风控面试

1.进件渠道(60%会问到)
线上业务:信息流、APP、微信公众号等线下业务:地摊导流、网点进件、合作企业团办、客户自己申请等

2.策略制定的步骤(20%会问到)
策略主要是根据业务中的风险点，寻找有效的特征进行防范。将变量进行特征重要性排序，用排名较高的/高 IV 的变量用作策略，一般命中策略的坏样本浓度要达到 3 倍以上，同时也要按月回溯策略的命中率和逾期率，尽可能少影响通过率的情况下框住坏的客群。弱变量/低 IV 的变量可以放到模型中，同时要注意策略用到的变量和模型用到的变量尽量不要有相似的，这样可以减少策略与模型的耦合

3.贷前策略包括哪些数据(80%会问到)
-般数据源类型分为决策类和排序类。决策类有黑名单类(多头、逾期、黑产、失信、罪犯等)，验证类(学历、社保公积金、运营商实名与在网时长、地址信息、收入信息等)，刻画类(关注类、消费画像、第三方规则)，排序类有评分类(芝麻信用分芝麻欺诈分等)。

4.说说策略是怎么做优化的?(100%会问到)策略调优分为几步:
(1)确认是 A 类调优还是 D 类调优。
D 类就是降逾期指标，在通过客群中找差客户拒绝;A 类就是提通过率回捞，在拒绝的客群中找好客户通过，
(2)量化分析调优闽值。
D 类调优离线即可完成分析，根据逾期指标选定 Y(FPD1/FSTPD1/M4+等)，比较逾期指标上升前后的客群异，找到逾期率发生变化的原因。然后寻找单变量或者组合变量进行分析，识别出逾期率较高的客户进行拒绝。
A 类调优需要决策引擎标记豁免样本，比较通过率下降前后的客群差异(新老客户/新老资产/渠道变化等)，寻找拒绝率较高的可放松的拒绝规则，放松阈值进行 AB 测试。
(3)预测策略调整的效果
根据历史数据回溯每月数据，分析策略调整对通过率、逾期率的变化。
(4)调整后观察和验证结果是否与预期一致试验一段时间后，对上与不上策略的样本进行 vintage 分析，观察策略上线是否对逾期指标有影响以及影响是否与预测一致。

5.怎么做数据清洗的?(80%会问到)缺失值处理:缺失值处理的方法有剔除、填补以及不处理三种方式，
异常值处理:了解异常值出现的原因，根据实际情况决定是否保留异常值。

常变量/同值化处理:对同值较高或者方差较低的变量作剔除，
分类变量降基处理:分类变量可以根据 bad rate 编码后再做分箱，也可以将少数类合并成一类，确保每一类中都有好坏样本。

6.怎么做特征衍生的?(60%会问到)RFM 方法。
R(Recency):客户最近一次交易消费时间的间隔。R 值越大，表示客户交易发生的日期越久，反之则表示客户交易发生的日期越近。
F(Frequency):客户在最近一段时间内交易消费的次数。F 值越大，表示客户交易越频繁，反之则表示客户交易不够活跃。
M(Monetary):客户在最近一段时间内交易消费的金额。M 值越大，表示客户价值越高，反之则表示客户价值越低，

常规统计特征:统计函数最大值、最小值、平均值、标准差来描述以上分布特征
时间距离特征:客户最远一次、最近一次或者某个特殊事件发生的时点。
行为波动特征:刻画客户某段连续时间内的行为变化特征。
集中度特征:用以刻画客户行为的偏好程度举一些根据征信报告还款历史衍生的例子:
近 3 个月总逾期次数、近 6 个月最大连续逾期次数、最近 1 次逾期距今月数、近 12 个月逾期连续增加次数、近 12 个月逾期增加次数、近 12 个月每两个月之间增长的最大值、近 12 个月取最大值距今月数等

7.怎么做特征筛选的?(60%会问到)
特征选择的话常见的有 IV 值、相关系数、稳定性 CS1、逻辑回归系数一致、逻辑回归变量显著性:xgb 特征重要度。逻辑回归评分卡筛选变量的步骤案例如下:
1、保留 IV 值大于 0.02 的变量，共 500 个;
2、把初筛的到的量进行 WOE 编码;
3、变量间两两相关检验并筛选，删除相关性大于 0.7 的变量 400 个，剩余 100;
4、变量稳定性检验，把稳定性大于 0.05 的变量删除，剩余 60 个;
5、逐步回归法筛选最终入模变量，剩余入模变量 10 个。

8.怎么做特征分箱的?(60%会问到)类别型变量进行降基处理(看是否需要)后分箱
数值型变量等频分箱、等距分箱、决策树分箱、卡方分箱、手工分箱。
分完箱之后看 woe 与坏账率是否单调或者符合业务意义，如不符合再手动进行调整

9.目标变量怎么定义?(100%会问到)
贷前模型的 Y 主要通过 vintage 和迁徙率。vintage 确定观察期，迁徙率确定逾期多少为坏。

10.模型是怎么调参的?
先用交叉验证方法初步检验模型可以达到的上限作为 baseline，调参方法可以从训练速度、精度过拟合三个方面回答，一般用网格搜索或者贝叶斯优化。