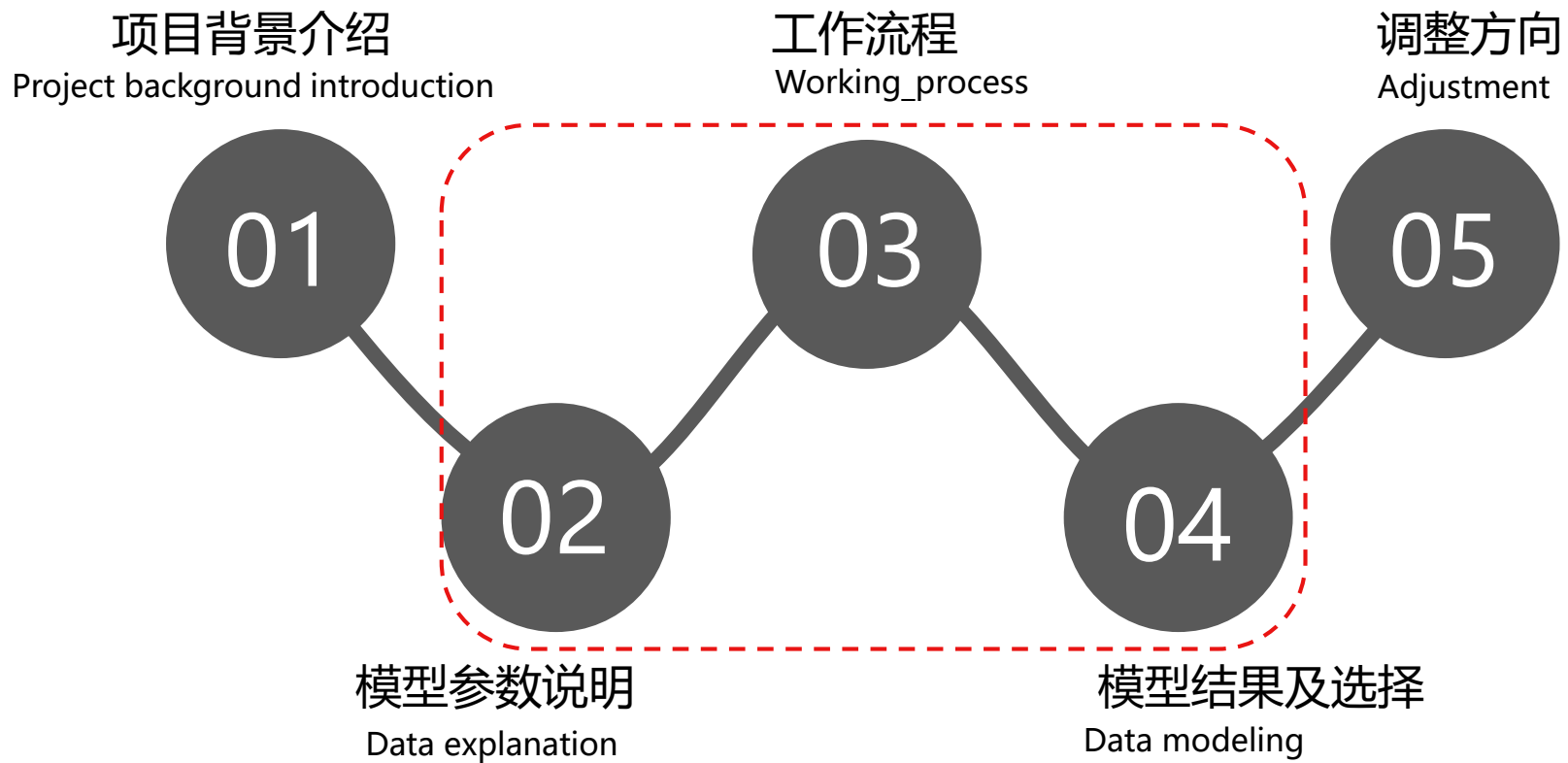


防盜刷及客群画像

防盜刷尝试性建模

商业研究部

主目录 CONTENTS



工作内容概述及范例展示



01 项目背景介绍

01.项目背景介绍

分析目的

- 深入了解盗刷用户画像及行为偏好 (客户画像, 聚类模型) 评分, 速度, 解释性
 - 挖掘出盗刷的关键因素 (关键节点挖掘)
 - 通过算法预测用户是否盗刷 (预测模型)
 - 增加盗刷成本进而减少盗刷情况 (风控策略)
-

