Journal of Business Intelligence Management Studies
Vol. 10, No.38, Winter 2022
ims.atu.ac.ir
DOI: 10.22054/IMS.2021.58551.1902

ATU PRESS

Original Research
Accepted: 2021-12-26
Revised: 2021-10-31
Received: 2021-01-24
eISSN: 2538-2527
ISSN: 1735/112X

Modified LRFM in Order to Bank Customer Clustering Based on Genetic Algorithm

Moahmmad Kazemi — PhD Student, Information Technology Management, Faculty of Management, Islamic Azad University, Central Tehran Branch, Tehran.

Mohammad Ali Keramati * — Faculty Member, Department of Industrial Management, Faculty of Management, Islamic Azad University, Central Tehran Branch, Tehran

Mehrzad Minooie — Faculty Member, Department of Industrial Management, Faculty of Management, Islamic Azad University, Central Tehran Branch, Tehran

Abstract

Clustering is a common method for analyzing various data that is used in many fields, including statistical pattern recognition, machine learning, data mining, image analysis, and bioinformatics. Clustering the process of grouping objects similar to different groups, or more precisely, partitioning and dividing a set of data, into separate subcategories, the main point of which is not to be specific. The number of classes is in clustering. One of its most widely uses is in the field of data, the clustering of which is performed by experts in taste. Bank customer clustering has been a challenge from the beginning, and it has been difficult to find consensus among experts to select a feature for grouping. This paper seeks to provide a solution for dynamic clustering of bank customers. This clustering will be based on a genetic algorithm and will decide on the number of categories, members of each category, and the similarity criteria used. The dynamics of the method are based on the improvement of the LRFM method using the genetic algorithm. In other words, the genetic algorithm will try to find different information fields about the bank's customers in the database; Put the right fields next to

* Corresponding Author: mohammadalikeramati@yahoo com
How to Cite: Kazemi, M., Keramati, M. A., Minooie, M. (2022). Modified LRFM in Order to Bank Customer Clustering Based on Genetic Algorithm , *Journal of Business Intelligence Management Studies*, 10(38), 317-356.

the features used in the LRFM method and get better results for clustering the bank's customers. This process leads to the criterion determination of similarity of one customer with another and the degree of similarity between them.

Keywords: Bank Customer Relationship Management; Clustering; LRFM Model; Genetic Algorithm.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

مقاله پژوهشی
تاریخ ارسال: ۱۳۹۹/۱۱/۰۵
تاریخ بازنگری: ۹۹/۰۴/۲۵
تاریخ پذیرش: ۹۹/۰۷/۲۷
ISSN: 1735/112X
eISSN: 2538-2527

مطالعات مدیریت کسب و کار هوشمند
سال دهم، شماره ۳۸، زمستان ۱۴۰۰، ص ۳۱۷ تا ۳۵۶
ims.atu.ac.ir
DOI: 10.22054/IMS.2021.58551.1902

ارایه مدل بهبودیافته LRFM به جهت خوشه‌بندی مشتریان بانک‌ها بر مبنای الگوریتم ژنتیک

محمد کاظمی دانشجوی دکتری مدیریت فناوری اطلاعات، دانشکده مدیریت، دانشگاه آزاد اسلامی واحد تهران مرکزی، تهران، ایران.

محمدعلی کرامتی * دانشیار گروه مدیریت صنعتی، دانشکده مدیریت، دانشگاه آزاد اسلامی واحد تهران مرکزی، تهران، ایران.

مهرزاد مینوئی استادیار گروه مدیریت صنعتی، دانشکده مدیریت، دانشگاه آزاد اسلامی واحد تهران مرکزی، تهران، ایران.

چکیده

خوشه‌بندی یک روش رایج برای تجزیه‌وتحلیل داده‌های مختلف در بسیاری از زمینه‌ها از جمله شناسایی آماری، بانکداری، داده‌کاوی، تجزیه‌وتحلیل تصویر و غیره می‌باشد. خوشه‌بندی فرآیند گروه‌بندی اشیای مشابه به گروه‌های مختلف یا به بیان دقیق‌تر، تقسیم مجموعه‌ای از داده‌ها به زیرمجموعه‌های مجزا است. نکته اصلی، مشخص نبودن تعداد گروه‌ها در خوشه‌بندی است؛ به‌گونه‌ای که در خوشه‌بندی متخصصین، سلیقه‌ای عمل می‌نمایند. این مقاله به دنبال ارایه راه‌حلی برای خوشه‌بندی پویای مشتریان بانک، بر مبنای الگوریتم ژنتیک با لحاظ نمودن روش LRFM می‌باشد. به عبارت دقیق‌تر، الگوریتم ژنتیک سعی خواهد کرد از بین فیلدهای اطلاعاتی مختلفی که در مورد مشتریان بانک در پایگاه داده وجود دارد؛ فیلدهای مناسبی را در کنار ویژگی‌های به کار رفته در روش LRFM قراردهد تا نتایج مناسب‌تری را در خوشه-بندی مشتریان بانک به دست آورد، تصمیم‌گیری در مورد تعداد گروه‌های موجود نیز بر عهده الگوریتم ژنتیک خواهد بود. پیاده‌سازی‌های مختلف الگوریتم ژنتیک با استفاده از تابع‌های مختلف جهش و پیوند انجام پذیرفته تا دستیابی به بهترین حالت پیاده‌سازی گردد. شایان ذکر است که تمامی روش‌های پیوند و

مقاله حاضر برگرفته از رساله دکتری رشته مدیریت فناوری اطلاعات، دانشگاه آزاد اسلامی واحد تهران مرکز می‌باشد.

* نویسنده مسئول: mohammadalikeramati@yahoo com

جهش (به دلیل تعدد زیاد) قابل آزمون نیستند. لیکن، روند به گونه‌ای طراحی‌شده تا در پیاده‌سازی، بهبود آن نسبت به روش پایه LRF و برخی روش‌های رقیب حاصل گردد.

کلیدواژه‌ها: مدیریت ارتباط با مشتریان بانک، خوشه¬بندی، مدل **LRFM**، الگوریتم ژنتیک.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

مقدمه

شناخت مشتریان و ایجاد تمایز بین آن‌ها از مسایل مهم در حوزه سازمان‌های مشتری‌مدار است. اولین نکته در شناخت مشتریان و ایجاد تمایز بین آن‌ها، استفاده از روش مناسب به همراه معیارهای لازم است (Lakshmanaprabu Et Al, 2018). این موضوع مقیاسی مناسبی را در اختیار قرار می‌دهد که بتوان به آن اعتماد کرد. این مساله، حل یکی از مشکلات اصلی حوزه بانکداری است که ارتباط تنگاتنگی با حوزه فناوری اطلاعات دارد. ترکیب بحث مدیریتی این موضوع با حیطه فناوری اطلاعات یکی از مباحث مهم حوزه مدیریت فناوری اطلاعات را رقم خواهد زد که می تواند به ارایه راه حلی برای خوشه بندی مشتریان بانک بیانجامد. خوشه‌بندی مشتریان بانک یکی از اهدافی است که بسیاری از محققین در راستای پاسخ به آن اقدام نموده اند (GHOSAL et al, 2020). این خوشه بندی باید توسط الگوریتم های رایانه ای انجام گردد تا خود سیستم، در مورد تعداد دسته ها، اعضای هر دسته و معیار شباهتی که استفاده می شود، تصمیم گیری نماید تا الگوریتم کاملا پویا گردد. به عبارتی دقیق تر، ممکن است مشتریان یک بانک با استفاده از N خوشه و معیار شباهت A خوشه بندی گردند و مشتریان بانک دیگری با استفاده از M خوشه و معیار شباهت B خوشه بندی شوند. لذا مدنظر است که پویایی موجود با استفاده از روش بهبود یافته LRFM[1] بر مبنای الگوریتم ژنتیک انجام پذیرد. مدل LRFM به عنوان یک روش قدرتمند در خوشه بندی مشتریان مطرح می باشد که تلاش های متعددی مبتنی بر این مدل انجام پذیرفته است (مصلحی و همکاران، ۱۳۹۳). این مدل برای خوشه بندی مشتریان بسیار کاربردی است و تحقیقات علمی و عملی زیادی نیز بر مبنای آن انجام پذیرفته است. با این حال یکی از مهم ترین مشکلاتی که هنوز در هر دو مدل RFM و LRFM مورد توجه است، امکان بکارگیری تعداد محدودی از متغیرهای برگزیده می باشد (خدابنده لو و همکاران، ۱۳۹۴). در حالی که بسیاری از ویژگی های مالی، روزمره و خانوادگی متاثر از پاسخ مشتری است. مدل LRFM شامل شاخص های طول (مدت زمان)

1. LRFM = Length, Recency, Frequency, Monetary

ارتباط با مشتری[1]، تازگی مبادله[2]، تعداد دفعات مبادله[3] و ارزش پولی مبادله شده[4] می باشد. این ویژگی ها در کنار سایر ویژگی های مشتری قرار خواهد گرفت و با استفاده از پویایی مبتنی بر الگوریتم ژنتیک[5] به ایجاد خوشه های مختلف می توان دست یافت (خدابنده لو و همکاران، ۱۳۹۴؛ Palaniappan et al, 2017). طبیعتا، با انتخاب جامعه هدف مناسب و متعاقب آن داشتن دیدی کلی از مشتریان امکان ارایه خدمت رسانی مطلوب جهت ایشان فراهم می گردد.

زمانی اهمیت این موضوع بیش تر مشخص می شود که بدانیم هزینه جذب یک مشتری جدید تقریبا پنج برابر هزینه نگهداری مشتریان قدیمی است. لذا، طبیعی است از یک سو در دنیای پر از رقابت امروز با رشد تکنولوژی های فناوری اطلاعات، احتمال خروج مشتریان قدیمی افزایش یافته و از سویی دیگر، نیازهای مشتریان نیز با افزایش سطح دانش آنها بیش تر و متنوع تر گردیده که به تبع آن، جست و جوی آن ها برای یافتن راه حل مناسب جهت رفع نیازهای پیش رو نیز بیش تر خواهد شد. این امر سبب می شود مشتریان بیش از بیش در معرض جذب توسط سایر رقیبا قرار گیرند (عبدالمنافی و همکاران، ۱۳۹۴؛ تقوی فرد و خواجوند، ۱۳۹۲).

در همین راستا هر چه بتوان خدمات را به جامعه هدف دقیق تری ارایه نمود، متعاقبا در هزینه و زمان نیز صرفه جویی می گردد. بر مبنای این هدف؛ سوالات مهم نیز در این بخش مطرح می گردند.

