

毕业设计开题报告

|  |  |
| --- | --- |
| **姓 名：** | 李保媛 |
| **学 号：** | 12422706026 |
| **学 院：** | 信息工程学院 |
| **专业班级：** | 大数据（专升本）2403 |
| **指导教师：** |  |

2025年9月3日

|  |  |
| --- | --- |
| **论文题目：** | 基于RFM模型的视频平台用户价值分群系统设计 |
| **一、选题背景与意义**   1. 国内外研究现状  1、国外研究现状  * RFM 模型起源于直复营销领域，用于基于顾客最近一次消费时间（Recency）、消费频次（Frequency）、消费金额（Monetary）对客户进行价值细分，便于差异化触达与资源优化配置。随着数据规模与计算能力提升，研究从单纯 RFM 拓展到 LRFM（加入关系长度/生命周期）、RFM-I（加入互动）、将 RFM 作为特征输入到聚类、概率模型与（准）实时运营系统中。 * 在方法层，K-Means、层次聚类、GMM、K-Prototypes 等无监督方法被广泛用于客户细分；肘部法、轮廓系数等用于聚类数选择；BLUP 与贝叶斯概率模型（如 Pareto/NBD、BG/NBD）用于生命周期价值（CLV）估计，并与 RFM 互为补充。 * 在视频/流媒体场景，学界与业界强调“观看行为”与“订阅价值”的协同：以会籍续费、观看时长、完播率、设备多样性与交互行为（点赞、评论、分享）作为忠诚与流失风险的关键信号；同时将分群结果输入推荐与营销自动化链路，实现分群驱动的排期策略、消息与 A/B 实验。  2、国内研究现状  * 互联网平台广泛将 RFM 用于会员运营与复购促进。研究从电商领域扩展至内容平台与长/短视频平台，逐步引入“观看频次/天”“时长与完播率”“互动参与度”“广告观看贡献”等指标，形成“行为 RFM + 价值 RFM”的组合；并与沉睡用户唤醒、拉新转化、续费保活与广告变现策略结合。 * 在工程实现上，企业多采用离线（批处理）+近实时（流式）双轨，以统一用户数据平台（CDP）承载画像与分群规则，通过可视化看板与自动化触达闭环支撑运营迭代。学术界对“视频行为信号的标准化”“非付费用户 M 维度的代理量化”“冷启动与跨端识别”等问题的探讨逐年增多，但将分群与在线推荐、动态实验联动的系统化落地研究仍相对不足。  1. 选题目的及意义  * 目的：构建适配视频平台业务特性的 RFM 指标体系与分群方法，设计可落地的用户价值分群系统，实现以分群为枢纽的精细化运营与策略闭环，并量化其对留存、续费与 ARPU 的提升。 * 意义： * 理论层面：在传统 RFM 框架下引入视频场景的行为代理变量与标准化方法，探索 RFM 与 CLV/流失预测模型的协同。 * 实践层面：形成可复用的分群流程与系统架构指南，提升用户经营效率，支持 A/B 实验与自动化触达，降低运营试错成本。 | |
| 1. **主要研究内容**  * 指标与度量体系设计： * R 维度：最近观看时间、最近互动时间、最近付费/续费时间（按业务线分别计算）。 * F 维度：最近30/60/90天观看活跃天数、有效会话数、付费/续费次数。 * M 维度：订阅与单次付费金额、广告观看贡献的代理货币化（如可视曝光×eCPM 估测）、增值收入（打赏、道具）。 * 扩展维度：观看时长、完播率、系列追更率、互动率、设备与网络多样性、内容品类多样性。 * 数据预处理与标准化：口径统一、异常与作弊过滤、分布稳定性评估（PSI）、跨周期可比性校准、对数/分箱与 Min-Max/Robust 标准化。 * 分群方法：基于标准化 RFM 的无监督聚类（K-Means/K-Prototypes/GMM），聚类数选择（肘部法、轮廓系数）、稳健性评估（多次重启、一致性指标）。 * 画像与命名：结合业务解读形成如“高价值忠诚”“高潜拉升”“新近高活”“沉睡挽回”“仅广告价值”等分群，并产出关键行为画像。 * 策略与闭环：为各分群配置续费激励、内容召回、冷启动扶持、价格梯度、广告频控、消息触达与推荐权重策略，并以 A/B 实验与分层归因评估效果。 * 系统设计：数据采集与治理层、特征与分群计算层（离线+近实时）、特征/分群服务层（API/特征库/CDP）、可视化与运营工作台、实验与监控模块。 * 评估与实验：以留存率、续费率、ARPU/Revenue Lift、广告收益、用户体验指标为核心，采用分群内对照与跨群对照、时间分层与多指标置信评估。 | |
