorical-分类的、Integer-整型、  Real-实数 |
| 相关任务 | Classification-分类、Regression-回归、Clustering-聚类、Other-其他 |
| 实例数 | 数字 |
| 允许缺省 | Yes/NO |
| 访问量 | 数字 |
| 相关论文 | 数字、字母、下划线 |

学生信息导入模板：

|  |  |
| --- | --- |
| **学生字段[标识符]** | **输入要求** |
| 学号 | 数字、字母 |
| 姓名 | 汉字、字母 |
| 年级 | 年份 |
| 班级 | 数字、字母 |
| 专业 | 计算机科学与技术、软件工程、人工智能、  网络工程 |

一键分组要求：

|  |  |
| --- | --- |
| **分组字段[标识符]** | **输入要求** |
| 学号 | 数字、字母 |
| 姓名 | 汉字、字母 |
| 年级 | 年份 |
| 班级 | 数字、字母 |
| 专业 | 计算机科学与技术、软件工程、人工智能、  网络工程 |
| 组名 | 数字、字母 |
| 分组编号 | 数字 |
| 组长 | 汉字、字母 |
| 执行任务 | 已有任务范围之内选择 |
| 任务状态 | 待分配、任务需推进、任务完成、任务进行中、任务未完成、新任务、新的提交 |

新建任务输入要求：

|  |  |
| --- | --- |
| **任务字段[标识符]** | **输入要求** |
| 任务名称 | 字母、数字、下划线 |
| 任务描述 | 字母、数字、下划线 |
| 任务编号 | 数字、字母 |
| 挖掘语言 | Java、C++、C、Go、Python、Matlab、Scala |
| 算法配置 | C4.5、k均值聚类算法、支持向量机、  Apriori关联算法、EM 最大期望算法、PageRank算法、AdaBoost 迭代算法、kNN：k最近邻算法、CART 分类算法 |
| 数据集合 | 已有数据集范围之内选择 |
| 任务状态 | 待分配、任务需推进、任务完成、任务进行中、任务未完成、新任务、新的提交 |
| 计划时间 | 年月日-年月日 |
| 阶段一目标描述 | 字母、数字 |
| 阶段一时间 | 年月日-年月日 |

数据挖掘分析要求：

|  |  |
| --- | --- |
| **数据集字段 [标识符]** | **输入要求** |
| 数据集 | 数据集文件 |

### 输入数据的现实背景

|  |  |  |
| --- | --- | --- |
| **输入数据** | **现实背景** | **数据来源** |
| 用户名 | 用户初始注册时填写相关信息 | 用户自行填写信息 |
| 密码 |
| 学号/工号 |
| 姓名 |
| 年级 |
| 班级 |
| 专业 |
| 数据集名称 | 为数据挖掘任务上传用到的数据集 | UCI等各数据集网站及其对应数据集描述文件 |
| 描述摘要 |
| 数据特征 |
| 属性类型 |
| 相关任务 |
| 实例数 |
| 允许缺省 |
| 访问量 |
| 相关论文 |
| 学号 | 教师批量上传学生信息用作分组分配任务 | 教师、学校 |
| 姓名 |
| 年级 |
| 班级 |
| 专业 |
| 组名 | 教师为分组设置组名组号与组长，并定义任务状态 | 教师依据情况自行设定 |
| 分组编号 |
| 组长 |
| 执行任务 |
| 任务状态 |
| 任务名称 | 教师新建一个数据挖掘实践任务为其填写配置信息，分阶段设立目标与时限 | 教师依据情况自行设定 |
| 任务描述 |
| 任务编号 |
| 挖掘语言 |
| 算法配置 |
| 数据集合 |
| 任务状态 |
| 计划时间 |
| 阶段一目标描述 |
| 阶段一时间 |
| 数据集 | 供实践任务使用 | 教师或学生上传 |

### 输入格式

注册要求：

|  |  |
| --- | --- |
| **学生字段[标识符]** | **输入格式** |
| 用户名 | 用户名4-16位字母、数字、下划线、减号组成 |
| 密码 | 密码必须由6-20个字母、数字、下划线组成 |
| 学号/工号 | 学号必须由字母、数字组成 |
| 姓名 | 姓名只含中文 |
| 年级 | 年级不能为空 |
| 班级 | 班级不能为空 |
| 专业 | 专业不能为空 |

