基于Python的数据分析及挖掘

# 课程背景

目前大数据领域处理编程语言大多数基于Python语言,语法简洁清晰,开发速度快,可以用来独立进行数据量化及数据挖掘。

# 课程收益

1. 使用Python可以进行数据的量化分析。
2. 使用Python进行数据分析与挖掘挖掘。

# 课程时长

2天

# 课程大纲

1. 数据量化分析（第1天）
   1. NumPy数组
      1. NumPy数组对象
      2. 创建多维数组
      3. 选择NumPy数组元素
      4. NumPy的数值类型
      5. 一维数组的切片与索引
   2. Pandas入门
      1. Pandas的安装与概览
      2. Pandas数据结构之DataFrame
      3. Pandas数据结构之Series
      4. 利用Pandas查询数据
      5. 利用Pandas的DataFrame进行统计计算
      6. 利用Pandas的DataFrame实现数据聚合
      7. DataFrame的串联与附加操作
      8. 连接DataFrames
      9. 处理缺失数据问题
      10. 处理日期数据
      11. 数据透视表
   3. 数据可视化
      1. Matplotlib的子库
      2. Matplotlib绘图入门
      3. 对数图
      4. 散点图
      5. 图例和注解
      6. 三维图
   4. 文本分析
      1. 词频统计 - 语料库的构建
      2. 词频统计 - 中文分词
      3. 词频统计 - 实现
      4. 使用sklearn包
      5. 案例1：关键词提取
      6. 案例2：关键词提取实现
      7. 案例3：相似文章推荐
      8. 案例4：自动提取文章摘要
2. 建模与数据挖掘（第2天）
   1. 回归
      1. 相关分析
      2. 简单线性回归
      3. 多重线性回归
      4. 一元非线性回归
      5. 逻辑回归
      6. 神经网络简介
      7. 案例1：餐厅营业额预测
      8. 案例2：广告投放与收益预测
      9. 案例3：住户身份归类
      10. 案例4：鸢尾花数据分类
      11. 案例5：学生求学意愿预测
      12. 案例6：葡萄酒品类分类
   2. 特征工程
      1. 数据处理
      2. 特征选择
      3. 维度压缩
   3. 聚类
      1. K-Means划分法
      2. DBSCAN密度法
      3. 层次聚类法
      4. 案例1：人群聚集点的分类
   4. 关联
      1. 关联规则
      2. 协同推荐
      3. 案例1：超市顾客购买行为关联性识别
   5. 时间序列
      1. 非季节性时间序列分解
      2. 季节性时间序列分解
      3. 序列预测
      4. 案例1：股价波动周期性预测
   6. 模型持久化
      1. 模型持久化