Journal of Soft Computing and Information Technology (JSCIT)
Babol Noshirvani University of Technology, Babol, Iran
Journal Homepage: jscit.nit.ac.ir
Volume 12, Number 4, Winter 2023, pp. 67-78
Received: 08/31/2023, Revised: 03/07/2024, Accepted: 08/14/2024

A Model for Time Series Clustering Optimization Based on Genetic Algorithm to Analyze the Behavior of Customers

Bahare Haghi[1], Hodjat Hamidi[2*]
1- Department of Industrial Engineering, Information Technology Group, K. N. Toosi University of Technology, Tehran, Iran.
2*- Department of Industrial Engineering, Information Technology Group, K. N. Toosi University of Technology, Tehran, Iran.
[1]b.haghi@email.kntu.ac.ir, [2*]h_hamidi@kntu.ac.ir

Corresponding author's address: Hodjat Hamidi, Tehran, Mollasadra St, K. N. Toosi University of Technology, Department of Industrial Engineering, Information Technology Group.

Abstract- Currently, organizations use data mining and business intelligence tools and techniques to analyze the behavior of their customers. Customer segmentation is a broad analytical tool used to identify distinct groups of customers. Because static segmentation leads to missing important patterns and trends in customer behavior over time. In this research, a model is presented that displays the behavior of each customer as a time sequence of the variables of purchase novelty, number of purchases, purchase amount, and customer cost. In fact, it also considers the time dimension of customer behavior. Then, using the genetic algorithm, optimal weights are found for each feature, and customers are segmented with clustering algorithms. The data used in this research is related to the transaction data of a payment service company during a period of thirty months. The results indicate that the best clustering result is achieved using spectral clustering algorithm by computing silhouette and Calinski-Harabasz metrics. These findings demonstrate that with optimal weighting, the genetic algorithm has been able to combine the features in a way that improves the silhouette metric to 0.91.

Keywords- Dynamic customer segmentation, Time series clustering, Analytical customer relationship management, Dynamic customer behavior, Data Mining, Genetic Algorithm

مجله علمی پژوهشی رایانش نرم و فناوری اطلاعات
دانشگاه صنعتی نوشیروانی بابل
صفحه مجله: jscit.nit.ac.ir
جلد ۱۲، شماره ۴، زمستان ۱۴۰۲ ، صفحه ۶۷-۷۸
دریافت: ۱۴۰۲/۰۶/۰۹، بازنگری: ۱۴۰۲/۱۲/۱۷، پذیرش: ۱۴۰۳/۰۵/۲۴

ارایه مدلی برای بهینه‌سازی خوشه‌بندی سری‌های زمانی بر اساس الگوریتم ژنتیک برای تحلیل رفتار مشتریان

بهاره حقی[۱]، حجت‌اله حمیدی*[۲]
۱- دانشکده مهندسی صنایع، گروه فناوری اطلاعات، دانشگاه صنعتی خواجه نصیرالدین طوسی، تهران، ایران.
*۲- دانشکده مهندسی صنایع، گروه فناوری اطلاعات، دانشگاه صنعتی خواجه نصیرالدین طوسی، تهران، ایران.
[1]b.haghi@email.kntu.ac.ir, *[2]h_hamidi@kntu.ac.ir

* نشانی نویسنده مسئول: حجت‌الله حمیدی، تهران، خیابان ملاصدرا، دانشگاه صنعتی خواجه نصیرالدین طوسی، دانشکده مهندسی صنایع، گروه فناوری اطلاعات.

چکیده- در حال حاضر، سازمان‌ها از ابزارها و تکنیک‌های داده‌کاوی و هوش تجاری برای تحلیل رفتار مشتریان خود استفاده می‌کنند. بخش‌بندی مشتریان یک ابزار تحلیلی گسترده است که برای شناسایی گروه‌های متمایز از مشتریان استفاده می‌شود. از آنجا که بخش‌بندی ثابت منجر به از دست دادن الگوها و روندهای مهم رفتار مشتری در طول زمان می‌شود. در این مقاله، یک مدل ارائه شده است که رفتار هر مشتری را به عنوان یک دنباله زمانی از متغیرهای خرید جدید، تعداد خرید، مبلغ خرید و هزینه مشتری نمایش می‌دهد در واقع بُعد زمانی رفتار مشتری را نیز در نظر می‌گیرد. سپس با استفاده از الگوریتم ژنتیک وزن‌های بهینه را برای هر ویژگی یافته و با الگوریتم‌های خوشه‌بندی، بخش‌بندی مشتریان انجام می‌شود. داده‌های مورد استفاده در این تحقیق مربوط به داده‌های تراکنشی یک شرکت خدمات پرداخت در طی دوره سی‌ماه است. نتایج نشان می‌دهد که بهترین نتیجه خوشه‌بندی با استفاده از الگوریتم خوشه‌بندی طیفی با محاسبه معیارهای سیلوهت و کالینسکی به دست می‌آید. این نتایج نشان می‌دهد که با توجه به وزن‌دهی بهینه، الگوریتم ژنتیک موفق به ترکیب ویژگی‌ها به گونه‌ای شد که معیار سیلوهت را به ۰/۹۱ ارتقا دهد.

واژه‌های کلیدی: بخش‌بندی پویا مشتریان، خوشه‌بندی سری‌های زمانی، مدیریت تحلیلی ارتباط با مشتری، داده‌کاوی، رفتار پویا مشتریان، الگوریتم ژنتیک

۱- مقدمه

امروزه مدیریت ارتباط با مشتریان به دلیل رقابت شدید بین شرکت‌ها در صنایع مختلف بسیار مهم شده است [۱] . با پیشرفت فناوری‌های اطلاعات و ارتباطات، حجم زیادی از داده‌ها درباره مشتریان در دسترس سازمان‌ها قرار می‌گیرد. به منظور استفاده از داده‌ها برای تصمیم‌گیری‌های استراتژیک، تکنیک‌های داده‌کاوی به عنوان یک ابزار قدرتمند برای تجزیه و تحلیل داده‌ها و خلق دانش درآمده‌اند. بخش‌بندی مشتریان ابزاری محبوب برای شناسایی گروه‌های مختلف مشتریان با نیازها و رفتارهای مشابه است[۲]. این بخش‌بندی به سازمان‌ها کمک می‌کند تا با بهره‌گیری از روش‌های تجزیه و تحلیل داده‌ها، به صورت کارآمدتر با مشتریان تعامل داشته باشند. اکثر مطالعات قبلی از یک رویکرد بخش‌بندی ثابت با جمع‌آوری داده‌های مشتریان در یک دوره ثابت استفاده کرده‌اند. محدودیت روش‌های بخش‌بندی ثابت این است که این روش‌ها قادر به مدل‌سازی رفتار پویای مشتریان و کشف الگوها و روندهای معنادار نیستند [۳]. بنابراین برای استخراج الگوهای مفید در مورد رفتار مشتری، رویکردهای بخش‌بندی پویا را نباید نادیده گرفت [۴] و [۵].در این مقاله، تلاش شده است یک مدل جدید جهت بخش‌بندی مشتریان با استفاده از تکنیک‌های داده‌کاوی ارائه شود. این مدل با استفاده از الگوریتم‌ها و روش‌های داده‌کاوی متناسب با مجموعه

داده‌های سازمان، قادر خواهد بود مشتریان را به گروه‌های مختلفی بر اساس ویژگی‌های مشترک تقسیم‌بندی کند. در مدل پیشنهادی، ویژگی‌های خرید جدید[۱]، تعداد[۲] خرید، مبلغ[۳] خرید و هزینه[۴] مشتری از داده‌های تراکنشی مشتریان استخراج می‌شود. سپس، تقسیم بندی مشتری با استفاده از خوشه بندی سری‌های زمانی[۵] انجام می‌شود. کیفیت خوشه‌های حاصل در خوشه‌بندی سری‌زمانی به انتخاب یک معیار فاصله مناسب و یک الگوریتم خوشه بندی موثر بستگی دارد. در مدل پیشنهادی از الگوریتم‌های خوشه‌بندی قدرتمند از جمله خوشه‌بندی سلسله مراتبی[۶]، خوشه‌بندی طیفی[۷]، خوشه‌بندی فازیC-means و خوشه‌بندی K-means برای به‌دست آوردن بهترین نتیجه خوشه‌بندی استفاده می‌شوند. پس از آن، هر بخش برچسب گذاری می‌شود و روندهای رفتاری آن تجزیه و تحلیل می‌شود تا الگوهای غالب آن آشکار شود. در این مقاله، یک مدل ارائه شده است که رفتار هر مشتری را به عنوان یک دنباله زمانی از متغیرهای خرید جدید، تعداد خرید، مبلغ خرید و هزینه مشتری نمایش می‌دهد در واقع بُعد زمانی رفتار مشتری را نیز در نظر می‌گیرد. سپس با استفاده از الگوریتم ژنتیک وزن‌های بهینه را برای هر ویژگی یافته و با الگوریتم‌های خوشه‌بندی، بخش‌بندی مشتریان انجام شده است. برای نشان دادن سودمندی این مدل، یک مطالعه موردی بر روی مشتریان یک شرکت خدمات پرداخت بانکی انجام شده است. نتایج مطالعه‌موردی الگوهای مفیدی را در مورد رفتارهای مشتریان بانک نشان می دهد.

