

2021中国大学生保险 数字挑战赛数字赛道

团队名：scu_team

成员：谢东霖、陆琪

学校：四川大学

赛题背景

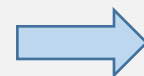
背景

2020年，疫情和车险综合改革下，车险保费增速下降，单非车险保费保持较高增速；且60%的非车险是在车险销售环境下购买。
而车险作为刚需产品，因此为非车险等多种保险产品销售提供了有效通道。



需求

通过对数据进行探索建模，挖掘客户在车险销售场景下购买非车险的意向



业务应用

结合挖掘建模得到的可解释性信息进行业务应用

赛题数据

用户基本信息

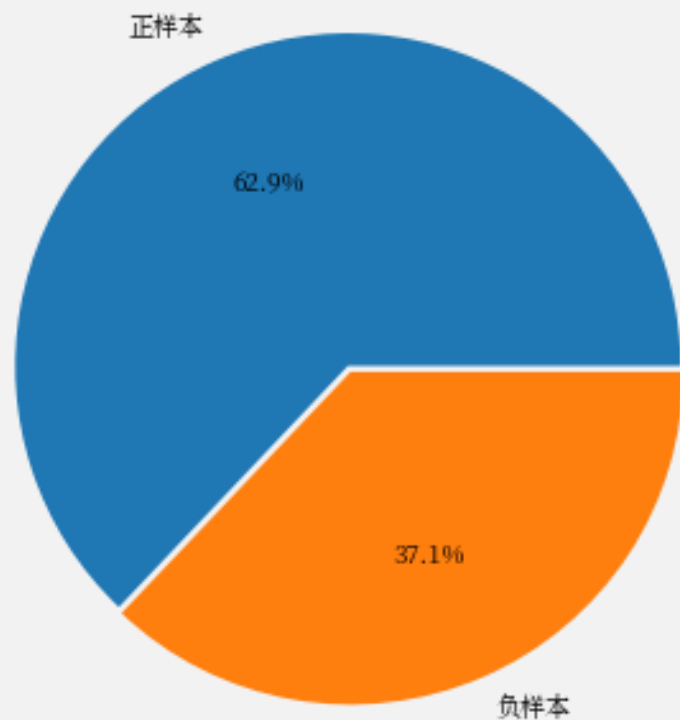
字段名	含义
p1_gender	性别
p1_age	年龄
p1_census_register	籍贯
birth_month	出生月份
p2_marital_status	婚姻状况
f1_child_flag	是否有子女
p2_is_child_under_15_family	家庭是否有 15 岁以下小孩
p2_is_adult_over_55_family	家庭是否有 55 岁以上老人
p1_is_bank_eff	是否银行有效客户
p2_is_enterprise_owner	是否为企业主
p2_is_smeowner	是否为小微企业主
f2_posses_house_flag	是否有房
f2_cust_housing_price_total	房产总价
newvalue	新车购置价
.....	

保险基本信息

字段名	含义
dpt	保险机构
xz/xb	险种/险别
nprem_ly	上年的车险保费
ncd_ly	上年ncd系数(无赔偿优待)
nprem_od/si_od	车损险保费/保额
nprem_bt/si_bt	盗抢险保费/保额
nprem_tp/si_tp	商业三者险
nprem_vld/si_vld	司机座位责任险保费/保额
nprem_vlp/si_vlp	乘客座位责任险保费/保额
.....	
suiche_nonauto_nprem(16-20)	16-20年随车非车保费
suiche_nonauto_amount(16-20)	16-20年随车非车保额
dur_personal_insurance_90	点击时长非车险