数据集上传要求：

|  |  |
| --- | --- |
| **数据集字段[标识符]** | **输入格式** |
| 数据集名称 | 255位数字、字母、下划线组合，下划线不可出现在数据集名称第一位 |
| 描述摘要 | 255位数字、字母、下划线组合 |
| 数据特征 | Multivariate-多变量、Univariate-单变量、Sequential-按序、Time-Series-时序、Text-文本、Domain-Theory-畴理论 |
| 属性类型 | Categorical-分类的、Integer-整型、Real-实数 |
| 相关任务 | Classification-分类、Regression-回归、Clustering-聚类、Other-其他 |
| 实例数 | 45位数字 |
| 允许缺省 | Yes/NO |
| 访问量 | 45位数字 |
| 相关论文 | 255位数字、字母、下划线及符号组合 |

学生信息导入模板：

|  |  |
| --- | --- |
| **学生字段[标识符]** | **输入格式** |
| 学号 | 学号必须由字母、数字组成 |
| 姓名 | 姓名只含中文 |
| 年级 | 年级不能为空 |
| 班级 | 班级不能为空 |
| 专业 | 专业不能为空 |

一键分组要求：

|  |  |
| --- | --- |
| **分组字段[标识符]** | **输入格式** |
| 学号 | 学号必须由字母、数字组成 |
| 姓名 | 姓名只含中文 |
| 年级 | 年级不能为空 |
| 班级 | 班级不能为空 |
| 专业 | 专业不能为空 |
| 组名 | 数字、字母 |
| 分组编号 | 11位纯数字长度之内编号 |
| 组长 | 10个汉字字母以内 |
| 执行任务 | 已有任务范围之内选择 |
| 任务状态 | 待分配、任务需推进、任务完成、任务进行中、任务未完成、新任务、新的提交 |

新建任务输入要求：

|  |  |
| --- | --- |
| **任务字段[标识符]** | **输入格式** |
| 任务名称 | 32位数字、字母、下划线组合，下划线不可出现在数据集名称第一位 |
| 任务描述 | 255位数字、字母、下划线组合 |
| 任务编号 | 11位纯数字长度之内编号 |
| 挖掘语言 | Java、C++、C、Go、Python、Matlab、Scala |
| 算法配置 | C4.5、k均值聚类算法、支持向量机、  Apriori关联算法、EM 最大期望算法、PageRank算法、AdaBoost 迭代算法、kNN：k最近邻算法、CART 分类算法 |
| 数据集合 | 已有数据集范围之内选择 |
| 任务状态 | 待分配、任务需推进、任务完成、任务进行中、任务未完成、新任务、新的提交 |
| 计划时间 | 2xxx年xx月xx日-2xxx年xx月xx日 |
| 阶段一目标描述 | 255位数字、字母组合 |
| 阶段一时间 | 2xxx年xx月xx日-2xxx年xx月xx日 |

数据挖掘分析要求：

|  |  |
| --- | --- |
| **项数据集字段[标识符]** | **输入格式** |
| 数据集 | 文件后缀为.exc或.csv的数据集文件 |

### 输入举例

注册要求：

|  |  |  |
| --- | --- | --- |
| **学生字段[标识符]** | **输入示例** | **是否缺省** |
| 用户名 | Admin | NO |
| 密码 | 1234567 | NO |
| 学号/工号 | 1915106840330 | NO |
| 姓名 | 潘少军 | NO |
| 年级 | 2015级 | NO |
| 班级 | 9151062301 | NO |
| 专业 | 软件工程 | NO |

