

关于举办“大数据高校教师寒假进修班”的通知

一、培训背景

随着互联网普及和移动互联网的突飞猛进，人们已然进入信息爆炸时代，我们已不再为如何获取信息而苦恼，但如何对已获取的信息进行有效处理，即在现有数据资源上进行“数据深加工”，进一步从中挖掘出有价值的内容却困扰着无数人。在此背景下，大数据人才也顺势成为当今社会最受欢迎的群体之一。

但现实是一方面大量社会企业很难寻找到合适的大数据人才，另一方面高校对于如何培养出符合企业需求的人才又深感困扰，因此如何针对性协助高校提升大数据教学实战水平，并使培养出的学生更符合企业实际需要，是本次培训的主要目的。

二、培训对象

对大数据感兴趣人员均可，如本科老师、研究生、本科生，高职老师、学生，以及“泰迪杯”和“全国高职大数据技术与应用比赛”参赛员。

三、培训意义

- 1、了解大数据岗位目前的就业形势、前景及所需相关技能，了解企业实际需求，并参与一个实际项目的全过程；
- 2、掌握 Python 数据挖掘技能和 Hadoop 开发技能，使用相关技能实现算法；
- 3、将培训转化成教学成果，运用到教师自己后续的教学当中去；
- 4、“泰迪杯”数据挖掘大赛辅导、Hadoop 方面全国高职大赛辅导；
- 5、了解大数据专业方面的教材、实验室\实训室建设内容、产品、科研和创新创业方面的合作。

四、主办方信息

广州泰迪智能科技有限公司是广州经济技术开发区内的一家专业从事数据挖掘软件基础研究、培训、咨询服务及数据挖掘应用开发的高新技术企业。公司技术团队成员均具有多年的系统开发与设计和多年数据仓库系统实施、数据挖掘咨询经验。迄今为止，公司共主持或参与国家级和省级重大项目达 10 多个，并拥有发明专利 12 项、软件著作权 18 项、商标 2 项，图书专著 10 本。成功地为中国电信、南方电网、广东电网、交通银行、司法厅、国税局、环保厅等单位提供顾问咨询及项目实施服务。

五、培训特点

本次大数据培训均采用实操方式，老师授课，学员上机操作。此次培训以企业真实案例为背景，让学员充分感受到真实的大数据项目流程及其所需要的技能和知识点，弥补高校在数据分析和大数据实战方面的短板，同时结合“泰迪杯”数据挖掘挑战赛和“全国高职大数据技术与应用比赛”，在实训中给予学员指导。

凡参与本次培训均可得到：1. 培训发票；2. 培训结业证书；3. 大数据教学实训平台试用账号；4. 课程建设与程序设计的相关资源；5. 老师可获得专业建设方案。

六、培训时间

2018 年 1 月 20 日-1 月 26 日（Python 培训），2018 年 1 月 27 日-2 月 2 日（Hadoop 培训），共计 14 天，每天培训 6 小时。

七、培训地点

西安铁路职业技术学院（西安市自强西路）。

八、培训费用

1、培训费用：

Python 7 天实操培训 1980 元\人，学生 1200 元\人；

Hadoop 7 天实操培训 1980 元\人，学生 1200 元\人；

可根据情况二选一，若两个培训均参加 9 折优惠，名额有限，先报先得。

2、报名联系方式：

陈老师：15202918652 张老师：18928815243

九、课程安排：

Python 数据挖掘分析课程

日期	主题	内容	学时
Day1	数据挖掘导论 准备工作	0.0 数据挖掘概论 0.1 数据科学人才培养探索 0.3 广州泰迪智能科技有限公司介绍 1.1 认识 Python 1.2 搭建 Python 环境 1.3 安装 PyCharm 并创建一个应声虫程序	2
	Python 基础知识	2.1 掌握 Python 固定语法 2.2 创建字符串变量并提取里面的数值 2.3 计算圆形的各参数	2
	Python 数据结构	3.1 创建一个列表 (list) 并进行增删改查操作 3.2 转换一个列表为元组 (tuple) 并进行取值操作 3.3 创建一个字典 (dict) 并进行增删改查操作 3.4 将两个列表转换为集合 (set) 并进行集合运算	2
Day2	程序流程控制语句	4.1 实现考试成绩划分 4.2 实现一组数的连加与连乘 4.3 使用冒泡排序法排序 4.4 实训 (猜数字游戏)	2
	函数	5.1 自定义函数实现输出方差 5.2 使用匿名函数添加列表元素 5.3 存储并导入函数模块	2
	文件基础	6.1 认识文件 6.2 读取 txt 文件中的数据 6.3 保存数据为 csv 格式文件 6.4 认识 os 模块	2
Day3	NumPy 数值计算 基础	7.1 认识 NumPy 数组对象 ndarray 7.2 认识 NumPy 矩阵与通用函数	2
	Matplotlib 数据可 视化基础	8.1 了解绘图基础语法与常用参数 8.2 分析特征间的关系 8.3 分析特征内部数据分布与分散状况	2