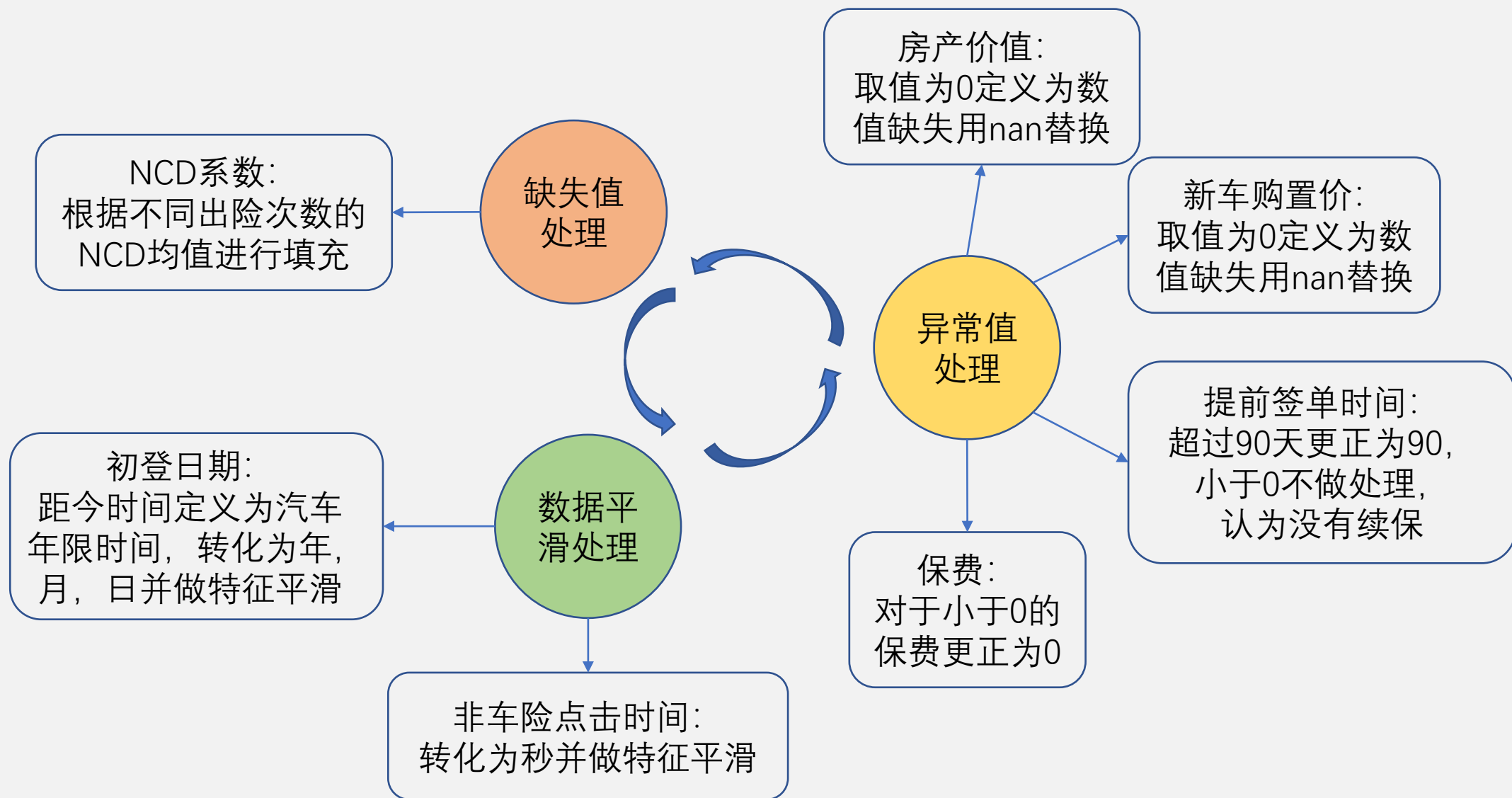
数据探索

一月份车险用户购买非车险的比例

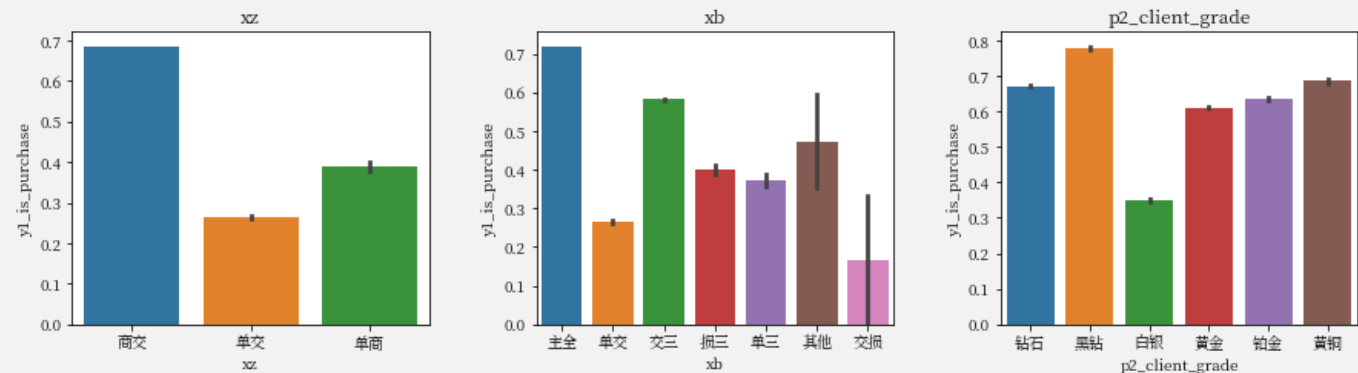


一月份在车险销售场景下购买非车险和未购买客户之间比例约为6: 4, 正负样本比例比较均衡, 且在车险销售场景下购买非车险的客户占比更多。

