KMAN PUBLISHING INSTITUTE
Journal Website

Article history:
Received 29 November 2024
Revised 04 December 2024
Accepted 20 December 2024
Published online 31 December 2024

Journal of Technology in Entrepreneurship and Strategic Management

In Press

E-ISSN: 3041-8585

Application of Machine Learning Methods in Analyzing Customer Preferences in Grocery Store Chains

Karim. Layegh Ahani[1], Younos. Vakil Alroaia[2*], Abolfazl. Danaei[3], Farshad. Faezi Razi[2]

[1] PhD Student in Marketing Management, Department of Management, Semnan Branch, Islamic Azad University, Semnan, Iran
[2] Associate Professor, Department of Management, Semnan Branch, Islamic Azad University, Semnan, Iran
[3] Associate Professor, Department of Media Management, Semnan Branch, Islamic Azad University, Semnan, Iran

*** Corresponding author email address**: Y.vakil@semnaniau.ac.ir

PROOF VERSION

A r t i c l e I n f o

Article type:
Original Research

How to cite this article:
Layegh Ahani, K., Vakil Alroaia, Y., Danaei, A., & Faezi Razi, F. (IN PRESS). Application of Machine Learning Methods in Analyzing Customer Preferences in Grocery Store Chains. *Journal of Technology in Entrepreneurship and Strategic Management.*

© 2024 the authors. Published by KMAN Publication Inc. (KMANPUB), Ontario, Canada. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) License.

A B S T R A C T

The primary objective of this research is to discover valuable patterns in customer data from grocery store chains to identify customer preferences and detect valuable customer churn. Using data mining and machine learning methods, valuable customers were identified based on RFM variables. This was achieved through clustering methods. Subsequently, valuable customers were classified using classification methods. The preferences of high-value customers were then identified through an association rule approach. Additionally, clustering and classification methods were employed to find useful patterns for analyzing the behavior of churned and non-churned customers. Following a series of data preparation and preprocessing operations, the data for each customer, along with their transactions, was determined. The initial data was based on each customer's transactions. However, after data preparation and preprocessing, datasets were created, capturing information specific to each customer. Using data mining and machine learning methods, alongside the three approaches of association rules, classification, and clustering, models and useful behavioral patterns were trained and constructed from the customer dataset of grocery store chains. Ultimately, a set of patterns and rules was developed to identify customer preferences, recognize valuable customers, and detect churned customers.

Keywords: *Customer preferences, machine learning, classification, clustering.*

Introduction

Customer preference analysis has emerged as a critical aspect of retail success, particularly for large-scale grocery store chains that must handle a wide variety of products and cater to different consumer segments (Chen et al., 2020; Zhou et al., 2020). Effectively pinpointing customer needs requires an understanding of both purchasing behaviors and consumption patterns, which can be extremely challenging given the massive volume of data. As machine learning algorithms become increasingly sophisticated, they provide robust tools for extracting meaningful insights from such data, thereby enhancing the ability of chain stores to deliver personalized offerings, improve satisfaction, and boost loyalty (Ram et al., 2017).

Alongside preference analysis, the issue of customer churn holds equal significance. Retailers face the risk of losing valuable patrons when dissatisfaction, competitive offerings, or unaddressed needs drive shoppers to other stores (Lalwani et al., 2021; Vafeiadis et al., 2015). By employing data-driven techniques—clustering, classification, and association rule mining—managers can anticipate churn, identify high-value customers, and determine the best interventions for retention (Mishra & Reddy, 2017). Therefore, the aim of this study is to propose an integrated machine learning framework that simultaneously addresses customer preference identification, high-value customer recognition, and churn prediction in the context of grocery store chains.

Methods and Materials

This study developed an analytical framework to process and extract insights from large-scale transaction data obtained from a chain of grocery stores. In the initial phase, a set of preprocessing steps was carried out to clean and unify the data. Tasks such as handling missing values, removing noise, and integrating multiple sources of customer transactions were completed to assure the reliability of downstream analysis.

Following data cleaning, the study built new features reflecting key behavioral indicators. Among these features were recency of the last purchase, transaction frequency, and monetary value. Pivot tables were used to transform transaction-level data into a customer-level dataset, where each row represented a unique customer, and each column captured a relevant feature (e.g., cumulative spending or number of visits). Once transformed, normalization techniques were adopted to handle differences in the scale of variables, ensuring appropriate inputs for clustering and classification methods.

Next, the research applied clustering to divide customers into meaningful groups. Iterative approaches with varying numbers of clusters were tested to optimize internal validity measures such as the Davies-Bouldin index. Candidate algorithms included k-means and self-organizing maps (SOM). In parallel, association rule mining was employed to detect frequent item sets and generate rules that highlighted co-purchase patterns within each cluster.

For classification tasks, multiple machine learning algorithms were explored to identify (1) whether a customer would buy a specific product, (2) whether a customer belonged to the high-value segment, and (3) whether a customer would churn. The algorithms tested included decision trees, naïve Bayes, logistic regression, random forests, k-nearest neighbors, support vector machines, and several neural network architectures. Hyperparameters, such as the number of hidden layers for neural networks or the number of estimators for ensemble methods, were refined through grid search and cross-validation.

PROOF VERSION

After model training, predictive performances were assessed using accuracy, precision, recall, and F1-score. Visualizations of decision boundaries, cluster distributions, and association rule networks were produced using tools like Tableau, SPSS Modeler, and RapidMiner. The final stage involved selecting the best algorithms for each task, integrating the outcomes into a business intelligence overview to support decision-makers in marketing, product assortment, and loyalty program design.

Findings and Results

Clustering revealed that two main segments were especially informative. One cluster encompassed a large proportion of high-value customers, while the other captured lower-value or one-time purchasers. A small subset of customers in the high-value cluster also displayed signs of churn, indicating that the store risks losing profitable clients if it does not act.

Association rule mining highlighted several frequent product sets with high support and confidence. Specific products emerged as leading indicators for co-purchasing behaviors, and the presence of these products in a customer's cart significantly correlated with higher monetary spending and longer-term store loyalty. Additionally, a few rules showed that certain product combinations served as predictors of customer churn, hinting at evolving preferences or dissatisfaction among loyal customers.

Classification models demonstrated varied performance. Decision trees, random forests, and neural networks generally exhibited strong accuracy in predicting whether a customer would buy a specified product or not. Moreover, for churn prediction among the high-value segment, naïve Bayes and logistic regression provided robust predictive scores, suggesting they capture underlying signals in the user's purchase history. Random forests and neural networks showed some advantages in terms of overall accuracy, but they often produced more complex models that reduced interpretability.

Analyses focusing on RFM-based classification confirmed that recency and frequency were closely connected to churn outcomes. Customers who had fewer transactions and long gaps since the previous purchase were more likely to stop shopping at the chain. Conversely, those with higher frequencies and recent visits contributed substantially to store revenue, making them prime targets for retention strategies.

Conclusion

The results underscore the importance of a comprehensive data-driven approach in retail management. By leveraging clustering methods, the study succeeded in capturing the distinct purchase patterns of high-value and low-value segments, revealing actionable insights for targeted marketing interventions. Additionally, association rules provided managers with a clear view of co-purchase patterns, guiding product recommendations and potential promotional bundling strategies.

Combining classification techniques with RFM variables enabled effective churn detection, highlighting that customers with extended time since their last purchase and fewer transactions were at higher risk of defection. Notably, the algorithms showing strong predictive power ranged from interpretable methods, such as decision trees, to more complex ensemble and neural network techniques that delivered higher predictive accuracy but required supplemental interpretive tools.

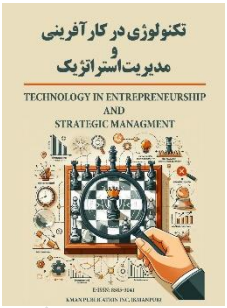
PROOF VERSION

E-ISSN: 3041-8585

Overall, the multi-pronged methodology outlined here allows for a 360-degree perspective on customer behavior, addressing preference analysis, high-value customer identification, and churn prediction in an integrated fashion. From a managerial standpoint, the findings guide resource allocation decisions, ensuring that marketing budgets are allocated to both retention of top spenders and revitalizing dormant accounts.

Future research could benefit from developing a decision support system that incorporates real-time streaming data, providing managers with immediate alerts on potential churn risks or emerging product preference trends. Moreover, exploring deep learning architectures for more nuanced models might reveal additional latent factors driving purchase decisions. By continually refining these machine learning and data mining techniques, grocery store chains can maintain a competitive edge in a rapidly evolving marketplace and ultimately enhance both customer satisfaction and business profitability.

PROOF VERSION



شاپای الکترونیکی: ۳۰۴۱-۸۵۸۵

تکنولوژی در کارآفرینی و مدیریت استراتژیک

در دست چاپ

HAMAN PUBLISHING INSTITUTE

وبسایت مجله

تاریخچه مقاله
دریافت شده در تاریخ ۹ آذر ۱۴۰۳
اصلاح شده در تاریخ ۱۴ دی ۱۴۰۳
پذیرفته شده در تاریخ ۳۰ دی ۱۴۰۳
منتشر شده در تاریخ ۱۱ دی ۱۴۰۳

بکارگیری روش‌های یادگیری ماشین در تحلیل ترجیحات مشتریان فروشگاه‌های زنجیره‌ای مواد غذایی

کریم لایق آهنی[۱]، یونس وکیل الرعایا[۲*]، ابوالفضل دانایی[۳]، فرشاد فائزی رازی[۲]

۱. دانشجوی دکتری مدیریت بازاریابی، گروه مدیریت، واحد سمنان، دانشگاه آزاد اسلامی، سمنان، ایران
۲. دانشیار گروه مدیریت، واحد سمنان، دانشگاه آزاد اسلامی، سمنان، ایران
۳. دانشیار گروه مدیریت رسانه ، واحد سمنان، دانشگاه آزاد اسلامی، سمنان، ایران

*ایمیل نویسنده مسئول: Y.vakil@semnaniau.ac.ir

اطلاعات مقاله

نوع مقاله

پژوهشی اصیل

نحوه استناد به این مقاله:

لایق آهنی، کریم.، وکیل‌الرعایا، یونس.، دلنایی، ابوالفضل. و فائزی رازی، فرشاد. (در دست چاپ). بکارگیری روش‌های یادگیری ماشین در تحلیل ترجیحات مشتریان فروشگاه‌های زنجیره‌ای مواد غذایی. *تکنولوژی در کارآفرینی و مدیریت استراتژیک*.

© ۱۴۰۳ تمامی حقوق انتشار این مقاله متعلق به نویسنده است. انتشار این مقاله به‌صورت دسترسی آزاد مطابق با گواهی (CC BY-NC 4.0) صورت گرفته است.

چکیده

هدف اصلی پژوهش کشف الگوهای ارزشمند در داده‌های مشتریان فروشگاه‌های زنجیره‌ای برای شناسایی ترجیحات و خروج مشتریان ارزشمند می‌باشد. با استفاده از روش‌های داده‌کاوی و یادگیری ماشین، مشتریان ارزشمند بر اساس متغیرهای RFM شناسایی شدند. این کار توسط روش‌های خوشه‌بندی صورت گرفت. سپس به طبقه‌بندی مشتریان ارزشمند پرداخته شد که این کار توسط روش‌های طبقه‌بندی صورت گرفت. در ادامه به شناسایی ترجیحات مشتریان با ارزش پرداخته که این کار توسط رویکرد قوانین وابستگی انجام شد. همچنین با استفاده از روش‌های طبقه‌بندی و خوشه‌بندی به یافتن الگوهای مفید برای شناسایی و تحلیل رفتار مشتریان خروجی و غیرخروجی پرداخته شد. بعد از یک مجموعه عملیات آماده‌سازی و پیش‌پردازش داده‌ها، اطلاعات هر مشتری به همراه تراکنش‌های آن‌ها مشخص شد. داده اولیه بر اساس تراکنش هر مشتری بود. لیکن بعد از عملیات آماده‌سازی و پیش‌پردازش داده‌ها، مجموعه داده‌هایی بدست آمد که مربوط به هر مشتری بوده و اطلاعات آن‌ها را ثبت کرده بود. با استفاده از روش‌های داده‌کاوی و یادگیری ماشین و سه رویکرد قوانین وابستگی، طبقه‌بندی و خوشه‌بندی، به آموزش و ساخت مدل‌ها و الگوهای رفتاری مفید در مجموعه داده مشتریان فروشگاه‌های زنجیره‌ای پرداخته شد. در نهایت، مجموعه‌ای از الگوها و قوانین برای شناسایی ترجیحات مشتری، تشخیص مشتریان ارزشمند و شناسایی مشتریان خروجی بدست آمد.

کلیدواژگان: *ترجیحات مشتری، یادگیری ماشین، طبقه بندی، خوشه بندی*

PROOF VERSION

مقدمه

تحقیقات نشان می‌دهد که یکی از مهم‌ترین گام‌ها در برآورده‌سازی نیازهای مشتری، تحلیل ترجیحات مشتری است (Arifin & Purnama, 2023). این موضوع از اهمیت اساسی در فروشگاه‌های زنجیره‌ای برخوردار است. زیرا مشتریان ترجیحات متفاوتی در خرید کالاها دارند. لذا باید یک ارتباط بین ویژگی‌های محصول یا خدمت و ترجیحات مشتری برقرار باشد (Chen et al., 2020). همچنین یکی از موضوعات مهم شناسایی نیازهای مشتری است. مقاله (Zhou et al., 2020) مشتریان فروشگاه بسیار زیاد بوده و شناسایی نیازهای آن‌ها در حجم بسیار زیاد داده بسیار سخت است. الگوریتم‌های یادگیری ماشین توانسته است نیازهای مشتری را شناسایی کند. تشخیص نیازهای مشتری می‌تواند به فروشگاه در پیش‌بینی خروج و ترجیحات مشتری کمک کند (Zandi, 2024; Zidehsaraei, 2024).

