LRFS Online Shoppers Behaviour-Based Efficient Customer Segmentation Model

Pooja K N [1] Swathi K V [2]

[1] *Assistant Professor,Department of MCA,BIET,Davanagere*

[2] *Student, 4th Semester MCA,Department of MCA,BIET,Davanagere*

Abstract

In the realm of digital commerce, online shopping has witnessed unprecedented growth globally, becoming a cornerstone of modern consumer behavior. This research introduces an advanced customer segmentation model, named LRFS, which builds upon the traditional LRF framework (Length of Relationship, Recency of Purchase, and Frequency of Purchase), specifically tailored for the e-commerce sector. The innovation of the LRFS model lies in the integration of a novel component, ‘‘S’’, which quantifies the Staying Rate relative to the revenue generated by customers on a specific website. This addition aims to enhance the granularity and efficacy of customer segmentation by leveraging data extracted from Google Analytics. To operationalize the LRFS model, this study employs two renowned clustering algorithms, KMeans and K-Medoids, analyzing the dataset through the lens of three distinct dimensionality reduction techniques: PCA (Principal Component Analysis), t-SNE (t-Distributed Stochastic Neighbor Embedding), and Autoencoder. This methodological approach facilitates a robust comparative analysis between the LRFS model and its predecessors — LR, LF, and LRF — utilizing K-Means clustering to evaluate the precision of customer cluster assignments. The empirical findings of this research underscore the superiority of the LRFS model in achieving more accurate and insightful customer segmentation

Keywords: *, KMeans and K-Medoids, analyzing the dataset*

I. INTRODUCTION

The exponential rise of e-commerce has transformed the global retail landscape, ushering in a new era of consumer behavior that is increasingly digital, data-driven, and dynamic. With the proliferation of online shopping platforms, businesses are inundated with vast amounts of customer interaction data, offering an unprecedented opportunity to analyze and understand consumer preferences, purchase patterns, and behavioral trends. In this context, customer segmentation has emerged as a critical strategy for e-commerce businesses aiming to deliver personalized experiences, optimize marketing efforts, and enhance customer retention.Traditional customer segmentation models such as LR (Length of Relationship), LF (Length and Frequency), and LRF (Length, Recency, and Frequency) have laid the foundation for analyzing customer behavior. However, these

models often fall short in capturing the complex dynamics of user engagement and value generation in modern online platforms. To address these limitations, this study proposes an enhanced segmentation model named **LRFS**, which introduces a new dimension—**Staying Rate** (S)—to complement the existing parameters. The "S" factor is designed to measure the duration and intensity of customer engagement relative to the revenue they generate, providing a more comprehensive view of customer loyalty and contribution.The LRFS model is supported by advanced clustering techniques—KMeans and K-Medoids—which are employed to identify meaningful patterns within customer data. Furthermore, three dimensionality reduction techniques—PCA, t-SNE, and Autoencoder—are integrated into the analysis to improve clustering performance and visualization. By leveraging Google Analytics data, this research aims to demonstrate how the LRFS model surpasses traditional segmentation frameworks in accuracy, relevance, and business applicability.

II. LITERATURE REVIEW

The use of e-commerce throughout the world in recent years is very rapid. The continuous increase in sales shows that e-commerce has huge market potential. Store profits are derived from the process of assessing data to identify and classify online shopper intentions. The process of assessing the data uses conventional machine learning algorithms and deep neural networks. Comparison of algorithms in this study using the python programming language by knowing the value of Accuracy, F1-Score, Precision, Recall, and ROC AUC. The test results show that the accuracy of the deep neural network algorithm is 98.48%, the F1 score is 95.06%, precision is 97.36%, recall is 96.81% and AUC is 96.81%. So, based on this research, deep neural network data mining techniques can be an effective algorithm for online shopper intention data sets with cross-validation folds of 10, six hidden layer decoder-encoder variations, relu-sigmoid activation function, adagrad optimizer, and learning rate of 0.01 and no dropout. The value of this deep neural network algorithm is quite dominant compared to conventional machine learning algorithms and related research.