商业价值

- 通过监测盗刷行为, 对盗刷玩家进行画像测评, 提前监控潜在盗刷玩家
 - 减少盗刷成功现象, 降低对营收负贡献概率
 - 实时监测玩家行为, 对潜在盗刷玩家进行及时预警
-

数据说明

- 此次数据是08项目组所有数据,因不可控因素导致数据质量较差,故测试模型采用拟合填充数据会导致此次模型存在过拟合状态



02.模型参数说明

01.特征字段介绍


共 175154 行 X 25 列数据 (含标签列)

用户属性

用户储值行为

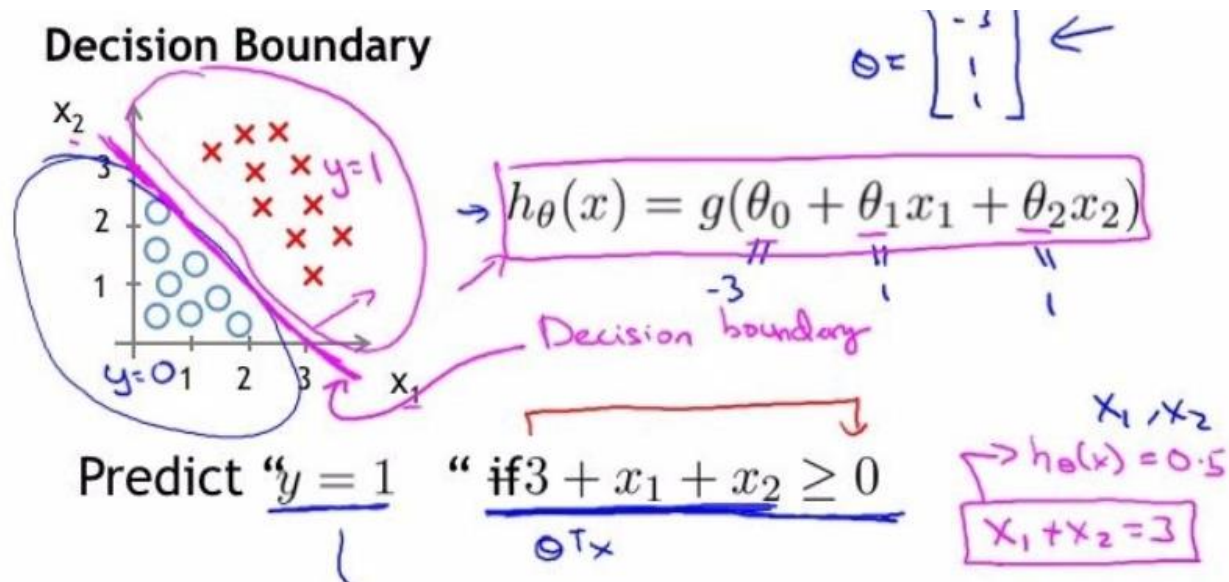
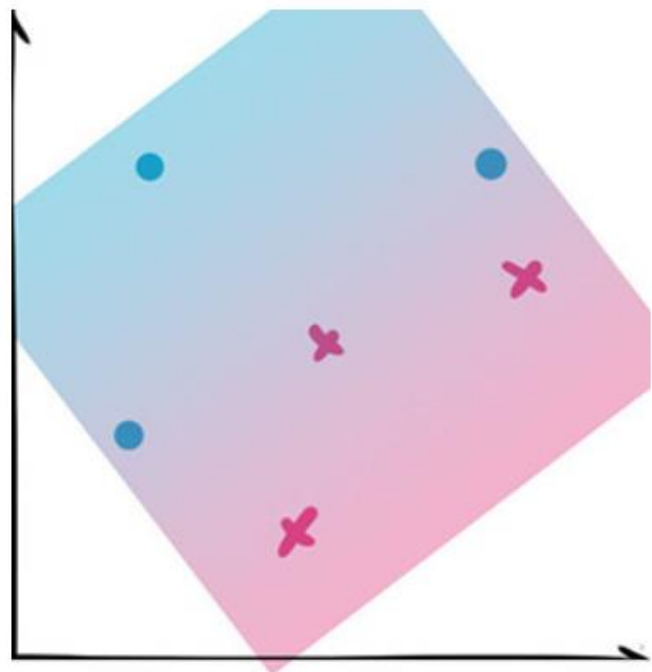
用户游戏