- چه متدولوژی برای پیگیری روند ارایه مدل پیشنهادی استفاده می گردد؟
- چه ویژگی هایی می تواند در کنار مدل LRFM قرار گیرد تا نتایج مناسبی را ارایه نمایند؟

1. Length
2. Recency
3. Frequency
4. Monetary
5. Genetic algorithm

- ساختار کروموزوم های[1] الگوریتم ژنتیک چگونه خواهد بود؟
- چه معیار شباهت یا روش خوشه بندی مشتریان مناسب خواهد بود؟
- چگونه می توان مدل LRFM را به وسیله الگوریتم ژنتیک بهبود داد؟
- روش های انتخاب، پیوند و جهش در الگوریتم ژنتیک پیاده سازی شده چگونه خواهند بود؟

در مدل پیشنهادی این مقاله تلاش خواهد شد روشی پویا به جهت بکارگیری متغیرهای با روش LRFM ارایه گردد تا امکان اجرای خوشه بندی های مختلفی بسته به زمان استفاده مهیا شود. این موضوع به انطباق بیش تر روش خوشه بندی پیشنهادی با واقعیت خواهد انجامید. لذا برای بالابردن کارایی خوشه بندی، می باید مشتریان را بر اساس ویژگی های مناسب گروه بندی کرد. لیکن، انتخاب ویژگی های کارا از بین تمامی حالات ممکن بسیار زمان بر است.

پیشینه تحقیق

در این بخش به تلاش های مرتبط سایر محققان که در راستای مساله مطروحه در این مقاله بوده پرداخته شده است تا نقاط ضعف و قوت هر کدام را به اختصار بررسی نماییم. مدل RFM یک مدل محبوب و معتبر در حوزه ارتباط با مشتری است که در زمان نیاز به مدیریت ارتباط با مشتری کارایی لازم خود را نشان می دهد. این مدل برای اولین بار توسط هاگز[2] در سال ۱۹۹۴ ارایه گردید (Hughes, 1994). این مدل بر پایه سه شاخص ذیل استوار است که در حالت کلی به تحلیل رفتار و ارزش گذاری بر رفتار مشتریان متمرکز می باشد.

- **تازگی مبادله (R)**: این شاخص به فاصله زمانی طی شده بین آخرین فعالیت انجام شده از سوی مشتری در سازمان تا حال حاضر اشاره دارد.
- **تعداد دفعات مبادله (F)**: نشان دهنده تعداد مبادلات (تراکنش های) مشتری

1. Chromosome
2. Hughes

است که در یک دوره زمانی مشخص انجام گرفته است.

- **ارزش پولی مبادله شده (M)**: منظور از ارزش پولی مبادله شده، ارزشی است که برای مبادلات در یک دوره زمانی مورد نظر صرف شده است.

نکته ای که در ذهن متواتر می شود آن است که خصوصیات و ویژگی هایی متنوعی از یک مشتری در پایگاه داده های سازمان های مختلف وجود دارد؛ لذا این امکان که می توان با جمع آوری خصوصیات یا سایر ویژگی های یک مشتری در کنار این سه ویژگی نسبت به ارایه پاسخ های مناسب تری اقدام نمود دور از ذهن نمی باشد البته، پژوهش های سایر محققین نیز در همین راستا بوده و نتایج مشابه ای را نیز ارایه نموده؛ به این معنی که با انجام تغییرات در روش RFM و در نظر گرفتن چند ویژگی بیشتر از مشتریان توانسته اند به نتایج مطلوبتری دست یابند (Zhang et al, 2015). این موضوع از آن جا ناشی می شود که اساسی فرضیه در مدل RFM آن است که الگوهای آینده خریدها و مبادلات یک مشتری نیز همانند الگوهای حال و گذشته او است. با توجه به این فرض است که وزن تمامی این ویژگی ها یکسان در نظر گرفته شده و به همین سه ویژگی از مشتریان نیز بسنده شده است.

با توجه به نقایص یاد شده، تلاش های فراوانی به منظور بهبود مدل RFM صورت پذیرفته است که یکی از مهم ترین آن ها مدل LRFM است (فتاحی و ربیعی، ۱۳۹۹). در مدل LRFM، علاوه بر سه متغیر مذکور (تازگی مبادله، تعداد دفعات مبادله و ارزش پولی مبادله شده) متغیر دیگری نیز اضافه شده است که طول همکاری مشتری با سازمان را لحاظ می کند. هدف از این تغییر در مدل RFM، برطرف کردن ضعف این مدل در لحاظ کردن ارتباط مشتریان وفادار بوده است. منظور از مشتریان وفادار مشتریانی هستند که دارای ارتباط بلندمدت با سازمان ها هستند. مدل LRFM توانایی تفکیک ارتباط مشتریان (بلندمدت و کوتاه مدت) از هم را دارد. مدت زمان ارتباط یک مشتری با سازمان می تواند تاثیر مستقیمی بر سودآوری و تاثیرگذاری یک مشتری بر سازمان داشته باشد. هر چه این ارتباط طولانی تر باشد، مبین این امر است که وفاداری مشتری به سازمان بیش تر می باشد

و هر چه بتوان ارتباط مشتریان با سازمان را طولانی تر کرد می توان وفاداری مشتری با سازمان را بهبود بخشید (Patil & Dharwadkar, 2017). سهرابی و خانلری تلاشی قابل توجه ای در راستای استفاده از مدل RFM برای سنجش ارزش دوره ارتباط مشتریان دارند (سهرابی و خانلری، ۱۳۸۶). در این مقاله، به منظور تعیین ارزش دوره ارتباط و چگونگی خوشه بندی مشتریان، شعبه ای از بانک های خصوصی تازه تاسیس در ایران رابا لحاظ نمودن معیارهای تاخر، فراوانی و ارزش مالی با بکارگیری روش خوشه بندی K-میانگین، مد نظر قرارداده اند، لذا به منظور صحت سنجی و تایید نتایج حاصله نسبت به استفاده از روش تحلیل ممیزی و در نهایت خوشه بندی مشتریان اقدام نموده اند که از عمده اشکالات قابل ذکر عدم پویایی روش مذکور می باشد. قنبری و رزمی در تلاشی دیگر، مدلی برای محاسبه ارزش دوره عمر مشتری ارایه کرده اند (قنبری و رزمی، ۱۳۸۸). مدلی که در آن، ضمن توجه به نقاط قوت روش‌های متداول به محاسبه ارزش عمر مشتری از طریق ترکیب دو روش RFM و ROI پرداخته شده است. منظور از ROI، احتمال عقد قرارداد مجدد مشتری در دوره‌های آینده و وفادار ماندن او به سازمان می باشد. این احتمالات پس از ترکیب شدن با میزان سود دهی مشتری برای سازمان، معیاری جهت اولویت‌بندی مشتریان خواهد بود، براین اساس اطلاعات یکی از شرکت‌های صنعت بیمه، نحوه استفاده از مدل و کارایی آن ارزیابی شده و نتایج آن نیز ارایه گردیده است.

پالانیاپان و همکاران[1] (۲۰۱۷) نیز توانسته اند اطلاعات مشتریان را بین سال های ۲۰۰۸ تا ۲۰۱۳، از یکی از بانک ها ی کشور پرتغال استخراج کرده و مدلی را برای دسته بندی مشتریان ارایه نمایند (Palaniappan et al, 2017). در این مدل تلاش شده از طریق دسته بندی مشتریان، نسبت به شناسایی مشتریانی که احتمال وفاداری بالاتری را نسبت به سایرین دارند اقدام شود. لذا این امر که از طریق دسته بند، بیز ساده[2]، جنگل تصادفی[3] ودرخت تصمیم[4] پیاده سازی گردیده امکان انجام مقایسه مناسبی را بین روش ها یاد شده، فراهم

1. Palaniappan
2. Naïve bayes
3. Random forest
4. Decision tree

نموده است. پاتیل و دارواد‌کار (۲۰۱۷) نیز روشی را برای دسته بندی مشتریان دو بانک از کشور هند ارایه نموده اند که با استفاده از اطلاعات در دسترس و بر مبنای شبکه های عصبی ارایه شده است (Patil & Dharwadkar, 2017). مطالعات آن ها نشان می دهد مدل پیشنهادی ایشان که بر مبنای دو پایگاه داده متفاوت ارایه شده از دقت های مطلوبی برخوردار است. لیکن، به دلیل عدم پویایی آن نمی توان تضمین نمودکه در سایر بانک ها با تنوع داده های و شرایط متفاوت نتیجه یکسانی حاصل گردد. البته، منظور آن نیست که مدل پیشنهادی این مقاله در تمامی شرایط و با هر داده بانکی صحیح عمل خواهد کرد. زیرا، به هر حال هر روش معایب خاص خود را دارد. ولی در مدل پیشنهادی این مقاله تلاش خواهد شد تا حد امکان، زمینه پویایی آن فراهم گردد. یانگ[1]و ژنگ[2] (۲۰۱۵) نیز روشی را برای دسته بندی مشتریان بانکی در کشور تایوان ارایه نموده اند. ایشان مشتریان بانک را به پنج دسته تقسیم و برای سنجش کارایی روش خود از نمودار ROC[3] استفاده کرده اند تا مقدار AUC[4]را به دست آورند (Yang & Zhang, 2015). در روشی دیگر، نیلوی[5] و نوید[6] نیز روشی را دسته بندی مشتریان از طریق داده های بانکی در تایوان ارایه کرده اند. در این روش برای دسته بند از درخت تصمیم و بیز ساده استفاده شده و در ادامه بین این دو دسته بندی مقایسه ای صورت گرفته است. آزمایش آنها نشان می دهد روش بکارگیری بیز ساده دقت مناسبتری دارد (Niloy & Navid, 2018). لیکن، مشکلات روش های دسته بندی و عدم پویایی روش های یاد شده همچنان پابرجا است.

در تلاشی دیگر، زوارم و مزیدی تلاش کرده اند تا روش LRFM را با استنتاج فازی ترکیب کرده و مشکلات این روش را تا حد امکان کاهش دهند (Zoeram & Karimi Mazidi, 2018). در این روش سعی شده است تا بر مبنای چهار متغیر مدل LRFM، میزان

1. Yang
2. Zhang
3. Receiver operating characteristic
4. Area under curve
5. Niloy
6. Navid

وفاداری یک مشتری با استفاده از استناج فازی تخمین زده شود. از نقاط قوت این روش آن است که با استفاده از سیستم استنتاج فازی توانسته است وزن تاثیر گذار به هر یک از چهار متغیر مدل LRFM را برای خوشه بندی تغییر دهند. در این مقاله تاکید شده است که در صورت ایجاد پویایی برای سایر ویژگی های مشتری و درگیرکردن آنها در خوشه بندی احتمال دست یابی به نتایج مطلوبتر وجود دارد که این امر سنگ بنای روش پیشنهادی این مقاله نیز قرار گرفته است. ماریسا و همکاران (۲۰۱۹) تلاش کرده اند روشی ارایه نمایند تا با استفاده از وزن دهی به هر کدام از المان های مدل LRFM پویایی اندکی را در روش ایجاد کنند (Marisa et al, 2019). در روش مذکور، پس از وزن دهی به هر کدام از المان های مدل LRFM که مبتنی بر حجم داده ها انجام می گیرد، از روش K-میانگین برای خوشه بندی استفاده شده است. در این مقاله نیز تغییر وزن المان های مدل LRFM و مبتنی بودن آن بر داده های موجود و حجم داده ها، توانسته است تاثیرگذاری مدل LRFM را در خوشه بندی K-میانگین افزایش دهد. بنابراین، می توان قدرت مدل LRFM را بادرگیرکردن خصوصیات داده ها افزایش داد. در روش پیشنهادی این مقاله نیز تلاش خواهد شد روشی پویا به منظور تاثیر گذاری از طریق افزایش وتعدد متغیرهای در روش LRFMارایه گردد تا امکان ارایه خوشه بندی های مختلف متناسب با زمان فراهم گردد. این موضوع به انطباق بیش تر روش خوشه بندی پیشنهادی با واقعیت می انجامد. در این راه از الگوریتم ژنتیک بهره برده می شود. زیرا، راه حل دقیق و آشکاری برای این موضوع به ذهن نمی رسد.