III. EXISTING SYSTEM

Author:K. Khalili-Damghani, F. Abdi, and S. Abolmakarem,The advancement of information technology has significantly broadened access to consumer behavior and purchase intent data, facilitating the development of increasingly sophisticated segmentation strategies[1]. Author:C. I. Agustyaningrum, M. Haris, R. Aryanti, and T. MisriatDespite this complexity, the value of simplicity in segmentation cannot be overstated . RFM analysis, focusing on recency, frequency, and monetary value, has long been a focal point for researchers, serving as a behavioral-based method to profile clients and evaluate the consistency of their purchasing behaviors over time This analysis not only aids in marketing decision-making but also in tailoring strategies to engage the right consumers effectively[2].Author:J. A. Al-Gasawneh, M. H. Al-Wadi, B. M. Al-Wadi, B. E. Alown, and N. M. Nuseirat Notably, some researchers have proposed a two-phased model technique as an innovative segmentation solution, leveraging the RFM model. [3], Author:S. Ahsain and M. A. Kbir, with comparative studies also exploring the juxtaposition

of AI methods and RFM analysis.Furthermore, empirical research employing a decision tree methodology for segmenting the electronic toll collection (ETC) customer base revealed its proficiency in analyzing travel behaviors, valuation, and potential appreciation for ETC services. This approach, integrating decision tree-based explainability with clustering via the ExKMC algorithm.[4],Author:K. Tabianan, S. Velu, and V. Ravi, enhances the interpretability of cluster assignments. The RFM model, alongside data modeling techniques, has beenapplied to identify customer behavior patterns, validated through various classification methods including multi-layer perceptron (MLP), support vector machine (SVM), and decision tree classification (DTC).LRFS [5]. Author:A. Balkaya, E. Tuzunkan, K. Ayaz, Y. Akçay, F. Abut, M. F. Akay, S. Erdem, and A. Alsaç, Beyond traditional RFM models, modifications incorporating new weights or dimensions have been explored to address inherent limitations and enhance segmentation outcomes. Such adaptations recognize that customers with higher RFM scores are generally more profitable and responsive, yet traditional RFM analyses do not account for changes in customer behavior over time or the varying emphasis on RFM components across different industries. [6], Author:B. Han, M. Kim, and J. Lee, [7]. Author:M. A. Khan and S. Khan,Since the importance of the three variables in the RFM model can vary based on product features and industry characteristics, researchers have experimented with the order of importance of each variable , while creating formulas to calculate the RFM score[8]. Author:K. Fang, Y. Jiang, and M. Song,Analytic Hierarchy Process(AHP) was implemented to determine the various weights of RFM variables.[9].Author:K. Khalili-Damghani, F. Abdi, and S. Abolmakarem WRFM or the Weighted RFM has also been used by . [10] along with a range of data mining approaches such as K-means, ARM, and neural networks. Their proposed model demonstrates how WRFM paired with clustering methods exceeds traditional RFM and enhances business strategy, resulting in higher company profits.In another study, the LRFMP model , where the ‘P’ feature denotes the periodicity of the customer return, was applied to access the consumers in the grocery retail industry. The periodicity of customer returns is a crucial consideration in various organizations, such as grocery stores, when analyzing the behavior of customers. LRFMP was also implemented in where the authors have used two-stage clustering along with the LRFMP model for segmenting customers and analyzing their characteristics across Iranian Fintech companies. Following the initial clustering, more clustering experiments were conducted in segments that required further variable analysis for a better understanding of the important clients.

3.1 DISADVANTAGES

The complexity of data: Most of the existing machine learning models must be able to accurately interpret large and complex datasets to detect Online Shoppers’ Behavior.

• Data availability: Most machine learning models require large amounts of data to create accurate predictions. If data is unavailable in sufficient quantities, then model accuracy may suffer.

• Incorrect labeling: The existing machine learning models are only as accurate as the data trained using the input dataset. If the data has been incorrectly labeled, the model cannot make accurate predictions.

IV. PROPOSED SYSTEM

The proposed system introduces the LRFS (Length, Recency, Frequency, Staying Rate for Revenue) model, a novel approach to customer segmentation that incorporates a ‘‘Staying Rate for Revenue’’ metric, designed to provide a deeper understanding of the online shopping community. By leveraging advanced clustering algorithms and dimensionality reduction techniques, the LRFS model aims to bridge the gap in existing segmentation models, offering a more nuanced reflection of customer interests derived from website-specific features. This research marks a significant advancement in the field of e-commerce customer segmentation through the introduction of the LRFS model. This model uniquely captures customer behaviors and preferences by analyzing data derived from online shopping interactions.

