Journal Website

Article history:
Received 09 May 2026
Revised 22 May 2026
Accepted 27 May 2026
Initial Publication 15 August 2026
Final Publication 23 August 2027

Dynamic Management and Business Analysis

Volume 6, Issue 3, pp 1-30

E-ISSN: 3041-8933

Validation of an Integrated Framework for Identifying and Analyzing Customer Behavior Using LRFFM Indicators and Data Mining Techniques

Mohammad. Ghasemi Zadeh[1], Seyed Abdollah Amin. Mousavi[2*], Mohammad. Maleki Nia[3]

[1] Department of Information Technology Management, KI.C., Islamic Azad University, Kish, Iran
[2] Department of Management, CT.C., Islamic Azad University, Tehran, Iran
[3] Department of Information Technology Management, ST.C., Islamic Azad University, Tehran, Iran

*** Corresponding author email address**: saa.mousavi@iau.ac.ir

A r t i c l e I n f o

Article type:
Original Research

How to cite this article:
Ghasemi Zadeh, M., Mousavi, S. A. A., & Maleki Nia, M. (2027). Validation of an Integrated Framework for Identifying and Analyzing Customer Behavior Using LRFFM Indicators and Data Mining Techniques. *Dynamic Management and Business Analysis*, *6*(3), 1-30.
https://doi.org/10.61838/dmbaj.428

A B S T R A C T

Objective: This study aimed to validate an integrated framework for identifying and analyzing retail customer behavior using extended LRFFM indicators combined with clustering, association-rule mining, and machine-learning classification techniques.

Methodology: This applied, quantitative, data-driven study was conducted according to the CRISP-DM process. The dataset consisted of 541,013 cleaned transactional records obtained from customers of a retail store. Transaction-level data were aggregated at the customer level, and Recency, two Frequency indicators, Monetary value, and Length of relationship were extracted. Customers with more than 93 days since their last purchase were defined as churners. K-Means was used for customer segmentation, FP-Growth and association rules were applied to identify basket-purchase patterns, and ten classification algorithms were compared, including decision tree, k-nearest neighbors, Naive Bayes, neural network, logistic regression, support vector machine, linear discriminant analysis, AdaBoost, bagging, and stacking.

Findings: Clustering of 26,428 customers identified three distinct segments. The most valuable cluster had a Recency of 19.29, an average of 144.35 transactions, a Monetary value of 95,455,030.90, an average of 260.49 purchased items, and a Length of 131.32. Among 12,433 churned customers, 880 were identified as valuable churners. Association-rule analysis revealed different basket structures between valuable churners and non-churners, with several rules reaching a confidence of 1.00. In churn classification, stacking achieved the best overall performance, with 86.47% accuracy, 84.07% precision, 90.00% recall, and an F1-score of 86.93%. Logistic regression and Naive Bayes ranked second and third, respectively.

Conclusion: The integrated LRFFM and data-mining framework successfully combined customer value assessment, churn identification, market-basket analysis, and predictive modeling, and can support more targeted retail decisions concerning customer retention, reactivation, and relationship management.

Keywords: *Customer Behavior, LRFFM, Customer Churn, Data Mining, Clustering, Association Rules, Machine Learning, Retail*

EXTENDED ABSTRACT

Introduction

The rapid expansion of transactional databases, digital retail systems, and data-driven customer relationship management has transformed the way retailers identify, evaluate, and respond to customer behavior. In contemporary retail environments, customer value can no longer be adequately understood through aggregate sales indicators alone, because customers differ substantially in terms of purchase recency, transaction frequency, basket size, monetary contribution, and length of relationship with the retailer. Machine learning and data mining techniques have therefore become increasingly important for converting large volumes of transactional data into actionable knowledge concerning customer retention, repurchase behavior, purchasing patterns, and churn risk (Manzoor et al., 2024; Mouli et al., 2024; Vemulapalli, 2024). Customer churn represents a particularly important managerial problem because the cessation or substantial decline of purchasing activity can reduce long-term customer value and increase the costs associated with acquiring replacement customers. Although churn can be explicitly observed in contractual industries such as telecommunications, banking, and subscription-based services, it is more difficult to define in non-contractual retail settings, where customers rarely formally terminate their relationship with a store. Consequently, churn in retail must often be inferred from behavioral signals such as prolonged inactivity, declining purchase frequency, changing transaction patterns, and increasing time since the most recent purchase (Matuszelański & Kopczewska, 2022; Sweidan et al., 2022; Zelenkov & Suchkova, 2023). Previous studies have demonstrated the usefulness of machine learning for churn prediction across telecommunications, banking, e-commerce, hospitality, pharmaceutical retail, and fast-moving consumer goods (Adeniran et al., 2024; Brito et al., 2024; Dursun-Cengizci & Caber, 2025; Espinoza-Vega & Roa, 2024; Mahdi & Jabbari, 2024; Phumchusri & Amornvetchayakul, 2024). However, a major methodological challenge is that many existing approaches focus primarily on a single predictive task rather than integrating customer valuation, behavioral segmentation, basket analysis, and churn classification within one analytical framework. RFM-based approaches remain among the most widely used methods for summarizing customer behavior, while recent research has shown that incorporating time-varying or extended behavioral measures can improve the representation of customer dynamics (Chou et al., 2022; Mena et al., 2024). Extending conventional RFM into an LRFFM structure allows customer behavior to be represented through Recency, two dimensions of Frequency, Monetary value, and Length of relationship, thereby distinguishing between purchase occasions and the total volume of items purchased. Such enriched behavioral indicators can be combined with clustering to identify heterogeneous customer groups and with classification techniques to distinguish churners from non-churners (Liu et al., 2022; Sina Mirabdolbaghi & Amiri, 2022). Moreover, association-rule mining can complement these approaches by revealing recurring product combinations and relationships within customer baskets, thereby connecting customer-level value analysis with product-level purchasing behavior. Ensemble learning has also emerged as a promising approach for churn prediction because combining multiple classifiers can improve robustness and balance predictive performance across accuracy, precision, recall, and F1-score (Bogaert & Delaere, 2023; Kimura, 2022). Accordingly, the present study aimed to validate an integrated framework for identifying and analyzing retail customer behavior

using LRFFM indicators in combination with clustering, association-rule mining, and machine-learning classification techniques.

Methods and Materials

The study adopted an applied, quantitative, data-driven design based on a real transactional dataset from Branch 1 of the Seraj retail store. The analytical workflow followed the CRISP-DM process, including business understanding, data understanding, data preparation, modeling, evaluation, and interpretation of outputs. The initial dataset contained approximately 541,000 transaction-level records and 13 variables, including store name, product name, barcode, customer identifier, season, month, day of week, day, net sales amount after returns, date of last sale, number of days since the last sale, total number of items sold, and number of sales invoices. Records with missing customer identifiers or product barcodes, invalid values, negative net sales amounts, and unusable transactional information were removed or corrected during preprocessing. After cleaning, 541,013 records remained. Transaction-level data were subsequently aggregated at the customer level to calculate LRFFM indicators. Recency represented the number of days since the customer's most recent purchase; the first Frequency measure represented the number of purchase transactions; the second Frequency measure represented the total number of products purchased; Monetary represented the customer's total net purchase value; and Length represented the duration between the customer's first and most recent observed purchases. Customer churn was operationally defined using a 93-day inactivity threshold: customers with Recency greater than 93 days were classified as churners, whereas the remaining customers were classified as non-churners. K-Means clustering was applied to LRFFM indicators without using churn status as an input feature, and the Davies–Bouldin index was used to select an appropriate number of clusters. A separate clustering analysis was also conducted among churned customers to identify the most valuable churned segment. For market-basket analysis, customer-level cluster and churn information was joined with barcode-level purchasing data, the dataset was converted into a binary transaction–product matrix, and FP-Growth was used to identify frequent itemsets and generate association rules. For churn classification, the analysis focused on customers belonging to the valuable segment. Because the original classes were imbalanced, random undersampling was used to create a balanced dataset containing 170 churners and 170 non-churners. Product-purchase variables were coded as binary indicators, and features with insufficient observations were removed, reducing the initial 8,441 product variables to 485 features. Ten classifiers were compared: decision tree, k-nearest neighbors, Naive Bayes, neural network, logistic regression, support vector machine, linear discriminant analysis, AdaBoost, bagging, and stacking. Their performance was evaluated using accuracy, precision, recall, and F1-score.

Findings

The final customer-level dataset contained 26,428 customers. Evaluation of K-Means solutions with two to ten clusters showed that the three-cluster solution produced the most favorable reported Davies–Bouldin value of -0.549 and was therefore retained. The three clusters contained 22,427, 256, and 3,745 customers, respectively. The smallest cluster represented the most valuable group: it had an average of 144.35 transactions, an average monetary value of 95,455,030.90, an average of 260.49 purchased items, a mean Recency of 19.29 days, and a mean Length of 131.32 days. In addition, 93% of customers in this cluster were non-churners. By contrast, the largest cluster exhibited substantially lower purchase frequency, monetary contribution, basket volume, and relationship length, together with a markedly higher

E-ISSN: 3041-8933

Recency. A separate clustering analysis of 12,433 churned customers identified two clusters. The more valuable churned cluster contained 880 customers and showed a mean transaction frequency of 42.83, mean monetary value of 22,566,464.85, mean purchased-item count of 70.15, Recency of 126.40, and Length of 27.35, whereas the other churn cluster contained 11,553 customers and had corresponding values of 10.73, 3,745,859.98, 15.98, 136.73, and 6.24. Association-rule analysis revealed distinct purchasing structures among valuable non-churners and valuable churners. Among valuable non-churners, several combinations involving product codes 2200170, 2203314, 2202312, 2204359, and 2204364 showed strong co-occurrence, with multiple rules reaching confidence levels above 0.90 and some reaching 1.00. For example, the combination of products 2202312 and 2204364 predicted the presence of 2200170 with confidence 1.00. Among valuable churners, a different group of products, particularly 2201609, 2200351, 2200348, 2201618, and 2201611, formed the core of recurrent purchasing patterns; the rule 2201609 + 2200348 → 2200351 reached confidence 1.00. In the classification stage, stacking achieved the best overall rank, with accuracy of 86.47%, precision of 84.07%, recall of 90.00%, and F1-score of 86.93%. Logistic regression ranked second, achieving accuracy of 85.88%, precision of 79.90%, recall of 95.88%, and F1-score of 87.17%. Naive Bayes ranked third with accuracy of 85.59%, precision of 83.06%, recall of 89.41%, and F1-score of 86.12%. Support vector machine, decision tree, and AdaBoost also produced acceptable results. Bagging yielded the highest recall at 98.82% but substantially lower precision of 64.62%, indicating a high rate of false-positive churn predictions. Linear discriminant analysis produced the weakest overall performance, with accuracy of 58.53%, precision of 57.92%, recall of 62.35%, and F1-score of 60.06%.

Discussion and Conclusion

The findings demonstrate that customer behavior in retail is multidimensional and that no single analytical technique is sufficient to capture its full complexity. LRFFM indicators successfully differentiated customers according to recency of interaction, transaction intensity, purchasing volume, financial value, and relationship duration, while clustering converted these continuous behavioral differences into managerially interpretable customer segments. Particularly important was the identification of a small but highly valuable segment characterized by recent purchases, high frequency, large basket volume, substantial monetary contribution, long-term interaction, and a high proportion of non-churners. The separate identification of valuable churned customers further showed that churn status alone is insufficient for customer prioritization: some churned customers had substantially stronger historical relationships and higher economic value than others and therefore represent more important targets for reactivation strategies. Association-rule analysis added another layer of information by showing that valuable churners and non-churners differed not only in aggregated behavioral indicators but also in the specific product combinations appearing in their baskets. This suggests that customer retention strategies can be enhanced by combining value-based segmentation with product-level behavioral knowledge. The classification results also confirmed the usefulness of ensemble learning. Stacking provided the most balanced overall performance, while logistic regression produced particularly high sensitivity to churners and bagging nearly maximized recall at the expense of a much higher false-positive rate. These differences indicate that model selection should depend not only on overall predictive accuracy but also on the managerial cost of false negatives and false positives. If missing a valuable churner is considered particularly costly, a high-recall model may be preferred; if retention resources are limited, a

more balanced model such as stacking may be more appropriate. Overall, the results validate the proposed integrated framework as a practical approach for transforming raw retail transactions into multiple forms of actionable knowledge. By linking LRFFM-based customer value assessment, behavioral clustering, market-basket analysis, and machine-learning churn classification, the framework enables simultaneous identification of valuable customers, valuable churners, recurrent product combinations, and customers at risk of disengagement. The framework can therefore support more targeted retention programs, personalized offers, cross-selling strategies, reactivation campaigns, and data-driven customer relationship management in retail settings.

E-ISSN: 3041-8933


شاپای الکترونیکی: ۸۹۳۳-۳۰۴۱

مدیریت پویا و تحلیل کسب و کار

دوره ۶، شماره ۳، صفحه ۳۰-۱

وبسایت مجله

تاریخچه مقاله
دریافت شده در تاریخ ۱۹ اردیبهشت ۱۴۰۵
اصلاح شده در تاریخ ۱ خرداد ۱۴۰۵
پذیرفته شده در تاریخ ۶ خرداد ۱۴۰۵
اولین انتشار در تاریخ ۲۴ مرداد ۱۴۰۵
انتشار نهایی در تاریخ ۱ شهریور ۱۴۰۶

اعتبارسنجی یک چارچوب یکپارچه برای شناسایی و تحلیل رفتار مشتریان با استفاده از شاخص‌های LRFFM و تکنیک‌های داده‌کاوی

محمد قاسمی زاده[۱]، دکتر سید عبدالله امین موسوی[۲*]، دکتر محمد ملکی نیا[۳]

۱. گروه مدیریت فن آوری اطلاعات، واحد بین المللی کیش، دانشگاه آزاد اسلامی، کیش، ایران
۲. گروه مدیریت، واحد تهران مرکزی، دانشگاه آزاد اسلامی، تهران، ایران
۳. گروه مدیریت فن آوری اطلاعات، واحد تهران جنوب، دانشگاه آزاد اسلامی، تهران، ایران

*ایمیل نویسنده مسئول: saa.mousavi@iau.ac.ir

اطلاعات مقاله

نوع مقاله
پژوهشی اصیل

نحوه استناد به این مقاله:
قاسمی زاده، محمد.، موسوی، سید عبدالله امین.، و ملکی نیا، محمد. (۱۴۰۶). اعتبارسنجی یک چارچوب یکپارچه برای شناسایی و تحلیل رفتار مشتریان با استفاده از شاخص‌های LRFFM و تکنیک‌های داده‌کاوی. *مدیریت پویا و تحلیل کسب و کار*، ۶(۳)، ۳۰-۱.

چکیده

هدف: هدف پژوهش حاضر، اعتبارسنجی یک چارچوب یکپارچه برای شناسایی و تحلیل رفتار مشتریان فروشگاه خرده‌فروشی با استفاده از شاخص‌های توسعه‌یافته LRFFM و ترکیب روش‌های خوشه‌بندی، قوانین وابستگی و الگوریتم‌های طبقه‌بندی بود. **روش‌شناسی:** پژوهش حاضر از نوع کاربردی، کمی و داده‌محور بود و بر اساس فرایند CRISP-DM انجام شد. داده‌ها شامل ۵۴۱,۰۱۳ رکورد تراکنشی پاک‌سازی‌شده مربوط به مشتریان یک فروشگاه خرده‌فروشی بود. داده‌های تراکنشی در سطح مشتری تجمیع و شاخص‌هایRecency ، دو شاخصFrequency ، Monetaryو Length استخراج شدند. مشتریانی که بیش از ۹۳ روز از آخرین خرید آنان گذشته بود، ریزشی در نظر گرفته شدند. برای بخش‌بندی مشتریان ازK-Means ، برای استخراج الگوهای سبد خرید از FP-Growth و قوانین وابستگی، و برای پیش‌بینی ریزش از ده الگوریتم طبقه‌بندی شامل درخت تصمیم، نزدیک‌ترین همسایه، نیوبیز، شبکه عصبی، رگرسیون لجستیک، ماشین بردار پشتیبان، تحلیل تمایزی خطی، آدابوست، بگینگ و استیکینگ استفاده شد. **یافته‌ها:** خوشه‌بندی ۲۶,۴۲۸ مشتری، سه خوشه متمایز را آشکار کرد و خوشه ارزشمندتر دارای Recency برابر با ۱۹.۲۹، میانگین ۱۴۴.۳۵ تراکنش، Monetaryبرابر با ۹۵,۴۵۵,۰۳۰.۹۰، میانگین ۲۶۰.۴۹ قلم خرید و Length برابر با ۱۳۱.۳۲ بود. در میان ۱۲,۴۳۳ مشتری ریزشی، ۸۸۰ مشتری به‌عنوان مشتریان ارزشمند ریزشی شناسایی شدند. قوانین وابستگی نیز الگوهای متفاوتی را میان مشتریان ارزشمند ریزشی و غیرریزشی نشان دادند و برخی قوانین به اطمینان ۱.۰۰ رسیدند. در طبقه‌بندی، استیکینگ با Accuracy برابر با ۸۶.۴۷ درصد، Precision برابر با ۸۴.۰۷ درصد، Recallبرابر با ۹۰.۰۰ درصد و F۱ برابر با ۸۶.۹۳ درصد بهترین عملکرد کلی را داشت. رگرسیون لجستیک و نیوبیز در رتبه‌های بعدی قرار گرفتند. **نتیجه‌گیری:** چارچوب یکپارچه LRFFM و داده‌کاوی توانست ارزش مشتری، وضعیت ریزش، الگوهای سبد خرید و قابلیت پیش‌بینی ریزش را به‌صورت هم‌زمان تحلیل کند و می‌تواند به‌عنوان ابزاری برای پشتیبانی از تصمیم‌گیری‌های مرتبط با نگهداشت، بازگشت و مدیریت هدفمند مشتریان در خرده‌فروشی استفاده شود.