用户属性	用户储值行为	用户游戏	用户游戏				
栏位	定义	栏位	定义				
☀️👉☀️ 👉👉☀️👉 👉	玩家编号	♂️📁♂️□ 📁♂️♂️♂️ ○♂️	玩家首储值时间	♂️📁♂️♂️ 📁○♂️♂️ ♂️♂️	玩家第一个玩的游戏👉👉	👉♂️♂️♂️♂️♂️ ♂️♂️♂️♂️♂️	首次转账前点数
♂️♂️♂️ 📁♂️♂️♂️ ♂️○♂️	账号创建时间	♂️📁♂️♂️ 📁♂️♂️♂️ ♂️♂️♂️	账号创建时间	♂️📁♂️♂️ 📁○♂️♂️ &♂️♂️♂️♂️	玩家第一个玩的游戏分类👉👉	👉♂️♂️♂️♂️♂️ ♂️♂️♂️♂️♂️	首次转账点数
♂️♂️♂️♂️ ♂️♂️♂️♂️ ♂️♂️♂️	玩家验证手机号时间	○📁♂️♂️ ♂️♂️♂️♂️ ♂️♂️♂️♂️♂️ ♂️♂️	玩家一天内最多储值次数	♂️♂️♂️♂️♂️ ♂️♂️♂️♂️ ○♂️♂️♂️♂️ ♂️	玩家第一次转账前游戏个数	👉♂️♂️♂️♂️♂️ ♂️♂️♂️♂️♂️ ♂️♂️♂️	首次转账时间
📁♂️♂️♂️ ♂️♂️♂️♂️ ♂️	玩家活跃值	♂️📁♂️♂️ ♂️♂️♂️♂️♂️ ♂️♂️	绑定手机后📁♂️天内储值次数	♂️♂️♂️♂️♂️ ♂️♂️♂️♂️ ○♂️♂️♂️♂️ ♂️♂️	玩家第一次转账前游戏分类个数	👉♂️♂️♂️♂️♂️ ♂️♂️♂️♂️♂️ ♂️♂️	首次转账时等级
♂️♂️♂️ ♂️♂️♂️♂️ ♂️	到达转账条件所需时间	♂️📁♂️♂️ ♂️♂️♂️♂️♂️ ♂️♂️	绑定手机后📁♂️天内玩家累计储值金额	♂️♂️♂️♂️♂️ ♂️♂️♂️♂️ ○♂️♂️♂️♂️ ♂️♂️♂️♂️♂️ ♂️♂️		♂️♂️♂️♂️♂️♂️ ♂️♂️♂️♂️♂️♂️ ♂️♂️♂️	首次转账账号是否为币商
👉♂️♂️♂️	是否次日刷玩	♂️📁♂️♂️ ♂️♂️♂️♂️♂️ ♂️♂️♂️				👉♂️♂️♂️♂️♂️♂️ ♂️♂️♂️♂️♂️♂️♂️ ♂️♂️♂️	首次转账后有无游戏行为