با استفاده از تحلیل ترجیحات مشتری می‌توان به سفارشی‌سازی و شخصی‌سازی مشتریان پرداخت. در این خصوص می‌توان از خوشه‌بندی استفاده کرد. زیرا هر خوشه، ویژگی‌های خاص خود را در خصوص مشتریان دارد (Zhou et al., 2020). شناسایی ترجیحات مشتری از این جهت اهمیت دارد که می‌توان با شناخت ترجیحات مشتری به اتخاذ تصمیمات بازاریابی و حتی روش ارائه محصولات و خدمات پرداخت (Ram et al., 2017).

یکی دیگر از موضوعات مهم در مدیریت مشتریان، پیش‌بینی میزان خروج مشتری است. این موضوع به طور فزاینده‌ای در تحقیقات اخیر مورد بررسی قرار گرفته است. به طور نمونه مقاله (Lalwani et al., 2021) با استفاده از روش‌های مشهور یادگیری ماشین به پیش‌بینی خروج مشتریان پرداخته است. مشتریان فروشگاه نیز ممکن است اقدام به خروج از فروشگاه داشته باشند. لذا بایستی از الگوهای داده‌کاوی و یادگیری ماشین برای شناسایی رفتار مشتریانی که از فروشگاه خارج می‌شود یا وفادار به فروشگاه هستند، استفاده نمود.

در مقاله (Vafeiadis et al., 2015) به مقایسه ۵ الگوریتم یادگیری ماشین برای پیش‌بینی خروج مشتری پرداخته شد که عبارتند از شبکه عصبی، درخت تصمیم، ماشین بردار پشتیبان، نیو بیز و رگرسیون لجستیک. همچنین عملکرد الگوریتم بوستینگ نیز بررسی شد. همچنین پارامترهای مدل‌ها نیز تغییر یافت. نتایج نشان داد که استفاده از بوستینگ توانسته است عملکرد مدل را بهبود دهد. در مقاله (Mishra & Reddy, 2017) به مقایسه چند الگوریتم یادگیری ماشین برای پیش‌بینی خروج مشتری پرداخته شد. از روش‌های تجمعی *ensemble* مانند جنگل تصادفی، *bagging* و *boosting* استفاده شد و نتایج نشان داد که روش جنگل تصادفی بهترین عملکرد را دارد. از شاخص‌های متعددی برای ارزیابی مدل‌ها استفاده شد. در مقاله (Nelaturi & Devi, 2019) با استفاده از یادگیری عمیق به ارائه یک سیستم پیشنهاددهنده بر اساس ترجیحات مشتری پرداخته شد و نتایج نشان داد این روش می‌تواند به ارائه پیشنهادات بر اساس ترجیحات مشتری بپردازد.

در مقاله (Dullaghan & Rozaki, 2017) به استفاده از روش‌های داده‌کاوی در پیش‌بینی خروج مشتری پرداخته شد. نکته مهم در این مقاله استفاده از روش *RFM* برای تحلیل خروج مشتری بود. همچنین به بخش‌بندی مشتریان نیز پرداخته شد. از روش‌های نیو بیز و درخت تصمیم برای طبقه‌بندی مشتریان استفاده شد. در مقاله (Verma, 2020) به مقایسه الگوریتم‌های شبکه‌های عصبی مبتنی بر یادگیری عمیق برای پیش‌بینی خرید مشتریان در یک شرکت خرده‌فروشی پرداخته شد. مدل‌ها بر اساس داده‌های شرکت خرده‌فروشی الکترونیکی ارزیابی شد. همچنین از سری‌های زمانی برای پیش‌بینی خرید مشتریان بهره گرفته شد. در مقاله (Selvakanmani et al., 2020) با استفاده از روش یادگیری تجمیعی به پیش‌بینی خروج مشتری در یک شرکت مخابراتی پرداخته شد. البته در این مقاله اشاره شده است که برخی مشتریان ارزش کمی دارند. اگر این مشتریان شرکت را ترک کنند شاید به گونه‌ای به سود شرکت باشد. یا اینکه برنامه‌هایی را اتخاذ کرد تا

PROOF VERSION

ارزش این مشتریان افزایش یابد. همچنین در این مقاله از استراتژی کاهش ابعاد نیز استفاده شده است. نتایج نشان داد الگوریتم‌های تجمیعی می‌توانند به خوبی به طبقه‌بندی مشتریان بپردازند.

در مقاله (Parre Amarnath, 2018) به استفاده از یک تکنیک خوشه‌بندی دو لایه برای بخش‌بندی و تحلیل رفتار مشتریان موبایل پرداخته شد. در این مقاله در ابتدا در لایه اول خوشه‌بندی ارزش مشتریان شناسایی شد. سپس در لایه دوم به گروه‌بندی مشتریان به منظور ارائه خدمات در آینده پرداخته شد و این کار بر اساس ویژگی‌های آن‌ها در هر خوشه انجام شد. مقاله (Chen et al., 2020) به ارائه یک مدل ترکیبی شبکه عصبی و روش معادلات ساختاری برای تحلیل ترجیحات مشتری پرداخت. ترجیحات مشتری دارای روابط غیرخطی و پیچیده است که می‌توان در حجم زیاد داده به شناسایی این روابط پرداخت. مقاله (Lalwani et al., 2021) به استفاده از روش‌های یادگیری ماشین برای ارائه یک سیستم پیش‌بینی خروج مشتری پرداخت. در یک مطالعه (Ram et al., 2017)به ارائه تکنیک‌های داده‌کاوی برای پیش‌بینی ترجیحات مشتری پرداخته شد. در این مقاله از روش خوشه‌بندی برای توصیف مشتریان در هر خوشه استفاده شد. سپس از روش بیزین برای پیش‌بینی ترجیحات مشتری استفاده شد.

در مقاله(Zhuravlev et al., 2020) با استفاده از روش‌های یادگیری ماشین به ارائه یک سیستم پیشنهاددهنده بر اساس ترجیحات مشتری پرداخته شد. ویژگی‌های محصولاتی که قبلا توسط مشتریان خریداری شده بود، مورد بررسی قرار گرفت. برای محصولا نرخ ترجیح تعیین شد. همچنین به بررسی این موضوع پرداخته شد که مشتریان در گذشته چه رفتاری در خصوص محصولات داشته‌اند. از تکنیک خوشه‌بندی برای بخش‌بندی محصولات استفاده شد. مقاله (Jing et al., 2017) به ارائه یک سیستم شخصی‌سازی‌شده ارائه پیشنهاد بر اساس کاوش ترجیحات مشتری پرداخت. در این خصوص ترجیح مشتری در ابتدا مدل شد و سپس بر اساس این مدل به کاوش ترجیح مشتری پرداخته شد. مقاله (Jiang et al., 2019) به ارائه یک سیستم ترجیح مشتری پویا با استفاده از سری زمانی و فازی و کاوش عقاید پرداخت. مقاله (Yan et al., 2020) با استفاده از داده‌کاوی به شناسایی مدل رفتار انتخاب مشتری در اینترنت اشیا پرداخت. در این خصوص از تکنیک درختان تصمیم‌گیری استفاده شد. مقاله (Kim et al., 2021) به ارائه یک مدل یادگیری عمیق ترکیبی برای تحلیل رفتار خرید مجدد مشتریان پرداخت. نتایج نشان داد که این مدل با بیش از ۹۰ درصد دقت می‌تواند به پیش‌بینی رفتار خرید مجدد مشتریان بپردازد.

با توجه به تحقیقات پیشین، هر یک از تحقیقات به طور جداگانه به شناسایی ترجیحات مشتری، تعیین مشتریان ارزشمند و شناسایی مشتریان خروجی پرداخته‌اند. لیکن هر کسب و کاری و از جمله فروشگاه زنجیره‌ای نیاز به این دارد که هر سه موضوع را به طور همزمان پیگیری کند. یک فروشگاه زنجیره‌ای باید بتواند علاوه بر اینکه مشتریان ارزشمند خود را شناسایی می‌کند، بررسی کند که کدام یک از مشتریان در حالی که ارزشمند هستند، وفادار نیز می‌باشند. همچنین لازم است برای این دسته از مشتریان، ترجیحات آن‌ها شناسایی شود. لذا پژوهش حاضر قادر است به فروشگاه‌های زنجیره‌ای در شناسایی مشتریان ارزشمند و وفادار و تعیین ترجیحات آن‌ها به طور همزمان کمک کند.

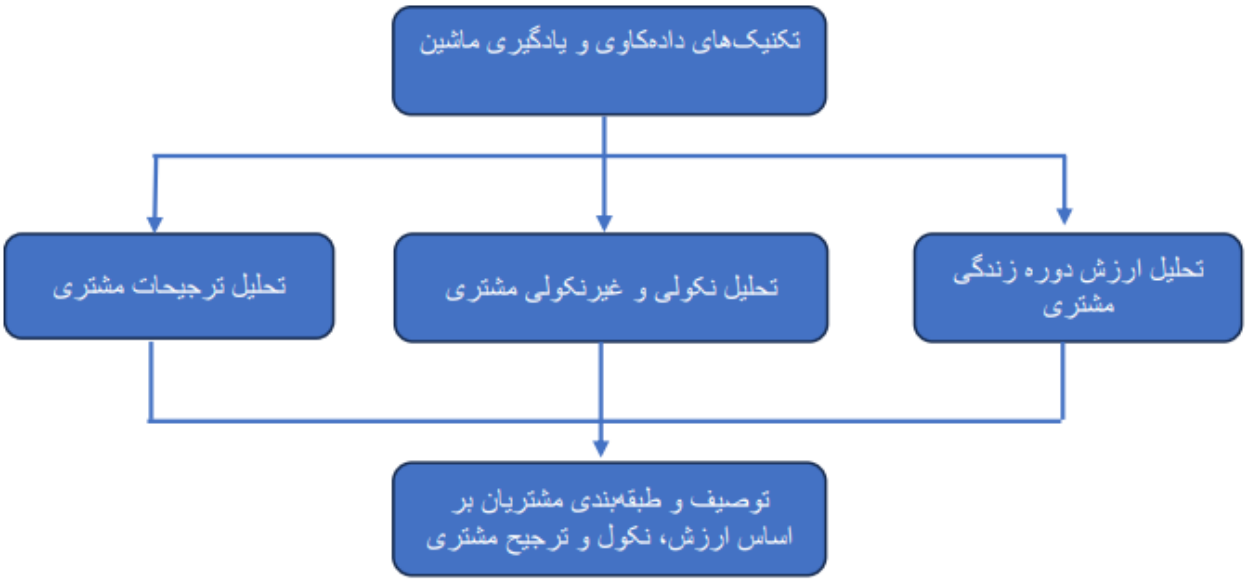
شکل زیر مدل مفهومی پژوهش را نشان می‌دهد.

PROOF VERSION

E-ISSN: 3041-8585

شکل ۱

مدل مفهومی پژوهش



همان‌طور که مشاهده می‌شود، در ابتدا از تکنیک‌های داده‌کاوی و یادگیری ماشین برای تحلیل مجموعه داده‌های مشتریان استفاده می‌شود. با توجه به تحقیقات مختلف اشاره‌شده در این فصل، می‌بایست بر اساس داده‌ها و مفاهیمی که در فصل سوم اشاره می‌شود به انتخاب تکنیک‌های مناسب یادگیری ماشین و داده‌کاوی برای این پژوهش پرداخت. بعد از بکارگیری الگوریتم‌های یادگیری ماشین، الگوهای ارزشمند در مجموعه داده شناسایی می‌شود. این الگوها از سه منظر مورد تحلیل و استفاده واقع می‌شوند. برخی از الگو برای تحلیل ارزش مشتری و شناسایی مشتریان ارزشمند مفید هستند. از این الگوها می‌توان برای بخش‌بندی و پیش‌بینی مشتریان ارزشمند استفاده نمود. برخی الگوه برای تحلیل نکول مشتری مفید هستند. به عبارت دیگر مشتریان بر اساس میزان نکولی‌بودن می‌توانند بخش‌بندی شوند. همچنین می‌توان به شناسایی و پیش‌بینی نکولی‌بودن مشتریان آینده پرداخت. در نهایت برخی مدل‌ها می‌توانند برای شناسایی ترجیحات مشتری مناسب باشند.

استفاده از این سه مفهوم به طور همزمان می‌تواند به شناسایی و پیش‌بینی و توصیف مشتریان بر اساس ارزش آن‌ها و میزان نکولی‌بودن و ترجیح آن‌ها منجر شود. لذا شرکت‌ها می‌توانند از مدل پیشنهادی این پژوهش برای شناسایی و تعیین ویژگی‌های مهم مشتریان خود استفاده کنند.

روش پژوهش

روش‌شناسی پژوهش بر اساس روش‌های داده‌کاوی و یادگیری ماشین انجام می‌شود.

۱. در ابتدا به تحلیل فروشگاه‌های زنجیره‌ای پرداخته می‌شود. سپس مسئله اصلی و اهداف اصلی برای بکارگیری داده‌کاوی تبیین می‌شود. هدف اصلی شناسایی ترجیحات مشتریان ارزشمند و خروج آن‌ها است. همچنین مشتریان باارزش شناسایی می‌شوند.

۲. در مرحله بعد به جمع‌آوری داده‌های مشتریان پرداخته می‌شود. این مجموعه داده در بانک‌های اطلاعاتی مشتریان فروشگاه قرار دارد.

