

关于举办“Hadoop 大数据及海量数据挖掘应用工程师培训班”的通知

随着云时代的来临，大数据技术将具有越来越重要的战略意义。大数据分析 & 挖掘技术已经渗透到每一个行业和业务职能领域，逐渐成为重要的生产要素，人们对于海量数据的运用将预示着新一轮生产率增长和消费者盈余浪潮的到来。大数据的出现催生出产业人才缺口瓶颈，在大数据挖掘项目的实施方面，被调查公司普遍缺乏相关的技术能力。75%以上的公司表示在人员和培训方面存在障碍，会大数据挖掘技术的人才很热门，但是比较难找而且昂贵，会 Hadoop 技术的数据挖掘人才更是奇缺。虽然大数据目前在国内还处于初级阶段，但是其商业价值已经显现出来，特别是有实践经验的大数据分析人才更是各企业争夺的热门。

广州泰迪智能科技有限公司和深圳市飞博远创科技有限公司联合开展“Hadoop 大数据及海量数据挖掘应用工程师培训班”。培训班的具体时间将公布在公司官网，望大家积极报名参加。

一、课程适用对象：

数学类、统计类、计算机类、经济金融类、物流运输等专业的在校学生；生物信息类、医学类、农业类、环境类、信息管理类等专业的研究生；需求分析及系统设计人员；数据挖掘开发人员；进行数据挖掘应用研究的科研人员；关注高级数据分析的人员等。

二、培训时间和地点：

（一）集中班：

上课时间为周一至周五，每天 9：00～17：00。

时间：广州开课时间：2015 年 07 月 20 日～2015 年 08 月 26 日

地点：广州科学城科珠路 232 号益民科技园（科汇金谷对面）

（二）周末班

时间：2015 年 06 月 27 日～2015 年 09 月 27 日期间的周六、周日。

地点： 深圳南山区科苑北科兴科学园

三、培训内容：

培训内容分以下两个阶段：

第一教学实践阶段：

全真案例实训，课程由企业开发，主讲教师及助教由公司委派。

实训场地：广州、深圳

第二项目实战阶段：

公司提供项目实习机会，学员自主选择是否参与项目实习。如果学员选择实习，学员将被分配到各个项目小组中，由项目经理分派数据挖掘建模任务，由学员独立或合作完成。

实训场地：泰迪科技或客户现场进行项目实施。

时间周期：约 4 个月，一般为 5 天工作制。

培训内容包含以下模块：

培训课程目录					
科目	课程模块	课程主要内容	理论学时	实践学时	备注
理论	数据挖掘基本理论	1、数据挖掘概述 2、数据探索:数据质量探索与数据特征分析 3、数据预处理：数据清洗、数据集成、数据变换、数据规约 4、数据挖掘建模：分类与预测、聚类分析、关联规则、时序模式、离群点检测、智能推荐	16	-	
	hadoop 基础	1、Hadoop 概述与 Mapreduce 简介 2、Hadoop 分布式文件系统与 Hadoop 的输入输出 3、Hadoop 应用开发与 MapReduce 的工作原理 4、MapReduce 的类型与格式、MapReduce 的类库 5、Hadoop 集群的安装与 Hadoop 的管理 6、Pig、Hive、Hbase、ZooKeeper 简介	8	8	
工具	MATLAB	1、MATLAB 概予与 MATLAB 基本语法 2、 MATLAB 数据分析工具箱介绍	4	4	网络课程 请自备电脑
	SPSS Modeler	1、 SPSS 的基本操作 2、 SPSS 函数详解与 SPSS 编程 3、 SPSS 统计分析与报表制作 4、 SPSS 数据挖掘算法：回归、聚类、决策树、神经网络、关联规则等	4	4	网络课程 请自备电脑

