**长春理工大学**

**毕业论文开题报告**

**经济管理 学院 信息管理与信息系统 专业**

**学 号 210925118**

**学生姓名 吴佳玮**

**指导教师 国 琳**

填 写 说 明

一、学生应认真阅读《题目论证书》，明确了解题目的具体要求。

二、开题报告由学生按要求填写完成后上传至教务管理系统，指导教师在教务管理系统审阅后填写意见。

三、指导教师依照开题报告中的进度安排进行检查，并在教务管理系统的工作记录上填写阶段指导性意见及解决性建议。工作记录从进入毕业设计（论文）阶段开始填写。

四、学号为9位阿拉伯数字，如150411316。

五、恪守学术道德规范，严禁毕业设计作假，一经查实按相关规定处理。

|  |
| --- |
| **开 题 报 告**   1. **本课题研究的目的、意义**   随着互联网和电子商务的迅猛发展，电商平台积累了大量的用户行为数据。这些数据不仅包括用户的购买记录，还包括浏览、评论、分享等互动行为。通过分析这些数据，可以深入理解用户的消费习惯和偏好，从而优化平台的服务和营销策略。特别是对于商品分享行为的研究，能够帮助平台识别哪些用户更有可能成为回流用户，并采取针对性措施提高用户留存率和复购率。   1. **拟采取的研究路线**   为了预测商品分享行为的回流用户，本研究将结合多种方法和技术，主要路线包括：  **数据预处理：**首先对电商平台的用户行为数据进行清洗和预处理，去除噪声和异常值，确保数据的质量。刘柏鳞（2016）在《基于客户网络购物行为分析及预测系统研究》中强调了数据预处理的重要性，认为高质量的数据是后续分析的基础，可以通过对用户的浏览记录、购买历史、评价反馈等多维度数据进行分析，可以构建出详细的用户画像，进而为个性化推荐和精准营销等提供支持。  1.数据清洗  去除噪声和异常值：识别并删除无效或异常的数据点，如重复记录、缺失值较多的记录等。  填补缺失值：对于部分缺失的数据，采用K近邻（KNN）、均值填充等方法进行填补。  数据标准化：对数值型特征进行归一化处理，确保不同特征之间的量纲一致。  2.数据集成  多源数据融合：将来自不同渠道的数据进行整合，形成统一的用户行为数据集。  时间序列处理：将用户的行为按时间顺序排列，生成用户的行为序列，便于后续的时间依赖性分析。  **特征工程：**从用户的历史行为数据中提取有用的特征，如用户的浏览记录、购买历史、分享次数、评论内容等。通过特征工程，可以构建出更具代表性的特征集，为模型训练提供有力支持。  1.用户行为特征：  RFM模型特征：基于Recency（最近一次购买时间）、Frequency（购买频率）、Monetary（购买金额）构建用户活跃度和忠诚度的特征。  分享行为特征：统计用户的分享次数、分享的商品类别、分享的时间间隔等，评估用户的社交影响力。  浏览行为特征：分析用户的浏览深度、停留时间、浏览路径等，了解用户的兴趣偏好。  评论行为特征：提取用户的评论内容，使用自然语言处理（NLP）技术分析评论的情感倾向（正面/负面），并计算用户的评论活跃度。  2.商品特征：  商品属性：如价格、品牌、类别、销量、评分等，评估商品的吸引力。  商品复购周期：根据历史销售数据，计算不同类型商品的平均复购周期，作为预测用户复购行为的重要依据。  商品关联性：通过关联规则挖掘（如FP-Growth算法），发现用户经常一起购买的商品组合，构建商品之间的关联特征。  3.用户-商品交互特征：  用户对商品的兴趣度：结合用户的浏览、收藏、加购物车等行为，计算用户对特定商品的兴趣度。  用户对品牌的忠诚度：统计用户对同一品牌的购买频率和复购率，评估用户的品牌忠诚度。  4.时间特征：  季节性特征：考虑节假日、促销活动等对用户行为的影响，构建时间窗口内的用户行为特征。  动态变化特征：引入时间衰减因子，反映用户行为随时间的变化趋势，捕捉用户的兴趣转移。  **模型构建：**采用机器学习和神经网络算法构建预测模型。具体来说，可以使用XGBoost、LightGBM、随机森林等分类算法，结合深度学习中的卷积神经网络（CNN）和循环神经网络（RNN），以提高模型的泛化能力和预测精度。张诗晨（2019）在《基于机器学习的电商在线消费者购买行为预测研究》中提到，基于机器学习的分类模型能够有效预测用户的复购行为，融合模型（如RNN-NB）相较于单一模型具有更高的预测准确率，尤其是在处理短序列数据时表现更为出色。为了进一步提升预测效果，可以采用模型融合技术，如Soft-Voting、Stacking等。通过融合多个模型的预测结果，可以减少单一模型的过拟合问题，提高模型的稳定性和准确性。吕泽宇, 李纪旋, 陈如剑, 等.（2019）在《电商平台用户再购物行为的预测研究》中表明，模型融合能够显著提升预测性能，特别是在处理大规模数据时表现更为突出。  1.模型选择：  传统机器学习模型：如逻辑回归、随机森林、XGBoost、LightGBM等，适用于处理结构化数据，能够快速建立基准模型。  