PROOF VER

۳. سپس به پیش‌پردازش داده‌ها پرداخته می‌شود. روش‌های مختلف پیش‌پردازش داده‌ها از جمله برخورد با داده‌های مفقود، نویز، و ترکیب و تجمیع داده‌ها و یکپارچه‌سازی آن‌ها صورت می‌گیرد. در صورت لزوم از روش‌های استخراج ویژگی و انتخای ویژگی و کشف نقاط آنومالی استفاده می‌شود.

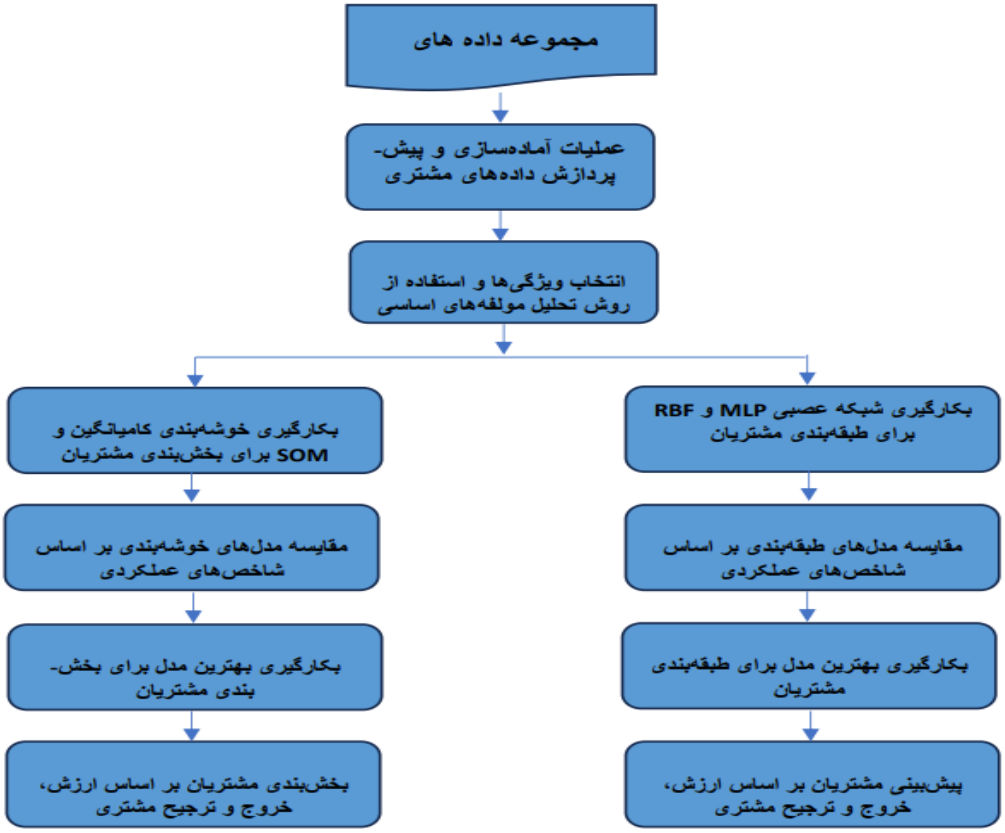
۴. در مرحله چهارم به مدل‌سازی پرداخته می‌شود. در ابتدا از روش‌های خوشه‌بندی به بخش‌بندی مشتریان بر اساس ترجیحات و خروج و ارزش مشتری پرداخته می‌شود. یعنی مشتریانی که شباهت بیشتری به هم دارند، ترجیحات و دلایل خروج آن‌ها نیز ممکن است مشابه باشد. برای خوشه‌بندی می‌توان از روش‌های مختلف استفاده نمود. به طور نمونه *k-means* و روش سلسله‌مراتبی و روش *DBSACN* و روش *SOM*. بعد از خوشه‌بندی مشتریان می‌توان برای هر خوشه به طبقه‌بندی ترجیحات و خروج مشتری با استفاده از شبکه‌های عصبی همچون *MLP*، *RBF* و روش‌های دیگر یادگیری ماشین پرداخت.

۵. نتایج با هم مقایسه می‌شوند و بهترین عملکرد برای شناسایی مشتری ارزشمند و پیش‌بینی ترجیح مشتری و خروج آن تعیین می‌شود.

۶. در مرحله آخر با استفاده از نتایج بدست‌آمده به ارائه پیشنهادات و استراتژی‌ها در خصوص ترجیحات مشتری و خروج آن‌ها و همچنین تدوین استراتژی برای مشتریان باارزش پرداخته می‌شود.

شکل ۲

مراحل انجام پژوهش



E-ISSN: 3041-8585

در مرحله پیش‌پردازش داده‌ها به انجام برخی عملیات مهم برای آماده‌سازی و پیش‌پردازش داده‌ها پرداخته می‌شود. در این مرحله از روش‌های متعدد انتخاب ویژگی‌ها استفاده می‌شود. این الگوریتم‌ها به شناسایی ویژگی‌های مهم در مجموعه داده می‌پردازند.

برای اجرای مدل پیشنهادی از یک مجموعه داده مشتریان فروشگاه‌های زنجیره‌ای استفاده می‌شود. مجموعه داده مربوط به مشتریان فروشگاه‌های زنجیره‌ای کاشان امیرکبیر شماره ۲۷ است. با توجه به اینکه اغلب مشتریان از فروشگاه نزدیک محل خود خریداری می‌کنند، بهتر است به تحلیل مشتریان و رفتار آن‌ها بر اساس هر فروشگاه به طور جداگانه پرداخته شود. ترکیب اطلاعات چند فروشگاه ممکن است نتواند الگوهای ارزشمندی توسط روش‌های یادگیری ماشین و داده‌کاوی ارائه دهد. مجموعه داده اولیه شامل تعداد ۱۳ متغیر یا ویژگی و همچنین تعداد ۵۴۱۰۳۲ رکورد یا *instance* است. هر رکورد اشاره به یک تراکنش دارد. جدول زیر چند نمونه از رکوردهای تکمیل‌شده از داده‌ها را ارائه می‌دهد.

جدول ۱

چند نمونه از رکوردهای تکمیل‌شده از داده‌ها

نام فروشگاه	نام کالا	بارکد کالا	شماره مشتری	فصل	ماه	روز هفته	روز	مبلغ فروش خالص (ریال)	آخرین تاریخ فروش (تاریخ)	فاصله از آخرین فروش (روز)	مجموع تعداد کالای فروش	تعداد فاکتور فروش
۲۷ کاشان امیرکبیر	پاک کره۲۵ گرمی	۶۲۶۰۰۶۳۲۰۰۸۲۱	۹۰۱۰۲۲۲۰۱۸	تابستان	مرداد	پنجشنبه	۲۰	۶۱,۰۰۹	۱۴۰۱۰۵۲۰	-	۱	۱
۲۷ کاشان امیرکبیر	مایع لباسشویی پلی واش اکتیو ۲۵۰۰ گرمی ۶ عددی بنفش	۶۲۶۲۸۲۵۸۰۵۶۹۵	۹۰۱۰۲۲۲۰۱۸	تابستان	مرداد	پنجشنبه	۲۰	۸۷۸,۶۴۲	۱۴۰۱۰۵۲۰	-	۱	۱
۲۷ کاشان امیرکبیر	مینو ترد ۱۲۵ گرم ۲۵ عددی	۶۲۶۰۱۰۰۱۰۳۳۰۶	۹۰۱۰۲۲۲۰۱۸	تابستان	مرداد	پنجشنبه	۲۰	۹۱,۷۴۳	۱۴۰۱۰۵۲۰	-	۱	۱
۲۷ کاشان امیرکبیر	پرسیل پودر لباسشویی ماشینی لوندر با پاکیزگی عمیق ۵۰۰ گرمی	۶۲۶۰۱۰۵۰۰۴۹۷۴	۹۰۱۰۲۲۲۰۱۸	تابستان	مرداد	پنجشنبه	۲۰	۱۴۲,۸۲۴	۱۴۰۱۰۵۲۰	-	۱	۱
۲۷ کاشان امیرکبیر	کاله کره پاستوریزه ۲۵ گرمی قالبی	۶۲۶۰۱۶۱۵۳۹۹۴۶	۹۰۱۰۲۲۲۰۱۸	تابستان	مرداد	پنجشنبه	۲۰	۶۰,۸۵۳	۱۴۰۱۰۵۲۰	-	۱	۱
۲۷ کاشان امیرکبیر	غنچه پلاس روغن مایع آفتابگردان توکوفرول ۱۳۵۰ گرمی	۶۲۶۰۱۰۱۸۰۲۶۵۹	۹۰۱۰۲۲۲۰۱۸	تابستان	مرداد	پنجشنبه	۲۰	۹۶۳,۹۰۰	۱۴۰۱۰۵۲۰	-	۱	۱
۲۷ کاشان امیرکبیر	ساعی روغن سرخ کردنی کم جذب ۸۱۰ گرمی	۶۲۶۰۲۹۵۱۰۱۷۰۵	۹۰۱۰۲۲۲۰۱۸	تابستان	مرداد	پنجشنبه	۲۰	۶۱۳,۰۴۰	۱۴۰۱۰۵۲۰	-	۱	۱

از ترکیب ویژگی‌های فصل، ماه، و روز می‌توان به ویژگی R یعنی recency در RFM رسید. یعنی آخرین تراکنش در چه زمانی بوده است. اگر مبنای تاریخ را اول فروردین ۱۴۰۱ قرار دهیم، می‌توان به ساخت ویژگی R اقدام کرد به این صورت که روز تهیه این تراکنش‌ها به صورت روز مبنا است و برای هر تراکنش مقدار R از این تاریخ محاسبه می‌شود. مثلا اگر مشتری در آخرین تراکنش، دو روز پیش به فروشگاه رجوع کرده است، مقدار R برابر با ۲ است.

همچنین از ترکیب این سه ویژگی می‌توان به ویژگی برون رفت مشتری رسید.

یکی از تکنیک‌های مشهور در تعیین مشتریان ارزشمند، تحلیل RFM است که در این تحلیل، منظور از R همان زمان آخرین تراکنش مشتری است. لذا می‌توان این ویژگی را به عنوان R در نظر گرفت. هر قدر خرید مشتری در تاریخ نزدیک‌تری نسبت به روز آخر شهریور صورت گرفته باشد، بهتر است. لذا هر قدر مقدار این متغیر بیشتر باشد، بهتر است. البته می‌توان این مقدار را به جای فروردین از آخر

PROOF VERSION

شهریور حساب کرد و لذا هر قدر مقدار این ویژگی کمتر باشد، بهتر است. پس هر قدر مقدار R کمتر باشد، نشان می‌دهد که مشتری اخیرا تراکنشی را انجام داده است

پس از ویژگی روز می‌توان علاوه بر متغیر R به عنوان ویژگی خروج مشتری یا عدم خروج مشتری نیز استفاده کرد.

یکی از تکنیک‌های برای شناسایی ارزش هر مشتری یا مشتری ارزشمند، تکنیک تحلیل RFM است. یکی از متغیرها در این تکنیک، متغیر M است که اشاره ارزش مالی تراکنش مشتری دارد. هر قدر مقدار این متغیر بیشتر باشد، نشان می‌دهد که تراکنش اهمیت بیشتری دارد. پس مشتری که دارای تراکنش‌های بالا در مقدار M باشد، می‌تواند به عنوان یک مشتری ارزشمند بالقوه تعیین شود. در این مجموعه داده نیز اگر برای هر تراکنش مقدار متغیر مبلغ فروش خالص پس از برگشتی را داشته باشیم، لذا آن تراکنشی ارزشمندتر است که مقدار این متغیر در آن بیشتر باشد

بعد از آماده‌سازی و پیش‌پردازش داده‌ها و بر اساس مدل پیشنهادی به استفاده از روش‌های یادگیری ماشین و داده‌کاوی برای تعیین ترجیحات مشتری، ارزش مشتری و خروج مشتری پرداخته می‌شود. تا اینجا تعداد رکوردها برابر با ۵۴۱۰۱۳ تراکنش و تعداد شش متغیر یا ویژگی شد. در ادامه بایستی با استفاده از عملیات pivot برای هر مشتری یک رکورد تعریف کنیم. یعنی مجموعه داده را طوری تغییر دهیم که سطرها نشان دهنده مشتریان و ستون‌ها نشان‌دهنده متغیرهای مربوط به هر مشتری باشد. برای این کار از نرم‌افزار Tableau استفاده می‌شود. البته قبل از آن برای اینکه مقادیر ویژگی ماه تبدیل به عدد شود، از روش فیلترکردن استفاده کرده و مثلا فروردین را برابر با عدد یک قرار می‌دهیم و تا شهریور که برابر با عدد شش است.

یافته‌ها

در این بخش به ارائه سه روش برای خوشه‌بندی مشتریان پرداخته می‌شود. هدف این سه روش بخش‌بندی مشتریان بر اساس ترجیحات مشتری، ارزش مشتری، و خروج مشتری است. می‌توان از تحلیل مولفه اساسی برای کاهش ابعاد سه متغیر RFM استفاده نمود. البته قبل از بکارگیری این روش، باید داده‌ها را نرمال کرد. سه روش پیشنهادی برای خوشه‌بندی به قرار زیر است.

پارامترهای مدل خوشه‌بندی کامیانگین و SOM به صورت جدول زیر است.

