第九届全国高校师资班六大专题

机器学习与深度学习案例及实战

人工智能-计算机视觉及知识图谱实战

Python 数据分析实战

区块链应用案例实战

大数据基础及实战

大数据案例及实战

单击即可查看课程

参加峰会，了解不可错过的新工科教育革命

关于普开

北京普开数据技术有限公司成立于 2010 年，是国内第一家专注大数据教育、大数据应用开发和大数据资源服务的国家高新技术企业。以服务高校、企业和个人为目标，以大数据实验平台、人工智能实验平台和大数据、人工智能教育为核心，为高校和企业培养人才提供教材、教案、应用场景、实训设备、课程体系等产品和服务，协助高校提升大数据及人工智能专业竞争力，培养应用型稀缺人才。

源于产业，集高校、研究院、大型企业多渠道研发力量

师资培训用户数量位居行业第一

1200+ 高校数量

3000+ 老师数量

大数据实验室高校用户口碑第一

50+ 高校已落地

课程体系完备
行业案例丰富
实验手册符合 3M 认知模型
Docker+OpenStack 混合架构节省硬件
管理平台灵活便捷……

全国高校师资班

课程最新、好评最多、讲师最牛、价格最低、开班最多

课程 PPT | 实验手册 | 案例代码 | 配套视频 | 教学研讨 | 教学软件 | 好书赠送

六大专题时间安排

序号	专题	时长	开课日期	地点	费用
1	机器学习与深度学习典型案例及实战□	6 天	7 月 27 日-8 月 1 日	北京	5200
2	人工智能-计算机视觉及知识图谱实战	6 天	8 月 3 日-8 月 8 日	北京	5200
3	Python 数据分析实战	5 天	8 月 3 日-8 月 7 日	北京	4900
4	□大数据应用案例及实战□	6 天	8 月 3 日-8 月 8 日	北京	4200
5	大数据基础及实战□	6 天	7 月 27 日-8 月 1 日	北京	4200
6	区块链应用案例实战	5 天	7 月 27 日-7 月 31 日	北京	4900

往届师资班回顾

1200 所高校共同的选择！

大数据教学研讨会

师资培训期间，普开组织的高校大数据及人工智能专业建设研讨会，来自 50 多所高校的老师齐聚一堂围绕“大数据专业建设课程方案”、“如何进行大数据课程教学”及“实践实验教学方案”进行了激烈的分组讨论并总结陈述。

人工智能教学研讨会

畅销书免费送

普开联合清华大学出版社、中国铁道出版社、电子工业出版社开展大型书展，现场为老师们带来了琳琅满目的书籍，其中有普开数据在清华大学出版社出版的畅销书《大数据技术基础》、《大数据离线分析》，所有书籍免费赠送，如您不方便会后出版社将为您免费邮寄。

校企合作签约仪式

普开数据与河北经贸大学、湖北文理学院、太原科技大学、哈尔滨广厦学院、北京联合大学、河北工业大学、贵州大学、北京邮电大学、临沂大学等 30 多所高校签订大数据校企合作协议，并举行了签约授牌仪式。

现场配备大数据教学实验科研平台

普开自主研发的智能大数据教学实验科研平台，打通了教学-实验-评价-监控-分析全流程，实现大数据教学、产教融合和智能分析的高效融合，完善全面的本科高职课程体系，及 9 类配套教学资源，全面满足大数据教学及科研需求，底层 Docker+OpenStack 混合架构，支持大规模集群并发，并行独立实验，节省硬件投入。目前普开大数据平台已在国内 50 多所高校落地并使用，现场部署了实验环境，欢迎老师们体验咨询。

答谢交流晚宴

每一届师资班普开数据都为参训教师举办专场交流答谢会，宽敞舒适的北理工国际交流中心宴会厅可容纳上千人，高校老师、主任以及院长齐聚一堂，一边用餐一边轻松愉快的交流，答谢宴上笑声不断，还有神秘大奖等您来拿！

专场一：机器学习与深度学习典型案例及实战

课程简介

课程强调动手操作；内容以代码落地为主，以理论讲解为根，以公式推导为辅。共 6 天 12 节，包括 6 节机器学习和 6 节深度学习，每节各 3~3.5 小时，讲解机器学习和深度学习的模型理论和代码实践，梳理机器学习、深度学习、计算机视觉的技术框架，从根本上解决如何使用模型、优化模型的问题；每次课中，首先阐述算法理论和少量公式推导，然后使用真实数据做数据挖掘、机器学习、深度学习的数据分析、特征选择、调参和结果比较。

课程目标

通过课程学习，可以理解机器学习的思维方式和关键技术；了解深度学习和机器学习在当前工业界的落地应用；能够根据数据分布选择合适的算法模型并书写代码，初步胜任使用 Python 进行数据挖掘、机器学习、深度学习等工作。

课程特点

课程重视代码实践，使用金融、气象、农业、交通、安防等工业界实际数据（数据已脱敏）进行机器学习模型的落地应用。虽然课程坚持推导公式，但更重视机器学习和深度学习的原理与实操；将实际工作中遇到的行业应用和痛点做最直观切实的展示；重视算法模型的同时，更强调实际问题中应该如何模型选择、特征选择和调参。

课程大纲

第一天：Python、TensorFlow、回归分析

第一节：Python 与 TensorFlow

- 解释器 Python2.7/3.6 与 IDE：Anaconda/Pycharm
- 列表/元组/字典/类/文件
- numpy/scipy/matplotlib/panda 的介绍和典型使用

代码和案例实践：

1. 卷积与(指数)移动平均线
2. 股票数据分析
3. 缺失数据的处理
4. 环境数据异常检测和分析

- scikit-learn 的介绍和典型使用
- TensorFlow 典型应用
- 典型图像处理
- 多种数学曲线
- 多项式拟合
- 快速傅里叶变换 FFT
- 奇异值分解 SVD
- Soble/Prewitt/Laplacian 算子与卷积网络

第二节：回归分析

- 线性回归
- Logistic/Softmax 回归
- 广义线性回归
- L1/L2 正则化
- Ridge 与 LASSO
- Elastic Net
- 梯度下降算法：BGD 与 SGD
- 特征选择与过拟合
- Softmax 回归的概念源头
- 最大熵模型
- K-L 散度

代码和案例实践：

1. 股票数据的特征提取和应用
2. 泰坦尼克乘客缺失数据处理和存活率预测
3. 环境检测数据异常分析和预测
4. 模糊数据查询和数据校正方法
5. PCA 与鸢尾花数据分类
6. 二手车数据特征选择与算法模型比较
7. 广告投入与销售额回归分析
8. 鸢尾花数据集的分类
9. TensorFlow 实现线性回归
10. TensorFlow 实现 Logistic 回归

第二天：决策树、随机森林、svm

第三节：决策树和随机森林

- 熵、联合熵、条件熵、KL 散度、互信息
- 最大似然估计与最大熵模型
- ID3、C4.5、CART 详解
- 决策树的正则化
- 预剪枝和后剪枝
- Bagging
- 随机森林
- 不平衡数据集的处理
- 利用随机森林做特征选择
- 使用随机森林计算样本相似度异常值检测

代码和案例实践：

1. 随机森林与特征选择
2. 决策树应用于回归
3. 多标记的决策树回归
4. 决策树和随机森林的可视化
5. 葡萄酒数据集的决策树/随机森林分类
6. 泰坦尼克乘客存活率估计

第四节：SVM

- 线性可分支持向量机
- 软间隔
- 损失函数的理解
- 核函数的原理和选择
- SMO 算法
- 支持向量回归 SVR
- 多分类 SVM

代码和案例实践：

1. 原始数据和特征提取
2. 调用开源库函数完成 SVM
3. 葡萄酒数据分类
4. 数字图像的手写体识别
5. MNIST 手写体识别
6. SVR 用于时间序列曲线预测
7. SVM、Logistic 回归、随机森林横向比较

第三天 ：聚类、隐马尔科夫模型 HMM

第五节：聚类

- 各种相似度度量及其相互关系
- Jaccard 相似度和准确率、召回率
- Pearson 相关系数与余弦相似度
- K-means 与 K-Medoids 及变种
- AP 算法(Sci07)/LPA 算法及其应用
- 密度聚类 DBSCAN/DensityPeak(Sci14)
- 谱聚类 SC
- 聚类评价和结果指标

代码和案例实践：

1. K-Means++算法原理和实现
2. 向量量化 VQ 及图像近似
3. 并查集的实践应用
4. 密度聚类的异常值检测
5. 谱聚类用于图片分割

第六节：隐马尔科夫模型 HMM

- 主题模型 LDA
- 词潜入和 word2vec
- 前向/后向算法
- HMM 的参数学习
- Baum-Welch 算法详解
- Viterbi 算法详解
- 隐马尔科夫模型的应用优劣比较
- 共轭先验分布
- Laplace 平滑
- Gibbs 采样详解
- Metropolis-Hastings 算法
- MCMC

代码和案例实践：

1. 敏感话题分析
2. 网络爬虫的原理和代码实现
3. LDA 开源包的使用和过程分析
4. HMM 用于中文分词
5. 文件数据格式 UFT-8、Unicode
6. 发现新词和分词效果分析
7. 高斯混合模型 HMM
8. GMM-HMM 用于股票数据特征提取

第四天 计算机视觉的关键任务及实践

第七节：卷积神经网络 CNN

- 神经网络结构，滤波器，卷积
- 池化，激活函数，反向传播
- 目标分类与识别、目标检测与追踪
- AlexNet、VGGNet、GoogleLeNet
- Inception-V3/V4
- ResNet、DenseNet