深度学习模型：如卷积神经网络（CNN）、循环神经网络（RNN）、长短期记忆网络（LSTM）等，适用于处理复杂的时序数据和文本数据，能够捕捉用户行为的长期依赖关系。  混合模型：结合传统机器学习和深度学习的优势，构建混合模型（如XGBoost+LSTM），以提高预测的准确性和鲁棒性。  2.模型训练：  数据划分：将数据集划分为训练集、验证集和测试集，确保模型的泛化能力。  超参数调优：使用网格搜索（Grid Search）、随机搜索（Random Search）或贝叶斯优化（Bayesian Optimization）等方法，寻找最优的模型超参数。  正负样本平衡：由于回流用户的比例通常较低，存在样本不平衡问题。可以采用SMOTE（Synthetic Minority Over-sampling Technique）等方法进行过采样，或者使用代价敏感学习（Cost-sensitive Learning）来调整分类权重。  3.模型融合：  Soft-Voting：通过对多个模型的预测结果进行加权平均，减少单一模型的过拟合风险，提升整体预测性能。  Stacking：将多个基础模型的输出作为新的特征，输入到元模型（如逻辑回归或XGBoost）中进行最终预测，进一步提高模型的准确性。  **实验验证：**通过实验验证模型的有效性。可以使用A/B测试或交叉验证的方法，评估模型在不同场景下的表现。同时，还可以通过ROC曲线、F1值等指标衡量模型的预测效果。  1.评估指标：  分类指标：如准确率（Accuracy）、精确率（Precision）、召回率（Recall）、F1值等，用于评估模型的分类效果。  AUC-ROC曲线：用于评估模型的区分能力，特别是在二分类任务中，AUC值越高，模型的预测性能越好。  混淆矩阵：详细展示模型的分类结果，帮助分析模型的误分类情况。  KS统计量：用于评估模型的排序能力，KS值越大，模型的区分能力越强。  4.2 实验设计：  A/B测试：在实际电商平台上进行A/B测试，比较不同模型的预测效果，确保模型在真实环境中的有效性。  交叉验证：使用k折交叉验证（如5折或10折）评估模型的稳定性和泛化能力，避免过拟合。  离线实验：在离线环境中使用历史数据进行实验，验证模型的预测性能，并根据实验结果进行调整。  4.3 模型优化：  特征选择：通过特征重要性分析（如SHAP值、Permutation Importance），筛选出对模型预测最有贡献的特征，减少冗余特征，提升模型效率。  模型简化：在保证预测性能的前提下，尽量简化模型结构，降低计算复杂度，提升模型的可解释性和实时性。  **论文提纲：**  **电商平台商品分享行为的回流用户预测研究**  摘要  一、引言   1. 研究背景与意义 2. 研究方法与技术路线   二、相关工作综述   1. 用户复购行为的研究现状 2. 商品分享行为的研究现状 3. 现有研究的不足与改进方向   三、数据收集与预处理   1. 数据来源 2. 数据清洗 3. 数据集成 4. 数据可视化   四、特征工程   1. 用户行为特征 2. 商品特征 3. 用户-商品交互特征 4. 时间特征   五、模型构建与训练   1. 模型选择 2. 模型训练 3. 模型融合   六、模型评估与优化   1. 评估指标 2. 实验设计 3. 模型优化   七、结论与展望   1. 研究总结 2. 研究局限 3. 未来研究方向   致谢  参考文献  附录   1. **进度安排**   毕业实习：开学第1-8周，共8周  第9周：撰写毕业论文初稿，撰写研究背景与目的意义，初步构建训练模型；  第10周：训练模型，参考国内外研究完成论文研究；  第11周：修改模型，优化模型数据；  第12-14 周：毕业论文修改，定稿；  第15 周：毕业论文集体评阅；  第16 周：毕业论文答辩；  第17 周：学院自查并提交后期自查材料、全面总结材料。   1. **文献综述**   随着互联网的普及和电子商务的发展，消费者的购物模式发生了显著变化。电商平台不仅提供了便捷的购物体验，还积累了大量的用户行为数据。这些数据为研究消费者的复购行为提供了丰富的资源。  研究表明，产品的质量、功能、设计等因素对消费者的复购意愿有重要影响。高质量的产品能够提高消费者的满意度，进而增加复购的可能性[14]。此外，产品的独特性和创新性也能吸引消费者进行重复购买[12]。  服务质量和品牌形象是影响消费者复购行为的重要因素。良好的客户服务可以增强消费者的信任感，提升其对品牌的忠诚度[14]。品牌知名度和美誉度的提升也有助于消费者在选择商品时优先考虑该品牌[9]。特别是对于社区电商，用户群体相对固定，品牌认同感的建立尤为重要[9]。  价格是消费者决策过程中的关键因素之一。合理的价格定位和有效的促销活动可以刺激消费者的购买欲望，促进复购行为[17]。例如，限时折扣、满减优惠等促销手段能够激发消费者的紧迫感，促使他们更快地做出购买决定[17]。  消费者的满意度、情感体验和品牌认同感等因素对复购行为有着直接的影响[11]。满意的购物经历会增强消费者的品牌忠诚度，而负面的情感体验则可能导致消费者流失[13]。