۱. یکی از روش‌ها این است که مشتریان را بر اساس ترجیحات آن‌ها خوشه‌بندی کنیم. ترجیحات مشتری در این روش، منظور خرید کالاها است. مثلا ممکن است در یک خوشه مشتریان الگوی خرید مشابهی داشته باشند. لذا ابتدا هر یک از کالاها را به عنوان یک ویژگی لحاظ کرده و سپس خرید یا عدم خرید را توسط هر مشتری تعیین می‌کنیم. در نهایت این کالاها به عنوان ویژگی‌های مجموعه داده شناسایی می‌شوند. با انجام عمل خوشه‌بندی می‌توان به تعیین قوانین انجمنی در هر خوشه اقدام نمود و الگوهایی مکرر خرید را در هر خوشه شناسایی کرد. در زیر نتایج حاصل از روش خوشه‌بندی کامیانگین و SOM ارائه شده است. همچنین می‌توان بر اساس متغیر RFM و متغیر خروج مشتری به شناسایی ترجیحات مشتریان پرداخت. مثلا اینکه مشتریان ارزشمند دارای چه ترجیحاتی هستند یا اینکه مشتریان خروجی چه ترجیحاتی دارند.

در ابتدا به خوشه‌بندی ترجیحات مشتریانی پرداخته می‌شود که نسبت به خوشه اول باارزش‌تر هستند و همچنین مشتریانی که خروج ندارند.

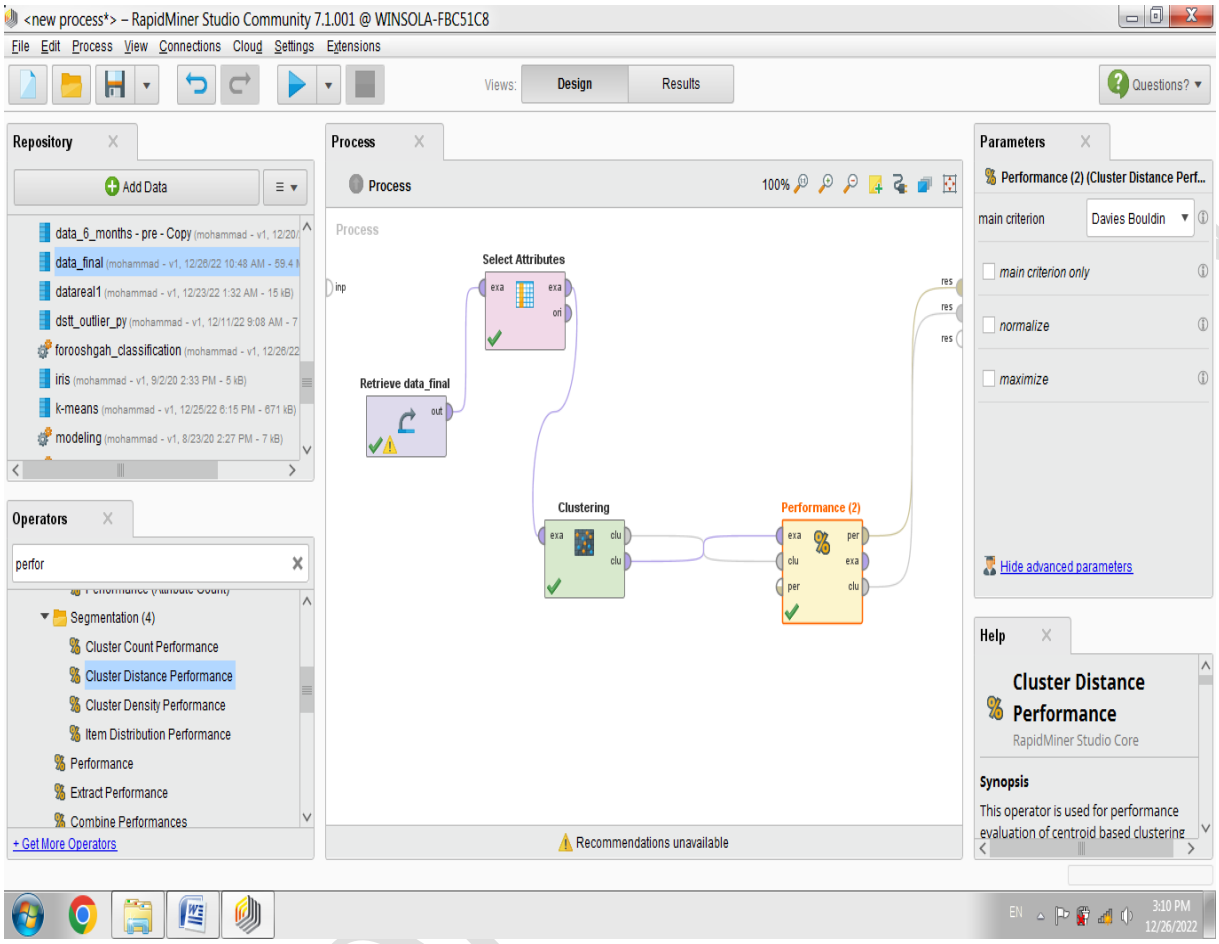
مراحل انجام کار به صورت زیر است:

PROOFVERSION

E-ISSN: 3041-8585

شکل ۳

مراحل کار در نرم افزار



در تعداد دو خوشه شاخص دیویس بولدین برابر با ۴.۹۱- است.

در تعداد خوشه ۳ مقدار این شاخص برابر با ۳.۴۶- است.

در تعداد خوشه ۴ مقدار این شاخص برابر با ۴.۸۴- است.

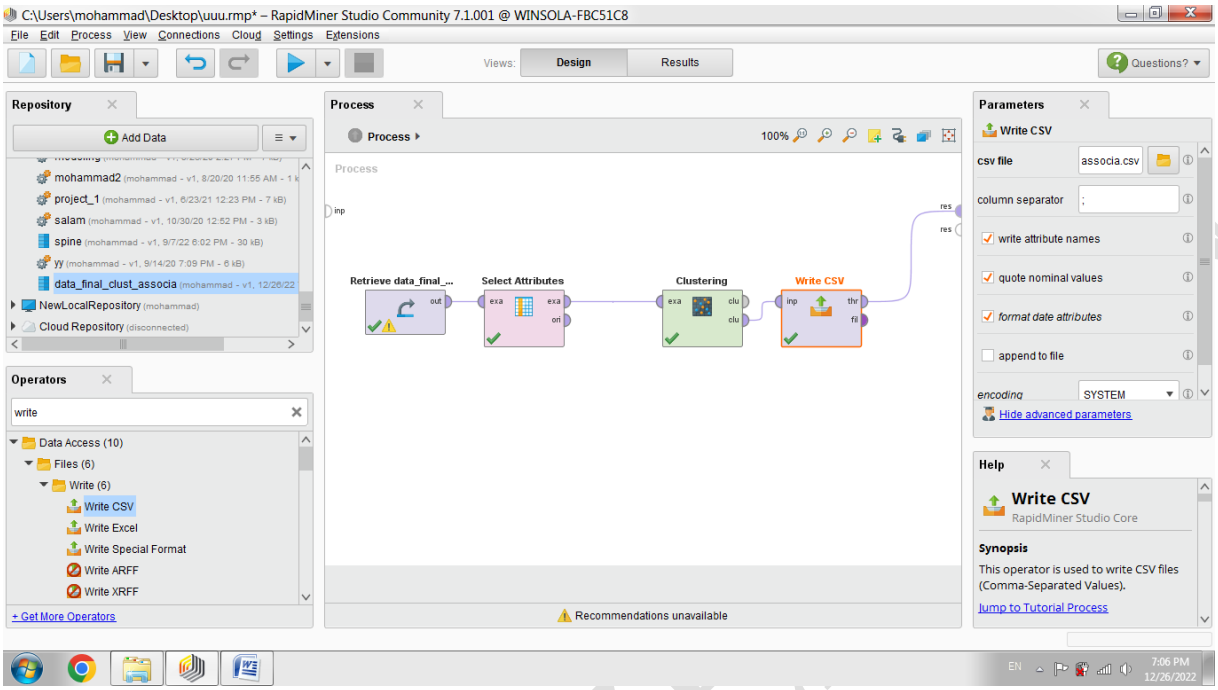
لذا بهتر است تعداد خوشه‌ها برابر با دو خوشه باشد. بعد از تعیین تعداد خوشه‌ها به دو خوشه می‌توان به تعیین قوانین وابستگی در هر خوشه اقدام نمود. تعداد مشتریان در هر خوشه به صورت زیر است.

تعداد مشتریان در خوشه اول برابر با ۱۶۴۴ و در خوشه دوم برابر با ۲۱۱ مشتری است. به طور نمونه به تعیین قوانین وابستگی برای خوشه دوم پرداخته می‌شود. البته باید این مجموعه داده را ایجاد کنیم. لذا مشتریان خوشه دوم را فیلتر می‌کنیم. در این خصوص از عملگر write csv استفاده می‌شود.

PROOF

شکل ۴

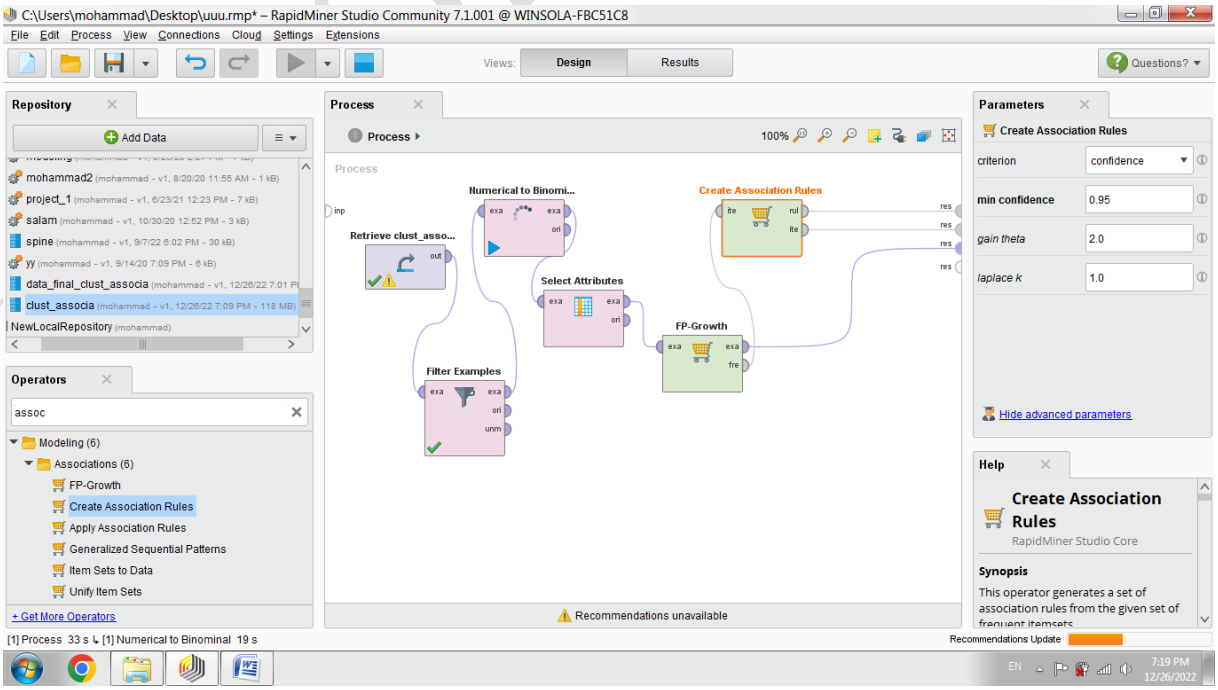
مرحله دوم تحلیل در نرم افزار



در ادامه مجموعه داده جدید را برای ایجاد قوانین وابستگی بکار می‌بریم. برای این کار از فرایند زیر استفاده می‌کنیم.

شکل ۵

افزودن مجموعه داده جدید

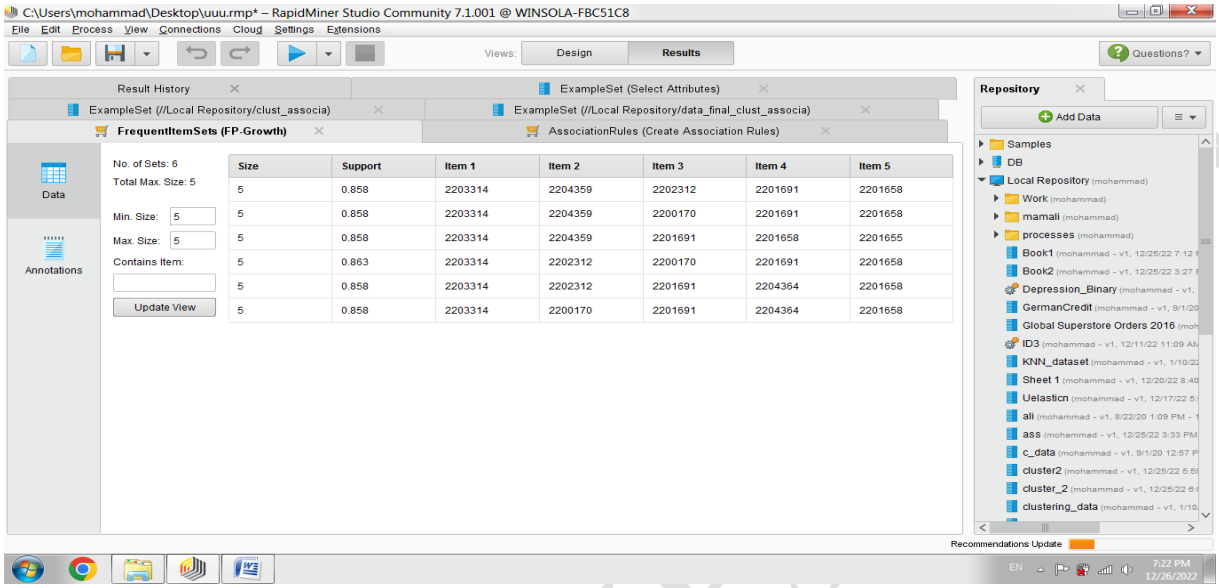


E-ISSN: 3041-8585

از filter example برای فیلتر کردن خوشه دوم استفاده شد. همچنین حداقل مقدار پشتیبان و اطمینان برابر با ۰.۹۵ است.