روش پیشنهادی

متدولوژی[1] های مختلفی برای انجام تجزیه و تحلیل در خصوص عملکرد مشتریان بانک و به طور کلی تحقیقات داده کاوی[2] بر روی داده های مشتریان قابل انجام می باشد (Wei et al, 2010). متدولوژیهای CRISP-DM، SEMMA و KDD از مهم ترین متدولوژی های

1. Methodology
2. Data Mining

کاربردی در این حوزه می باشند که هر یک مزایا و معیاب خاص خود را دارند. لیکن، با توجه به تحقیقات سایر محققان و نتایج به دست آمده از آن ها در روش های مشابه، برای تحلیل مشتریان و خوشه بندی در کارهای داده کاوی، عموما" از متدولوژی CRISP-DM که محبوب تر نیز می باشد استفاده می گردد (سعدالهی اسرمی و سلاجقه، ۱۳۹۴). این موضوع از نتایج تحقیقات انجام پذیرفته توسط انجمن تحقیقات داده کاوی نیز قابل استنباط است. این انجمن نیز ذکر می نماید که متدولوژی CRISP-DM در مقایسه با سایر متدولوژی ها پرکاربردتر بوده است (Azevedo & Santos, 2008). بنابراین، در صورت استفاده از متدولوژی CRISP-DM می توان انتظار داشت روش پیشنهادی از سایر روش ها کامل تر باشد.

با توجه به مطالب عنوان شده، می توان از از روش CRISP-DM بعنوان یکی از قدرتمندترین و محبوب ترین متدولوژی ها نام برد که پس از عبور از شش مرحله روش CRISP-DM، نتیجه ای علمی برپایه داده های مساله حاصل گردد (تقوی فرد و خواجوند، ۱۳۹۲). چارچوب پیشنهادی تحقیق حاضر نیز بر اساس این متدولوژی ارایه شده که پس از هر یک از این شش مرحله در روش پیشنهادی این مقاله مراتب مورد شرح و بسط قرار خواهد گرفت مراحل این متدولوژی عبارتند از:

مرحله اول: شناخت بانکداری خرد

مرحله ابتدایی به عنوان مهم ترین مرحله این فرایند بین محققین مختلف شناخته می شود. در گام نخست، هدف از طرح پروژه و نیازهای آن به همراه موانع موجود در کار مورد ارزیابی قرار می گیرند. همچنین فاکتورهای تاثیرگذار در حل مساله مورد ارزیابی و جست وجو قرار می گیرند. سازمانی که در این مقاله مورد بررسی قرار می گیرد بانک صادرات ایران است. مشتریانی که داده های آن ها استخراج گردیده مربوط به مشتریان حقیقی شعب استان تهران و حوزه انتخابی این تحقیق بانکداری خرد است.

مرحله دوم: شناخت داده های تراکنشی، استفاده از خدمات، استفاده از تسهیلات و انجام امور مالی مربوطه

در مرحله شناخت داده ها به جمع آوری داده و ارزیابی کیفیت داده های جمع آوری شده پرداخته می شود. مشخص است که در مسایل این حوزه، زیرمجموعه ای از کل داده های موجود به عنوان نمونه ای از کل داده ها در نظر گرفته خواهد شد. پس از بررسی ها و پیگیری های انجام شده و کسب مجوزهای لازم، یک پایگاه داده حاوی اطلاعات ۱۰ درصد از مشتریان بانک صادرات ایران (ساکن شهر تهران) در اختیار پژوهشگران این مقاله قرار گرفت. اطلاعات این ۱۰ درصد مشتری به صورت تصادفی و بدون در نظر گرفتن هیچ فاکتوری از بین مشتریان شهر تهران انتخاب شده اند. تعداد کل مشتریان ساکن شهر تهرانحدود ۹ میلیون نفر (۸. ۹۲۵. ۳۴۴ نفر) است که ۱۰ درصد آن برابر با ۹۰۰ هزار نفر می باشند که جهت بررسی رقم بسیار بالایی را شامل می گردد. تاریخ استخراج اطلاعات این مشتریان مورخه۱۳۹۹/۰۳/۰۱ می باشد. لذا تاریخ انجام تراکنش ها، کارکردها و فعالیت های مشتری مربوط به بازه زمانی مورخه۱۳۹۹/۰۱/۰۱ الی۱۳۹۹/۰۲/۳۱ می باشد. لازم به یادآوری است که هیچ کدام از خصوصیات مشتری که محرمانه بوده و مربوط به اطلاعات هویتی باشد در اختیار پژوهشگران این مقاله نبوده، بنابراین، مشکلی در قبال حریم شخصی و محرمانگی داده ها وجود نخواهد داشت. ویژگی های استفاده شده در جدول (۱) بصورت مشروح عنوان شده است.

مرحله سوم: آماده سازی داده های تراکنشی،

استفاده از خدمات، استفاده از تسهیلات و انجام امور مالی مربوطه در این مرحله تلاش می شود داده های استخراج شده با شرایط منطبق گردند، بگونه ای که داده ها توسط الگوریتم قابل استفاده باشند. لذا می بایست داده ها با ویژگی های مناسب انتخاب[1]، تمیز شده[2]و در

1. Selection
2. Cleaning

صورت لزوم از پایگاه های مختلف تلفیق[1] گردند. این موضوع نشان می دهد ارایه خدمات هدفمند به مشتریان بانک از اهمیت ویژه ای برخوردار است. هر چه بتوان جامعه هدف را دقیق تر شناسایی نمود، متعاقبا" در ارایه هدفمند خدمات، هزینه و زمان صرفه جویی خواهد شد. خوشه بندی مشتریان بانک یکی از اهدافی است که بسیاری از محققین در راستای پاسخ به آن اقدام کرده اند (Ghosal et al, 2020).

مرحله چهارم: مدلسازی و ارایهLRFM-GA

در این بخش به چگونگی ترکیب مدل LRFM و الگوریتم ژنتیک پرداخته خواهد شد که انتظار می رود نتیجه حاصله منجر به بهبود نتایج مدل LRFM گردد. لذا از این پس، روش پیشنهادی این مقاله با نام LRFM-GA خطاب می گردد. در ادامه تمامی گام های الگوریتم ژنتیک که بر پایه مساله پیش رو است می بایست طی شده و تمامی پارامترها و خصوصیات روش پیشنهادی تنظیم و ذکرگردد. با توجه به آن که تعداد مشتریان حاضر در نمونه گیری ۹۰۰ هزار نفر می باشند؛ باید تدبیری اتخاذ شود تا از لحاظ پیاده سازی مشکلی بروز ننماید. روند کلی مدل LRFM-GA پیشنهادی در چهار گام زیر خلاصه خواهد شد.

جدول ۱. بازه تاریخی اخذ و محاسبه ویژگی های مشتریان.

دسته بندی	ویژگی	بازه تاریخی اخذ و محاسبه
داده های تراکنشی	شماره مشتری	در تاریخ ۱۳۹۹/۲/۳۱
	طول ارتباط مشتری با بانک صادرات	تا تاریخ ۱۳۹۹/۲/۳۱
	تازگی مبادله	۱۳۹۹/۱/۱ تا ۱۳۹۹/۲/۳۱
	تعداد دفعات مبادله	۱۳۹۹/۱/۱ تا ۱۳۹۹/۲/۳۱
	ارزش پولی مبادله شده	۱۳۹۹/۱/۱ تا ۱۳۹۹/۲/۳۱
	کد نوع عملیات	۱۳۹۹/۱/۱ تا ۱۳۹۹/۲/۳۱
	کد پایانه	۱۳۹۹/۱/۱ تا ۱۳۹۹/۲/۳۱
	بانک میزبان	۱۳۹۹/۱/۱ تا ۱۳۹۹/۲/۳۱

1. Integration

دسته بندی	ویژگی	بازه تاریخی اخذ و محاسبه
	تاریخ انجام تراکنش	۱۳۹۹/۱/۱ تا ۱۳۹۹/۲/۳۱
داده های استفاده از خدمات	ارزش انتقال وجه شتابی	۱۳۹۹/۱/۱ تا ۱۳۹۹/۲/۳۱
	مبلغ کل خرید کارت هدیه از بانک صادرات	۱۳۹۹/۱/۱ تا ۱۳۹۹/۲/۳۱
	مبلغ کل خرید کالا از کارتخوان بانک صادرات	۱۳۹۹/۱/۱ تا ۱۳۹۹/۲/۳۱
	دفعات استفاده مشتریان از همراه بانک صادرات	۱۳۹۹/۱/۱ تا ۱۳۹۹/۲/۳۱
	دفعات استفاده مشتریان از اینترنت بانک صادرات	۱۳۹۹/۱/۱ تا ۱۳۹۹/۲/۳۱
	دفعات پرداخت قبض به وسیله بانک صادرات	۱۳۹۹/۱/۱ تا ۱۳۹۹/۲/۳۱
	مبلغ کل خرید شارژ تلفن همراه	۱۳۹۹/۱/۱ تا ۱۳۹۹/۲/۳۱
داده های استفاده از تسهیلات	مبلغ تسهیلات دریافتی از بانک صادرات	تا تاریخ ۱۳۹۹/۲/۳۱
	میزان تسهیلات معوق از بانک صادرات	تا تاریخ ۱۳۹۹/۲/۳۱
داده های مالی	میانگین تراکنش ماهیانه	تا تاریخ ۱۳۹۹/۲/۳۱
	مبلغ کل سپرده مشتری در بانک صادرات	تا تاریخ ۱۳۹۹/۲/۳۱
	مبلغ کل چک های وصول نشده از بانک صادرات	۱۳۹۹/۱/۱ تا ۱۳۹۹/۲/۳۱

- **ایجاد جمعیت اولیه:**

جمعیت اولیه به صورت باینری ایجاد می گردند. هر کدام از ژن های داخل کروموزوم ها مربوط به یکی از ویژگی هایی است که می توان در مورد آن تصمیم گیری نمود که آیا در کنار ویژگی های مدل LRFM قرار بگیرد یا نه. بنابراین، کروموزوم الگوریتم ژنتیک دارای ۱۷ ژن باینری می باشند. زمانی که ژن باینری یک در نظر گرفته شود به معنای در نظر گرفتن ویژگی مربوطه در کنار ویژگی های مدل LRFM است، در صورتی که صفر باشد به معنای عدم وجود و کاربرد ویژگی مربوطه خواهد بود. لازم به ذکر است جمعیت اولیه به تعداد ۱۰۰۰ کروموزوم می باشد. بنابراین، ابعاد مساله پیش رو، دارای ۱۷ ویژگی خواهد بود که بسیار حجیم است و تعداد حالت هایی که می توان به عنوان پاسخ در نظر گرفت بسیار زیاد خواهد شد.

