



Python机器学习小案例源码

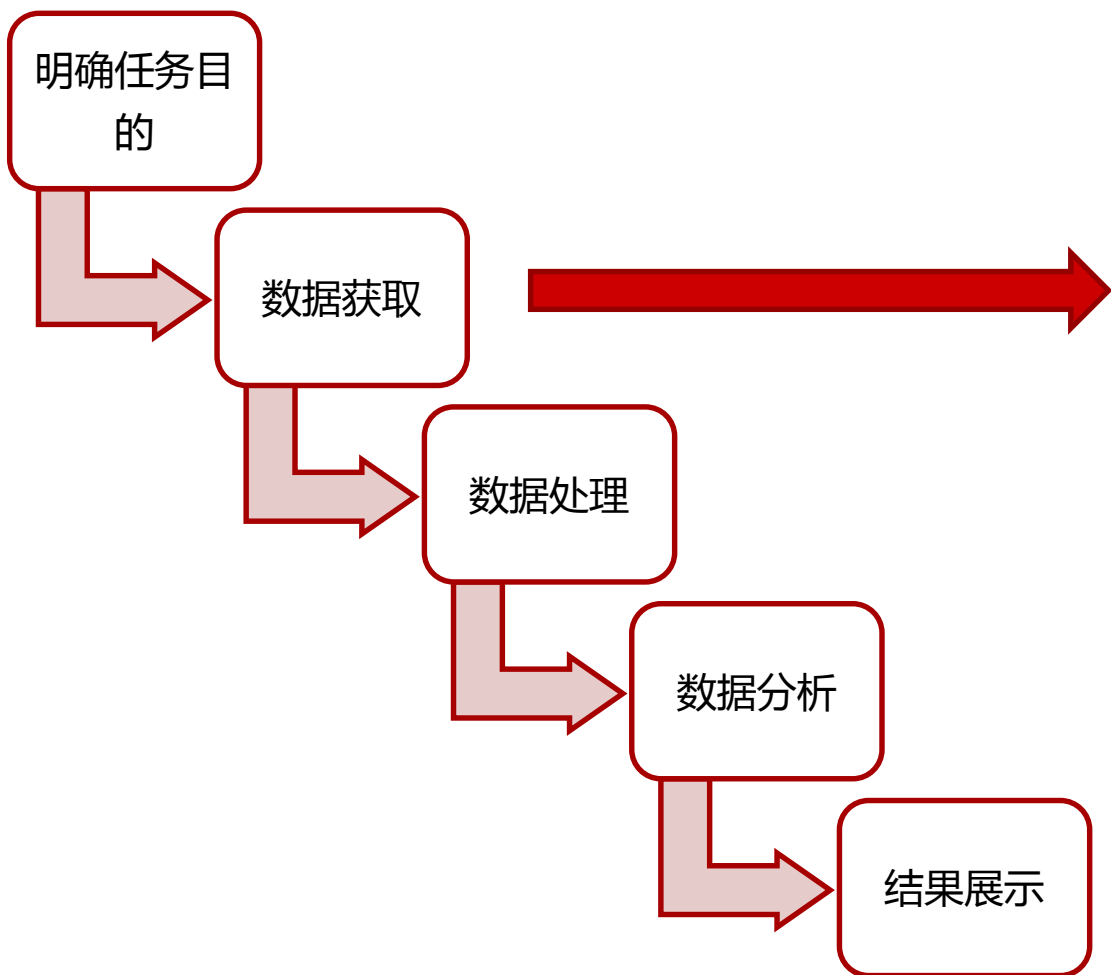
— 骨科疾病预测

Robin

目录