شکل ۶

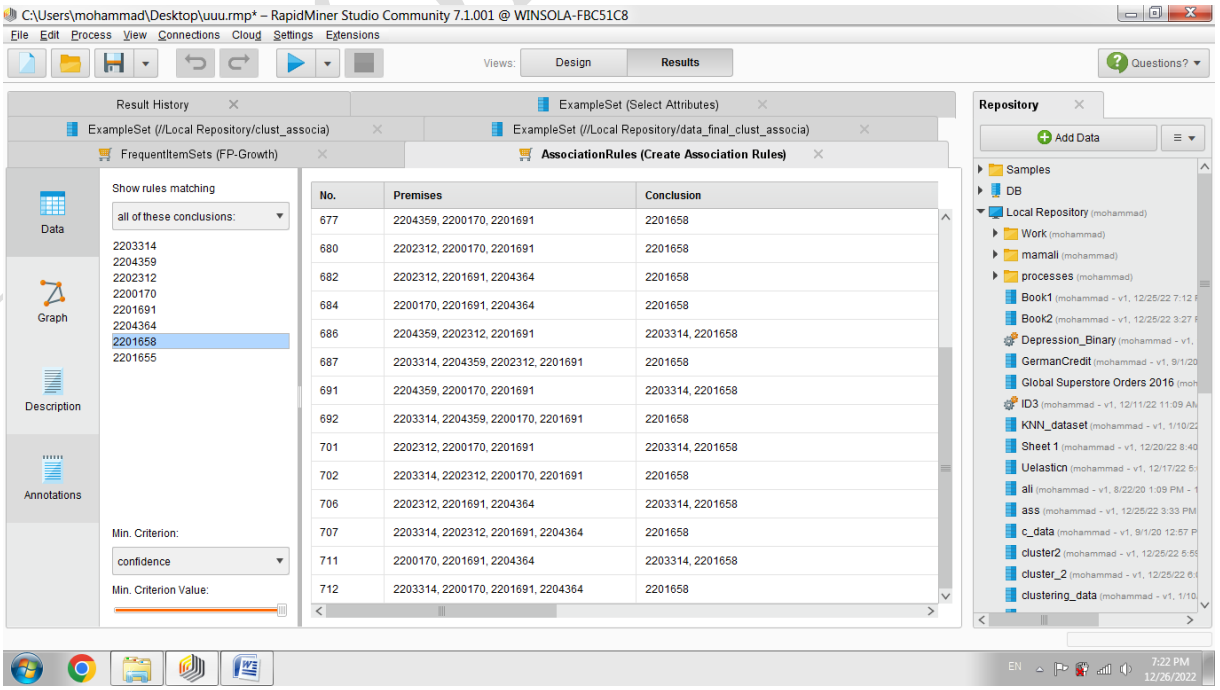
نمونه اقلام مکرر (بعد از فیلتر شدن خوشه دوم)



همچنین برخی الگوهای مکرر به صورت زیر است:

شکل ۷

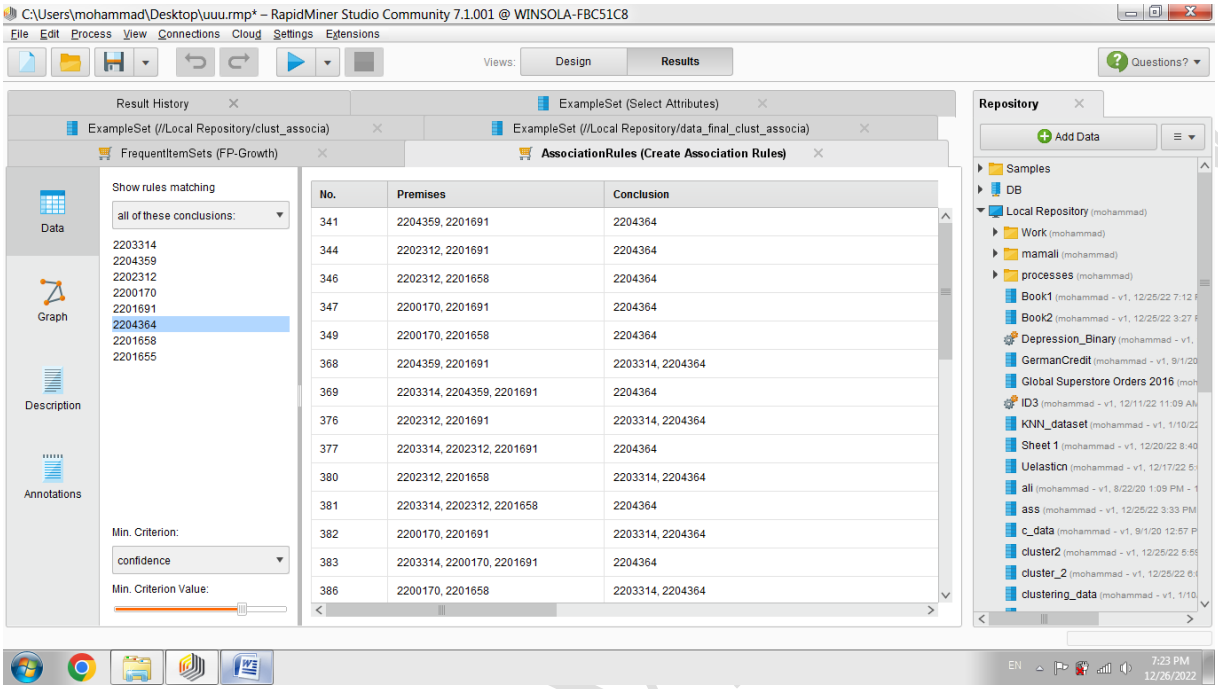
نمونه الگوهای مکرر بعد از فیلترسازی



به طور نمونه برخی الگوها برای کالای ۲۲۰۴۳۶۴ به صورت زیر است.

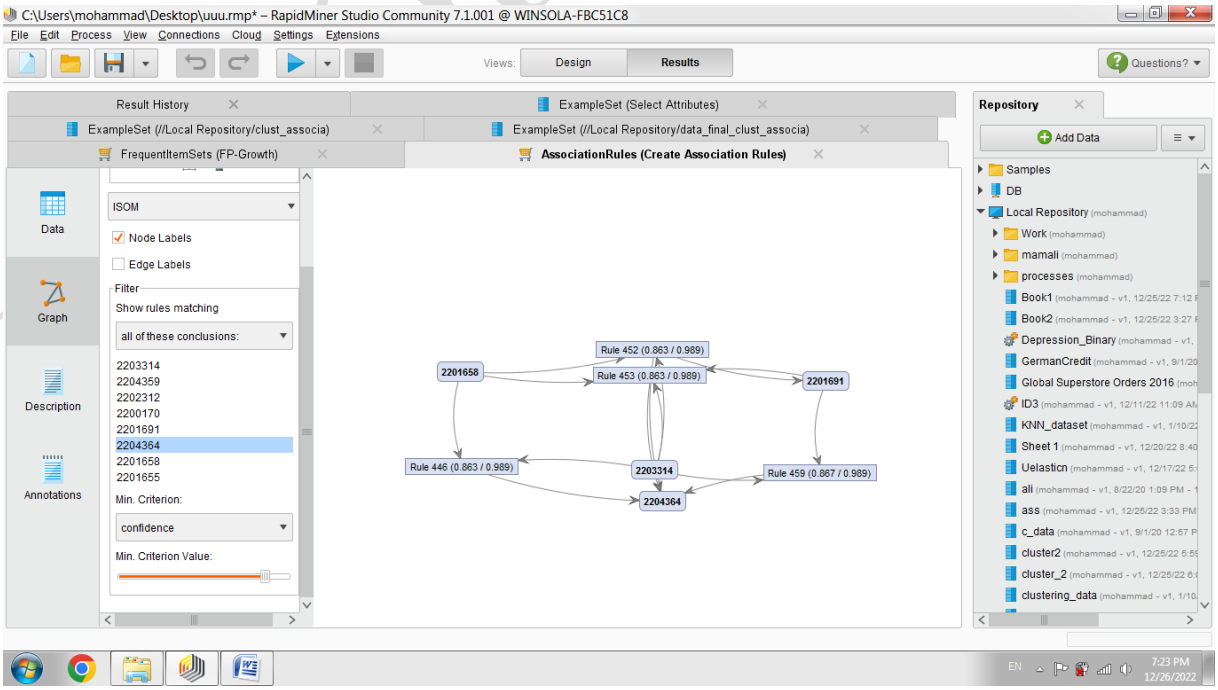
شکل ۸

نمونه الگوها برای کالای ۲۲۰۴۳۶۴



شکل ۹

نمایش گرافیکی برخی از قوانین وابستگی



E-ISSN: 3041-8585

به عنوان یک کار دیگر می‌توانیم قوانین وابستگی را مربوط به مشتریانی که در خوشه دوم یعنی خوشه ارزشمند بوده ولی خروج داشته‌اند پیدا کنیم.

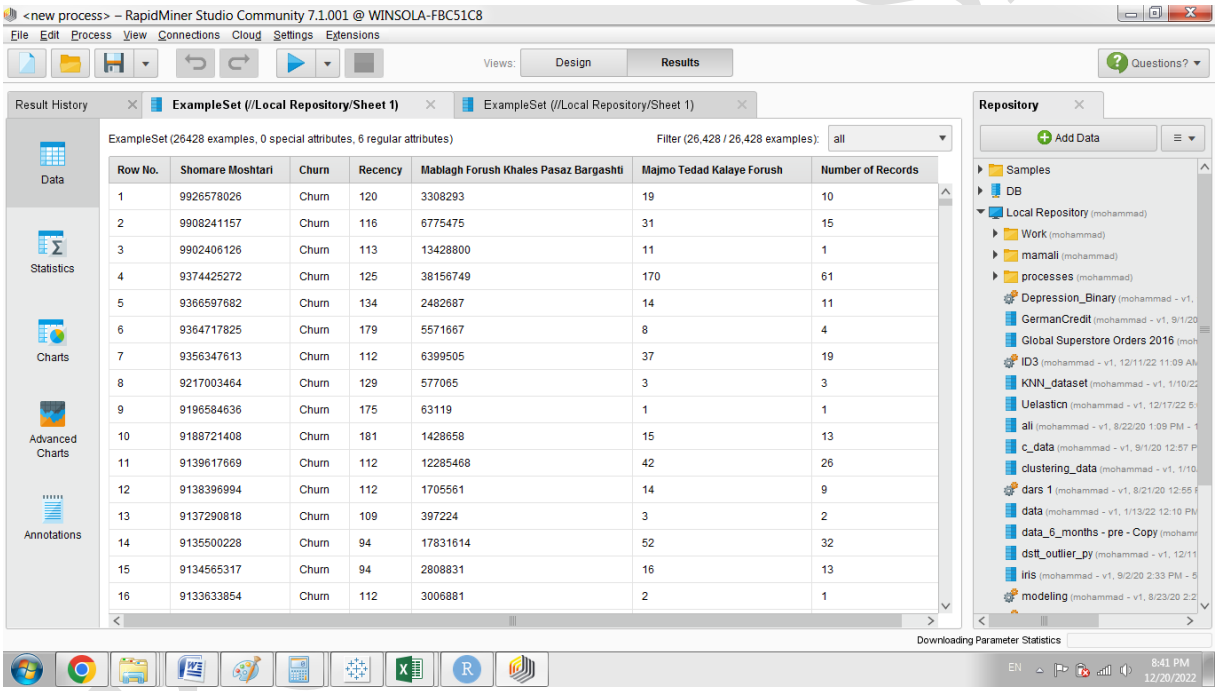
البته در این بخش می‌توانیم از روش‌های دیگر خوشه‌بندی مثل SOM و DBSCAN و سلسله‌مراتبی نیز استفاده کنیم.

۲. روش بعدی، خوشه‌بندی مشتریان بر اساس متغیرهای RFM است. بعد از خوشه‌بندی به تعیین میزان ارزش هر خوشه پرداخته می‌شود. برای محاسبه ارزش هر خوشه می‌توان به جمع نرمال مقادیر هر یک از سه متغیر RFM برای هر مشتری پرداخت. برای تعیین ارزش هر خوشه به جمع ارزش کل مشتریان پرداخته و سپس نتیجه را بر تعداد مشتریان در خوشه می‌توان تقسیم نمود. جدول زیر به ارائه نتایج حاصل از اجرای الگوریتم کامیانگین و SOM می‌پردازد.