代码和案例实践：

1. 数字图片分类
2. 卷积核与特征提取
3. 以图搜图
4. 人证合一
5. 卷积神经网络调参经验分享

第八节：图像视频的定位与识别

- 视频关键帧处理
- 物体检测与定位
- RCNN，Fast-RCNN，Faster-RCNN，MaskRCNN
- YOLO
- FaceNet

代码和案例实践：

1. 迁移学习
2. 人脸检测
3. OCR 字体定位和识别
4. 睿客识云
5. 气象识别

第五天 循环神经网络 RNN、自然语言处理

第九节：循环神经网络 RNN

- RNN 基本原理
- LSTM、GRU
- Attention
- CNN+LSTM 模型
- Bi-LSTM 双向循环神经网络结构
- 编码器与解码器结构
- 特征提取：word2vec
- Seq2seq 模型

代码和案例实践：

1. 看图说话
2. 视频理解
3. 藏头诗生成
4. 问答对话系统
5. OCR
6. 循环神经网络调参经验分享

第十节：自然语言处理

- 语言模型 Bi-Gram/Tri-Gram/N-Gram
- 分词、词性标注
- 依存句法分析
- 语义关系抽取
- 词向量

代码和案例实践：

1. 输入法设计
2. HMM 分词
3. 文本摘要的生成
4. 智能对话系统和 SeqSeq 模型
5. 阅读理解的实现与 Attention

■ 文本分类 ■ 机器翻译 ■ 文本摘要 ■ 阅读理解 ■ 问答系统 ■ 情感分析	

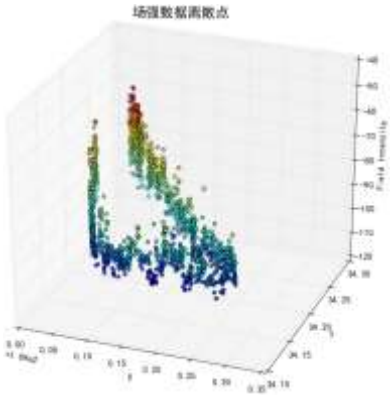
第六天 深度学习与计算机视觉的未来发展趋势

第十一节：生成对抗网络 GAN ■ 生成与判别 ■ 生成模型：贝叶斯、HMM 到深度生成模型 ■ GAN 对抗生成神经网络 ■ DCGAN、InfoGan ■ Conditional GAN ■ Wasserstein GAN	**代码和案例实践：** 1．图片生成 2. 看图说话 3. 对抗生成神经网络调参经验分享
第十二节：强化学习 RL ■ 为何使用增强学习 ■ 马尔科夫决策过程 ■ 贝尔曼方程、最优策略 ■ 策略迭代、值迭代 ■ Q Learning、SarsaLamda ■ DQN、A3C、ELF	**代码和案例实践：** 1. OpenAI 2. 飞翔的小鸟游戏 3. 基于增强学习的游戏学习 4. DQN 的实现

现场实验

实验案例全部更新，赠送课件、代码、数据、配套实践手册

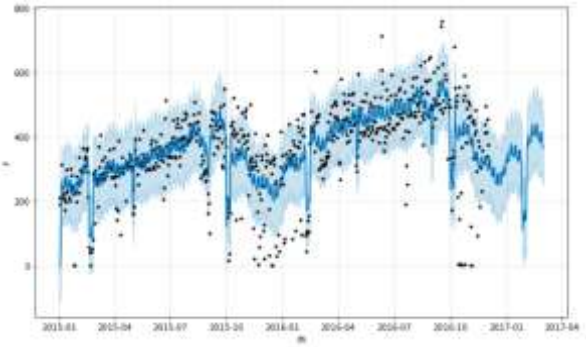
实验一、无线电磁波频率场强预测



实验二、气象设备海量雨量筒图片处理和识别

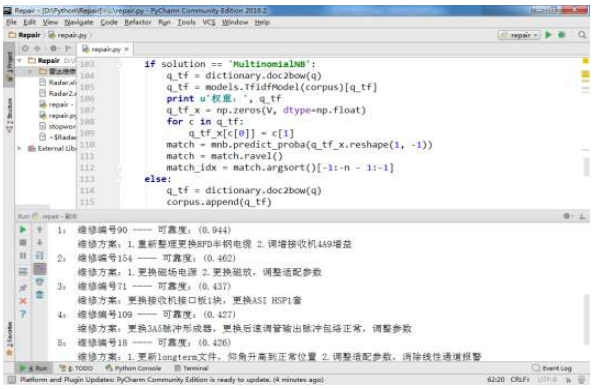
实验三、视频检测与图像定位

实验四、时间序列分析

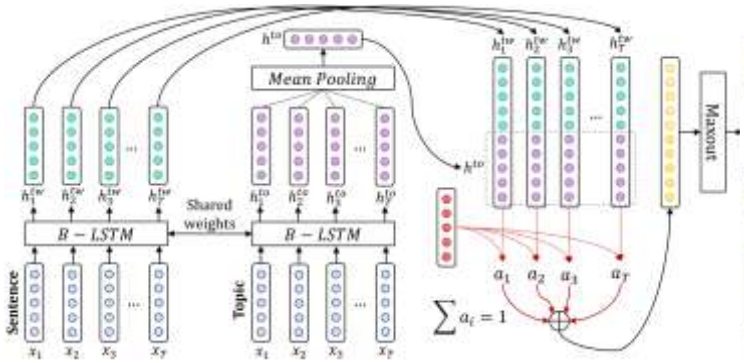


实验五、车辆跟踪和目标定位

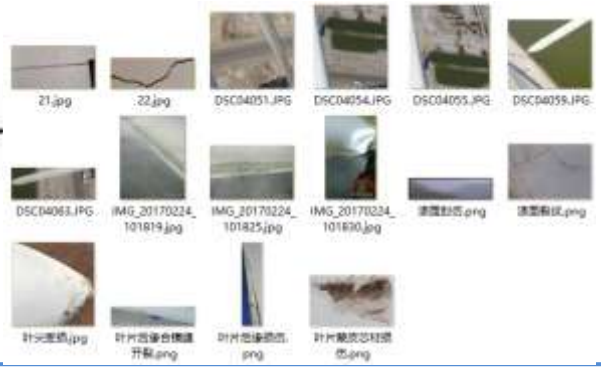
实验六、雷达维修方案智能客服推送算法



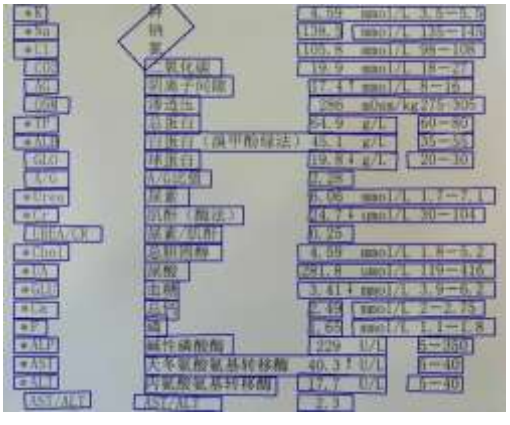
实验七、Topic Bi-LSTM+Attention 模型框架



实验八、风机发电设备缺陷检测



实验九、睿客 OCR



实验十、睿客识云

专家介绍

邹老师：计算机博士，中国科学院副研究员；主持国家级科研项目2个，副负责1个，国家专利2项，在多家公司担任技术顾问、首席科学家、大学创业导师，研究方向机器学习、数据挖掘、人工智能、计算几何，应用于大型气象设备的图像与文本挖掘、股票交易与预测、医药图像识别、智能畜牧、量子化学医药路径寻优、传统农资产品价格预测和决策等领域。擅长机器学习核心算法分析和代码实现。具备成熟授课经验，讲课清晰易懂，扎实的理论知识和出色表达获得学员的广泛好评。

专场二：人工智能-计算机视觉及知识图谱实战

课程简介

本课程围绕计算机视觉、深度学习和知识图谱领域最新发展，通过形象生动的课程讲授、以微项目工程实践为导向的实践环节设计、基于小组的学习及互动三位一体的教学方式，讲授计算机视觉与深度学习的基本原理、实例应用及前沿进展，内容包括图像获取、视觉预处理与特征提取、运动估计、三维视觉、深度学习原理、基于深度学习的图像识别、分类等原理，同时结合项目实例讲解视觉跟踪、目标检测等应用；知识图谱以问答系统为例，介绍知识图谱的基本概念及工程化实现方法。课程注重教学内容的可视化和工程实践，在激发学习兴趣的同时，切实掌握教学内容并能动手实践，为创新教学及深入研究、工程项目实践奠定坚实基础。

课程大纲

第一天

第 1 节　课程引论

- 计算机视觉和深度学习的概念，发展历史，特点
- 计算机视觉和深度学习应用

- 课程结构、需要理论基础和相关知识

第 2 节　图像获取

- 图像和视觉系统概念、构成、特点、重点介绍应用
- 照明模型，光通量和辐照度定义；颜色空间与颜色模型，RGB、CMYK 与 HSI 颜色空间
- 图像与视频的采集，彩色图像传感器的基本原理
- 视频压缩与图像显示

第 3 节　图像预处理及特征提取（配合动手实验）

- 图像预处理及空域滤波的常用方法，包括平均滤波，高斯滤波，中值滤波等；**结合实例介绍 OpenCV 实现及结果对比**
- 图像边缘检测方法，包括 Sobel 算子，Laplace 算子，Canny 算子；OpenCV 实现
- 灰度直方图的基本概念；基于灰度的直方图分与大津算法，**结合 OpenCV 实例**
- 图像的直线检测方法，Hough 变换及 OpenCV 实现
- 图像特征点提取：特征角点的基本定义，Harris 角点检测算法，SIFT 角点检测算法（选讲），**OpenCV 示例**

