

《大数据导论》课程

教学内容、教学计划和考核要求

1. 教学内容（专题）

- 1) 大数据的概念 (Understanding of Big Data)
- 2) 大数据分析、动机和目标(Data Analytics and its Motivation)
- 3) 数据预处理 (Data Preprocessing)
- 4) 特征工程 (Feature Engineering)
- 5) 分类模型及算法(Classification)
- 6) 聚类模型及算法(Clustering)
- 7) 线性模型及算法(Linearly models)
- 8) 集成学习(Ensemble Learning)
- 9) 神经网络与深度学习(Neural Network and Deep Learning)
- 10) 推荐系统(Recommendation System)
- 11) 深度强化学习 (Deep enforcement learning)
- 12) 迁移学习 (Transfer Learning)
- 13) 转换器 (Transformer and BERT)
- 14) 因果机器学习 (Causal Machine Learning)
- 15) AI 代理 (AI Agent for Data Analysis)

2. 教学模式：

- 1) 上述专题 2), 3), 4) 为必选, 其它 12 个专题, 每位同学根据自己的兴趣和爱好, 在其中至少选取 5 个专题 (可以另外增加专题)。
- 2) 课堂教学采用“翻转课堂 (Flipped Classroom)”模式, 专题 2), 3), 4) 随机选取同学用 20 分钟左右的时间, 介绍自己对主题中理论、方法、技术及其应用的理解和体会, 教师组织全体同学进行研讨, 课堂允许同学使用大语言模型参与讨论。
- 3) 其它 11 个专题, 同学自愿结合 2—3 人组成一个学习小组, 每组选出一名组长, 每个小组负责主讲一个专题, 要求每位小组成员每次用 20 分钟左右的时间, 相对完整讲解主讲其中的一个小节。
- 4) 主讲内容以围绕专题内容为线索, 所列的问题为引导, 通过阅读中文或英文, 线上或线下的有关文献资料, 总结出每个专题的重要理论、模型、方法和技术, 特别是它们的适用范围, 优势与缺陷。可做 PPT 或其它方式, 以讲清楚为目标。每个专题内容原则上 4-6 学时讲完; 课堂鼓励其它同学, 借助 AI 进行讨论。
- 5) 作业:
A) 在针对上述 3 个必选专题和 5 个 (至少) 自选专题

中，分别撰写每个专题“内容小结”，小结要求用自己的语言、系统化、结构化、图表化地：**归纳和总结**每个专题的重要概念、模型与算法的理论、思想、方法与技术等，它们之间的逻辑关系，所要解决的问题，以及解决问题的思路 and 过程等。重点考察同学**自己的**理解，体会、看法和观点，以及应用的经验等，请勿“抄袭”。

- B)** 针对每个专题（包括必选和自选）内容，选择具有实际业务背景的数据和问题，完成一个微型项目（Mini-projects），撰写相应的微型项目的研究报告。
- C)** 课程结束时，每个小组的完成一个完整的数据分析项目，并撰写研究报告。
- D)** 每个专题内容小结及微型项目报告，请在相应专题研讨结束后的**1 周末之内**递交，递交以 PDF 电子文档，文档名称：姓名+章节名称，递交到

<https://zhbit.fanya.chaoxing.com/portal>

3. 考核要求

课程期末成绩：

课堂讲解（10%）+课堂讨论（5%）+专题小结及微型项目报告（70%）+小组完整项目分析报告（15%）。