

大数据建模与应用实战培训大纲

主讲

王老师

报名电子邮件

Liang1994bj@126.com

培训地点

北京

计划培训时间

2天

课程定制

可根据用户需求调整课程内容。

课程概述

大数据分析、挖掘、建模技术已经逐步地应用到城市管理、医疗、物流、金融、汽车、零售、餐饮、体育、娱乐、新兴互联网企业、证券企业、电信等行业，给这些行业带来了一定的数据价值增值作用。

本次课程面向对大数据分析、挖掘等感兴趣的研究人员，带大家实践大数据分析挖掘与建模系统的项目训练，成体系地讲解：

- (1) 大数据建模、分析挖掘技术体系及其平台方案实现；
- (2) 大数据支持的应用-方法与实际案例讲解；

设备要求

学员需要准备的笔记本配置 i5 及以上 CPU，4GB 及以上内存，硬盘空间预留 50GB（可用移动硬盘）。

培训参考教材

- 《数据挖掘导论（完整版）》书号：9787115241009
- 《大数据技术原理与应用（第2版）》书号：9787115443304
- 《利用 Python 进行数据分析》书号：9787111436737

课程安排

第一天

- 大数据、人工智能与区块链（上午）

1. 大数据 ABCD

- (1) 大数据发展历程
- (2) 大数据的概念
- (3) 大数据处理流程
- (4) 大数据架构-Hadoop 分布式架构
- (5) 大数据架构-spark 框架
- (6) spark 原理
- (7) 我国大数据产业发展现状
- (8) 大数据产业链
- (9) 计算机、人工智能之父图灵：图灵测试
- (10) 人工智能 3 个主义
- (11) 大数据与人工智能的关系
- (12) 2012 年是大数据元年
- (13) 人工智能已有 60 余年历史
- (14) 当前是人工智能第三次浪潮
- (15) 大数据成就了新人工智能时代
- (16) 人工智能发展现状--企业发展趋势
- (17) 人工智能发展现状--企业及融资规模分布
- (18) 人工智能发展现状--专利分布
- (19) 人工智能发展现状--应用领域分布
- (20) 什么是区块链？
- (21) 区块链典型架构与核心技术
- (22) 区块链可被应用于多种场景，但目前还没有成熟商业场景

2. 大数据典型应用

- (1) 谷歌票房预测
 - (2) 谷歌票房预测启示
 - (3) 谷歌流感预测
-

-
- (4) 亚马逊推荐算法
 - (5) 智能风控：金融反欺诈
 - (6) 零售交易反欺诈系统整体设计
 - (7) P2P 平台网民评论（情感分析；2016 建模大赛 1 等奖论文）
 - (8) P2P 平台-加入大数据指标后模型精度提高 5 到 8 个百分点
 - (9) P2P 平台-大数据指标的重要性也被模型所证实
 - (10) 投资：期货价格预测
 - (11) 医疗：肿瘤识别
 - (12) 医疗：AI+基因大数据
 - (13) 安防：人脸识别
 - (14) 大数据在物流领域中的应用
 - (15) 大数据在城市管理中的应用
 - (16) 大数据在金融行业中的应用
 - (17) 大数据在汽车行业中的应用
 - (18) 大数据在零售业中的应用
 - (19) 大数据在餐饮行业中的应用
 - (20) 大数据在能源行业中的应用
 - (21) 大数据在体育和娱乐领域中的应用
 - (22) 大数据在安全领域中的应用
 - (23) 大数据在政府领域中的应用

• 机器学习与 AI 方法：典型应用场景与新进展（下午）

1. 大数据颠覆了传统的理念和技术吗？
 - (1) 大数据确实带来了新的理念和技术体系
 - (2) 也推动了一些新算法的诞生
 - (3) 那么，大数据时代，传统理念还有用吗？
 - (4) 传统技术还有用吗？
2. 常见分类模型
 - (1) Bagging
 - (2) 随机森林 RF

-
- (3) SVM 和 RF
 - (4) Adaboost
 - (5) 近几年 GBDT\XGBOOST\LightGBM 渐成流行
 - (6) 逻辑斯特回归 (LR)
 - (7) KNN
 - (8) 朴素贝叶斯
3. 常见聚类模型
- (1) K 均值聚类
 - (2) 均值漂移算法
 - (3) 基于密度的聚类算法 (DBSCAN)
 - (4) 利用高斯混合模型进行最大期望估计
 - (5) 凝聚层次聚类
4. 常见回归模型
- (1) 线性回归
 - (2) 多项式回归
 - (3) Ridge 回归
 - (4) Lasso 回归
 - (5) ElasticNet 回归
5. 常见关联模型
- (1) 关联分析的基本概念
 - (2) 发现频繁项集
 - (3) 发现关联规则
6. 有监督、无监督和半监督
- (1) 有监督学习
 - (2) 无监督学习
 - (3) 半监督学习

第二天

• 大数据实战--典型场景与解决方案（上午）

1. 精准营销：大数据支持的快速获客
 - (1) 场景还原
 - (2) 解决方案
 - (3) 代码讲解
 - (4) 上机实操
 2. 精准营销：潜在客户识别
 - (1) 场景还原
 - (2) 解决方案
 - (3) 代码讲解
 - (4) 上机实操
 3. 精准营销：精准广告投放
 - (1) 场景还原
 - (2) 解决方案
 - (3) 代码讲解
 - (4) 上机实操
 4. 客户挽留：客户流失预警与挽留
 - (1) 场景还原
 - (2) 解决方案
 - (3) 代码讲解
 - (4) 上机实操
 5. 风险识别 I-大数据支持的反欺诈与失联修复
 - (1) 场景还原
 - (2) 解决方案
 - (3) 代码讲解
 - (4) 上机实操
 6. 风险识别 2-基于机器学习方法的客户风险评估
 - (1) 场景还原
 - (2) 解决方案
-

-
- (3) 代码讲解
 - (4) 上机实操
 - NLP 实战--典型场景与解决方案 (下午)
 - 1. 认识 NLP
 - (1) NLP 起源
 - (2) NLP 基础--词特征
 - (3) NLP 基础--词向量
 - (4) NLP 常见应用
 - ①关键词提取
 - ②文本摘要
 - ③情感分析
 - ④阅读理解
 - ⑤聊天机器人
 - ⑥知识图谱与 KBQA
 - 2. 典型案例：舆情监控与情感分析
 - (1) 场景还原
 - (2) 解决方案
 - (3) 代码讲解
 - 3. 上机实操
 - (4) 上机操作
 - (5) 操作答疑
-

第三天

考试安排

考试方式： 闭卷考试

考试时间： 培训结束后考试，具体日期以公告为准

考试及别和科目： 中级数据分析师

科目一： 中级数据分析方法
科目二： 数据挖掘
科目三： 大数据技术原理与应用
科目四： Python 数据分析
科目五： SQL 数据库应用与实践

初级数据分析师

科目一： 初级数据分析方法
科目二： Excel 数据处理与分析
科目三： 大数据技术概论
科目四： R 语言

专家介绍

王老师：高级统计师、经济学博士、数据挖掘方向博士后、北京大学客座教授，现为某大型金融集团 AI 部门数据挖掘团队负责人，曾在国家统计局工作 10 年，研究方向为金融与医疗大数据挖掘、统计建模、机器学习与深度学习算法，在大数据应用、人工智能领域有丰富实战项目经验。