03. 工作流程

03.说明




拟合的不是线性函数，拟合的是一个概率学中的函数，即对是否为盗刷玩家进行分类，将输出标签赋予对应UID，输出预测结果

03.工程处理流程

特征筛选

数据处理

建模



04.模型结果

04.模型结果展示

逻辑回归

准确率: 0.96
AUC: 0.499

最优得分: 0.9644874899112187
测试得分: 0.9605263157894737
全部及最优系数: LogisticRegression(C=0.3)
定义搜索的最优系数: {'C': 0.3, 'solver': 'lbfgs'}

朴素贝叶斯

准确率: 0.647
AUC: 0.587

参数:
priors=None

决策树

准确率: 0.378
参数C: 0.478

Ⓜ □ ✦ ◆ Ⓜ □ ✦ □ ■ ❏
⑥ ♯ ✦ ■ ✦ ⑦
◆ □ ✦ ◆ ◆ Ⓜ □ ❏ ⑥ ♯ Ⓜ
◆ ◆ ⑦
○ ⊗ ☒ Ⓜ ♯ Ⓜ □ ◆ Ⓜ ❏
Ⓜ □ ■ Ⓜ □
○ ✦ ■ Ⓜ ♯ ⊗ ○ □ ● Ⓜ ◆ ✦
◆ □ ● ✦ ◆ ❏ ❏
○ ✦ ■ Ⓜ ♯ ⊗ ○ □ ● Ⓜ ◆ ✦

随机森林

准确率: 0.960
参数C: 0.407

■ Ⓜ ♯ ◆ ✦ ○ ⊗ ◆ □ □
◆ ❏ □ □ □
Ⓜ □ ✦ ◆ Ⓜ □ ✦ □ ■ ❏
⑥ ♯ ✦ ■ ✦ ⑦
○ ⊗ ☒ Ⓜ ♯ Ⓜ □ ◆ Ⓜ ❏
❏ ❏ □ ■ Ⓜ
○ ✦ ■ Ⓜ ♯ ⊗ ○ □ ● Ⓜ ◆
Ⓜ ◆ □ ● ✦ ◆ ❏ ❏
○ ✦ ■ Ⓜ ♯ ⊗ ○ □ ● Ⓜ ◆

XGBoost

准确率: 0.960
AUC: 0.407
参数:

□ ♯ er Ⓜ Ⓜ ✦ ✦ ❏ Ⓜ ❏
⑥ ♯ ✦ ■ ⊗ □ ☒ ❏ ● □ ♯
✦ ◆ ✦ Ⓜ ⑦
■ Ⓜ ♯ ◆ ✦ ○ ⊗ ◆ □ □ ◆
❏ □ □ □
◆ ◆ ♯ ◆ ⊗ ○ □ ● Ⓜ ❏ ❏ ❏
♯
○ ⊗ ☒ Ⓜ ♯ Ⓜ □ ◆ Ⓜ ❏
❏ □ □

