# 课程大纲
# COURSE SYLLABUS

| | | |
|---|---|---|
| **1.** | **课程代码/名称**<br>**Course Code/Title** | **金融数据挖掘 Financial Data Mining** |
| **2.** | **课程性质**<br>**Compulsory/Elective** | 选修课 Elective Courses |
| **3.** | **课程学分/学时**<br>**Course Credit/Hours** | 3/48 |
| **4.** | **授课语言**<br>**Teaching Language** | 中英双语 English & Chinese |
| **5.** | **授课教师**<br>**Instructor(s)** | 陈琨 Chen Kun |
| **6.** | **先修要求**<br>**Pre-requisites** | 无 No |
| **7.** | **教学目标**<br>**Course Objectives** | |
| | 此课程的目的是讲授数据分析以及数据挖掘的基本过程、模型和工具，及其在金融中的应用。此课程将培养学生软件包（如 Excel 和 weka 软件）、编程（如 Python、JAVA）的实用技巧以及通过分析最新的文献来讨论如何在技术领域分析和解决金融数据问题。通过本课程的学习，学生可学习金融科技技术，为从事创新金融行业奠定基础。<br>The course aims to teach students the process, models, and tools for data analytics and data mining in finance. The course will teach students the practical skills to employ software packages (such as Excel and weka) , programming（such as python or JAVA） and apply necessary extensions to analytic framework and tackle financial data analysis problems. The course will equip students the basic skills in social network analysis and models to pursue further study in the Fintech domain. | |
| **8.** | **教学方法**<br>**Teaching Methods** | |
| | 理论课形式主要为课堂讨论研究文献及方法，并辅以研究方法实际操作，共 48 学时。<br>The lecture is mainly for the discussion of research articles and methods, and the practical operation of the research methods is supplemented. There are total 48 hours of lectures. | |
| **9.** | **教学内容**<br>**Course Contents** | |
| | **理论 lecture** | |
| | **Section 1** | **Data type, data quality, data preprocessing**<br>This section explains how to describe data objects from attributes and metrics, explains the detection and correction of data quality problems, and the ideas and methods for preprocessing data. |
| | **Section 2** | **Classification: Basic Concepts, Decision Trees**<br>This section mainly describes the basic concepts of classification and introduces the general approach to solving classification problems, and explains the working principle of the decision tree classification method and the method of establishing the decision tree. |
| | **Section 3** | **Classification: Model Evaluation**<br>This section mainly introduces some commonly methods for evaluating the performance of a classifier, such as retention methods, random subsampling, and cross-validation. |
| | **Section 4** | **Classification: Alternative Techniques**<br>This section mainly explains some alternative classification techniques: Rule-Based Classifier、Nearest-Neighbour classifiers、Bayesian Classifiers、 |

| | | |
|---|---|---|
| | | Artificial Neural Network (ANN)、Support Vector Machine (SVM). |
| | **Section 5** | **Lab-Text Mining and Classification**<br>This section mainly explains the use of different classification algorithms in open sourced software to analyze the textual data, as well as in this chapter, students will learn the features of different classification algorithms. |
| | **Section 6** | **Classification: Literature**<br>Select reading Fintech related literature, especially about classification. |
| | **Section 7** | **Association Analysis: Apriori**<br>This section mainly explains the effective techniques for generating frequent itemsets and rules in the association rule mining algorithm, focusing on the frequent itemsets and rules generation of the Apriori algorithm. |
| | **Section 8** | **Association Analysis: FP-Growth Algorithm**<br>This section mainly introduces other algorithms that generate frequent itemsets—the FP growth algorithm. |
| | **Section 9** | **Association Analysis: Evaluation**<br>This section mainly discusses the evaluation metrics about association analysis. |
| | **Section 10** | **Lab—Basket Analysis**<br>This chapter mainly explains the analysis of dataset using the association rules in open sourced software. |
| | **Section 11** | **Association Analysis: Literature**<br>Select reading Fintech related literature, especially about association analysis. |
| | **Section 12** | **Cluster Analysis: Basic Concepts and Algorithms**<br>This section mainly introduces the K-means, condensed hierarchical clustering algorithm, the basic principles of the DBSCAN algorithm, and the advantages and disadvantages of each algorithm. |
| | **Section 13** | **Cluster Analysis: Cluster Evaluation**<br>This section mainly introduces methods for evaluating clusters generated by clustering algorithms. |
| | **Section 14** | **Lab—Clustering**<br>This section explains how to use K-means to process the real data using Python. |
| | **Section 15** | **Cluster Analysis: Literature**<br>This section mainly discusses the current financial application of cluster analysis technology in practice cases through reading literature. |
| | **Section 16** | **Project Demonstration and Financial Review** |
| **10.** | **课程考核**<br>**Course Assessment** | |
| | 20%平时作业 + 40%期末报告 I + 40%期末报告 II<br>20% Assignments + 40% Final Report Ⅰ + 40% Final Report ⅠⅠ | |
| **11.** | **教材及其它参考资料**<br>**Textbook and Supplementary Readings** | |
| | Pang-Ning Tan, Michael Steinbach, Vipin Kumar, Introduction to Data Mining, posts & telecom press. | |