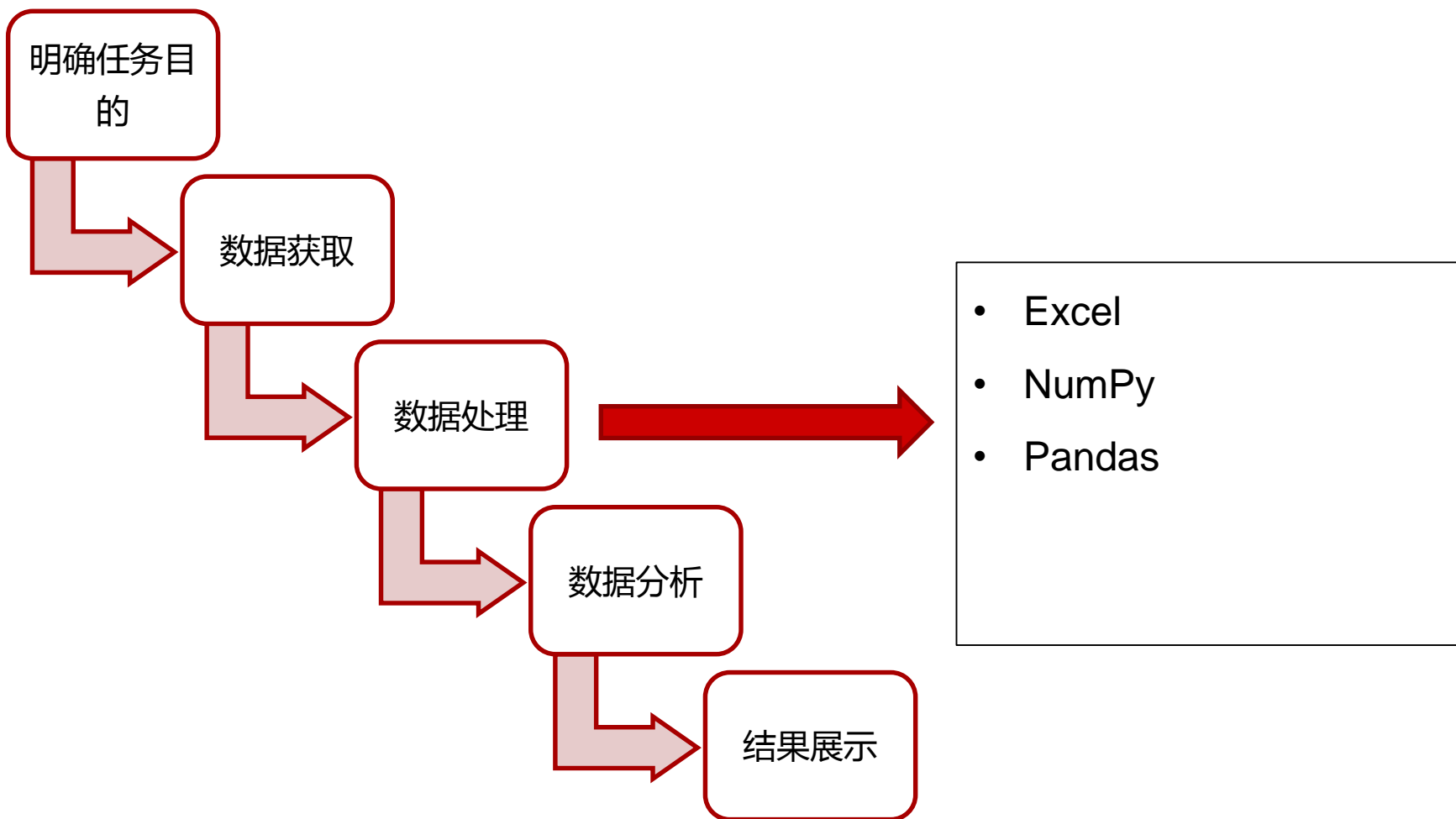
- 数据分析基本步骤
- 数据分析建模理论基础
- k-近邻算法
- 案例实战

数据分析基本步骤

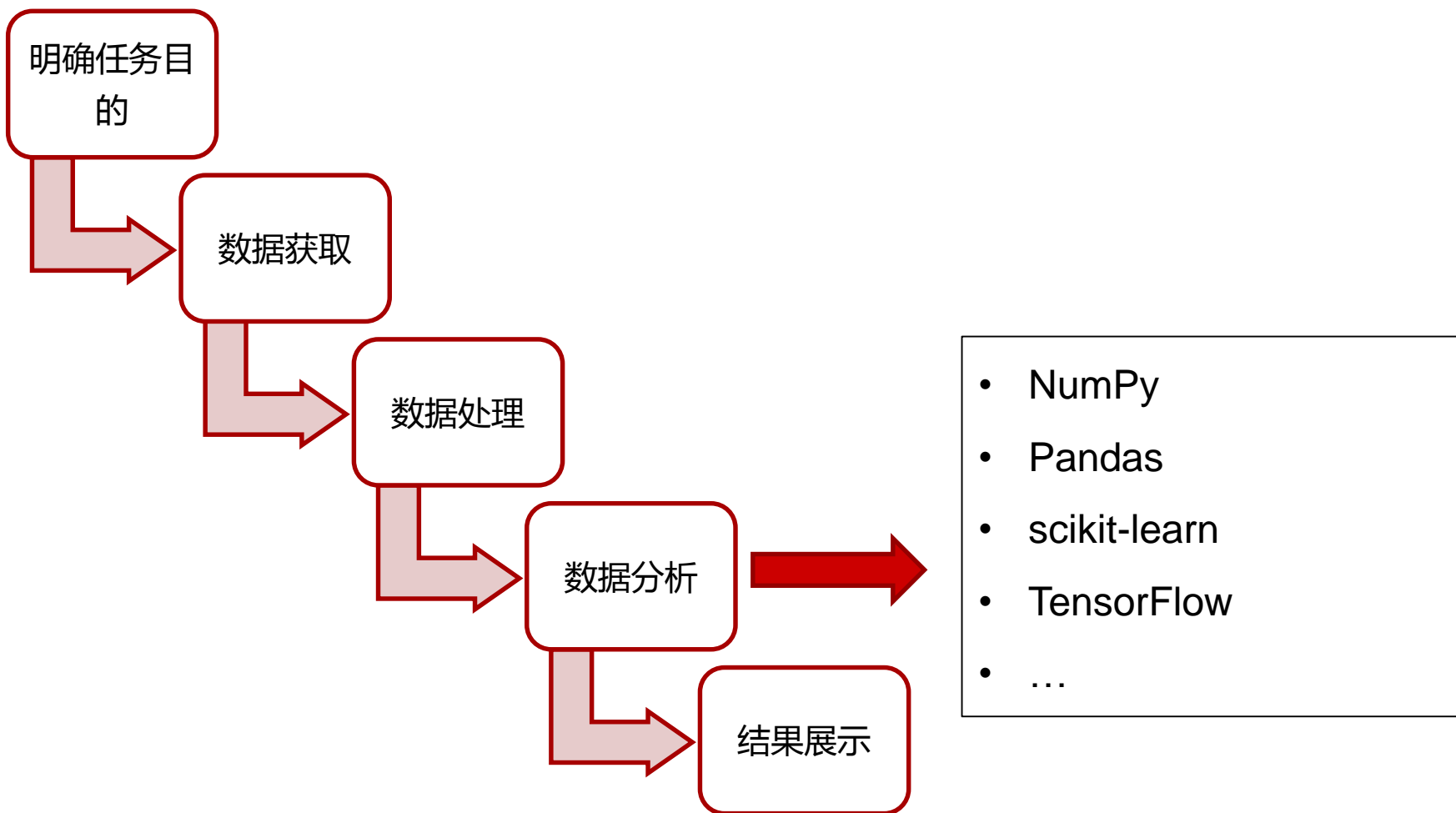


- 客户提供
- 自动化收集
 - 网络爬虫
- 手动收集
 - 调查问卷
- 途径
 - 公开信息
 - 外部数据库
 - 自有数据库
 - 调查问卷
 - 客户数据