数据集上传要求：

|  |  |  |
| --- | --- | --- |
| **数据集字段[标识符]** | **输入示例** | **是否缺省** |
| 数据集名称 | Autistic Spectrum Disorder Screening Data for Adult | NO |
| 描述摘要 | Autistic Spectrum Disorder (ASD) is a neurodevelopment condition associated with significant healthcare costs, and early diagnosis can significantly reduce these. Unfortunately, waiting times for an ASD diagnosis are lengthy and procedures are not cost effective. | YES |
| 数据特征 | Multivariate-多变量 | YES |
| 属性类型 | Categorical-分类的 | YES |
| 相关任务 | Regression-回归 | YES |
| 实例数 | 701 | YES |
| 允许缺省 | Yes/NO | YES |
| 访问量 | 21 | YES |
| 相关论文 | 1) Tabtah, F. (2017). Autism Spectrum Disorder Screening: Machine Learning Adaptation and DSM-5 Fulfillment. Proceedings of the 1st International Conference on Medical and Health Informatics 2017, pp.1-6. Taichung City, Taiwan, ACM.  2) Thabtah, F. (2017). ASDTests. A mobile app for ASD screening. www.asdtests.com [accessed December 20th, 2017]. | YES |

学生信息导入模板：

|  |  |  |
| --- | --- | --- |
| **学生字段[标识符]** | **输入示例** | **是否缺省** |
| 学号 | 915106840330 | NO |
| 姓名 | 潘少军 | NO |
| 年级 | 2015级 | YES |
| 班级 | 9151062301 | YES |
| 专业 | 软件工程 | YES |

一键分组要求：

|  |  |  |
| --- | --- | --- |
| **分组字段[标识符]** | **输入示例** | **是否缺省** |
| 学号 | 915106840401  915106840337  915106840345  915106840344 | NO |
| 姓名 | 陈清梅  文皓宇  郑鑫宇  张书豪 | NO |
| 年级 | 2015级 | NO |
| 班级 | 9151069501  9151065801  9151065801  9151069501 | NO |
| 专业 | 软件工程 | NO |
| 组名 | 对自闭症的探索小组 | NO |
| 分组编号 | 01 | NO |
| 组长 | 陈清梅 | NO |
| 执行任务 | 孤独症谱系障碍儿童与成人筛查数据 | YES |

新建任务输入要求：

|  |  |  |
| --- | --- | --- |
| **任务字段[标识符]** | **输入示例** | **是否缺省** |
| 任务名称 | 印度新德里癌症患者平均年龄与五年存活率的关系 | NO |
| 任务描述 | 根据近五年印度新德里一所医院的癌症患者年龄与后续医疗状况的信息数据集来探究患者平均年龄与其五年存活率是否存在关联及这种关联是一种什么样的关系。 | YES |
| 任务编号 | 01 | NO |
| 挖掘语言 | Java、C++ | YES |
| 算法配置 | k均值聚类算法、Apriori关联算法 | YES |
| 数据集合 | pima-indians-diabetes.data.csv | YES |
| 任务状态 | 任务需推进 | YES |
| 计划时间 | 2018年8月15日-2018年8月22日 | NO |
| 阶段一目标描述 | 建立线性回归方程。 | NO |
| 阶段一时间 | 2018年8月15日-2018年8月22日 | NO |

数据挖掘分析要求：

|  |  |  |
| --- | --- | --- |
| **数据集字段[标识符]** | **输入示例** | **是否缺省** |
| 数据集 | Autism-Adult-Data.arff  pima-indians-diabetes.data.csv | NO |

## 输出

注册生成账号：

|  |  |
| --- | --- |
| **学生字段[标识符]** | **输出精度** |
| 用户名 | 字母、数字、下划线 |
| 密码 | 字母、数字、下划线 |
| 学号/工号 | 数字、字母 |
| 姓名 | 汉字 |
| 年级 | 年份 |
| 班级 | 数字、字母 |
| 专业 | 计算机科学与技术、软件工程、人工智能、  网络工程之内 |

一键分组输出：

|  |  |
| --- | --- |
| **分组字段[标识符]** | **输出精度** |
| 学号 | 数字、字母 |
| 姓名 | 汉字、字母 |
| 年级 | 年份 |
| 班级 | 数字、字母 |
| 专业 | 计算机科学与技术、软件工程、人工智能、  网络工程 |
| 组名 | 数字、字母 |
| 分组编号 | 数字 |
| 组长 | 汉字、字母 |
| 执行任务 | 已有任务范围之内 |
| 任务状态 | 待分配、任务需推进、任务完成、任务进行中、任务未完成、新任务、新的提交 |