04.逻辑回归结果可视化

```
precision    recall  f1-score   support

非盗刷      0.96      1.00      0.98      511
盗刷        0.00      0.00      0.00       21

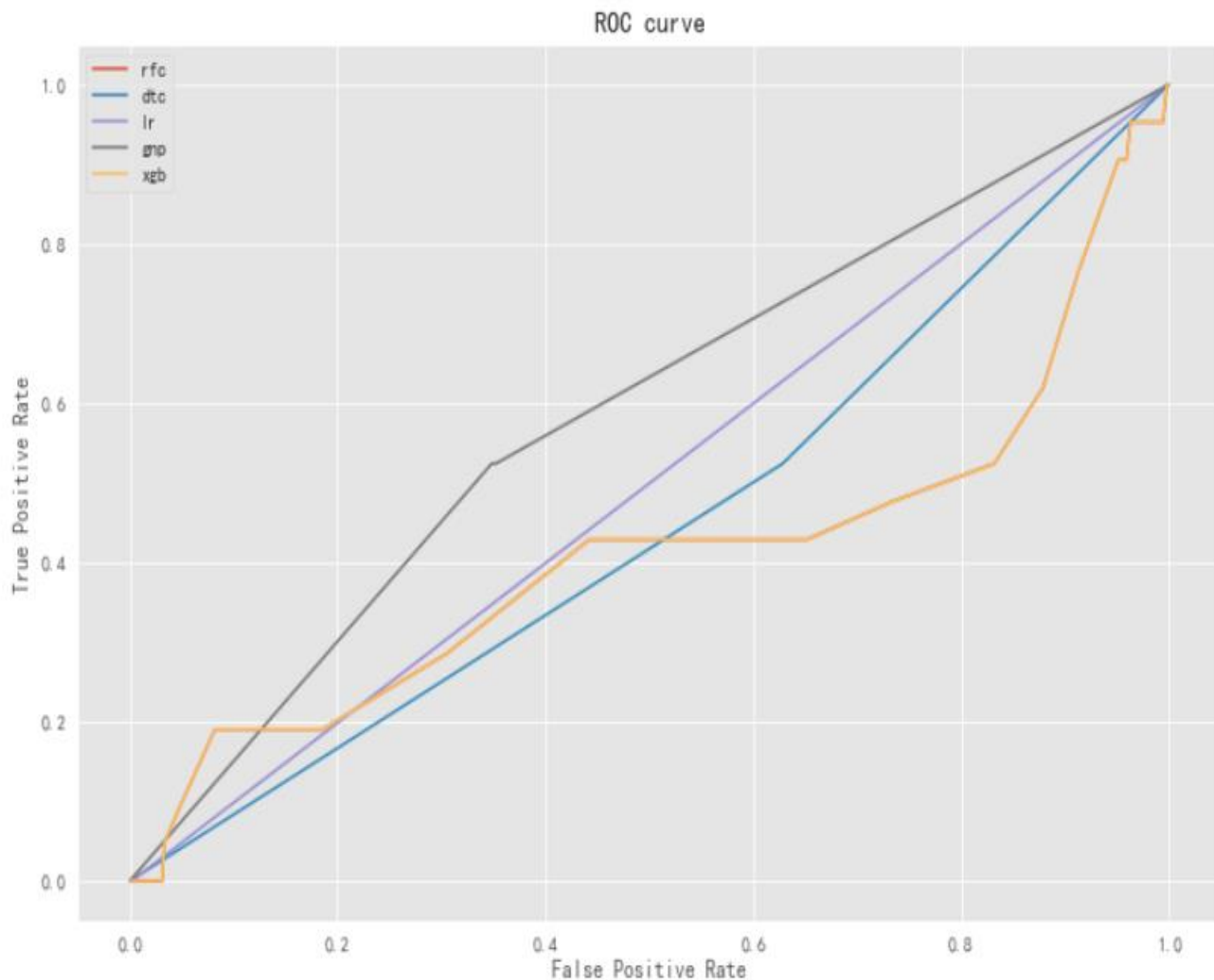
accuracy          0.96      532
macro avg      0.48      0.50      0.49      532
weighted avg   0.92      0.96      0.94      532

[[511  0]
 [ 21  0]]
```

```
auc  accuracy  precision  recall  f1
score 0.451278  0.949      1.0      0.05  0.095
```

说明：模型因数据填充存在过拟合状态，不具备落地价值，数据仅供分享交流使用

04. ROC曲线模型比较



使用归一化数据的模型:

- 逻辑回归 (lr)
- 朴素贝叶斯 (gnb)

