变量解释：

|  |  |
| --- | --- |
| 变量名 | 变量解释 |
| recency | 上次购买距今月数 |
| history | 平均开销 |
| used\_discount | 是否使用过折扣 |
| used\_bogo | 是否使用过买一送一 |
| zip\_code | 所属地区  Suburban/Urban/Rural |
| is\_referral | 是否被引荐 |
| channel | 购买渠道  Phone/Web/Multichannel |
| offer | 给予了哪些优惠  Discount/But One Get One/No Offer |
| conversion | 反馈（是否购买） |

特别说明：

1.本作业的不同之处在于，对于目标客户的分类采用二分类的方法

在原项目的基础上，将TR+CN视为目标客户，TN+CR视为非目标客户

2.在模型的选择上，原项目采用了XGBoost，本作业提供了随机森林和逻辑回归两种策略，并在比较两种方法的效果最终选择了随机森林

3.在模型评估方面，由于本作业将原项目简化成了二分类问题，所以选择了二维混淆矩阵及其相关指标F1分数、准确率、精确率、召回率和ROC曲线及其AUC分数作为评价指标

模型说明：

1. 模型概述

本文档描述了基于随机森林算法构建的Uplift模型。该模型用于预测顾客对BOGO和Discount策略的响应，并帮助优化营销策略，以增加销售和提高客户忠诚度。

2. 数据集

使用的数据集包含顾客的历史交易记录，其中包括购买记录、优惠券使用情况等信息。数据集经过预处理和特征工程处理，用于训练和评估Uplift模型。

3. 特征工程

将定类变量进行独热编码。

将目标变量转换为二分类变量，其中0表示CN和TR，1表示CR和TN。

4. 模型构建

使用随机森林算法构建Uplift模型，将其应用于BOGO和Discount两种营销策略。模型基于历史交易数据，通过学习顾客的行为模式来预测其对不同营销策略的响应。

5. 模型评估

使用准确率、混淆矩阵、精确率、召回率、F1分数以及ROC曲线和AUC值等指标对模型进行评估。评估结果显示，模型在预测顾客对BOGO和Discount策略的响应方面表现良好。

结果展示：

BOGO 策略的 Uplift 模型评估结果:

准确率: 0.49789194253591507

混淆矩阵:

[[3090 3612]

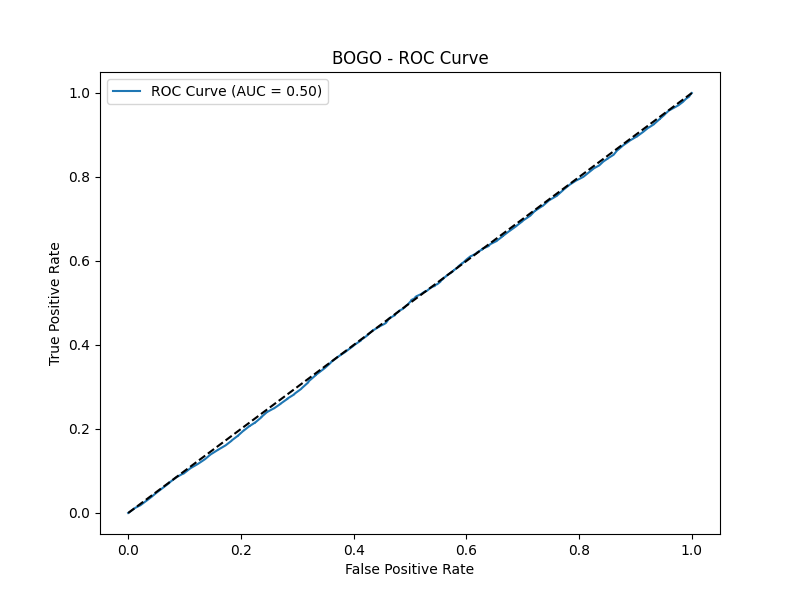
[2819 3287]]

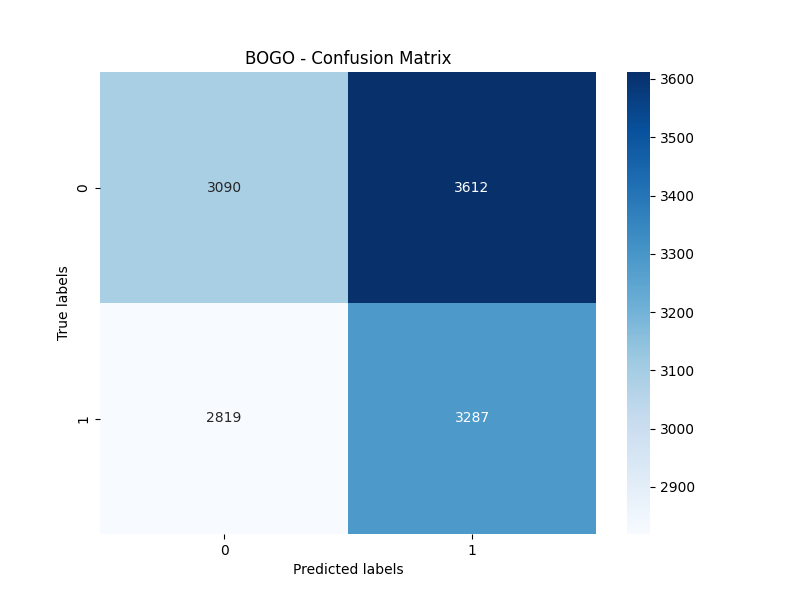
精确率: 0.4764458617190897

召回率: 0.5383229610219457

F1 分数: 0.5054978854286812

AUC 值: 0.4959443739533242





discount策略的 Uplift 模型评估结果:

准确率: 0.49679286608260326

混淆矩阵:

[[3116 3746]

[2687 3235]]

精确率: 0.4634006589313852

召回率: 0.5462681526511314

F1 分数: 0.5014337750910641

AUC 值: 0.5010846705215013