در ادامه، ساماندهی مقاله به این صورت است که در بخش ۲ به مرور کارهای گذشته پرداخته می‌شود. تکنیک‌های مختلف مورد استفاده در مقاله و سپس چارچوب پیشنهادی برای تجزیه تحلیل رفتار مشتری در بخش ۳ ارائه می‌گردد. نتایج چارچوب پیشنهادی در یک مطالعه موردی گزارش می‌شود و الگوهای رفتاری کشف‌شده در بخش ۴ تحلیل می‌شوند. بخش ۵ به ارزیابی کارایی الگوریتم پیشنهادی اختصاص داده شده است. در نهایت، در بخش ۶ با نتیجه‌گیری و بیان چند جهت‌دهی تحقیقاتی برای آینده به پایان می‌رسد.

۲- کارهای گذشته

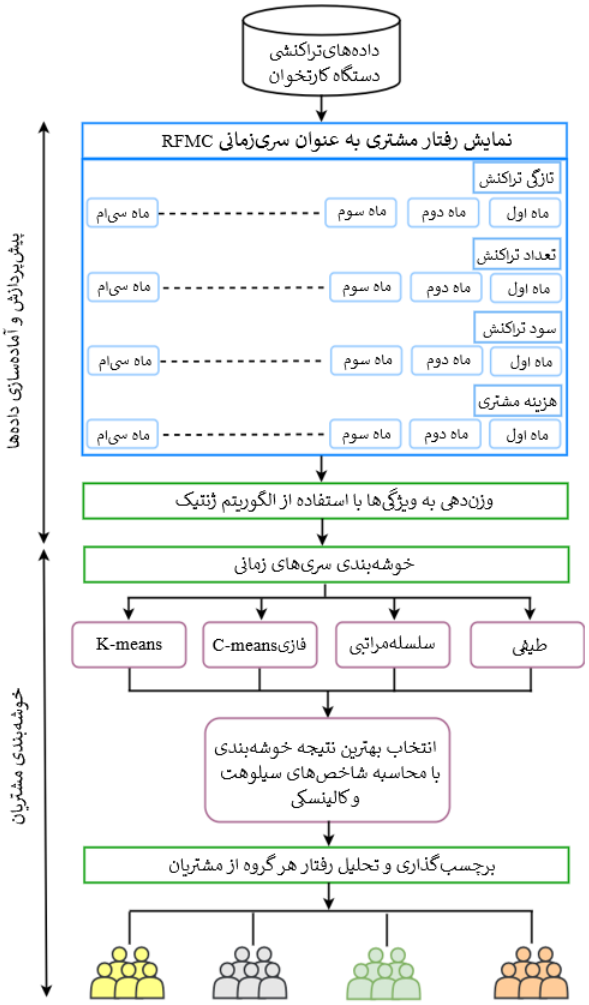
نویسندگان در [۶] مروری بر روش‌های بخش‌بندی مشتریان داشته‌اند و اعلام نمودند که بخش‌بندی مشتریان بر اساس ویژگی‌های جمعیت شناختی مانند سن، جنسیت، تحصیلات، شغل و درآمد به تحلیل و درک نقش و تأثیر ویژگی‌های جمعیت شناختی در رفتار و الگوی خرید مشتریان کمک می‌کند. با بخش‌بندی مشتریان بر اساس ویژگی‌های جمعیت شناختی، شرکت‌ها می‌توانند بازار هدف خود را بهینه‌سازی کرده و هزینه‌های بازاریابی را کاهش دهند. نویسندگان در [۷] تحقیقی در حوزه بانکداری انجام دادند که مدلی برای تقسیم مشتریان در بخش‌های همگن، محاسبه ارزش طول عمر مشتری در بخش‌های مختلف بر اساس تجزیه و تحلیل خرید جدید، تکرار خرید و مبلغ خرید و درنهایت مدلی برای پیش‌بینی هر بخش از مشتریان از روش سری‌زمانی فراهم می‌کنند. در [۸] دو روش برای بخش‌بندی مشتریان و محاسبه ارزش طول عمر مشتری برای هر بخش ارائه کرده‌اند: ۱. روش خرید جدید، تکرار خرید و مبلغ خرید و ۲. روش پیشنهادی خرید جدید، تکرار خرید و مبلغ خرید پیشرفته با پارامتر اضافی تعداد اقلام. نتایج این روش‌ها نشان داد که افزودن تعداد شمارش به عنوان یک پارامتر جدید در روش خرید جدید، تکرار خرید، مبلغ خرید هیچ نتیجه‌ای برای خوشه‌بندی ندارد. در [۹] به ارائه مدل خرید جدید، تکرارخرید، مبلغ خرید به ازای هر محصول پرداختند. نتایج این تحقیق نشان داد که داده‌های محصول می‌تواند اطلاعات مفیدی را برای مدیریت دارایی‌های بازاریابی و برآورد دقیق طول عمر مشتری فراهم کنند؛ همچنین می‌تواند خطای پیش‌بینی ارزش پایه مشتریان را کاهش دهد. در [۱۰] تحقیقی بر پایه خرید جدید، تکرار خرید، مبلغ خرید در صنعت خرده‌فروشی انجام دادند و در آن برای طبقه‌بندی مجموعه داده‌ها از الگوریتم K-means استفاده کردند. در این تحقیق تعداد و ارزیابی خوشه‌ها با استفاده از شاخص سیلوهت برای الگوریتم خوشه‌بندی صورت گرفت. در [۱۱] روش جدید خوشه‌بندی چند معیاره را برای رفع نقایص روش خوشه‌بندی با استفاده از بهینه‌سازی دوفازی ارائه دادند. آنها با هوشمندسازی عملگر جهش[۸]، الگوریتم ژنتیک[۹] را ارتقا داده و یک الگوریتم ژنتیک ابتکاری معرفی کرده‌اند. نتایج نشان داد که الگوریتم خوشه‌بندی بهینه‌سازی دوفازی با استفاده از الگوریتم فراابتکاری[۱۰] مبتنی بر جمعیت ژنتیک برای حل مسئله خوشه‌بندی، یک روش کارا اما نیازمند زمان زیادی است. در [۱۲] مدلی ترکیبی از خرید جدید، تکرار خرید، مبلغ خرید و سری‌زمانی ارائه دادند که در این روش آنها دو ویژگی تکرار خرید و مبلغ خرید را ترکیب خطی نموده تا تجزیه و تحلیل مشتریان را بر اساس تغییرات کسب‌وکار انجام دهند و یک مدل تقسیم‌بندی که بر اساس الگوهای جدید رفتاری به‌روز شود ارائه دادند. آنها مدل پیشنهادی خود را با مدل خرید جدید، تکرار خرید ، مبلغ خرید معمولی مقایسه کرده و برای هر بخش استراتژی مناسبی تعیین کرده و یک کمپین پیامکی بر اساس این استراتژی‌ها اجرا کرده‌اند. نتایج کمپین نشان داد که مدل تجزیه و تحلیل آنها، تعداد خریدها و میانگین مبلغ سبد خرید را بهبود بخشیده است. در [۱۳] یک مدل ارائه دادند که با استفاده از الگوریتم‌های پیشرفته خوشه‌بندی

و خوشه‌بندی سری‌های زمانی، بُعد زمانی رفتار مشتری را در نظر می‌گیرد. در این روش ابتدا از داده‌های تراکنشی مشتریان سه ویژگی خرید جدید، تکرار خرید، مبلغ خرید استخرج شده سپس، با استفاده از الگوریتم‌های خوشه‌بندی سری‌های زمانی، بخش‌بندی مشتریان انجام می‌شود. در این روش، از تکنیک‌های خوشه‌بندی سری‌های زمانی و معیارهای فاصله سری‌زمانی استفاده شده است. سپس، این مفاهیم را ترکیب کرده و روش پیشنهادی را ارائه داده‌اند. در [۱۴] ابتدا تحلیل خرید جدید، تکرار خرید، مبلغ خرید را بر روی داده‌های تراکنشی مشتریان اجرا کردند؛ سپس مشتریان با استفاده از روش‌های خوشه‌بندی K-means و فازیC-means خوشه‌بندی شدند. در این تحقیق همچنین یک ایده جدید برای انتخاب مراکز اولیه خوشه در الگوریتم‌های فوق معرفی شد. در [۱۵] مدلی برای برآورد ارزش طول عمر مشتری در صنعت هواپیمایی معرفی کردند. آنها در ابتدا با استفاده از اطلاعات پرواز مشتریان و سپس با افزودن داده‌های شبکه اجتماعی، مدل رگرسیون چندگانه را تهیه و سپس دو مدل به‌دست آمده را با هم مقایسه کردند. نتایج نشان داد که با افزودن عوامل اجتماعی به مدل پایه، هم دقت مدل و هم توانایی پیش‌بینی آن می‌تواند به‌طور بالقوه بهبود یابد و می‌تواند با مدل‌های موجود در حوزه‌های مختلف تطبیق داده شود. در [۱۶] ، مشتریان یک شرکت را توسط مدل خرید جدید، تکرارخرید و مبلغ خرید بررسی کردند. آنها از الگوریتم خوشه‌بندی فازی و تحلیل پورتفولیوی مشتریان بهره گرفتند. آنها از مدل شاخص سودآوری مشتری برای اعتبارسنجی استفاده کردند و رتبه‌بندی با روش تجزیه و تحلیل سلسله مراتبی برای یافتن ارزش طول عمر مشتری صورت گرفت. نتایج این تحقیق نشان داد که مشتریان شرکت بر اساس مقدار ارزش طول عمر مشتری به دست آمده به سه خوشه (سوپراستار) رتبه اول، (معمولی) رتبه دوم و (خفته) رتبه سوم طبقه‌بندی می‌شوند.