第二天

第 4 节　运动估计（配合动手实验）

- 基于背景建模的运动目标检测：混合高斯模型，基于混合高斯模型的背景提取算法，运动目标检测与跟踪；
- 光流估计：光流基本概念，光流方程，Lucas-Kanade 光流估计方法，Horn-Schunck 光流估计方法(选讲)

第 5 节　图像处理综合实例（配合动手实验）

- 实例 1：米粒记数及特征统计
- 实例 2：光流跟踪
- 实例 3：基于背景建模的的多目标检测与跟踪

第 6 节　视觉成像几何与位姿估计（配合动手实验）

- 相机成像几何与常见坐标变换
- 根据特征点对应确定相对位姿(PnP)：基本原理，算法步骤，**结合 OpenCV 实现**
- 双目视觉基础：双目视觉基本问题介绍，极线几何原理

第三天

第 7 节　位姿估计(续)与相机标定（配合动手实验）

- 双目视觉基础续：基于双目视觉的三维重构，基本思想与算法示意
- 相机标定方法：Zhang 方法的基本思想，OpenCV 实现
- 相机标定及位姿估计示例

第 8 节　卷积神经网络概要

- 问题背景：介绍神经网络深度学习的发展历程；深度学习出现的必然性和优势
- 二分类问题和 Logistic 回归；
- 损失函数与梯度下降法；
- 多层前馈网络与误差反传算法
- 卷积层与池化层
- 卷积神经网络构成，通过 LeNet-5 实例介绍卷积神经网络构成及各层详细解释

第四天

第 9 节　Tensorflow 与 MNIST 识别实例（结合 Tensorflow 动手实验）

- Tensorflow 介绍
- MNIST 数据集与基本操作
- 使用 LeNet 进行 MNIST 数字识别

第 10 节　图像识别与分类（结合 Tensorflow 动手实验）

- AlexNet、VGG 与 Resnet
- CIFAR-10 与 ImageNet 图像识别
- CIFAR-10 数据集与 ImageNet 图像识别，包括使用 Tensorflow 如何读取 CIAR-10
- 卷积神经网络图像识别

第五天，以问答系统为例，介绍知识图谱的基本概念及工程化实现方法。

第 11 节 知识图谱及行业应用

- 什么是知识图谱
- 知识图谱的典型应用场景
- 知识图谱相关资料汇总

第 12 节 知识图谱基本概念及方法

- 知识图谱和人工智能的关系
- 三元组

第 13 节 图数据库与 SQL 数据库简介

- 关系型数据库简介

- NoSQL
- Neo4j 为例的图数据库

第 14 节 三元组在图数据库与 SQL 数据库中的存储与检索（动手实验）

第六天，以问答系统为例，介绍知识图谱的基本概念及工程化实现方法。

第 15 节 基于知识图谱的 QA 系统简介

- 传统 QA 系统概述
- 传统 QA 系统的算法，及其局限性
- 基于知识图谱的 QA 系统的架构

第 16 节 问题解析的自然语言基础

- 自然语言处理概述
- 分词、词性标注和依存语法分析

第 17 节 问题解析的方法：人工规则与机器学习

- 基于模板的知识抽取
- 基于模板的实体抽取
- 基于深度学习的关系抽取

动手实验：使用模板进行问题解析

第 18 节 知识图谱的构建与查询

- 知识的抽取
- 知识图谱工程化构建
- 知识的检索

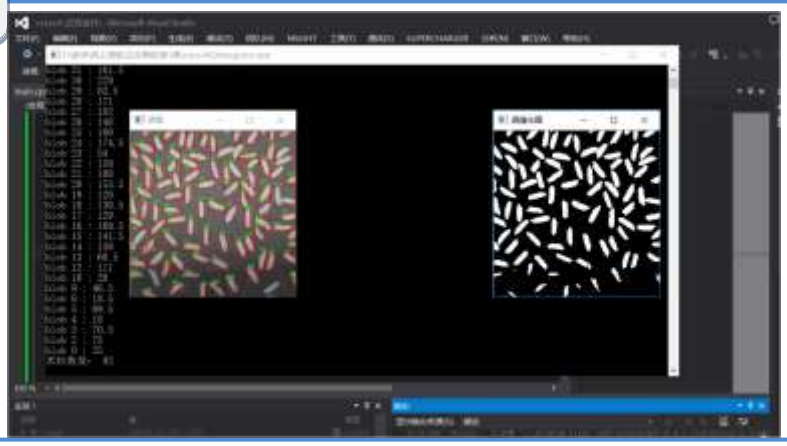
现场实验

实验一、米粒检测

要求： 检测图中所有面积>50 的米粒

检测流程：

1. 图像采集(取到图像)
2. 图像预处理
3. 基于灰度的阈值分割
4. 图像特征描述及目标分析



实验二、运动目标检测（静止背景）

要求： 检测红虚线中的所有运动目标，并标识每一目标的外接包围矩形，判断目标类别

检测流程：

1. 视频采集(取到视频中当前帧图像)
2. 图像预处理
3. 背景提取
4. 前景分割
5. 图像特征描述及目标分析
6. 得到最终结果

实验三、光流跟踪

要求： 使用光流法跟踪给定视频或摄像头中的运动特征点

检测流程：

8. 视频采集(取到视频中当前帧图像)
9. 图像预处理
10. 提取特征点
11. 使用光流法估计特征点运动
12. 相邻帧及特征点交换
13. 得到最终结果

专家介绍

屈老师，知名大学副教授，博士学位，视觉技术研究室负责人。从事计算机视觉和神经网络方向的教学和科研实践工作 20 年。目前承担本科生及研究生多项课程，完成及承担国家及省、部级科研项目多项，在运动跟踪、视觉检测等方面取得一系列突破性研究成果，累计发表视觉相关方向论文 40 余篇，获省部级奖励两项。

陶老师， 现任北京大数据基地首席科学家，曾任 IBM 资深数据科学家，高级项目总监；东华软件大数据高管，从事工业大数据建模分析与平台搭建。曾经为众多世界 500 强企业提供算法和数据平台服务。本科和研究生均毕业于北京大学数学科学学院。 陶老师常年以特聘讲师身份，在清华大学、武汉大学、南京大学、浙江大学、天津大学，中国科学院大学，北京交通大学等授课。 中国计算机协会青年科学家论坛委员。目前从事教育、知识图谱、工业、舆情、医疗健康、生态和政府大数据的项目实施和研究。

专场三：Python 数据分析实战

课程简介

本课程首先介绍 Python 运算符、表达式、内置对象 ,列表、元组、字典、集合以及 range 对象、map 对象、filter 对象、zip 对象等常用对象，选择结构与循环结构，函数定义与使用，面向对象程序设计基础，字符串与正则表达式，文件操作等内容。学习完这些基础内容之后，重点学习网络爬虫原理与应用、numpy 数组运算与矩阵运算、pandas 数据分析、matplotlib 数据可视化以及 sklearn 机器学习等方面的内容。

课程目标

通过本课程的学习，能够理解 Python 的编程模式，熟练运用 Python 列表、元组、字典、集合、字符串等基本数据类型以及相关列表推导式、切片等操作来解决实际问题，熟练掌握 Python 分支结构、循环结构、函数的应用，理解正则表达式处理字符串的原理，熟练使用 Python 读写文本文件与二进制文件以及 Word、Excel、PowerPoint 等文件操作，掌握网络爬虫原理并能够自己编写爬虫在网上抓取数据，掌握使用 numpy 进行数组运算与矩阵运算，能够使用 pandas 分析业务数据并使用 matplotlib 绘制折线图、散点图、饼状图、柱状图等等，能够使用 sklearn 实现线性回归、系统聚类、KNN、KMeans 等算法并解决实际问题,同时还能养成代码优化与安全编程意识和习惯。

课程特点

授课模式为边讲边练，案例丰富，实用性强，讲解透彻。

课程大纲

第一天

第 1 节　Python 概述

- Python 是这样一种语言
- Python 版本之争
- Python 编程规范与代码优化建议
- Anaconda3 开发环境的安装与使用
- 安装扩展库的几种方法
- 标准库与扩展库中对象的导入与使用

代码和案例实战

Python 安装与开发环境搭建。

第 2 节　运算符、表达式与内置对象

- Python 常用内置对象
- Python 运算符与表达式
- Python 关键字简要说明
- Python 常用内置函数用法精要
- 代码和案例实战
- 基本输入输出、内置函数、函数式编程，曼哈顿距离计算。

第 3 节　详解 Python 序列结构

- 列表
- 元组

  元组对象的创建与删除，其他类型到元组的转换，元组与列表的区别，生成器推导式。
- 字典

  字典对象的创建与删除，字典元素、键、值的访问方法，字典元素的添加与修改，使用字典对象的 get()方法获取指定键对应的值，有序字典对象的使用，使用字典来计数，有序字典。
- 集合

  集合对象的创建与删除，其他类型数据到集合的转换，使用 add()方法为集合增加元素，使用

pop()和 remove()方法删除元素，使用 clear()方法清空集合，集合的并、交、差以及对称差等运算，使用集合来提取序列中的唯一元素。

- 序列解包的多种形式和用法

  序列解包的本质——对多个变量同时赋值；序列解包的要求——等号右侧值的数量与等号左侧变量的数量一致。

代码和案例实战

词频统计，电影评分与推荐，过滤无效书评。

第二天

第 4 节　程序控制结构

- 条件表达式

Python 关系运算符可以连用，枚举序列元素时多使用成员测试运算符 in，算数运算符/和//的区别，任何合法的 Python 表达式都可以作为条件表达式，False、0（或 0.0、0j 等）、空值 None、空列表、空元组、空集合、空字典、空字符串、空迭代对象作为条件表达式与 False 等价。