- **سنجش برازش هر کروموزوم:**

با استفاده از هر کروموزوم حاضر در جمعیت، عمل خوشه بندی با استفاده از روش DBSCAN++ انجام می گیرد. میزان مناسب بودن خوشه بندی انجام شده بر مبنای معیار دیویس بولدین سنجیده می شود که متناظر با برازش هر کروموزوم است.

- **تولید نسل جدید:**

با استفاده از چرخه رولت به انتخاب کروموزوم های والد پرداخته و دو کروموزوم جهت تولید دو فرزند برای جایگزینی در نسل بعد انتخاب می شوند.

- **خاتمه الگوریتم:**

تمامی کروموزوم های جمعیت جدید مورد ارزیابی قرار می گیرند. شرط خاتمه الگوریتم، تعداد نسل مشخص است. پاسخی که بیش ترین برازندگی را در جمعیت پایانی دارد به عنوان پاسخ نهایی انتخاب می گردند. به عبارت دقیق تر، ویژگی هایی که این کروموزم پیشنهاد می دهند در کنار ویژگی های مدل LRFM قرار می گیرد. بر اساس تحقیقات انجام شده، روش DBSCAN با میزان ابعاد مساله (۱۷ ویژگی) مشکلی در اجرا ندارد. تنها موردی که باقی می ماند آن است که با وجود ۹۰۰ هزار مشتری و ۱۷ ویژگی در مساله، ماتریس مربوط به داده های مساله ۱۷ x ۹۰۰۰۰۰ می باشد. هر کدام از کروموزوم های موجود در جمعیت نشان از انتخاب تعدادی ویژگی از بین ۱۷ ویژگی را دارند که در کنار چهار ویژگی مدل LRFM قرار گرفته و عمل خوشه بندی حاصل می شود. انجام خوشه بندی با استفاده از روش DBSCAN برای این تعداد داده و برای هر کدام از کروموزوم های موجود در جمعیت کاری زمان بر خواهد بود. بنابراین، باید متدی برای پیاده سازی روش DBSCAN اتخاذ گردد تا بتوان مدل LRFM را با استفاده از الگوریتم ژنتیک بهبود بخشید تا روش LRFM-GAحاصل گردد. بنابراین، با توجه به مزایای الگوریتم DBSCAN برای روش پیشنهادی این مقاله، می توان انتظار داشت نتایج حاصل از اجرای آن برای داده های مساله پیش رو مناسب باشند. لیکن، استفاده از روش DBSCAN برای

پایگاه داده هایی با اندازه بزرگ از نظر زمان اجرا مناسب نیست. دلیل این موضوع نیز به پیچیدگی مربوط به جست و جوی همسایگی در پایگاه داده های بزرگ مربوط می شود. این موضوع با ترکیب روش DBSCAN و الگوریتم ژنتیک وخیم تر می گردد. زیرا، برای هر کروموزوم باید مناسب بودن خوشه بندی پیشنهادی توسط آن کروموزوم مورد بررسی قرار گیرد. به همین سبب، دو اقدام صورت پذیرفته است:

۱) اجرای دو قانون اولیه قبل از بررسی خوشه بندی پیشنهادی توسط کروموزوم:

این قوانین قبل از بررسی خوشه بندی پیشنهادی توسط کروموزوم برروی مقادیر پیشنهادی کروموزوم مورد ارزیابی قرار می گیرد. این قوانین توسط متخصصین امر مشخص گردیده و کمک می کند زمان اجرا برای ارزیابی یک کرومزوم که به احتمال بالا خوشه بندی مناسبی را پیشنهاد نمی دهد تلف نگردد. دو قانون به شرح زیر است:

با توجه به جدول (۱) چهار گروه ویژگی در نظر گرفته شده است. در صورتی که کروموزوم انتخابی از دو یا سه گروه از این چهار گروه ویژگی نداشته باشند مورد ارزیابی قرار نمی گیرد.

کروموزوم حتما می بایست شامل ویژگی های بازه ای و غیربازه ای باشند. کروموزوم باید از اطلاعات غیربازه ای که رفتار مشتری را در طولانی مدت مدل می کند در کنار ویژگی های غیربازه ای که رفتار حال مشتری را مدل می کند استفاده کند.

۲) استفاده از روش DBSCAN++ برای افزایش سرعت و کاهش زمان در کار با پایگاه داده های بزرگ:

این روش در تلاش های مشابه نیز دیده می شود. مهم ترین نکته در DBSCAN++ آن است که توانسته با استفاده از نمونه برداری n داده از N داده به نتیجه مطلوب دست یابد. در روش DBSCAN میزان چگالی برای تمامی داده ها محاسبه می گردد که این محاسبات زمان زیادی را به خود اختصاص می دهد و برای پایگاه داده های حجیم مناسب نخواهد

بود. روش DBSCAN++ بر اساس زیرمجموعه ای انتخاب شده از داده ها کار می نماید. لازم به ذکر است تمامی روند اجرایی زیر از نظر کارایی۱، پیچیدگی زمانی۲ و حافظه ۳ مورد اثبات بوده و توسط محققین مختلف مورد استفاده قرار گرفته است. روش DBSCAN++ می تواند برای M داده که زیر مجموعه ای از N داده می باشد عمل کرده و شرط ۴M<<N نیز می تواند برقرار است.

در این گام، تابعی برای سنجش کیفیت پاسخ ها و راه حل های پیشنهادی در جمعیت بیان شده است که بررسی می نماید آیا با سنجش یک کروموزوم DBSCAN، قادر به ارایه پیشنهاد خواهیم بود و یا متعاقب آن می باید عمل خوشه بندی انجام گیرد.

سه مزیت اصلی روش LRFM متغیرهای مناسبی را برای ترکیب با روشی که برای حل مساله پیش روی ما دراین مقاله مناسب می باشند به شرح ذیل بیان می نماید (ویجایالاسمی۵و پونیزاوالی۶، ۲۰۱۲):

برخلاف روش هایی همانند K-میانگین نیازی به وارد کردن تعداد خوشه ها نیست.

خوشه با هر شکل یا خصوصیتی قابل یافتن است.

با وجود داده های نامناسب (پرت) از قدرت و قابلیت روش کاسته نمی شود. به همین سبب، مشتریانی که خصوصیاتی بسیار متفاوت دارند نیز از قدرت روش کم نمی کنند.

لذا پس از انجام خوشه بندی داده ها بر اساس ویژگی های مدل LRFM و ویژگی های پیشنهاد شده توسط یک کروموزوم از جمعیت به وسیله روش DBSCAN، بررسی صورت پذیرفته مبین آن است که خوشه بندی چگونه انجام شده است. هر چه خوشه بندی بر اساس ویژگی های مدل LRFM و ویژگی های پیشنهاد شده توسط یک کروموزوم از جمعیت مطلوبتری برخوردار باشد نتیجه می شود که کروموزوم مربوطه ویژگی های

---

1. Performance
2. Time Complexity
3. Memory

۴. مقدار M می تواند خیلی کوچک تر از N باشد.

5. Vijayalaksmi
6. Punithavalli

مناسب تری را برای ترکیب با ویژگی های مدل LRFM پیشنهاد نمود. نکته قابل تامل آن است که چگونه نتیجه خوشه بندی به دست آمده ارزیابی شود. در همین خصوص جهت ارزیابی نتیجه خوشه بندی از معیار دیویس بولدین[1]بهره خواهیم برد که بر اساس پراکندگی یک خوشه (siو sj) و عدم شباهت بین دو خوشه i و j (di,j) تعریف می‌شود (Vergani et al, 2018). شباهت بین دو خوشه را می‌توان به صورتهای مختلفی تعریف کرد ولی می بایست شرایط ذیل را دارا باشند (دهقان و همکاران، ۱۳۹۱):

Ri,j۰ ≤

Ri,j=Rj,i

اگر si و sj هر دو برابر صفر باشند آنگاه Ri,j نیز برابر صفر باشد.

اگر sj>skو di,j=di,k آن گاه Ri,k<Ri,j

اگر sj=skو di,j<di,k آن گاه Ri,k<Ri,j

معمولا شباهت بین دو خوشه به صورت:

$$R_{i,j} = \frac{s_i + s_j}{d_{i,j}} \qquad (۱)$$

تعریف می‌شود و در آن di,jو si به ترتیب، با روابط (۲) و (۳) محاسبه می‌شوند. v_i, v_j میانگین خوشه های i و j هستند.

$$d_{i,j} = d\,(v_i, v_j) \qquad (۲)$$

$$s_i = \frac{1}{||c_i||}\sum_{x \in c_i} d\,(x, v_i) \qquad (۳)$$

با توجه به مطالب عنوان شده و تعریف شباهت بین دو خوشه شاخص دیویس بولدین به صورت ذیل قابل بررسی می باشد:

1. DaviesBouldin Index

$$DB = \frac{1}{n_c}\sum_{i=1}^{n_c} R_i \qquad (۴)$$

و Ri به صورت:

$$R_i = \max_{i,j=1,\ldots,n_c \text{ and } i\neq j} (R_{i_و j}) \qquad (۵)$$

محاسبه می‌شود. برای هر کدام از خوشه های به دست آمده، این معیار محاسبه می گردد. به عبارت دقیق تر، اگر سه خوشه در نظر گرفته شده باشد:

یک بار خوشه های ۲ و۳ را یکی در نظر گرفته و معیار دیویس بولدین برای خوشه ۱ و در برابر آن ها محاسبه می گردد.

دومین بار خوشه های ۱ و۳ را یکی در نظر گرفته و معیار دیویس بولدین برای خوشه ۲ و در برابر آن ها محاسبه می نماییم.

سومین بار خوشه های ۲ و ۱ را یکی در نظر گرفته و معیار دیویس بولدین برای خوشه ۳ و در برابر آن ها محاسبه خواهیم نمود.

سپس، سه مقدار به دست آمده را جمع نموده و به عنوان میزان برازش کروموزوم مربوطه در نظر می گیریم.

