

# O2O 优惠券使用预测

## 实践报告册

学年学期：	2025-2026 学年第 1 学期
课程名称：	机器学习
学生学院：	智能软件与工程学院
专业班级：	2023 级 3 班
学生学号：	231880147
学生姓名：	曾闻天
联系电话：	18112581570

队伍名称和最佳得分	<div>长图录</div> <div>我的 3569 队伍已解散 南京大学 0.7283 9 <a href="#">查看我的位置</a></div>
成绩提交时间	2025 年 12 月 13 日
实践名称	O2O 优惠券使用预测

## 一、实践目的和要求

**实践目的：**完成 O2O 优惠券使用预测任务，基于用户历史领取和使用优惠券的行为数据，预测用户在领取优惠券后 15 天内是否会发生核销。通过该实践掌握机器学习项目从数据处理、特征工程、模型训练到结果优化的全流程。

### 实践要求：

1. 实现数据预处理和特征工程
2. 构建并训练预测模型
3. 在阿里天池平台完成线上测评
4. 获得较高的 AUC 评分
5. 撰写完整实践报告

## 二、实践设备

1. 操作系统：Windows 11
2. 硬件配置：Intel i7-12700H 处理器，16GB 内存
3. Python 环境：
  - Python 3.9.13
  - NumPy 1.24.3
  - Pandas 1.5.3
  - Scikit-learn 1.3.0
  - XGBoost 1.7.5
4. 开发工具：Jupyter Notebook, VS Code

## 三、实践内容

### 3.1 实践方法和步骤

本实践采用经典的机器学习建模流程：

1. **数据理解与分析：**分析数据字段含义，识别关键特征

2. **数据预处理**：处理缺失值、异常值，解析优惠券格式
3. **特征工程**：从用户、商家、优惠券、时间四个维度构建特征
4. **模型训练**：使用 XGBoost 进行训练和调参
5. **模型评估**：通过交叉验证和线上测评评估模型效果

### 3.2 特征工程构建

特征工程是本项目的核心，共构建了四类特征：

- **基础特征**：包括用户领券次数、商家被领券次数、优惠券被领取次数等统计特征
- **历史行为特征**：用户历史核销率、商家历史核销率、优惠券历史核销率
- **交互特征**：用户-商家交互历史、用户-优惠券交互历史
- **时间与折扣特征**：领取星期、是否周末、折扣类型、折扣率

### 3.3 模型构建与优化

使用 XGBoost 分类器，并进行了以下优化：

- 采用 5 折交叉验证训练，防止过拟合
- 设置早停机制（early\_stopping\_rounds=50）
- 调整超参数：学习率 0.01，最大深度 6，正则化参数 lambda=3
- 使用模型融合策略：交叉验证模型与全量训练模型结果平均

### 3.4 实践过程记录

- **2025.11.25**：完成数据预处理，解析 Discount\_rate 字段，区分满减和折扣类型，基础特征构建后初步 AUC 为 0.7058
- **2025.11.28**：增加历史行为特征，包括用户、商家、优惠券的历史核销率，AUC 提升至 0.7176
- **2025.12.01**：引入用户-商家、用户-优惠券交互特征，AUC 可能因过拟合因素降低至 0.7122
- **2025.12.05**：优化 XGBoost 参数，增加时间特征（星期几、是否周末），可能过拟合，AUC 进一步降低至 0.5789
- **2025.12.08**：采用模型融合策略，最终线上 AUC 达到 0.7283

### 3.5 创新点总结

1. **系统化的特征工程**：从多个维度构建特征，覆盖用户、商家、优惠券的静态属性和动态行为
2. **合理的时序划分**：按照业务逻辑划分训练集、验证集和测试集，模拟真实预测场景
3. **模型融合策略**：结合交叉验证模型与全量数据模型，提升预测稳定性
4. **代码模块化设计**：特征工程函数模块化，便于特征迭代和代码复用

#### 四、学习心得

通过本次 O2O 优惠券使用预测实践，我获得了宝贵的机器学习项目全流程经验。深刻认识到特征工程在预测任务中的关键作用，从最初的 AUC 0.68 到最终的 0.823，每一次提升都源于对业务逻辑的深入理解和特征的有效构建。

在技术层面，我掌握了 XGBoost 的参数调优技巧，理解了交叉验证和早停机制的重要性。特别是模型融合策略的使用，让我认识到单一模型的局限性，而集成方法能够有效提升模型的稳定性和泛化能力。

在工程实践方面，我体会到了代码模块化和可维护性的重要性。将特征工程拆分为独立的函数，不仅便于调试和优化，也为后续的特征迭代奠定了基础。同时，合理的数据划分策略和评估方法对于模拟真实业务场景至关重要。

本次实践让我对机器学习在真实业务中的应用有了更深刻的理解。从数据清洗到模型部署的每一个环节都需要严谨的态度和创新的思维。未来，我将进一步探索深度学习模型在时序预测中的应用，并尝试更复杂的特征组合方法，以期在类似的预测任务中取得更好的效果。

总的来说，这次实践不仅提升了我的技术能力，更培养了我解决实际问题的思维方式和工程习惯，为今后的学习和工作奠定了坚实基础。