با توجه به مرور کارهای گذشته می‌توان برداشت کرد که تحقیق‌های پیشین در مورد بخش‌بندی مشتریان از روش خرید جدید، تکرارخرید، و مبلغ خرید بیشتر استفاده کرده‌اند، زیرا این روش در عین سادگی کارایی بالایی دارد. تحقیقات انجام شده، نگاه مجزا به سه ویژگی خرید جدید، تکرار خرید و مبلغ خرید داشته و روابط بین این ویژگی‌ها را نادیده گرفته‌اند و با اعطای وزن یکسان به ویژگی‌ها بخش‌بندی را انجام داده‌اند، در حالی که اهمیت هر ویژگی متفاوت است. به منظور پر کردن این خلاء تحقیقاتی، در این مقاله از الگوریتم ژنتیک برای وزن‌دهی به ویژگی‌های مشتریان استفاده شده است، که منجر به بهبود بخش‌بندی مشتریان نسبت به کارهای مشابه شده است.

۳- روش‌شناسی تحقیق

هدف اصلی روش پیشنهادی ارائه یک مدل برای شناسایی بخش‌های مختلف مشتریان با مدل‌سازی رفتار مشتری به عنوان سری‌های زمانی است. همانطور که در شکل ۱ نشان داده شده است. مدل پیشنهادی به طور موثر مدل تراکنش جدید، تعداد تراکنش، سود تراکنش و هزینه را با خوشه‌بندی سری‌های زمانی ادغام می کند. داده‌های سری‌زمانی را از جریانی از داده‌های تراکنشی دارای مهر زمانی استخراج می‌کند تا رفتار پویا مشتری را نشان دهد. سپس، سری‌های زمانی مشتریان را با استفاده از تکنیک‌های پیشرفته خوشه‌بندی سری‌های‌زمانی خوشه‌بندی می‌کند. در ادامه، جزئیات هر مرحله در مدل توضیح داده شده است.



شکل۱. مدل پیشنهادی تحقیق

۳-۱- نمایش رفتار مشتری به عنوان سری‌های زمانی RFMC[11]

در این مرحله، داده‌های تراکنشی مشتریان با مهر زمانی بر اساس بازه‌های زمانی یک ماهه تجمیع می‌شوند. سری‌های زمانی مربوط به هر متغیر برای مشتری محاسبه می شود. تعریف سری زمانی برای متغیرهای تازگی تراکنش، تعداد تراکنش، سود تراکنش و هزینه به شرح زیر است:

تراکنش جدید(t): تعداد ماه‌های گذشته شده از زمان آخرین تراکنش مشتری در بازه زمانی مشخص.

تعداد تراکنش(t): تعداد تراکنش‌هایی که با دستگاه کارتخوان در طول بازه زمانی مشخص شده انجام شده است.

سود تراکنش(t): میزان درآمدی که از تراکنش‌های مشتری در طول بازه زمانی مشخص به سازمان منتقل می‌شود.

هزینه مشتری(t): هزینه‌هایی که سازمان برای خدمات و محصولات ارائه شده به مشتری در طول بازه زمانی مشخص می‌پردازد. این هزینه‌ها شامل هزینه‌های رول دستگاه کارتخوان، بازدیدهای دوره‌ای، هزینه‌های مرتبط با خدمات فنی یا مشاوره، هزینه‌های مرتبط استفاده از تجهیزات و تکنولوژی‌های خاص و سایر هزینه‌های مرتبط با ارائه خدمات به مشتری می‌شود. در این روش‌شناسی تحلیل رفتار مشتری با در نظر گرفتن چهار ویژگی تراکنش جدید، تعداد تراکنش، سود تراکنش و هزینه مشتری صورت گرفته است. این رویکرد گسترده و جامع این امکان را می‌دهد تا تأثیر هر یک از این ویژگی‌ها بر رفتار مشتریان بررسی شود.

۳-۲- وزن‌دهی به ویژگی‌ها با استفاده از الگوریتم ژنتیک

در این مرحله، از الگوریتم ژنتیک برای وزن‌دهی به ویژگی‌ها در خوشه‌بندی سری‌های زمانی استفاده شده است. هدف این فرایند، یافتن مجموعه‌ای از وزن‌های بهینه است که به ویژگی‌ها اهمیت مناسبی در خوشه‌بندی بدهد. با استفاده از این وزن‌ها، می‌توان خوشه‌بندی دقیق‌تری انجام داد.

۳-۳- خوشه‌بندی سری‌های زمانی

خوشه‌بندی سری‌های زمانی، یک روش تحلیلی است که امکان دسته‌بندی سری‌های زمانی مشابه را در گروه‌های جداگانه فراهم می‌کند. در این روش، ابتدا سری‌های زمانی به عنوان نمونه‌های داده‌های زمانی استخراج می‌شوند. سپس با استفاده از معیارهای فاصله مناسب، فاصله بین سری‌های زمانی محاسبه می‌شود. این معیارها می‌توانند شامل فاصله اقلیدسی[12]، فاصله منهتن[13] و سایر معیارهای مشابه باشند که بر اساس ویژگی‌های سری‌های زمانی به دست آمده‌اند [۱۷]. پس از آن، از چهار الگوریتم خوشه‌بندی قدرتمند (سلسله مراتبی، طیفی ،K-means و فازی C-means) برای خوشه‌بندی استفاده می‌شود.

۳-۴- انتخاب بهترین نتیجه خوشه‌بندی با محاسبه شاخص‌های سیلوهت[14] و کالینسکی[15]

در این مرحله، پس از انجام بخش‌بندی مشتریان با استفاده از ترکیب‌های مختلفی از الگوریتم‌ها و وزن ویژگی‌ها، کیفیت مدل‌های بخش‌بندی نیاز به ارزیابی با استفاده از یک شاخص مناسب برای اعتبار خوشه‌بندی دارد. خوشه‌بندی یک روش بدون ناظر است که هدف آن تقسیم داده‌ها به بخش‌هایی است که شباهت داخلی بالا و شباهت بین بخشی کم دارند [۱۸]. شاخص‌های اعتبار خوشه‌بندی داخلی بررسی می‌کنند که نقاط داده در هر گروه به چه اندازه درون خود گروه با یکدیگر شباهت دارند و به انتخاب بهترین نتیجه خوشه‌بندی کمک می‌کنند.

۳-۵- برچسب گذاری و تحلیل رفتار هر گروه از مشتریان

در این مرحله، هر یک از خوشه‌های حاصل شده برچسب‌گذاری می‌شوند و رفتار آن گروه بررسی می‌شود تا الگوهای غالب آنها در طول زمان شناسایی شوند. برای خلاصه کردن مراحل اصلی روش‌شناسی پیشنهادی، شکل ۱ ترسیم شده است. این مدل ورودی‌ها را به عنوان تراکنش‌های مشتری با نشان زمانی دریافت می‌کند و برای هر مشتری یک سری زمانی از ویژگی‌ها ایجاد می‌کند که نمایانگر رفتار مشتری در طول زمان است. سپس با استفاده از الگوریتم ژنتیک وزن مناسب برای هر ویژگی استخراج شده و بخش‌بندی مشتریان از طریق خوشه‌بندی سری‌زمانی انجام می‌شود. به‌طورخاص، ترکیب‌های مختلفی از الگوریتم‌ها و وزن‌های ویژگی بررسی می‌شوند تا بهترین مدل خوشه‌بندی برای شاخص اعتبار خوشه‌بندی داده پیدا شود. در نهایت، خوشه‌های حاصل تحلیل می‌شوند تا الگوهای غالب آنها در طول زمان آشکار شود.

۴- مطالعه تجربی

این بخش نمونه‌ای واقعی از اجرای چارچوب پیشنهادی با استفاده از داده‌های یک شرکت خدمات پرداخت بانکی است. چارچوب پیشنهادی و تمام روش‌های موجود در آن با استفاده از پایتون ۳.۷ پیاده‌سازی شده است. داده‌های ورودی شامل داده‌های ۳۰ ماهه تراکنش‌های دستگاه‌های کارتخوان هستند که شامل اطلاعات شناسه کارتخوان، شناسه تراکنش، تاریخ و مبلغ تراکنش می‌شود. این داده‌ها شامل ۸۴۴,۰۵۸, ۱۹۵ ریز تراکنش خرید توسط مشتریان می‌باشد.

۴-۱- نمایش رفتار مشتری به عنوان سری‌های زمانی RFMC

در این مرحله، ابتدا داده‌ها با استفاده از نرمال‌سازی مین-مکس[۱۶] نرمال‌سازی شدند. سپس داده‌های تراکنش به بازه‌های ماهانه تجمیع شدند. سپس با در نظر گرفتن تعریف متغیرهای RFMC، سری‌های زمانی R، F، M و C برای هر ماه استخراج می‌شود. از آنجا که تمامی مشتریان انتخابی در تمام ماه‌ها تراکنش دارند، متغیر R برای تمام مشتریان صفر خواهد بود. بنابراین، متغیر R در تحلیل‌ها در نظر گرفته نمی‌شود زیرا توانایی تمیزدهی به تقسیم‌بندی مشتریان را ندارد. علاوه بر این، تکرار تراکنش هر مشتری نشان‌دهنده تعداد تراکنش‌هایی است که هر مشتری انجام داده است. همچنین، واضح است که مقدار پول (مالی) هر تراکنش می‌تواند مقداری دلخواه داشته باشد. بنابراین، تعداد بالای تراکنش‌ها به معنی ارزش بالای مالی نیست. به همین دلیل، برای تجزیه و تحلیل رفتار مشتری، متغیر مالی به جای مبلغ تراکنش میزان سود حاصل از تراکنش در نظر گرفته شده است. زیرا هدف نهایی هر شرکت کسب سودآوری مطلوب است. در بخش‌بندی مشتریان، به ویژگی هزینه نیز توجه شده است. در نمونه مورد بررسی که شامل تراکنش‌های دستگاه کارتخوان است، هزینه‌هایی که مشتریان در طول رابطه با سازمان دارند در نظر گرفته شده‌اند. هزینه‌ها شامل هزینه‌های رول دستگاه کارتخوان، هزینه‌های بازدیدهای دوره‌ای، خرابی دستگاه کارتخوان و سایر هزینه‌هایی است که سازمان برای مشتریان خود می‌پردازد. با در نظر گرفتن هزینه در بخش‌بندی مشتریان، می‌توان گروه‌های مشتری با الگوهای هزینه‌ای مشابه را شناسایی کرده و راهکارهای بهینه برای مدیریت هزینه‌ها و بهره‌وری بیشتر سازمان ارائه داد.