- 选择结构

单分支选择结构、双分支选择结构、多分支选择结构、选择结构的嵌套，条件表达式后必须有冒号且换行缩进后开始写代码块，elif 关键字表达 else if 的意思。

- 循环结构

for 循环、while 循环，range 对象在循环中的使用，成员测试符 in 与序列解包在循环语句中的使用，带有 else 子句的循环结构，循环代码的优化，break 与 continue 语句的作用。

代码和案例实战

尼姆游戏，蒙蒂霍尔悖论，决赛现场成绩计算，汉诺塔问题，任务调度。

第 5 节　函数

- 函数定义与使用

  Python 函数的定义形式，def 关键字，不需要声明函数的参数类型和返回值类型，括号不可以省略，冒号必须有。函数嵌套定义。函数递归调用。

- 函数参数

  形参与实参的概念，在函数内部直接修改形参不会影响实参。位置参数、默认值参数、关键参数和长度可变参数的用法。传递参数时的序列解包。

- 变量作用域

  在函数内如果只引用某个变量的值而没有为其赋新值，该变量为（隐式的）全局变量；如果在函数内任意位置有为变量赋新值的操作，该变量即被认为是（隐式的）局部变量，除非在函数内显式地用关键字 global 进行声明。

- lambda 表达式

  使用 lambda 表达式声明匿名函数和命名函数，在 lambda 表达式中调用函数，把 lambda 表达式作为函数参数。

- 生成器函数设计要点

yield 语句和 return 语句的区别，生成器对象的惰性求值特点。

代码和案例实战

猜数游戏，抓狐狸游戏，轮盘抽奖，报数游戏，递归算法，经典排序算法。

第三天

第 6 节　字符串

- 字符串编码格式简介

  ASCII、UTF8、GBK 的区别。

- 转义字符与原始字符串

  转义字符的用法，字符串前面加字母 r 的含义。

- 字符串格式化

  format()方法，格式化的字符串常量。

- 字符串常用操作

  使用 find()、index()、count()等方法以及关键字 in 确定字符串中是否包含另一个字符串，使用 split()、partition()方法对字符串进行切分，使用 join()方法连接字符串，使用 replace()、translate()对字符串进行“修改”，使用 strip()方法删除字符串开始、结束或两侧的指定字符，使用 lower()、upper()、capitalize()、title()、swapcase()等方法进行大小写转换，使用 startswith()、endswith()方法测试字符串是否以另一个字符串开始或结束，使用 center()、ljust()、rjust()等方法对字符串进行排版。

- 字符串常量
- 中英文分词

  Python 扩展库 jieba 和 snownlp 的用法。

- 汉字到拼音的转换
- Python 扩展库 pypinyin 的用法。

代码和案例实战

凯撒加密，压缩，密码生成器，密码安全强度检查，批量随机信息生成，文本嗅探。

第 7 节　正则表达式

- 正则表达式语法

  正则表达式基本语法与扩展语法，常用正则表达式的写法。

- 直接使用正则表达式模块 re 处理字符串

  compile()、match()、search()、findall()、split()、sub()、escape()等 re 模块常用方法。

- 使用正则表达式对象处理字符串

  正则表达式对象方法的用法。

- match 对象

  match 对象方法的用法。

代码和案例实战

提取最长数字子串，提取电话号码，特定类型字符串检查。

第四天

第 8 节　文件内容操作

- **文件操作基本知识**

文本文件与二进制文件的概念与访问方法的区别 ,文件对象 ,open()方法以及 flush()、close()、read()、readline()、readlines()、tell()、write()、writelines()等常用方法，上下文管理语句 with 的优点与用法。

- **文本文件基本操作案例精选**

文本文件读写方法，read()、write()、seek()等文件对象常用方法的运用。

- **二进制文件操作案例精选**

二进制文件读写方式，对象序列化，pickle、struct、shelve 等模块的用法，Excel、Word 等常见类型文件的操作。

代码和案例实战

文本文件读写，二进制文件序列化与反序列化，JSON 文件读写，Word、Excel、PowerPoint 文件读写。

第 9 节　文件与文件夹操作

- os 模块

  文件夹遍历

- os.path 模块

  路径类型判断，路径连接，获取文件属性

- shutil 模块

复制文件，压缩与解压缩，删除目录树

代码和案例实战

目录树深度优先遍历，目录树广度优先遍历，获取文件夹属性，磁盘清理器。

第五天

第 10 节　数据采集、分析与科学计算可视化

- 网络爬虫

  urllib，scrapy，requests，beautifulsoup

- 扩展库 numpy 简介

  数组运算，矩阵运算。

- 学计算扩展库 scipy

  constants 模块，special 模块，signal 模块，ndimage 模块。

- 扩展库 pandas 简介

  DataFrame 基本操作，缺失值、重复值、异常值处理，数据差分，读写文件。

- matplotlib

  绘制折线图、散点图、饼状图、柱状图，设置图例，绘制三维图形。

- sklearn

  线性回归、系统聚类、KNN、KMeans，贝叶斯

代码和案例实战

中国工程院院士信息爬取，指定城市当前天气爬取，天涯小说爬取，饭店营业额数据处理与可视化，儿童身高预测，肥胖程度分类，基于 SVM 的图像识别，垃圾邮件分类。

现场实验

实验一、抓狐狸游戏的设计与开发

重点练习 Python 内置函数、列表、选择结构、

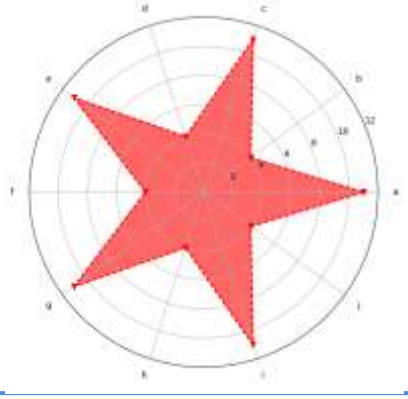
循环结构、异常处理结构、序列解包以及标准库

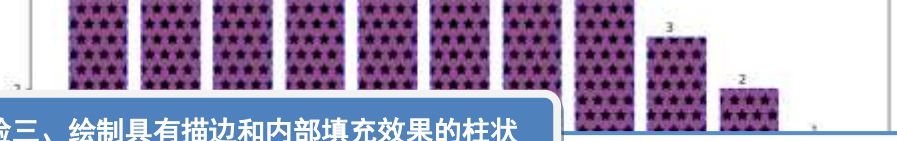
random 的用法。

实验二、绘制雷达图

使用 matplotlib 扩展库绘制

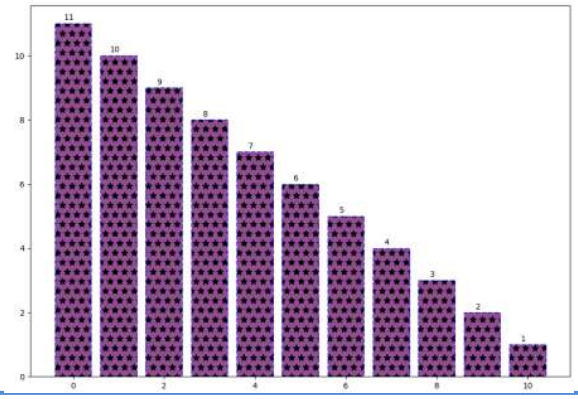
雷达图的方法与要点。





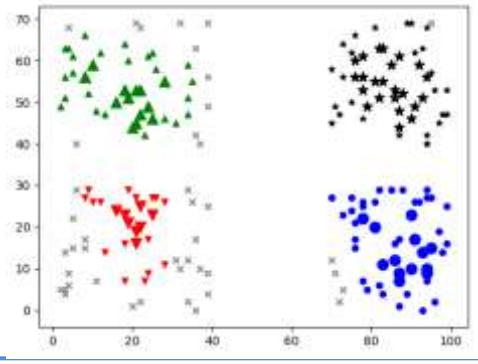
实验三、绘制具有描边和内部填充效果的柱状

介绍使用 matplotlib 扩展库绘制柱状图的方法与要点。



实验四：使用 DBSCAN 算法进行分类

介绍 DBSCAN 算法原理与实现，通过实际的案例演示该算法的用法。



董老师 ,副教授 ,先后出版《Python 程序设计(第 2 版)》、《Python 可以这样学》、《Python 程序设计开发宝典》、《中学生可以这样学 Python》、《Python 程序设计基础（第 2 版)》、《玩转 Python 轻松过二级》等系列教材，被超过百所高校选作教材，其中《Python 可以这样学》已发行繁体版。多次获得校级教学优秀效果一等奖，分别使用 C#和 Python 开发过适用于煤矿、油田、淘宝店的多套系统，以及在线考试系统和课堂教学管理系统。发表科研论文 40 余篇，近 30 篇被 EI 收录。长期维护微信公众号“Python 小屋”并已免费分享 550 多篇 Python 技术文章，关注人数近 3 万。近几年来应邀为国内高校和企业作报告或培训 30 多次，多次出席全国教学研讨会并做专题报告。

专场四：区块链应用案例实战

课程简介

当下比特币，以太坊，区块链备受世人关注，掀起国际新思潮。第四次产业革命在即，区块链的思想和技术必将掀起时代变革。区块链技术已经开始被全球广泛认同，预示着人类将跨入“Token 经济”时代。