| 1. **研究方法与手段**  * 文献研究与比较分析：梳理 RFM、CLV、客户细分与流媒体运营的代表性研究与最佳实践。 * 数据分析方法：探索性数据分析（EDA）、稳健分布检验、异常检测与抗噪处理。 * 统计与机器学习： * 无监督聚类（K-Means、GMM、K-Prototypes）、聚类有效性评估（轮廓系数、Calinski–Harabasz、Davies–Bouldin）。 * 辅助模型：BG/NBD 等概率模型估计消费/观看到达过程；逻辑回归/梯度提升用于流失/续费倾向预测与策略联动。 * 特征工程：时间窗滚动特征、分箱与WOE、目标编码（谨慎防泄漏）、多源归并与身份解析。 * 系统实现与工程： * 计算与存储：Python/Pandas、Spark/Flink（批/流）、ClickHouse/BigQuery/MySQL、Redis、Kafka。 * 服务与可视化：FastAPI/Flask、Superset/Metabase/内部 BI、实验平台与消息队列。 * 可靠性与合规：指标血缘、权限与脱敏、GDPR/数据最小化与用途限制、审计与告警。 * 评估与实验设计：A/B/n、准实验（断点回归/时间序列对照）、提升度与置信区间计算、长期效应与反作弊控制。 | |
| 1. **论文撰写提纲**  * 摘要 * 第1章 绪论：研究背景、问题定义、研究意义、技术路线与创新点 * 第2章 理论基础与文献综述：RFM/CLV/客户细分与聚类理论、视频平台相关研究 * 第3章 研究设计：数据来源与口径、指标体系、方法框架与系统总体设计 * 第4章 数据预处理与指标标准化：数据治理、异常处理、分布稳定性与标准化 * 第5章 分群模型构建与结果分析：聚类实现、聚类数选择、稳健性与画像解读 * 第6章 系统设计与实现：数据管道、服务接口、可视化与运营工作台 * 第7章 实证与策略评估：实验设计、策略联动、效果评估与业务复盘 * 第8章 结论与展望：研究结论、局限性与未来工作（如与推荐系统的深度联动、实时分群） | |
| 五、参考文献   * Hughes, A. M. Strategic Database Marketing. McGraw-Hill, 1996/2000. * Blattberg, R. C., Kim, B.-D., Neslin, S. A. Database Marketing: Analyzing and Managing Customers. Springer, 2008. * Gupta, S., Lehmann, D. R., Stuart, J. A. Valuing Customers. Journal of Marketing Research, 2004. * Fader, P. S., Hardie, B. G. S. Probability Models for Customer-Base Analysis. Journal of Interactive Marketing, 2009. * Buckinx, W., Van den Poel, D. Customer base analysis: partial defection of behaviorally loyal clients. Expert Systems with Applications, 2005. * Ngai, E. W. T., Xiu, L., Chau, D. C. K. Application of data mining techniques in CRM. Expert Systems with Applications, 2009. * Kim, H.-S., Jung, S.-H., Suh, E.-H. A LRFM model for customer segmentation. (常见引用于客户细分研究，2006，相关会议/期刊) * Gómez-Uribe, C. A., Hunt, N. The Netflix Recommender System. ACM Transactions on Management Information Systems, 2016. * Kaufman, L., Rousseeuw, P. J. Finding Groups in Data: An Introduction to Cluster Analysis. Wiley, 1990. * Rousseeuw, P. J. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics, 1987. * Chin, T.-W., et al. Customer lifetime value prediction and its use for customer segmentation. (相关综述/实证论文，选取于营销科学文献) * Han, J., Kamber, M., Pei, J. Data Mining: Concepts and Techniques. Morgan Kaufmann, 2011/2012. * 周志华. 机器学习. 清华大学出版社, 2016. * 李航. 统计学习方法（第二版）. 清华大学出版社, 2019. * 陈超美, 刘鸿亮. 应用多元统计分析（第二版）. 高等教育出版社, 2014. | |
| 1. **指导教师评语**   是否同意开题： （是、否）  指导教师（签字）：  年 月 日 | |
| **七、学院审核意见**  负责人（签字） ：  年 月 日 | |