۴-۲- وزن‌دهی به ویژگی‌ها با استفاده از الگوریتم ژنتیک

در مرحله وزن‌دهی به ویژگی‌ها، از الگوریتم ژنتیک برای تعیین وزن‌های مناسب ویژگی‌ها استفاده شده است. الگوریتم ژنتیک، یک روش محاسباتی الهام‌گرفته از مکانیسم تکامل و انتخاب طبیعی در طبیعت است. با استفاده از این الگوریتم، وزن‌های مناسب برای هر ویژگی استخراج می‌شوند تا تأثیر آنها در تحلیل و بخش‌بندی مشتریان بهینه شود. در این مرحله، پس از آماده‌سازی داده‌ها، به الگوریتم ژنتیک داده شدند. الگوریتم ژنتیک با استفاده از فرایند تکاملی، بهترین وزن برای هر ویژگی را مشخص می‌کند. ابتدا، یک جمعیت اولیه از وزن‌ها تولید می‌شود. سپس با استفاده از عملگرهای ترکیب و جهش، نسل‌های بعدی ایجاد می‌شوند. در هر نسل، وزن‌ها براساس تابع ارزیابی بهبود می‌یابند و تنظیم می‌شوند تا بهترین وزن برای هر ویژگی برآورده شود [۱۹]. پس از اجرای الگوریتم ژنتیک و بهبود تدریجی وزن‌ها، بهترین مقادیر وزن برای هر ویژگی استخراج می‌شود. در این روش، از یک ترکیب خطی از ویژگی‌ها استفاده شده است، به این معنا که ویژگی‌ها با وزن‌های متفاوتی ترکیب می‌شوند تا یک مقدار نهایی برای تحلیل و بخش‌بندی مشتریان به‌دست آید. مقادیر مجاز برای این ویژگی‌ها در بازه [۱, ۱-] قرار دارند، به این معنا که وزن‌های مربوط به ویژگی‌ها بین یک و منفی یک می‌باشند. مقادیر مرتبط با پارامترهای الگوریتم ژنتیک استفاده شده، در جدول ۱ آورده شده است.

جدول۱. پارامترهای الگوریتم ژنتیک

نام پارامتر	مقدار
حداکثر تعداد نسل ها	۱۰۰
تعداد افراد در هر نسل	۲۰۰
احتمال جهش	۰/۱
نسبت نخبگان	۰/۰۱
احتمال ترکیب	۰/۵
نسبت والدین	۰/۳
نوع ترکیب	uniform
حداکثر تعداد نسل‌های بدون بهبود	None

الگوریتم ژنتیک با استفاده از چهار الگوریتم سلسله مراتبی، طیفی، K-means و فازیC-means، اجرا شده است و چهار نمودار خروجی مربوط به هر تابع تناسب به دست آمد. نمودارهای ۲ تا ۵ نمایانگر عملکرد الگوریتم ژنتیک در بخش‌بندی مشتریان با استفاده از هر یک از این الگوریتم‌ها می‌باشند. بر اساس خروجی‌های به دست آمده از این نمودارها، الگوریتم خوشه‌بندی طیفی با دقت بالاتری نسبت به دیگر الگوریتم‌ها عمل کرده است. بنابراین، به منظور بهبود دقت خوشه‌بندی، از الگوریتم طیفی و وزن‌های ارائه شده توسط آن استفاده شده است. این نتایج نشان‌دهنده این است که الگوریتم ژنتیک با تابع تناسب محاسبه شده توسط الگوریتم طیفی، توانایی بیشتری در بخش‌بندی مشتریان را ارائه می‌دهد و به افزایش دقت و کارایی فرایند بخش‌بندی کمک می‌کند. به علاوه، بهترین راه‌حل پیدا شده توسط الگوریتم ژنتیک برای وزن‌دهی به ویژگی‌ها به شکل زیر است: وزن‌ها برای سه ویژگی تعداد تراکنش، درآمد تراکنش و هزینه، به ترتیب برابر با ۰.۷۰- ، ۰.۸۰، و ۰.۹۰ می‌باشند و تابع هدف (شاخص دان) برابر با ۰.۹۱ به ازای الگوریتم طیفی می‌باشد. در مراحل بعدی، وزن‌ها در ویژگی‌ها ضرب می‌شوند و خوشه‌بندی با مقادیر جدید انجام می‌شود.

۴-۳- خوشه‌بندی سری‌های زمانی

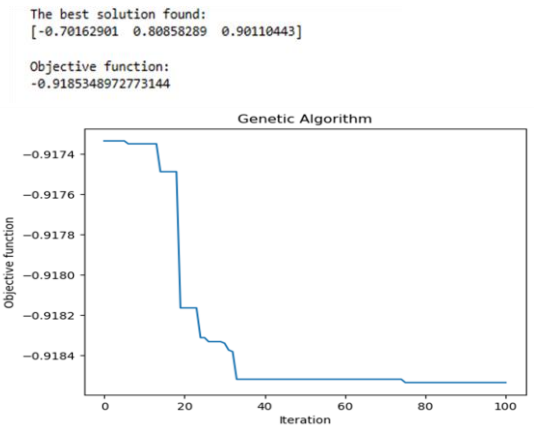
پس از نرمال‌سازی داده‌ها و وزن‌دهی به ویژگی‌ها با استفاده از الگوریتم ژنتیک، نوبت به بخش‌بندی مشتریان با استفاده از چهار الگوریتم محبوب شناسایی شده در مرور ادبیات می‌رسد. داده‌های آماده‌سازی شده با استفاده از هر چهار الگوریتم طیفی، سلسه‌مراتبی، فازی C-means و K-means بخش‌بندی شده و بهترین الگوریتم با توجه به معیار سیلوهت و کالینسکی معرفی می‌شود.

۴-۳-۱- تعیین تعداد بهینه خوشه‌ها با استفاده از شاخص دان[۱۷]

در این بخش، با استفاده از شاخص دان، تعداد بهینه خوشه‌ها برای بخش‌بندی مشتریان تعیین شده است. شاخص دان یک معیار کمی است که بر اساس فاصله بین خوشه‌ها و داخل خوشه‌ها، تعداد بهینه خوشه‌ها را تعیین می‌کند. با محاسبه این شاخص برای تعداد خوشه‌های مختلف، تعداد بهینه خوشه‌ها بدست می‌آید. هر چه شاخص دان مقدار بیشتری داشته باشد نتایج خوشه‌بندی بهتر است [۲۰]. مقادیر مختلف شاخص دان محاسبه شده در جدول ۲ قابل مشاهده است. همچنین نمودار تغییرات شاخص دان نیز در شکل ۶ مشاهده می‌شود که بیشترین مقدار شاخص دان به ازای تعداد خوشه سه بدست آمده است.

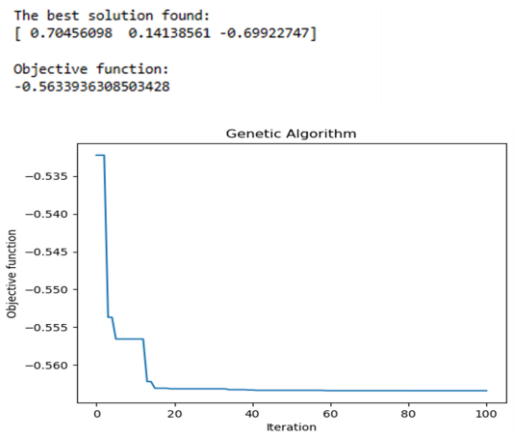
۴-۴- انتخاب بهترین نتیجه خوشه‌بندی با محاسبه شاخص‌های سیلوهت و کالینسکی

در این مرحله، با استفاده از الگوریتم ژنتیک دقت خوشه‌بندی بهبود می‌یابد. ابتدا، پیش از اجرای الگوریتم ژنتیک، خوشه‌بندی با استفاده از چهار الگوریتم طیفی، سلسه مراتبی، فازی C-means و K-means انجام می‌شود انجام می‌شود سپس پس از اجرای الگوریتم ژنتیک با وزن‌های بهینه، مرحله دیگر خوشه‌بندی با همان چهار الگوریتم انجام شده است. با استفاده از معیارهای سیلوهت و کالیسنکی، کیفیت خوشه‌بندی ارزیابی گردید. این مقایسه مراحل مختلف نشان می‌دهد که الگوریتم ژنتیک با وزن‌های بهینه، بهبود قابل ملاحظه‌ای در دقت خوشه‌بندی به ارمغان آورده است.