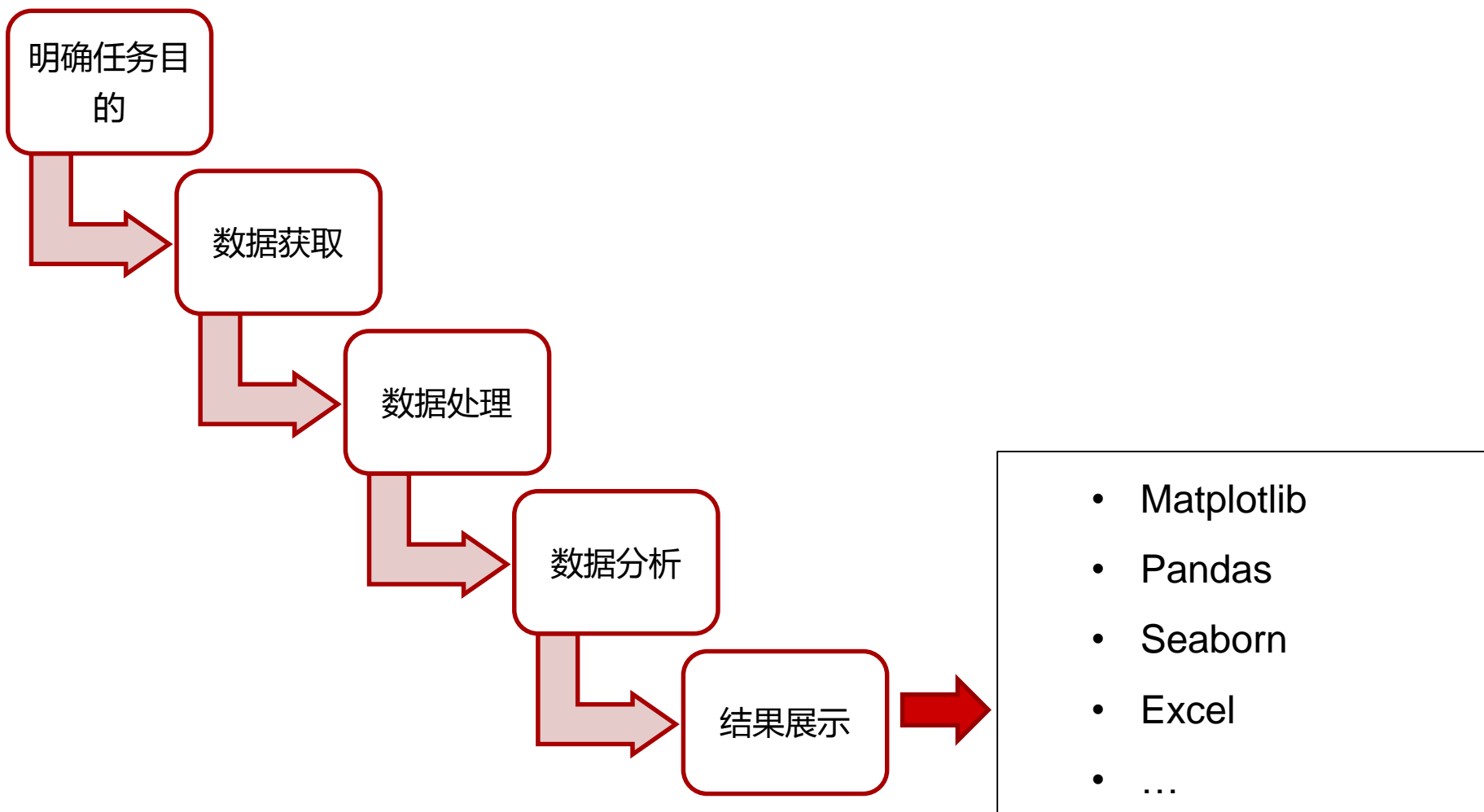
数据分析基本步骤



数据分析基本步骤



数据分析基本步骤



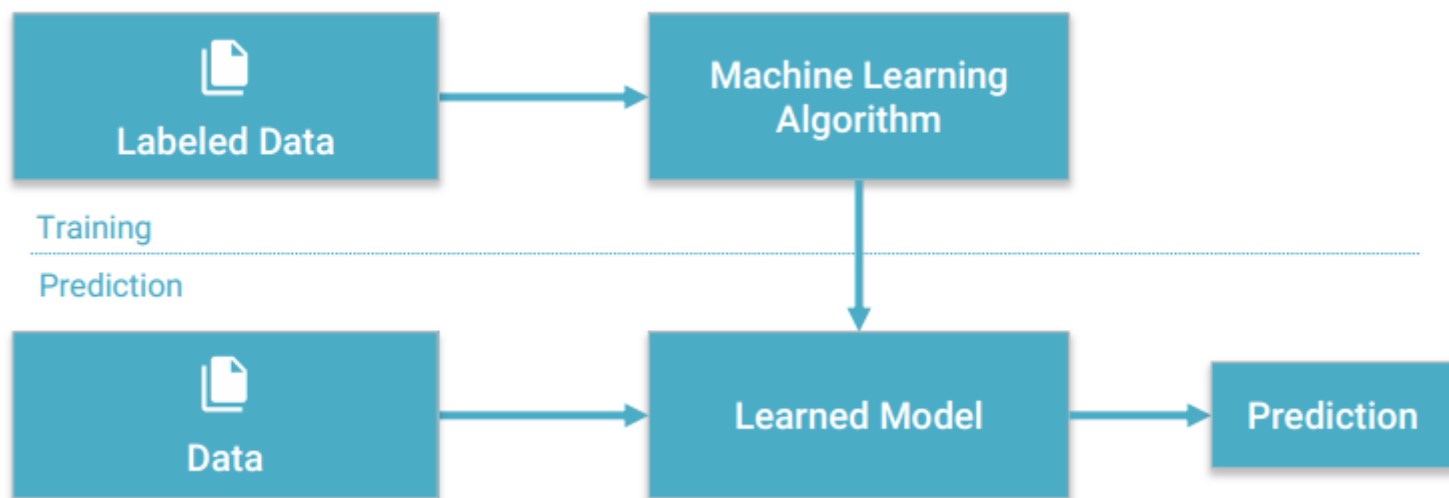
目录

- 数据分析基本步骤
- 数据分析建模理论基础
- k-近邻算法
- 案例实战

数据分析建模理论基础

定义

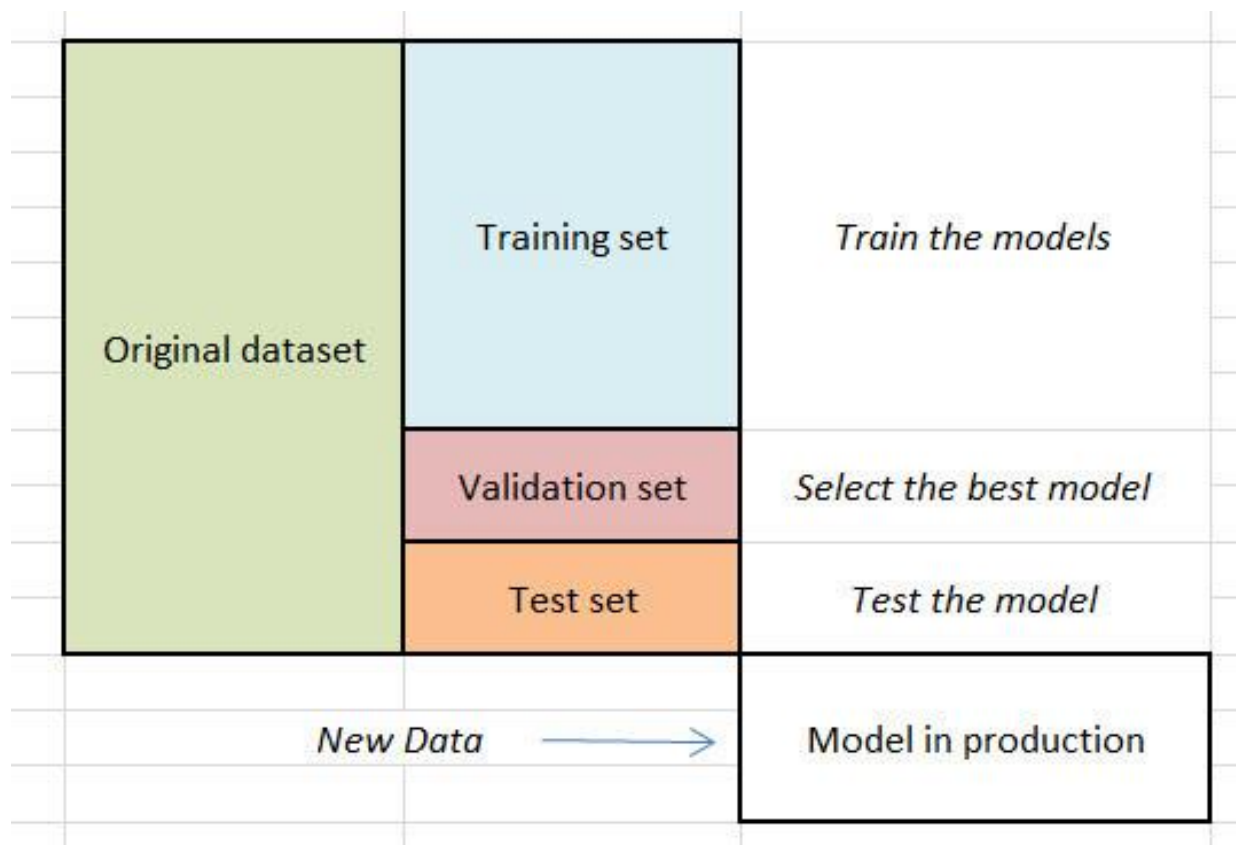
- Machine Learning is a type of Artificial Intelligence that provides computers with the ability to **learn without being explicitly programmed**.