کلیدواژگان: *رفتار مشتری، LRFFM، ریزش مشتری، داده‌کاوی، خوشه‌بندی، قوانین وابستگی، یادگیری ماشین، خرده‌فروشی*

مقدمه

تحولات گسترده در فناوری اطلاعات، توسعه سامانه‌های فروش الکترونیکی و حضوری و افزایش توان سازمان‌ها در ثبت و ذخیره داده‌های تراکنشی، ماهیت تصمیم‌گیری در مدیریت ارتباط با مشتری را به‌طور اساسی تغییر داده است. امروزه فروشگاه‌های خرده‌فروشی با حجم عظیمی از داده‌های مربوط به خرید، زمان مراجعه، تعداد تراکنش‌ها، ارزش مالی خرید، اقلام انتخاب‌شده و الگوهای تکرار خرید مشتریان مواجه‌اند که در صورت تحلیل مناسب می‌توانند از داده‌های خام به دانش قابل‌استفاده برای تصمیم‌گیری مدیریتی تبدیل شوند. در چنین محیطی، شناخت رفتار مشتری دیگر نمی‌تواند صرفاً بر قضاوت‌های شهودی یا شاخص‌های کلی فروش استوار باشد، بلکه مستلزم استفاده از رویکردهای تحلیلی است که بتوانند تغییرات رفتاری مشتریان را در طول زمان شناسایی کرده و احتمال تداوم یا توقف تعامل آنان با کسب‌وکار را آشکار سازند. مطالعات اخیر نشان داده‌اند که الگوریتم‌های یادگیری ماشین می‌توانند با استفاده از تاریخچه رفتار مشتری، الگوهایی را شناسایی کنند که در روش‌های تحلیلی سنتی کمتر قابل تشخیص هستند و از این طریق زمینه پیش‌بینی ریزش، حفظ مشتری و توسعه راهبردهای مدیریت ارتباط با مشتری را فراهم آورند (Manzoor et al., 2024; Mouli et al., 2024; Vemulapalli, 2024).

در فضای رقابتی خرده‌فروشی، مسئله حفظ مشتری اهمیت ویژه‌ای دارد؛ زیرا ارزش مشتری صرفاً به یک خرید منفرد محدود نمی‌شود و استمرار تعامل او با فروشگاه، تعداد دفعات مراجعه، حجم اقلام خریداری‌شده و ارزش مالی تراکنش‌ها می‌تواند سهم وی را در عملکرد بلندمدت کسب‌وکار تعیین کند. از این منظر، یکی از مسائل اصلی مدیریت مشتری، شناسایی مشتریانی است که الگوی تعامل آنان رو به کاهش است یا احتمال دارد ارتباط خرید خود را با فروشگاه متوقف کنند. مطالعات حوزه نگهداشت مشتری نشان داده‌اند که استفاده از داده‌های رفتاری می‌تواند به شناسایی زودهنگام مشتریان در معرض ریزش و طراحی تصمیمات هدفمند برای حفظ آنان کمک کند (Adeniran et al., 2024; Das et al., 2023). همچنین، توسعه سامانه‌های پیش‌بینی ریزش بر مبنای یادگیری ماشین نشان داده است که ترکیب انتخاب مناسب متغیرها، پیش‌پردازش داده و انتخاب الگوریتم طبقه‌بندی می‌تواند توان سازمان در شناسایی مشتریان پرریسک را بهبود دهد (Lalwani et al., 2022; Sana et al., 2022). ازاین‌رو، ریزش مشتری نه‌فقط یک پیامد بازاریابی، بلکه یک مسئله تحلیلی چندبعدی است که باید از طریق بررسی تغییرات رفتار خرید، فراوانی تعامل، ارزش مالی و الگوهای زمانی مشتری مطالعه شود.

بااین‌حال، تعریف و شناسایی ریزش در تمام صنایع یکسان نیست. در صنایع قراردادی مانند مخابرات، بانکداری یا نرم‌افزارهای اشتراکی، خاتمه قرارداد، لغو اشتراک یا بسته‌شدن حساب می‌تواند معیار نسبتاً مشخصی برای ریزش باشد. پژوهش‌های انجام‌شده در صنعت مخابرات نشان داده‌اند که الگوریتم‌های یادگیری ماشین با استفاده از متغیرهای رفتاری و خدماتی می‌توانند مشتریان ریزشی را با دقت مناسبی پیش‌بینی کنند (Saleh & Saha, 2023; Srivastava & Sinha, 2024). در صنعت بانکداری نیز مدل‌های طبقه‌بندی برای شناسایی مشتریان در معرض ترک خدمات و افزایش عملکرد مدل‌های پیش‌بینی ریزش توسعه یافته‌اند (Brito et al., 2024; Tran et al., 2023). در شرکت‌های ارائه‌دهنده نرم‌افزار به‌عنوان خدمت نیز الگوهای استفاده مشتری و داده‌های تعامل به‌عنوان ورودی مدل‌های یادگیری ماشین برای پیش‌بینی ریزش مورد استفاده قرار گرفته‌اند (Phumchusri & Amornvetchayakul, 2024). در مقابل، در خرده‌فروشی غیرقراردادی، مشتری الزام رسمی برای اعلام قطع رابطه ندارد و ممکن است صرفاً برای مدتی طولانی خریدی انجام ندهد؛ بنابراین، ریزش باید از روی تغییرات رفتار تراکنشی و فاصله زمانی از آخرین خرید استنباط شود.

این ویژگی موجب می‌شود پیش‌بینی ریزش در خرده‌فروشی از نظر مفهومی و محاسباتی پیچیده‌تر باشد. پژوهش‌های انجام‌شده در تجارت الکترونیکی و خرده‌فروشی نشان داده‌اند که الگوی خرید مشتری، موقعیت زمانی تراکنش‌ها و تغییرات تدریجی رفتار، اطلاعات مهمی

E-ISSN: 3041-8933

برای تشخیص کاهش تعامل فراهم می‌کنند (Matuszelański & Kopczewska, 2022; Sweidan et al., 2022). همچنین، استفاده از چارچوب‌های یادگیری عمیق متوالی نشان داده است که ترتیب زمانی رفتار مشتری می‌تواند اطلاعاتی فراتر از وضعیت ایستای او در یک مقطع زمانی ارائه کند (Equihua et al., 2023). مطالعه تغییر در الگوهای رفتاری مشتریان نیز نشان داده است که توجه به روند تغییر رفتار می‌تواند برای پیش‌بینی ریزش ارزشمند باشد، زیرا مشتریان پیش از قطع کامل تعامل ممکن است علائمی مانند کاهش فراوانی خرید، افزایش فاصله میان تراکنش‌ها یا تغییر در ترکیب خرید نشان دهند (Zelenkov & Suchkova, 2023). ازاین‌رو، در محیط خرده‌فروشی، ریزش باید در چارچوب رفتار پویا و تاریخچه تراکنش‌های مشتری تفسیر شود.

یکی از شناخته‌شده‌ترین رویکردها برای تبدیل داده‌های خام خرید به متغیرهای رفتاری قابل‌تحلیل، استفاده از شاخص‌های مبتنی بر RFM است. این رویکرد رفتار مشتری را از طریق تازگی خرید، فراوانی خرید و ارزش مالی تراکنش‌ها خلاصه می‌کند و زمینه مناسبی برای بخش‌بندی و مدل‌سازی مشتریان فراهم می‌آورد. بااین‌حال، پژوهش‌های جدید نشان داده‌اند که استفاده از مقادیر RFM به‌صورت ایستا ممکن است بخشی از پویایی رفتار مشتری را نادیده بگیرد. بررسی معیارهای RFM متغیر در طول زمان نشان داده است که استفاده از اطلاعات زمانی و تغییرپذیر می‌تواند قدرت مدل‌های پیش‌بینی ریزش را افزایش دهد (Mena et al., 2024). از سوی دیگر، رویکردهای پیش‌بینی خرید مجدد نیز نشان می‌دهند که ترکیب مدل‌های رفتاری مشتری با یادگیری ماشین می‌تواند درک دقیق‌تری از احتمال استمرار تعامل فراهم کند (Chou et al., 2022). بر همین مبنا، توسعه RFM به چارچوب LRFFM، که علاوه بر تازگی و ارزش مالی، دو بعد متفاوت از فراوانی و همچنین طول رابطه مشتری با کسب‌وکار را در نظر می‌گیرد، می‌تواند تصویر غنی‌تری از ارزش و وضعیت رفتاری مشتری ایجاد کند.

تمایز میان تعداد دفعات خرید و حجم اقلام خریداری‌شده در چنین چارچوبی اهمیت خاصی دارد. مشتریانی که تعداد مراجعات زیادی دارند لزوماً در هر مراجعه حجم خرید بالایی ندارند و، برعکس، برخی مشتریان ممکن است با دفعات مراجعه کمتر، ارزش مالی یا حجم سبد خرید بیشتری ایجاد کنند. همچنین، طول رابطه مشتری با فروشگاه می‌تواند نشان‌دهنده سابقه تعامل باشد، اما این شاخص فقط در کنار تازگی خرید، فراوانی و ارزش مالی معنا پیدا می‌کند؛ زیرا ممکن است مشتری سابقه‌ای طولانی داشته باشد ولی مدت زیادی از آخرین خرید او گذشته باشد. بنابراین، استفاده هم‌زمان از شاخص‌های LRFFM می‌تواند از تقلیل رفتار پیچیده مشتری به یک متغیر منفرد جلوگیری کند و زمینه مناسب‌تری برای تحلیل چندبعدی فراهم سازد. مطالعاتی که بر انتخاب ویژگی‌ها و کاهش ابعاد در مدل‌های پیش‌بینی ریزش تمرکز داشته‌اند نیز نشان داده‌اند که کیفیت نمایش داده و انتخاب متغیرهای معنادار می‌تواند به اندازه انتخاب الگوریتم در عملکرد مدل مؤثر باشد (Sina Mirabdolbaghi & Amiri, 2022).

در کنار شاخص‌های رفتاری، خوشه‌بندی یکی از روش‌های مهم برای شناخت ناهمگونی مشتریان است. همه مشتریان یک فروشگاه از نظر تازگی خرید، تعداد تراکنش‌ها، حجم کالا، ارزش مالی و سابقه تعامل یکسان نیستند؛ بنابراین، تحلیل کل مشتریان به‌عنوان یک جمعیت همگن می‌تواند تفاوت‌های مدیریتی مهم را پنهان کند. الگوریتم‌های خوشه‌بندی قادرند گروه‌هایی از مشتریان با الگوهای رفتاری مشابه ایجاد کنند و این گروه‌ها سپس از نظر ارزش، سطح فعالیت یا احتمال ریزش تفسیر شوند. رویکردهای ترکیبی خوشه‌بندی و طبقه‌بندی نشان داده‌اند که استفاده از خوشه‌بندی پیش از مرحله پیش‌بینی می‌تواند ساختار پنهان داده را آشکار سازد و در برخی شرایط کیفیت تشخیص ریزش را افزایش دهد (Liu et al., 2022). اهمیت این موضوع در خرده‌فروشی بیشتر است؛ زیرا ممکن است یک مشتری ریزشی از نظر ارزش گذشته با مشتری ریزشی دیگری تفاوت اساسی داشته باشد و در نتیجه، اولویت مدیریتی یکسانی برای بازگرداندن این دو مشتری وجود نداشته باشد.

در همین راستا، مفهوم «مشتری ارزشمند ریزشی» اهمیت ویژه‌ای پیدا می‌کند. تمامی مشتریان ریزشی هزینه و پیامد یکسانی برای کسب‌وکار ایجاد نمی‌کنند. مشتری‌ای که پیش از ریزش دارای فراوانی خرید بالا، مبلغ خرید زیاد، حجم سبد بزرگ و رابطه طولانی با فروشگاه

بوده است، از نظر مدیریتی با مشتری کم‌فعال و کم‌ارزش متفاوت است. ازاین‌رو، چارچوب تحلیلی مطلوب باید بتواند علاوه بر تشخیص ریزش، سطح ارزش گذشته مشتری را نیز مشخص کند. مطالعات خرده‌فروشی دارویی و کالاهای تندمصرف نشان داده‌اند که مدل‌های یادگیری ماشین می‌توانند در شناسایی مشتریان پرریسک و حمایت از راهبردهای نگهداشت در محیط‌های خرده‌فروشی نقش مؤثری داشته باشند (Espinoza-Vega & Roa, 2024; Mahdi & Jabbari, 2024). همچنین، مطالعات پیش‌بینی ریزش در هتلداری نشان داده‌اند که تاریخچه رفتار مشتریان تکراری می‌تواند برای پیش‌بینی مشتریانی که احتمال قطع تعامل آنان وجود دارد، به‌طور مؤثری مورد استفاده قرار گیرد (Dursun-Cengizci & Caber, 2025). این یافته‌ها اهمیت ترکیب دو مفهوم ارزش مشتری و خطر ریزش را برجسته می‌کنند.

بعد دیگری که در بسیاری از مطالعات پیش‌بینی ریزش کمتر با تحلیل ارزش مشتری ادغام شده است، ساختار سبد خرید و روابط میان کالاهاست. داده‌های تراکنشی فقط نشان نمی‌دهند که مشتری چه مقدار یا چند بار خرید کرده است؛ بلکه مشخص می‌کنند چه کالاهایی را همراه یکدیگر انتخاب کرده است. قوانین وابستگی با شناسایی مجموعه اقلام پرتکرار و روابط اگر–آنگاه میان کالاها، امکان استخراج الگوهای هم‌وقوعی در سبد خرید را فراهم می‌کنند. در چارچوب مدیریت مشتری، این اطلاعات می‌تواند برای شناخت کالاهای مکمل، طراحی فروش متقاطع، پیشنهادهای شخصی‌سازی‌شده و بررسی تفاوت سبد خرید میان گروه‌های مشتریان به کار رود. در نتیجه، ترکیب قوانین وابستگی با خوشه‌بندی مشتریان می‌تواند تحلیل سبد خرید را از یک تحلیل کلی در سطح فروشگاه به تحلیلی هدفمند برای گروه‌های ارزشمند، فعال یا ریزشی تبدیل کند. این موضوع با یافته‌های پژوهش‌های پیش‌بینی رفتار خرید نیز همسو است که نشان داده‌اند روش‌های طبقه‌بندی یادگیری ماشین می‌توانند از اطلاعات رفتاری و خرید برای پیش‌بینی تصمیمات و الگوهای مشتریان استفاده کنند (Chaubey et al., 2023).

علاوه بر خوشه‌بندی و قوانین وابستگی، طبقه‌بندی بخش اساسی چارچوب‌های پیش‌بینی ریزش را تشکیل می‌دهد. مطالعات متعدد، الگوریتم‌هایی مانند درخت تصمیم‌گیری، نزدیک‌ترین همسایه، نیوبیز، رگرسیون لجستیک، ماشین بردار پشتیبان، شبکه‌های عصبی و روش‌های تجمیعی را برای مسئله ریزش بررسی کرده‌اند. مقایسه مدل‌های طبقه‌بندی نشان داده است که هیچ الگوریتمی در همه مجموعه داده‌ها به‌صورت مطلق بهترین نیست و عملکرد مدل به ماهیت داده، نوع ویژگی‌ها، عدم تعادل کلاس‌ها و معیار ارزیابی بستگی دارد (Louro et al., 2024; Mouli et al., 2024). همچنین، مرور مطالعات یادگیری ماشین در پیش‌بینی ریزش نشان می‌دهد که ارزیابی هم‌زمان Accuracy، Precision، Recall و F1 برای انتخاب مدلی که از نظر کسب‌وکار نیز قابل‌استفاده باشد ضروری است (Manzoor et al., 2024). به‌ویژه در مسئله ریزش، اتکا به Accuracy به‌تنهایی می‌تواند گمراه‌کننده باشد؛ زیرا مدل ممکن است با پیش‌بینی کلاس اکثریت دقت کلی بالا ولی توان اندکی در شناسایی مشتریان واقعاً در معرض ریزش داشته باشد.

عدم تعادل کلاس‌ها یکی از چالش‌های مهم این حوزه است. در بسیاری از داده‌های واقعی، تعداد مشتریان غیرریزشی بیشتر از مشتریان ریزشی است و این ناهمگونی می‌تواند موجب گرایش مدل به کلاس اکثریت شود. رویکردهای ترکیبی بازنمونه‌گیری و یادگیری تجمیعی نشان داده‌اند که مدیریت عدم تعادل می‌تواند توان تشخیص کلاس اقلیت را بهبود بخشد (Kimura, 2022). همچنین، استفاده از هوش مصنوعی و مدل‌های پیش‌بینی ریزش در بازارهای کسب‌وکار نشان داده است که طراحی مناسب فرایند مدل‌سازی، از آماده‌سازی داده تا انتخاب الگوریتم و ارزیابی خروجی، برای دستیابی به عملکرد پایدار ضروری است (Faritha Banu et al., 2022). بنابراین، در یک چارچوب کاربردی برای خرده‌فروشی، متعادل‌سازی داده‌ها و مقایسه چند مدل طبقه‌بندی باید بخشی از فرایند اعتبارسنجی باشد.

در این میان، روش‌های تجمیعی یا Ensemble Learning جایگاه مهمی در ادبیات جدید پیش‌بینی ریزش یافته‌اند. این روش‌ها به‌جای اتکا به یک طبقه‌کننده منفرد، خروجی چند مدل را ترکیب می‌کنند و می‌توانند از نقاط قوت الگوریتم‌های مختلف بهره ببرند. مطالعات مقایسه‌ای در زمینه پیش‌بینی ریزش نشان داده‌اند که روش‌های تجمیعی در بسیاری از مجموعه داده‌ها توان رقابتی بالایی دارند، اگرچه عملکرد

E-ISSN: 3041-8933

آن‌ها همچنان به معماری ترکیب، ویژگی‌های داده و مدل‌های پایه وابسته است (Bogaert & Delaere, 2023). استفاده از استیکینگ، بگینگ و تقویت‌کننده‌هایی مانند آدابوست می‌تواند امکان مقایسه عملکرد مدل‌های منفرد و ترکیبی را فراهم کند و انتخاب مدل را بر مبنای شواهد تجربی انجام دهد. این مسئله در کاربردهای عملی مهم است، زیرا شناسایی نادرست یک مشتری غیرریزشی به‌عنوان ریزشی می‌تواند منابع نگهداشت را هدر دهد و عدم شناسایی یک مشتری واقعاً ریزشی نیز می‌تواند به از دست رفتن فرصت مداخله منجر شود.

