

第一章 Python与数据化运营

- 1、数据化运营
 - 意义
 - 提高运营决策效率
 - 提高运营决策正确性
 - 优化运营执行过程
 - 提高投资回报
 - 两种方式
 - 辅助决策式数据化运营
 - 数据驱动式数据化运营
 - 数据需求沟通
 - 需求产生
 - 需求沟通
 - 业务需求沟通
 - 产生背景，要解决的问题，预期效果
 - 数据现状沟通
 - 统计分析、OLAP分析、回归、聚类、分类、关联、异常检测、时间序列、协同过滤、主题模型、路径分析、漏斗分析等
 - 数据与分析的关联性沟通
 - 数据分析建模
 - 获取数据
 - 数据预处理
 - 质量检验、样本均衡、分类汇总、合并数据集、删除重复项、数据分区、排序、数据离散化、数据标准化、过滤变量、转置、查找转换、数据脱敏、数据转换、抽样、异常值和缺失值处理
 - 数据分析建模
 - 统计分析、OLAP分析、回归、聚类、分类、关联、异常检测、时间序列、协同过滤、主题模型、路径分析、漏斗分析等
 - 数据结论输出
 - 数据分析或挖掘建模报告，另外还包括Excel统计结果、数据API输出、数据结果返回数据库、数据结果直接集成到应用程序中进行自动化运营
 - 数据落地应用
 - 数据结论沟通
 - 通过沟通来初步验证结论的正确性、可靠性和可行性，并对结果进行修正
 - 数据部署应用
 - 后续效果监测和反馈
- 2、python库

- 计算库: numpy, pandas, sklearn
- 标准库: re、time、datetime、json、base64、os、sys、cPickle、tarfile
- 统计分析: Statsmodels
- 中文处理: 结巴分词
- 文本挖掘: Gensim
- 爬虫和解析: requests、Beautiful Soup、xml
- 图像处理: OpenCV和PIL
- 数据读取: xlrd、pymongo、mysql.connector
- 数据预处理: imblearn
- 展示美化类: Matplotlib、graphviz、prettytable、wordcloud、mpl_toolkits、pydotplus
- 3、python库分类详解
 - 科学计算库
 - Numpy
 - Scipy
 - Pandas
 - Statsmodels
 - 机器学习库
 - scikit-learn
 - 自然语言处理库
 - 结巴分词
 - 实现分词、词典管理、关键字抽取、词性标注
 - Gensim
 - 可扩展统计语义、分析纯文本语义结构以及检索语义上类似的文档
 - 数据库连接库
 - MySQL连接库
 - MongoDB连接库
 - 由C++语言编写的分布式文件存储的数据库，它是以Key-Value（键值对）形式面向文档存储的非关系型数据库
 - HTML处理库
 - BeautifulSoup
 - 图形展示库
 - Matplotlib
 - 图像处理库
 - PIL
 - 常用的图像输入、处理和分析的库，提供了多种数据处理、变换的操作方法和属性
 - OpenCV
 - OCR工具: Tesseract-ocr
 - Google支持的开源OCR图文识别项目
 - 机器学习框架: TensorFlow

- 谷歌基于DistBelief进行研发的第二代人工智能学习系统，它使用图模型将复杂的数据结构传输至人工智能神经网络中进行分析 and 处理的框架
- 4、案例分析
 - 第一步 导入库
 - 第二步 导入数据
 - 第三步 数据预处理
 - 列表（新建、追加）、for循环、变量赋值、字符串分割、数值转换。列表转数组、重新设置数组形状
 - 第四步 数据分析
 - 第五步 数据建模
 - 第六步 模型评估
 - 第七步 销售预测
- 5、小结
 - 《利用Python进行数据分析 [Python for Data Analysis]》
 - 适合数据分析和挖掘工作者作为入门阅读
 - 《Python数据分析与挖掘实战》
 - 适合对Python有一定了解的数据工作者阅读
 - 《机器学习实战 [Machine learning in action]》
 - 需要读者具有相当的算法、程序和模型专业知识，适合中高级数据挖掘和建模工程师阅读
 - 《集体智慧编程》
 - 使用Python进行机器学习的专业书籍，需要读者了解工程、算法和模型知识，适合中高级数据挖掘、建模工程师、程序员阅读
 - 《Python基础教程（第2版 修订版）》
 - 主要围绕每个方法、条件、函数、对象、属性等进行介绍，适合程序员以及想深入了解python工作原理和逻辑的读者查阅
 - 《深入浅出MySQL》（第2版）
 - 完整的阐述了MySQL的开发、设计、运维、管理等方面的内容，内容全面，并且有适合初学者的章节
 - 《TensorFlow实战》