附件1.

**2021年第一期大数据实战&人工智能应用开发高校师资能力提高线下培训课程表**

|  |  |  |  |
| --- | --- | --- | --- |
| **大数据实战高校师资能力提高培训** | | | |
| **日期** | | **时间**  **安排** | **课程内容** |
| **第一天** | 上午 | 09:00  12:00 | **模块一：大数据理论框架（Hadoop体系）**  **理论讲解：**  1、大数据的架构  2、大数据采集与预处理  3、大数据的存储  4、数据挖掘  **操作演示**  1、Hadoop平台上的数据存储评估  **学员实验**  1、基于Uber出租车数据分析市民行为特征：数据上传Hadoop大数据平台 |
| 下午 | 14:30  17:30 | **模块二：Hadoop大数据分析项目实战**  **理论讲解**   1. 分析决策需求：数据仓库 2. 数据库MySQL的统计分析功能 3. 大数据数据导入工具 4. 大数据的数据仓库工具Hive   5、大数据的数据清洗工具Pig  **实验讲解**   1. Uber出租车数据分析市民行为特征：使用Hive分析离线数据   **学员实验**   1. Uber出租车数据分析市民行为特征：数据提取，数据拆分，数据可视化   **实验验收**  1、作业分享、作业验收、作业讲解、作业点评 |
| **第二天** | 上午 | 09:00  12:00 | **模块三：大数据理论框架（Spark体系）**  **理论讲解**  1、分布式内存计算系统  2、分布式流处理平台  3、Scala语言  4、Spark Core  5、Spark SQL  **实验讲解**  1、Spark RDD编程接口对数据进行分析  2、Spark DataFrame接口对数据进行分析  **学员实验**  1、Uber出租车数据分析市民行为特征：使用Spark-SQL分析Uber市民乘车数据 |
| 下午 | 14:30  17:30 | **模块四：Spark计算任务优化**  **理论讲解**  1、Spark的计算任务提交方式  2、Hadoop分布式计算的原理  3、Spark 任务DAG的构建  4、数据缓存  5、Spark计算任务的Stage  6、Spark计算任务中的依赖  **实验讲解**   1. Uber出租车数据分析市民行为特征：公平调度策略的实现   **学员实验**   1. Uber出租车数据分析市民行为特征：分布式计算资源池优化 |
| **第三天** | 上午 | 09:00  12:00 | **模块五：数据分析与机器学习**  **理论讲解**  1、数据分析基础（Python）  2、多维数据分析方法与特征工程  3、机器学习算法分类（分类、聚类和回归）  4、模型选择与评估  5、数据分析到机器学习  **实验讲解**  1、Python机器学习库实现鸢尾花分类  **学员实验**   1. 鸢尾花分类：使用Spark DataFrame接口导入鸢尾花数据 |
| 下午 | 14:30  17:30 | **模块六：大数据数据挖掘项目实战：**  **项目讲解**  1、Spark机器学习决策树算法  2、Spark机器学习关联规则算法  3、Spark机器学习K-Means算法  4、Spark机器学习PageRank算法  **学员实验**   1. 使用Spark机器学习库实现鸢尾花分类 2. 使用Spark机器学习库预测信用卡欺诈   **实验验收**  1、作业分享、作业验收、作业讲解、作业点评 |
| **第四天** | 上午 | 09:00  12:00 | **Kaggle竞赛—MovieLens电影推荐系统**  **竞赛实战：根据MovieLens评分数据为每个用户推荐10部电影**  **1、问题定义**  1）项目概述  2）问题陈述  3）评价指标  **2、分析**  1）数据的探索  2）探索性可视化  3）算法和技术  4）基准模型 |
| 下午 | 14:30  17:30 | **竞赛实战：推荐系统ALS的原理和代码详解**  **3、实现方法**  1）数据预处理  2）执行过程  3）完善  **4、输出结果**  1）模型的评估与验证  2）合理性分析  **5、项目结论**  1）结果可视化  2）对项目的思考  3）需要做出的改进 |
| **晚上** | 19：00-21：00 | **竞赛项目完成情况点评** |
| 第五天  **第五天** | **上午** | 09:00  12:00 | **企业参观、调研**  **1）大数据行业解决方案与产品展示**  **2）世界技能大赛云计算竞赛平台体验**  **3）大数据系列教材、教学资源和实验平台体验**  **4）大数据1+X认证体系介绍** |
| **下午** | 14:30  17:30 | **结业典礼**  **交流互动 & 结班** |

