

# 研究生《金融数据挖掘与分析》课程教学大纲

## 一、教师信息

姓名： 傅毅

职称： 副教授

办公室： 316

电话： 021-64322246

电子信箱： fuyi@shnu.edu.cn

答疑时间： 根据学生时间安排

## 二、课程基本信息

课程名称（中文）： 金融数据挖掘与分析

课程名称（英文）： Financial data mining and analysis

课程性质： ☐ 专业必修课 ☐ 基础选修课 ☒ 专业选修课

课程类别\*： ☒ 案例教学 ☐ 实验教学 ☐ 其他

课程代码：

周学时： 3

总学时： 48

学分：

先修课程： Python 的安装与基础语法（要求预习完成）

开设专业： 一年级专硕

## 三、课程简介

本课程以数据挖掘基础理论与具体操作为主要内容，讲述数据挖掘的各主要步骤、算法和应用，并通过对实际案例的分析，帮助学生更加深入地理解常用的数据挖掘模型。本课程的要求学生能够使用 PYTHON 语言数据挖掘平台，通过对学生在获取数据、清理数据、分析数据和知识发现等不同环节的训练，帮助学生掌握在应用场景下独立使用数据挖掘理论与方法解决现实问题的能力。

## 四、课程目标

课程的研究对象主要是各种主流的数据挖掘模型，主要有以下几个方面的内容，分别是 PYTHON 可视化、分类决策树模型、回归决策树模型、聚类模型、关联分析等。学生通过数据挖掘方面的学习，不仅可以掌握数据挖掘的相关理论和

模型，而且可以通过不同案例的实务操作，掌握运用不同模型运用的技巧，为从事数据量化分析的工作提供必要的准备。通过了解我国各个行业转型升级的现状，加强学生的中国特色社会主义道路自信、理论自信、制度自信、文化自信。坚定学生投身社会主义建设的理想。

## 五、教学内容与进度安排

### 前言 数据科学驱动我国各行业的转型与升级（课程思政）

#### 1. 课时数：3

#### 2. 讲授内容或训练技能，重点、难点

讲授：数据科学概念。案例介绍数据科学+新零售；数据科学+医疗；数据科学+金融；数据科学+制造；数据科学+交通。数据科学与数据挖掘的关系。

重点、难点：理解数据科学驱动我国各行业的转型与升级中的作用，理解数据挖掘在数据科学中的重要地位。

#### 3. 学生学习任务

通过了解我国各个行业转型升级的现状，加强学生的中国特色社会主义道路自信、理论自信、制度自信、文化自信。坚定学生投身社会主义建设的理想。

#### 4. 教学方法

案例教学。

#### 5. 课外学习要求

自己查找资料，介绍数据科学给我们生活带来巨大改变的案例。

### 第一章 Python 的数据可视化

#### 1. 课时数：12

#### 2. 讲授内容或训练技能，重点、难点：

##### 1) 讲授内容：

（1）数据的导入导出的学习与应用。

（2）数据生成。

(3) Python 的数据可视化。

2) 重难点:

数据的读取, 数据的变形, 数据的清洗

3. 学生学习任务

1) 数据的合并与分解。

2) 数据的筛选与补充。

3) 数据的生成与可视化。

4. 教学方法

在线完成, 多媒体教学、教师讲授、学生讨论、提问、上机练习。

5. 课外学习要求

对于课堂要求完成数据的整理与实现 (案例: 自选数据可视化)。

## 第二部章 聚类分析 (城市特征分析)

1. 课时数: 3

2. 讲授内容或训练技能, 重点、难点:

1) 讲授内容:

(1) K 均值聚类的基本原理。

(2) 两步聚类的原理。

(3) 密度聚类的原理。

3. 学生学习任务

1) 理解不同聚类方法的原理。

2) 掌握不同聚类方法的区别。

3) 熟悉基于 Python 平台上聚类方法的应用

4. 教学方法

多媒体教学、教师讲授、学生讨论、上机实验。

5. 课外学习要求

学习聚类方法在 Python 平台上的建模 (案例: 城市特征分析)。

## 第三部章 逻辑回归 (信用评分卡设计)

1. 课时数：6
2. 讲授内容或训练技能，重点、难点：
  - 1) 讲授内容：
    - (1) 逻辑回归的基本原理。
    - (2) 评分卡设计的原理。
    - (3) WOE 计算的原理。
3. 学生学习任务
  - 1) 理解逻辑回归方法。
  - 2) 掌握评分卡设计的方法。
  - 3) 熟悉基于 Python 平台上逻辑回归的应用
4. 教学方法  
多媒体教学、教师讲授、学生讨论、上机实验。
5. 课外学习要求  
学习聚类方法在 Python 平台上的建模（案例：评分卡设计）。