از سوی دیگر، ادبیات نشان می‌دهد که کاربرد مدل‌های ریزش به یک صنعت خاص محدود نیست. در تجارت الکترونیکی، روش‌های هوش تجاری مبتنی بر یادگیری ماشین برای پیشگیری از ریزش مشتری توسعه یافته‌اند (Shobana et al., 2023). در خرده‌فروشی، تحلیل ریزش در سطح دسته‌های کالایی نیز نشان داده است که رفتار مشتری می‌تواند در سطوح مختلف محصولی بررسی شود و الگوهای ترک خرید ممکن است در میان گروه‌های محصول تفاوت داشته باشند (Sousa, 2023). در محیط هتلداری، استفاده از متن‌کاوی و جنگل تصادفی برای تحلیل پروژه‌های پیش‌بینی ریزش نشان‌دهنده امکان ادغام انواع مختلف داده برای شناخت رفتار مشتری است (Taherkhani et al., 2023). همچنین، مطالعه ریزش در خرده‌فروشی با رویکرد یادگیری ماشین نشان داده است که فناوری‌های تحلیلی می‌توانند به تحول تصمیم‌گیری در مدیریت مشتری کمک کنند (Jung et al., 2023). گستردگی این کاربردها نشان می‌دهد که مسئله اصلی دیگر امکان استفاده از یادگیری ماشین نیست، بلکه چگونگی طراحی چارچوبی است که خروجی الگوریتم‌ها را به دانش منسجم و قابل‌تفسیر مدیریتی تبدیل کند.

نکته مهم دیگر این است که بخش بزرگی از پژوهش‌های پیشین بر یک وظیفه تحلیلی منفرد تمرکز داشته‌اند؛ برخی صرفاً ریزش را پیش‌بینی کرده‌اند، برخی بر بهبود الگوریتم طبقه‌بندی متمرکز شده‌اند و برخی دیگر تنها رفتار خرید یا تغییرات زمانی را بررسی کرده‌اند. برای مثال، مطالعات یادگیری عمیق در محیط خرده‌فروشی و چارچوب‌های متوالی بر بهبود پیش‌بینی ریزش تأکید داشته‌اند (Equihua et al., 2023; Mahdi & Jabbari, 2024)، درحالی‌که پژوهش‌های دیگری بر پیش‌بینی خرید مجدد یا تغییر الگوی رفتار تمرکز کرده‌اند (Chou et al., 2022; Zelenkov & Suchkova, 2023). پژوهش‌های مربوط به بهینه‌سازی پیش‌بینی نیز عمدتاً کاهش ویژگی‌ها، انتخاب متغیرها یا بهبود عملکرد مدل را هدف قرار داده‌اند (Sina Mirabdolbaghi & Amiri, 2022). چنین رویکردهایی ارزشمند هستند، اما در محیط واقعی مدیریت خرده‌فروشی، مدیران به مجموعه‌ای از پاسخ‌های مکمل نیاز دارند: چه مشتریانی ارزشمند هستند، کدام مشتریان در معرض ریزش‌اند، مشتریان ریزشی ارزشمند چه کسانی‌اند، چه کالاهایی را این گروه‌ها خریداری می‌کنند و کدام مدل می‌تواند وضعیت ریزش را با عملکرد مناسب تشخیص دهد.

این نیاز، ضرورت حرکت از «یک مدل پیش‌بینی» به سوی «یک چارچوب یکپارچه تحلیل مشتری» را مطرح می‌کند. چنین چارچوبی باید بتواند داده‌های خام تراکنشی را در یک فرایند منظم آماده کند، شاخص‌های رفتاری معنادار ایجاد نماید، مشتریان را بر اساس ارزش و رفتار بخش‌بندی کند، الگوی سبد خرید گروه‌های مهم را استخراج کند و در نهایت با مقایسه چند الگوریتم، مشتریان ریزشی و غیرریزشی را طبقه‌بندی نماید. این یکپارچگی از آن جهت اهمیت دارد که خروجی یک مرحله می‌تواند ورودی مرحله دیگر باشد؛ برای مثال، خوشه‌بندی می‌تواند ابتدا مشتریان ارزشمند را شناسایی کند، سپس قوانین وابستگی بر سبد خرید همان گروه اجرا شود و در مرحله بعد مدل‌های طبقه‌بندی برای تشخیص ریزش در میان مشتریان دارای اهمیت مدیریتی مقایسه شوند. رویکردهای ترکیبی در پژوهش‌های پیشین ظرفیت چنین ادغامی را تا حدودی نشان داده‌اند، اما هنوز نیاز به چارچوب‌هایی وجود دارد که خوشه‌بندی، پیش‌بینی و تفسیر رفتار خرید را در یک ساختار واحد در محیط خرده‌فروشی ترکیب کنند (Brito et al., 2024; Liu et al., 2022).

همچنین، اعتبار چنین چارچوبی صرفاً با بالا بودن یک شاخص عملکرد اثبات نمی‌شود. اعتبارسنجی باید متناسب با ماهیت هر جزء انجام گیرد: کیفیت خوشه‌بندی باید بر اساس میزان فشردگی و تفکیک خوشه‌ها بررسی شود؛ قوانین وابستگی باید بر اساس شاخص‌هایی مانند

پشتیبان، اطمینان و قدرت ارتباط ارزیابی شوند؛ و مدل‌های طبقه‌بندی نیز باید با چند معیار عملکرد به‌طور هم‌زمان مقایسه شوند. مطالعات روش‌شناختی و مرورهای اخیر بر پیش‌بینی ریزش نیز بر اهمیت فرایند شفاف، قابل بازتولید و مرحله‌بندی‌شده از پیش‌پردازش تا ارزیابی مدل تأکید دارند (Louro et al., 2024). از سوی دیگر، استفاده از چند الگوریتم و مقایسه تجربی آن‌ها نسبت به انتخاب پیشینی یک مدل، می‌تواند اعتبار نتیجه‌گیری درباره مناسب‌ترین روش طبقه‌بندی را افزایش دهد (Mouli et al., 2024). این موضوع به‌خصوص در مجموعه داده‌های خرده‌فروشی با ابعاد زیاد و ویژگی‌های دودویی مرتبط با خرید کالاها اهمیت دارد.

در نهایت، از منظر کاربرد مدیریتی، ارزش اصلی داده‌کاوی زمانی آشکار می‌شود که نتایج آن بتواند به تصمیم‌های مشخص در مدیریت مشتری منجر شود. شناسایی مشتریان ارزشمند غیرریزشی می‌تواند مبنای حفظ و تقویت وفاداری قرار گیرد؛ تشخیص مشتریان ارزشمند ریزشی می‌تواند برنامه‌های بازگشت را هدفمند سازد؛ قوانین سبد خرید می‌توانند برای پیشنهادهای مکمل و فروش متقاطع استفاده شوند؛ و مدل‌های پیش‌بینی می‌توانند در اولویت‌بندی مشتریان برای اقدامات نگهداشت نقش داشته باشند. شواهد موجود در زمینه خرده‌فروشی، تجارت الکترونیکی، مخابرات، بانکداری و خدمات نشان می‌دهد که داده‌کاوی و یادگیری ماشین زمانی بیشترین ارزش را ایجاد می‌کنند که از یک ابزار صرفاً پیش‌بینی‌کننده به یک سامانه پشتیبان تصمیم تبدیل شوند (Adeniran et al., 2024; Chaudhary et al., 2024; Sweidan et al., 2022). بر این اساس، خلأ اصلی مورد توجه پژوهش حاضر، فقدان یک چارچوب منسجم است که ارزش رفتاری مشتری، احتمال ریزش، ساختار سبد خرید و عملکرد الگوریتم‌های مختلف را در سطح یک مطالعه خرده‌فروشی واقعی به‌طور هم‌زمان و به‌صورت یکپارچه بررسی کند.

بنابراین، هدف پژوهش حاضر اعتبارسنجی یک چارچوب یکپارچه برای شناسایی و تحلیل رفتار مشتریان فروشگاه خرده‌فروشی با استفاده از شاخص‌های توسعه‌یافته LRFFM و تلفیق تکنیک‌های خوشه‌بندی، قوانین وابستگی و الگوریتم‌های طبقه‌بندی داده‌کاوی است.

روش پژوهش

پژوهش حاضر از نظر هدف، کاربردی و از نظر رویکرد، کمی و داده‌محور است و با استفاده از مطالعه موردی و تحلیل داده‌های تراکنشی مشتریان یک فروشگاه خرده‌فروشی انجام شد. چارچوب کلی اجرای پژوهش بر اساس فرایند استاندارد CRISP-DM طراحی شد و مراحل فهم کسب‌وکار، فهم داده، آماده‌سازی داده، مدل‌سازی، ارزیابی و به‌کارگیری و تفسیر نتایج را در بر گرفت. انتخاب این چارچوب با توجه به ماهیت مسئله پژوهش صورت گرفت؛ زیرا هدف اصلی، اعتبارسنجی یک چارچوب یکپارچه برای شناسایی و تحلیل رفتار مشتریان از طریق ترکیب شاخص‌های رفتاری LRFFM با روش‌های داده‌کاوی بود. در مرحله فهم کسب‌وکار، مسئله اصلی در قالب شناخت رفتار خرید مشتریان، بخش‌بندی آنان، شناسایی مشتریان ارزشمند و کم‌فعال، تشخیص مشتریان ریزشی و غیرریزشی و استخراج الگوهای موجود در سبد خرید تعریف شد. در مرحله فهم داده نیز ساختار داده‌های تراکنشی، نوع متغیرها، کیفیت اولیه رکوردها و میزان تناسب اطلاعات موجود با اهداف پژوهش بررسی شد. ماهیت فرایند CRISP-DM در این مطالعه تکرارشونده در نظر گرفته شد؛ به این معنا که در صورت مشاهده مشکلات مربوط به کیفیت داده، انتخاب ویژگی‌ها یا عملکرد مدل‌ها، امکان بازگشت به مراحل آماده‌سازی داده و اصلاح فرایند مدل‌سازی وجود داشت.

جامعه مورد بررسی شامل مشتریان فروشگاه خرده‌فروشی بود که سوابق تراکنشی آنان در پایگاه داده فروشگاه ثبت شده بود و واحد اصلی تحلیل در بخش‌های بخش‌بندی و پیش‌بینی، هر مشتری بود. داده‌های اولیه در سطح تراکنش قرار داشتند، اما پس از پاک‌سازی و تجمیع، سوابق خرید متعلق به هر مشتری به یک مجموعه از شاخص‌های رفتاری تبدیل شد تا امکان تحلیل رفتار مشتری در سطح فردی فراهم شود. مشتریانی در مجموعه داده تحلیلی باقی ماندند که دارای شناسه قابل‌ردیابی در پایگاه داده، حداقل اطلاعات لازم برای استخراج شاخص‌های رفتاری و سوابق تراکنشی قابل‌استفاده بودند. رکوردهای فاقد اطلاعات ضروری، تراکنش‌های تکراری یا ناسازگار، داده‌های

E-ISSN: 3041-8933

غیرقابل‌استفاده و مواردی که امکان انتساب معتبر آن‌ها به مشتری یا سبد خرید وجود نداشت، در مرحله آماده‌سازی داده بررسی و حسب مورد اصلاح یا حذف شدند. از آنجا که در محیط خرده‌فروشی رابطه مشتری و فروشگاه ماهیت قراردادی ندارد، وضعیت ریزش بر اساس رفتار مشاهده‌شده در تراکنش‌ها به‌صورت عملیاتی تعریف شد. بر این اساس، مشتریانی که طی ۹۳ روز اخیر هیچ تراکنش خریدی نداشتند، در گروه مشتریان ریزشی و مشتریانی که در این بازه حداقل یک تراکنش خرید داشتند، در گروه مشتریان غیرریزشی قرار گرفتند. این متغیر دوحالتی به‌عنوان متغیر هدف در بخش طبقه‌بندی مورد استفاده قرار گرفت، در حالی که در مرحله خوشه‌بندی وضعیت ریزش در تشکیل خوشه‌ها دخالت داده نشد و تنها پس از تشکیل خوشه‌ها برای تفسیر ویژگی‌های رفتاری گروه‌های به‌دست‌آمده بررسی شد.

ابزار اصلی گردآوری داده در این پژوهش، پایگاه داده تراکنشی فروشگاه خرده‌فروشی و سوابق ثبت‌شده خرید مشتریان بود؛ بنابراین برای گردآوری داده‌ها از پرسشنامه یا ابزار خودگزارشی استفاده نشد و اطلاعات مورد نیاز مستقیماً از داده‌های واقعی خرید استخراج گردید. داده‌های اولیه شامل شناسه مشتری، تاریخ و زمان تراکنش، شناسه یا نام کالا، گروه کالا، تعداد اقلام خریداری‌شده، مبلغ تراکنش و اطلاعات مرتبط با ترکیب سبد خرید بود. این اطلاعات امکان بازسازی سابقه خرید هر مشتری و استخراج متغیرهای مرتبط با رفتار خرید را فراهم ساخت. برای حفظ محرمانگی، اطلاعات هویتی مستقیم مشتریان در فرایند تحلیل مورد استفاده قرار نگرفت و هر مشتری بر اساس شناسه یا کد اختصاصی موجود در مجموعه داده شناسایی شد. در مرحله بررسی اولیه داده، تعداد و نوع متغیرها، قالب مقادیر، دامنه متغیرهای عددی، توزیع داده‌ها، وجود مقادیر مفقود، تراکنش‌های تکراری، ناسازگاری‌های احتمالی و مقادیر غیرعادی بررسی شد و سپس داده‌های معتبر برای تشکیل مجموعه داده نهایی انتخاب شدند.

مبنای اصلی سنجش رفتار مشتریان در پژوهش حاضر، شاخص‌های LRFFM بود که از داده‌های تراکنشی خام استخراج شدند. برای این منظور، اطلاعات مربوط به سابقه زمانی رابطه مشتری با فروشگاه، زمان آخرین خرید، تعداد و فراوانی تراکنش‌ها، فاصله و الگوی تکرار خریدها و ارزش مالی خرید مشتریان در دوره مورد بررسی محاسبه و در سطح هر مشتری تجمیع شد. همچنین متغیرهایی مانند تاریخ نخستین و آخرین خرید، تعداد خریدها، فاصله زمانی بین خریدها، مجموع ارزش مالی تراکنش‌ها و سایر شاخص‌های مورد نیاز برای توصیف رفتار خرید از داده‌های اولیه استخراج شدند. تبدیل داده‌های سطح تراکنش به شاخص‌های سطح مشتری موجب شد که داده‌های خام خرید به مجموعه‌ای از ویژگی‌های رفتاری قابل استفاده در مدل‌های خوشه‌بندی و طبقه‌بندی تبدیل شوند. پیش از اجرای الگوریتم‌ها، مقادیر مفقود و ناسازگار بررسی شد، رکوردهای نامعتبر و تکراری اصلاح یا حذف شدند و متغیرهایی که فاقد ارتباط مستقیم با اهداف تحلیل رفتار مشتریان بودند، از مجموعه داده کنار گذاشته شدند. همچنین ویژگی‌های عددی مورد استفاده در مدل‌هایی که نسبت به تفاوت مقیاس متغیرها حساس بودند، نرمال‌سازی شدند تا تأثیر تفاوت دامنه شاخص‌های LRFFM بر نتایج مدل‌سازی کاهش یابد. برای تحلیل قوانین وابستگی نیز شکل دیگری از مجموعه داده تهیه شد که در آن اقلام متعلق به هر تراکنش یا سبد خرید در کنار یکدیگر قرار گرفتند تا امکان شناسایی روابط و هم‌وقوعی کالاهای خریداری‌شده فراهم شود. به این ترتیب، دو ساختار تحلیلی اصلی از داده‌های خام ایجاد شد: داده‌های تجمیع‌شده در سطح مشتری برای خوشه‌بندی و طبقه‌بندی، و داده‌های سبد خرید در سطح تراکنش برای استخراج قوانین وابستگی.

تحلیل داده‌ها بر اساس منطق یکپارچه CRISP-DM و با ترکیب سه رویکرد اصلی داده‌کاوی شامل خوشه‌بندی، طبقه‌بندی و قوانین وابستگی انجام شد. پس از پاک‌سازی، تبدیل، تجمیع و نرمال‌سازی داده‌ها، شاخص‌های LRFFM به‌عنوان متغیرهای اصلی توصیف‌کننده رفتار مشتریان وارد مرحله مدل‌سازی شدند. در بخش خوشه‌بندی از الگوریتم K-Means برای بخش‌بندی مشتریان استفاده شد. هدف این مرحله، شناسایی گروه‌هایی از مشتریان بود که از نظر الگوی خرید، تازگی خرید، فراوانی تعامل، ارزش مالی و سایر شاخص‌های رفتاری شباهت بیشتری به یکدیگر داشتند. وضعیت ریزشی یا غیرریزشی مشتری در تشکیل خوشه‌ها به‌عنوان متغیر هدف وارد نشد؛ بلکه پس از تشکیل

خوشه‌ها، مشخصات هر خوشه بر اساس مقادیر شاخص‌های LRFFM و میزان حضور مشتریان ریزشی و غیرریزشی مورد تفسیر قرار گرفت. به این ترتیب، امکان شناسایی خوشه‌هایی مانند مشتریان باارزش، مشتریان فعال، مشتریان کم‌فعال و گروه‌های در معرض ریزش فراهم شد. کیفیت راه‌حل خوشه‌بندی با استفاده از شاخص دیویس–بولدین ارزیابی شد. این شاخص میزان فشردگی درون‌خوشه‌ای و جدایی میان خوشه‌ها را به‌طور هم‌زمان در نظر می‌گیرد و مقادیر کمتر آن به‌عنوان نشان‌دهنده ساختار خوشه‌ای مطلوب‌تر در نظر گرفته شد. بررسی پروفایل رفتاری خوشه‌ها نیز برای ارزیابی قابلیت تفسیر و کاربرد مدیریتی بخش‌بندی انجام گرفت.