در ابتدا داده‌ها را فراخوانی می‌کنیم. شکل زیر داده‌ها را نشان می‌دهد:

شکل ۱۰

فراخوانی داده‌ها



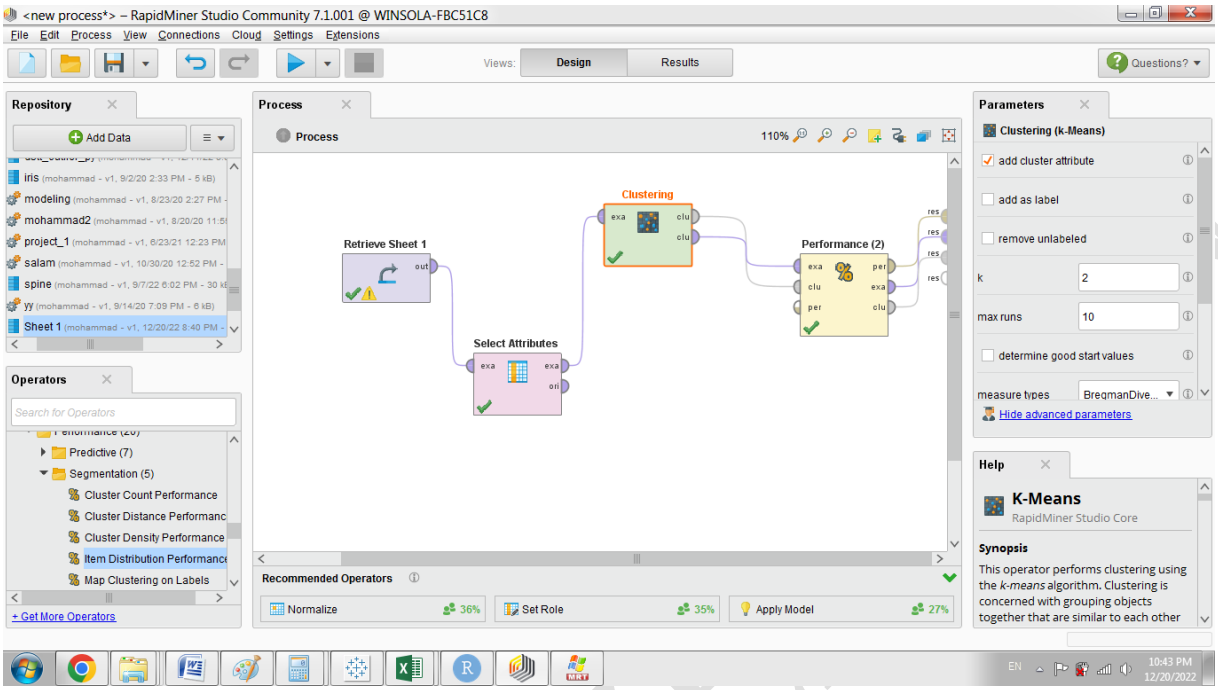
همان‌طور که مشاهده می‌شود، تعداد ویژگی‌ها برابر با ۶ ویژگی است و هر سطر مربوط به هر مشتری می‌باشد. در این مجموعه داده مشخص است که چه مشتری برون‌رفت داشته است. همچنین مقادیر R و F1 و F2 و M برای هر مشتری مشخص است. می‌توان با روش‌های مختلف خوشه‌بندی به بخش‌بندی مشتریان بر اساس متغیرهای RFM پرداخت و سپس بهترین خوشه را شناسایی کرد. همچنین می‌توان برای هر مشتری برچسب خوشه‌ای را تعیین کرد.

برای این کار مسیر زیر را در نرم‌افزار رپیدماینر انجام می‌دهیم.

در ابتدا تعداد خوشه را برابر با ۲ قرار داده و مرتب تغییر می‌دهیم تا به بهترین شاخص دیویس‌بولیدین که کمترین مقدار آن است، دست یابیم.

شکل ۱۱

دستیابی به بهترین شاخص دیویس بولدین



جدول ۲

مقادیر دیویس بولدین برای تعداد خوشه‌های مختلف

تعداد خوشه	مقدار دیویس بولدین
۲	-۰.۵۵
۳	-۰.۵۴۹
۴	-۰.۴۳۳
۵	-۰.۴۴۸
۶	-۰.۴۸۴
۷	-۰.۴۴۷

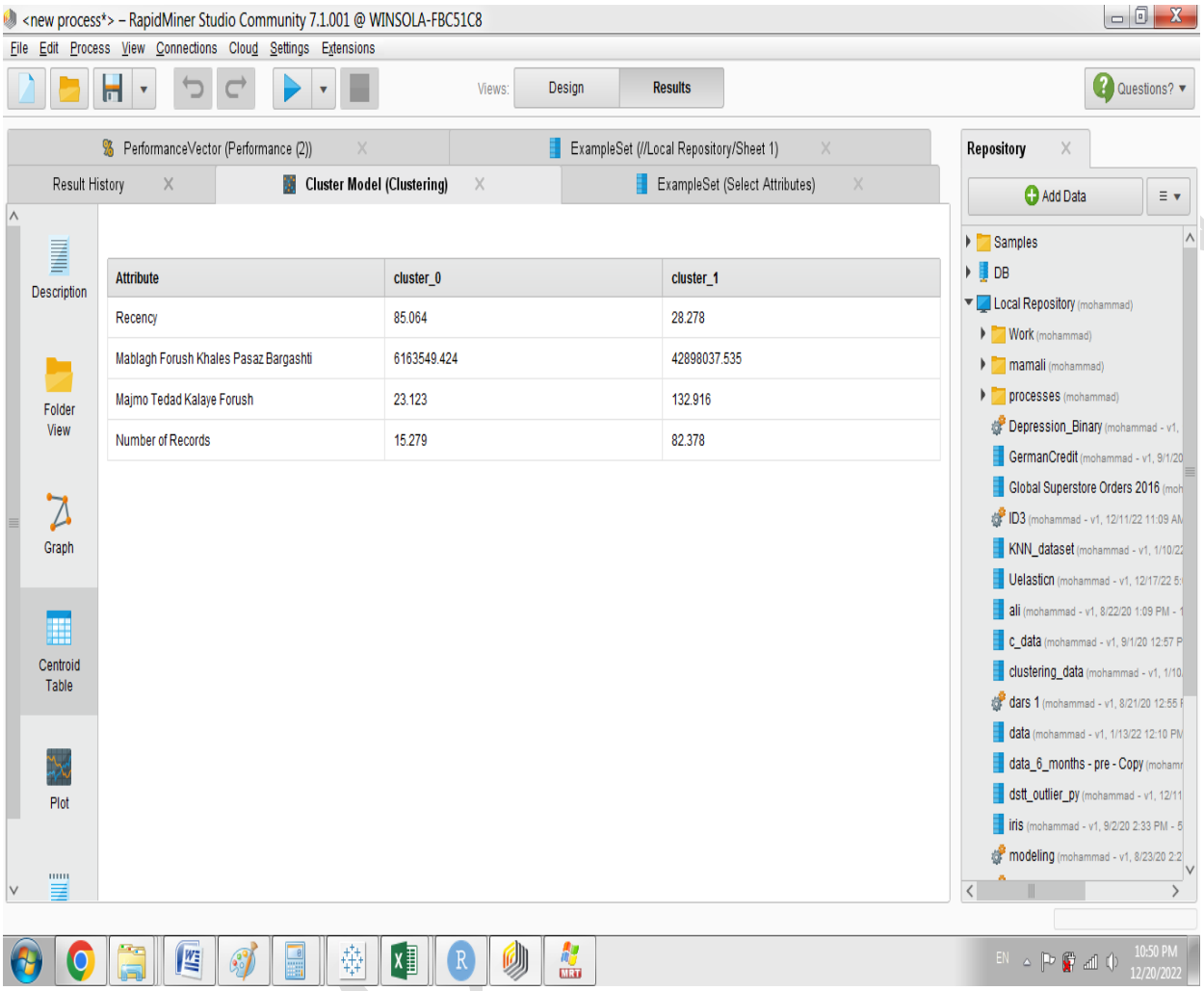
مقادیر دیگر خوشه‌ها نیز بررسی شد. ولی مقدار شاخص دیویس بولدین کمتر نمی‌شد. بهترین تعداد خوشه برابر با عدد ۲ است. لذا مشتریان به دو خوشه تقسیم می‌شوند. نتایج به قرار زیر است.

PROOF VERSION

E-ISSN: 3041-8585

شکل ۱۲

تقسیم مشتریان به دو خوشه

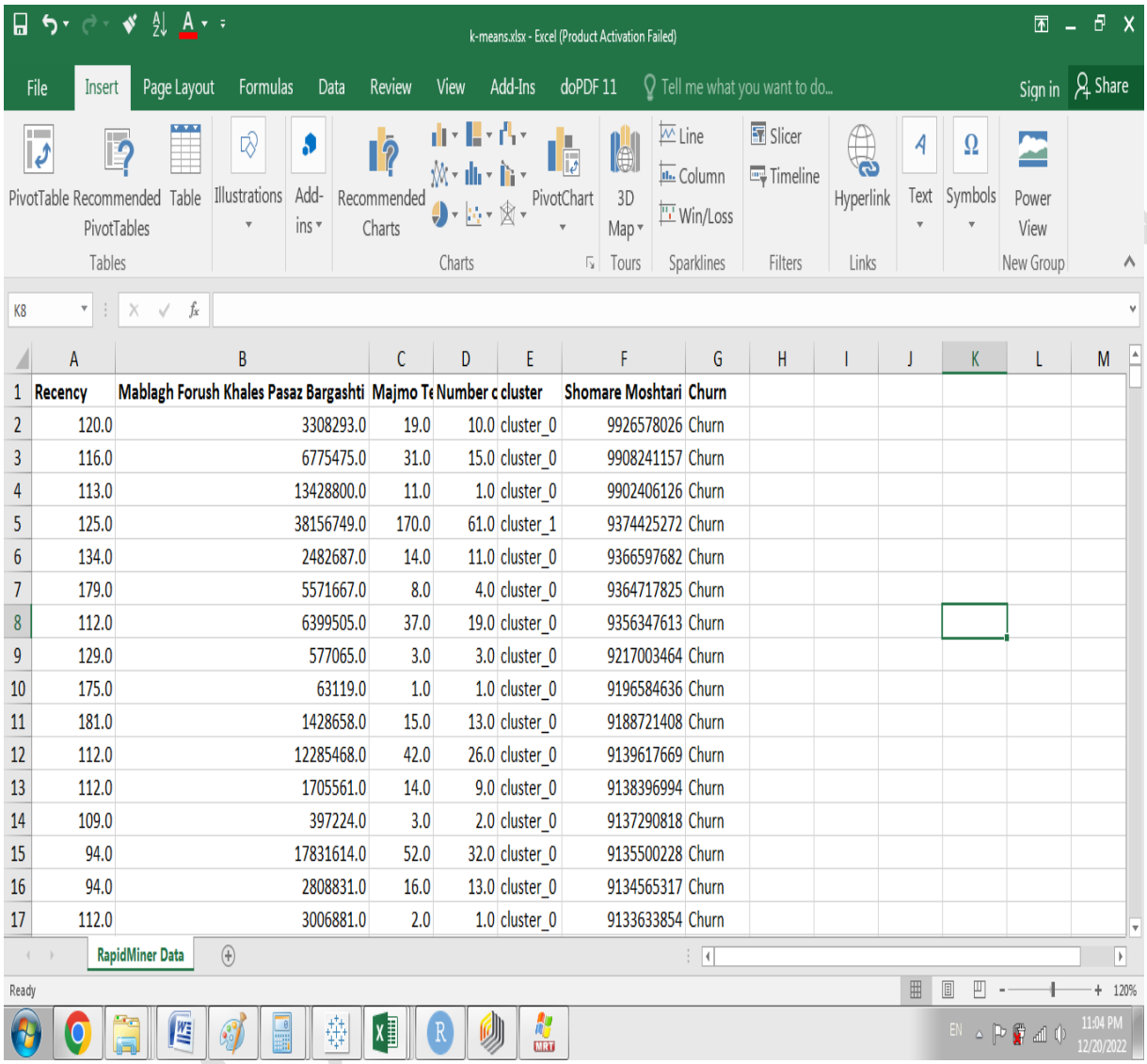


مشاهده می‌شود که به نظر می‌رسد که خوشه اول عملکرد بهتری دارد. زیرا پارامترهای RFM در آن نسبت به خوشه دوم بهتر است. لذا این خوشه به عنوان خوشه ارزشمند مشتریان فروشگاه زنجیره‌ای انتخاب می‌شود. تعداد مشتریان در خوشه اول برابر با ۲۰۴۵ مشتری است. تعداد مشتریان در خوشه دوم برابر با ۲۴۳۸۳ مشتری است.

بعد از تعیین خوشه مناسب و ارزشمند می‌توان مشتریان را بر اساس خوشه برچسب‌گذاری کرد. البته می‌توان ارزش هر خوشه را نیز محاسبه کرد. برای این کار باید مقادیر RFM برای مشتریان را در خوشه جمع کرده و بر تعداد مشتریان در هر خوشه تقسیم نمود. بدین منظور خروجی کار را وارد نرم‌افزار اکسل کرده و مقادیر را جمع و بر تعداد تقسیم می‌کنیم. داده جدید به صورت زیر است:

شکل ۱۳

ورود داده‌ها به نرم افزار اکسل



	A	B	C	D	E	F	G
1	Recency	Mablagh Forush Khales Pasaz Bargashti	Majmo Te	Number c	cluster	Shomare Moshtari	Churn
2	120.0	3308293.0	19.0	10.0	cluster_0	9926578026	Churn
3	116.0	6775475.0	31.0	15.0	cluster_0	9908241157	Churn
4	113.0	13428800.0	11.0	1.0	cluster_0	9902406126	Churn
5	125.0	38156749.0	170.0	61.0	cluster_1	9374425272	Churn
6	134.0	2482687.0	14.0	11.0	cluster_0	9366597682	Churn
7	179.0	5571667.0	8.0	4.0	cluster_0	9364717825	Churn
8	112.0	6399505.0	37.0	19.0	cluster_0	9356347613	Churn
9	129.0	577065.0	3.0	3.0	cluster_0	9217003464	Churn
10	175.0	63119.0	1.0	1.0	cluster_0	9196584636	Churn
11	181.0	1428658.0	15.0	13.0	cluster_0	9188721408	Churn
12	112.0	12285468.0	42.0	26.0	cluster_0	9139617669	Churn
13	112.0	1705561.0	14.0	9.0	cluster_0	9138396994	Churn
14	109.0	397224.0	3.0	2.0	cluster_0	9137290818	Churn
15	94.0	17831614.0	52.0	32.0	cluster_0	9135500228	Churn
16	94.0	2808831.0	16.0	13.0	cluster_0	9134565317	Churn
17	112.0	3006881.0	2.0	1.0	cluster_0	9133633854	Churn

همان‌طور که مشاهده می‌شود، برای هر مشتری، شماره خوشه مشخص شده است. همچنین مشخص است که این مشتری از نوع برون‌رفت است یا خیر.