این شاخص در واقع میانگین شباهت بین هر خوشه با شبیه‌ترین خوشه را محاسبه می‌کند. می‌توان دریافت که هرچه مقدار این شاخص بیشتر باشد، خوشه‌های مناسب تری تولید شده است. در کنار معیار دیویس-بولدین از معیار سیلهوتی۱ نیز استفاده خواهد شد. این روش سنجش بر اساس میانگین فاصله ها است، هر یک از نمونه های یک خوشه با تمام نمونه های موجود در همان خوشه و کم ترین فاصله آن با کل نمونه های موجود در خوشه های دیگر مشخص و تعریف می شود (Mirjalili & Thinsungnoena, 2019).

1. Silhouette

رابطه مربوط به این شاخص را می توان به شرح ذیل بیان نمود:

$$sil\ (i) = \frac{b(i) - a\ (i)}{\max\{b(i)\ and\ a(i)\}} \qquad (۶)$$

پارامتر a(i) نشان دهنده میانگین فاصله یک نمونه با نمونه های دیگر در همان خوشه و پارامتر b(i) به معنای کم ترین فاصله بین یک نمونه و همه نمونه های موجود در خوشه ها ی دیگر است. هر چه مقدار این شاخص به یک نزدیک تر باشد نشان از مطلوب بودن خوشه بندی و هر چه به منفی یک نزدیک تر باشد نشان از نامناسب بودن خوشه بندی است.

لذا با انتخاب کروموزوم های والد می توان کروموزوم های جدیدی را برای جمعیت جدید ایجاد کرد. یکی از محبوب ترین روش های موجود برای انتخاب کروموزوم ها، روش چرخ رولت[1] است. مقادیر به دست آمده از طریق معیار دیوس بولدین برای چرخ رولت و انتخاب کروموزوم های والد استفاده می گردد.

زمانی که دو والد برای تولید یک یا چند کروموزوم انتخاب شدند؛ نیاز به عملگرهایی برای تولید فرزند است. پیوند برای ایجاد فرزنداز ترکیب کروموزوم های دو والد از طریق عملگر چهار-نقطه ای صورت می پذیرد. لازم به ذکر است، مفهومی به نام احتمال جهش به پیاده سازی الگوریتم ژنتیک در روش پیشنهادی اضافه شده است. منظور از احتمال جهش که با Pm نشان داده می شود، احتمال تغییر یک ژن در یک کروموزوم است. به عبارت دقیق تر، برای هر کدام از ژن های موجود در کروموزوم یک عدد تصادفی تولید می گردد. اگر این عدد تصادفی از Pm کم تر بود جهش برای آن ژن انجام می شود؛ در غیر این صورت جهشی انجام نمی گیرد. این مقدار احتمال باعث ایجاد حالات مختلف و گوناگونی در جمعیت می شود تا فضای پاسخ به شکل مناسب تری قابل جست و جو گردد.

فرزندان حاصل از عملگرهای پیوند و جهش برای ایجاد نسل جدید استفاده می

1. Roulette Wheel

گردند. تمامی اعضای جمعیت جدید از فرزندان ایجاد می شوند تا تمامی جواب های بررسی شده در جمعیت قبلی از بین برود. زیرا، هدف ایجاد تنوع بیش تر در پاسخ ها به دلیل فضای گسترده پاسخ و تنوع جواب است.

مرحله پنجم: ارزیابی مدل LRFM-GA

با ارزیابی انجام شده در این مرحله، تعیین خواهد شد که کدام یک از اهداف تعیین شده واصل شده اند، به این مفهوم که نتایج حاصله از روش ها و تکنیک های انتخاب شده در مرحله چهارم استخراج، مورد بررسی قرار می گیرند. اهدافی که وصول شده اند یا نشده اند سنجش گردیده و تغییرات احتمالی لازم در تکنیک انتخاب شده انجام می گیرد. دو نوع ارزیابی در این بخش انجام خواهد گرفت:

۱) **خوشه بندی مشتریان بدون پیش فرض:** در این ارزیابی، مقادیر ویژگی های مشتریان در اختیار روش قرار داده شده است تا طبق روند بیان شده در فصل سوم به خوشه بندی مشتریان در خوشه های مختلف بپردازد.

۲) **خوشه بندی با پیش فرض:** در این نوع از ارزیابی، هدف مشخص است. به عنوان مثال، هدف یافتن خوشه های مشتریانی است که از منظر عملکردی، در بازپرداخت وام خوش حساب می باشند. البته این موضوع بدان معنا نیست که تمامی مشتریان به دو خوشه بد حساب و خوش حساب تقسیم شوند. زمان بدون پیش فرض مربوط به ارزیابی هایی است که تلاش داریم مشتریان بانک را بر اساس شرایطی که در حال حاضر دارا می باشند به خوشه های متعددی تقسیم کنیم و سپس، تصمیم در خصوص آن ها اتخاذ گردد. زمان با پیش فرض مربوط به ارزیابی هایی است که هدف از خوشه بندی مشتریان ارایه خدماتی معین به آن ها است. به عبارت دقیق تر، شناسایی مشتریانی با ویژگی های خاص که جامعه هدف خدمت مورد ارایه می باشند.

در روش بدون پیش فرض، داده های جمع آوری شده در اختیار روش LRFM-GA

قرار می گیرد. بنابراین، بر مبنای ۱۷ ویژگی آمده در جدول (۱) که قرار است در کنار چهار ویژگی طول ارتباط مشتری با بانک صادرات، تازگی مبادله، تعداد دفعات مبادله و ارزش پولی مبادله شده قرار گیرند عمل خوشه بندی با استفاده از روش LRFGM-GA انجام می گیرد. پس از انجام خوشه بندی، میزان مناسب بودن خوشه ها به شرح ذیل سنجیده خواهند شد.

پس از انجام عمل خوشه بندی به وسیله روش DBSCAN++، خوشه هایی به دست می آیند که تعداد آن ها از طریق روش DBSCAN++ و با توجه به چگالی داده ها مشخص می گردند، همچنین میزان مناسب بودن خوشه بندی انجام شده با استفاده از معیار دیویس بولدین انجام خواهد پذیرفت (داده ها با استفاده از چهار ویژگی روش LRFM، به وسیله روش DBSCAN++ خوشه بندی می شوند و برای ارزیابی خوشه بندی از معیار دیویس بولدین استفاده خواهند شد). این مقادیر کارایی روش استاندارد LRFM را نشان خواهند داد. همچنین، روند LRFM-GA با استفاده از ویژگی های پیشنهادی الگوریتم ژنتیک که در کنار ویژگی های استاندارد LRFM قرار می گیرند برای خوشه بندی داده ها استفاده خواهند شد. متعاقب آن، مقادیر دیویس بولدین روش LRFM-GA نیز به دست خواهد آمد تا با مقادیر مربوط به استاندارد LRFM قابل مقایسه باشد. معیار دیویس بولدین یک معیار کلی برای سنجش میزان مناسب بودن خوشه بندی است که روابط (۱) تا (۵) به این معیار اختصاص دارد. مزیت این شاخص آن است که به روش خوشه بندی و تعداد خوشه ها وابسته نیست (PETROVIC, 2006).

جدول (۲) نشان از مقادیر دیویس بولدین برای هر دو روش LRFM و LRFM-GA دارد و تعداد خوشه هایی که به وسیله روش DBSCAN++ برای هر دو روش مشخص شده نیز در این جدول بیان گردیده است.

جدول ۲. مقایسه ای بین دو روش LRFM و LRFM-GA در خوشه بندی بدون پیش فرض.

روش	LRFM	LRFM-GA
مقدار دیویس بولدین	۴۱۵ .۱	۲۳۱ .۱
مقدار سیلهوتی	۷۰ .۰	۸۳ .۰
تعداد خوشه ها	۱۷	۱۵

مقایسه مقادیر شاخص دیویس بولدین برای هر دو روش در جدول شماره (۲) مبین این امر است که مقدار شاخص برای روش LRFM-GA کم تر از استاندارد LRFM می باشد. لذا می توان نتیجه گرفت کم تر بودن این شاخص در مدل LRFM-GA در مقایسه با مدل LRFM بیان گر کیفیت مطلوبتر خوشه ها در خوشه بندی با روش پیشنهادی این مقاله است.

به جهت صحه گذاشتن بر نتایج به دست آمده می توان در کنار مقایسه حاضر، نسبت به انجام مقایسه ای دیگری با استفاده از پیاده سازی مراتب از طریق روش RFM اقدام نمود. زیرا، نتایج به دست آمده از تلاش های سایر محققان حاکی از برتری روش LRFM در مقایسه با روش RFM است. در صورتی که نتایج حاصل از مقایسه روش RFM با روش های مذکور نیز این روند را تایید نماید می توان انتظار داشت، نتایج به دست آمده بر مبنای جدول شماره (۲) نیز از اعتماد قابل قبولی برخوردار باشد. پس، داده ها با استفاده از سه ویژگی روش RFM (تازگی مبادله، تعداد دفعات مبادله و ارزش پولی مبادله شده) و به وسیله روش DBSCAN++ خوشه بندی می شوند و برای ارزیابی خوشه بندی نیز از معیار دیویس بولدین استفاده خواهد شد. این مقادیر کارایی روش RFM را نشان خواهد داد. جدول شماره (۳) به صورت کاملتری به مقایسه بین هر سه روش RFM، LRFM و LRFM-GA می پردازد.

جدول ۳. مقایسه خوشه بندی بدون پیش فرض از طریق سه روش RFM، LRFM و LRFM-GA.

روش	RFM	LRFM	LRFM-GA
مقدار دیویس بولدین	۹۳۲ .۱	۴۱۵ .۱	۲۳۱ .۱
مقدار سیلهوتی	۶۴ .۰	۷۰ .۰	۸۳ .۰
تعداد خوشه ها	۱۴	۱۷	۲۶

نتایج حاصل از جدول شماره (۳) نیز حاکی از برتری روش LRFM نسبت به RFM دارد. لذا این موضوع نیز می تواند مبین اجرای روند صحیحی اقدامات انجام پذیرفته باشد. با

توجه به شاخص دیویس بولدین ترتیب کارایی روش ها را می توان به صورت RFM < LRFM <LRFM-GA ارایه نمود. ارتباط بین LRFM و RFM بر اساس تحقیقات سایر محققان نیز قابل استناد است. پس، نتایج به دست آمده از اعتماد قابل قبولی برخوردار خواهد بود (بابائیان و سرفرازی، ۲۰۱۹). نکته دیگری که در جدول شماره (۳) قابل تامل است تعداد خوشه هایی است که مشتریان به آن تعداد خوشه تقسیم شده اند. مشخص است که روش پیشنهادی LRFM-GA، مشتریان را به ۲۶ خوشه تقسیم کرده در حالی که روش های LRFM و RFM به ترتیب به ۱۷ و ۱۴ خوشه تقسیم نموده اند.