新建任务输出：

|  |  |
| --- | --- |
| **任务字段[标识符]** | **输出精度** |
| 任务名称 | 数字、字母、下划线 |
| 任务描述 | 数字、字母、下划线 |
| 任务编号 | 数字 |
| 挖掘语言 | Java、C++、C、Go、Python、Matlab、Scala |
| 算法配置 | C4.5、k均值聚类算法、支持向量机、  Apriori关联算法、EM 最大期望算法、PageRank算法、AdaBoost 迭代算法、kNN：k最近邻算法、CART 分类算法 |
| 数据集合 | 已有数据集范围之内 |
| 任务状态 | 待分配、任务需推进、任务完成、任务进行中、任务未完成、新任务、新的提交 |
| 计划时间 | 年月日-年月日 |
| 阶段一目标描述 | 数字、字母 |
| 阶段一时间 | 年月日-年月日 |

数据挖掘分析输出：

|  |  |
| --- | --- |
| **数据集字段 [标识符]** | **输出精度** |
| 数据集各个属性的统计值 | 小数点后一位 |
| 数据集各个属性的回归值 | 小数点后一位 |
| 数据集各个属性的三维特征分布 | 小数点后一位 |
| 数据集各个属性的直方分布 | 小数点后一位 |
| 数据集各个属性的密度分布 | 小数点后一位 |

### 输出数据的现实背景

|  |  |
| --- | --- |
| **输出数据** | **现实背景** |
| 用户名 | 用户注册以后产生一个账号。 |
| 密码 |
| 学号/工号 |
| 姓名 |
| 年级 |
| 班级 |
| 专业 |
| 学号 | 教师选择一部分学生进行分组以后产生的分组信息。 |
| 姓名 |
| 年级 |
| 班级 |
| 专业 |
| 组名 |
| 分组编号 |
| 组长 |
| 执行任务 |
| 任务状态 |
| 任务名称 | 教师新建一个数据挖掘实践任务后产生的任务相关信息和数据集、算法配置信息和分阶段目标时间信息。 |
| 任务描述 |
| 任务编号 |
| 挖掘语言 |
| 算法配置 |
| 数据集合 |
| 任务状态 |
| 计划时间 |
| 阶段一目标描述 |
| 阶段一时间 |
| 阶段二目标描述 |
| 阶段二时间 |
| 数据集各个属性的统计值 | 数据集各个属性的统计值、回归值、三维特征分布、直方分布、密度分布等从数值上反映数据集特征及内部关联。 |
| 数据集各个属性的回归值 |
| 数据集各个属性的三维特征分布 |
| 数据集各个属性的直方分布 |
| 数据集各个属性的密度分布 |

### 输出格式

注册生成账号：

|  |  |
| --- | --- |
| **学生字段[标识符]** | **输出格式** |
| 用户名 | 4-16位字母、数字、下划线组合，下划线不可出现在用户名第一位 |
| 密码 | 6-20位字母、数字、下划线组合，下划线不可出现在密码第一位 |
| 学号/工号 | 11位纯数字或第五位为字母其他位为数字的字串/8位纯数字工号 |
| 姓名 | 10个汉字字母以内 |
| 年级 | 入学年份+级 |
| 班级 | 9位纯数字或第五位为字母其他位为数字的字串 |
| 专业 | 计算机科学与技术、软件工程、人工智能、网络工程范围之内 |

一键分组输出：

|  |  |
| --- | --- |
| **分组字段[标识符]** | **输出格式** |
| 学号 | 11位纯数字或第五位为字母其他位为  数字的字串 |
| 姓名 | 10个汉字字母以内 |
| 年级 | 入学年份+级 |
| 班级 | 9位纯数字或第五位为字母其他位为数字的字串 |
| 专业 | 计算机科学与技术、软件工程、人工智能、  网络工程 |
| 组名 | 数字、字母 |
| 分组编号 | 11位纯数字长度之内编号 |
| 组长 | 10个汉字字母以内 |
| 执行任务 | 已有任务范围之内选择 |
| 任务状态 | 待分配、任务需推进、任务完成、任务进行中、任务未完成、新任务、新的提交 |