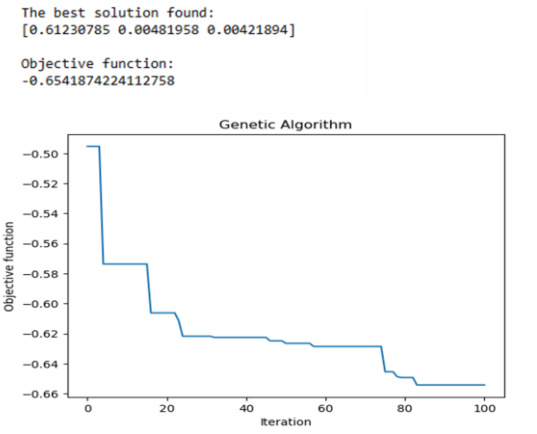


شکل ۲: خروجی الگوریتم ژنتیک با تابع تناسب K-Means

شکل ۳: خروجی الگوریتم ژنتیک با تابع تناسب Spectral



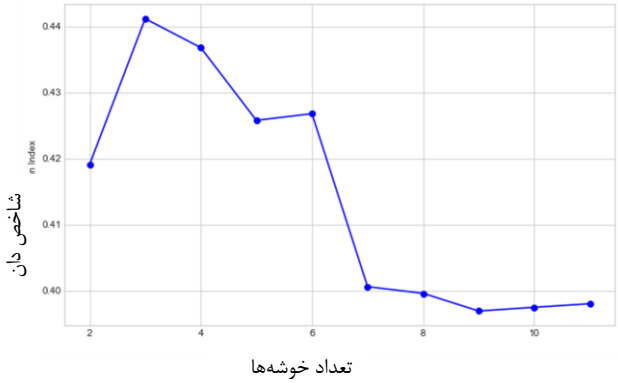
شکل ۴: خروجی الگوریتم ژنتیک با تابع تناسب فازی C-menas



شکل ۵: خروجی الگوریتم ژنتیک با تابع تناسب سلسله مراتبی

جدول۲. نتایج محاسبه شاخص دان

تعداد خوشه بهینه	شاخص دان
۲	۰/۴۱۹۰
۳	۰/۴۴۱۱
۴	۰/۳۹۸۳
۵	۰/۴۲۵۷
۶	۰/۴۲۶۸
۷	۰/۴۰۰۶
۸	۰/۳۹۹۷
۹	۰/۳۹۸۷
۱۰	۰/۳۹۶۸
۱۱	۰/۳۹۸۳

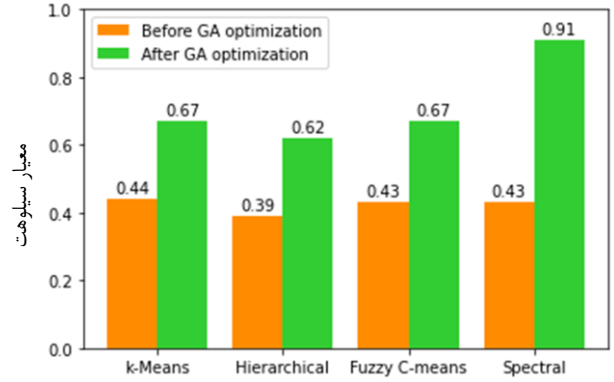


شکل۶.نمودار تغییرات شاخص دان بر حسب تعداد خوشه‌ها

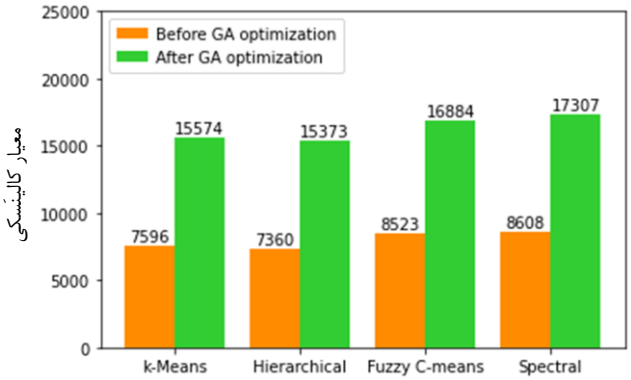
در شکل‌های ۷ و ۸، مقایسه میان معیارهای سیلوهت و کالینسکی برای الگوریتم‌های مختلف قبل و بعد از اجرای الگوریتم ژنتیک آمده است. این نمودارها نشان می‌دهند که پس از اعمال بهینه‌سازی توسط الگوریتم ژنتیک، افزایش چشمگیری در مقادیر معیارهای سیلوهت و کالینسکی برای تمام الگوریتم‌ها مشاهده شده است. این افزایش نشان‌دهنده بهبود و بهتر شدن تقسیم‌بندی خوشه‌ها پس از اجرای الگوریتم ژنتیک می‌باشد. نتایج معیارهای سیلوهت و کالینسکی نشان می‌دهد که الگوریتم خوشه‌بندی طیفی با دقت بالاتری نسبت به سایر الگوریتم‌ها عمل کرده است. لذا در پی بهبود و افزایش دقت خوشه‌بندی در ادامه خوشه‌بندی مشتریان با استفاده از الگوریتم طیفی انجام شده است.

۴-۵- برچسب‌گذاری هر بخش مشتری و تحلیل رفتار هر گروه

این بخش از مقاله، به تحلیل گروه‌های مختلف مشتریان پرداخته است. با محاسبه مرکز هر گروه، اطلاعات زمانی آن بخش مشخص می‌شود. شکل‌های ۹، ۱۰ و۱۱ نمودار تغییرات میانگین مقادیر هر

