

Lesson 18 supply chain

关于上节课二手车项目的处理:

在 train set 上对 brand 做特征工程: 首先在 train set 上对 price > 0 做 filter 然后聚类做统计形成新特征, 然后在 test set 上用 brand 做 merge 到 test 上, 这样 train 和 test 的维度就一样了

标签泄露: 特征与 label 进行了关联, 如果某个构造的特征与 label 进行关联 model 进行预测的时候, 训练集结果会很准确
适当的标签泄露是有价值的, 但取大于你 value-count 个数

欠拟合的原因: $n_estimator$ 太少, learning rate 太小

RFM 指标:

Recency 最近一次消费时间间隔 Frequency 消费频率, 一段时间内消费次数

Monetary 消费金额, 一段时间内的消费金额

supply chain: order status \rightarrow 分类 sales Order Item Quant } 回归

判断欺诈: 人工记录有没有发货

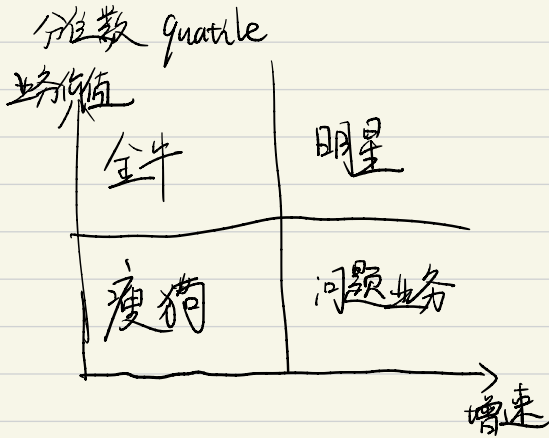
RFM 把用户分为了 8 种: 重要价值用户

	R	F	M
重要	H	H	H
发展	H	L	H
保持	L	H	H
挽留	L	L	H
一般	H	H	L
价值低	H	L	L
发展低	L	H	L
保持低	L	L	L

MSRF 重要客户被挖走的原因: 没有用户分类; 识别有价值的用户

业务规则来划分: ①等频 ②等距

R value 越小越好 R score 越大越好



2x2

四象限法则

方法论 → 策略

算法 → 工具