در بخش طبقه‌بندی، وضعیت ریزش مشتری به‌عنوان متغیر هدف دوحالتی تعریف شد و شاخص‌های رفتاری و تراکنشی مشتریان به‌عنوان متغیرهای پیش‌بین وارد مدل‌ها شدند. برای بررسی استحکام چارچوب پیشنهادی و مقایسه توان الگوریتم‌های مختلف در تشخیص مشتریان ریزشی و غیرریزشی، از روش‌های درخت تصمیم‌گیری، نزدیک‌ترین همسایه، نیوبیز، شبکه عصبی، رگرسیون لجستیک، ماشین بردار پشتیبان، تحلیل تمایزی خطی، آدابوست، بگینگ و استیکینگ استفاده شد. عملکرد مدل‌های طبقه‌بندی بر مبنای ماتریس اغتشاش و شاخص‌های دقت کلی، صحت، یادآوری، ویژگی، معیار F1، نرخ خطا و سطح زیر منحنی ROC ارزیابی و با یکدیگر مقایسه شد. در ماتریس اغتشاش، مشتریان ریزشی که به‌درستی ریزشی تشخیص داده شدند به‌عنوان مثبت واقعی، مشتریان غیرریزشی که به‌درستی غیرریزشی تشخیص داده شدند به‌عنوان منفی واقعی، مشتریان غیرریزشی که به‌اشتباه ریزشی پیش‌بینی شدند به‌عنوان مثبت کاذب و مشتریان ریزشی که به‌اشتباه غیرریزشی تشخیص داده شدند به‌عنوان منفی کاذب در نظر گرفته شدند. مقایسه هم‌زمان معیارهای ارزیابی به این دلیل انجام شد که انتخاب مدل مناسب صرفاً بر اساس دقت کلی صورت نگیرد و توان مدل در شناسایی واقعی مشتریان در معرض ریزش نیز مورد توجه قرار گیرد. مدل دارای عملکرد مناسب‌تر بر اساس مجموعه معیارهای ارزیابی به‌عنوان مدل مطلوب‌تر برای تشخیص وضعیت ریزش مشتریان در چارچوب پیشنهادی در نظر گرفته شد.

در بخش قوانین وابستگی، از الگوریتم Apriori برای استخراج الگوهای تکرارشونده و روابط میان اقلام موجود در سبدهای خرید استفاده شد. داده‌های تراکنشی در این مرحله به‌گونه‌ای سازمان‌دهی شدند که اقلام خریداری‌شده در هر سبد به‌صورت مجموعه‌ای از کالاهای هم‌وقوع قابل تحلیل باشند. قوانین استخراج‌شده به شکل روابط «اگر–آنگاه» تفسیر شدند و برای ارزیابی اهمیت آن‌ها از معیارهای پشتیبان، اطمینان و لیفت استفاده شد. پشتیبان نشان‌دهنده میزان فراوانی وقوع یک مجموعه اقلام یا قانون در کل تراکنش‌ها، اطمینان بیانگر احتمال مشاهده بخش پیامد قانون در صورت وقوع بخش مقدم آن، و لیفت نشان‌دهنده میزان قدرت ارتباط میان مقدم و پیامد در مقایسه با حالت استقلال آن‌ها بود. قوانین دارای مقادیر مناسب این شاخص‌ها برای شناسایی الگوهای معنادار سبد خرید مورد توجه قرار گرفتند و در صورت امکان، الگوهای خرید مشتریان ریزشی و غیرریزشی نیز از نظر نوع اقلام و روابط میان کالاها مقایسه شدند.

اعتبارسنجی چارچوب یکپارچه پژوهش بر اساس عملکرد مکمل سه بخش مدل انجام شد؛ به این معنا که اعتبار بخش بخش‌بندی مشتریان از طریق کیفیت خوشه‌های K-Means و شاخص دیویس–بولدین، اعتبار بخش تشخیص ریزش از طریق عملکرد مدل‌های طبقه‌بندی و معیارهای مبتنی بر ماتریس اغتشاش و ROC، و اعتبار بخش تحلیل الگوهای خرید از طریق مقادیر پشتیبان، اطمینان و لیفت قوانین وابستگی بررسی شد. سپس نتایج سه بخش در کنار شاخص‌های LRFFM تفسیر شدند تا مشخص شود چارچوب پیشنهادی تا چه اندازه قادر است تصویری یکپارچه از ارزش مشتری، الگوی رفتاری، وضعیت فعالیت یا ریزش و ساختار سبد خرید ارائه کند. در مرحله نهایی، خروجی‌های مدل‌ها از منظر قابلیت تبدیل به دانش مدیریتی بررسی شدند تا مشتریان ارزشمند و در معرض ریزش، گروه‌های رفتاری متفاوت و الگوهای خرید قابل‌استفاده برای تصمیم‌گیری در زمینه نگهداشت مشتری، برنامه‌های تشویقی و پیشنهادهای فروش شناسایی شوند. در این پژوهش،

E-ISSN: 3041-8933

خروجی مدل‌ها نقش پشتیبان تصمیم‌گیری داشتند و اجرای مستقیم راهکارهای بازاریابی یا مداخلات نگهداشت مشتری بخشی از فرایند تحلیل داده محسوب نشد.

یافته‌ها

در مرحله فهم داده‌ها بر اساس فرایند استانداردCRISP-DM ، ساختار مجموعه داده تراکنشی فروشگاه خرده‌فروشی، ماهیت رکوردها، متغیرهای موجود و میزان تناسب آن‌ها با اهداف چارچوب پیشنهادی بررسی شد. داده‌های مورد استفاده مربوط به خرید مشتریان فروشگاه سراج شعبه ۱ بود و هر رکورد اولیه بیانگر یک قلم کالا در یک تراکنش خرید بود؛ بنابراین، یک مشتری می‌توانست در دوره مورد بررسی چندین تراکنش و هر تراکنش نیز چندین قلم کالا داشته باشد. مجموعه داده اولیه تقریباً ۵۴۱,۰۰۰ رکورد و ۱۳ ویژگی داشت و اطلاعات مربوط به نام فروشگاه، نام کالا، بارکد کالا، شماره مشتری، فصل، ماه، روز هفته، روز خرید، مبلغ فروش خالص پس از برگشتی، آخرین تاریخ فروش، تعداد روز سپری‌شده از آخرین فروش، مجموع تعداد کالای فروش و تعداد فاکتور فروش را شامل می‌شد. بررسی این ساختار نشان داد که داده‌ها ظرفیت لازم برای شناسایی مشتریان، بازسازی سبد خرید، استخراج شاخص‌های رفتاریLRFFM ، تحلیل ریزش مشتری، خوشه‌بندی، طبقه‌بندی و استخراج قوانین وابستگی را دارند.

جدول ۱

نمونه‌ای از داده‌های تراکنشی فروشگاه خرده‌فروشی

نام فروشگاه	نام کالا	بارکد کالا	شماره مشتری	فصل	ماه	روز هفته	روز	مبلغ فروش خالص پس از برگشتی (ریال)	آخرین تاریخ فروش	روز سپری‌شده از آخرین فروش	تعداد کالا	تعداد فاکتور
سراج شعبه ۱	ساعی روغن سرخ‌کردنی کم‌جذب ۸۱۰ گرمی	۶۲۶۰۲۹۵۱۰۱۷۰۵	۹۰۱۰۲۲۲۰۱۸	تابستان	مرداد	پنج‌شنبه	۲۰	۶۱۳,۰۴۰	۱۴۰۱۰۵۲۰	-	۱	۱
سراج شعبه ۱	سافتلن پودر لباسشویی ماشینی طلایی ۵۰۰ گرمی	۶۲۶۰۰۱۰۵۰۰۳۹۴	۹۰۱۰۲۲۲۰۱۸	تابستان	مرداد	پنج‌شنبه	۲۰	۱۵۳,۵۴۶	۱۴۰۱۰۵۲۰	-	۱	۱
سراج شعبه ۱	پگاه کره ۲۵ گرمی	۶۲۶۰۷۰۷۴۰۰۵۸۷	۹۰۱۰۲۲۲۰۱۸	تابستان	مرداد	پنج‌شنبه	۲۰	۵۹,۰۳۷	۱۴۰۱۰۵۲۰	-	۱	۱
سراج شعبه ۱	شون شامپو فروتی میکس ml۴۰۰	۶۲۶۰۴۸۲۵۲۶۳۷۳	۹۰۱۰۲۲۲۰۱۸	تابستان	مرداد	پنج‌شنبه	۲۰	۴۳۶,۰۵۵	۱۴۰۱۰۵۲۰	-	۱	۱
سراج شعبه ۱	کملیون باتری معمولی قلمی ۲ عددی	۴۲۶۰۰۳۳۱۵۱۰۳۲	۹۰۱۰۲۲۲۰۱۸	تابستان	مرداد	پنج‌شنبه	۲۰	۹۰,۸۲۶	۱۴۰۱۰۵۲۰	-	۱	۱
سراج شعبه ۱	سیگنال خمیردندان ضدپوسیدگی M۱۰۰	۶۲۶۰۵۲۶۴۱۸۱۷۶	۹۰۱۰۲۲۲۰۱۸	تابستان	مرداد	پنج‌شنبه	۲۰	۲۳۴,۸۶۲	۱۴۰۱۰۵۲۰	-	۱	۱
سراج شعبه ۱	کاله پنیر یو اف سفید ۴۰۰ گرم	۶۲۶۰۱۶۱۵۲۹۲۲۰	۹۰۱۰۲۲۲۰۱۸	تابستان	مرداد	پنج‌شنبه	۲۰	۲۸۸,۰۰۰	۱۴۰۱۰۵۲۰	-	۱	۱

تمرکز تحلیل بر یک شعبه مشخص موجب شد رفتار خرید مشتریان در یک محیط فروشگاهی نسبتاً همگن بررسی شود. در عین حال، ساختار چارچوب از نظر فنی قابلیت اجرا در سایر شعب را دارد، هرچند تعمیم مستقیم مقادیر عددی، ساختار خوشه‌ها و الگوهای خرید به سایر شعب بدون اجرای مجدد مدل مناسب نیست؛ زیرا عواملی نظیر موقعیت مکانی، ترکیب جمعیت مشتریان، نوع تقاضا و شرایط عملیاتی هر شعبه می‌توانند ساختار رفتار خرید را تغییر دهند.

بررسی ویژگی‌ها نشان داد که «نام فروشگاه» به دلیل ثابت بودن مقدار آن در تمام رکوردهای شعبه فاقد قدرت تفکیک‌کنندگی بود و از تحلیل‌های بعدی حذف شد. «شماره مشتری» به‌عنوان شناسه اصلی برای اتصال تراکنش‌ها و تجمیع رکوردها در سطح مشتری به کار رفت، اما خود به‌عنوان متغیر توضیحی وارد مدل نشد. «بارکد کالا» شناسه پایدار هر قلم کالا بود و همراه با «نام کالا» نقش اصلی را در ایجاد ماتریس تراکنش–کالا و استخراج قوانین وابستگی داشت. نام کالا متغیری اسمی بود که برای تفسیر سبد خرید مناسب بود، ولی به‌صورت مستقیم در مدل‌های عددی خوشه‌بندی و طبقه‌بندی استفاده نشد.

متغیرهای فصل، ماه، روز هفته و روز خرید اطلاعات زمانی تراکنش‌ها را فراهم کردند. این متغیرها برای شناخت الگوهای زمانی و بازسازی زمان خرید به کار رفتند. در مقابل، شاخص Recency بر اساس فاصله از آخرین خرید محاسبه شد و صرفاً متکی بر نام روز هفته نبود. «مبلغ فروش خالص پس از برگشتی» مبنای استخراج شاخص Monetary بود و «مجموع تعداد کالای فروش» برای سنجش حجم اقلام خریداری‌شده استفاده شد. ویژگی «تعداد فاکتور فروش» در سطح خام در بسیاری از رکوردها تنوع اندکی داشت؛ بنابراین، فراوانی واقعی خرید هر مشتری در مرحله تجمیع با شمارش تراکنش‌ها یا رکوردهای مرتبط با مشتری استخراج شد.

مدل توسعه‌یافته LRFFM در این پژوهش پنج بعد رفتاری را پوشش داد .R بیانگر Recency یا فاصله از آخرین خرید بود .F اول تعداد دفعات خرید یا تعداد تراکنش‌های مشتری را نشان داد .F دوم تعداد کل کالاهای خریداری‌شده را اندازه‌گیری کرد .M مجموع ارزش مالی خریدهای مشتری و L طول دوره تعامل او با فروشگاه را نشان داد. تمایز میان دو شاخص F اهمیت داشت؛ زیرا ممکن بود مشتری تعداد مراجعه کمی داشته، اما در هر مراجعه حجم زیادی کالا خریداری کرده باشد.

جدول ۲

متغیرهای نهایی و نقش آن‌ها در چارچوب تحلیلی

متغیر	نقش در تحلیل
شماره مشتری	شناسه تجمیع تراکنش‌ها در سطح مشتری
بارکد و نام کالا	تشکیل سبد خرید و اجرای قوانین وابستگی
RecencyیاR	فاصله زمانی از آخرین خرید و مبنای تعیین ریزش
Fاول	تعداد تراکنش‌ها یا دفعات خرید
Fدوم	مجموع تعداد اقلام خریداری‌شده
MonetaryیاM	مجموع مبلغ خرید مشتری
LengthیاL	طول دوره تعامل مشتری با فروشگاه
Churn	متغیر هدف دوحالتی برای طبقه‌بندی

E-ISSN: 3041-8933

بر اساس این ساختار، هرچه R کمتر بود، خرید مشتری اخیرتر محسوب می‌شد؛ درحالی‌که مقادیر بالاتر F اول، Fدوم، Mو L عموماً نشان‌دهنده فعالیت، حجم خرید، ارزش مالی و سابقه تعامل بیشتر مشتری بودند. وضعیت ریزش نیز به‌صورت عملیاتی با آستانه ۹۳ روز تعریف شد؛ به این معنا که مشتری دارای Recency بیش از ۹۳ روز، ریزشی و سایر مشتریان غیرریزشی در نظر گرفته شدند.

در مرحله پیش‌پردازش، نخست مقادیر مفقود در متغیرهای کلیدی بررسی شدند. رکوردهای فاقد شماره مشتری امکان انتساب به فرد مشخص را نداشتند و رکوردهای فاقد بارکد معتبر نیز برای تحلیل سبد خرید قابل‌اتکا نبودند؛ ازاین‌رو، مواردی که امکان اصلاح معتبر آن‌ها وجود نداشت حذف شدند. همچنین، مبلغ فروش خالص پس از برگشتی بررسی شد و رکوردهای دارای مقادیر منفی که عمدتاً می‌توانستند بیانگر برگشت کالا یا اصلاح تراکنش باشند، با توجه به فراوانی اندک از تحلیل حذف شدند. مقادیر صفر یا نامناسب نیز بررسی و در موارد غیرقابل‌استفاده اصلاح یا حذف شدند.

پس از حذف داده‌های ناقص و نامعتبر، تعداد رکوردهای قابل‌استفاده به ۵۴۱,۰۱۳ رکورد رسید و ۶ ویژگی اصلی برای ادامه عملیات حفظ شد. مقدار اسمی ماه نیز به مقدار عددی متناظر تبدیل شد؛ برای مثال، فروردین با ۱ کدگذاری شد. این تبدیل صرفاً برای پردازش محاسباتی انجام گرفت و معنای زمانی متغیر ماه حفظ شد.

برای تبدیل داده‌های تراکنشی به سطح مشتری از عملیات Pivot استفاده شد. شماره مشتری در سطح ردیف قرار گرفت و متغیرهای عددی شامل تعداد رکوردها، روز، مبلغ فروش خالص و مجموع تعداد کالای فروش تجمیع شدند. به این ترتیب، هر مشتری به یک رکورد تبدیل شد و شاخص‌های رفتاری در سطح فرد قابل محاسبه شدند. خروجی Summary به‌جای Full Data استخراج شد تا داده‌های نهایی در سطح مشتری باقی بمانند.

در محاسبهRecency ، بیشترین ماه و بیشترین روز خرید هر مشتری استخراج شد و با در نظر گرفتن مقدار مرجع ۱۸۶، شاخص Rاز رابطه زیر محاسبه شد:

[

R = ۱۸۶ - ((MAX(Mah)-۱)\times۳۱)-MAX(Rooz)

]

برای تعیین طول رابطه مشتری با فروشگاه، ابتدا فاصله اولین خرید تا زمان مرجع با رابطه زیر محاسبه شد:

[

lastly = ۱۸۶ - ((MIN(Mah)-۱)\times۳۱)-MIN(Rooz)

]

سپس شاخص Length از اختلاف فاصله اولین خرید و Recency به دست آمد:

[

Long = lastly - Recency

]

در نهایت، متغیر ریزش به‌صورت زیر تعریف شد:

[

Churn=

\begin{cases}
Churn & Recency>۹۳\
No_Churn & Recency\leq۹۳
\end{cases}
]

در نتیجه این مرحله، مجموعه داده سطح مشتری شامل شاخص‌های LRFFM و متغیر Churn ایجاد و به قالب اکسل صادر شد. برای قوانین وابستگی نیز ساختار داده مجدداً به ماتریس تراکنش–کالا تبدیل شد. در این ساختار، خانه خالی به معنای عدم خرید کالا بود و با صفر جایگزین شد، درحالی‌که مقدار یک خرید کالا را نشان می‌داد. این عملیات با fillna(۰) انجام شد. در بخش طبقه‌بندی نیز ویژگی‌هایی که داده کافی نداشتند، با آستانه حداقل ۱۰ مشاهده معتبر حذف شدند؛ این کار معادل استفاده از dropna(thresh=۱۰, axis=۱) بود.

خوشه‌بندی مشتریان با الگوریتم K-Means و بر اساس شاخص‌های LRFFM انجام شد و متغیر Churn در ساخت خوشه‌ها وارد نشد. در یک تحلیل مقدماتی مرتبط با آماده‌سازی گروه‌های مورد استفاده در قوانین وابستگی، مدل دوخوشه‌ای نیز بررسی شد. در آن اجرا، شاخص دیویس–بولدین ۰.۵۴۸- گزارش شد و ۲,۰۰۱ مشتری در خوشه اول و ۲۴,۴۲۷ مشتری در خوشه دوم قرار گرفتند. میانگین‌های خوشه اول برای R ، Fاول، Fدوم، Mو L به‌ترتیب ۲۷.۷۴۴، ۸۲.۹۹۶، ۱۳۳.۹۸۶، ۴۳,۲۹۹,۷۱۷.۱۹۲ و ۱۱۴.۳۸۳ بود، درحالی‌که مقادیر متناظر در خوشه دوم ۸۵.۰۰۶، ۱۵.۳۴۹، ۲۳.۲۳۴، ۶,۱۹۶,۸۱۴.۱۱۰ و ۲۹.۹۵۰ به دست آمد. بنابراین، خوشه اول در اجرای دوخوشه‌ای به دلیل Recency کمتر و مقادیر بیشتر سایر شاخص‌ها، گروه ارزشمندتر محسوب شد.

بااین‌حال، برای تعیین ساختار نهایی خوشه‌بندی کل مشتریان، تعداد خوشه‌ها از ۲ تا ۱۰ بررسی شد. مقدار گزارش‌شده شاخص دیویس–بولدین برای سه خوشه ۰.۵۴۹- بود و از نظر معیار مورد استفاده، بهترین مقدار در میان گزینه‌های بررسی‌شده محسوب شد.