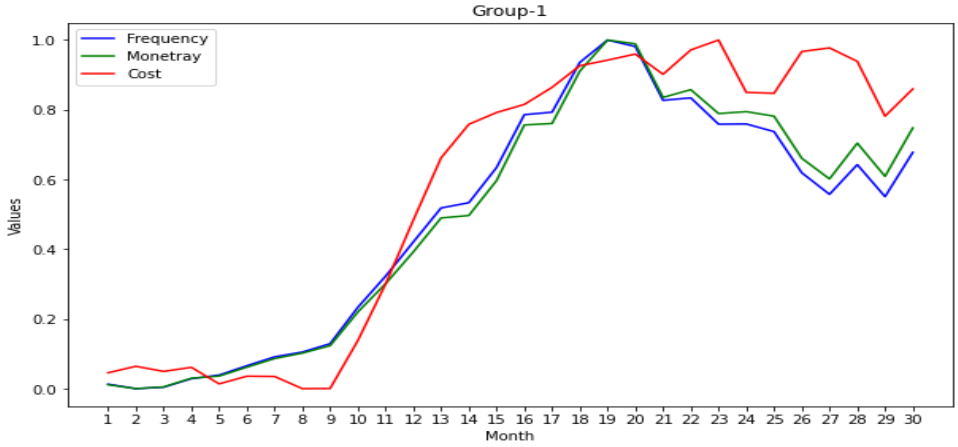


شکل۷. مقایسه معیار سیلوهت قبل و بعد از بهینه‌سازی خوشه‌بندی با الگوریتم ژنتیک

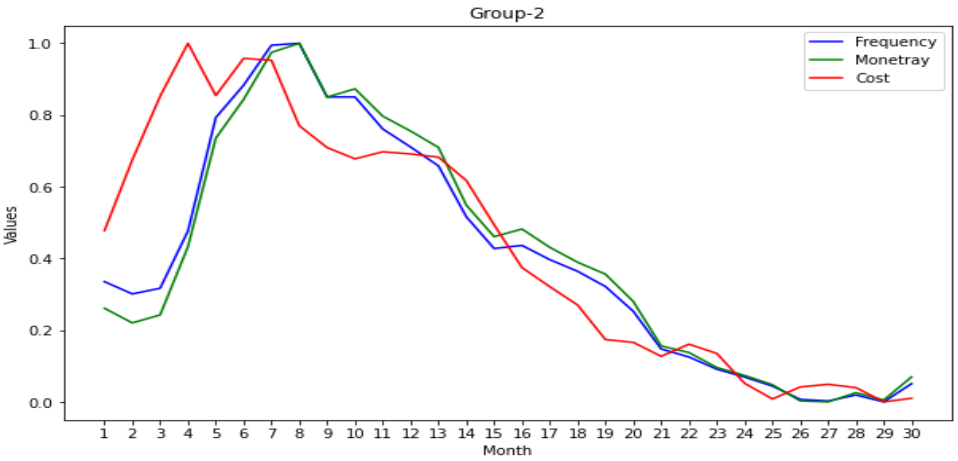


شکل۸. مقایسه معیار کالینسکی قبل و بعد از بهینه‌سازی خوشه‌بندی با الگوریتم ژنتیک

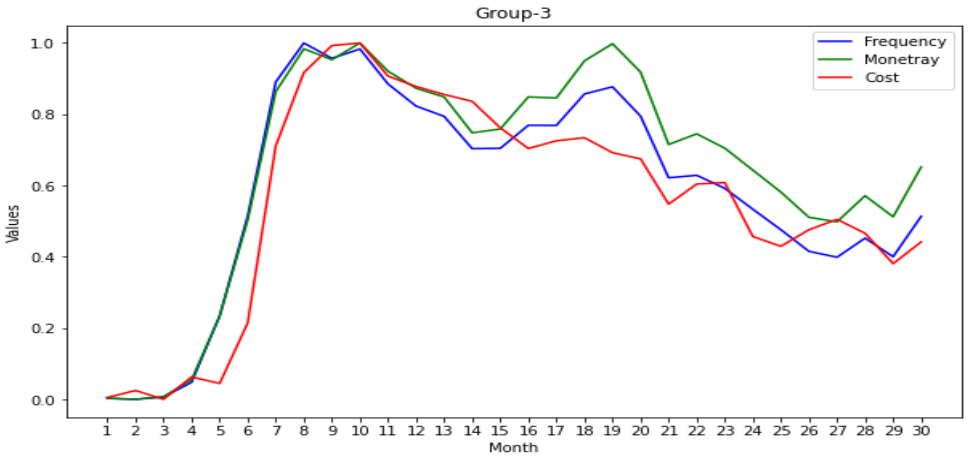
سه ویژگی سود تراکنش، تعداد تراکنش و هزینه را بر اساس گروه‌های خوشه‌بندی نشان می‌دهد. در این نمودار، هر یک از ویژگی‌های سود تراکنش، تعداد تراکنش و هزینه با یک رنگ خاص نمایش داده شده است. محور x نمودار، تاریخ‌ها را نمایش می‌دهد و محور y نمودار، میانگین مقادیر ویژگی‌ها برای هر گروه در هر تاریخ را نشان می‌دهد. با مشاهده این نمودار می‌توان تغییرات میانگین مقادیر ویژگی‌ها را برای هر یک از گروه‌ها در طول زمان مشاهده نمود. این نمودار نشان می‌دهد که عملکرد گروه‌های خوشه‌بندی چگونه در طول زمان تغییر کرده‌اند و آیا میانگین مقادیر ویژگی‌های آن‌ها افزایش یا کاهش یافته است. همچنین می‌توان با مقایسه گروه‌های مختلف، مشاهده نمود که کدام گروه بهترین عملکرد را از نظر میانگین مقادیر ویژگی‌ها داشته است. در کدام بازه‌های زمانی تغییرات قابل مشاهده‌ای داشته و از این اطلاعات در تحلیل و بررسی داده‌های مشتریان بهره‌مند شد. شکل ۱۲ نشان‌دهنده مقادیر ویژگی‌های تراکنش، سود و هزینه برای هر یک از گروه‌های خوشه‌بندی شده است. در این نمودار، هر ستون مربوط به یک ویژگی است و برای هر ویژگی سه میله به ترتیب برای گروه‌های یک، دو و سه کشیده شده است. همچنین، هر گروه با یک رنگ



شکل۹. نمودارخطی میانگین مقادیر ویژگی‌های خوشه اول



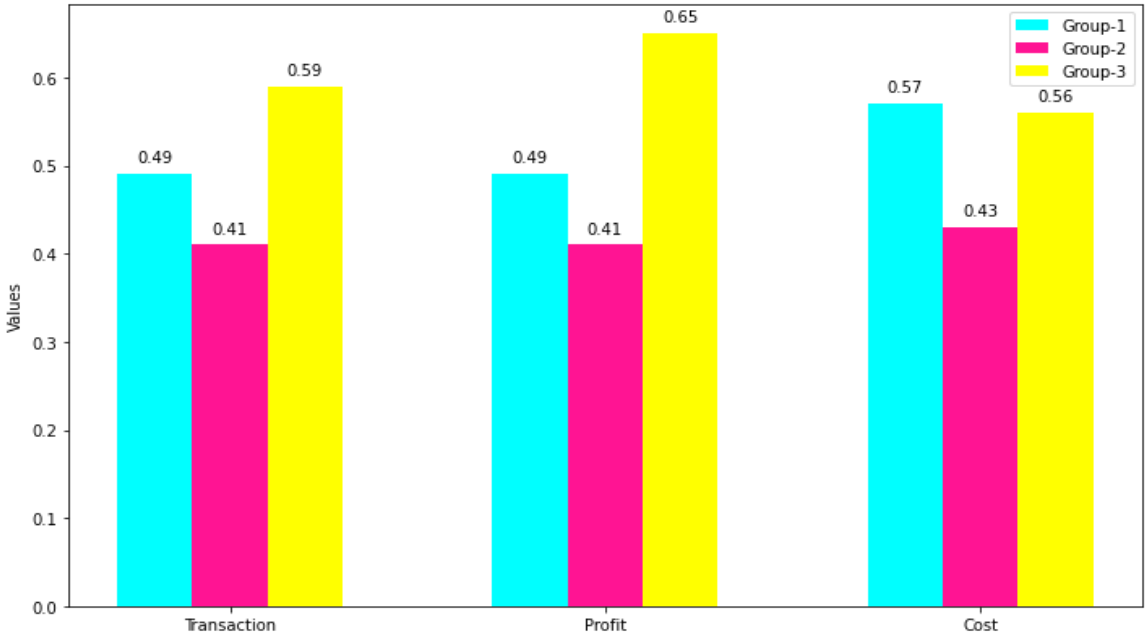
شکل۱۰. نمودارخطی میانگین مقادیر ویژگی‌های خوشه دوم



شکل۱۱. نمودارخطی میانگین مقادیر ویژگی‌های خوشه سوم

متمایز شده است. مقادیر نمایش داده شده در این نمودار، نمایان‌گر میانگین مقادیر مربوط به هر ویژگی برای هر گروه می‌باشد. این نمودار به‌طور کلی نشان می‌دهد که هر یک از گروه‌ها چه میزان از نظر ویژگی‌های مورد بررسی با سایر گروه متفاوت هستند. با توجه به این اطلاعات آماری، می‌توان نتیجه گرفت که گروه سوم نسبت به دو گروه دیگر، میانگین تعداد تراکنش و سود تراکنش بیشتری دارد. لذا این گروه مشتریان فعال‌تری نسبت به دو گروه دیگر دارد. بعد از گروه سوم، گروه اول دارای سهم تراکنشی و سود بیشتری در مجموع می‌باشند. بر اساس این دو ویژگی گروه دوم در جایگاه سوم قرار می‌گیرد. زیرا مشتریان گروه دوم دارای کمترین سهم تراکنشی

و سود نسبت به سایر گروه‌ها هستند. از طرف دیگر، مشاهده می‌شود که هزینه مشتریان در گروه دوم کمتر از دو گروه دیگر است. این نتیجه نشان می‌دهد که گروه دوم سوم قرار دارد و سپس گروه اول. این نکته نشان‌دهنده این است که مشتریان گروه دوم به طور میانگین با هزینه کمتری تراکنش‌های خود را نسبت به دو گروه دیگر انجام می‌دهند (جدول۳). به این ترتیب با تحلیل این ویژگی‌ها و مقادیر آماری مرتبط با هر گروه، می‌توان تفاوت‌ها و مزیت‌های هر گروه در کسب و کار را بهتر درک کرد و تصمیم‌گیری‌های بهتری در مورد استراتژی‌ها و نگهداری مشتریان انجام داد. البته زمان هم در این تحلیل بسیار مهم است با توجه به شکل‌های ۹ تا ۱۱، که نمودار میانگین مقادیر بر اساس گروه‌ها را نشان می‌دهد، می‌توان مشاهده کرد که مشتریان گروه سوم در بازه سی‌ماهه مورد بررسی رفتاری متعادل و یکنواخت داشته‌اند. این نشان دهنده وفاداری آن‌هاست. مشتریان گروه دوم درماه‌های اولیه، سود و تراکنش‌های بالایی داشتند اما پس از مدتی، مقادیر این ویژگی‌ها کاهش یافتند و آن‌ها در شرف دست رفتن می‌باشند. این نکته نشان‌دهنده نیاز به اتخاذ استراتژی‌های مؤثر جهت حفظ مشتریان گروه دوم و جلوگیری از، از دست دادن آن‌هاست. به علاوه، مشتریان گروه اول در ابتدا



شکل ۱۲. نمودار مقایسه‌ای میانگین مقادیر ویژگی‌ها بر اساس هر گروه از مشتریان

جدول۳. تحلیل رفتار هر گروه از مشتریان

تحلیل رفتار	بخش	سایز	برچسب
مشتریان این گروه در ابتدا سود کمی داشتند اما اخیراً رشد قابل‌توجهی داشته‌اند و به مشتریان پرسودی تبدیل شده‌اند. برای این گروه می‌توان استراتژی‌هایی اندیشید تا این مشتریان ترغیب به انجام تراکنش‌های بیشتری شوند. به طور کلی، استراتژی‌های محوری بانک باید بر روی ارتقاء و تقویت تجربه و رضایت مشتریان این گروه تمرکز کنند.	خوشه اول	۱۵۵۱۳	نقره‌ای
مشتریان این گروه در ماه‌های اولیه، سود و تراکنش‌های بالایی داشتند اما پس از مدتی، مقادیر این ویژگی‌ها کاهش یافتند و آن‌ها در شرف از دست رفتن می‌باشند. این موضوع نشان‌دهنده نیاز بانک به در پیش گرفتن استراتژی‌های مؤثر جهت حفظ مشتریان و جلوگیری از، از دست دادن آن‌هاست.	خوشه دوم	۳۵۵۷	برنزی
این گروه که به عنوان پرسودترین گروه شناخته می‌شود شامل مشتریان طلایی است. زیرا منبع بالقوه درآمدی برای سازمان محسوب می‌شود. همچنین مشتریان این گروه در بازه زمانی سی‌ماهه رفتاری متعادل و یکنواخت داشتند این نشان‌دهنده وفاداری آنهاست. با توجه به اهمیت این دسته مشتریان خدمات ویژه‌ای برای نگه‌داشت آنها توسط بانک باید طراحی شود.	خوشه سوم	۱۸۱۴	طلایی

تراکنش‌های کمی داشتند اما اخیراً رشد قابل‌توجهی داشته‌اند و به مشتریان پرسودی تبدیل شده‌اند. به طور کلی، با تحلیل این داده‌های آماری و نمودارها، می‌توان تفاوت‌ها و رفتارهای هر گروه را بهتر درک کرد و تغییرات لازم را در کسب و کار انجام داد.