4.1 ADVANTAGES

• Introduction of the ‘‘S’’ component within the LRFS model, demonstrating its potential to significantly enhance revenue generation by improving the precision of customer segmentation

• A detailed exploration of the LRFS model, emphasizing the importance of preprocessing, feature investigation, and the creation of new metrics tailored to the objectives of the model.

• A comprehensive comparative analysis among various unsupervised machine learning algorithms and dimensionality reduction techniques, identifying the

most effective strategies for implementing the LRFS model.

• Evidence of the LRFS model’s superiority over traditional segmentation models, highlighting its capacity for more accurate customer clustering

• The application of a combined customer classification and relationship matrix to analyze clustered groups, yielding insights into customer behaviors and preferences,and illustrating potential use cases

System Architecture

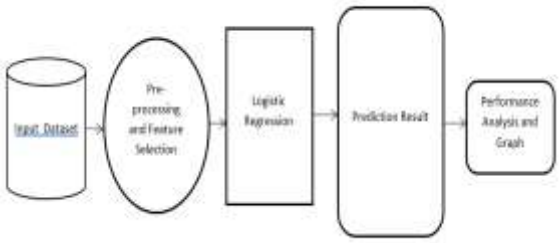


Fig4.1. System Architecture

V. MODULE DESCRIPTION

5.1 User Authentication and Server Login

Secure server login functionality enables authorized access to the system.

Users can log in to manage datasets, initiate training processes, and view results.

5.2 Dataset Management Module

Allows users to browse, upload, and manage datasets (e.g., customer behavioral data collected via Google Analytics).

Supports preprocessing operations such as cleaning, normalization, and feature selection.

5.3 Training and Testing Module

Enables users to split the dataset into training and testing sets.

Supports training using clustering algorithms like **KMeans** and **K-Medoids**.

Integrates **dimensionality reduction techniques** such as PCA, t-SNE, and Autoencoder for enhanced clustering accuracy.

5.4 Accuracy Visualization Module

Displays **trained and tested clustering accuracy** using visual tools such as **bar charts** and statistical performance metrics.

Enables easy comparison between different models (LR, LF, LRF, LRFS) and techniques.

5.5 Prediction and Classification Module

Predicts customer group or cluster based on LRFS metrics.

Visualizes prediction outcomes and highlights group characteristics using graphical representation.

5.6 Tweet Type Analysis (Optional or Add-On Module)

This module analyzes and classifies tweets or text data relevant to e-commerce customer sentiment or behavior.

Shows **predicted tweet types** (e.g., positive, negative, promotional).

Displays **tweet type distribution** in graphical form for better understanding of social sentiment trends.

VI.RESULT

The proposed LRFS model was rigorously evaluated using real-world customer data extracted from Google Analytics. A comparative study was conducted between the traditional models (LR, LF, LRF) and the newly introduced LRFS model, utilizing **KMeans** and **K-Medoids** clustering algorithms in conjunction with **PCA**, **t-SNE**, and **Autoencoder** for dimensionality reduction. The following key observations were derived:

6.1 Enhanced Accuracy: The LRFS model achieved higher clustering precision compared to traditional models. It was particularly effective in distinguishing between high-value and low-value customer segments.

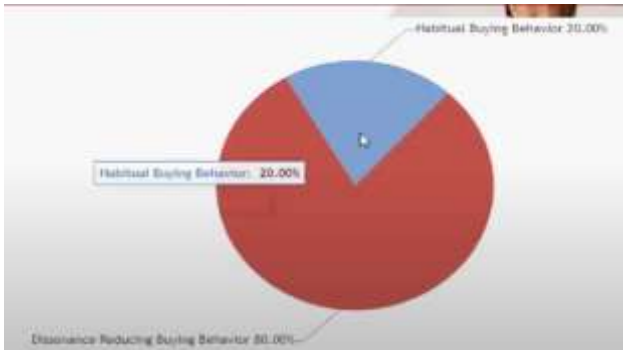
6.2 Impact of the 'S' Component: The inclusion of the **Staying Rate for Revenue (S)** component led to a notable improvement in customer group classification. This metric effectively captured user engagement and loyalty metrics tied directly to revenue.