|  |  |  |  |
| --- | --- | --- | --- |
| **人工智能应用开发高校师资能力提高培训** | | | |
| **日期** | | **时间**  **安排** | **课程内容** |
| **第一天** | 上午 | 09:00  12:00 | **模块一：机器学习入门**  **理论讲解：**  1、课程概述  2、机器学习概述  3、机器学习开发环境  **操作演示**  1、机器学习分类与回归  **实验讲解**  1、快速入门实验：预测乐高玩具套装的价格（线性回归）  **学员实验**  1、波士顿房价预测 |
| 下午 | 14:30  17:30 | **模块二：数据处理、特征工程与数据降维**  **理论讲解**   1. 特征提取和特征工程   2. 数值幅度缩放  3. PCA降维  **实验讲解**  1、电信行业脱敏数据特征提取、维度建模  2、数据统计分析：Python分析企业员工工资表  **学员实验**  1、使用PCA对半导体行业制造数据降维  **实验验收**  1、作业分享、作业验收、作业讲解、作业点评 |
| **第二天** | 上午 | 09:00  12:00 | **模块三：机器学习经典算法—分类算法理论讲解**  1、K近邻算法介绍  2、决策树算法介绍  3、随机森林算法介绍  4、朴素贝叶斯算法介绍  5、Logistic回归与SVM算法介绍  6、应用场景的选择和分类算法误差评估  **实验讲解**  1、使用朴素贝叶斯过滤垃圾邮件  **学员实验**  1、使用K-近邻算法改进交友网站数据匹配效果 |
| 下午 | 14:30  17:30 | **模块四：机器学习经典算法—无监督学习算法与回归算法理论讲解**  1、K-Means聚类算法介绍  2、线性回归算法介绍  3、决策树回归算法介绍  4、随机森林回归介绍  5、关联规则算法介绍  **实验讲解**  1、Apriori算法原理与编程实现  **学员实验**   1. 根据某电商企业的销售数据生成商品推荐规则   **实验验收**  1、作业分享、作业验收、作业讲解、作业点评 |
| **第三天** | 上午 | 09:00  12:00 | **模块五：深度学习概述**  **理论讲解：**  1、神经网络算法原理与应用场景  2、深度学习框架  3、卷积神经网络及其应用  **实验讲解**  1、使用TensorFlow深度学习框架实现分类  **学员实验**  1、基于CNN破解验证码 |
| 下午 | 14:30  17:30 | **模块六：项目实战：手写数字识别**  **理论讲解：**   1. 卷积神经网络的原理 2. 常见的卷积神经网络模型   **实验讲解**   1. CNN卷积神经网络的设计和优化 2. ResNet网络模型构建   **学员实验**   1. 利用卷积神经网络识别手写数字 |
| **第四天** | 上午 | 09:00  12:00 | **Kaggle竞赛实战—猫狗分类检测**  **竞赛实战：猫狗图片识别**  **1、问题定义**  1）项目概述  2）问题陈述  3）评价指标  **2、分析**  1）数据的探索  2）探索性可视化  3）算法和技术  4）基准模型 |
| 下午 | 14:30  17:30 | **综合实战：猫狗图片识别**  **3、方法**  1）数据预处理  2）执行过程  3）完善  **4、结果**  1）模型的评估与验证  2）合理性分析  **5、项目结论**  1）结果可视化  2）对项目的思考  3）需要做出的改进 |
| 晚上 | 19：00-20：30 | **竞赛项目完成情况点评** |
| 第  **第五天** | 上午 | 09:00  12:00 | **企业参观、调研**  **1）人工智能行业解决方案与产品展示**  **2）世界技能大赛云计算竞赛平台体验**  **3）人工智能系列教材、教学资源和实验平台体验**  **4）人工智能认证体系介绍** |
| 下午 | 14:30  17:30 | **交流互动 & 结班** |