نتیجه حاصله آن است که هر چه ویژگی های بیش تری برای خوشه بندی لحاظ شود، خوشه بندی مشتریان با تعداد و جزئیات بیش تری انجام خواهد گرفت. به همین سبب، تعداد خوشه ها در روش LRFM-GA بیش تر از دو روش دیگر می باشد. لازم به ذکر است که تعداد بالاتر خوشه ها الزاما به معنای دقت بالاتر نیست. لیکن، با توجه به بهترشدن شاخص دیویس بولدین می توان انتظار داشت افزایش تعداد خوشه ها به معنای استفاده از ویژگی های بیش تر و رسیدن به خوشه بندی مناسبتر باشد. لیست ویژگی هایی که در نهایت به وسیله الگوریتم ژنتیک برای قرارگرفتن در کنار ویژگی های روش LRFM پیشنهاد گردیده در جدول شماره (۴) ارایه شده است. این ویژگی ها توسط بهترین کروموزوم در آخرین نسل الگوریتم ژنتیک پیشنهاد شده است.

جدول ٤. ویژگی های انتخاب شده توسط روش LRFM-GA در آزمایش خوشه بندی بدون پیش فرض.

دسته بندی	ویژگی	ویژگی های انتخاب شده توسط روش LRFM-GA
داده های تراکنشی	شماره مشتری	
	طول ارتباط مشتری با بانک صادرات	✓
	تازگی مبادله	✓
	تعداد دفعات مبادله	✓
	ارزش پولی مبادله شده	✓

دسته بندی	ویژگی	ویژگی های انتخاب شده توسط روش LRFM-GA
	کد نوع عملیات	
	کد پایانه	
	بانک میزبان	
	تاریخ انجام تراکنش	
داده های استفاده از خدمات	ارزش انتقال وجه شتابی	
	مبلغ کل خرید کارت هدیه از بانک صادرات	
	مبلغ کل خرید کالا از کارتخوان بانک صادرات	
	دفعات استفاده مشتریان از همراه بانک صادرات	✓
	دفعات استفاده مشتریان از اینترنت بانک صادرات	✓
	دفعات پرداخت قبض به وسیله بانک صادرات	✓
	مبلغ کل خرید شارژ تلفن همراه	
داده های استفاده از تسهیلات	مبلغ تسهیلات دریافتی از بانک صادرات	✓
	میزان تسهیلات معوق از بانک صادرات	✓
داده های مالی	میانگین تراکنش ماهیانه	✓
	مبلغ کل سپرده مشتری در بانک صادرات	
	مبلغ کل چک های وصول نشده از بانک صادرات	✓

با توجه به مشروح ویژگی های جدول شماره (۴) می توان دریافت که نتیجه پایانی روش LRFM-GA دارای یازده ویژگی است که البته با توجه به داده هایی که در اختیارش قرار داده شده انتخاب گردیده اند. با تغییر داده ها ممکن است روند تصمیم گیری روش

پیشنهادی در مورد ویژگی ها تغییر کند. این موضوع خود مبین پویایی روش می باشد. (هفت ویژگی در کنار ویژگی های LRFM قرار گرفته است) در خوشه بندی بدون پیش فرض مشتریان، ویژگی هایی که انتخاب آن ها تحت اختیار مشتری نبوده انتخاب نشده که این امر موضوعی قابل تامل می باشد. از آن جا که روش پیشنهادی تلاش می کند در خوشه بندی بدون پیش فرض، مشتریان را به خوشه هایی تقسیم کند تا در هر خوشه افراد شبیه حضور داشته باشند؛ ویژگی هایی که انتخاب آن ها توسط مشتری مدیریت نمی شود (به طور کلی) انتخاب نشده اند. این موضوع صحت روش پیشنهادی را می رساند و نتایج را قابل اعتماد تر می نماید. از جمله ویژگی هایی که تحت اختیار مشتری نیستند و توسط روش پیشنهادی LRFM-GA انتخاب نشده اند عبارتند از: شماره مشتری، کد نوع عملیات و کد پایانه. ویژگی های انتخاب شده توسط روش LRFM-GA در جدول شماره (۴) با علامت مشخص شده اند. در ادامه سه هدف خاص را در دستور کار قرار داده و سعی می شود تا مشتریان با توجه به این اهداف خاص به خوشه های مختلف تقسیم شوند. این سه هدف عبارتند از:

۱) مشتریان خوش حساب در پرداخت تسهیلات دریافتی از بانک
۲) مشتریان خوش حساب از منظر معدود بودن چک های وصول نشده از بانک
۳) مشتریان وفادار به خدمات تلفن همراه و اینترنت، بانک

با توجه به اهداف مشخص شده باید روندی اتخاذ گردد تا خوشه بندی با روش LRFM-GA بر مبنای این اهداف انجام گیرد.

روند مورد نظر که به استثناء چهار ویژگی استفاده شده در LRFM می باید در تمامی ویژگی هایی که در روند تصمیم گیری (به جهت وصول هدفی خاص) موثر هستند حضور داشته باشند و همچنین در تمام کروموزوم های جمعیت برابر با یک در نظر گرفته شوند. به عبارت دقیق تر از بین ۱۷ ویژگی موجود در جدول شماره (۱) که الگوریتم ژنتیک از بین آن ها به یافتن ویژگی های مناسب اقدام می نماید، تمام ویژگی های مرتبط با هدف پیش فرض نیز برابر با یک در نظر گرفته شده و الزاما در کنار ویژگی های روش LRFM قرار

گیرند. سپس، الگوریتم ژنتیک تلاش خواهد کرد از بین ویژگی های باقی مانده، ویژگی هایی را در راستای بهبود خوشه بندی انتخاب نماید.

۱) مشتریان خوش حساب در پرداخت تسهیلات دریافتی از بانک

برای خوشه بندی مشتریان بر اساس خوش حساب بودن آن ها در پرداخت تسهیلات دریافتی، ویژگی آن ها در الگوریتم ژنتیک برابر یک در نظر گرفته شده و پیاده سازی می شود. بنابراین، ویژگی های طول ارتباط مشتری با بانک، تازگی مبادله، تعداد دفعات مبادله، ارزش پولی مبادله شده و میزان تسهیلات معوق از بانک برابر یک در نظر گرفته می شوند. مابقی ۱۶ ویژگی دیگر برای اینکه در کنار این ویژگی ها قرار بگیرند یا خیر در اختیار روش LRFM-GA قرار دارد. نتایج به دست آمده بر اساس خوشه بندی انجام شده بر مبنای روش های پیشنهادی LRFM-GA، LRFM و RFM در جدول شماره (۵) بیان شده است. همچنین، ویژگی هایی که توسط روش LRFM-GA انتخاب شده اند تا در کنار پنج ویژگی دیگر ذکر شده قرار داشته باشند نیز در جدول شماره (۶) بیان گردیده است. این ویژگی ها در کنار پنج ویژگی یاد شده قرار می گیرند تا خوشه بندی برای شناسایی مشتریان خوش حساب در پرداخت تسهیلات دریافتی از بانک صادرات با دقت بیش تری انجام پذیرد.

پیاده سازی دیگری که در کنار پیاده سازی روش های پیشنهادی LRFM-GA، LRFM و RFM انجام پذیرفته است مربوط به خوشه بندی مشتریان بر مبنای پنج ویژگی طول ارتباط مشتری با بانک صادرات، تازگی مبادله، تعداد دفعات مبادله، ارزش پولی مبادله شده و میزان تسهیلات معوق می باشد که نتایج آن نیز در جدول شماره (۵) ثبت گردیده. دلیل این پیاده سازی اثبات ویژگی های دیگری است که در روش LRFM-GA در کنار این پنج ویژگی بکار گرفته شدهو به منظور خوشه بندی مشتریان بانک مناسب بوده است. همان طور که از جدول شماره (۵) بر می آید، میزان شاخص دیویس بولدین در استفاده از روش LRFM-GA بر مبنای ویژگی های موجود در جدول شماره (۶) کم تر از سایر پیاده سازی های می باشد. این موضوع نیز مبین برتری روش پیشنهادی در انجام

خوشه بندی با پیش فرض خوش حسابی در پرداخت تسهیلات است. اعداد ۲. ۸۶۰ و ۲. ۷۹۲ که به ترتیب مربوط به مقادیر شاخص دیویس بولدین در پیاده سازی های RFM و LRFM است در مقایسه با مقدار این شاخص برای روش LRFM-GAدارای مقادیر بالاتری می باشد. دلیل این موضوع را می توان بر عدم تاکید این دو روش بر هدف عنوان نمود. به عبارت دقیق تر، پویایی خاصی در این دو روش وجود ندارد تا بسته به هدف موجود و شرایط داده ها بتوانند خروجی مناسب تری تولید کنند و خود را با شرایط وفق دهند.

جدول ۵. مقایسه خوشه بندی با پیش فرض مربوط به خوش حسابی در پرداخت تسهیلات.

روش	RFM	LRFM	LRFM-GA	استفاده از ۵ ویژگی
مقدار دیویس بولدین	۲. ۸۶۰	۲. ۷۹۲	۱. ۳۰۸	۱. ۹۳۷
مقدار سیلهوتی	۰. ۷۰	۰. ۷۴	۰. ۸۱	۰. ۷۷
تعداد خوشه ها	۲۹	۳۷	۱۴	۳۱

۲) مشتریان خوش حساب از منظر معدود بودن چک های وصول نشده از بانک

برای خوشه بندی مشتریان بر اساس خوش حساب بودن آن ها در مبلغ کل چک های وصول نشده، ویژگی مربوط به این مورد نیز در الگوریتم ژنتیک برابر یک در نظر گرفته می شود. بنابراین، ویژگی های طول ارتباط مشتری با بانک، تازگی مبادله، تعداد دفعات مبادله، ارزش پولی مبادله شده و مبلغ کل چک های وصول نشده از بانک صادرات برابر یک در نظر گرفته خواهد شد و مابقی ۱۶ ویژگی موجود در جدول شماره (۱) برای تشخیص این که در کنار این ویژگی ها قرار بگیرند یا خیر در اختیار روش LRFM-GA خواهند بود. نتایج به دست آمده بر اساس خوشه بندی انجام شده بر مبنای روش های پیشنهادی LRFM-GA، LRFM و RFM در جدول شماره (۷) قابل مشاهده می باشند. همچنین، ویژگی هایی که توسط روش LRFM-GA انتخاب شده اند تا در کنار پنج ویژگی ذکر شده قرار گیرند نیز در جدول شماره (۸) ثبت گردیده است. این ویژگی ها در

کنار پنج ویژگی یاد شده برای شناسایی مشتریان خوش حساب در پرداخت تسهیلات دریافتی از بانک صادرات ایران قرار خواهند گرفت.