جدول ۳

مقایسه شاخص دیویس–بولدین برای تعداد مختلف خوشه‌ها

تعداد خوشه	شاخص دیویس–بولدین
۲	-۰.۵۴۸
۳	-۰.۵۴۹
۴	-۰.۴۳۳
۵	-۰.۴۴۸
۶	-۰.۴۸۴
۷	-۰.۴۴۷
۸	-۰.۴۵۵
۹	-۰.۴۶۹
۱۰	-۰.۴۷۹

بر این اساس، مدل نهایی با سه خوشه اجرا شد. تعداد مشتریان در خوشه‌های ۰، ۱ و ۲ به‌ترتیب ۲۲,۴۲۷، ۲۵۶ و ۳,۷۴۵ نفر بود که جمع آن‌ها ۲۶,۴۲۸ مشتری را تشکیل داد. در یکی از گزارش‌های مقدماتی مقدار ۳,۴۷۵ برای خوشه سوم ثبت شده بود، اما با توجه به مجموع کل ۲۶,۴۲۸ مشتری و جدول نهایی خوشه‌بندی، مقدار سازگار ۳,۷۴۵ است.

E-ISSN: 3041-8933

جدول ۴

مشخصات رفتاری و مالی خوشه‌های نهایی مشتریان

شاخص	خوشه ۰	خوشه ۱	خوشه ۲
تعداد مشتری	۲۲,۴۲۷	۲۵۶	۳,۷۴۵
نسبت ریزشی	۰.۵۳	۰.۰۷	۰.۱۵
نسبت غیرریزشی	۰.۴۷	۰.۹۳	۰.۸۵
تعداد تراکنش	۱۲.۶۹	۱۴۴.۳۵	۵۸.۶۰
مبلغ خرید	۴,۹۸۲,۴۷۸.۸۵	۹۵,۴۵۵,۰۳۰.۹۰	۲۷,۱۹۱,۹۲۸.۲۶
تعداد کالا	۱۹.۱۱	۲۶۰.۴۹	۹۰.۹۱
Recency	۸۸.۵۱	۱۹.۲۹	۳۷.۸۹
Length	۲۱.۸۹	۱۳۱.۳۲	۹۶.۸۳

نتایج نشان داد مشتریان خوشه ۱، با وجود تعداد کم‌تر، بیشترین تعداد تراکنش، بیشترین تعداد کالا، بالاترین مبلغ خرید، کمترین Recencyو بیشترین Length را داشتند. همچنین ۹۳ درصد مشتریان این خوشه غیرریزشی بودند. در نتیجه، این خوشه به‌عنوان گروه ارزشمندتر شناسایی شد. خوشه ۲ در مرتبه بعد قرار داشت و خوشه ۰ از نظر فعالیت و ارزش مالی ضعیف‌ترین وضعیت را نشان داد.

در تحلیل‌های بعدی که صرفاً روی مشتریان دارای اطلاعات کامل سبد خرید و قابل‌استفاده در طبقه‌بندی انجام شد، زیرمجموعه ارزشمند قابل‌تحلیل شامل ۱,۹۹۱ مشتری بود. این کاهش تعداد ناشی از اعمال فیلترهای لازم برای موجود بودن اطلاعات کالا و تشکیل ماتریس خرید بود. در این مجموعه، ۱۷۰ مشتری در کلاس ریزشی قرار داشتند.

به‌منظور شناسایی مشتریانی که اگرچه ریزش کرده‌اند اما در گذشته ارزش رفتاری و مالی بالایی داشته‌اند، ۱۲,۴۳۳ مشتری ریزشی به‌صورت مستقل خوشه‌بندی شدند. مقادیر دیویس–بولدین برای دو تا ده خوشه به‌ترتیب ۰.۵۹۰-، ۰.۳۷۹-، ۰.۳۹۰-، ۰.۴۲۰-، ۰.۴۳۲-، ۰.۴۴۶-، ۰.۴۶۹-، ۰.۴۷۶- و ۰.۴۸۳- بود. بر این اساس، ساختار دوخوشه‌ای انتخاب شد. ۸۸۰ مشتری در خوشه ارزشمندتر و ۱۱,۵۵۳ مشتری در خوشه دوم قرار گرفتند.

جدول ۵

مقایسه شاخص‌های LRFFM *در خوشه‌های مشتریان ریزشی*

شاخص	خوشه ارزشمند ریزشی	خوشه دوم ریزشی
تعداد مشتری	۸۸۰	۱۱,۵۵۳
تعداد تراکنش	۴۲.۸۳	۱۰.۷۳
مبلغ خرید	۲۲,۵۶۶,۴۶۴.۸۵	۳,۷۴۵,۸۵۹.۹۸
تعداد کالا	۷۰.۱۵	۱۵.۹۸
Recency	۱۲۶.۴۰	۱۳۶.۷۳
Length	۲۷.۳۵	۶.۲۴

خوشه اول مشتریان ریزشی از نظر تعداد تراکنش، تعداد کالا، مبلغ خرید و طول رابطه با فروشگاه به‌طور محسوسی بهتر بود و Recencyآن نیز اندکی پایین‌تر از خوشه دوم قرار داشت. ازاین‌رو، این گروه به‌عنوان «مشتریان ارزشمند ریزشی» شناسایی شد و از نظر برنامه‌های بازگشت مشتری اهمیت بیشتری داشت.

برای تحلیل سبد خرید، داده‌های خوشه‌بندی‌شده بر اساس شماره مشتری با داده‌های بارکد کالا به روش Inner Join ترکیب شدند. بدین ترتیب، برای هر مشتری علاوه بر شاخص‌هایLRFFM ، خوشه، وضعیت ریزش و اقلام خریداری‌شده نیز در دسترس قرار گرفت. سپس داده‌ها به ماتریس صفر و یک تبدیل شدند و الگوریتم FP-Growth برای شناسایی مجموعه اقلام پرتکرار و استخراج قوانین وابستگی اجرا شد. در تنظیمات اولیه، حداقل پشتیبان ۰.۵ و حداقل اطمینان ۰.۵ در نظر گرفته شد. بااین‌حال، خروجی‌های ثبت‌شده نهایی شامل قوانین و مجموعه اقلامی با پشتیبان کمتر نیز بود و تحلیل نتایج بر اساس همان مقادیر خروجی ثبت‌شده انجام شد.

در مشتریان ارزشمند غیرریزشی، مجموعه چهارقلمی ۲۲۰۰۱۷۰، ۲۲۰۳۳۱۴، ۲۲۰۲۳۱۲ و ۲۲۰۴۳۵۹ دارای پشتیبان ۰.۱۲۱ بود و ترکیب ۲۲۰۰۱۷۰، ۲۲۰۳۳۱۴، ۲۲۰۲۳۱۲ و ۲۲۰۴۳۶۴ پشتیبان ۰.۱۲۷ داشت. مجموعه‌های سه‌قلمی ۲۲۰۰۱۷۰–۲۲۰۳۳۱۴–۲۲۰۲۳۱۲، ۲۲۰۰۱۷۰–۲۲۰۳۳۱۴–۲۲۰۴۳۵۹، ۲۲۰۰۱۷۰–۲۲۰۳۳۱۴–۲۲۰۴۳۶۴، ۲۲۰۰۱۷۰–۲۲۰۲۳۱۲–۲۲۰۴۳۵۹، ۲۲۰۰۱۷۰–۲۲۰۲۳۱۲–۲۲۰۴۳۶۴، ۲۲۰۰۱۷۰–۲۲۰۳۳۱۴–۲۲۰۴۳۵۹ و ۲۲۰۰۱۷۰–۲۲۰۳۳۱۴–۲۲۰۲۳۱۲–۲۲۰۴۳۶۴ به‌ترتیب پشتیبان ۰.۱۳۹، ۰.۱۲۱، ۰.۱۲۸، ۰.۱۲۱، ۰.۱۲۷، ۰.۱۲۲ و ۰.۱۲۸ داشتند.

مجموعه‌های دوقلمی مهم این گروه شامل ۲۲۰۱۰۳۵–۶۲۶۰۰۵۵۸۴۳۰۳۶ با پشتیبان ۰.۱۱۷، ۲۲۰۰۱۷۰–۲۲۰۳۳۱۴ با ۰.۱۴۰، ۲۲۰۰۱۷۰–۲۲۰۲۳۱۲ با ۰.۱۳۹، ۲۲۰۰۱۷۰–۲۲۰۴۳۵۹ با ۰.۱۲۳، ۲۲۰۰۱۷۰–۲۲۰۴۳۶۴ با ۰.۱۲۸، ۲۲۰۳۳۱۴–۲۲۰۲۳۱۲ با ۰.۱۴۱، ۲۲۰۳۳۱۴–۲۲۰۴۳۵۹ با ۰.۱۲۵، ۲۲۰۳۳۱۴–۲۲۰۴۳۶۴ با ۰.۱۳۰، ۲۲۰۲۳۱۲–۲۲۰۴۳۵۹ با ۰.۱۲۴ و ۲۲۰۲۳۱۲–۲۲۰۴۳۶۴ با ۰.۱۲۸ بود.

در میان اقلام منفرد مشتریان ارزشمند غیرریزشی، بیشترین پشتیبان متعلق به ۲۲۰۱۰۳۵ با ۰.۳۳۲ و پس از آن ۶۲۶۰۰۵۵۸۴۳۰۳۶ با ۰.۲۹۵ بود. سایر اقلام پرتکرار شامل ۶۲۶۱۹۰۶۱۰۱۱۷۶ با ۰.۲۲۰، ۲۲۰۸۱۲۱ با ۰.۱۹۴، ۲۲۰۶۸۶۱ با ۰.۱۹۲، ۶۲۶۰۰۱۰۵۰۰۳۹۴ با ۰.۱۹۰، ۶۲۶۷۶۰۴۴۱۴۶۹۳ با ۰.۱۸۶، ۱۶۴۲۰۰۲۷۵۳۰۰۰۴۴ و ۲۲۰۱۹۴۶ هر یک با ۰.۱۷۶، ۱۶۴۲۰۰۲۷۵۳۰۰۰۰۱ با ۰.۱۷۴، ۶۲۶۰۰۴۳۹۰۰۰۱۷ با ۰.۱۷۱، ۶۲۶۲۹۱۶۶۰۰۷۴۱ با ۰.۱۶۰، ۲۲۰۰۱۳۴ با ۰.۱۵۶، ۶۲۶۰۰۱۶۳۴۲۰۱۱ با ۰.۱۵۳، ۲۲۰۰۱۷۰ و ۲۲۰۰۹۱۷ هر یک با ۰.۱۵۲، ۲۲۰۳۳۱۴ با ۰.۱۵۱، ۶۲۶۵۴۹۹۵۰۰۰۴۰، ۱۶۴۲۰۰۲۷۵۳۰۰۰۲ و ۲۲۰۱۱۵۷ هر یک با ۰.۱۴۹، ۶۲۶۹۵۲۳۹۰۰۰۵۹ و ۲۲۰۲۳۱۲ هر یک با ۰.۱۴۷، ۶۲۶۰۴۵۹۸۰۰۰۷۹ با ۰.۱۴۶، ۶۲۶۱۸۴۷۵۰۲۲۰۹ و ۶۲۶۱۰۶۱۱۰۰۵۳۳ هر یک با ۰.۱۴۲، ۶۲۶۰۰۵۳۱۸۶۲۴۱، ۲۲۰۴۳۵۹ و ۶۲۶۰۱۷۶۸۰۲۹۲۹ هر یک با ۰.۱۴۱، ۲۲۰۱۰۰۰، ۶۲۶۳۸۱۲۸۰۱۲۴۹ و ۱۶۴۲۰۰۲۰۳۲۵۰۲۳۲ هر یک با ۰.۱۳۸ و ۲۲۰۴۳۶۴ با ۰.۱۳۶ بود. پشتیبان سایر اقلام منفرد گزارش‌شده بین ۰.۱۱۵ و ۰.۱۳۱ قرار داشت و شامل ۶۲۶۰۱۰۵۰۰۵۱۵۵، ۲۲۰۱۶۵۵، ۲۲۰۰۸۹۳، ۶۲۶۱۹۸۵۲۰۰۰۵۰، ۶۲۶۰۱۶۱۵۲۹۲۷۵، ۶۲۶۰۵۳۲۸۲۲۰۹۷، ۲۲۰۱۶۹۱، ۶۲۶۲۱۵۷۹۰۰۱۰۵، ۶۲۶۰۰۵۳۱۸۳۹۷۴، ۶۲۶۰۰۵۳۱۸۳۹۹۸، ۶۲۶۰۶۶۷۳۰۲۵۳۲، ۶۲۶۰۷۰۷۴۰۰۵۸۷، ۲۲۰۱۰۱۴، ۶۲۶۳۸۱۲۸۰۱۳۸۶، ۶۲۶۰۶۶۷۳۰۲۵۲۵، ۲۲۰۱۰۱۵، ۶۲۶۸۱۸۷۱۰۲۱۱۳، ۶۲۶۵۵۴۷۶۲۰۲۸۷، ۲۲۰۱۰۱۶ و ۲۲۰۱۶۵۸ بود.

در مجموعه دیگری از تحلیل مشتریان ارزشمند غیرریزشی، الگوهای بزرگ‌تر ۷ تا ۹ قلمی نیز مشاهده شدند. هسته اصلی این مجموعه‌ها از کدهای ۲۲۰۴۳۵۹، ۲۲۰۳۳۱۴، ۲۲۰۰۱۷۰، ۲۲۰۲۳۱۲، ۲۲۰۴۳۶۴، ۲۲۰۱۶۵۵، ۲۲۰۱۶۹۱، ۲۲۰۱۶۵۸ و ۲۲۰۰۴۹۵ تشکیل

E-ISSN: 3041-8933

می‌شد. پشتیبان مجموعه‌های هفت‌قلمی عمدتاً ۰.۱۸۱ یا ۰.۱۸۶ بود، مجموعه‌های هشت‌قلمی نیز پشتیبان ۰.۱۸۱ یا ۰.۱۸۶ داشتند و مجموعه کامل نه‌قلمی شامل هر ۹ کد فوق پشتیبان ۰.۱۸۱ را نشان داد.

در مشتریان ارزشمند ریزشی، مجموعه سه‌قلمی ۲۲۰۱۶۰۹–۲۲۰۰۳۵۱–۲۲۰۰۳۴۸ دارای پشتیبان ۰.۰۷۶ بود. مجموعه‌های دوقلمی شامل ۶۲۶۱۹۰۶۱۰۱۱۷۶–۱۶۴۲۰۰۲۷۵۳۰۰۰۱ با ۰.۰۹۴، ۶۲۶۰۱۷۶۸۰۲۹۲۹–۶۲۶۱۹۰۶۱۰۱۱۷۶ با ۰.۰۷۱، ۱۶۴۲۰۰۲۷۵۳۰۰۰۱–۲۲۰۱۹۴۶ با ۰.۰۸۲، ۱۶۴۲۰۰۲۷۵۳۰۰۰۴۴–۱۶۴۲۰۰۲۷۵۳۰۰۰۱ با ۰.۰۷۱، ۶۲۶۱۱۰۶۰۱۰۰۵۶–۲۲۰۱۹۴۶ با ۰.۰۷۱، ۲۲۰۰۳۵۱–۲۲۰۱۶۰۹ با ۰.۰۹۴، ۲۲۰۰۳۴۸–۲۲۰۱۶۰۹ با ۰.۰۷۶، ۲۲۰۱۶۱۸–۲۲۰۱۶۰۹ با ۰.۰۷۱، ۲۲۰۱۶۱۱–۲۲۰۱۶۰۹ با ۰.۰۷۶، ۲۲۰۰۳۴۸–۲۲۰۰۳۵۱ با ۰.۰۹۴، ۲۲۰۱۶۱۸–۲۲۰۰۳۵۱ با ۰.۰۷۶ و ۲۲۰۱۶۱۱–۲۲۰۰۳۵۱ با ۰.۰۷۱ بودند.

در تحلیل دیگری بر روی خوشه ارزشمند ریزشی، مهم‌ترین مجموعه‌های دوقلمی شامل ۶۲۶۱۹۰۶۱۰۱۱۷۶–۱۶۴۲۰۰۲۷۵۳۰۰۰۱ با پشتیبان ۰.۰۸، ۲۲۰۱۰۳۵–۶۲۶۱۹۰۶۱۰۱۱۷۶ با ۰.۰۵، ۲۲۰۱۹۴۶–۱۶۴۲۰۰۲۷۵۳۰۰۰۱ با ۰.۰۵، ۲۲۰۱۶۱۸–۲۲۰۰۳۵۱ با ۰.۰۵، ۲۲۰۱۶۱۱–۲۲۰۰۳۵۱ با ۰.۰۵، ۲۲۰۰۳۴۸–۲۲۰۰۳۵۱ با ۰.۰۶، ۲۲۰۱۶۱۸–۲۲۰۱۶۱۱ با ۰.۰۵ و ۲۲۰۱۶۱۱–۲۲۰۰۳۴۸ با ۰.۰۵ بود.

در میان کالاهای منفرد مشتریان ارزشمند ریزشی، بیشترین پشتیبان به کد ۶۲۶۱۹۰۶۱۰۱۱۷۶ با ۰.۲۸۲ تعلق داشت و پس از آن ۱۶۴۲۰۰۲۷۵۳۰۰۰۱ با ۰.۲۲۴، ۶۲۶۸۱۸۷۱۰۲۱۱۳ و ۶۲۶۱۱۰۶۰۱۰۰۵۶ هر یک با ۰.۱۸۲، ۶۲۶۰۴۵۹۸۰۰۰۷۹ با ۰.۱۷۶، ۲۲۰۱۹۴۶ با ۰.۱۶۵، ۶۲۶۹۵۲۳۹۰۰۰۵۹ و ۱۶۸۱۰۰۱۹۰۶۱۰۰۰۵ هر یک با ۰.۱۵۹، ۶۲۶۰۱۷۶۸۰۲۹۲۹ و ۱۶۴۲۰۰۲۷۵۳۰۰۰۴۴ هر یک با ۰.۱۵۳، ۶۲۶۰۰۵۵۸۴۳۰۳۶، ۶۲۶۰۰۱۶۳۴۲۰۱۱ و ۲۲۰۰۹۱۷ هر یک با ۰.۱۴۷، ۶۲۶۵۴۹۹۵۰۰۰۴۰ و ۶۲۶۰۵۷۳۱۰۱۲۸۱ هر یک با ۰.۱۴۱ و ۶۲۶۰۰۱۰۵۰۰۳۹۴ و ۲۲۰۸۰۹۴ هر یک با ۰.۱۳۵ قرار داشتند. سایر اقلام منفرد این گروه پشتیبان‌هایی در دامنه تقریبی ۰.۰۷۱ تا ۰.۱۲۹ نشان دادند.

