**杭州电子科技大学计算机学院**

**数据仓库与数据挖掘（魏丹老师授课班级）**

**课程大作业**

# 一、实验目的

1. 将数据挖掘的概念具体化，以巩固和补充数据挖掘技术的基本原理；

2．利用数据挖掘技术来解决现实中相关的问题；

3. 锻炼分析问题、解决问题及动手实践能力。

# 二、组织：

以小组协作方式进行。

# 三、评分标准及注意事项：

(1)本次实践题目包含一个实验题目，程序完成调试通过后，完成项目对应的**文档**（参照“**数据仓库与数据挖掘课程项目报告**”完成），**文档文件命名为：组号+ \*\*\* ；**

**(2)验收完成后**将**组内评分表、答辩PPT**、**课程项目报告**、**项目源文件和所用数据集**打包并提交；

**(3)提交网址是：**https://www.wjx.cn/jq/46478225.aspx

# 四、实验内容：

## 实验题目

* 课题自选，选择数据挖掘相关课题，可参考以下：
* **分类**
* **预测**

房价，贷款，交通，空气质量，微博等

* **推荐**

新闻，音乐，广告，产品，电影，购物等

* **参考竞赛**

Kaggle <https://www.kaggle.com/competitions>

阿里云-天池 <https://tianchi.aliyun.com/>

DataCastle <http://www.dcjingsai.com/>

DataFountain [http://www.datafountain.cn/#/competitions](http://www.datafountain.cn/)

* 实施方式：以小组协作方式进行，自由组合分组，每个小组不超过4人，自设一名组长组织小组的分工合作；
* 课堂演示、答辩、验收；
* 提交材料：①组内评分表 ②答辩PPT ③课程项目报告（要求：格式规范、语句通顺、内容完整、数据详实）④项目源文件和所用数据集
* 课程项目要求：

1. 综合应用数据挖掘技术（数据预处理、分类、聚类、关联规则等）解决现实中相关的问题；
2. 结果可视化；
3. 程序代码必须具有可读性，关键性变量、算法或者代码必须要有清楚的注释。

* **最后评定：**

1、根据平时表现(平时考勤和课堂表现)、上机作业（上机实验和实验报告）期末项目（问题规模、技术难度、实现程度、答辩情况、结果完整性和准确性）及设计文档（格式规范、语句通顺、报告内容完整、数据详实）综合评定成绩。

2、程序雷同者，视情况扣分或记零分。

****

**计算机学院**

**数据仓库与数据挖掘课程项目报告**

**设计题目**：

**学生姓名**：

**学 号**：

**指导教师**：

**目 录**

1.项目背景 5

2.项目分工 5

3.实验环境 5

4.主要设计思想 5

4.1 实验工具介绍 5

4.2数据预处理方法设计 5

4.3数据挖掘算法的选择与设计 5

4.4性能评估方法 6

5.项目实验过程 6

5.1 已实现的功能点 6

5.2创新点 6

5.3项目实现的主要步骤 6

5.4建模与测试 6

5.5使用说明 6

6.结果分析 6

7. 收获和建议 7

8.项目源代码（每个部分写明是哪位同学完成） 7

# 1.项目背景

(项目背景功能简介、涉及技术介绍等)

# 2.项目分工

(参与人员及分工)

# 3.实验环境

(操作系统、开发环境)

# 4.主要设计思想

(实验工具介绍、数据预处理方法设计、数据挖掘算法的选择与设计、性能评估方法，可根据项目自行编号)

## 4.1 实验工具介绍

## 4.2数据预处理方法设计

（以文字+图+实现代码的形式详细说明每部分是怎样进行设计，以及设计之后的实现效果,如果没有图，则可以不画。）

## 4.3数据挖掘算法的选择与设计

（以文字+图+实现代码的形式详细说明每部分是怎样进行设计，以及设计之后的实现效果,如果没有图，则可以不画。）

## 4.4性能评估方法

（性能评估指标，例如：准确率、召回率、F-measure等）

# 5.项目实验过程

(项目已实现的**功能点**，**创新点**（项目中方法有自己的创新方法的话，记得写明创新点），**项目实现的主要步骤**（程序运行截图及说明），**建模与测试**，**使用说明**)

## 5.1 已实现的功能点

（以文字+图的形式说明）

## 5.2创新点

(备注：请写出项目中的创新点，例如优化了\*\*\*算法)

## 5.3项目实现的主要步骤

（以文字+程序运行截图的形式说明）

## 5.4建模与测试

（数据集说明，数据挖掘模型等）

## 5.5使用说明

（使用方法和步骤、参数设置等）

# 6.结果分析

# 7. 收获和建议

备注：每个人单独写自己的收获和建议。例如：学到了什么？遇到了什么困难？如何解决？（黄色填充部分，正式文档需删掉）

# 8.项目源代码（每个部分写明是哪位同学完成）