数据探索——数据预处理

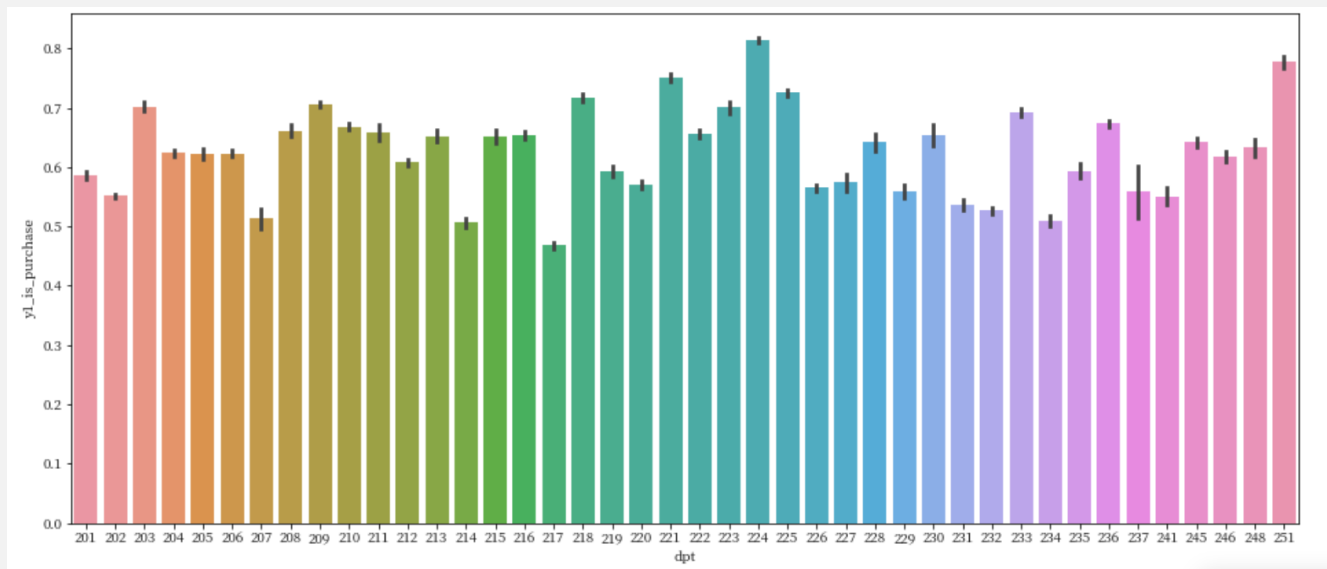


特征工程——单因子分析



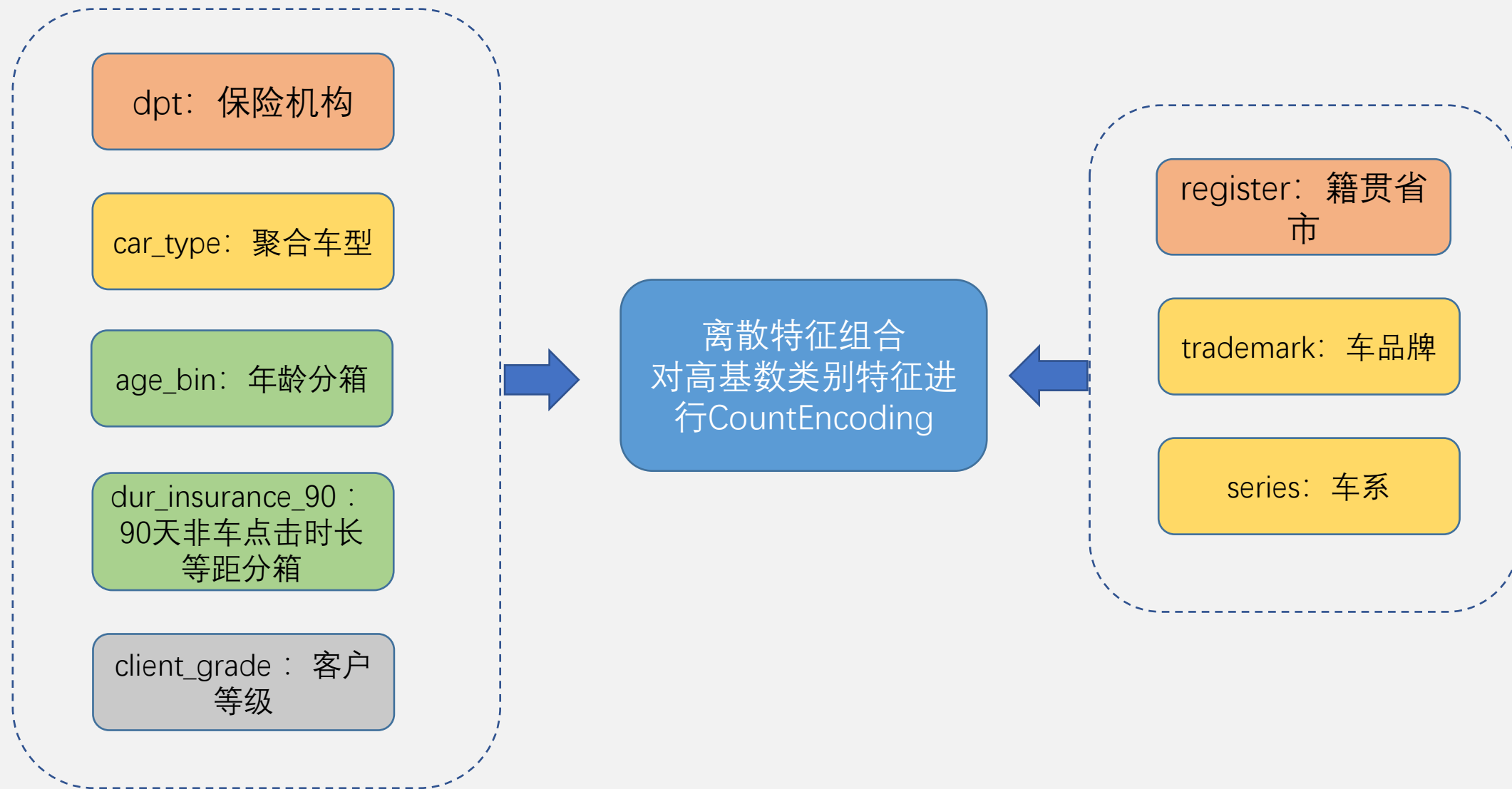
1.离散属性

对于**险别、险种、机构以及用户俱乐部等级**而言，在每个类别上购买非车险的比例差异比较明显