6.3 Best Performing Combination: Among all combinations tested, the **K-Medoids + t-SNE** pairing yielded the most well-defined and interpretable clusters, offering a deeper understanding of nuanced customer behaviors.

6.4 Visual Insights: Bar charts and cluster plots provided clear evidence of improved segmentation

quality under the LRFS model, making it easier for businesses to target and strategize marketing efforts.

6.5 Use Case Demonstrations: Various practical scenarios were evaluated to showcase how the LRFS model supports personalized marketing, churn prediction, and customer lifecycle management.



Screenshot 6.1

VII.CONCLUSION

The development and implementation of the LRFS customer segmentation model marks a significant advancement in the domain of e-commerce analytics. By integrating the Staying Rate for Revenue **(S)** into the traditional LRF framework, the LRFS model provides a more granular and insightful view of customer behavior. The findings from this study confirm that:The LRFS model outperforms conventional segmentation models in terms of clustering accuracy, customer insight, and strategic applicability.Combining LRFS with K-Medoids and t-SNE offers the best results, capturing the depth and diversity of online customer interactions.Businesses adopting the LRFS model can refine marketing strategies, optimize resource allocation, and enhance customer relationship management by targeting the right customer segments more effectively.In conclusion, the LRFS model serves as a powerful, data-driven tool for online businesses aiming to stay competitive by understanding and acting upon the evolving preferences and behaviors of their digital clientele.

REFERENCES

I. K. Rachmawati, M. Bukhori, F. Nuryanti, and S. Hidayatullah, “Collaboration technology acceptance model, subjective norms and personal innovations on buying interest online,” Int. J. Innov. Sci. Res. Technol., vol. 5, no. 11, pp. 115–122, 2020.

2. C. I. Agustyaningrum, M. Haris, R. Aryanti, and T. Misriati, “Online shopper intention analysis using conventional machine learning and deep neural network classification algorithm,” Jurnal Penelitian Pos dan Informatika, vol. 11, no. 1, pp. 89–100, Nov. 2021.

3. J. A. Al-Gasawneh, M. H. Al-Wadi, B. M. Al-Wadi, B. E. Alown, and N. M. Nuseirat, “The interaction effect of comprehensiveness between social media and online purchasing intention in Jordanian pharmacies,” Int. J. Interact. Mobile Technol., vol. 14, no. 15, p. 208, Sep. 2020.

4. S. Ahsain and M. A. Kbir, “Predicting the client's purchasing intention using machine learning models,” in Proc. E3S Web Conf., vol. 351. Les Ulis, France : EDP Sciences, 2022, p. 01070.

5. K. Tabianan, S. Velu, and V. Ravi, “K-means clustering approach for intelligent customer segmentation using customer purchase behavior data,” Sustainability, vol. 14, no. 12, p. 7243, Jun. 2022.

6. A. Balkaya, E. Tuzunkan, K. Ayaz, Y. Akçay, F. Abut, M. F. Akay, S. Erdem, and A. Alsaç, "Developing customer segmentation models for digital marketing campaigns using machine learning," in Proc. 6th Int. Medit. Sci. Eng. Congr. (IMSEC), Alanya, Turkiye, 2022, pp. 697–701, Paper 318.

7. B. Han, M. Kim, and J. Lee, "Exploring consumer attitudes and purchasing intentions of cross-border online shopping in Korea," J. Korea Trade, 2018, doi: 10.1108/JKT-10-2017-0093.

8. M. A. Khan and S. Khan, "Service convenience and post-purchase behaviour of online buyers: An empirical study," J. Service Sci. Res., vol. 10, no. 2, pp. 167–188, Dec. 2018.

9.K. Fang, Y. Jiang, and M. Song, "Customer profitability forecasting using big data analytics: A case study of the insurance industry," Comput. Ind. Eng., vol. 101, pp. 554–564, Nov. 2019.

10. K. Khalili-Damghani, F. Abdi, and S. Abolmakarem, "Hybrid soft computing approach based on clustering, rule mining, and decision tree analysis for customer segmentation problem: Real case of customer-centric industries," Appl. Soft Comput., vol. 73, pp. 816–828, Dec. 2018.