#### 第四章 决策树 （银行贷款违约识别与车位估值）

1. 课时数：6
2. 讲授内容或训练技能，重点、难点：
  - 1) 讲授内容：
    - (1) 分类决策树算法。
    - (2) 回归决策树算法。
  - 2) 重难点：  
算法的基本原理和推导。
3. 学生学习任务
  - 1) 掌握不同算法的推导。
  - 2) 掌握不同决策树模型在 Python 上的操作和选项设置。
4. 教学方法  
多媒体教学、教师讲授、学生讨论、上机实验
5. 课外学习要求

决策树模型在 Python 平台上的建模与实现（案例：银行贷款违约识别、车位估值）。

## 第五章 关联分析 （产品的智能推荐）

1. 课时数：6

2. 讲授内容或训练技能，重点、难点：

1) 讲授内容：

- （1）理解关联分析的应用背景。
- （2）学习关联分析的基本原理。
- （3）关联分析的主要指标：支持度，可信度，提升度。
- （4）关联规则的简化。

2) 重难点：

高频集的推导，关联规则的简化。

3. 学生学习任务

- 1) 理解关联规则的基本应用。
- 2) 掌握高频集的推导
- 3) 关联分析的代码实现。

4. 教学方法

多媒体教学、教师讲授、学生讨论、上机实验。

5. 课外学习要求

课外上机练习（案例：产品的智能推荐）。

## 第六章 神经网络 （数据挖掘项目的完整步骤）

1. 课时数：6

2. 讲授内容或训练技能，重点、难点：

1) 讲授内容：

- （1）理解神经网络的应用背景。
- （2）学习神经网络的基本原理。
- （3）熟悉数据挖掘流程中的各个步骤。

2) 重难点:

非对称问题处理, 哑变量处理。

3. 学生学习任务

1) 理解神经网络的基本应用。

2) 掌握激活函数的使用方法

3) MLP 的代码实现。

4. 教学方法

多媒体教学、教师讲授、学生讨论、上机实验。

5. 课外学习要求

课外上机练习(案例: **数据挖掘项目的完整步骤**)。

**作业汇报: 期中期末各一次(共 6 课时)**

注: 根据学生的实际情况, 内容和课时适当调整。

## 六、修读要求

学生在修读期间需遵守学校与学院的各项课堂要求, 认证完成作业, 掌握实际问题向理论模型迁移的能力, 学会知识的场景迁移和实际应用。

## 七、学习评价方案

**最终成绩=过程成绩(60%)(其中: 考勤分 10%, 作业等共 50%)+期末成绩 40%。**

## 八、课程资源

(教材; 补充材料和扩展阅读; 专业资源网站; 课件; 作业交流、学习讨论、思考题等。)

(一) 教材

自编讲义

(二) 参考书目

参考书:

[1] 陈封能, 斯坦巴赫, 库玛尔 著, 数据挖掘导论, 人民邮电出版社, 2011。

[2] Jiawei Han, Micheling Kamber, Jian Pei, 数据挖掘概念与技术, 机械工业出版社, 2012。

[3] Luís Torgo, 数据挖掘与 R 语言, 机械工业出版社, 2013。

[4] 薛薇, 基于 SPSS Modeler 的数据挖掘, 中国人民大学出版社, 2014。

[5] [美] 埃里克·马瑟斯 (Eric Matthes) 著, 袁国忠 译, Python 编程 从入门到实践, 人民邮电出版社, 2016。

[6] [美] 韦斯·麦金尼 (Wes McKinney) 著, 徐敬一 译, 利用 Python 进行数据分析, 机械工业出版社, 2018。

## 九、其他需要说明的事宜

无