۵- ارزیابی کارایی

در این بخش، به ارزیابی کارایی چارچوب پیشنهادی پرداخته می‌شود که شامل سه زیربخش است که در ادامه توضیح داده می‌شود.

۵-۱- نوآوری‌های تحقیق

از جمله مهم‌ترین نوآوری‌های مقاله این است که در مقابل اکثر روش‌های قبلی که وزن یکسان برای پارامترهای خرید جدید، تکرار خرید و مبلغ خرید در نظر می‌گرفتند، در این مقاله با استفاده از الگوریتم ژنتیک وزن‌های بهینه برای ویژگی‌های مشتریان به دست آمده است. در این مدل، با بهینه‌سازی و تنظیم وزن‌ها با استفاده از الگوریتم ژنتیک، ویژگی‌های مختلف به صورت مناسبی در بخش‌بندی مشتریان استفاده شده‌اند. یکی دیگر از نوآوری‌های چارچوب پیشنهادی این است که بُعد زمان را به مدل تراکنش جدید، تعداد تراکنش، مبلغ تراکنش اضافه می‌کند تا با استفاده از سری‌های زمانی، پویایی‌های مشتری را ثبت و تحلیل کند. یکی دیگر از نوآوری‌های تحقیق حاضر این است که در بخش‌بندی مشتریان، به ویژگی هزینه نیز توجه شده است. در نمونه مورد بررسی که شامل تراکنش‌های دستگاه کارتخوان است، هزینه‌هایی که مشتریان در طول رابطه با سازمان دارند در نظر گرفته شده‌اند. هزینه‌ها شامل هزینه‌های رول برای تراکنش‌ها، هزینه‌های بازدیدهای دوره‌ای، خرابی دستگاه کارتخوان و سایر هزینه‌هایی است که سازمان برای مشتریان خود می‌پردازد. با در نظر گرفتن ویژگی هزینه در بخش‌بندی مشتریان، می‌توان بهترین راهکارها را برای مدیریت هزینه‌ها و بهینه‌سازی منابع سازمان پیدا کرد. با تحلیل هزینه‌های مشتریان و تاثیر آنها بر درآمد و سود سازمان، می‌توان راهبردهایی تدوین کرد تا هزینه‌ها را کاهش داد. همچنین، می‌توان گروه‌های مشتری با الگوهای هزینه‌ای مشابه را شناسایی کرده و راهکارهای بهینه برای مدیریت هزینه‌های سازمان ارائه داد. مدل بخش‌بندی مشتریانی که در این مقاله توسعه داده شده است با موفقیت در یکی از شرکت‌های پرداخت الکترونیکی به کار گرفته شده و نتایج بهتری نسبت به مدل بخش‌بندی پیشین که در شرکت استفاده می‌شد، ارائه داده است. لذا این مدل تجاری‌پذیر بوده و قابلیت استفاده در شرایط واقعی را دارد.

۵-۲- ادغام چارچوب پیشنهادی در یک سیستم اطلاعاتی برای بهبود تصمیم گیری

درصورتی که ورودی یک سیستم اطلاعاتی داده‌های تراکنش‌های مشتریان باشد، ابزارهای تحلیل سری‌های زمانی مانند خوشه‌بندی طبقه‌بندی سری‌های زمانی و پیش‌بینی سری‌های زمانی می‌توانند به مدیران در تصمیم‌گیری‌های مرتبط با مشتریان در طول دوره عمر آن‌ها کمک کنند. برخی از کاربردهایی که بر اساس این رویکرد پیشنهادی قابل اجرا هستند عبارتند از:

۱. تقسیم‌بندی مشتریان[۱۸]: با استفاده از تکنیک‌های خوشه‌بندی سری‌های زمانی، تجزیه و تحلیل گروه‌های مختلف رفتاری مشتریان در طول زمان انجام می‌شود.
۲. شناسایی خروج مشتریان[۱۹]: ریزش مشتری یک مشکل مهم در شرکت ها است. شناسایی مشتریان در حال ریزش و اتخاذ اقدام مناسب برای نگهداری آنها ضروری است[۲۰]. تشخیص افرادی که قصد خروج از لیست مشتریان را دارند و انجام اقدامات مناسب برای حفظ آنها، بسیار ضروری است. مشابه تقسیم‌بندی مشتریان، اکثر مطالعات در زمینه شناسایی خروج مشتریان بر اساس رویکردهای ثابت انجام می‌شود. اما با استفاده از تداوم رفتار پویای مشتریان و با استفاده از سری‌های زمانی و برچسب‌گذاری هر مشتری به عنوان وفادار یا خروج‌کننده، می‌توان ابزاری قدرتمند برای شناسایی خروج مشتریان ایجاد کرد.
۳. پیش‌بینی رفتار مشتریان[۲۰]: پیش‌بینی رفتار مشتریان به عنوان یکی از زمینه‌های حیاتی در مدیریت ارتباط با مشتریان دارای اهمیت بسیاری است [۲۱]. با در نظر گرفتن رفتار مشتری به عنوان یک دنباله زمانی، روش‌های قوی پیش‌بینی سری‌زمانی می‌توانند به نیروی کمکی برای مدیران تبدیل شوند. اصلی‌ترین تفاوت این رویکرد با تحقیقات پیشین، بهره‌برداری از خصوصیت پیش‌بینی پذیری در آن است.

۵-۳- محدودیت‌های تحقیق

اولین محدودیت تحقیق پردازش حجم بزرگ داده‌ها بوده است. این مقاله با مواجهه با حجم گسترده داده‌ها نیازمند منابع قوی از نظر سخت‌افزاری بوده و برای پردازش و تحلیل داده‌ها نیاز به منابع پردازشی قدرتمند داشته است. دومین محدودیت تحقیق این است که تنها به داده‌های تراکنشی مشتریان دسترسی داشته و اطلاعات دیگری از مشتریان مانند مکان، اندازه فروشگاه و تعداد پرسنل برای

تجزیه و تحلیل دقیق‌تر بخش‌های به دست آمده، در اختیار ما نبوده‌اند.

۶- نتیجه‌گیری و کارهای آینده

در این مقاله یک مدل بخش‌بندی مشتریان برای بخش‌بندی پویای مشتریان مبتنی بر مدل‌سازی رفتار مشتریان ارائه شده است. بخش‌بندی مشتریان، با در نظر گرفتن طول عمر مشتری از ابتدا تا لحظه حال انجام شده است در واقع تنها به بررسی رفتار گذشته مشتریان توجه نداشته است و نگاهی به رفتار آتی مشتریان نیز داشته است. لذا در این تحقیق طول عمر مشتری به صورت کامل در بخش‌بندی دیده شده است. این تحقیق یک مدل پویا برای بخش‌بندی مشتریان ارائه داده است. این مدل، با استفاده از خوشه‌بندی سری‌های‌زمانی، قادر به رصد تغییرات رفتاری مشتریان است. این تحول از بخش‌بندی ثابت به بخش‌بندی پویا، امکان بهبودهای مداوم و سازگار با رفتار مشتریان را فراهم می‌کند. این تحقیق به ویژگی‌های خرید جدید، تکرار خرید و مبلغ خرید به صورت ترکیب خطی از آنها توجه کرده و روابط بین این ویژگی‌ها را نادیده نگرفته است. با استفاده از الگوریتم ژنتیک، وزن‌های بهینه به ویژگی‌های مشتریان داده شده است و این منجر به بهبود بخش‌بندی مشتریان نسبت به مدل‌های قبلی شده است.از جمله نتایج این تحقیق می‌توان به مدل‌سازی بخش‌بندی مشتریان بر اساس تغییر رفتار در طول عمر مشتری، استفاده از تحلیل سری‌زمانی، در نظر گرفتن ویژگی هزینه در بخش‌بندی و توانایی بهبود مداوم بخش‌بندی مشتریان با توجه به تغییرات رفتاری اشاره کرد. این مدل با موفقیت در یکی از شرکت‌های پرداخت الکترونیکی استفاده شده و نتایج بهتری نسبت به مدل‌های پیشین ارائه داده است.الگوریتم ژنتیک با مقداردهی بهینه به وزن‌ها، توانسته است یک مدل خوشه‌بندی با دقت بالا ارائه دهد. مقدار شاخص دان بدست آمده نشان‌دهنده‌ی انطباق بسیار خوب خوشه‌بندی با داده‌ها و جداپذیری مناسب بین خوشه‌ها است. این به این معناست که وزن‌های بهینه که توسط الگوریتم ژنتیک به دست آمده‌اند، به خوبی ویژگی‌ها را در خوشه‌بندی مشتریان نمایان می‌سازند و مدل خوشه‌بندی حاصل، نتایج دقیق‌تری را ارائه می‌دهد. سپس، با استفاده از این وزن‌های بهینه، الگوریتم‌های خوشه‌بندی اجرا شدند و بخش‌بندی مشتریان انجام شد. این وزن‌های بهینه موفق به تنظیم و انتخاب بهترین ویژگی‌ها برای خوشه‌بندی مشتریان شده‌اند. نتیجه حاصل از خوشه‌بندی با این وزن‌ها در بخش خوشه‌بندی قابل مشاهده است. با استفاده از این نتایج، می‌توان ادعا کرد که الگوریتم ژنتیک با پارامترهای انتخابی موفق به پیدا کردن وزن‌های بهینه برای خوشه‌بندی مشتریان شده است و خوشه‌بندی حاصل دقیق‌تر و بهتر از قبل انجام شده است. در این مقاله، با استفاده از الگوریتم ژنتیک و بهینه‌سازی وزن‌ها، به یک روش ترکیبی و کارا برای بخش‌بندی مشتریان دست یافته شد. همانطور که در جدول ۴ مشخص است دقت بخش‌بندی بهبود معناداری نسبت به روش‌های دیگر داشته است. به علاوه، با تحلیل‌های انجام شده، مشخص گردید که حجم بزرگ داده‌ها (۱۹۵،۸۴۴،۰۵۸ تراکنش و ۴۸۹،۴۸ مشتری) در تحلیل‌ها و بخش‌بندی مشتریان، شامل جامعیت و گستردگی بیشتری است و این مسئله نتایج دقیق‌تر و معتبرتری را برای بخش‌بندی مشتریان ارائه داده است. نکته دیگری که متمایز کننده‌ی کار حاضر از مقالات مرجع است، استفاده از ویژگی هزینه در بخش‌بندی مشتریان است. برخلاف سایر مقالات که این ویژگی را در بخش‌بندی لحاظ نکردند. این تحقیق ترکیب خطی از ویژگی‌ها را در بخش‌بندی مشتریان استفاده کرده است و همچنین داده‌های مشتریان را در طول سی‌ماه مورد بررسی قرار داده است. در صورتی که در کارهای مشابه بازه زمانی مورد بررسی هفت یا یازده ماه بوده است. بنابراین، با توجه به بهبود دقت بخش‌بندی، حجم بزرگ داده‌ها، لحاظ کردن ویژگی هزینه و تفاوت در طول مدت مورد بررسی، نتیجه‌گیری می‌شود که نتایج این تحقیق موفقیت‌آمیز و مؤثر بوده و بهبود قابل توجهی نسبت به مقالات مرجع داشته است.