پیاده سازی دیگری که در کنار پیاده سازی روش های پیشنهادی LRFM-GA، LRFM و RFM انجام گرفته، مربوط به خوشه بندی مشتریانی است که بر مبنای پنج ویژگی طول ارتباط مشتری با بانک، تازگی مبادله، تعداد دفعات مبادله، ارزش پولی مبادله شده و مبلغ کل چک های وصول نشده از بانک صادرات ارایه گردیده تا مشخص گردد آیا ویژگی های دیگری که در روش LRFM-GA در کنار این پنج ویژگی استفاده شده برای خوشه بندی مشتریان بانک مناسب بوده است.

همان گونه که از جدول شماره (۷) بر می آید، میزان شاخص دیویس بولدین در استفاده از روش LRFM-GA بر مبنای ویژگی های موجود در جدول شماره (۸) کم تر از سایر پیاده سازی های انجام شده است، که این موضوع به نوبه خود مبین برتری روش پیشنهادی در انجام خوشه بندی با پیش فرض خوش حسابی در پرداخت چک می باشد.

۳) مشتریان وفادار به خدمات تلفن همراه و اینترنت بانک صادارت ایران

برای خوشه بندی مشتریان بر اساس وفاداربودن و استفاده کردن آن ها از خدمات تلفن همراه و اینترنت، ویژگی های مربوط به این مورد نیز، در الگوریتم ژنتیک برابر با یک در نظر گرفته می شود. در پایگاه داده این مقاله، هر یک از خدمات همراه بانک و اینترنت بانک بعنوان دو ویژگی جدا از هم در نظر گرفته شده. زیرا ممکن است فردی به یکی از این دو خدمت اقبال بیش تری داشته باشد. بنابراین، ویژگی های طول ارتباط مشتری با بانک، تازگی مبادله، تعداد دفعات مبادله، ارزش پولی مبادله شده، دفعات استفاده مشتریان از همراه بانک و دفعات استفاده مشتریان از اینترنت بانک برابر یک در نظر گرفته می شوند.

جدول ٦. ویژگی های انتخاب شده توسط روش LRFM-GA در آزمایش خوشه بندی با پیش فرض مربوط به خوش حسابی در پرداخت تسهیلات.

دسته بندی	ویژگی	ویژگی های انتخاب شده توسط روش LRFM-GA
داده های تراکنشی	شماره مشتری	
	طول ارتباط مشتری با بانک صادرات	✓
	تازگی مبادله	✓
	تعداد دفعات مبادله	✓
	ارزش پولی مبادله شده	✓
	کد نوع عملیات	
	کد پایانه	
	بانک میزبان	
	تاریخ انجام تراکنش	
داده های استفاده از خدمات	ارزش انتقال وجه شتابی	✓
	مبلغ کل خرید کارت هدیه از بانک صادرات	
	مبلغ کل خرید کالا از کارتخوان بانک صادرات	
	دفعات استفاده مشتریان از همراه بانک صادرات	✓
	دفعات استفاده مشتریان از اینترنت بانک صادرات	✓
	دفعات پرداخت قبض به وسیله بانک صادرات	
	مبلغ کل خرید شارژ تلفن همراه	
داده های استفاده از تسهیلات	مبلغ تسهیلات دریافتی از بانک صادرات	✓
	میزان تسهیلات معوق از بانک صادرات	✓
داده های مالی	میانگین تراکنش ماهیانه	✓
	مبلغ کل سپرده مشتری در بانک صادرات	✓
	مبلغ کل چک های وصول نشده از بانک صادرات	✓

جدول ۷. مقایسه خوشه بندی با پیش فرض مربوط به خوش حسابی در پرداخت چک.

روش	RFM	LRFM	LRFM-GA	استفاده از ۵ ویژگی
مقدار دیویس بولدین	۷۵۲ .۲	۷۱۰ .۲	۴۰۳ .۱	۸۴۲ /۱
مقدار سیلهوتی	۷۱ .۰	۷۳ .۰	۸۲ .۰	۷۸ .۰
تعداد خوشه ها	۳۱	۲۷	۱۹	۲۵

الباقی ۱۵ ویژگی موجود در جدول شماره (۱) برای تشخیص این که در کنار این ویژگی ها قرار بگیرند یا خیر نیز مجددا" در اختیار روش LRFM-GA قرار خواهند گرفت. نتایج به دست آمده بر اساس خوشه بندی انجام شده بر مبنای روش های پیشنهادی LRFM-GA، LRFM و RFM نیز در جدول شماره (۹) ارایه شده است. همچنین، ویژگی هایی که توسط روش LRFM-GA انتخاب شده اند تا در کنار شش ویژگی ذکر شده قرار بگیرند نیز در جدول شماره (۱۰) موجود می باشند. همان طور که از جدول شماره (۹) بر می آید، میزان شاخص دیویس بولدین در استفاده از روش LRFM-GA بر مبنای ویژگی های موجود در جدول شماره (۱۰) کم تر از سایر پیاده سازی های انجام شده است. این موضوع نیز خود مبین برتری روش پیشنهادی در انجام خوشه بندی با پیش فرض خوش حسابی در پرداخت چک است.

جدول ۸. ویژگی های انتخاب شده توسط روش LRFM-GA در آزمایش خوشه بندی با پیش فرض مربوط به خوش حسابی در پرداخت چک.

دسته بندی	ویژگی	ویژگی های انتخاب شده توسط روش LRFM-GA
داده های تراکنشی	شماره مشتری	
	طول ارتباط مشتری با بانک صادرات	✓
	تازگی مبادله	✓
	تعداد دفعات مبادله	✓
	ارزش پولی مبادله شده	✓
	کد نوع عملیات	
	کد پایانه	

دسته بندی	ویژگی	ویژگی های انتخاب شده توسط روش LRFM-GA
	بانک میزبان	
	تاریخ انجام تراکنش	
داده های استفاده از خدمات	ارزش انتقال وجه شتابی	✓
	مبلغ کل خرید کارت هدیه از بانک صادرات	
	مبلغ کل خرید کالا از کارتخوان بانک صادرات	
	دفعات استفاده مشتریان از همراه بانک صادرات	✓
	دفعات استفاده مشتریان از اینترنت بانک صادرات	✓
	دفعات پرداخت قبض به وسیله بانک صادرات	
	مبلغ کل خرید شارژ تلفن همراه	
داده های استفاده از تسهیلات	مبلغ تسهیلات دریافتی از بانک صادرات	
	میزان تسهیلات معوق از بانک صادرات	
داده های مالی	میانگین تراکنش ماهیانه	✓
	مبلغ کل سپرده مشتری در بانک صادرات	✓
	مبلغ کل چک های وصول نشده از بانک صادرات	✓

جدول ۹. مقایسه خوشه بندی با پیش فرض مربوط به استفاده از خدمات همراه بانک و اینترنت بانک

روش	RFM	LRFM	LRFM-GA	استفاده از ۵ ویژگی
مقدار دیویس بولدین	۸۲۹ .۲	۷۹۳ .۲	۳۷۳ .۱	۱۰۳ .۲
مقدار سیلهوتی	۷۰ .۰	۷۱ .۰	۷۹ .۰	۷۷ .۰
تعداد خوشه ها	۳۵	۳۰	۲۵	۲۲

با توجه به مطالب مطروحه می توان استنباط نمود روش LRFM-GA توانسته کارایی مناسبی را در خوشه بندی مشتریان بانک ارایه نماید. این خوشه بندی می تواند با توجه به کلیت خصوصیات مشتریان باشد یا به هدفی خاص بستگی داشته باشد. لازم به ذکر است مقادیر پارامترهای مربوط به روش پیشنهادی در جدول شماره (۱۱) ارایه گردیده است.

جدول ۱۰. ویژگی های انتخاب شده توسط روش LRFM-GA در آزمایش خوشه بندی با پیش فرض مربوط وفاداری به خدمات اینترنت و همراه بانک.

دسته بندی	ویژگی	ویژگی های انتخاب شده توسط روش LRFM-GA
داده های تراکنشی	شماره مشتری	
	طول ارتباط مشتری با بانک صادرات	✓
	تازگی مبادله	✓
	تعداد دفعات مبادله	✓
	ارزش پولی مبادله شده	✓
	کد نوع عملیات	
	کد پایانه	
	بانک میزبان	
	تاریخ انجام تراکنش	
داده های استفاده از خدمات	ارزش انتقال وجه شتابی	✓
	مبلغ کل خرید کارت هدیه از بانک صادرات	✓
	مبلغ کل خرید کالا از کارتخوان بانک صادرات	✓
	دفعات استفاده مشتریان از همراه بانک صادرات	✓
	دفعات استفاده مشتریان از اینترنت بانک صادرات	✓
	دفعات پرداخت قبض به وسیله بانک صادرات	✓
	مبلغ کل خرید شارژ تلفن همراه	✓

دسته بندی	ویژگی	ویژگی های انتخاب شده توسط روش LRFM-GA
داده های استفاده از تسهیلات	مبلغ تسهیلات دریافتی از بانک صادرات	
	میزان تسهیلات معوق از بانک صادرات	
داده های مالی	میانگین تراکنش ماهیانه	
	مبلغ کل سپرده مشتری در بانک صادرات	✓
	مبلغ کل چک های وصول نشده از بانک صادرات	✓

جدول ۱۱. مقادیر پارامترهای روش پیشنهادی.

پارامتر	مقدار
EPS	۲۰
MINPTS	۳۰
تعداد جمعیت اولیه	۱۰۰۰
احتمال جهش	۰۴ .۰
تعداد تکرار	۵۰
نوع پیوند	۴ نقطه ای

مرحله ششم: توسعه مدل LRFM-GA

در این مرحله در خصوص چگونگی گسترش[1] روش، نگهداری[2] برنامه، تولید گزارش نهایی و تغییراتی که می توان در راستای استفاده و ارایه این روش پیشنهادی در سایر حوزه ها بکار گرفت بحث می گردد.

نتیجه گیری و پیشنهادات

هدف اساسی این مقاله، خوشه بندی مشتریان بانک می باشد. در ابتدا، تمامی ویژگی های

1. Development
2. Maintenance

مشتریان از پایگاه داده بانک استخراج و متعاقب آن الگوریتم ژنتیک با جمعیت اولیه (حالات مختلفی از انتخاب هایی را که می توان برای ویژگی های مختلف انتظار داشت) ایجاد می گردد. جهت بررسی میزان مناسب بودن ویژگی های پیشنهادی هر کروموزوم، یکسری قانون ارزیابی وضع گردیده است. اگر در این مرحله کروموزوم پیشنهادی قوانین وضع شده را رعایت کرده باشند در مرحله بعد با استفاده از ویژگی های پیشنهادی نسبت به انجام خوشه بندی به وسیله روش DBSCAN++ پرداخته خواهد شد. میزان مناسب بودن خوشه بندی با استفاده از معیار دیویس بولدین مشخص می گردد (هر چه این شاخص کم تر باشد، خوشه بندی مناسب تری توسط کروموزوم پیشنهاد صورت پذیرفته است). در ادامه با استفاده از چرخ رولت و برازش تمام کروموزوم ها می توان نسل جدید را ایجاد نمود (پس از انجام عمل های پیوند وجهش، نسل جدید ساخته می شود) حال، مراحل الگوریتم ژنتیک برای نسل جدید ادامه می یابد. این تکرار به تعداد مشخصی انجام می پذیرد تا در آخرین نسل، بهترین پیشنهاد برای انتخاب ویژگی ها برگزیده شوند. از این پیشنهاد برای خوشه بندی به وسیله روش DBSCAN++ استفاده شده است و کارایی خوشه بندی نیز با شاخص دیویس بولدین مشخص می شود.