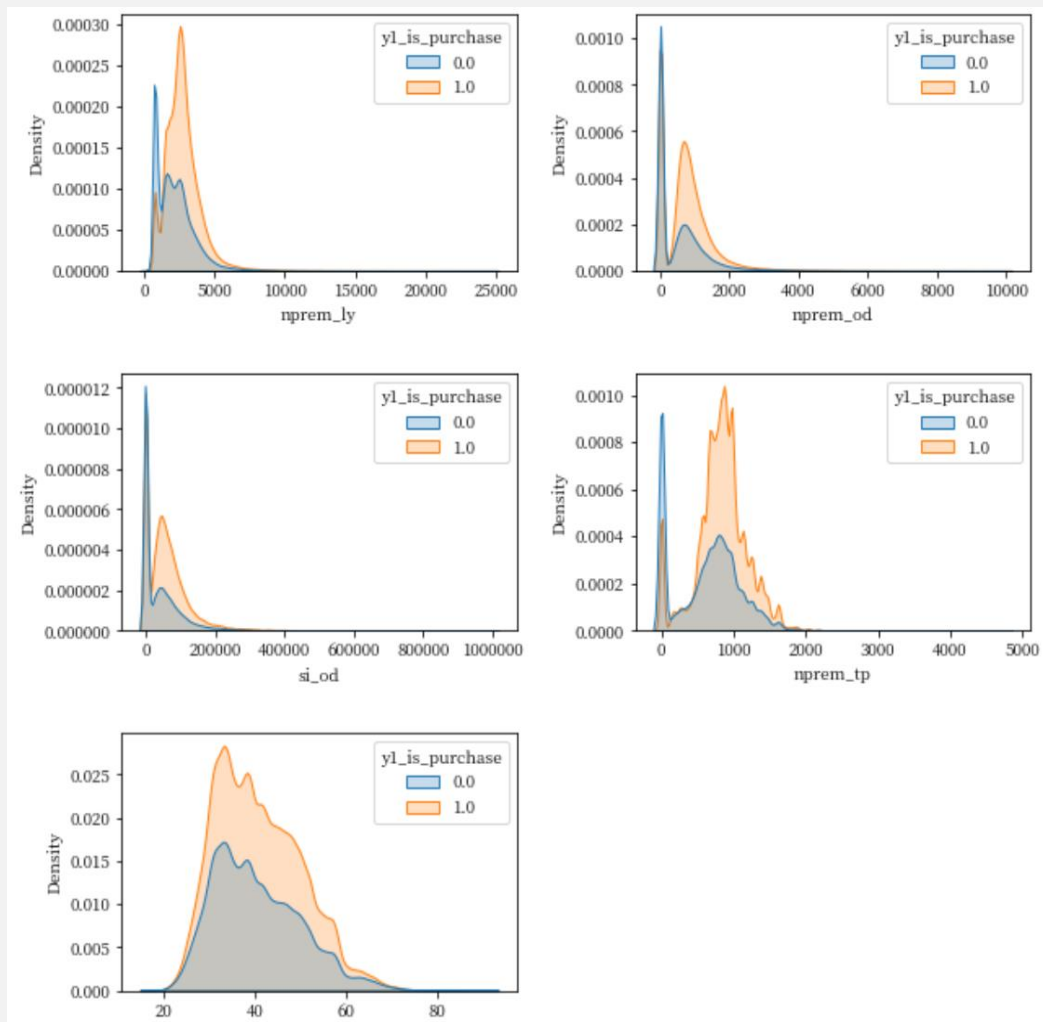


其中对于**机构**而言，在模型的特征重要程度上是最大的。

特征工程——离散化特征构造



特征工程——单因子分析



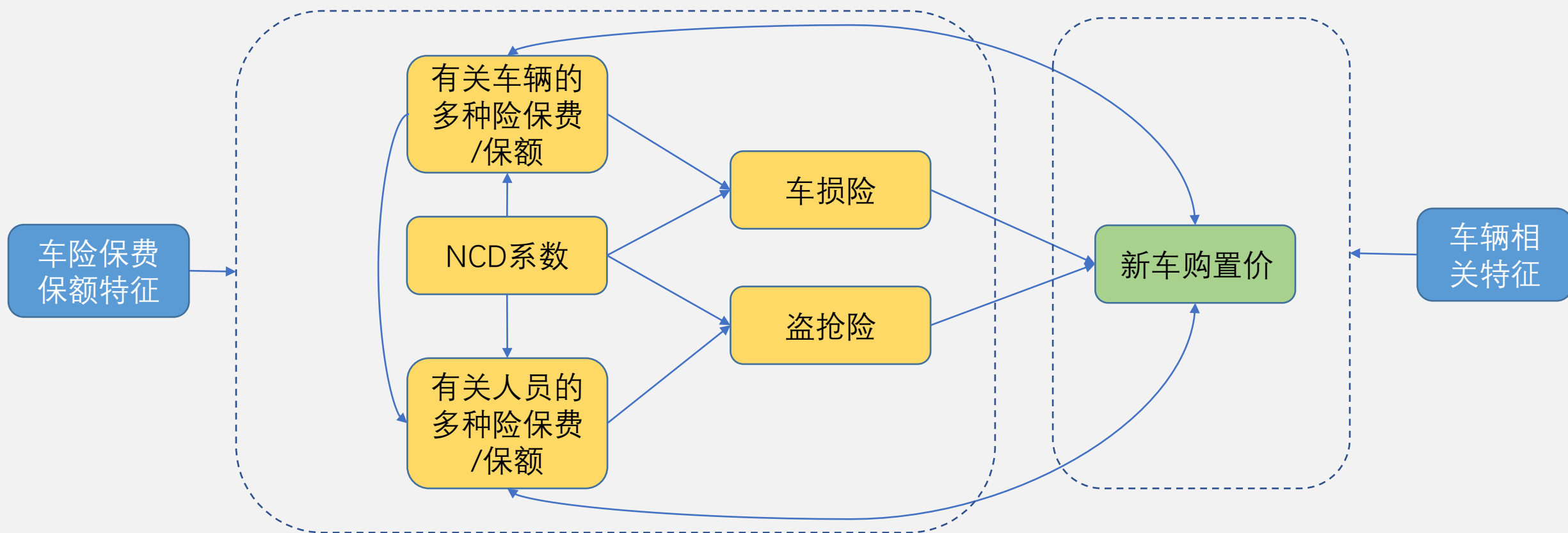
2.连续纸属性

购买非车险和未购买非车险的客户在上年车险保费、车损险保费/保额、商业三者险保费上的分布存在一定的差异

