

介绍 1-1 辅助海关红酒定价决策的背景及解决方案

建议课时：10 分钟

一、 教学目标

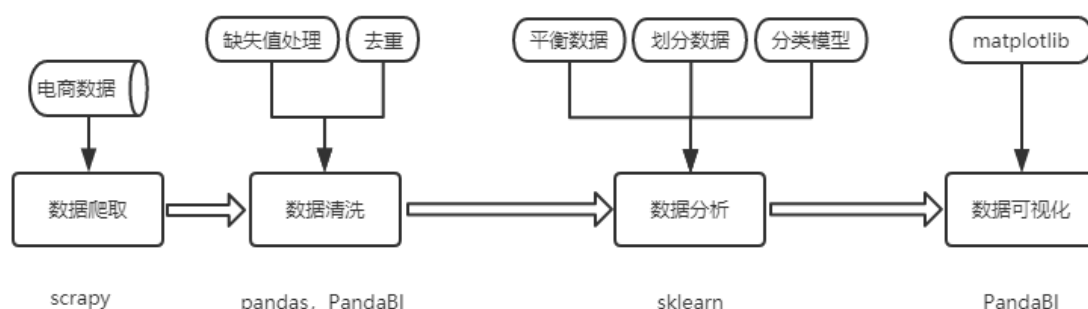
1. 了解海关背景下红酒定价预测的影响
2. 回顾数据分析的一般流程及每个阶段的基本处理

二、 背景介绍

随着时代的发展，国内外交易甚是频繁，进出口贸易的检查也需要越来越多的人工审核。以红酒类商品为例，为避免谎报价格，出现逃税现象，人工需要对商品进行拆箱核实才可出关，很大程度上阻碍了商品的流通速度和降低其工作效率，但这又是不得不检查的部分。本案例将从红酒类商品的定价维度利用机器学习的算法进行预估区间，为审价类风险提供了判断依据，很大程度上能避免谎报价格发现的遗漏且降低人工开箱的频率，提高工作效率。该流程将来也可套用在其他商品上。

三、 架构分析

主要解决的问题是红酒定价预测问题，以辅助海关决策。本应是个回归问题，但网络上的电商平台售卖信息有不合理的部分，将改用分类问题来解决，即对价格规范一个区间，对区间进行预测。完整解决方案如下：



具体解释如下：

步骤	处理内容	工具/第三方包
数据爬取	根据预先整理的红酒品牌列表，检索京东的相关产品界面，爬取商品的描述信息	Scrapy
数据清洗	数据清洗是关键，具体涉及的操作有 <ul style="list-style-type: none">● 去重● 缺失值处理● 剔除非红酒信息数据● 特征选择● 处理红酒的单瓶价格● 定义价格区间● 等等	Pandas, numpy, matplotlib, PandaBI
数据分析	拆分训练集和测试集，构建算法，查看特征的重要性，再进行特征选择，运行算法，择优选模型	sklearn xgboost
数据可视化	本案例中可视化有两个用处：一是展示效果，二是查看特征分布	matplotlib, PandaBI