- Provides **various techniques** that can learn from and make predictions on **DATA**.



数据分析建模理论基础

机器学习

- 训练集 vs 验证集 vs 测试集



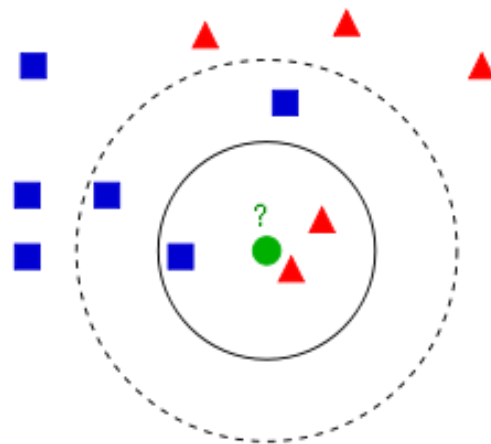
目录

- 数据分析基本步骤
- 数据分析建模理论基础
- **k-近邻算法**
- 案例实战

k-近邻算法

kNN (k-NearestNeighbor), k-近邻算法

- 是一种基于样本/实例的算法
- 步骤：
 1. 计算出测试样本和所有训练样本的距离；
 2. 为测试样本选择k个与其距离最小的训练样本；
 3. 统计出k个训练样本中大多数样本所属分类；
 4. 这个分类就是待分类数据所属分类



k-近邻算法

kNN (k-NearestNeighbor), k-近邻算法

- 需要明确的问题
 1. 相似性度量，如：欧氏距离等
 2. 紧邻点个数，通过交叉验证得到最优紧邻点个数
- kNN优缺点

优点	缺点
<ul style="list-style-type: none">• 算法简单直观，易于实现• 不需要额外的数据，只依靠数据（样本）本身	<ul style="list-style-type: none">• 计算量较大，分类速度慢• 需要预先指定k值