此外，消费者的情感依恋和社会认同感也会影响其复购意愿[9]。  在线评论和口碑传播是影响消费者购买决策的重要信息来源。正面的评价和推荐可以增加消费者的信任感，促进复购行为；相反，负面的评价可能会降低消费者的购买意愿[14]。因此，企业应重视用户反馈，及时处理负面评价，以维护良好的品牌形象[18]。  内容营销通过提供有价值的信息和娱乐内容，增强消费者的品牌认同感，从而促进复购行为[9]。研究表明，内容营销不仅能提高品牌的知名度，还能通过情感共鸣的方式，使消费者更愿意再次购买该品牌的产品或服务[9]。  对于复购行为有多种预测方法，传统的统计方法如回归分析、结构方程模型（SEM）等常用于分析消费者复购行为的影响因素。这些方法通过构建数学模型，量化各因素对复购行为的影响程度[12]。然而，传统方法在处理大规模、复杂的数据时存在局限性，难以捕捉到消费者行为的动态变化。  近年来，随着大数据和机器学习技术的发展，越来越多的研究开始采用数据挖掘和机器学习方法来预测消费者的复购行为。常见的算法包括决策树、支持向量机（SVM）、随机森林、梯度提升树（GBDT）、神经网络等[4]。这些算法能够自动提取特征，识别出影响复购行为的关键因素，并进行精准预测[2]。神经网络的消费者复购行为预测模型[7]，实现了用户属性、商品属性及用户行为特征的自动抽取与选择，并以此对消费者复购行为进行预测。实验结果表明，基于神经网络的预测模型F1值比基准模型平均提升了7%~11%。  分类模型如逻辑回归、朴素贝叶斯、支持向量机等被广泛应用于复购行为的预测。这些模型通过将消费者分为不同的类别，如“可能复购”和“不太可能复购”，从而实现对复购行为的预测[3]。研究表明，融合模型（如RNN-NB）相较于单一模型具有更高的预测准确率，尤其是在处理短序列数据时表现更为出色[4]。  时间序列分析也是一种常用的预测方法，尤其适用于处理具有时间依赖性的消费者行为数据。循环神经网络（RNN）、长短期记忆网络（LSTM）等深度学习模型能够有效地捕捉消费者行为的时间动态特征，从而提高预测的准确性[5]。此外，引入时间奖惩因子和商品复购周期的改进推荐算法也显示出较好的效果[10]。  为了减少单一模型可能出现的过拟合现象，可以使用多模型融合的复杂模型进行预测。实验研究表明，多lightGBM模型的融合方法能够达到很高的再购物行为预测准确度[1]，其AUC值能够达到0.7018，同时，基于这种方法实现的预测器只需要少数特征就能对预测结果产生很好的贡献。  用户画像和行为特征的构建是复购行为预测的基础。通过对用户的浏览记录、购买历史、评价反馈等多维度数据进行分析，可以构建出详细的用户画像，进而为个性化推荐和精准营销提供支持[6]。RFM模型（Recency, Frequency, Monetary）作为一种经典的用户细分工具，能够有效识别高价值用户，提升复购率[8]。  结构方程模型（SEM）是一种综合性的统计方法，能够同时处理多个变量之间的关系，广泛应用于消费者行为研究。Hellier等人[12]通过SEM模型研究了消费者复购意图的影响因素，发现服务质量和顾客满意度对复购意图有显著影响。此外，Mittal等人[16]进一步探讨了客户特征对复购行为的调节作用，指出不同类型的客户在复购行为上表现出不同的倾向。同时，通过RFM模型和数据建模技术构建的复购行为的预测模型在检测消费者的行为模式上亦有出色表现。研究表明，八次交易足以以高精度分类消费者，RFM模型在客户分类和细分方面具有较高的准确性[15]。  同时，社交媒体和内容营销在消费者复购行为中扮演着重要角色。通过社交媒体平台，企业可以与消费者进行互动，了解他们的需求和偏好，进而制定更加个性化的营销策略[9]。内容营销则通过提供有价值的信息和娱乐内容，增强消费者的品牌认同感，促进复购行为[9]。  Wijarnoko等人[18]通过对大量文献的系统性回顾，总结了影响消费者复购意图的主要因素，包括产品属性、服务质量、品牌形象、价格、促销策略、消费者满意度等。他们指出，未来的研究应更加关注这些因素之间的相互作用，以及如何通过数据挖掘和机器学习技术来更好地预测复购行为。  Chatzoglou等人[17]通过对零售商店消费者的实证研究，验证了服务质量、顾客体验、感知价值等因素对复购意图的影响。研究表明，良好的店内购物体验和顾客满意度是影响复购行为的关键因素。此外，消费者的情感反应（如满意感和后悔感）也在复购意图中起到了重要作用[13]。  综上所述，影响消费者复购行为的因素是多方面的，既包括产品本身的特性，也涉及服务质量、品牌形象、价格策略、消费者满意度等外部因素。随着大数据和机器学习技术的不断发展，复购行为的预测方法也在不断创新和完善。未来的研究可以进一步探索如何结合多种预测方法，提升预测的准确性和鲁棒性，同时深入挖掘社交媒体和内容营销对复购行为的影响机制，为企业制定更加科学的营销策略提供理论支持。  **参考文献：**  [1]吕泽宇,李纪旋,陈如剑,等.电商平台用户再购物行为的预测研究[J].