区块链技术的出现，使得互联网发展从信息互联网到价值互联网、秩序互联网转型，其

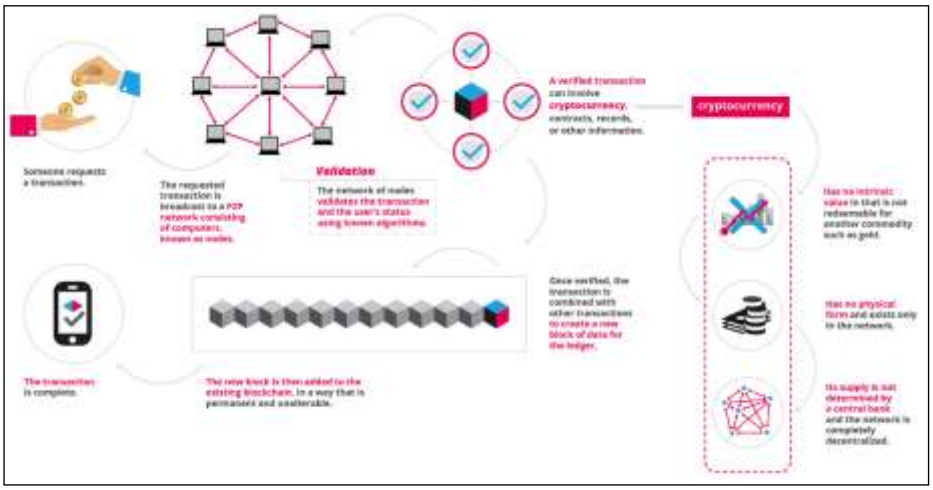
以共享经济与价值链接为主要特征的“分布式商业”模式，将催生大量创新合作场景，构建起以区块链技术为虚拟中心的新产业生态系统，将对经济社会治理、产业变革与创新产生颠覆式影响。