	R	1、 R 概述与基本操作 2、 基础统计与统计绘图 3、 统计模型 4、 数据挖掘与文本挖掘	4	4	网络课程 请自备电脑
	SAS	1、 SAS 程序基础篇：SAS 总体概述、SAS 编程基本语法 2、 SAS 统计篇：检验、方差、相关分析、回归分析、因子分析、聚类分析、关联规则 3、 SAS/EM 模块：执行数据挖掘	4	4	网络课程 请自备电脑
案例	电力窃漏电用户自动识别	通过采集电力计量自动化系统的电流、电压、功率因数等用电负荷数据，用电异常等终端报警信息以及电力营销系统提供的历史窃漏电用例，归纳出窃漏电用户的关键特征，构建出窃漏电用户的识别模型，最终通过监测用户用电数据，实现窃漏电用户的实时在线诊断。	8	8	实践演练环境： MATLAB、SAS、SPSS、R
	航空公司客户价值分析	航空市场竞争激烈，某航空公司面临着常旅客流失、竞争力下降、航空资源未充分利用等经营危机。通过积累的大量的会员档案信息和其乘坐航班记录，建立合理的客户价值评估模型对客户进行分群，分析比较不同客户群的客户价值，对不同价值的客户类别提供个性化服务，并制定相应的营销策略。	8	8	实践演练环境： MATLAB、SAS、SPSS、R
	中医证型的关联规则挖掘	借助三阴乳腺癌患者的病理信息，发现中医症状间的关联关系和诸多症状间的规律性，挖掘潜在证素，并且依据规则分析病因、预测病情发展以及为未来临床诊治提供有效借鉴。这样患者在治疗过程中，医生可以有效的减少西医以及化疗治疗的毒副作用，为后续治疗打下基础。	8	8	实践演练环境： MATLAB、SAS、SPSS、R
	基于水色图像的水质评价	由于水色能反映水中浮游植物的种类和多少，因此可通过观察水色变化来调控水质，维持养殖水体生态系统中浮游植物、微生物类、浮游动物等合理的动态平衡。通过历史水产专家经验判断结果以及用数码相机按照标准进行水色采集的数据，利用图像处理技术，实现基于水色图像的水质自动评价。	8	8	实践演练环境： MATLAB、SAS、SPSS、R
	基于关联规则的电子商务智能推荐服务	根据电子商务网站采集到的用户访问数据，对用户的访问内容作关联规则分析，探寻用户访问网站的业务逻辑关系，为用户推荐精确的知识内容与服务，从而建立稳定的网站忠实顾客群，实现客户链式反应增值，为企业实现更为人性化的客户定位，提高消费者满意度。	8	8	实践演练环境： MATLAB、SAS、SPSS、R
	应用系统负载分析及磁盘容量预测	与应用系统关联的任何一种资源负载过大，都可能会引起系统性能下降甚至瘫痪。通过分析磁盘容量相关数据，预测应用系统服务器磁盘空间是否满足系统健康运行的要求。根据用户需求设置不同的预警等级，将预测值与容量值进行比较，对其结果进行预警判断，为系统管理员提供定制化的预警提示	8	8	实践演练环境： MATLAB、SAS、SPSS、R
	面向网络	网络舆情对政治生活秩序和社会稳定的影响与日俱增，	8	8	实践演练

	舆情的关联度分析	一些重大的网络舆情事件使人们开始认识到网络对社会监督起到的巨大作用。通过从社区网站上采集到网络舆情信息和分析对象信息，建立分析对象与舆情资源之间关联度的计算规则，寻找出与给定舆情资源联系最紧密的分析对象。			环境： MATLAB、SAS、SPSS、R
	家用电器用户行为分析及事件识别	家用电器在使用过程中，因地区气候、区域不同、用户年龄性别差异，形成不同的机组使用行为，称之为用户的使用习惯。根据用户使用热水器的使用行为构建出洗浴事件识别模型，进而深入了解热水器在用户所在区域的真实使用习惯，优化新产品的设计和布局营销。	8	8	实践演练环境： MATLAB、SAS、SPSS、R
	基于基站定位数据的商圈分析	商圈是现代市场中企业市场活动的空间，也是商品和服务享用者的区域。商圈划分的目的之一是为了研究潜在的顾客的分布以制定适宜的商业对策。从通信运营商提供的基站定位数据，利用基站小区的覆盖范围作为商圈区域的划分，归纳出商圈的人流特征和规律，识别出不同类别的商圈，实现营销指导。	8	8	实践演练环境： MATLAB、SAS、SPSS、R
	气象与输电线路缺陷关联分析	由于输电线路的组成部件大都暴露在外，极端恶劣的气候对输电线路造成非常大的危害。通过分析鸟害、接地装置受损、保护区安全隐患、部件发热异常、锈蚀损伤等缺陷与气象因子的关联关系，构建出气象因子与缺陷关联关系的预警模型。实现在极端气象条件下缺陷在高发区发生的概率并进行预警。	8	8	实践演练环境： MATLAB、SAS、SPSS、R
认证考核	数据挖掘能力认证	数据挖掘能力认证是数据挖掘方面能力的综合认证。 考核形式：以三人项目组的为单位，进行项目实战考核。 评分标准：数据预处理的完整性、对相关领域背景问题的理解程度、对挖掘模型评价的客观性、模型应用的创新性、结果的正确性和文字表述的清晰程度。	8	-	现场认证，考核合格者发放权威认证证书
实践	上机练习	真实案例实操，项目式管理	-	-	-