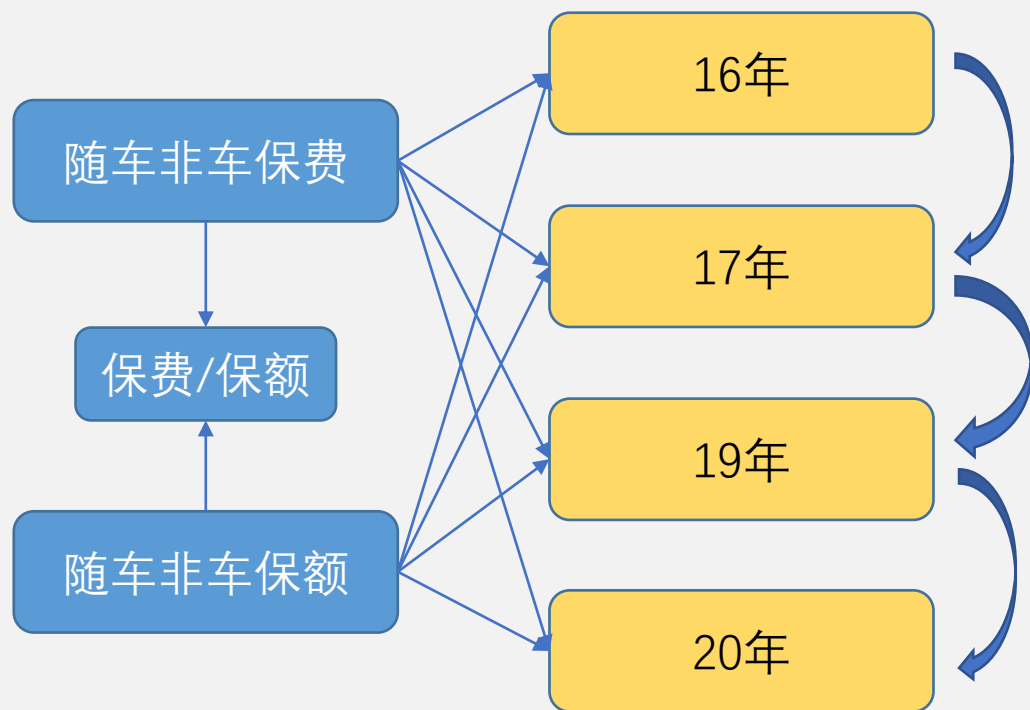
(1)在车险销售场景下，购买非车险的用户的车险保费、车损险保费保额、商业三者险保费普遍更高。

(2)年龄在30-55区间内，购买非车险的用户比例明显高于其他年龄段

特征工程——连续特征构造



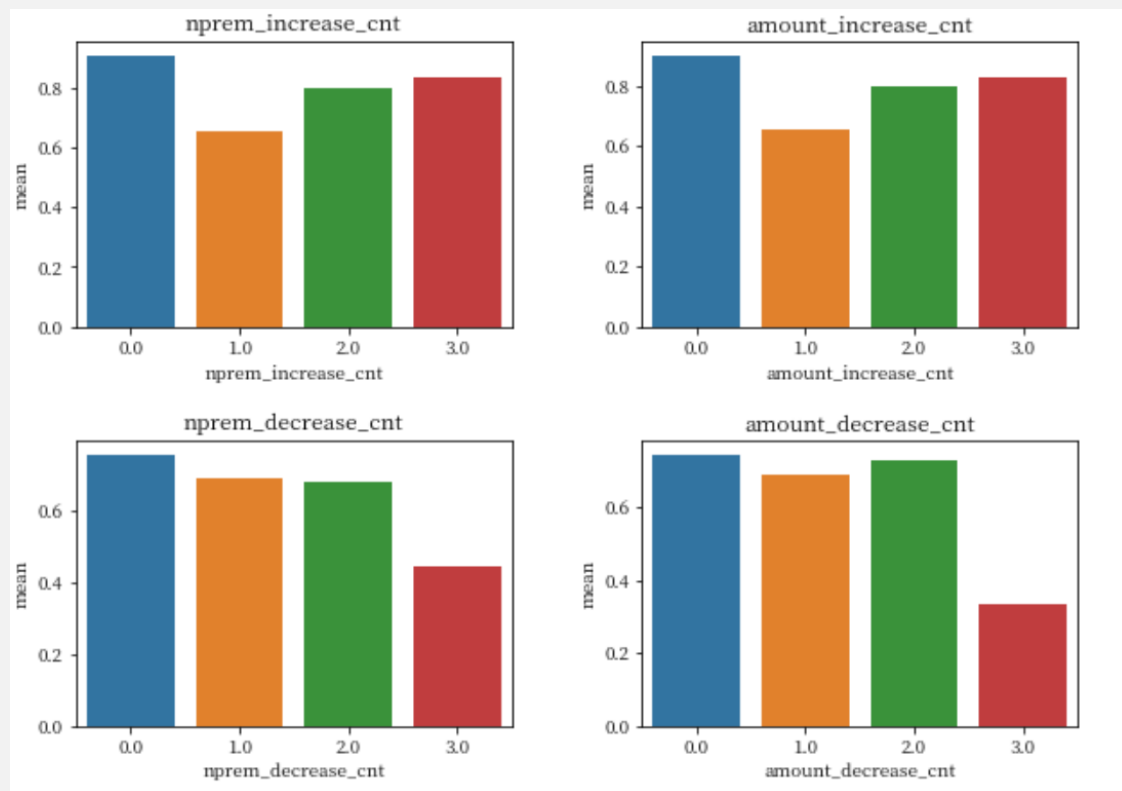
特征工程——历史特征



考虑历史购买非车险可能会直接影响用户现在购买非车险的意愿。

- 当前年份相比与前一年,保费差值及增加/下降的年份计数;
- 当前年份相比于之前历年,非车保费/保额的比例;
- 近五年保费, 保额, 保费/保额的变化趋势 (逐年增加/逐年减少/有增有减) ;
- 每年非车保费变化的差值, 方差及均值。

