

# 《大数据和 Python 分析应用实践》

## 培训大纲



### 一、 课程简介

大数据建模与分析挖掘技术已经逐步地应用到新兴互联网企业（如电子商务网站、搜索引擎、社交网站、互联网广告服务提供商等）、银行金融证券企业、电信运营等行业，给这些行业带来了一定的数据价值增值作用。

本次课程面向有一定的数据分析挖掘算法基础的工程师，带大家实践大数据分析挖掘平台的项目训练，系统地讲解：

- (1) Hadoop 和 Spark 大数据建模、分析挖掘技术体系及其平台方案实现；
- (2) 大数据分析挖掘和机器学习的模型、建模方法和案例，重点是决策树机器学习模型和关联规则机器学习模型及其应用；
- (3) SPSS 大数据建模与挖掘软件的应用实践，包括决策树，关联规则，朴素贝叶斯网络，SVM 模型的实践与操作；
- (4) Python 朴素贝叶斯、决策树、逻辑回归机器学习模型及应用实践操作；
- (5) Spark MLlib 大数据建模与机器学习的应用实践；
- (6) 机器学习和大数据建模在电商、电信领域的用户画像应用案例。

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），基本的大数据分析平台所依赖的软件包和依赖库等。

本课程基本的实践环境是操作系统：Windows 或 Linux。

大数据建模挖掘软件包括：SPSS Modeler 15，Python 3，Anaconda 4，Hadoop

2.7.\* , Spark 2.1.\*。

## 二、培训人群

1. 大数据分析应用开发工程师
2. 大数据分析项目的规划咨询管理人员
3. 大数据分析项目的 IT 项目高管人员
4. 大数据分析挖掘处理算法应用工程师
5. 大数据分析集群运维工程师
6. 大数据分析项目的售前和售后技术支持服务人员

## 三、培训特色

定制授课+ 实战案例训练+ 互动咨询讨论，共 4 天（可扩展到 6 天和 8 天）

本课程采用技术原理与项目实战相结合的方式进行教学，在讲授原理的过程中，穿插实际的系统操作，本课程讲师也精心准备的实际的应用案例供学员动手训练。

## 四、培训目标

1. 让学员掌握在大数据、人工智能下的大数据建模、分析挖掘技术和机器学习算法模型的应用。
2. 让学员掌握大数据建模与分析挖掘软件的应用实践。
3. 让学员掌握 SPSS Modeler , Python , Spark MLlib 等大数据建模、分析、挖掘的应用实战。

## 五、培训大纲

时间	专题	详细内容与知识点
第一天	Hadoop 大数据技术平台 实战	<ol style="list-style-type: none"><li>1. Hadoop 大数据平台</li><li>2. HDFS 分布式存储</li><li>3. MapReduce 分布式并行计算</li><li>4. Hadoop 解决方案</li></ol>

		5. Hive On Hadoop 数据仓库方案
第二天	Spark 大数据处理平台实战	6. Spark 实时处理分析 7. Spark RDD 编程 8. Spark SQL 数据分析 9. Spark MLlib 介绍 10. Spark Streaming 流处理实践
第三天	大数据建模、分析挖掘技术体系及其平台方案实现	11. 大数据挖掘模型 12. 机器学习技术和模型 13. 有监督学习、无监督学习和半监督学习 14. 大数据建模和挖掘的应用场景、挖掘过程 15. 常见的大数据建模工具和机器学习工具 (SPSS Modeler 和 Python) 16. 大数据分析建模的案例 (运营商案例)
	Python 大数据决策树算法、模型、建模、挖掘和分析实践	17. 决策树的原理 18. 决策树的算法实现 19. 决策树用于分析挖掘的建模实现 20. 决策树的评估 21. 决策树的应用：客户流失预测
第四天	Python 大数据建模与挖掘软件的应用实践	22. Python 决策树建模实践操作 23. Python 朴素贝叶斯分类实践操作 24. Python 关联规则建模实践操作
	大数据关联规则算法、模型、建模、挖掘和分析实	25. 关联规则和频繁模式挖掘的原理 26. 关联规则的算法实现、建模

