第7章 竞赛网站用户行为分析

教案

**课程名称：**《Python数据分析与应用》

**课程类别：**必修

**适用专业：**大数据技术类相关专业

**总学时：**64学时（其中理论28学时，实验36学时）

**总学分：**4.0学分

**本章学时**：8学时

# 材料清单

* 1. 《Python数据分析与应用（第2版）（微课版）》教材。
  2. 配套PPT。
  3. 引导性提问。
  4. 探究性问题。
  5. 拓展性问题。

# 教学目标与基本要求

### 教学目标

根据竞赛网用户访问的原始数据，在数据中构建用户标签以及对网页进行分类。其次结合实际业务情况构建了聚类特征，最后使用K-Means聚类算法建立聚类分群模型。并对聚类得到的结果分析各个群体的特征，从而结合网站的实际营运情况提出相对应的运营建议。

### 基本要求

1. 熟悉竞赛网站用户行为分析的步骤和流程。
2. 掌握竞赛网站用户数据的预处理方法。
3. 掌握使用K-Means聚类算法的方法。

# 问题

### 引导性提问

引导性提问需要教师根据教材内容和学生实际水平，提出问题，启发引导学生去解决问题，提问，从而达到理解、掌握知识，发展各种能力和提高思想觉悟的目的。

* 1. 为什么要对竞赛网站的用户进行分群呢？
  2. 数据分析是什么？
  3. 数据分析的主要流程是哪些？

### 探究性问题

探究性问题需要教师深入钻研教材的基础上精心设计，提问的角度或者在引导性提问的基础上，从重点、难点问题切入，进行插入式提问。或者是对引导式提问中尚未涉及但在课文中又是重要的问题加以设问。

1. 什么是聚类算法？
2. 为什么要进行数据清洗？
3. 如何进行特征的构建？
4. K-Means聚类算法基本原理是怎么样的？

### 拓展性问题

拓展性问题需要教师深刻理解教材的意义，学生的学习动态后，根据学生学习层次，提出切实可行的关乎实际的可操作问题。亦可以提供拓展资料供学生研习探讨，完成拓展性问题。

1. 还有其他的预处理方法吗？
2. 除了K-Means聚类算法还能举出其他的聚类算法吗？
3. 聚类算法与分类算法的区别是什么？

# 主要知识点、重点与难点

### 主要知识点

1. 访问数据的特征值转换。
2. 竞赛网站用户的识别。
3. 异常访问数据的数据清洗。
4. 竞赛网站网页的分类。
5. 竞赛网站用户的特征构造。
6. K-Means聚类算法原理。
7. 使用K-Means聚类算法对竞赛网站用户进行分群。
8. K-Means聚类模型算法的结果分析。

### 重点

1. 竞赛网站用户异常访问的数据清洗。
2. 竞赛网站用户的特征构造。
3. K-Means聚类算法原理。
4. 使用K-Means聚类算法对竞赛网站用户进行分群。

### 难点

1. 访问数据的特征值转换。
2. 异常访问数据的数据清洗。
3. 使用K-Means聚类算法对竞赛网站用户进行分群。

# 教学过程设计

### 理论教学过程

1. 了解竞赛网站背景。
2. 认识用户行为分析。
3. 熟悉竞赛网站用户行为分析的步骤与流程。
4. 了解预处理竞赛网站用户访问数据的操作方法。
5. 了解K-Means聚类算法。

### 实验教学过程

1. 连接数据库，读取竞赛网站数据。
2. 数据预处理的特征值转换。
3. 数据预处理的用户识别。
4. 数据预处理的数据清洗。
5. 数据预处理的网页分类。
6. 数据预处理的构造特征。
7. 使用K-Means聚类算法进行用户分群。
8. K-Means聚类算法的模型应用。

# 教材与参考资料

### 教材

曾文权，张良均．Python数据分析与应用（第2版）（微课版）[M]．北京：人民邮电出版社．2022．

### 参考资料

[1] 黄红梅，张良均．Python数据分析与应用[M]．北京：人民邮电出版社．2018．

[2] 张良均，谭立云．Python数据分析与挖掘实战（第2版）[M]．北京：机械工业出版社．2019．

[3] 张健，张良均．Python编程基础[M]．北京：人民邮电出版社．2018．