特征工程——历史特征



近五年内，随车非车保费/保额相比前一年保费/保额增加的年份数。

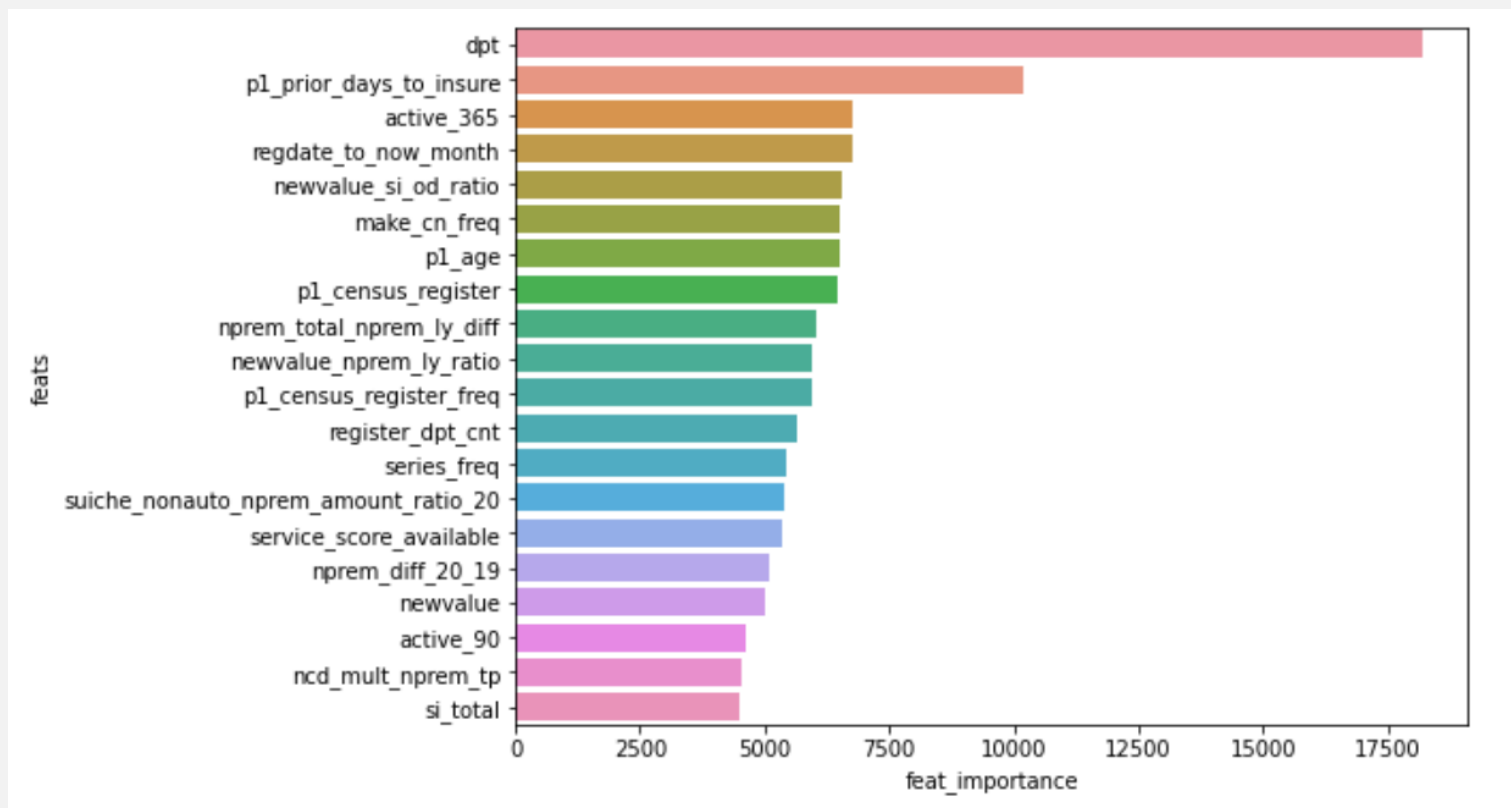
例如.近五年保费增加的年份数：
其中0表示没有增加: 可能是部分用户近五年都可能没有随车非车保费的购买，这部分用户不易从历史数据中推断，因此取值为0的不易于直接分析。