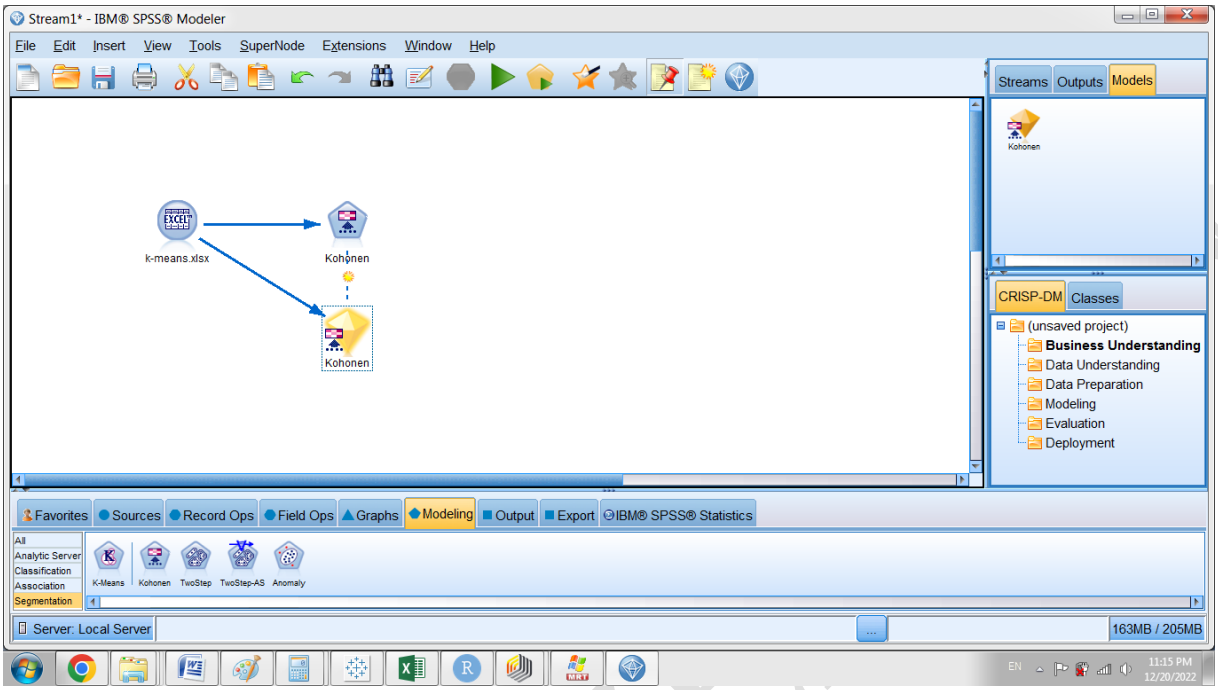
می‌توان از روش SOM نیز به بخش‌بندی مشتریان پرداخت. در این خصوص از نرم‌افزار SPSS Modeler استفاده می‌شود.

نمودار دیاگرام به صورت زیر است:

E-ISSN: 3041-8585

شکل ۱۴

استفاده از روش SOM

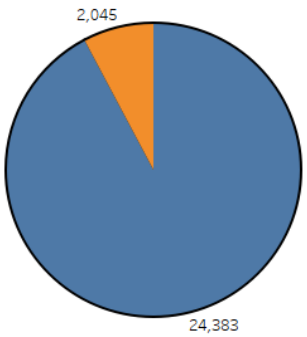


البته خروجی به این صورت است که الگوریتم تعداد خوشه‌ها را برابر با ۷۰ خوشه لحاظ کرد. لذا این نوع خوشه‌بندی چندان خوشایند نیست. به همین دلیل روش کامیانگین به عنوان خوشه‌بندی برتر انتخاب می‌شود.

در ادامه به بررسی ویژگی مشتریان در خوشه دوم پرداخته می‌شود. لذا مجموعه داده را به دو قسمت تقسیم کرده و به توصیف خوشه دوم می‌پردازیم. برای این کار از نرم‌افزار Tableau استفاده می‌شود.

شکل ۱۵

فراوانی خوشه‌ها

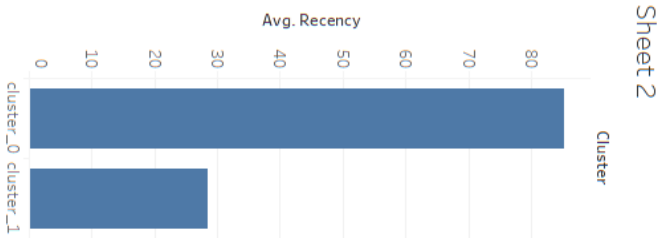


همان‌طور که در شکل فوق مشاهده می‌شود، خوشه اول، دارای تعداد مشتریان از نوع برون‌رفت به مراتب بیشتر دارد. شکل زیر نیز به مقایسه دو خوشه بر اساس متوسط اولین تعامل با فروشگاه می‌پردازد.

PROOF VERSION

شکل ۱۶

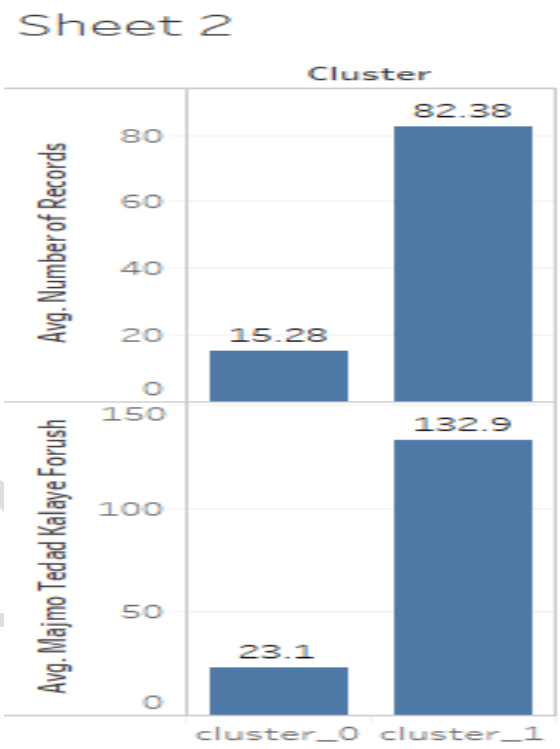
زمان اولین تعامل مشتریان با فروشگاه



نمودار نشان می‌دهد که متوسط زمان اولین تعامل مشتریان با فروشگاه در خوشه دوم کمتر است. لذا مقدارR کمتری دارد.

شکل ۱۷

مقایسه خوشه اول و دوم در شاخص تناوب

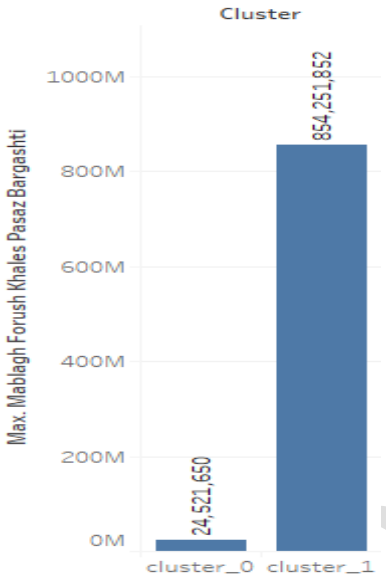


شکل فوق نشان می‌دهد که متوسط F1 و F2 در خوشه دوم بیشتر است. لذا مشتریان خوشه دوم به لحاظ شاخص تناوب نیز در وضعیت بهتری هستند. شکل زیر نیز نشان می‌دهد که بیشترین مقدار M مربوط به خوشه دوم بوده است.

E-ISSN: 3041-8585

شکل ۱۸

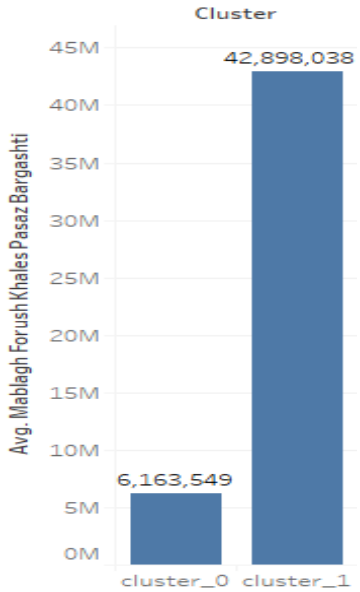
مقایسه بیشترین مقدار M



البته با توجه به شکل زیر، متوسط مقدار M برای خوشه دوم بیشتر است.

شکل ۱۹

مقایسه متوسط مقدار M



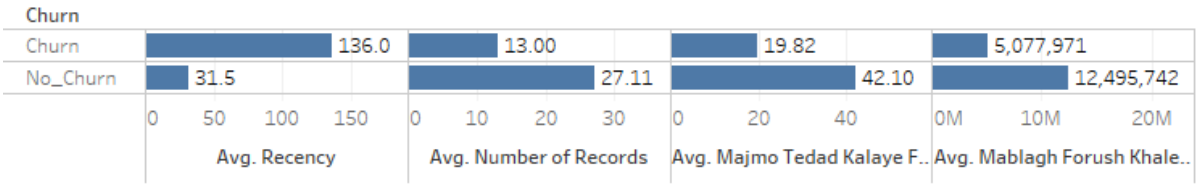
لذا در مجموع خوشه دوم از مشتریان ارزشمندتری برخوردار است.

PROOF VERSION

۳. یکی از روش‌های دیگر این است که به خوشه‌بندی مشتریان بر اساس خروج یا عدم خروج آن‌ها از فروشگاه پرداخته شود. یعنی اینکه کدام خوشه بیشترین و کمترین نرخ خروج را از فروشگاه داشته است. سپس می‌توان به توصیف هر خوشه پرداخت. جدول زیر به ارائه نتایج خوشه‌بندی بر اساس کامیانگین و SOM می‌پردازد. در ادامه می‌توان به تحلیل ویژگی‌های مشتریان بر اساس برون‌رفت پرداخت.

شکل ۲۰

مقایسه RFM مشتریان

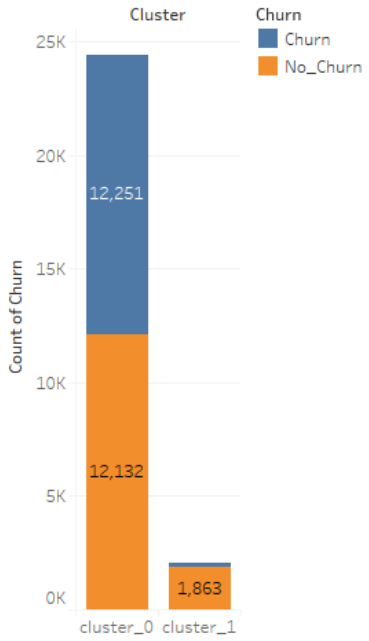


شکل فوق نشان می‌دهد که مشتریانی که برون‌رفت نداشته‌اند، در متغیرهای RFM نیز عملکرد بهتری دارند.

شکل زیر، خوشه‌ها را بر اساس مشتریان برون‌رفت و غیربرون‌رفت نشان می‌دهد.

شکل ۲۱

خوشه‌ها بر اساس مشتریان برون رفت و غیر برون رفت



همان‌طور که مشاهده می‌شود، اکثر مشتریان در خوشه دوم، غیربرون‌رفت هستند.

شکل زیر نشان می‌دهد که تقریبا نیمی از مشتریان در خوشه اول از نوع برون‌رفت هستند.

E-ISSN: 3041-8585

شکل ۲۲

مقایسه نوع برون رفت

Sheet 1

	Cluster		
Churn	cluster_0	cluster_1	Grand Total
Churn	12,251	182	12,433
No_Churn	12,132	1,863	13,995
Grand Total	24,383	2,045	26,428

Count of Churn broken down by Cluster vs. Churn.

بحث و نتیجه‌گیری

در این پژوهش به ارائه روشی برای تحلیل رفتار مشتریان فروشگاه زنجیره‌ای بر اساس ترجیحات مشتری، ارزش مشتری و خروج مشتری پرداخته شد. این کار توسط روش‌های داده‌کاوی و یادگیری ماشین صورت گرفت. در ابتدا مجموعه عملیات آماده‌سازی و پیش‌پردازش داده‌ها صورت گرفت. مجموعه داده دارای تعداد بسیار زیادی تراکنش مشتری و تعدادی ویژگی تراکنش‌ها بود. از سه رویکرد برای دستیابی به الگوهای یادگیری ماشین استفاده شد که عبارتند از قوانین وابستگی، طبقه‌بندی و خوشه‌بندی. در ادامه به تفسیر نتایج و یافته‌ها در این سه رویکرد پرداخته می‌شود.

با استفاده از رویکرد قوانین وابستگی به یافتن اقلام مکرر در یک سبد کالا پرداخته شد و سپس توسط الگوریتم قوانین وابستگی، رایج‌ترین قوانین در بین الگوهای خرید شناسایی شد با توجه به تعداد بسیار زیاد ویژگی‌ها در این مجموعه داده، شبکه عصبی قادر به اجرای مدل طبقه‌بندی نیست و لذا از روش انتخاب ویژگی‌ها به انتخاب ۱۰۰ ویژگی یا ۱۰۰ کالای مهم‌تر پرداخته شد.