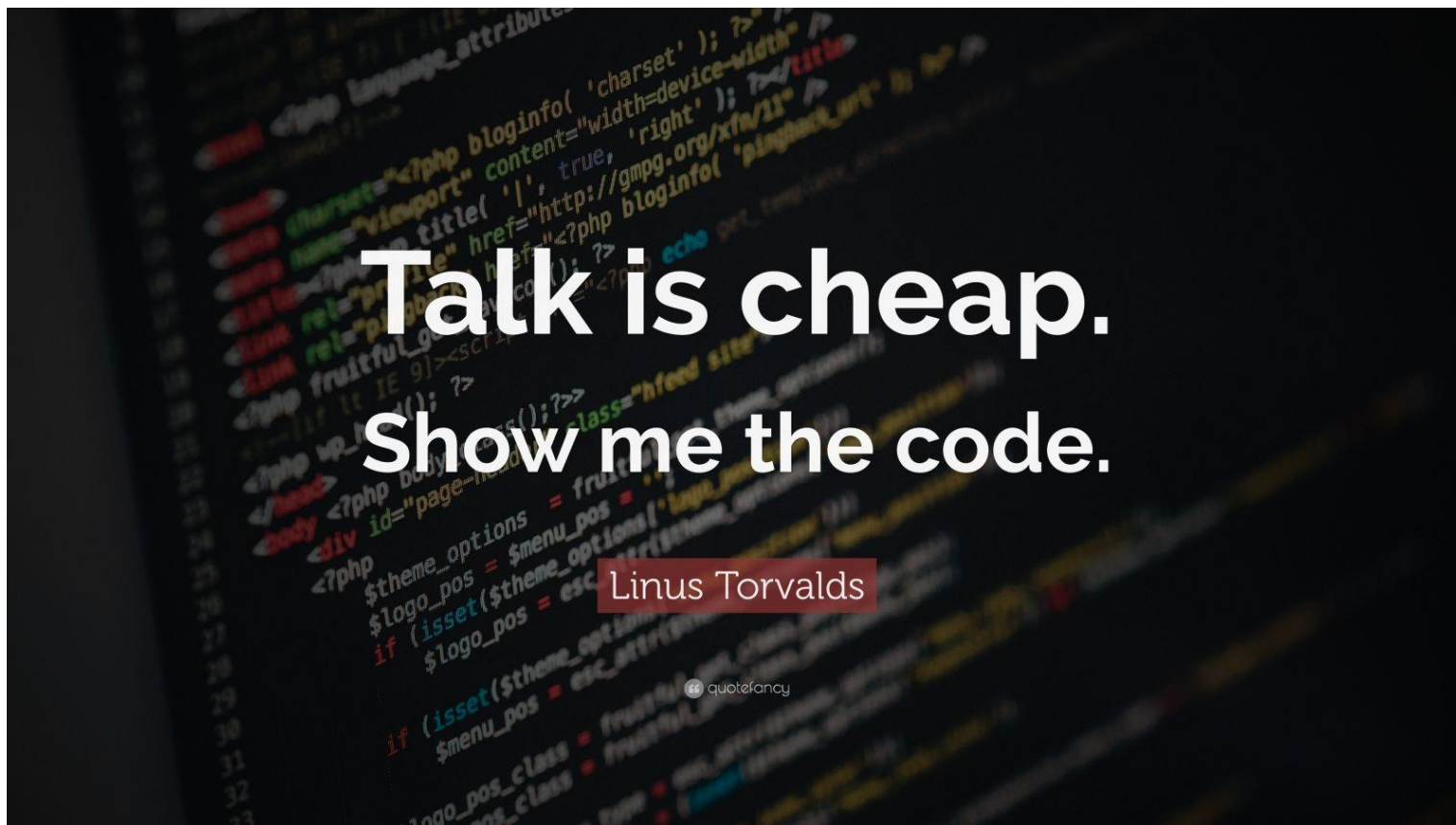
目录

- 数据分析基本步骤
- 数据分析建模理论基础
- k-近邻算法
- 案例实战

实战案例

项目名称：python机器学习小案例源码--骨科疾病预测

- 请参考相应的配套代码及案例讲解文档



疑问

□ 问题答疑：<http://www.xxwenda.com/>

■ 可邀请老师或者其他人回复问题

小象问答邀请 @Robin_TY 回答问题