در تحلیل نهایی‌تر مشتریان ارزشمند ریزشی، کالاهای منفرد با بالاترین پشتیبان شامل ۶۲۶۱۹۰۶۱۰۱۱۷۶ با ۰.۲۲۹، ۱۶۴۲۰۰۲۷۵۳۰۰۰۱ با ۰.۱۸۷، ۶۲۶۰۴۵۹۸۰۰۰۷۹ و ۲۲۰۱۹۴۶ هر یک با ۰.۱۴۸، ۶۲۶۰۰۵۵۸۴۳۰۳۶ با ۰.۱۴۰، ۶۲۶۹۵۲۳۹۰۰۰۵۹ با ۰.۱۳۸، ۶۲۶۱۱۰۶۰۱۰۰۵۶ با ۰.۱۳۷، ۶۲۶۰۰۱۶۳۴۲۰۱۱ با ۰.۱۱۶، ۶۲۶۰۰۱۰۵۰۰۳۹۴ با ۰.۱۱۴، ۶۲۶۸۱۸۷۱۰۲۱۱۳ و ۶۲۶۵۴۹۹۵۰۰۰۴۰ هر یک با ۰.۱۱۰، ۲۲۰۱۰۳۵ با ۰.۱۰۷، ۶۲۶۱۸۴۷۵۰۲۲۰۹ با ۰.۱۰۶، ۶۲۶۲۷۵۳۰۰۰۷۷۳ با ۰.۱۰۱ و ۶۲۶۰۰۴۳۹۰۰۰۱۷ با ۰.۱۰۰ بود. پشتیبان ۲۲۰۰۳۵۱ برابر ۰.۰۸۸، ۲۲۰۱۶۱۸ و ۲۲۰۱۶۱۱ هر یک ۰.۰۸۳، ۲۲۰۰۳۴۸ برابر ۰.۰۸۱ و ۲۲۰۱۶۰۹ برابر ۰.۰۶۴ بود.

جدول ۶

نمونه مهم‌ترین الگوها و قوانین وابستگی در دو گروه مشتریان

گروه	مقدم قانون یا مجموعه اقلام	نتیجه	پشتیبان	اطمینان
ارزشمند غیرریزشی	۲۲۰۲۳۱۲ + ۲۲۰۳۳۱۴	۲۲۰۰۱۷۰	۰.۱۴	۰.۹۹
ارزشمند غیرریزشی	۲۲۰۲۳۱۲ + ۲۲۰۴۳۶۴	۲۲۰۰۱۷۰	۰.۱۳	۱.۰۰
ارزشمند غیرریزشی	۲۲۰۲۳۱۲ + ۲۲۰۳۳۱۴ + ۲۲۰۴۳۶۴	۲۲۰۰۱۷۰	۰.۱۳	۱.۰۰
ارزشمند غیرریزشی	۲۲۰۳۳۱۴ + ۲۲۰۲۳۱۲	۲۲۰۴۳۶۴	۰.۱۳	۰.۹۱
ارزشمند غیرریزشی	۲۲۰۰۱۷۰ + ۲۲۰۳۳۱۴ + ۲۲۰۲۳۱۲	۲۲۰۴۳۶۴	۰.۱۳	۰.۹۲
ارزشمند غیرریزشی	۲۲۰۳۳۱۴	۲۲۰۴۳۵۹	۰.۲۳	۰.۹۰

ارزشمند غیرریزشی	۲۲۰۲۳۱۲	۲۲۰۴۳۵۹	۰.۲۳	۰.۹۳
ارزشمند ریزشی	۲۲۰۱۶۱۸	۲۲۰۱۶۰۹	۰.۰۷	۰.۷۱
ارزشمند ریزشی	۲۲۰۰۳۵۱	۲۲۰۱۶۰۹	۰.۰۹	۰.۷۶
ارزشمند ریزشی	۲۲۰۰۳۴۸ + ۲۲۰۰۳۵۱	۲۲۰۱۶۰۹	۰.۰۸	۰.۸۱
ارزشمند ریزشی	۲۲۰۰۳۴۸	۲۲۰۰۳۵۱	۰.۰۹	۰.۸۰
ارزشمند ریزشی	۲۲۰۰۳۴۸ + ۲۲۰۱۶۰۹	۲۲۰۰۳۵۱	۰.۰۸	۱.۰۰
ارزشمند ریزشی	۲۲۰۱۶۱۸	۲۲۰۰۳۵۱	۰.۰۵	۰.۶۲
ارزشمند ریزشی	۲۲۰۰۳۴۸	۲۲۰۰۳۵۱	۰.۰۵	۰.۶۸

قوانین کامل‌تر مشتریان ارزشمند غیرریزشی نشان دادند که ارتباط بین کدهای ۲۲۰۰۱۷۰، ۲۲۰۳۳۱۴، ۲۲۰۲۳۱۲، ۲۲۰۴۳۵۹ و ۲۲۰۴۳۶۴ بسیار قوی است. اطمینان قوانین در این شبکه از حدود ۰.۷۹۴ آغاز شد و در چند رابطه به ۱.۰۰۰ رسید. برای نمونه، قانون ۲۲۰۰۱۷۰ ⟶ ۲۲۰۳۳۱۴ دارای اطمینان ۰.۹۲۱، قانون ۲۲۰۳۳۱۴ ⟶ ۲۲۰۰۱۷۰ دارای ۰.۹۲۷، قانون ۲۲۰۳۳۱۴ ⟶ ۲۲۰۲۳۱۲ دارای ۰.۹۳۱، قانون ۲۲۰۲۳۱۲ ⟶ ۲۲۰۰۱۷۰ دارای ۰.۹۴۸، قانون ۲۲۰۲۳۱۲ ⟶ ۲۲۰۳۳۱۴ دارای ۰.۹۵۵ و قانون ۲۲۰۴۳۶۴ ⟶ ۲۲۰۳۳۱۴ دارای ۰.۹۵۵ بود. همچنین، ترکیب ۲۲۰۰۱۷۰ و ۲۲۰۴۳۶۴ با نتیجه ۲۲۰۳۳۱۴ اطمینان ۱.۰۰۰ و ترکیب ۲۲۰۲۳۱۲ و ۲۲۰۴۳۶۴ با نتیجه ۲۲۰۳۳۱۴ نیز اطمینان ۱.۰۰۰ داشت.

قوانین مربوط به ۲۲۰۴۳۵۹ در خوشه ارزشمند غیرریزشی نیز نشان دادند که مقدم‌های ۲۲۰۳۳۱۴ و ۲۲۰۰۱۷۰ هر یک با اطمینان ۰.۹۰ به حضور ۲۲۰۴۳۵۹ منجر می‌شدند و مقدم ۲۲۰۲۳۱۲ با اطمینان ۰.۹۳ به حضور این کالا مرتبط بود. در مجموعه بزرگ‌تر قوانین این خوشه، روابطی مانند ۲۲۰۴۳۵۹–۲۲۰۳۳۱۴ ⟶ ۲۲۰۱۶۵۵، ۲۲۰۱۶۹۱ یا ۲۲۰۱۶۵۸ و نیز ۲۲۰۴۳۵۹–۲۲۰۰۱۷۰ ⟶ همین اقلام با اطمینان ۰.۸۰۰ مشاهده شد. این یافته‌ها نشان دادند که برخی کدهای کالا به‌صورت یک هسته خرید مشترک در میان مشتریان فعال و ارزشمند ظاهر می‌شوند.

در مشتریان ارزشمند ریزشی، روابط متمرکزتری میان ۲۲۰۱۶۰۹، ۲۲۰۰۳۵۱، ۲۲۰۰۳۴۸، ۲۲۰۱۶۱۸ و ۲۲۰۱۶۱۱ مشاهده شد. قانون ۲۲۰۱۹۴۶ ⟶ ۱۶۴۲۰۰۲۷۵۳۰۰۰۱ دارای اطمینان ۰.۵۰۰ بود. قوانین ۲۲۰۱۶۰۹ ⟶ ۲۲۰۱۶۱۸ و ۲۲۰۰۳۵۱ ⟶ ۲۲۰۱۶۱۱ اطمینان ۰.۵۷۱ داشتند. روابط ۲۲۰۱۶۰۹ ⟶ ۲۲۰۰۳۴۸، ۲۲۰۱۶۰۹ ⟶ ۲۲۰۱۶۱۱، ۲۲۰۰۳۵۱ ⟶ ۲۲۰۱۶۱۸، ۲۲۰۱۶۰۹ ⟶ ۲۲۰۰۳۵۱+۲۲۰۰۳۴۸ و ۲۲۰۰۳۵۱ ⟶ ۲۲۰۱۶۰۹+۲۲۰۰۳۴۸ اطمینان ۰.۶۱۹ نشان دادند. قانون ۲۲۰۰۳۴۸ ⟶ ۲۲۰۱۶۰۹ دارای اطمینان ۰.۶۵۰ بود. اطمینان قوانین ۲۲۰۱۶۱۸ ⟶ ۲۲۰۱۶۰۹ و ۲۲۰۱۶۱۱ ⟶ ۲۲۰۰۳۵۱ برابر ۰.۷۰۶، ۲۲۰۱۶۰۹ ⟶ ۲۲۰۰۳۵۱ و ۲۲۰۰۳۵۱ ⟶ ۲۲۰۱۶۰۹ یا ۲۲۰۰۳۴۸ برابر ۰.۷۶۲ و ۲۲۰۰۳۴۸ ⟶ ۲۲۰۰۳۵۱ برابر ۰.۸۰۰ بود. در نهایت، قانون ۲۲۰۱۶۰۹+۲۲۰۰۳۴۸ ⟶ ۲۲۰۰۳۵۱ اطمینان کامل ۱.۰۰۰ نشان داد.

در خروجی دیگر مشتریان ارزشمند ریزشی، قانون ۶۲۶۱۹۰۶۱۰۱۱۷۶ ⟶ ۲۲۰۱۰۳۵ دارای اطمینان ۰.۲۰۴، ۱۶۴۲۰۰۲۷۵۳۰۰۰۱ ⟶ ۲۲۰۱۹۴۶ برابر ۰.۲۶۸، ۶۲۶۱۹۰۶۱۰۱۱۷۶ ⟶ ۱۶۴۲۰۰۲۷۵۳۰۰۰۱ برابر ۰.۳۲۸، ۲۲۰۱۹۴۶ ⟶ ۱۶۴۲۰۰۲۷۵۳۰۰۰۱ برابر ۰.۳۳۸، ۱۶۴۲۰۰۲۷۵۳۰۰۰۱ ⟶ ۶۲۶۱۹۰۶۱۰۱۱۷۶ برابر ۰.۴۰۲ و ۶۲۶۱۹۰۶۱۰۱۱۷۶ ⟶ ۲۲۰۱۰۳۵ برابر ۰.۴۳۶ بود. در مجموعه کدهای ۲۲۰۰۳۵۱، ۲۲۰۱۶۱۸، ۲۲۰۱۶۱۱ و ۲۲۰۰۳۴۸، اطمینان روابط بین ۰.۵۶۲ تا ۰.۶۷۶ قرار داشت. این الگوها نشان دادند سبد خرید مشتریان ارزشمند ریزشی از نظر کالاهای پرتکرار و روابط میان آن‌ها با گروه غیرریزشی یکسان نیست و می‌توان از این تفاوت برای طراحی پیشنهادهای بازگشت مشتری استفاده کرد.

E-ISSN: 3041-8933

در بخش طبقه‌بندی، هدف تشخیص مشتریان ریزشی و غیرریزشی بر اساس خرید یا عدم خرید کالاها بود. از آنجا که کلاس‌های اولیه نامتوازن بودند، مشتریان خوشه ارزشمند مبنای تحلیل قرار گرفتند و ۱۷۰ مشتری ریزشی با ۱۷۰ مشتری غیرریزشی که به‌صورت تصادفی انتخاب شده بودند ترکیب شدند. در نتیجه، مجموعه داده متعادل شامل ۳۴۰ مشتری ایجاد شد. در ساختار اولیه ۸,۴۴۱ ویژگی کالایی وجود داشت، اما پس از حذف کالاهایی که در نمونه انتخاب‌شده مشاهده کافی نداشتند، ۴۸۵ ویژگی باقی ماند. مقدار یک نشان‌دهنده خرید کالا و مقدار صفر نشان‌دهنده عدم خرید آن بود و Churn متغیر هدف قرار گرفت.

درخت تصمیم‌گیری باGain Ratio ، حداکثر عمق ۱۵، حداقل اندازه برگ ۲ و هرس اجرا شد. ماتریس اغتشاش آن شامل ۱۴۲ غیرریزشی درست، ۲۸ غیرریزشی که به‌اشتباه ریزشی تشخیص داده شدند، ۱۴۱ ریزشی درست و ۲۹ ریزشی که به‌اشتباه غیرریزشی تشخیص داده شدند بود. درخت ایجادشده قواعد قابل‌تفسیر متعددی استخراج کرد. برای مثال، خرید کد ۶۲۶۱۰۶۱۱۰۰۵۳۳ مستقیماً به کلاس No_Churnمنتهی شد و برای ۳۹ نمونه غیرریزشی بدون نمونه ریزشی مشاهده شد. در ادامه مسیر، کدهای ۶۲۶۰۱۰۰۱۰۳۹۵۵، ۶۲۶۶۰۶۱۱۰۳۹۱۱، ۲۲۰۲۰۲۲، ۶۲۶۰۰۶۳۲۰۰۸۲۱، ۲۲۰۰۴۹۵ و ۶۲۶۰۰۶۳۲۰۰۸۳۸ نیز شاخه‌هایی با غلبه کلاس غیرریزشی تشکیل دادند. در بخش‌های عمیق‌تر، کالاهایی مانند ۶۲۶۰۱۶۱۵۰۴۹۱۳، ۶۲۶۵۵۴۷۶۲۰۲۸۷، ۲۲۰۰۰۲۹ و ۲۲۰۱۰۱۴ در برخی ترکیبات به کلاس Churnمنجر شدند. در شاخه نهایی که اکثر شرایط خرید برقرار نبود، ۱۴۹ مشتری ریزشی در برابر ۲۸ مشتری غیرریزشی قرار گرفتند.

برای نزدیک‌ترین همسایه، kبرابر ۳ و فاصله اقلیدسی استفاده شد. ماتریس اغتشاش شامل ۷۸ غیرریزشی درست، ۹۲ غیرریزشی اشتباه، ۱۴۹ ریزشی درست و ۲۱ ریزشی اشتباه بود. این مدل Recall بالایی برای کلاس ریزشی داشت، اما تعداد زیادی مثبت کاذب ایجاد کرد.

در نیوبیز، ۱۳۹ مشتری غیرریزشی و ۱۵۲ مشتری ریزشی به‌درستی تشخیص داده شدند و به‌ترتیب ۳۱ و ۱۸ مورد خطا در این دو کلاس مشاهده شد. این مدل تعادل مناسب‌تری میان Precision و Recall نشان داد.

در شبکه عصبی با نرخ یادگیری ۰.۳ و مومنتوم ۰.۲، تعداد ۱۲۶ مشتری غیرریزشی و ۱۵۰ مشتری ریزشی درست طبقه‌بندی شدند و ۴۴ غیرریزشی و ۲۰ ریزشی به‌اشتباه طبقه‌بندی شدند.

در رگرسیون لجستیک با C برابر ۱، تعداد ۱۲۹ مشتری غیرریزشی و ۱۶۳ مشتری ریزشی درست طبقه‌بندی شدند و فقط ۷ مشتری ریزشی از دست رفتند؛ هرچند ۴۱ مشتری غیرریزشی به‌اشتباه به‌عنوان ریزشی شناسایی شدند.

در ماشین بردار پشتیبان، ۱۳۷ مشتری غیرریزشی و ۱۵۱ مشتری ریزشی درست تشخیص داده شدند و تعداد خطاها برای دو کلاس به‌ترتیب ۳۳ و ۱۹ بود.

تحلیل تمایزی خطی ضعیف‌ترین عملکرد را داشت؛ ۹۳ مشتری غیرریزشی و ۱۰۶ مشتری ریزشی درست طبقه‌بندی شدند، درحالی‌که ۷۷ غیرریزشی و ۶۴ ریزشی به‌اشتباه تشخیص داده شدند.

در آدابوست با پنج تکرار و استفاده از درخت تصمیم‌گیری، رگرسیون لجستیک، SVMو نیوبیز به‌عنوان مدل‌های پایه، ۱۳۷ مشتری غیرریزشی و ۱۴۹ مشتری ریزشی درست طبقه‌بندی شدند و تعداد خطاها ۳۳ و ۲۱ بود.

بگینگ از نظر کشف مشتریان ریزشی حساسیت بسیار بالایی داشت. این مدل ۱۶۸ مورد از ۱۷۰ مشتری ریزشی را شناسایی کرد و فقط ۲ مشتری ریزشی را از دست داد، اما ۹۲ مشتری غیرریزشی را نیز به‌اشتباه در کلاس ریزشی قرار داد. بنابراین، Recallبسیار بالا به قیمت کاهش Precision حاصل شد.

استیکینگ با ترکیب چهار طبقه‌کننده پایه، ۱۴۱ مشتری غیرریزشی و ۱۵۳ مشتری ریزشی را درست شناسایی کرد و فقط ۲۹ غیرریزشی و ۱۷ ریزشی را اشتباه طبقه‌بندی کرد. این مدل در مجموع متعادل‌ترین عملکرد را در شاخص‌های ارزیابی نشان داد.