计算机科学,2020,47(S1):424-428.  [2]杨澜.机器学习在电商用户购买行为预测中的应用研究[D].天津商业大学,2021.DOI:10.27362/d.cnki.gtsxy.2021.000342.  [3]杜世民.基于分类模型的电商用户复购行为预测研究[D].杭州师范大学,2019.  [4]张诗晨.基于机器学习的电商在线消费者购买行为预测研究[D].吉林大学,2019.  [5]张天鑫.基于机器学习的网购用户复购预测[D].中南财经政法大学,2023.DOI:10.27660/d.cnki.gzczu.2023.002985.  [6]刘柏鳞.基于客户网络购物行为分析及预测系统研究[D].北京工业大学,2016.  [7]张靖轩,张微微.基于神经网络的电商消费者复购行为预测[J].应用数学进展, 2021, 10(10):7.DOI:10.12677/AAM.2021.1010354.  [8]薛红松,苏国伟,张李义.基于数据挖掘的客户复购行为分析[J].武汉大学学报(工学版),2019,52(09):841-846.DOI:10.14188/j.1671-8844.2019-09-014.  [9]罗萧,蒋明华.内容营销、品牌认同与消费者品牌忠诚[J].商业经济研究,2019,(23):73-76.  [10]石力.社区电商用户复购行为预测及推荐算法研究[D].北京化工大学,2021.DOI:10.26939/d.cnki.gbhgu.2021.000008.  [11]汪鑫.线上消费者体验与复购行为研究[D].厦门大学,2019.DOI:10.27424/d.cnki.gxmdu.2019.001054.  [12] Hellier P K , Geursen G M , Carr R A ,et al.Customer Repurchase Intention: A General Structural Equation Model[J].European Journal of Marketing, 2003, 37(11/12):1762-1800.DOI:10.1108/03090560310495456.  [13] Liao C , Lin H N , Luo M M ,et al.Factors influencing Online Shoppers' Repurchase Intentions: The Roles of Satisfaction and Regret[J].Information & Management, 2016, 54(5):651-668.DOI:10.1016/j.im.2016.12.005.  [14] Zhang J , Zheng H , Liu J ,et al.Research on factors influencing the consumer repurchase intention: Data mining of consumers' online reviews based on machine learning[J].Neural Computing and Applications, 2024, 36(17):9837-9848.DOI:10.1007/s00521-024-09591-4.  [15] C M A R A , B M M , C S K ,et al.RFM-based repurchase behavior for customer classification and segmentation[J].Journal of Retailing and Consumer Services, 61[2024-11-30].DOI:10.1016/j.jretconser.2021.102566.  [16]Mittal,Vikas,Kamakura,et al.Satisfaction, Repurchase Intent, and Repurchase Behavior: Investigating the Moderating Effect of Customer Characteristics.[J].Journal of Marketing Research (JMR), 2001.  [17] Chatzoglou P, Chatzoudes D, Savvidou A, et. Factors affecting repurchase intentions in retail shopping: An empirical study[J]. Heliyon, 2022, 8(9): e10619.  [18] Wijarnoko, Muhammad, Pramana, Edwin, Santoso, Joan, et al. Factors That Influence Repurchase Intention: A Systematic Literature Review [J]. Teknika, 2023, 12:252-260. |