通过分析增加年份数1，2，3发现，保费增加的年份数越高，购买非车的比例也会更高。

反之同理，对于保费减少的年份数而言，减少的年份数越多，购买非车的比例是会更低。

特征工程——特征选择

Lightgbm模型输出特征重要度，降序排序，选择top20可视化



1.机构

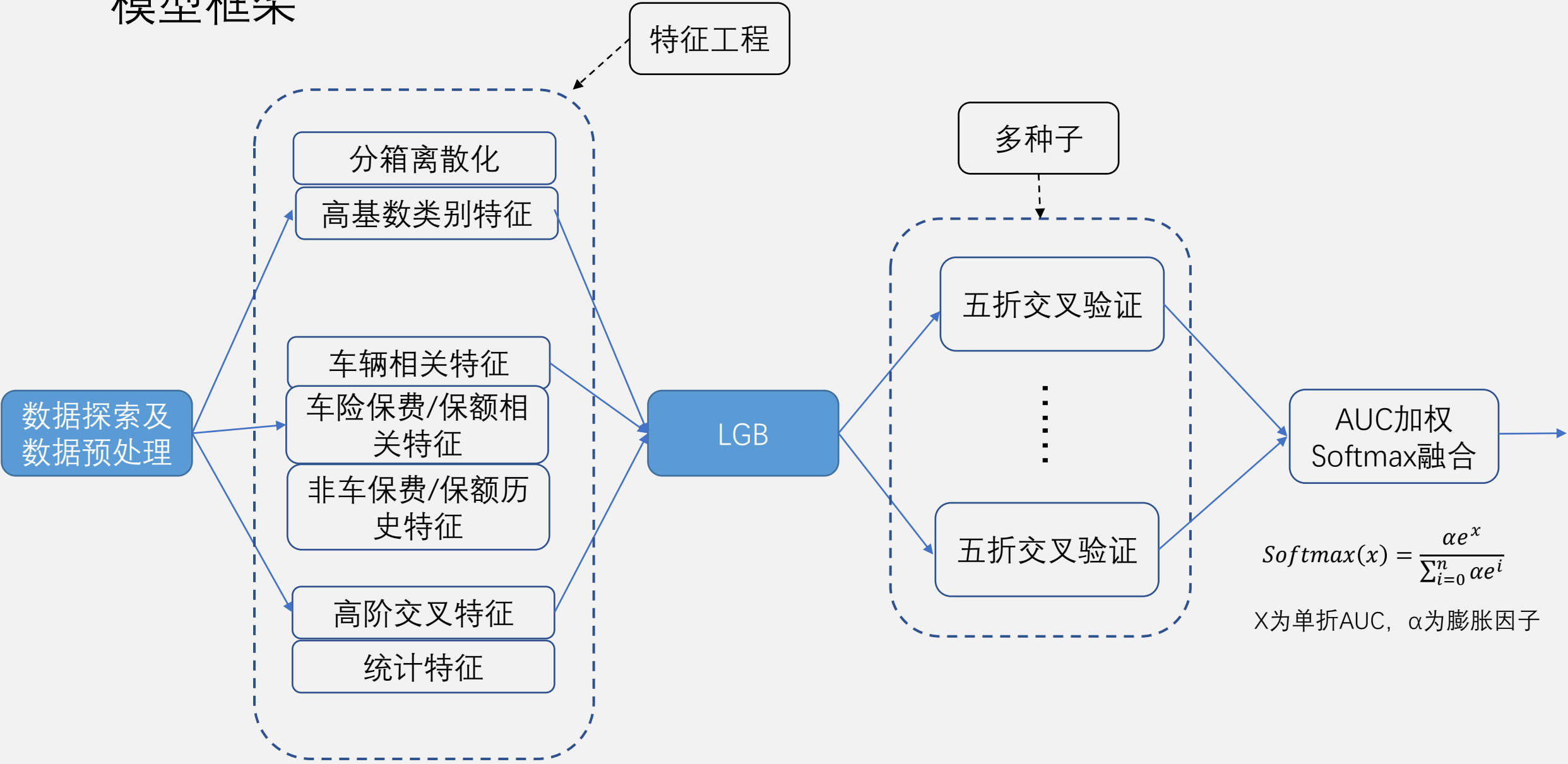
2.提前续保时间

3.近一年的活跃
天数

4.新车价格和车
损保额的比例

5.车型的
CountEncoding

模型框架



$$Softmax(x) = \frac{\alpha e^x}{\sum_{i=0}^n \alpha e^i}$$

X为单折AUC, α 为膨胀因子

业务应用