جدول ۷

مقایسه نهایی عملکرد الگوریتم‌های طبقه‌بندی

الگوریتم	Accuracy (٪)	Precision (٪)	Recall (٪)	F۱ (٪)	رتبه نهایی
استیکینگ	۸۶.۴۷	۸۴.۰۷	۹۰.۰۰	۸۶.۹۳	۱
رگرسیون لجستیک	۸۵.۸۸	۷۹.۹۰	۹۵.۸۸	۸۷.۱۷	۲
نیوبیز	۸۵.۵۹	۸۳.۰۶	۸۹.۴۱	۸۶.۱۲	۳
ماشین بردار پشتیبان	۸۴.۷۱	۸۲.۰۷	۸۸.۸۲	۸۵.۳۱	۴
درخت تصمیم‌گیری	۸۳.۲۴	۸۳.۴۳	۸۲.۹۴	۸۳.۱۹	۵
آدابوست	۸۴.۱۲	۸۱.۸۷	۸۷.۶۵	۸۴.۶۶	۶
بگینگ	۷۲.۳۵	۶۴.۶۲	۹۸.۸۲	۷۸.۱۴	۷
شبکه عصبی	۸۱.۱۸	۷۷.۳۲	۸۸.۲۴	۸۲.۴۲	۸
نزدیک‌ترین همسایه	۶۶.۷۶	۶۱.۸۳	۸۷.۶۵	۷۲.۵۱	۹
تحلیل تمایزی خطی	۵۸.۵۳	۵۷.۹۲	۶۲.۳۵	۶۰.۰۶	۱۰

رتبه‌بندی بر اساس عملکرد هم‌زمان چهار شاخص نشان داد میانگین رتبه استیکینگ ۱.۷۵ بود و بهترین جایگاه را کسب کرد. رگرسیون لجستیک با میانگین رتبه ۲.۷۵ و نیوبیز با ۳.۲۵ در رتبه‌های دوم و سوم قرار گرفتند. میانگین رتبه ماشین بردار پشتیبان ۴.۲۵، درخت تصمیم‌گیری و آدابوست هر یک ۵.۷۵، بگینگ ۶.۲۵، شبکه عصبی ۶.۷۵، نزدیک‌ترین همسایه ۸.۵۰ و تحلیل تمایزی خطی ۱۰.۰۰ بود. استیکینگ در Accuracy رتبه نخست، در Precision رتبه نخست، در Recall رتبه سوم و در F۱ رتبه دوم را کسب کرد. رگرسیون لجستیک با وجود Precision پایین‌تر، در Recall رتبه دوم و در F۱ رتبه نخست قرار گرفت. بگینگ بالاترین Recall را به دست آورد، اما به دلیل Precision پایین‌تر در رتبه‌بندی کلی جایگاه هفتم را کسب کرد.

در مجموع، یافته‌های حاصل از سه جزء اصلی چارچوب یکپارچه نشان دادند که شاخص‌های LRFFM امکان تمایز روشن میان گروه‌های رفتاری مشتریان را فراهم کردند؛ به‌ویژه خوشه‌ای با Recency کمتر و Frequency ، Monetary و Length بیشتر به‌عنوان بخش ارزشمند مشتریان شناسایی شد. تحلیل مستقل مشتریان ریزشی نیز گروهی از مشتریان با سابقه خرید و ارزش مالی بالا را آشکار کرد که برای اقدامات بازگشت مشتری اهمیت بیشتری داشتند. قوانین وابستگی نشان دادند الگوهای سبد خرید مشتریان ارزشمند غیرریزشی و ارزشمند ریزشی از نظر اقلام پرتکرار و روابط هم‌وقوعی کالاها تفاوت دارند. در نهایت، طبقه‌بندی مبتنی بر الگوی خرید نشان داد روش استیکینگ بهترین عملکرد کلی را در تفکیک مشتریان ریزشی و غیرریزشی ارائه می‌دهد، درحالی‌که رگرسیون لجستیک، نیوبیز و ماشین بردار پشتیبان نیز عملکرد مطلوبی داشتند. این نتایج در کنار یکدیگر نشان دادند که ترکیب LRFFM ، خوشه‌بندی، قوانین وابستگی و طبقه‌بندی می‌تواند چارچوبی یکپارچه برای شناخت ارزش مشتری، شناسایی ریزش، تحلیل سبد خرید و پشتیبانی از تصمیم‌گیری مدیریتی در محیط خرده‌فروشی فراهم سازد.

E-ISSN: 3041-8933

بحث و نتیجه‌گیری

یافته‌های پژوهش حاضر نشان داد که چارچوب یکپارچه مبتنی بر شاخص‌های توسعه‌یافته LRFFM و تکنیک‌های داده‌کاوی توانایی مناسبی برای شناسایی و تحلیل چندبعدی رفتار مشتریان فروشگاه خرده‌فروشی دارد. نتایج خوشه‌بندی نشان داد که مشتریان از نظر تازگی خرید، تعداد تراکنش‌ها، تعداد اقلام خریداری‌شده، ارزش مالی خرید و طول دوره تعامل با فروشگاه ناهمگن هستند و می‌توان آنان را در گروه‌هایی با الگوهای رفتاری متفاوت قرار داد. در ساختار سه‌خوشه‌ای، یک خوشه با Recency پایین‌تر و مقادیر بالاتر Frequency، Monetary و Length به‌عنوان خوشه مشتریان ارزشمندتر شناسایی شد. این خوشه علاوه بر فعالیت خرید بیشتر، سهم مالی بالاتری برای فروشگاه ایجاد کرده و نسبت مشتریان غیرریزشی در آن نیز بیشتر بود. این یافته نشان می‌دهد که استفاده هم‌زمان از ابعاد مختلف LRFFM نسبت به اتکا به یک شاخص منفرد، تصویر جامع‌تری از ارزش و وضعیت مشتری ارائه می‌کند. پژوهش‌های اخیر نیز نشان داده‌اند که متغیرهای رفتاری مبتنی بر RFM و نسخه‌های توسعه‌یافته آن می‌توانند در پیش‌بینی ریزش و شناخت پویایی رفتار مشتری نقش مؤثری داشته باشند و استفاده از تغییرات زمانی این شاخص‌ها قدرت تشخیص مدل را افزایش می‌دهد (Chou et al., 2022; Mena et al., 2024).

نتایج خوشه‌بندی همچنین نشان داد که تعداد اندکی از مشتریان می‌توانند از نظر ارزش مالی، فراوانی تراکنش و سابقه تعامل، تفاوت چشمگیری با بخش بزرگ‌تر مشتریان داشته باشند. این موضوع از منظر مدیریت مشتری اهمیت زیادی دارد، زیرا ارزش اقتصادی مشتریان به‌طور یکنواخت در جامعه مشتریان توزیع نمی‌شود و یک گروه کوچک می‌تواند سهم قابل‌توجهی از تعامل و درآمد فروشگاه را تشکیل دهد. یافته حاضر با پژوهش‌هایی همسو است که بر استفاده از خوشه‌بندی برای آشکارسازی ساختارهای پنهان رفتاری پیش از اجرای مدل‌های پیش‌بینی تأکید کرده‌اند. در مدل‌های ترکیبی، خوشه‌بندی می‌تواند ناهمگونی مشتریان را کاهش دهد و گروه‌هایی با رفتار مشابه ایجاد کند که تحلیل آن‌ها از نظر مدیریتی معنادارتر است (Liu et al., 2022). همچنین، مطالعات خرده‌فروشی نشان داده‌اند که بررسی رفتار مشتریان در سطح گروه‌های متفاوت، نسبت به تحلیل یکپارچه کل مشتریان، امکان شناسایی دقیق‌تر مشتریان پرارزش و مشتریان در معرض ریزش را فراهم می‌سازد (Jung et al., 2023; Sweidan et al., 2022).

یکی دیگر از یافته‌های مهم پژوهش، شناسایی گروه «مشتریان ارزشمند ریزشی» بود. در خوشه‌بندی اختصاصی مشتریان ریزشی، گروهی شناسایی شد که با وجود قرار گرفتن در وضعیت ریزش، تعداد تراکنش، تعداد کالاهای خریداری‌شده، مبلغ خرید و طول رابطه بالاتری نسبت به سایر مشتریان ریزشی داشت. Recency این گروه نیز نسبت به خوشه دیگر مشتریان ریزشی کمتر بود. این نتیجه از نظر مدیریتی اهمیت ویژه‌ای دارد، زیرا نشان می‌دهد همه مشتریان ریزشی از یک ارزش اقتصادی و رفتاری برخوردار نیستند. در نتیجه، اقدامات نگهداشت و بازگشت مشتری نباید به‌صورت یکسان برای تمامی مشتریان ریزشی اجرا شود. مطالعات پیشین نیز نشان داده‌اند که شناسایی مشتریان پرارزش در معرض ریزش می‌تواند منابع سازمان را به سوی گروه‌هایی هدایت کند که احتمال بازده بالاتری از اقدامات نگهداشت دارند (Espinoza-Vega & Roa, 2024; Mahdi & Jabbari, 2024). همچنین، مطالعات مربوط به مشتریان تکراری در صنعت هتلداری نشان داده‌اند که سابقه رفتاری گذشته مشتری اطلاعات مهمی برای پیش‌بینی ریزش آینده فراهم می‌کند و مشتریانی که سابقه تعامل قابل‌توجهی داشته‌اند باید به‌طور ویژه پایش شوند (Dursun-Cengizci & Caber, 2025).

یافته دیگر پژوهش مربوط به اهمیت Recency در تشخیص کاهش تعامل مشتریان بود. در خوشه‌های ارزشمندتر، مقدار Recency کمتر و در گروه‌های ریزشی مقدار آن به‌طور محسوسی بیشتر بود. این نتیجه نشان می‌دهد که افزایش فاصله زمانی از آخرین تراکنش یکی از شاخص‌های کلیدی کاهش ارتباط مشتری با فروشگاه است. چنین یافته‌ای با ماهیت غیرقراردادی خرده‌فروشی سازگار است، زیرا در این

محیط‌ها مشتری معمولاً قطع ارتباط خود را به‌صورت رسمی اعلام نمی‌کند و ریزش باید از روی رفتار واقعی خرید استنباط شود. پژوهش‌هایی که تغییرات رفتار مشتری را در طول زمان بررسی کرده‌اند نیز نشان داده‌اند که افزایش فاصله میان تعاملات و تغییر الگوهای خرید از نشانه‌های اولیه ریزش محسوب می‌شود (Matuszelański & Kopczewska, 2022; Zelenkov & Suchkova, 2023). همچنین، بهره‌گیری از داده‌های متوالی و تاریخچه تراکنش‌ها برای پیش‌بینی ریزش در خرده‌فروشی توانسته است عملکرد مناسبی نشان دهد، که اهمیت بعد زمانی رفتار مشتری را تأیید می‌کند (Equihua et al., 2023).

نتایج قوانین وابستگی نیز نشان داد که سبد خرید مشتریان ارزشمند غیرریزشی دارای مجموعه‌هایی از کالاهای پرتکرار با روابط هم‌وقوعی نسبتاً قوی است. در برخی قوانین، اطمینان روابط به بیش از ۰.۹۰ و حتی ۱.۰۰ رسید که نشان می‌دهد خرید برخی اقلام با خرید اقلام دیگر همراهی بسیار بالایی دارد. این یافته حاکی از آن است که در میان مشتریان فعال و ارزشمند، رفتار خرید صرفاً بر اساس حجم و مبلغ خرید قابل توصیف نیست، بلکه ترکیب کالاها نیز ساختار مشخصی دارد. ازاین‌رو، قوانین وابستگی می‌توانند مکمل شاخص‌های LRFFM باشند و ابعادی از رفتار مشتری را آشکار کنند که صرفاً از طریق Recency، Frequency یا Monetary قابل مشاهده نیست. مطالعات مربوط به پیش‌بینی رفتار خرید نشان داده‌اند که الگوهای کالاهای انتخاب‌شده توسط مشتری می‌توانند اطلاعات ارزشمندی برای شناخت رفتار آتی، طراحی پیشنهادهای مکمل و شخصی‌سازی تعاملات فراهم کنند (Chaubey et al., 2023). همچنین، کاربرد یادگیری ماشین در تجارت الکترونیکی نشان داده است که تحلیل جزئی رفتار خرید می‌تواند به طراحی راهبردهای هوش تجاری برای حفظ مشتری کمک کند (Shobana et al., 2023).

مقایسه قوانین وابستگی مشتریان ارزشمند غیرریزشی با مشتریان ارزشمند ریزشی نیز نشان داد که ساختار کالاهای پرتکرار و روابط میان اقلام در این دو گروه یکسان نیست. در مشتریان غیرریزشی، مجموعه‌ای از کدهای کالا با روابط بسیار قوی و پرتکرار مشاهده شد، درحالی‌که در مشتریان ریزشی ترکیب دیگری از اقلام غالب بود. این تفاوت می‌تواند نشان‌دهنده آن باشد که تغییر در ترکیب سبد خرید یا تعلق مشتریان به الگوهای متفاوت خرید، اطلاعات تکمیلی درباره وضعیت ارتباط آنان با فروشگاه فراهم می‌کند. مطالعات پیشین درباره ریزش در سطح گروه‌های کالایی نیز تأکید کرده‌اند که ریزش ممکن است نه‌تنها در سطح کل مشتری، بلکه در سطح دسته‌های محصول و نوع خرید نیز قابل مشاهده باشد (Sousa, 2023). ازاین‌رو، ترکیب خوشه‌بندی مشتریان با تحلیل سبد خرید یکی از نقاط مهم چارچوب حاضر است؛ زیرا امکان می‌دهد روابط میان کالاها برای گروه‌هایی با اهمیت مدیریتی متفاوت بررسی شوند، نه اینکه قوانین وابستگی صرفاً برای کل مشتریان بدون توجه به ارزش و وضعیت ریزش استخراج شوند.

در بخش طبقه‌بندی، نتایج نشان داد که الگوریتم‌های مختلف توانایی متفاوتی در تفکیک مشتریان ریزشی و غیرریزشی دارند. بالاترین عملکرد کلی مربوط به استیکینگ بود که Accuracy برابر با ۸۶.۴۷ درصد، Precision برابر با ۸۴.۰۷ درصد، Recall برابر با ۹۰.۰۰ درصد و F1 برابر با ۸۶.۹۳ درصد داشت. این نتیجه نشان می‌دهد که ترکیب خروجی چند طبقه‌کننده می‌تواند نسبت به اتکا به یک مدل منفرد، تعادل مناسبی میان شاخص‌های عملکرد ایجاد کند. پژوهش‌های مقایسه‌ای نیز نشان داده‌اند که روش‌های Ensemble می‌توانند در مسئله ریزش عملکرد قوی داشته باشند، زیرا نقاط قوت چند مدل را ترکیب کرده و ضعف‌های یک طبقه‌کننده منفرد را کاهش می‌دهند (Bogaert & Delaere, 2023). همچنین، مدل‌های ترکیبی و روش‌های بازنمونه‌گیری در مسئله ریزش مشتری توانسته‌اند عملکرد مناسبی در شرایط وجود ناهمگونی و عدم تعادل داده‌ها ارائه کنند (Kimura, 2022).

رگرسیون لجستیک در این پژوهش نیز عملکرد مطلوبی داشت و Recall آن به ۹۵.۸۸ درصد رسید. این موضوع نشان می‌دهد که این الگوریتم توانسته است بخش بزرگی از مشتریان ریزشی واقعی را شناسایی کند. بااین‌حال، Precision آن نسبت به استیکینگ و برخی

E-ISSN: 3041-8933

مدل‌های دیگر پایین‌تر بود که نشان می‌دهد افزایش حساسیت در شناسایی ریزش با افزایش تعداد مثبت‌های کاذب همراه شده است. این یافته اهمیت استفاده هم‌زمان از چند شاخص ارزیابی را نشان می‌دهد. در مطالعات مربوط به پیش‌بینی ریزش نیز تأکید شده است که Accuracy به‌تنهایی برای قضاوت درباره عملکرد کافی نیست و شاخص‌هایی مانند Precision، Recall و F1 باید به‌صورت هم‌زمان بررسی شوند (Louro et al., 2024; Manzoor et al., 2024). پژوهش‌هایی در بانکداری، مخابرات و سایر صنایع نیز نشان داده‌اند که رگرسیون لجستیک همچنان می‌تواند در کنار مدل‌های پیچیده‌تر، به‌ویژه زمانی که تفسیرپذیری اهمیت دارد، عملکرد رقابتی ارائه دهد (Srivastava & Sinha, 2024; Tran et al., 2023).

مدل نیوبیز نیز در رتبه سوم قرار گرفت و تعادل مناسبی میان Accuracy، Precision، Recall و F1 نشان داد. این نتیجه حاکی از آن است که علی‌رغم فرض ساده استقلال شرطی میان ویژگی‌ها، رویکرد احتمالاتی نیوبیز در داده‌های دودویی مربوط به خرید یا عدم خرید کالاها کارایی قابل‌قبولی دارد. یافته حاضر با مطالعاتی همسو است که مقایسه چند مدل طبقه‌بندی را بر ضرورت ارزیابی تجربی الگوریتم‌ها در هر مجموعه داده تأکید کرده‌اند و نشان داده‌اند هیچ الگوریتمی را نمی‌توان پیشاپیش به‌عنوان بهترین گزینه معرفی کرد (Lalwani et al., 2022; Mouli et al., 2024). به بیان دیگر، ساختار ویژگی‌ها، تعداد متغیرها، نوع کدگذاری و میزان عدم تعادل کلاس‌ها می‌توانند رتبه الگوریتم‌ها را تغییر دهند.

ماشین بردار پشتیبان و درخت تصمیم‌گیری نیز عملکرد قابل قبولی نشان دادند. مزیت درخت تصمیم‌گیری در پژوهش حاضر تنها به دقت آن محدود نبود، بلکه قواعد اگر–آنگاه استخراج‌شده از آن امکان شناسایی کالاهایی را فراهم کرد که در مسیر تفکیک مشتریان ریزشی و غیرریزشی نقش داشتند. این تفسیرپذیری از نظر کاربرد مدیریتی اهمیت دارد، زیرا مدیران می‌توانند علاوه بر دریافت برچسب پیش‌بینی، منطق برخی تصمیمات مدل را مشاهده کنند. مطالعات مربوط به استفاده از یادگیری ماشین در پیش‌بینی رفتار خرید و ریزش نیز نشان داده‌اند که الگوریتم‌های تفسیرپذیر می‌توانند بین عملکرد پیش‌بینی و قابلیت استفاده مدیریتی تعادل مناسبی ایجاد کنند (Adeniran et al., 2024; Chaubey et al., 2023). در مقابل، تحلیل تمایزی خطی ضعیف‌ترین عملکرد را در مطالعه حاضر داشت که می‌تواند ناشی از پیچیدگی روابط میان صدها ویژگی دودویی کالایی و ناکافی بودن فرض ساختار خطی برای جداسازی دو کلاس باشد.