	践	<p>27. Apriori 和 FP-growth 关联分析建模</p> <p>28. 关联规则分析挖掘模型的评估</p> <p>29. 关联规则挖掘的应用：交叉销售、个性化推荐</p>
第五天	Python 朴素贝叶斯、决策树、逻辑回归机器学习模型及应用实践操作	<p>30. Python：大数据建模挖掘和机器学习概述</p> <p>31. Python 朴素贝叶斯建模与应用</p> <p>32. Python 决策树建模与应用</p> <p>33. Python 逻辑回归建模与应用</p> <p>34. Python 大数据建模与机器学习的实践操作</p>
	Python TensorFlow 深度学习应用实践	<p>35. TensorFlow 深度学习框架</p> <p>36. CNN 深度学习实践</p> <p>37. 非结构化文本数据分类实战</p>
	大数据、Python 数据分析机器学习应用案例	<p>38. 案例分享</p> <p>39. 讨论、交流</p>

## 师资简介

钟运琴老师，男，目前在中国科学院某研究所工作，高级工程师，副研究员，课题组长，团队成员二十余人，博士毕业于中国科学院，博士、博士后，获工学博士学位（计算机专业），博士后研究方向为人工智能与数据分析科学，曾在国内某高校和某大型通信企业工作过。人工智能、大数据、云计算系列课程建设与教学专家，新技术课程开发组长。近八年来带领团队主要从事大数据管理与高性能分析处理（**Hadoop**、**Spark**、**Storm**）、大数据仓库（**HIVE**）和实时数据仓库（**SparkSQL**、**Shark**），大数据建模挖掘与机器学习（**Mahout**、**Mlib**、**Oryx**、**Pentaho BI**、**SAS**、**SPSS**、**R**等）、**MPP** 并行数据仓库（**Greenplum etc**）、**NoSQL** 与 **NewSQL** 分布式数据库（**HBase**、**MongoDB**、**Cassandra etc**）、（移动）电子商务平台、大数据搜索平台（**ElasticSearch**、**Solr**、**Lucene**等）、云计算与虚拟化

(**OpenStack** , **VMware** , **XenServer** , **CloudStack** , **KVM** , **Docker** , **SaaS 服务**)、云存储系统、**Swift** 对象存储系统、网络 **GIS** 地图服务器、互联网+在线教育云平台方面的项目研发与管理工作。钟老师参与了十余项云计算、大数据分析处理、银行与证券公司金融客户分析系统、金融防诈骗分析预警系统、用户行为分析系统和银行理财产品推荐系统、移动互联网和网络地图服务器相关的国家 **863**、**973** 计划项目，以及承担了二十余项企业合作的横向项目，有着丰富的项目咨询与规划论证经验，参与承担过十余个大型信息化系统咨询项目。在学术期刊和国际会议上发表了 **10** 余篇 **SCI/EI** 索引学术论文以及 **10** 余项发明专利和软件著作。参与过互联网产品创业公司的创建与研发运营全过程，并担任若干家互联网公司的咨询顾问以及两家上市公司的信息化顾问。钟老师有着九年多的企业内训和公开课培训讲师经历，主要讲授大数据平台、云计算、移动互联网、电子商务、**IT** 信息软件项目管理、企业信息化规划与管理、**IT** 战略规划与企业架构、数据中心主机规划与 **IDC** 系统运营等企业实战类培训课程。钟老师将原理技术剖析和应用实战相结合的授课风格受到广大公开课学员和企业内训客户学员的欢迎，钟老师已授课的企业内训客户包括：中海油信息中心、中国移动总部和若干个省级分公司、国家电网总部以及若干下属企业、中石化信息中心、中国联通、广东移动、中移动南方基地、胜利油田、天津石化、**PICC** 人财保险总部信息部、重庆移动、中国邮政、湖北移动、东软总部、江苏移动、徽商银行、中移动南方基地、攀商银行、江苏南瑞集团总部信息部、北京银行、燕山石化总部信息部、航天科技集团某院所、航天科工某院所、公安部出入境管理局、中国人寿总部信息技术部、中石油总部信息部与若干直属企业信息部、上海宝康电子、中电集团某所、中国移动咪咕公司总部、广东联合电子、瑞中科技、国家电网研究院、广西移动、广东电信、汕头移动、惠州移动、江苏移动、浙江移动、海南省移动、广东电信研究院、海口市移动、中国联通研究院、贵州省移动、北京市移动、兴业银行（上海），中移动研究院杭州研发中心、大庆油田、广东电信实业公司等九十余家国内大型国有企事业单位和民营上市公司。钟老师还讲授过一百三十余场云计算、大数据、移动互联网公开课培训课程。