**BI第六期名企班笔记**

**20210411\_1.1\_模型融合与供应链预测**

**1.Overview**

**1.1 学习方法论：**Thinking & Action，复习为王

**1.2 Project Overview:** 三个阶段，八个主流业务&比赛场景（①智能供应链预测②智慧金融-金融风控-银行评分卡模型③资源调度优化场景- Santa服务调度安排④智慧交通-车辆/订单路径规划）

**1.3技能点Overview:** 三个阶段，四个类别（①②③④）（考虑自创建BI技能树：模型树 & 算法树)

**1.4学习目标:**

第一部分：二手车价格预测

Note:

重要知识点：回归分析算法总结回顾、模型融合、分类中的voting策略

第二部分：供应链预测项目

**2. 项目一：二手车价格预测**

**2.1回顾总结回归分析算法（Page6-8）**

**关键知识点：**

①什么是回归分析？

研究因变量（目标）与自变量（特征）之间的关系。帮助数据科学家，更好的选择最佳的变量集，用于建立预测模型。

使用场景：价格预测，数量预测

②常见的回归算法？

1. 线性回归 / 逻辑回归
2. 多项式回归 Polynomial Regression
3. 多项式是一种常用的特征构造方法
4. 岭回归 Ridge Regression
5. 套索回归 Lasso Regression
6. 弹性回归 ElasticNet Regression
7. 岭回归和套索回归的混合技术，同时使用L2和L1正则

③ 还可以使用哪些回归算法？

A. 分类用于回归：SVM，KNN，CART

B. 集成学习：RF，GBDT, XGBoost, LightGBM

**2.2项目实战：二手车价格预测（Page9-21）**

**关键知识点：四个步骤**

**#步骤一 数据加载及预处理，查看数据整体情况，查看缺失值，缺失值可视化，查看label的分布**

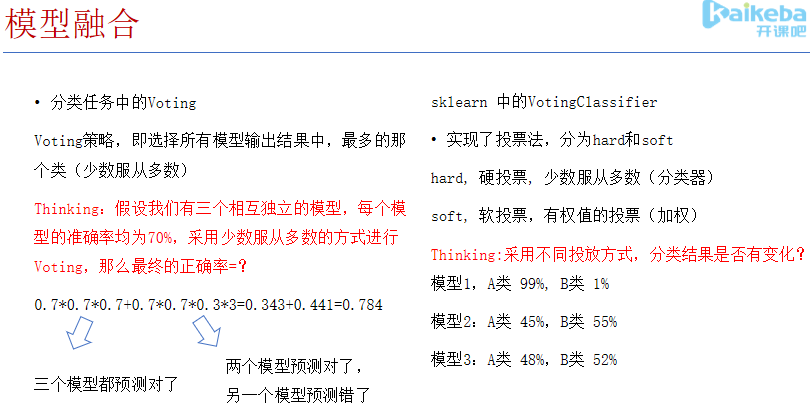
**#步骤二 特征选择**

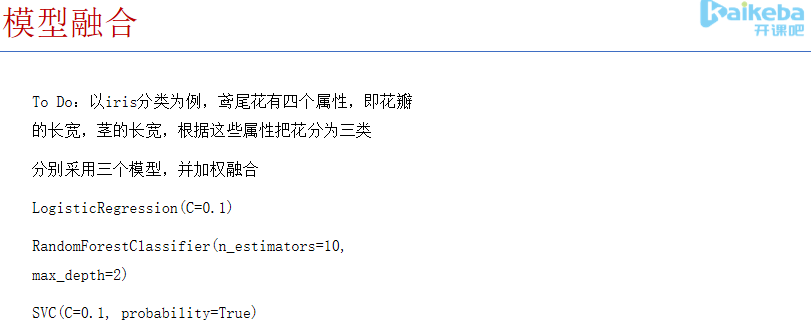
**#步骤三 模型训练，使用XGBoost，超参数设置**

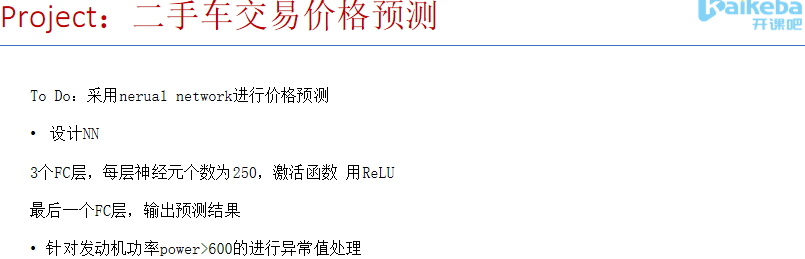
**#步骤四 模型预测**

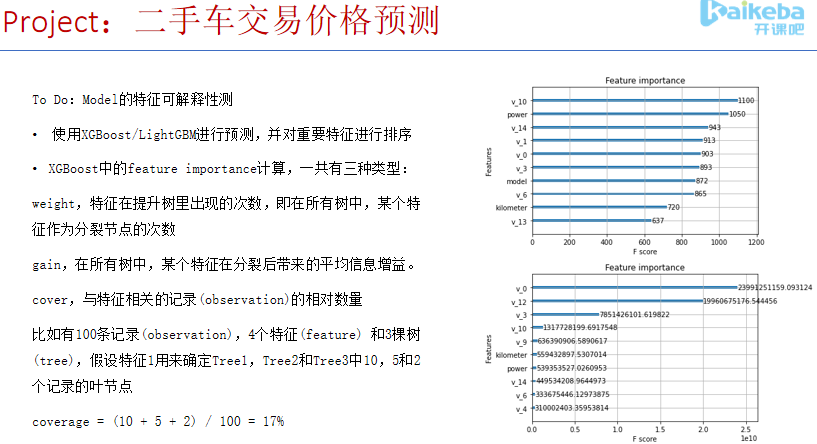
Note:

1. 计算偏度和峰度（与正态分布进行比较）：峰度用来描述取值分布形态陡缓程度的统计量。峰度为0表示该分布与正态分布的陡缓程度相同。峰度大于0表示分布与正态分布相比较为陡峭，为尖顶峰。峰度小于0表示分布与正态分布相比较为平坦，为平顶峰。
2. 有哪些提分（减少MAE）的方式？ 1、时间类型数据(regDate, creatDate)；2、模型融合，使用XGBoost,LightGBM；3、增加新的特征，对brand进行统计
3. 模型融合：回归任务中的加权融合，根据各个模型的最终预测表现分配不同的权重，比如准确率高的模型给予更高的权重，准确率低的模型给予较小的权重
4. 模型融合- 分类任务中的voting：









**3. 项目二：智能供应链分析**

**3.1项目实战：智能供应链分析（Page22-33）**

**关键知识点：**

**#步骤一 数据加载及预处理，查看数据整体情况，查看缺失值，缺失值可视化，查看label的分布**

**#步骤二 创建矩阵&绘制热力图**

**#步骤三 创建时间戳**

**#步骤四 使用RFM对用户进行分层**

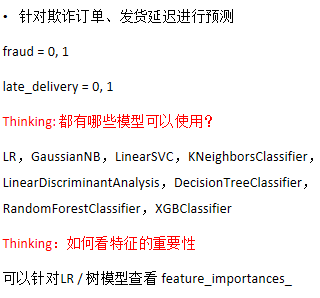
**#步骤五 取数值最大的Top10进行条形图展示**

**#步骤六 求得并设置分位数**

**#步骤七 显示不同地区的支付类型情况**

Note:

延申思考1：针对欺诈订单、发货延迟进行预测



延申思考2：针对销量、订单数量进行预测