بگینگ یافته متفاوتی ارائه کرد. این مدل Recall بسیار بالای ۹۸.۸۲ درصد داشت و تقریباً تمام مشتریان ریزشی را شناسایی کرد، اما Precision آن تنها ۶۴.۶۲ درصد بود. بنابراین، بگینگ برای موقعیتی که هدف اصلی از دست ندادن تقریباً هیچ مشتری ریزشی باشد مناسب‌تر به نظر می‌رسد، ولی هزینه آن تولید تعداد بیشتری هشدار کاذب است. این نتیجه نشان می‌دهد انتخاب «بهترین مدل» باید با هدف کسب‌وکار هماهنگ باشد. اگر هزینه از دست دادن یک مشتری ارزشمند بسیار زیاد باشد، سازمان ممکن است مدل با Recall بالاتر را ترجیح دهد؛ اما اگر منابع برنامه‌های نگهداشت محدود باشند، Precision و F1 اهمیت بیشتری پیدا خواهند کرد. این ملاحظه با مطالعات مربوط به حفظ مشتری و پیش‌بینی ریزش همسو است که بر ضرورت اتصال معیارهای فنی مدل به هزینه‌ها و اهداف واقعی کسب‌وکار تأکید دارند (Brito et al., 2024; Phumchusri & Amornvetchayakul, 2024).

یکی دیگر از یافته‌های قابل‌توجه پژوهش آن بود که طبقه‌بندی بر اساس الگوی خرید یا عدم خرید کالاها توانست مشتریان ریزشی و غیرریزشی را با عملکرد نسبتاً بالا از یکدیگر تفکیک کند. این نتیجه بیانگر آن است که سبد خرید مشتری اطلاعاتی فراتر از شاخص‌های کلی ارزش مشتری دارد. الگوی انتخاب اقلام می‌تواند حامل اطلاعاتی درباره ترجیحات، سبک خرید، میزان وابستگی به دسته‌های خاص کالا و احتمال استمرار تعامل باشد. مطالعات مربوط به پیش‌بینی خرید مجدد نیز نشان داده‌اند که ترکیب ویژگی‌های رفتاری با مدل‌های یادگیری ماشین می‌تواند اطلاعات بیشتری درباره تداوم خرید فراهم آورد (Chou et al., 2022). همچنین، پژوهش‌های حوزه خرده‌فروشی کالاهای

تندمصرف و دارویی نقش داده‌های تراکنشی را در تشخیص مشتریان در معرض ترک تأیید کرده‌اند (Espinoza-Vega & Roa, 2024; Mahdi & Jabbari, 2024).

در مجموع، نتایج پژوهش حاضر از اعتبار رویکرد یکپارچه حمایت می‌کند؛ زیرا هر یک از تکنیک‌ها بعد متفاوتی از رفتار مشتری را آشکار کرد. LRFFM امکان خلاصه‌سازی ارزش و تعامل مشتری را فراهم ساخت؛ K-Means گروه‌های رفتاری و به‌ویژه مشتریان ارزشمند و مشتریان ارزشمند ریزشی را آشکار کرد؛ FP-Growth و قوانین وابستگی ساختار سبد خرید و روابط میان کالاها را مشخص کردند؛ و الگوریتم‌های طبقه‌بندی امکان تشخیص مشتریان ریزشی و غیرریزشی را فراهم آوردند. این یکپارچگی با جهت‌گیری مطالعات جدید در حرکت از مدل‌های منفرد به چارچوب‌های چندمرحله‌ای و داده‌محور همسو است (Faritha Banu et al., 2022; Sana et al., 2022; Sina Mirabdolbaghi & Amiri, 2022). همچنین، نتایج با مطالعاتی که بر استفاده از هوش مصنوعی و مدل‌های پیش‌بین برای طراحی راهبردهای فعال نگهداشت مشتری تأکید دارند، همخوانی دارد (Das et al., 2023; Vemulapalli, 2024). بنابراین، ارزش اصلی چارچوب حاضر در آن است که ریزش را جدا از ارزش مشتری و رفتار خرید مطالعه نمی‌کند، بلکه این ابعاد را در یک ساختار تحلیلی واحد به یکدیگر متصل می‌سازد.

این پژوهش با چند محدودیت همراه بود. نخست، داده‌ها متعلق به یک شعبه فروشگاه خرده‌فروشی بود و بنابراین الگوهای استخراج‌شده ممکن است تحت تأثیر ویژگی‌های مکانی، ترکیب مشتریان، تنوع کالا و سیاست‌های همان شعبه قرار گرفته باشند. دوم، تعریف ریزش بر اساس آستانه ۹۳ روز انجام شد و هرچند این تعریف با ماهیت داده‌های غیرقراردادی خرده‌فروشی سازگار بود، انتخاب آستانه‌های دیگر ممکن است ترکیب مشتریان ریزشی و عملکرد مدل‌ها را تغییر دهد. سوم، در بخش طبقه‌بندی برای رفع عدم تعادل کلاس‌ها از کم‌نمونه‌گیری استفاده شد که موجب حذف بخشی از اطلاعات مشتریان غیرریزشی شد. چهارم، برخی متغیرهای بالقوه مؤثر مانند ویژگی‌های جمعیت‌شناختی، تخفیف‌ها، نوع کمپین بازاریابی، قیمت نسبی کالا، کانال خرید، مناسبت‌های زمانی و داده‌های رضایت مشتری در دسترس نبودند. همچنین، بخشی از فرایند آماده‌سازی داده‌ها بر ساختار زمانی ساده‌شده مبتنی بود و تحلیل‌های پیچیده‌تر توالی زمانی می‌توانست اطلاعات بیشتری از رفتار مشتری استخراج کند.

پیشنهاد می‌شود در پژوهش‌های آینده چارچوب حاضر بر داده‌های چند شعبه و دوره‌های زمانی طولانی‌تر اجرا شود تا پایداری خوشه‌ها، قوانین وابستگی و عملکرد الگوریتم‌های طبقه‌بندی در شرایط متفاوت ارزیابی شود. همچنین، می‌توان آستانه‌های مختلف ریزش را مقایسه و از روش‌های داده‌محور برای تعیین نقطه بهینه عدم فعالیت مشتری استفاده کرد. استفاده از روش‌های متعادل‌سازی پیشرفته‌تر مانند نمونه‌سازی مصنوعی، روش‌های هزینه‌محور و Ensembleهای متوازن نیز می‌تواند با حفظ اطلاعات بیشتر، عملکرد مدل را بهبود دهد. بررسی مدل‌های گرادیان بوستینگ، جنگل تصادفی، XGBoost، مدل‌های عمیق و مدل‌های توالی‌محور نیز پیشنهاد می‌شود. علاوه بر این، افزودن متغیرهای جمعیت‌شناختی، نوع تخفیف، زمان خرید، دسته محصول، شدت استفاده از برنامه‌های وفاداری و پاسخ مشتری به کمپین‌های گذشته می‌تواند تصویر کامل‌تری از ریزش ایجاد کند. مطالعه تغییرات LRFFM در طول زمان و استفاده از پنجره‌های متحرک نیز می‌تواند به شناسایی زودهنگام تغییرات رفتاری پیش از وقوع ریزش کمک کند.

پیشنهاد می‌شود مدیران فروشگاه به‌جای اجرای برنامه‌های یکسان برای همه مشتریان، مشتریان را بر اساس شاخص‌های LRFFM، خوشه رفتاری و وضعیت ریزش اولویت‌بندی کنند. مشتریان ارزشمند ریزشی باید در اولویت برنامه‌های بازگشت قرار گیرند، زیرا سابقه خرید و ارزش مالی بالاتری دارند. می‌توان از کالاهای پرتکرار و قوانین وابستگی استخراج‌شده برای طراحی پیشنهادهای مکمل، بسته‌های خرید، تخفیف‌های شخصی‌سازی‌شده و فروش متقاطع استفاده کرد. همچنین، استقرار مدل طبقه‌بندی استیکینگ به‌عنوان سامانه هشدار اولیه می‌تواند

E-ISSN: 3041-8933

مشتریان در معرض ریزش را شناسایی کند، درحالی‌که در شرایطی که از دست دادن مشتری بسیار پرهزینه است می‌توان از مدلی با Recall بالاتر برای غربالگری اولیه استفاده کرد. پیشنهاد می‌شود شاخص‌های مشتریان به‌صورت دوره‌ای به‌روزرسانی شوند و داشبوردی مدیریتی شامل Recency، Frequency، Monetary، Length، خوشه مشتری، احتمال ریزش و کالاهای مرتبط طراحی شود تا تصمیم‌های نگهداشت، بازگشت مشتری و پیشنهادهای فروش بر مبنای داده‌های به‌روز اتخاذ شوند.

تعارض منافع

در انجام مطالعه حاضر، هیچ‌گونه تضاد منافعی وجود ندارد.

مشارکت نویسندگان

در نگارش این مقاله تمامی نویسندگان نقش یکسانی ایفا کردند.

موازین اخلاقی

در انجام این پژوهش تمامی موازین و اصول اخلاقی رعایت گردیده است.

شفافیت داده‌ها

داده‌ها و مآخذ پژوهش حاضر در صورت درخواست از نویسنده مسئول و ضمن رعایت اصول کپی رایت ارسال خواهد شد.

حامی مالی

این پژوهش حامی مالی نداشته است.

References

Adeniran, I. A., Efunniyi, C. P., Osundare, O. S., Abhulimen, A. O., & OneAdvanced, U. (2024). Implementing Machine Learning Techniques for Customer Retention and Churn Prediction in Telecommunications. *Computer Science & It Research Journal*, *5*(8), 2011-2025. https://doi.org/10.51594/csitrj.v5i8.1489

Bogaert, M., & Delaere, L. (2023). Ensemble Methods in Customer Churn Prediction: A Comparative Analysis of the State-of-the-Art. *Mathematics*, *11*(5), 1137. https://doi.org/10.3390/math11051137

Brito, J. B. G., Bucco, G. B., & Heldt, R. (2024). A Framework to Improve Churn Prediction Performance in Retail Banking. *Financial Innovation*, *10*, 17. https://doi.org/10.1186/s40854-023-00558-3

Chaubey, G., Gavhane, P. R., Bisen, D., & Arjaria, S. K. (2023). Customer Purchasing Behavior Prediction Using Machine Learning Classification Techniques. *Journal of Ambient Intelligence and Humanized Computing*, *14*(12), 16133-16157. https://doi.org/10.1007/s12652-022-03837-6

Chaudhary, M., Afaq, A., Singh, G., & Kapoor, S. (2024). Unboxing the Mystery: Employee Churn in the Retail Industry Using Machine Learning Approach. *International Journal of System Assurance Engineering and Management*, 1-11. https://doi.org/10.1007/s13198-024-02490-w

Chou, P., Chuang, H. H. C., Chou, Y. C., & Liang, T. P. (2022). Predictive Analytics for Customer Repurchase: Interdisciplinary Integration of Buy-Till-You-Die Modeling and Machine Learning. *European Journal of Operational Research*, *296*(2), 635-651. https://doi.org/10.1016/j.ejor.2021.04.021

Das, L., Salman, R., Sabeer, S., Ansari, S. K., Aarif, M., & Rana, A. (2023). *Customer Retention Using Machine Learning* 2023 10th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON), https://doi.org/10.1109/UPCON59197.2023.10434812

Dursun-Cengizci, A., & Caber, M. (2025). Using Machine Learning Methods to Predict Future Churners: An Analysis of Repeat Hotel Customers. *International Journal of Contemporary Hospitality Management*, *37*(1), 36-56. https://doi.org/10.1108/IJCHM-06-2023-0844

Equihua, J. P., Nordmark, H., Ali, M., & Lausen, B. (2023). *Modelling Customer Churn for the Retail Industry in a Deep Learning Based Sequential Framework*.

Espinoza-Vega, A., & Roa, H. N. (2024). Boosting Customer Retention in Pharmaceutical Retail: A Predictive Approach Based on Machine Learning Models. In *Science and Information Conference* (pp. 97-117). Springer Nature. https://doi.org/10.1007/978-3-031-62277-9_7

Faritha Banu, J., Neelakandan, S., Geetha, B. T., Selvalakshmi, V., Umadevi, A., & Martinson, E. O. (2022). Artificial Intelligence-Based Customer Churn Prediction Model for Business Markets. *Computational Intelligence and Neuroscience*, *2022*, 1703696. https://doi.org/10.1155/2022/1703696

Jung, S. K., Tsang, S. S., & Phumchusri, N. (2023). Predicting Customer Churn: Revolutionizing Retail with Machine Learning. *International Journal for Applied Information Management*, *3*(1), 46-57. https://doi.org/10.47738/ijaim.v3i1.50

Kimura, T. (2022). Customer Churn Prediction with Hybrid Resampling and Ensemble Learning. *Journal of Management Information and Decision Sciences*, *25*(1).

Lalwani, P., Mishra, M. K., Chadha, J. S., & Sethi, P. (2022). Customer Churn Prediction System: A Machine Learning Approach. *Computing*, *104*(2), 271-294. https://doi.org/10.1007/s00607-021-00908-y

Liu, R., Ali, S., Bilal, S. F., Sakhawat, Z., Imran, A., Almuhaimeed, A., & Sun, G. (2022). An Intelligent Hybrid Scheme for Customer Churn Prediction Integrating Clustering and Classification Algorithms. *Applied Sciences*, *12*(18), 9355. https://doi.org/10.3390/app12189355

Louro, A. C., Pugirá, C. G., & Murari, R. S. (2024). A Scoping Review for Churn Prediction: Step-by-Step Tutorial and Reproducible R Code. *International Journal of Business Forecasting and Marketing Intelligence*, *9*(2), 160-178. https://doi.org/10.1504/IJBFMI.2024.137643

Mahdi, M., & Jabbari, M. (2024). Predicting Customer Churn in the Fast-Moving Consumer Goods Segment of the Retail Industry Using Deep Learning. *Mathematics and Computational Sciences*, *5*(3), 58-79.

Manzoor, A., Qureshi, M. A., Kidney, E., & Longo, L. (2024). A Review on Machine Learning Methods for Customer Churn Prediction and Recommendations for Business Practitioners. *IEEE Access*, *12*, 70434-70463. https://doi.org/10.1109/ACCESS.2024.3402092

Matuszelański, K., & Kopczewska, K. (2022). Customer Churn in Retail E-Commerce Business: Spatial and Machine Learning Approach. *Journal of Theoretical and Applied Electronic Commerce Research*, *17*(1), 165-198. https://doi.org/10.3390/jtaer17010009

Mena, G., Coussement, K., De Bock, K. W., De Caigny, A., & Lessmann, S. (2024). Exploiting Time-Varying RFM Measures for Customer Churn Prediction with Deep Neural Networks. *Annals of Operations Research*, *339*(1), 765-787. https://doi.org/10.1007/s10479-023-05259-9

Mouli, K. C., Raghavendran, C. V., Bharadwaj, V. Y., Vybhavi, G. Y., Sravani, C., Vafaeva, K. M., & Hussein, L. (2024). An Analysis on Classification Models for Customer Churn Prediction. *Cogent Engineering*, *11*(1), 2378877. https://doi.org/10.1080/23311916.2024.2378877

Phumchusri, N., & Amornvetchayakul, P. (2024). Machine Learning Models for Predicting Customer Churn: A Case Study in a Software-as-a-Service Inventory Management Company. *International Journal of Business Intelligence and Data Mining*, *24*(1), 74-106. https://doi.org/10.1504/IJBIDM.2024.135146

Saleh, S., & Saha, S. (2023). Customer Retention and Churn Prediction in the Telecommunication Industry: A Case Study on a Danish University. *SN Applied Sciences*, *5*(7), 173. https://doi.org/10.1007/s42452-023-05389-6

Sana, J. K., Abedin, M. Z., Rahman, M. S., & Rahman, M. S. (2022). A Novel Customer Churn Prediction Model for the Telecommunication Industry Using Data Transformation Methods and Feature Selection. *PLoS One*, *17*(12), e0278095. https://doi.org/10.1371/journal.pone.0278095

Shobana, J., Gangadhar, C., Arora, R. K., Renjith, P. N., Bamini, J., & Chincholkar, Y. D. (2023). E-Commerce Customer Churn Prevention Using Machine Learning-Based Business Intelligence Strategy. *Measurement: Sensors*, *27*, 100728. https://doi.org/10.1016/j.measen.2023.100728

Sina Mirabdolbaghi, S. M., & Amiri, B. (2022). Model Optimization Analysis of Customer Churn Prediction Using Machine Learning Algorithms with Focus on Feature Reductions. *Discrete Dynamics in Nature and Society*, *2022*, 5134356. https://doi.org/10.1155/2022/5134356

Sousa, C. F. F. (2023). *Categories' Churn: A Machine Learning Approach in Retail* Universidade do Minho].

Srivastava, S., & Sinha, S. (2024). Machine Learning Techniques: Predictive Modeling for Customer Churn in Telecommunications. *Nanotechnology Perceptions*, *20*(7), 1151-1166. https://doi.org/10.62441/nano-ntp.v20iS7.95

Sweidan, D., Johansson, U., Gidenstam, A., & Alenljung, B. (2022). *Predicting Customer Churn in Retailing* 2022 21st IEEE International Conference on Machine Learning and Applications (ICMLA), https://doi.org/10.1109/ICMLA55696.2022.00105

E-ISSN: 3041-8933

Taherkhani, L., Daneshvar, A., Amoozad Khalili, H., & Sanaei, M. R. (2023). Analysis of the Customer Churn Prediction Project in the Hotel Industry Based on Text Mining and the Random Forest Algorithm. *Advances in Civil Engineering*, *2023*, 6029121. https://doi.org/10.1155/2023/6029121

Tran, H., Le, N., & Nguyen, V. H. (2023). Customer Churn Prediction in the Banking Sector Using Machine Learning-Based Classification Models. *Interdisciplinary Journal of Information, Knowledge, and Management*, *18*. https://doi.org/10.28945/5086

Vemulapalli, G. (2024). AI-Driven Predictive Models Strategies to Reduce Customer Churn. *International Numeric Journal of Machine Learning and Robots*, *8*(8), 1-13.

Zelenkov, Y. A., & Suchkova, A. S. (2023). Predicting Customer Churn Based on Changes in Their Behavior Patterns. *Business Informatics*, *17*(1), 7-17. https://doi.org/10.17323/2587-814X.2023.1.7.17