جدول ۳

مقایسه الگوریتم‌های طبقه‌بندی بر اساس دقت طبقه‌بندی

الگوریتم طبقه‌بندی	دقت طبقه‌بندی
روش درخت تصمیم‌گیری	۹۸.۴۴
روش نیو بیز	۹۱
روش کا نزدیک‌ترین همسایه	۸۵
روش درخت ID۳	۹۷
روش شبکه عصبی	۹۷.۵
روش شبکه عصبی RBF	۹۵
روش جنگل تصادفی	۸۶.۴۹
روش رگرسیون لجستیک	۹۴.۰۵
روش ماشین بردار پشتیبان	۹۱.۸۹

PROOF VERSION

همان‌طور که در جدول فوق مشاهده می‌شود، دقت مدل طبقه‌بندی و روش شبکه عصبی و درخت تصمیم ID3 دارای بالاترین دقت هستند. لیکن روش شبکه عصبی قابلیت تفسیر ندارد. زیرا الگوهایی مثل قوانین اگر آنگاه تولید نمی‌کند. ولی درخت تصمیم‌گیری این قوانین را تولید می‌کند و لذا تفسیر راحت‌تری دارد. همچنین پیچیدگی مدل ایجادشده توسط ID3 بسیار بالا است و لذا روش اول برای محاسبات مربوطه لحاظ می‌شود.

در بخش بعدی به طبقه‌بندی مشتریان بر اساس متغیرهای RFM پرداخته شد. یعنی هر یک از متغیرهای RFM را به تنهایی می‌توان به عنوان ویژگی هدف لحاظ نمود. در این مطالعه، متغیر F1 یعنی تعداد تراکنش‌هایی که یک مشتری انجام داده است، به عنوان متغیر هدف لحاظ شد. هدف این بود که مدل طبقه‌بندی طراحی شود که بتواند تعیین کند که مشتریان چه تعداد تراکنشی را در فروشگاه انجام می‌دهند.

در این خصوص الگوریتم‌های مختلف طبقه‌بندی بر روی مجموعه داده مورد آموزش و آزمون قرار گرفت.

جدول ۴

نتایج مقایسه دقت طبقه‌بندی

الگوریتم طبقه‌بندی	دقت طبقه‌بندی
روش درخت تصمیم‌گیری	۶۲.۴۵
روش نیو بیز	۷۸.۹۸
روش کا نزدیک‌ترین همسایه	۵۲.۸۳
روش درخت ID۳	۶۱.۹۴
روش شبکه عصبی	۸۵.۹
روش شبکه عصبی RBF	۸۲.۲
روش جنگل تصادفی	۵۶.۹۸
روش رگرسیون لجستیک	۵۹.۶۸
روش ماشین بردار پشتیبان	۵۸.۵۴

همان‌طور که در جدول فوق مشاهده می‌شود، بالاترین دقت در طبقه‌بندی تعداد تراکنش‌های مشتریان مربوط به الگوریتم شبکه عصبی است. سپس الگوریتم شبکه عصبی RBF نیز توانسته است مشتریان را با دقت تقریبا مناسبی طبقه‌بندی کند.

یکی دیگر از روش‌های طبقه‌بندی که در این پژوهش مورد استفاده قرار گرفت از این قرار بود که مدلی را طبقه‌بندی کرد که مشخص کند مشتریان، یک کالای خاص را خریداری می‌کنند یا خیر. یعنی چه قوانین و الگوهایی برای خرید یک کالای خاص در این فروشگاه حاکم است. در این روش، متغیرهای مستقل متغیرهای RFM بودند. کالای ۲۲۰۰۱۷۰ به عنوان یک کالایی در نظر گرفته شد که بررسی شود آیا مشتریان بر طبق قوانین و الگوها، این کالا را خریداری می‌کنند یا خیر.

PROOF VERSION

E-ISSN: 3041-8585

جدول ۵

نتایج مدل‌های طبقه‌بندی

الگوریتم طبقه‌بندی	دقت طبقه‌بندی
روش درخت تصمیم‌گیری	۸۴.۹۱
روش نیو بیز	۸۱.۷۸
روش کا نزدیک‌ترین همسایه	۸۴.۲۶
روش شبکه عصبی	۸۳.۴
روش شبکه عصبی RBF	۸۳.۸
روش جنگل تصادفی	۸۵.۰۷

همان‌طور که در جدول فوق مشاهده می‌شود، مدل‌های جنگل تصادفی و روش درخت تصمیم‌گیری دارای بالاترین دقت در طبقه‌بندی کالای ۲۲۰۰۱۷۰ در تعیین خرید یا عدم خرید این کالا هستند. به عبارت دیگر می‌توان با استفاده از مدل‌های تولیدشده توسط این دو الگوریتم به پیش‌بینی و طبقه‌بندی خرید یا عدم خرید این کالا پرداخت. لذا برای مشتریانی که در فروشگاه وجود دارند، می‌توان پیشنهاد خرید این کالا را در صورتی که در مدل طبقه‌بندی لحاظ شوند، ارائه داد.

روش بعدی برای بکارگیری تکنیک‌های طبقه‌بندی، استفاده از این تکنیک‌ها برای طبقه‌بندی مشتریان بر اساس خروج یا عدم خروج آن‌ها از فروشگاه است. به عبارت دیگر مشتریانی که وفادار هستند و مشتریانی که کمتر وفادار بوده و از فروشگاه خارج شده‌اند را شناسایی می‌کند. در این روش متغیرهای کالاها به عنوان متغیرهای ورودی یا مستقل لحاظ شدند. به عبارت دیگر هدف از این نوع طبقه‌بندی این است که بررسی شود که اگر مشتریان از کالاها خریداری کنند یا خریداری نکنند، وضعیت آن‌ها به لحاظ خروج یا عدم خروج از فروشگاه چگونه است. برای ساخت مدل‌های طبقه‌بندی صرفا مجموعه داده مشتریان ارزشمند لحاظ شد. یعنی به بررسی خرید یا عدم خرید مشتریان ارزشمندی پرداخته شود که خروج یا عدم خروج از فروشگاه را داشته‌اند. با توجه به خوشه‌بندی که صورت گرفته است، خوشه دوم به عنوان مشتریان ارزشمند لحاظ شد. لذا به تحلیل خروج یا عدم خروج مشتریان در این خوشه پرداخته شد.

جدول ۶

دقت مدل‌های طبقه‌بندی

الگوریتم طبقه‌بندی	دقت طبقه‌بندی
روش درخت تصمیم‌گیری	۷۰.۲۸
روش نیو بیز	۷۳.۶۱
روش کا نزدیک‌ترین همسایه	۵۱.۶۷
روش درخت ID۳	۵۹.۷۲
روش جنگل تصادفی	۵۱.۳۹
روش رگرسیون لجستیک	۷۳.۶۱
روش ماشین بردار پشتیبان	۶۴.۱۷

همان‌طور که در جدول فوق مشاهده می‌شود، دقت روش طبقه‌بندی نیوبیز و رگرسیون لجستیک نسبت به بقیه الگوریتم‌ها بالاتر است. لذا از این دو الگوریتم می‌توان برای پیش‌بینی و طبقه‌بندی خروج مشتریان باارزش استفاده نمود.

PROOF VERSION

سه روش برای بخش‌بندی مشتریان فروشگاه بر اساس رویکردهای ترجیح مشتری، ارزش مشتری و خروج مشتری مورد استفاده قرار گرفت.

در ابتدا به بخش‌بندی مشتریان بر اساس ترجیحات آن‌ها پرداخته شد. در این خصوص هر یک از کالاها به عنوان یک ویژگی لحاظ شدند. سپس به استفاده از روش‌های خوشه‌بندی کامیانگین و روش SOM پرداخته شد در ادامه به خوشه‌بندی مشتریان بر اساس متغیرهای RFM پرداخته شد. در ادامه به بخش‌بندی مشتریان بر اساس خروج و عدم خروج آن‌ها در تعامل با فروشگاه زنجیره‌ای پرداخته شد. مشتریانی که خروج نداشته‌اند دارای عملکرد بهتری در متغیرهای RFM هستند و لذا مشتریانی که خروج نداشته‌اند، ارزشمندتر می‌باشند. همچنین مشتریان در خوشه دوم به میزان بیشتری وفادار هستند.

برای پژوهش‌های آتی می‌توان به طراحی یک سیستم پشتیبان تصمیم‌گیری پرداخت. همچنین می‌توان از روش‌های یادگیری عمیق برای طبقه‌بندی مشتریان استفاده نمود.

تقدیر و تشکر

این پژوهش برگرفته از رساله دکتری دانشگاه آزاد اسلامی واحد کاشان و به لطف مشارکت مدیران ریسک شرکت‌های دانش بنیان به انجام رسیده است که نویسندگان مقاله بر خود لازم می‌دانند از همکاری و مساعدت همه عزیزانی که در این پژوهش یاری نموده‌اند، سپاسگزاری نمایند.

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 پژوهش حاضر تمامی موازین 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.

PROOF VERSION

E-ISSN: 3041-8585

References

Arifin, R., & Purnama, D. A. (2023). Identifying Customer Preferences on Two Competitive Startupproducts: An Analysis of Sentiment Expressions and Textmining From Twitter Data. *Jurnal Infotel*, *15*(1), 66-74. https://doi.org/10.20895/infotel.v15i1.906

Chen, H., Liu, H., Chu, X., Zhang, L., & Yan, B. (2020). A two-phased SEM-neural network approach for consumer preference analysis. *Advanced Engineering Informatics*, *46*, 101156. https://doi.org/10.1016/j.aei.2020.101156

Dullaghan, C., & Rozaki, E. (2017). Integration of Machine Learning Techniques to Evaluate Dynamic Customer Segmentation Analysis for Mobile Customers. *International Journal of Data Mining & Knowledge Management Process*, *7*(1), 13-24. https://doi.org/10.5121/ijdkp.2017.7102

Jiang, H., Kwong, C. K., Okudan Kremer, G. E., & Park, W. Y. (2019). Dynamic modelling of customer preferences for product design using DENFIS and opinion mining. *Advanced Engineering Informatics*, *42*. https://doi.org/10.1016/j.aei.2019.100969

Jing, N., Jiang, T., Du, J., & Sugumaran, V. (2017). Personalized recommendation based on customer preference mining and sentiment assessment from a Chinese e-commerce website. *Electronic Commerce Research*, *18*(1), 159-179. https://doi.org/10.1007/s10660-017-9275-6

Kim, J., Ji, H., Oh, S., Hwang, S., Park, E., & del Pobil, A. P. (2021). A deep hybrid learning model for customer repurchase behavior. *Journal of Retailing and Consumer Services*, *59*, 102381.

Lalwani, P., Mishra, M. K., Chadha, J. S., & Sethi, P. (2021). Customer churn prediction system: a machine learning approach. *Computing*, *104*(2), 271-294. https://doi.org/10.1007/s00607-021-00908-y

Mishra, A., & Reddy, U. S. (2017). *A comparative study of customer churn prediction in telecom industry using ensemble based classifiers*. https://doi.org/10.1109/ICICI.2017.8365230

Nelaturi, N., & Devi, G. (2019). A Product Recommendation Model Based on Recurrent Neural Network. *Journal Européen des Systèmes Automatisés*, *52*(5), 501-507. https://doi.org/10.18280/jesa.520509

Parre Amarnath, M. C. (2018). A Two-Layer Clustering Model for Mobile Customer Analysis. *International Journal of Scientific Engineering and Technology Research*, *07*(05), 927-930.

Ram, S. V., Kempf, K. G., & Wu, T. (2017). *A Data Mining Approach to Modeling Customer Preference: A Case Study of Intel Corporation* ARIZONA STATE UNIVERSITY]. ASU Electronic Theses and Dissertations.

Selvakanmani, S., Nanda, P., K, D., A, K., & Afroz, S. (2020). Churn Prediction using Ensemble Learning: An Analytical CRM Application. *International Journal of Advanced Science and Technology*, *29*, 9192-9200.

Vafeiadis, T., Diamantaras, K. I., Sarigiannidis, G., & Chatzisavvas, K. C. (2015). A comparison of machine learning techniques for customer churn prediction. *Simulation Modelling Practice and Theory*, *55*, 1-9.

Verma, A. (2020). Consumer behaviour in retail: next logical purchase using deep neural network. *arXiv preprint arXiv:2010.06952*.

Yan, Y., Huang, C., Wang, Q., & Hu, B. (2020). Data mining of customer choice behavior in internet of things within relationship network. *International Journal of Information Management*, *50*, 566-574.

Zandi, A. (2024). Gamification in Retail: Enhancing Grocery Customer Experience With Location-Based Strategies. *International Journal of Serious Games*, *11*(2), 43-63. https://doi.org/10.17083/ijsg.v11i2.721

Zhou, F., Ayoub, J., Xu, Q., & Jessie Yang, X. (2020). A machine learning approach to customer needs analysis for product ecosystems. *Journal of Mechanical Design*, *142*(1), 011101.

Zhuravlev, Y., Dokukin, A., Senko, O., Stefanovsky, D., & Saenko, I. (2020). On a Novel Machine Learning Based Approach to Recommender Systems. Conference of Open Innovations Association, FRUCT,

Zidehsaraei, M. (2024). Do Pro-Environmental Factors Lead to Customers' Purchase Intention of Home Energy Management System? The Moderating Effects of Energy-Efficient Habits. *Green and Low-Carbon Economy*. https://doi.org/10.47852/bonviewglce42022907

PROOF VERSION