در تحقیقات آتی، می‌توان دامنه داده‌ها را افزایش داده و تاثیر سایر متغیرهای مشتریان مانند موقعیت جغرافیایی، سن، جنسیت و سایر ویژگی‌ها را نیز در بخش‌بندی بررسی کرد. این گسترش دامنه داده‌ها این امکان را می‌دهد تا به صورت جامع‌تری رفتار مشتریان تحلیل شده و بررسی‌های دقیق‌تری انجام شود. همچنین می‌توان با استفاده از روش‌های یادگیری عمیق مانند شبکه‌های عصبی عمیق، بهبود قابل توجهی در پیش‌بینی رفتار مشتریان داشت. این روش‌ها می‌توانند الگوهای پیچیده‌تر رفتار مشتریان را تشخیص دهند و دقت پیش‌بینی را افزایش دهند. همانطور که در این تحقیق عنوان شد، استفاده از الگوریتم های تکاملی مانند الگوریتم ژنتیک می‌تواند تاثیر بسزایی در بهبود کارایی استخراج ویژگی داشته باشد. به همین منظور یکی از جدیدترین اعضای خانواده الگوریتم‌های تکاملی به نام برنامه نویسی ژنتیک کارتیزین[۲۱] معرفی می‌گردد. با توجه به بهبود نتایج در مقاله‌های مختلف که از روش برنامه نویسی ژنتیک کارتیزین استفاده کرده‌اند نسبت به نمونه مشابه که از روش برنامه نویسی ژنتیک استفاده کرده‌اند، پیشنهاد می‌شود ترکیب ویژگی‌ها با کمک برنامه نویسی ژنتیک کارتیزین انجام شود. با ادامه تحقیقات و انجام تحقیقات آینده بر اساس نتایج مقاله، می‌توان به بهبود بخش‌های مختلف مدل بخش‌بندی مشتریان پرداخته و راهکارهای

بهینه‌تری برای بهبود رابطه با مشتریان و افزایش ارزش عمر آنها ارائه داد.

مراجع

[1] V. Kumar and W. Reinartz, "Customer relationship management: Concept, strategy, and tools", Springer, 2018.

[2] A. Parvaneh, M. Tarokh and H. Abbasimehr, "Combining data mining and group decision making in retailer segmentation based on LRFMP variables", International Journal of Industrial Engineering & Production Research, 25(3), 197–206, 2014.

[3] E. Akhondzadeh-Noughabi, and A. Albadvi, "Mining the dominant patterns of customer shifts between segments by using top-k and distinguishing sequential rules", Management Decision, 53(9), 1976-2003, 2015.

[4] A. Seret, S.K. vanden Broucke, B. Baesens and J. Vanthienen, "A dynamic understanding of customer behavior processes based on clustering and sequence mining", Expert Systems with Applications, 41(10), 4648–4657, 2014.

[5] I. Yanovitzky and A. VanLear, "Time series analysis: Traditional and contemporary approaches", The Sage Sourcebook of Advanced Data Analysis Methods for Communication Research, 89–124, 2008.

[6] J.N. SARI, L.E. NUGROHO, R. FERDIANA and P.I. SANTOSA, "Review on customer segmentation technique on ecommerce", Advanced Science Letters, 22, 3018-3022, 2016.

[7] M. Khajvand and M.J. Tarokh, "Estimating customer future value of different customer segments based on adapted RFM model in retail banking context", Procedia Computer Science, 3, 1327-1332, 2011.

[8] M. Khajvand, K. Zolfaghar, S. Ashoori and S. Alizadeh, "Estimating customer lifetime value based on RFM analysis of customer purchase behavior: Case study", Procedia Computer Science, 3, 57-63, 2011.

[9] R. Heldt, C.S. Silveira and F.B. Luce, "Predicting customer value per product: From RFM to RFM/P", Journal of Business Research, 0148-2963.

[10] P. Anitha, and M.M. Patil, "RFM model for Customer Purchase Behavior using K-Means Algorithm", Journal of King Saud University-Computer and Information Sciences, 1319-1578, 2019.

[11] A.Daneshvar, M. Homayounfar and A. FarahmandNezhad, "Development of an intelligent multi-criteria clustering method based on Promethee", Industrial Management Perspective, 36, 41-6, 2020.

[12] M. Tavakoli, M. Molavi and V. Masoumi, "Customer Segmentation and Strategy Development based on User Behavior Analysis RFM model and Data Mining Techniques: A Case Study", IEEE e-Business Engineering Conf, 2018.

[13] H. ABBASIMEHR & A. BAHRINI, "An analytical framework based on the recency, frequency, and monetary model and time series clustering techniques for dynamic segmentation", Expert Systems with Applications, 192, 116373, 2022.

[14] A.J. Christy, A. Umamakeswari, L. Priyatharsini and A. Neyaa, "RFM ranking–An effective approach to customer segmentation", Journal of King Saud University-Computer and Information Sciences, 2018.

[15] A.B. Çavdar and N. Ferhatosmanoğlu, "Airline customer lifetime value estimation using data analytics supported by social network information", Journal of Air Transport Management, 67, 19-33, 2018.

[16] H. Emami, F. Derakhshan, Integrating Fuzzy K-Means, Particle Swarm Optimization, and Imperialist Competitive Algorithm for Data Clustering. Arab J Sci Eng 40, 3545–3554 , 2015.

[17] G.E. Batista, E.J. Keogh, O.M. Tataw and V.M. De Souza, "CID: an efficient complexity-invariant distance for time series", Data Mining and Knowledge Discovery, 28(3), 634–669, 2014.

[18] O. Arbelaitz, I. Gurrutxaga, J. Muguerza, J.M. Pérez and I. nigo Perona, "An extensive comparative study of cluster validity indices. Pattern Recognition", 46(1), 243–256, 2014.

[19] J.C. Dunn, "A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters", Journal of Cybernetics. 3 (3): 32–57, 1973.

[20] N. Alboukaey, A. Joukhadar and N. Ghneim, "Dynamic behavior based churn prediction in mobile telecom", Expert Systems with Applications, 162, 2020.

[21] P. Montero-Manso and R.J. Hyndman, "Principles and algorithms for forecasting groups of time series: Locality and globality", International Journal of Forecasting, 2021.

[22] H. ABBASIMEHR & M. Shabani, "A new methodology for customer behavior analysis using time series clustering A case study on a bank's customers", Emerald , 50 (2), 221-242, 2021.

[23] H. ABBASIMEHR & F. SheikhBaghery, "A novel time series clustering method with fine-tuned support vector regression for customer behavior analysis", Expert Systems with Applications, 204, 117584, 2022.

پاورقی‌ها:

[1] Recency (R)
[2] Frequency (F)
[3] Monetray (M)
[4] Cost (C)
[5] TimeSeries Clusring
[6] Hierarchical Clustring
[7] Spectral Clustring
[8] Mutation
[9] Genetic Algorithm
[10] Meta-heuristic algorithm
[11] Recency, Frequency, Monetray, Cost
[12] Euclidean distance
[13] Manhattan distance
[14] Silhouette measure
[15] Caliński–harabasz measure
[16] Min-Max Normalization
[17] Dunn Index
[18] Customer Segmentation
[19] Churn Detection
[20] Customer Behavior Forecasting
[21] Cartesian Genetic Programming