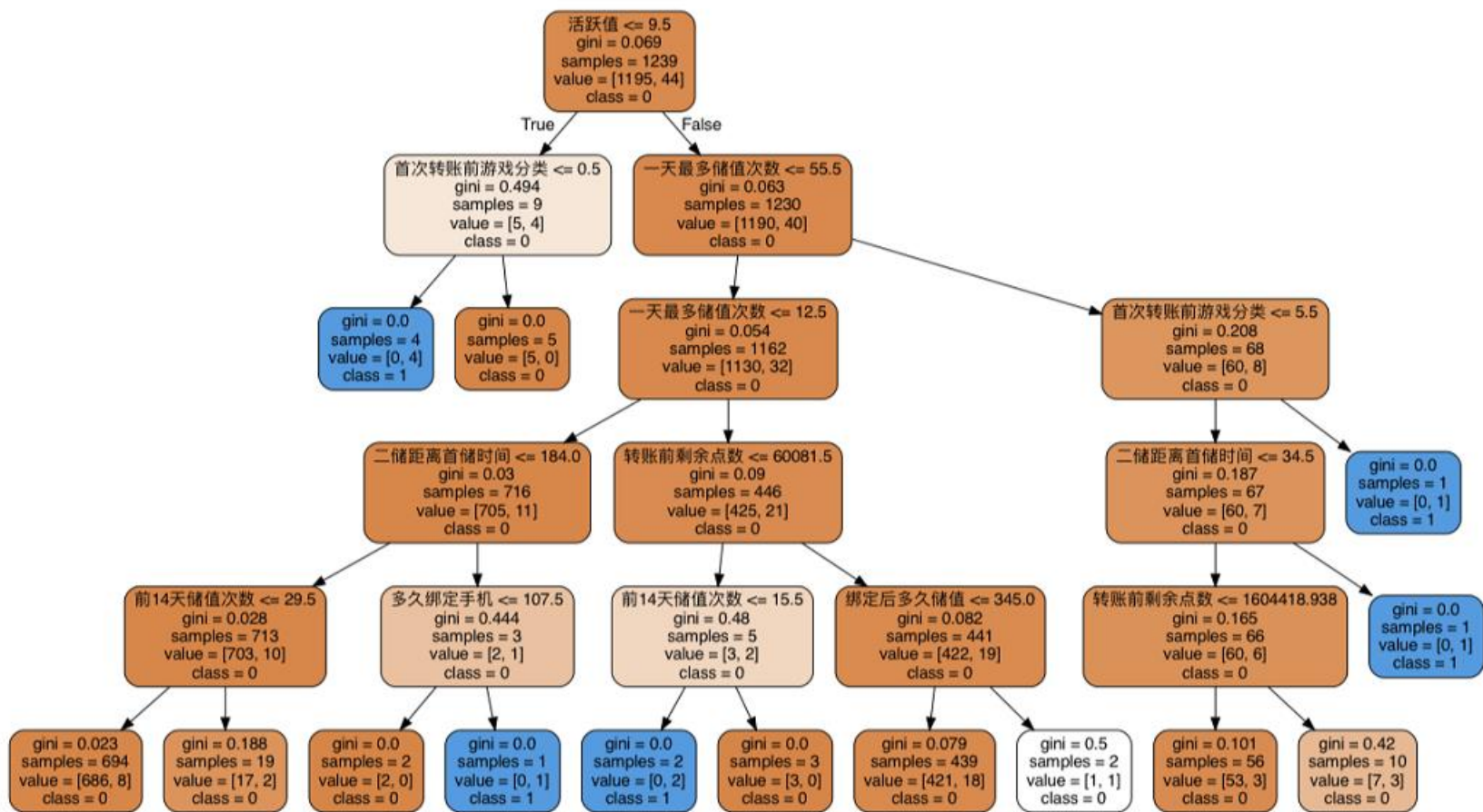
使用未归一化数据的模型:

- 决策树 (dtc)
- 随机森林 (rfc)
- XGBoost (xgb)

通过比较发现

Lr, 随机森林和XGBoost的效果更好

04.树模型指导业务





06.规划及框架搭建

规划及框架搭建

后续方向:

盗刷玩家用户画像

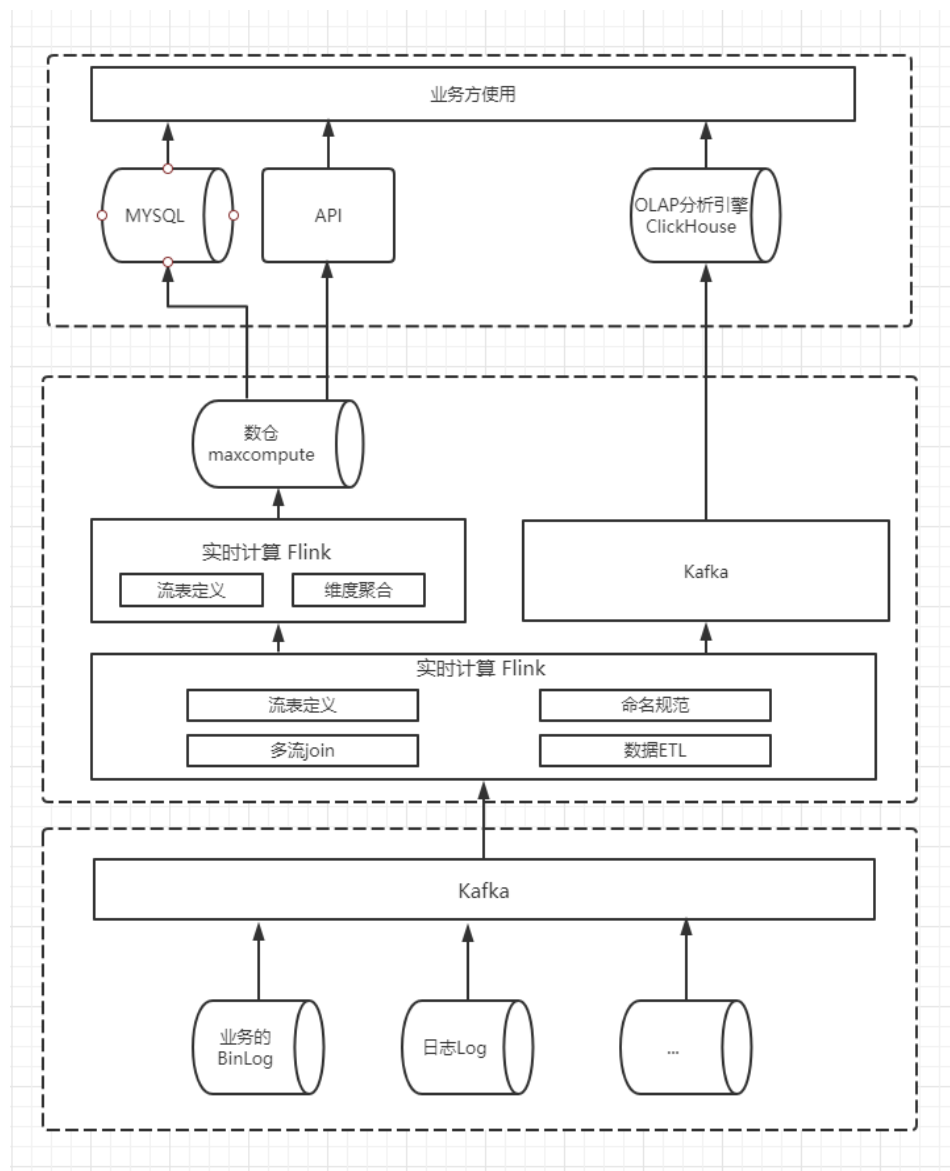
--待分析，对特征字段再进行深入探索研究

数据质量准备

--已接入数仓，对数据进行存储准备后续建模

用户信用评分体系

--衡量玩家未来一段时间内盗刷概率的预测





Thank You