具体课程详见： <http://www.tipdm.org/sj/638.jhtml>

实训案例动手实战环境： MATLAB、SPSS Modeler、R、SAS

四、师资力量



蒋老师

华南理工大学博士，兼有大型高科技企业和高校的工作经历，主要从事计算智能与数据挖掘、医学图像处理、并行计算、计算机入侵检测等方面的研究。主编或参编《神经网络实用教程》、《医药数据库应用基础教程》、《医学数据挖掘实用教程》等专著数篇。熟悉程序开发框架与开发平台，对计算机程序体系结构和算法设计有独到的见解。目前主要工作是从事大数据挖掘及其应用的策划、研发及咨询培训。



张老师

高级信息系统项目管理师，华南师范大学、广东工业大学兼职教授，广东省工业与应用数学学会理事。全国计算机技术与软件专业技术资格（水平）考试继续教育和CDA数据分析师培训讲师。发表数据挖掘相关论文数十篇，已取得国家发明专利12项，主编图书《神经网络实用教程》、《数据挖掘-实用案例分析》、《MATLAB数据分析与挖掘实战》等畅销图书。获得SAS、SPSS数据挖掘认证及Hadoop开发工程师证书。



李老师

著名互联网公司深度神经网络实验室数字多媒体技术国家重点实验室的深度学习专家。多年来专注于海量数据挖掘（大数据），人工智能，机器学习（深度学习），信息（文本和图像）检索，模式识别和并行计算。美国麻省理工大学媒体实验室访问科学家、加拿大约克大学和加拿大圣弗朗西斯泽维尔大学博士后，为电信、互联网、生产制造、零售、银行、生物、医药等多个行业提供过数据挖掘应用与咨询服务。



肖老师

应用数学博士，医学影像学博士后，兼有大型高科技企业和高校的工作经历。主要研究领域为最优化方法、分子影像学、大数据应用。多年从事数学建模和数据分析的咨询与培训。参加国家级课题1项，主持省级课题1项，专著有《MATLAB数据分析与挖掘实战》，在国内外学术期发表论文40余篇，具有电力、电信、银行、制造企业、电子商务和电子政务的项目经验和行业背景。



樊老师

Hortonworks授权Apache Hadoop 开发者认证培训讲师，Hadoop、Mahout技术实践者和研究者，对Hadoop的MapReduce编程模型有深刻理解和，同时对Mahout技术有较深认识和理解，熟悉Mahout中数据挖掘的K均值聚类算法、贝叶斯分类算法、FP树关联规则算法等。对Mahout源码有深入研究，主编图书《Mahout算法解析与案例实战》。具有电信行业和银行业的项目经验和行业知识。



刘老师

资深大数据挖掘专家和模式识别专家，韩国全北大学博士，加拿大圣西维尔大学博士后。在国内外重要学术期刊或会议上发表论文20余篇，发明专利10余项。主要研究领域为 物联网、大数据、普适计算、智能控制等。参与并负责多个国内外物联网，大数据相关智能化项目。近10年数据仓库系统管理与实施经验，超过10年的系统开发与设计经验。



云老师

中山大学理学硕士，从事用户数据分析和数据挖掘工作多年。对用户行为分析方法、市场发展情况监控、数据库和数据集市、精确营销有深厚的积累。对数据具有较高的敏感度，根据数据对其进行全面的统计分析。有为中国移动、南方电网、南方航空等大型企业长期提供成功顾问咨询及实施服务的经验，参编图书《MATLAB数据分析与挖掘实战》。

五、费用及资料

1. 现场班的培训费用为 15600 元/人。
2. 培训班提供根据多年数据挖掘应用案例整理出来的《MATLAB 数据分析与挖掘实战》及具有自主知识产权的多媒体材料、PPT 讲稿等其他相关培训资料。

六、开户银行及帐号：

开户银行：中国工商银行广州花城支行

单位名称：广州泰迪智能科技有限公司

帐号：3602028509201663221

七、报名方式：

1、徐老师(18922381327) , E-mail: xu.yinggang@qq.com

2、张老师(13560356095) , E-mail: 13560356095@qq.com

报名步骤：（发邮件）报名表→收到参会通知→转账交费→培训

您还可以加 QQ：532366581、324020430 咨询或报名！

报名截止时间：开课前一个星期

附件：

培训班报名回执表

姓 名		性别		从事工作	
联系电话		到达时间			
电子信箱					
发票抬头					
发票内容	☐ 培训费 ☐ 会议费 ☐ 技术服务费 ☐ 技术咨询费				
详细地址					
备 注	请将此表通过邮箱发至广州泰迪智能科技有限公司 邮箱：xu.yinggang@qq.com ， 13560356095@qq.com				