		8.4 实训：分析 1996-2015 年人口数据特征间关系	
	Pandas 统计分析基础	9.1 读取不通数据源的数据 9.2 掌握 DataFrame 的常用操作 9.3 转换与处理时间序列数据	2
Day4	使用 pandas 进行数据预处理	10.1 合并数据 10.2 清洗数据 10.3 标准化数据 10.4 转换数据	2
	使用 scikit-learn 构建模型	11.1 使用 sklearn 转换器处理数据 11.2 构建并评价聚类模型 11.3 构建并评价分类模型 11.4 构建并评价回归模型	2
	案例一：航空公司客户价值分析	12.1 了解航空公司现状与客户价值分析 12.2 预处理航空客户数据 12.3 使用 K-Means 算法进行客户分群 12.4 聚类结果讨论与评价	2
Day5	神经网络算法	13.1 算法的应用案例 13.2 算法的原理 13.3 算法的实现	3
	智能推荐算法	14.1 算法的应用案例 14.2 算法的原理 14.3 算法的实现	3
Day6	正则表达式	15.1 严格的字符匹配 15.2.正则表达式的广义化 15.3.例子：电话目录的整理 15.4.其他实用函数 15.5.Python 爬虫实现	3
Day6-7	案例二：电子商务网站智能推荐服务	16.1 案例背景 16.2 挖掘目标 16.3 分析方法与过程 16.4 数据预处理 16.5 模型构建 16.6 模型评价与优化	9

Hadoop 大数据挖掘课程

日期	主题	内容	课时
Day1	Hadoop 简介及其生态系统	1.1 Hadoop 简介 1.1.1 大数据基础 1.1.2 Hadoop、分布式文件系统 HDFS、并行计算框架 MapReduce、集群资源管理框架 YARN 简介 1.1.3 Hadoop 使用场景	3



		1.2 Hadoop 生态圈 1.2.1 Hive、HBase、Spark、Pig、Oozie 等简介 1.2.2 Hadoop 发行版：CDH、HDP、MapR 简介	
Day1~Day2	Hadoop 基础操作	2.1 Hadoop 安装配置 2.2 Hadoop 常用监控及集群启动与关闭 2.3 Hadoop 基本 Shell 命令 2.3.1 Hadoop FS 基本命令 2.3.2 Hadoop YARN 基本命令	9
Day3	Hadoop 编程基础	3.1 Hadoop Mapper、Reducer、Driver 原理及其实现分析 3.2 Hadoop 基础数据类型 3.3 Hadoop 输入输出类型	6
Day4	Hadoop 编程进阶	4.1 编程基础之 Combiner、Partitioner、FileInputFormat、FileOutputFormat、自定义键值对 4.2 MapReduce 实现 KMeans 算法及编程优化 4.3 Hadoop 性能调优-Combiner 使用 4.4 Hadoop 分区器 Partitioner 使用 4.5 Hadoop 自定义键值对实现高效编程	6
Day5	Hadoop MapReduce 算法实现	5.1 鸢尾花分类项目背景 5.2 KNN 算法原理分析 5.3 MapReduce 实现 KNN 算法	3
Day5~Day6	Hive 简介&架构	6.1 Hive 安装配置 6.2 Hive 三种客户端介绍 6.3 Hive 架构分析	4
Day6	Hive DDL 及 DML 操作	7.1 常用 Hive 命令介绍 7.2 Hive 常用查询分析	5
Day7	大数据企业案例分析	基于 Hive 的 SOGO 大数据分析	6



“大数据高校教师寒假进修班” 报名申请表

单位名称					
部门/院系					
通讯地址			邮编		
联系人		电话		邮箱	
姓名	性别	职务	身份证号码	手机号码	电子邮箱
是否需要安排住宿					
费用支付备注	1、请将报名表邮件至会务处，培训费报名后电汇至指定帐号，汇款说明底单上注明“大数据培训”字样，方便查账备案。将汇款底单传真或电邮至我处。 2、报到时现金或刷卡支付。				
银行信息	开户名	广州泰迪智能科技有限公司			
	开户行	中国工商银行广州花城支行			
	帐号	3602028509201663221			
联系人	电话	张老师：18928815243			
		陈老师：15202918652			
	邮箱	zhangwenjie@tipdm.com			
对本次培训关注的问题及相关建议、意见					