اشکال پیاده سازی روش LRFM-GA را می توان در کندی اجرای آن دانست علی رغم وضع قوانین ارزیابی کروموزوم و استفاده از روش DBSCAN++ که سریع تر از روش استاندارد DBSCAN می باشد، لیکن هنوز کندی اجرا عیب اصلی روش پیشنهادی است که برای برطرف کردن این اشکال می توان از طریق پیاده سازی های نرم افزاری یا قابلیت های سخت افزاری پردازش موازی[1] سرعت را به شدت افزایش داد تا اشکال یاد شده به حداقل میزان کاهش یابد.

ORCID
Mohammad Kazemihttps **http://orcid.org/0000-0002-0382-3921**
Mohammad Ali Keramati **http://orcid.org/0000-0002-6339-6562**
Mehrzad Minooie **http://orcid.org/0000-0002-8474-2938**

1. Parallel

منابع

تقوی فرد، م. خواجوند، س. (۱۳۹۲)، "بخش بندی خوشه ای مشتریان بانک (مورد مطالعه: بانک صادرات ایران) "، *کاوش های مدیریت بازرگانی*، ۵ (۹)، ۶۴-۳۹. doi: 20.1001.1.2645386.1392.5.9.3.7

خدابنده لو، س. نیک نفس، ع. قاضی زاده، ف. (۱۳۹۴)، *"بررسی مزایا، معایب و نحوه کاربرد RFM در ارزیابی وفاداری مشتری"*، دومین همایش ملی پژوهش های کاربردی در علوم کامپیوتر و فناوری اطلاعات.

دهقان، ن. زارع، آ. پیوندی، پ. هادی زاده، م. (۱۳۹۱)،*"استفاده از روش خوشه بندی K-mean جهت گروه بندی فرم بدن"*، هشتمین کنفرانس ملی مهندسی نساجی ایران.

سعدالهی اسرمی، ن. سلاجقه، ا. (۱۳۹۴)،*"مقایسه داده کاوی متدولوژی CRISP و متدولوژی SEMMA"*، دومین کنفرانس بین المللی پژوهش در علوم و تکنولوژی.

سهرابی، ب. خانلری، ا. (۱۳۸۶)، "سنجش ارزش دوره زندگی مشتری(CLV)بر اساس مدلRFM"، *بررسی های حسابداری و حسابرسی*، ۱۴ (۴۷)، ۲۰-۷..

عبدالمنافی، س. نخعی، م. پورجمالی، ف. (۱۳۹۴). *"مدیریت روابط با مشتری"*، کنفرانس بین المللی پژوهش های نوین در مدیریت و مهندسی صنایع.

فتاحی، س. ربیعی، م.(۱۳۹۹) ، "استفاده از مدل LRFM برای خوشه بندی کاربران بر اساس تحلیل رفتار جست و جو (مورد مطالعه: پایگاه اطلاعات علمی ایران (گنج)"، *پژوهش نامه پردازش و مدیریت اطلاعات*، ۳۶ (۲)، ۴۴۲-۴۱۹.

قنبری، ر. مهری، ب. (۱۳۸۸)،*"آشنایی با الگوریتم ژنتیک"*، پایان نامه کارشناسی ارشد، دانشگاه صنعتی شریف، دانشکده ریاضی.

مصلحی، ن. کفاش پور، آ. ناجی عظیمی، ز. (۱۳۹۳)، "استفاده از مدل LRFM برای بخش بندی مشتریان بر اساس ارزش چرخه عمر آن ها"، *پژوهش های مدیریت عمومی*، ۷ (۲۵)، ۱۴۰۱۱۹.

References

Azevedo, A. I. R. L. , Santos, M. F.2008, *"Kdd, Semma And Crisp-Dm: A Parallel Overview"*, Iadis European Conference On Data Mining, Amsterdam, The Netherlands, July 24-26, 2008. Proceedings

Ghosal, A. , Nandy, A. , Das, A. K. , Goswami, S. , Panday, M. ,2020, "A

Short Review On Different Clustering Techniques And Their Applications", *Technology In Modelling And Graphics*, Doi: 10.1007/978-981-13-7403-6_9

Hughes, A. M., 2004, *"Boosting Response With Rfm"*, Marketing Tools

Lakshmanaprabu, S. K. , Shankar, K. , Gupta, D. , Khanna, A. , Rodrigues, J. J. , Pinheiro, P. R. , De Albuquerque, V. H. C. ,2018, "Ranking Analysis For Online Customer Reviews Of Products Using Opinion Mining With Clustering", *Complexity*, (Article Id 3569351):1-9, Doi: 10.1155/2018/3569351

Marisa, F. , Ahmad, S. S. S. , Yusof, Z. I. M. , Hunaini, F. , Aziz, T. M. A. 2019, "Segmentation Model Of Customer Lifetime Value In Small And Medium Enterprise (Smes) Using K-Means Clustering And Lrfm Model", *International Journal Of Integrated Engineering*, 11(3) , Doi 10.30880/Ijie.2019.11.03.018

Mirjalili, S. , Thinsungnoena, T. 2019, "Genetic Algorithm. In Evolutionaryalgorithms And Neural Networks", *Springer, Cham*, Doi.Org/10.1007/978-3-319-93025-1

Niloy, N. H. , Navid, M. A. I. ,2018, "Naïve Bayesian Classifier And Classification Trees For The Predictive Accuracy Of Probability Of Default Credit Card Clients",*American Journal Of Data Mining Knowledge Discovery*, 3 (1), 1-12. Doi: 10.11648/J.Ajdmkd.20180301.11

Palaniappan, S. , Mustapha, A. Foozy, C. F. M. Atan, R. , Navid, S. , 2017, "Customer Proifling Using Classiifcation Approach For Bank Telemarketing" *International Journal Of Computer Vision*, 1 (4-2), 214- 217, Doi: 10.30630/Joiv.1.4-2.68

Patil, P. S. , Dharwadkar, N. V. ,2017, "Analysis Of Banking Data Using Machine Learning",*In International Conference On I-Smac (Iot In Social, Mobile, Analytics And Cloud)* , Doi: 10.1109/I-Smac.2017.8058305

Petrovic. S., 2006, *"A Comparison Between The Silhouette Index And The Davies-Bouldin Index In Labelling Ids Clusters"*,In Proceedings Of The 11th Nordic Workshop Of Secure It Systems, Corpus Id: 14657641

Vergani, A. A. , Binaghi, E. , Niloy, A. , Petrovic, H. , 2018, "A Soft Davies-Bouldin Separation Measure",In 2018 *Ieee International Conference On Fuzzy Systems (Fuzz-Ieee)* , Doi: 10.1109/Fuzz-Ieee.2018.8491581

Wei, J. T. , Lin, S. Y. , & Wu, H. H., 2010, "A Review Of The Application Of Rfm Model", *African Journal Of Business Management* 4(19):4199-4206

Yang, S. , Zhang, H. , 2018, "Comparison Of Several Data Mining Methods

In Credit Card Default Prediction",*Intelligent Information Management*, 10 (5), Doi: 10.4236/Iim.2018.105010

Zhang, Y. , Bradlow, E. T. , Small, D. S. , Silhouette, G. ,2015, "Predicting Customer Value Using Clumpiness: From Rfm To Rfmc",*Marketing Science*, 34 (2), 179-307. Doi.Org/10.1287/Mksc.2014.0873

Zoeram A. A. , Karimi Mazidi, A. R. 2018, "New Approach For Customer Clustering By Integrating The Lrfm Model And Fuzzy Inference System*", Iranian Journal Of Management Studies*, 11 (2), 351-378. Doi: 10.22059/Ijms.2018.242528.672839

References [in Persian]

Taghi Taghavifard, M., Khajvand, S., 2011, Customer Segmentationof Bank Customers (Case: Saderatbank Of Iran), *Business Management Explorations*, 5 (9), 64-39.Doi: 20.1001.1.2645386.1392.5.9.3.7 [In Persian]

Khodabandehl;, S.,. Nik Nafs, A. Ghazizadeh, F. 2014, *"Study Of Advantages, Disadvantages And Application Of Rfm In Assessing Customer Loyalty",* The Second National Conference On Applied Research In Computer Science And Information Technology. [In Persian]

Dehghan, N., Zare, A. Peyvandi, P. Hadizadeh, M. 2012, "*Using K-Mean Clustering Method To Group Body Shape",* 8th National Conference On Textile Engineering In Iran. [In Persian]

Sadollahi Asrami, N., Solajgheh, 2015, *A., Comparison Of Data Mining Crisp Methodology And Semma Methodology*", The Second International Conference On Research In Science And Technology. [In Persian]

Sohrabi, B. Khanlari, A. 2007, *"Measuring The Value Of The Customer Life Cycle (Clv) Based On The Rfm Model",* Accounting And Auditing Reviews, 14 (47), 20-7 .. [In Persian]

Abdolmanafi, S., Nakhai, M. Pourjamali, F. 2015, "Customer Relationship Management", *International Conference On New Research In Industrial Management And Engineering*[In Persian].

Fatahi, S., Rabiee, M., 2020, Using The Lrfm Model For User Clustering Based On Search Behavior Analysis (Case Study: Iran Scientific Database (Ganj)", *Journal Of Information Processing And Management*, 36 (2), 442-419. [In Persian]

Ghanbari, R. Mehri, B. 2009, *"Introduction To Genetic Algorithm",* Master Thesis, Sharif University Of Technology, Faculty Of Mathematics. [In Persian]

Moslehi, N., Kafashpour, A. Naji Azimi, Z. 2014, "Using The Lrfm Model To Segment Customers Based On The Value Of Their Life Cycle",

Public Management Research, 7 (25), 140119. [In Persian]

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

استناد به این مقاله: کاظمی، محمد.، کرامتی، محمدعلی.، مینوئی، مهرزاد. (۱۴۰۰). ارایه مدل بهبودیافته LRFM به جهت خوشه‌بندی مشتریان بانک‌ها بر مبنای الگوریتم ژنتیک، *مطالعات مدیریت کسب وکار هوشمند*، ۱۰(۳۸)، ۳۵۶-۳۱۷.

DOI: 10.22054/IMS.2021.58551.1902

Journal of Business Intelligence Management Studies is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License..