课程特点

了解区块链前世今生，发展脉络，未来趋势

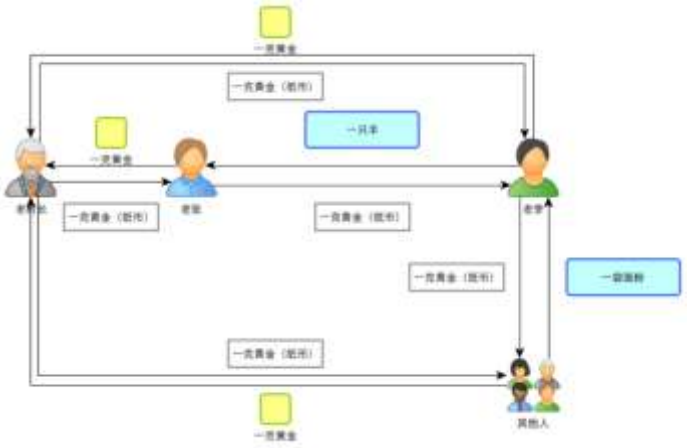
培养区块链应用场景业务分析能力

具备 Dapp 开发能力。



课程大纲

第一天：区块链基础



1.1 初识区块链家族

1.1.1 区块链的父母

1.1.2 P2P 下载

1.1.3 P-CDN

1.1.4 分布式计算

1.1.5 结交媒体

1.1.6 P2P 借贷

1.1.7 众筹及 ICO

1.1.8 区块链及比特币

1.1.9 自组织 DAO

1.2 区块链入门综合篇

1.2.1 比特币简介

1.2.2 区块链发展简史

1.2.3 区块链起源思想

1.2.4 区块链本质及核心特征

1.2.5 区块链应用场景

1.3 比特币核心概念

1.2.1 比特币核心术语

1.3.2 节点、钱包

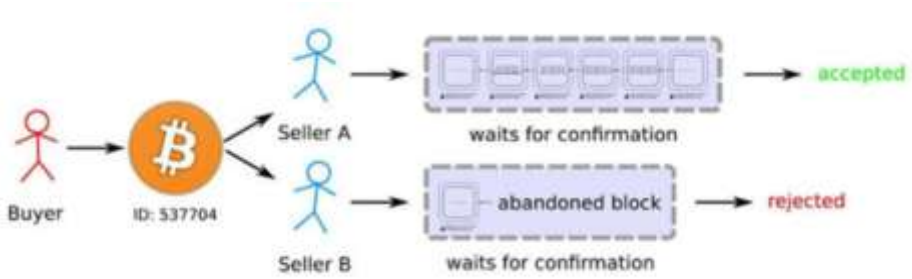
1.3.3 私钥和公钥

1.3.4 区块及区块结构

1.3.5 挖矿、双花

1.2.6 区块链不足

1.3.6 BIPs



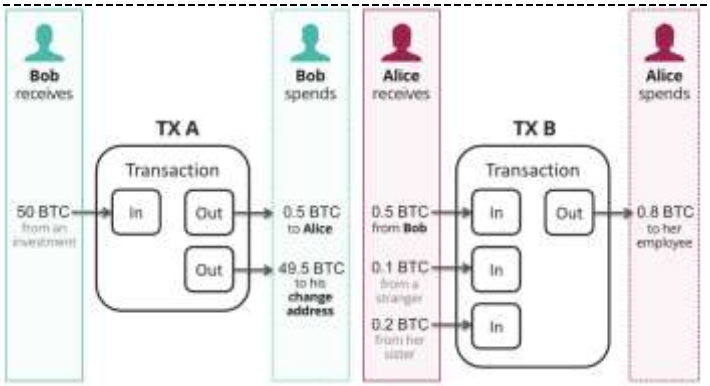
1.4 比特币交易初识

1.4.1 比特币交易概述

1.4.2 比特币交易过程

1.4.3 比特币系统如何运行

1.4.4 比特币的安全问题



第二天：区块链密码算法及共识算法

2.1 哈希算法

2.1.1 哈希的概念

2.1.2 主流的哈希算法

2.1.3 哈希与加密的区别

2.1.4 封装哈希算法工具类

2.2 非对称加密算法

2.2.1 非对称加密算法发展史

2.2.2 非对称加密与对称加密的区别

2.2.3 RSA 加密的原理

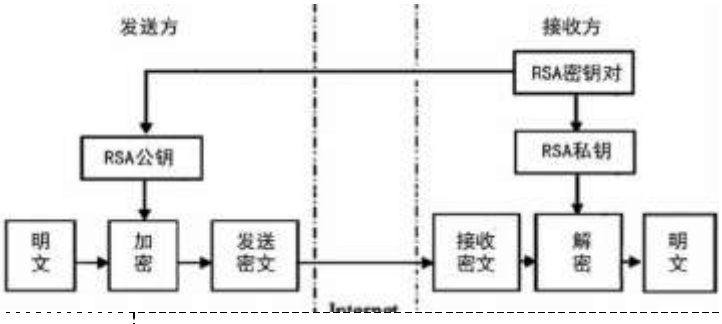
2.2.4 RSA 加密核心代码

2.3 数字签名

2.3.1 数字签名的概念

2.3.2 数字签名的原理

2.3.4 RSA 实现数字签名与验证



2.4 Base64 及 Base58 编码与解码

2.4.1 Base64 字符集

2.4.2 Base64 编码的实现原理

2.4.3 Base58 字符集

2.4.4 Base58 编码的核心代码

2.5 比特币地址的生成算法

2.5.1 比特币地址的生成步骤

2.5.2 生成 WIF 及 WIF-compressed 格式私钥

2.5.3 Base58Check

2.5.4 Secp256k1 算法生成公钥与私钥

2.5.5 封装比特币地址工具类

2.6 ECDSA 实现交易签名及签名验证

2.6.1 初步认识比特币交易

2.6.2 签名序列化 DER

2.6.3 Secp256k1 实现交易签名及签名验证

2.7 知名共识算法

2.7.1 PoW

2.7.2 PoS

2.7.3 DPoS

2.7.4 PBFT

2.7.5 Paxos

2.7.6 Raft

2.8 比特币挖矿的 PoW 算法

2.8.1 挖矿原理

2.8.2 计算难度目标的实现步骤及核心代码

2.8.3 编程验证挖矿结果的步骤及核心代码

2.8.4 难度目标 Hash 的快速算法

2.8.5 计算区块挖矿难度 difficulty

第三天：比特币交易及环境搭建

3.1 比特币客户端 Bitcoin Core 及 Bitcoin-cli

3.1.1 区块链节点及比特币钱包

3.1.2 比特币钱包 Bitcoin Core 的安装

3.1.3 轻节点钱包 Multibit 的安装

3.1.4 测试网络中用钱包实现比特币交易

3.1.5 Bitcoin-cli 基本命令

3.2 比特币客户端 bitcoind 的安装及高级用法

3.2.1 什么是 bitcoind

3.2.2 bitcoind 的安装和编译

3.2.3 bitcoin-cli 高级命令

3.3 比特币交易流程

3.3.1 比特币交易的本质

3.3.2 比特币交易的生命周期

3.3.3 比特币交易规则

3.3.4 交易验证过程

3.3.5 交易优先级

3.4 比特币交易脚本

3.4.1 比特币交易脚本概述

3.4.2 比特币交易脚本语法及常用指令

3.4.3 比特币交易脚本执行过程

3.4.4 比特币交易脚本类型

3.5 比特币交易的数据结构

3.5.1 比特币区块数据结构

3.5.2 P2PKH 交易数据结构

3.5.3 P2PK 交易及 coinbase 交易的数据结构

3.6 比特币交易编程

3.6.1 bitcoin-cli 实现交易

3.6.2 RPC 编程接口调用

第四天　以太坊环境搭建及项目实战-web3. js

4.1 以太坊私有链搭建

4.1.1 以太坊安装配置创世区块

4.1.2 启动私有链

4.2 node.js 开发环境搭建

4.2.1 node.js 安装

4.2.2 Web3.js 安装

4.2.3 Express 安装

4.3 以太坊实战案例

4.3.1 Express 前端网页制作

4.3.2 Solidity 编译部署和运行

4.4 solidity 语法

4.4.1 安装

4.4.2 数据类型

4.4.3 单位和全局变量

4.4.4 表达式和控制结构

4.4.5 异常

4.4.6 创建合约

4.5 web3.js

4.5.1 导入 web3.js

4.5.2 连接至节点

4.5.3 API

4.5.4 BigNumber.js

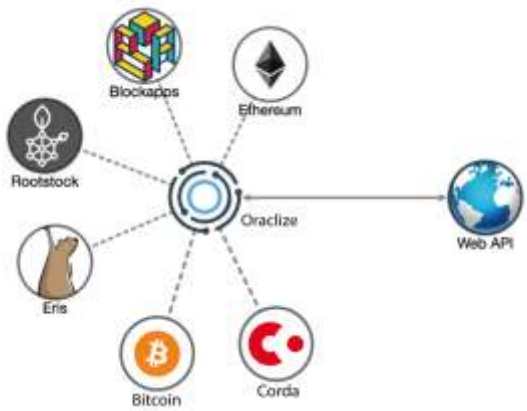
4.5.5 单位转换

4.5.6 检索 gas 价格、余额和交易

4.5.7 发送以太币

4.5.8 处理合约

4.5.9 检索和监听合约事件



4.6 创建客户端

4.6.1 项目结构

4.6.2 创建后端

4.6.3 创建前端

4.6.4 测试客户端

第五天　以太坊项目实战-投注 APP

5.3 创建投注合约

5.1 Oraclize

5.1.1 Oraclize 工作原理

5.1.2 数据源

5.1.3 真实性证明

5.1.4 定价

5.1.5 开始使用

5.2 处理字符串

5.4 创建投注 APP 客户端

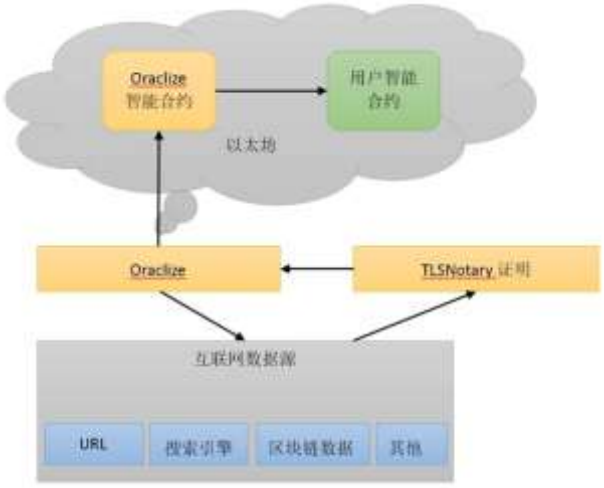
5.4.1 项目结构

5.4.2 创建后端

5.4.3 创建前端

5.4.4 测试客户端

项目实战



在区块链应用中，智能合约如何和外部世界交互是一个问题，很多的智能合约应用场景是根据一些外部事件，输出相应的结果，而传统的 IT 数据交互方式实际上并不能投入真正的工作。Oraclize 是一个在区块链和互联网 API 中间的一个服务，通过 Oraclize 可以使合约获取可信的互联网数据。

Oraclize 是一个独立的服务提供商，提供免费的数据输送服务，其目的是在区块链和互联网之间建立一道可信的数据网关，其目标是打破智能合约获取数据的束缚，在保证可信的情况下，使其具有访问互联网数据的能力。Oraclize 的运行状态如下图。

专家介绍

王老师

大数据教育解决方案领军者

会务咨询：肖艺：15210061600

从事IT行业15年，7年编程开发经验，8年计算机培训经验。自2003年始，分别从事Asp、Asp.net、PHP、JavaEE、Android、Web前端开发等开发相关工作。出版发行《变身程序猿——Android应用开发》一书，以及《Steven陪你学Android》《零基础学编程》等系列视频。

孔老师

7年开发和教育经验，对web前端，java，php，游戏开发有丰富经验。参与大型网络游戏《擒王》和京东小金库的开发工作，有亿万级访问量的项目经验。

专场五：大数据基础及实战

课程大纲

6 天大容量+典型案例+动手实验（赠送课件、代码、数据、配套的实验手册）

第一天

模块一、高校如何开设大数据教学课程

- 大数据相关专业人才需求缺口
- 大数据相关招聘岗位需求分析
- 大数据技术演进与变革
- 应对 IT 新技术变革，教师知识的储备与提升
- 在哪个层面进行教学

精彩案例

- ✧ 开展大数据分析挖掘教学的形式
- ✧ 高校开设大数据的教学可在多个层面上进行
- ✧ 有条件的高校可以开设云计算专业
- ✧ 不成熟的可在在计算机相关专业上增设数据挖掘和分析方向
- ✧ 不具备条件的高校可以开设大数据方面的课程

模块二、大数据的发展现状

- 大数据起源和产生、大数据概念的发展与解析
- 大数据在国内外发展现状、大数据在互联网发展现状
- 大数据四个特点分析

模块三、大数据带来的机遇和挑战

- 大数据能带来什么 、引领社会进入“大数据时代”
- 大数据对国家、社会的作用 、大数据将推动经济发展
- 大数据将推动科技发展进程、开启商业智能新阶段
- 数据分析的发展——从数据到知识
- 大数据如何让商业更智能、大数据应用案例
- 带来数据处理新变革 、大数据的关键技术
- 大数据与云计算 、大数据技术的发展趋势

精彩案例

- ✧ 智慧政府
- ✧ 智慧城市与大数据
- ✧ 车联网
- ✧ 大数据舆情分析
- ✧ 数据治理
- ✧ 智能推荐
- ✧ 大数据与银行
- ✧ 大数据营销分析

模块四、大数据核心组件的安装部署

- Hadoop V2 核心组件简单介绍
- Hadoop V2 部署角色简单介绍
- Hadoop V2 试验集群的部署结构

实战内容（现场实验环节）

- ✧ Hadoop V2 安装部署实验
- ✧ Red hat Linux 基础环境搭建
- ✧ Hadoop V2 单机系统版本安装配置

● Hadoop V2 安装依赖关系 ● Hadoop V2 生产环境的部署结构 ● Hadoop V2 集群部署 ● Hadoop V2 高可用配置方法 ● Hadoop V2 集群简单测试方法 ● Hadoop V2 集群异常 Debug 方法	✧ Hadoop V2 集群系统版本安装和启动配置 ✧ 使用 Hadoop MapReduceV2 样例代码快速测试系统 ✧ Hadoop V2 配置文件 core-site.xml，hdfs-site.xml，mapred-site.xml 和 yarn-site.xml 详解 ✧ Hadoop V2 环境变量 Hadoop-env.sh 和 yarn-env.sh 详解 ✧ 单机环境&模拟线上环境，Hadoop V2 集群安装 ✧ Hadoop V2 集群初始化、启动和停止操作 ✧ 结合修改 Hadoop V2 配置文件，讲解 Hadoop 运维和优化真实集群下验证 ✧ Hadoop 启动是否成功及实际启动部署问题现场解决

第二天

模块五、电商日志评论数据分析系统 ● 京东案例剖析 ● 京东数据介绍 ● 系统的技术框架选型 ● 日志采集系统介绍	**实战内容（现场实验环节）** ✧ Linux 下日志清洗、加载（实战） ✧ Web 系统最终结果数据展示（演示实验）
模块六、YARN 架构设计和核心概念 ● YARN 出现的背景 ● YARN 的设计思想和基本概念 ● YARN 的基础架构 ● YARN 的工作流程 ● YARN 基础类库详解 ● YARN 核心模块 ApplicationMaster 详解 ● YARN 核心模块 ResourceManager 剖析 ● YARN 核心模块 NodeManager 详解 ● Hadoop V2 资源调度器详细介绍	**实战内容（现场实验环节）** ✧ 基于 YARN 的应用程序设计和简单实现案例 ✧ YARN 使用第三方类库，包含通信、服务、时间、状态等 ✧ 从资源管理角度理解 YARN 框架 ✧ ApplicationMaster 核心源码 ✧ ResourceManager 核心源码 ✧ NodeManager 核心源码
模块七、基于电商日志数据的 MapReduce 高级编程 ● MapReduce V2 基本架构、原理和相关角色介绍 ● MapReduce V1 和 MapReduce V2 ● MapReduce V2 中 MRAppMaster 的工作流程 ● MapReduce V2 作业生命周期理解 ● MapReduce V2 资源调度理解 ● MapReduce V2 作业恢复和推测执行机制介绍 ● 剖析 MapReduce V2 样例程序代码流程 ● 基本 MapReduce V2 API 概念 ● 驱动代码 Mapper、Reducer ● Hadoop 流在 MapReduce V2 中的使用 ● 使用 Eclipse 进行快速开发 ● MapReduce V2 的编程优化 ● 满足解决实际数据分析问题的高级 MapReduce V2 多语言编程实践	**实战内容（现场实验环节）** ✧ Hadoop Streaming 和 Java MapReduce V2 编程的差异 ✧ 利用 Combiner 来减少中间数据 ✧ 编写 Partitioner 来优化负载平衡 ✧ 直接访问 Hadoop 分布式文件系统 HDFS ✧ MapReduce V2 的 join 操作 ✧ 辅助排序在 Reducer 方的合并 ✧ 定制开发 Writables 和 WritableComparable 高级类 ✧ 使用 SequenceFiles 和 Avro 文件保存二进制数据 ✧ 定制开发 InputFormats 和 OutputFormat ✧ 基于 MapReduce V2 的海量日志分析 ✧ 熟悉 Java 编程 IDE Eclipse ✧ MapReduce V2 wordcount 实例测试 ✧ MapReduce V2 wordcount 源代码查看讲解 ✧ MapReduce V2 实现单词统计扩展代码逻辑并测试结果 ✧ 基于 MapReduce V2 编写海量日志的简单分析需求