新建任务输出：

|  |  |
| --- | --- |
| **任务字段[标识符]** | **输出格式** |
| 任务名称 | 32位数字、字母、下划线组合，下划线不可出现在数据集名称第一位 |
| 任务描述 | 255位数字、字母、下划线组合 |
| 任务编号 | 11位纯数字长度之内编号 |
| 挖掘语言 | Java、C++、C、Go、Python、Matlab、Scala |
| 算法配置 | C4.5、k均值聚类算法、支持向量机、  Apriori关联算法、EM 最大期望算法、PageRank算法、AdaBoost 迭代算法、kNN：k最近邻算法、CART 分类算法 |
| 数据集合 | 已有数据集范围之内选择 |
| 任务状态 | 待分配、任务需推进、任务完成、任务进行中、任务未完成、新任务、新的提交 |
| 计划时间 | 2xxx年xx月xx日-2xxx年xx月xx日 |
| 阶段一目标描述 | 255位数字、字母组合 |
| 阶段一时间 | 2xxx年xx月xx日-2xxx年xx月xx日 |

数据挖掘分析输出：

|  |  |
| --- | --- |
| **数据集字段[标识符]** | **输出格式** |
| 数据集各个属性的统计值 | 一位小数数值 |
| 数据集各个属性的回归值 | 一位小数数值 |
| 数据集各个属性的三维特征分布 | 图表 |
| 数据集各个属性的直方分布 | 图表 |
| 数据集各个属性的密度分布 | 图表 |

### 输出举例

注册生成账号：

|  |  |
| --- | --- |
| **学生字段[标识符]** | **输出示例** |
| 用户名 | Admin |
| 密码 | 1234567 |
| 学号/工号 | 1915106840330 |
| 姓名 | 潘少军 |
| 年级 | 2015级 |
| 班级 | 9151062301 |
| 专业 | 软件工程 |

一键分组输出：

|  |  |
| --- | --- |
| **分组字段[标识符]** | **输出示例** |
| 学号 | 915106840401  915106840337  915106840345  915106840344 |
| 姓名 | 陈清梅  文皓宇  郑鑫宇  张书豪 |
| 年级 | 2015级 |
| 班级 | 9151069501  9151065801  9151065801  9151069501 |
| 专业 | 软件工程 |
| 组名 | 对自闭症的探索小组 |
| 分组编号 | 01 |
| 组长 | 陈清梅 |
| 执行任务 | 孤独症谱系障碍儿童与成人筛查数据 |
| 任务状态 | 915106840401  915106840337  915106840345  915106840344 |

新建任务输出：

|  |  |
| --- | --- |
| **任务字段[标识符]** | **输出示例** |
| 任务名称 | 印度新德里癌症患者平均年龄与五年存活率的关系 |
| 任务描述 | 根据近五年印度新德里一所医院的癌症患者年龄与后续医疗状况的信息数据集来探究患者平均年龄与其五年存活率是否存在关联及这种关联是一种什么样的关系。 |
| 任务编号 | 01 |
| 挖掘语言 | Java、C++ |
| 算法配置 | k均值聚类算法、Apriori关联算法 |
| 数据集合 | pima-indians-diabetes.data.csv |
| 任务状态 | 任务需推进 |
| 计划时间 | 2018年8月15日-2018年8月22日 |
| 阶段一目标描述 | 建立线性回归方程。 |
| 阶段一时间 | 2018年8月15日-2018年8月22日 |

数据挖掘分析输出：

|  |  |
| --- | --- |
| **数据集字段[标识符]** | **输出示例** |
| 数据集各个属性的统计值 | 465 |
| 数据集各个属性的回归值 | 7.334 |
| 数据集各个属性的三维特征分布 | 图表 |
| 数据集各个属性的直方分布 | 图表 |
| 数据集各个属性的密度分布 | 图表 |

## 出错处理和恢复

当用户在使用系统中操作太过频繁或切换页面过于迅速，系统会弹出“操作太过频繁，请稍后重试”的提示窗口，如果继续频繁操作用户需刷新系统才可继续正常使用。

## 登录与注销

当用户在使用系统中长时间（超过3小时）未进行操作或无动作，当前页面将出现会话过期，与后端的连接将自动中断，用户需重新登陆。