第三天

模块八、基于 Hadoo2.*的 Hive 实战技巧

（基于京东数据）

- Hive 和 Pig 架构和理论基础
- Hive 的作用和原理说明
- Hadoop 仓库和传统数据仓库的协作关系
- Hadoop/Hive 仓库数据数据流
- Hive 部署和安装
- Hive Cli 的基本用法
- HQL 基本语法
- 京东评论数据分析

开发技巧详解

- ✧ 使用 JDBC 连接 Hive 进行查询和分析
- ✧ 使用正则表达式加载数据
- ✧ HQL 高级语法
- ✧ 编写 UDF 函数
- ✧ 编写 UDAF 自定义函数
- ✧ 使用 Sqoop 进行数据分析
- ✧ 使用 Oozie 和 HUE 介绍
- ✧ 动手安装 Hive，并验证是否安装成功
- ✧ 现场解决启动、部署问题，并总结解决问题的方法
- ✧ 使用 Hive 创建自己的数据仓库
- ✧ 使用 Hive 在数据仓库上进行增删改查操作

第四天

模块九、基于运营商数据的 HBase 海量实时处理实战

- HBase 简介和架构
- HBase 核心知识点
  HBase 高级应用
- HBase 应用场景
- HBase（For Hadoop V2）安装、部署、启动
- HBase 常用接口和 SQL 引擎层实战
- 新进展、新技术
- 基于 Hadoop V1 和 V2 使用 HBase 的异同
- HBase 调优
- 京东评论实时数据入库、查询
- HBase 存储逻辑结构介绍
- HBase 与其他 NoSQL 的比较
- 列式存储核心：LSM
- 日志系统：WAL

实战内容（现场实验环节）

- ✧ 动手安装 HBase，并验证是否安装成功
- ✧ 现场解决启动、部署问题，总结解决问题的方法
- ✧ 使用 HBase 设计特定需求下的表结构
- ✧ 使用 Java API 对 HBase 进行增删改查操作
- ✧ NoSQL 数据库与关系数据库的对比
- ✧ 基于列簇的 NoSQL
- ✧ 集群下安装部署 HBase
- ✧ 启动 HBase 和启动顺序
- ✧ 如何通过命令和 WebUI 验证启动是否成功

实战内容（现场实验环节）

- ✧ 通过 HBase Shell 增删改查数据
- ✧ 通过 HBase Shell 进行管理表和 Region 操作
- ✧ 通过 HBase Shell 执行 Java 类中方法
- ✧ 解决 Master 单点问题
- ✧ Native Java 增删改查实战
- ✧ Thrift&Thrift2 增删改查实战
- ✧ MapReduce 访问 HBase 编程实战
- ✧ 数据批量导入（Bulk Load）实战
- ✧ 读、写优化中需要参数的调整
- ✧ hbase-site.xml 参数调优
- ✧ JVM 优化中需要调整的参数
- ✧ split & compact 优化相关参数
- ✧ 表设计优化相关参数
- ✧ HBase 客户端优化相关参数
  监控工具使用方法及注意事项

第五天

模块十、Spark 安装部署与运行模式

- Spark 简介
- Spark 是什么
- Spark 生态系统
- Spark 架构

实战内容（现场实验环节）

- ✧ 现场解决启动、部署问题，并总结解决问题的方法
- ✧ 导入数据，并使用 Java 语言编程进行查询操作
- ✧ Spark Streaming 及 SparkSQL 安装与使用

- Spark 集群安装与部署
- 在 Linux 集群上安装与配置
- 查看集群状态

模块十一、内存计算框架 Spark 实战

- 深入 Spark 核心架构
- Spark 集群配置介绍开发环境配置
- 搭建 Intellij 开发环境
- Spark 应用案例编程

- ✧ RDD、缓存策略介绍
- ✧ Transformation、action
- ✧ Lineage、容错处理
- ✧ 宽依赖与窄依赖
- ✧ Spark on Yarn 实践
- ✧ Python 简介
- ✧ PySpark API

第六天

- **模块十二、Spark 计算模型与核心机制**

- Spark 计算模型
- park 程序模型
- 弹性分布式数据集
- Spark 算子分类及功能
- Spark 工作机制
- Spark 应用执行机制
- Spark 调度与任务分配模块
- Spark I/O 机制
- 序列化、压缩
- Spark 块管理
- Spark 通信模块
- 通信框架 AKKA
- Client、Master 和 Worker 间的通信
- 容错机制、Lineage 机制
- Checkpoint 机制
- Shuffler 机制

实战内容（现场实验环节）

✧ 实践 RDD 的转换函数：map、filter、flatMap、sample、
✧ groupByKey、reduceByKey、 union、join、distinct、cartesian、mapValue、subtract、cache、persist
✧ 实践 RDD 的动作函数：foreach collect collectAsMap
✧ reduceByKeyLocally lookup counmt top reduce fold
✧ aggregate saveAsTextFile saveAsObject
✧ Spark 应用提交工具（spark-submit，spark-shell）
✧ RDD 特性、常见操作、缓存策略
✧ RDD Dependency、Stage 常、源码分析
✧ Spark on YARN 运行模式及测试

模块十三、基于京东评论数据的 Spark SQL 开发

- 使用 Spark SQL 的原因
- Spark SQL 的发展历程
- Spark SQL 的性能
- Spark SQL 运行架构
- Tree 和 Rule
- sqlContext 的运行过程
- hiveContext 的运行过程
- catalyst 优化器

- 解析 Spark SQL 组件
- LogicalPlan
- SqlParser
- Analyzer
- Optimizer
- 深入了解 Spark SQL 运行计划
- 不同数据源的运行计划
- 不同查询的运行计划
- Spark SQL 应用案例

京东评论数据基于 Sparl-sql 的分析

实战内容（现场实验环节）

✧ 通过 Spark SQL 对微博数据（JSON 格式）进行分析处理

- ✧ 通过 Spark SQL 对农产品数据进行分析处理
- ✧ sqlContext 应用
- ✧ hiveContext 应用、混合使用、缓存的使用、DSL 的使用

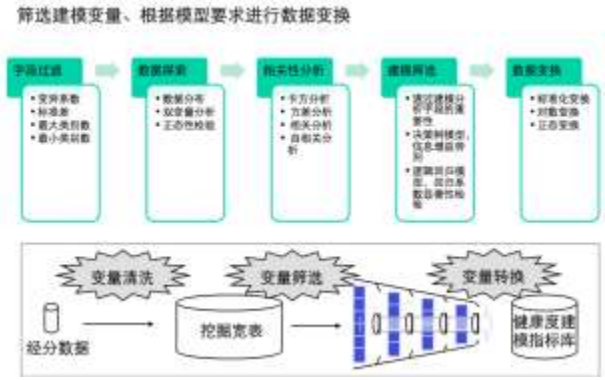
模块十四、流式计算组件 Spark Streaming 使用

- ● Spark streaming 的产生动机及概念
- ● Streaming 多数据源的使用
- ● Spark streaming 的程序设计
- ● Streaming 处理 hdfs 数据程序
- ● Streaming 的 transformation 操作
- ✧ 无状态 transformation 与有状态 transformation 对比
- ✧ Spark Streaming 性能优化
- ✧ Spark HistoryServer 历史应用监控
- ✧ Spark Streaming 流式计算
- ✧ Spark Streaming 原理、DStream 设计
- ✧ Spark Streaming 常见 input、out
- ✧ Spark Streaming 与 Kafka 集成

项目实战一

某运营商新增客户维系大数据项目。运营商建立新增用户维系闭环管理体系，准确分群，分级预警，制定针对性的维系策略和方案并实施，实现新增用户保有。

项目使用数据挖掘挖掘实现，包含 KMeans 等。



项目实战二

京东商城手机销量数据统计分析项目，旨在为电子商务平台卖家提供以客户实际应用产品的感受为核心的服务平台，使得卖家商城能够从海量购买用户的使用感受评论数据中获取到用户购买习惯及购买商品的偏好，进而为客户提供最优质最适合自己的商品。

该项目为电商务平台中的商品做商品的画像，如，在本系统中对手机商品做手机画像，进而统计出用户最喜欢的手机款式、颜色、性能等信息，全方位的分析一款手机的各种属性，统计这些属性用户在使用过程中的偏好，使得电商平台卖家为用户提供最优质的服务。该项目能够从海量用户评论信息的地域数据中，获取用户购买该产品的地区，帮助卖家了解商品的销售区域及区域的销售量等重要信息

以商业实战项目作为驱动来学习大数据技术，使得学员能够亲身体会大数据项目的使用场景和开发场景及其所产生的商业价值，零距离接触企业实战型项目，学以致用，不在停留在大数据的概念环节而是进入大数据技术实战项目开发的阶段。

项目实战三

某银行大数据平台项目，是针对银行的客户进行用户画像，用户开发等。满足了银行对客户进行分类、聚类等研究，最终为银行实现精准营销等方式实现了支撑。

用户画像的核心工作是为用户打标签，打标签的重要目的之一

专场六：大数据应用案例及实战

课程大纲

6 天大容量+2 大行业案例+现场实验（赠送课件、代码、数据、配套的实践手册）

第一天：电子商务推荐系统

电商推荐系统与大数据

- **推荐系统与大数据**
  - ✓ 大数据概述
  - ✓ 推荐系统技术栈
  - ✓ 大数据经典架构： lambda 架构
- **认识推荐系统**
  - ✓ 什么是推荐系统
  - ✓ 推荐系统现状 ，介绍电子商务推荐系统在国内应用情况，以及产生的收益如何。电子商务推荐系统大数架构
  - ✓ 推荐系统评测
  - ✓ 推荐系统带来的收益
- **推荐系统设计**
  - ✓ 推荐系统设计方法
  - ✓ 用户界面的重要性
- **用户画像系统**

第二天

推荐算法实践

- **推荐系统算法模型**
  - ✓ 推荐模型构建流程
  - ✓ 基于协同过滤的推荐算法

- ✓ 基于模型的推荐算法

- **Mahout 推荐算法实战**
  - ✓ 推荐系统模型架构
  - ✓ 推荐系统实现
  - ✓ 推荐系统评估
- **Spark 推荐算法实战**
  - ✓ 推荐系统构建流程
  - ✓ ALS 推荐的实现
- **用户活跃度和物品流行度关系**
  - ✓ 实验设计与算法评测
  - ✓ 推荐算法介绍
- **推荐系统冷启动问题**

第三天

构建推荐系统平台

- **数据采集**
  - ✓ 利用 Flume 和 Sqoop 进行分布式数据收集
- **数据预处理**
  - ✓ 数据清洗和建模
  - ✓ 利用 MapReduce 和 Hive 进行用户行为分析
- **构建离线推荐模型**
  - ✓ 用户行为数据收集
  - ✓ 用户行为数据存储
  - ✓ 利用 Mahout 和 Spark 来构建推荐模型
- **构建服务层**
- **利用 kafka 消息队列存储用户行为流式数据**
  - ✓ 用户行为实时获取
- **利用 Spark Streaming 完成实时推荐**

案例分析

电子商务推荐系统案例

第四天 电信运营商工单数据分析系统

项目概述、原始数据分析、数仓构建

- **了解实训项目的背景和基本概念，从最终成果展示开始反推整个项目的实现过程，并且根据具体的需求设计整体架构，对每个组件做选型。介绍每个组件的基本概念、架构和使用场景，利于判断比较。**

- ✓ 大数据行业的应用案例 、大数据背景
- ✓ 项目需求及功能设计
- ✓ 项目数据格式
- ✓ 项目整体演示 （实战）
- ✓ 项目整体架构（三大模块：数据源、数据存储、Web 展示）
- ✓ 原始数据演示 （实战）
- ✓ 原始数据基本分析（实战）
- ✓ 程序实现与部署（演示实战）
- ✓ 数仓构建（演示实战）

第五天

关键业务数据分析

- **HDFS 基本概念和架构分析**
- **Hive 与 Spark-SQL 介绍**
- **Scala、Spark 基本概念和简单实用（演示实战）**
- **关键业务模型分析**
- **基于 SparkSQL 的关键业务实现（实战）**
- **(1)服务用户分布图**
- **(2)服务能力**
- **(3)服务管控**
- **基于 Scala、Spark 的难点业务模型实现：服务能力（实战）**

第六天

生产部署和实施

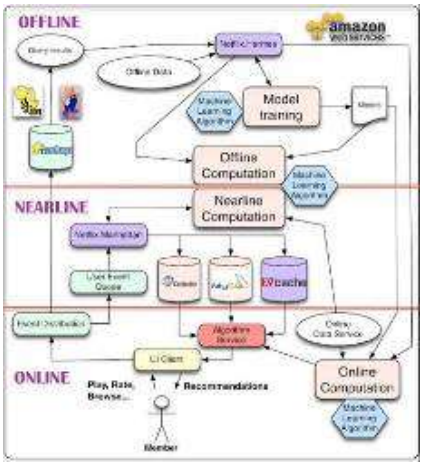
- **WEB 前端功能详解**
- **JSP 和 Servlet 的应用（实战）**
- **Tomcat 安装部署（实战）**
- **MySQL 与 jdbc 访问数据库（实战）**
- **javascript 与 echarts 的介绍**
- **完整项目生产部署流程（实战）**
- **解决项目中的疑难问题**

现场项目咨询

现场解决项目中的疑难问题

项目实战一：电子商务推荐系统

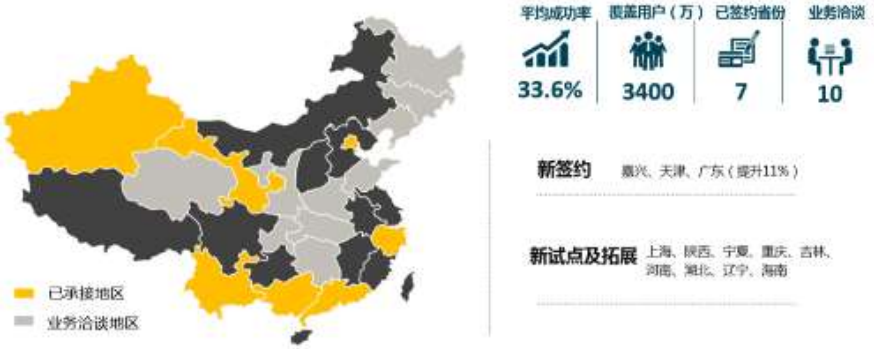
本项目将以一个实际的系统“电子商务推荐系统”为案例，从无到有，简单到深入，从容易到复杂，循序渐进地介绍一个大数据应用是如何落地的，包括如何进行技术选型，如何设计一个大数据架构，如何验证结果的有效性等。



	1	1	0	1	0	1	0
	1	0	1	0	1	0	1
	1	0	0	1	0	0	1
	0	1	0	1	0	1	0
	0	1	1	0	1	0	0
	1	1	0	0	1	0	0

项目实战二：电信运营商工单数据分析

本项目在对客户（各省电信公司）提供有价值的数据，关键元素：地址位置、群障、高危用户，查中自复；精确掌握各个服务团队（中心二、中心三等）的服务情况，对于可能不达标的情况能及时知道或者预警。



专家团队

任大勇

IBM大数据架构师，Cloudera高级认证讲师。

8年以上IT行业从业经验，某知名全球500强企业大数据平台架构师，大数据处理专家。长期从事大数据基础架构建设、大数据分析、挖掘等领域工作。一直致力于大数据行业相关技术的研究。对Hadoop生态系统，特别是Hive、Spark等开源框架的业务应用、可靠性、基础架构和高级应用等方面有着丰富经验。

项目经验

✧ 车联网大数据台构建

- 海量数据实时计算系统
- 客户价值体系分析系统
- 广告活动推荐系统、HBase 监控系统
- 中文搜索引擎系统
- 化妆品个性化推荐项目
- 用户触点个性化推荐项目
- Greenplum&Hadoop 大数据平台项目

游大海

大数据高级讲师，具有多年开发和教育培训经验。曾领导了多个大数据项目的研发，在Spark、Hadoop等方面有丰富的源码、实务和性能优化经验，授课通俗易懂，参与过多个大数据项目的架构设计、技术选型及模块的开发；精通Hadoop生态圈的常用框架，如Hbase, Hive　Flume, kafka, zookeeper等； 以及精通Spark RDD, Streaing, Sql，尤其擅长Spark Streaming实时计算。善于讲解复杂的概念，注重学员学习体验，能很快的让学员把所学的知识应用到实际项目中。

项目经验

✧ 门户网站的用户点击流日志分析
✧ 游戏日志分析项目
✧ 店商用户日志分析系统
✧ 随身 wifi 设备状态监控告警系统

工信部培训证书

食宿安排：食宿统一协助安排，费用自理；

培训费：含培训费、教材费、资料费、场地费、午餐费、答谢晚宴等；

颁发证书：参加相关培训并通过考试的学员，可以获得：

工业和信息化部颁发《Hadoop 高级开发工程师》、《大数据高级分析师》、《深度学习工程师》、《Python 数据分析工程师》、《区块链工程师》、《人工智能工程师》培训证书。该证书可在工信部相关网站查询，可作为能力评价、考核和任职的重要依据。考试及证书费用（可选）：500 元/人。

配套教材及实验手册

提供实验环境：安装实验和安装部署好的实验平台

每次师资班都设置了大量动手实验，并提供数据供实验使用，整个教学过程中采用理论结合实践的方式，以帮助学员快速掌握机器学习案例及应用。

【配套教材+实验手册】

培训教材、配套实验手册、本次培训全套教学课件、代码、数据、课程回放等。

住宿餐饮

普开数据为老师们统一安排住宿，费用自理。培训地点位于北京理工大学国际交流中心，由于参加老师较多交流中心房间数量有限，需提前与会务组确认。步行 1 分钟-10 分钟周边宾馆比较多，以下酒店也可自行安排预订。

1、北京理工大学国际交流中心（上课地点）

2、北京理工延园招待所（距上课地点 800 米），价格 360 元左右，(010)68911191。

3、速 8 酒店（苏州桥店）（距上课地点 200 米）010-68716717，价格在 280 元左右。

4、如家快捷酒店(北京苏州桥人民大学店) 距上课地点 340 米，010-62612266，价格 340 元左右。

5、汉庭酒店(人民大学店)（距上课地点 900 米）010-82656868，价格 400 元左右。

6、北京冀中金牛酒店（距上课地点 350 米），(010)56985555，价格 600 元左右。