

Bank Customer Segmentation Using Extended RFM Indicators and MiniBatchKMeans Algorithm

Zahra Shahi Shavoon ¹, Mohammadreza Rostami ^{2*}, Bjian Nahavandi ³

¹ Department of Financial Management, Science and Research Branch, Islamic Azad University, Tehran, Iran

^{2*} Associate Professor, Financial Management, Faculty of Social and Economic Sciences, Alzahra University, Tehran, Iran (Corresponding author), Email: m.rostami@alzahra.ac.ir

³ Department of Industrial Management and Technology, Science and Research Branch, Islamic Azad University, Tehran, Iran

ARTICLE INFO

Article history:

Received:31/07/2026

Received in revised form:18/08/2026

Accepted:29/08/2026

Available online:23/09/2026

Keywords:

Financing
customer clustering
extended RFM, MiniBatchKMeans
data-driven banking
customer relationship management
banking data mining

ABSTRACT

Accurate customer knowledge and robust segmentation are prerequisites for data-driven banking. This study designs and evaluates a scalable clustering model for individual bank customers using extended RFM concepts and MiniBatchKMeans. Operational data from a private bank were aggregated at customer level and processed through data-quality controls, signed logarithmic transformations for skewed financial variables, outlier control, and standardized preprocessing fitted without leakage. The final clustering space contained ten robust and traceable features: signed-log financial value, profitability and average balance; log-transformed recency and frequency; digital ratio and digital intensity; channel diversity and service diversity; and a robust profit-margin indicator. Because only 257 positive risk events were observed, risk variables were excluded from the clustering space and retained as a separate post-clustering governance layer. Cluster counts from $k=2$ to $k=8$ were assessed using Silhouette, Davies-Bouldin, Calinski-Harabasz, multi-seed ARI, aligned Jaccard stability, and minimum cluster size. The final two-cluster solution achieved Silhouette=0.5519, Davies-Bouldin=0.9009, Calinski-Harabasz=11999.13, mean ARI=0.9976 and mean Jaccard=0.9986. Among 378,312 clusterable customers, 66.74% belonged to a less-digital/lower-financial-value group and 33.26% to a more-digital/higher-financial-value group; 9,307 records with simultaneously zero financial base were kept in a separate data-quality queue rather than interpreted as a customer segment. The results support MiniBatchKMeans as a scalable and interpretable tool for banking customer segmentation, while emphasizing that cluster labels are descriptive and should not be treated as causal, credit-risk, or future-behavior predictions.

Article Type: Research Paper



©Authors

Cite: Shahi Shavoon , Z., Rostami, M. & Nahavandi, B. (2026). Bank Customer Segmentation Using Extended RFM Indicators and MiniBatchKMeans Algorithm. Journal of Intelligent Financial Management, 2(3), 11-32.

Journal of Intelligent Financial Management,
2025, Vol. 2, No.3, pp. 11- 32

Publish by:

Tolou-e Binsh-e Ayandeh Scientific Institute

[https://doi.org/ 10.25843/JIFM.2025.8563.78453](https://doi.org/10.25843/JIFM.2025.8563.78453)

بخش بندی مشتریان حقیقی بانک با شاخص های توسعه یافته RFM و الگوریتم MiniBatchKMeans

زهرا شاهی شاوون^۱، محمد رضا رستمی^{۲*}، بیژن نهاوندی^۳

۱ - گروه مدیریت مالی، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران

۲ و ۳ - دانشیار، مدیریت مالی، دانشکده علوم اجتماعی و اقتصادی دانشگاه الزهراء، تهران، ایران (نویسنده مسئول)، ایمیل نویسنده مسئول: m.rostami@alzahra.ac.ir

۳ - گروه مدیریت صنعتی و فناوری، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران

اطلاعات مقاله

تاریخچه مقاله:

تاریخ دریافت: ۱۴۰۵/۰۵/۰۹

تاریخ بازنگری: ۱۴۰۵/۰۵/۲۷

تاریخ پذیرش: ۱۴۰۵/۰۶/۰۷

تاریخ انتشار: ۱۴۰۵/۰۷/۰۱

کلیدواژه ها:

خوشه بندی مشتریان

RFM توسعه یافته

MiniBatchKMeans

بانکداری داده محور

مدیریت ارتباط با مشتری

داده کاوی بانکی

چکیده

شناخت دقیق مشتریان و تفکیک آنان به گروه های همگن، یکی از پیش نیازهای اصلی تصمیم گیری هوشمند در بانکداری داده محور است. پژوهش حاضر با هدف طراحی و ارزیابی مدل خوشه بندی مقیاس پذیر مشتریان حقیقی بانک مبتنی بر منطق RFM توسعه یافته و الگوریتم MiniBatchKMeans انجام شد. داده های عملیاتی یکی از بانک های خصوصی کشور پس از تجمع در سطح مشتری، کنترل کیفیت، مهندسی ویژگی، تبدیل های لگاریتمی مناسب، کنترل مقادیر افراطی و استانداردسازی بدون نشت داده برای مدل سازی آماده شدند. فضای نهایی خوشه بندی شامل ده ویژگی مقاوم و قابل ردیابی بود: لگاریتم علامت دار ارزش مالی، سودآوری و میانگین مانده؛ لگاریتم تأخر و تناوب؛ نسبت و شدت دیجیتال؛ تنوع کانال و تنوع خدمات؛ و حاشیه سود مقاوم. با توجه به وجود تنها ۲۵۷ رخداد مثبت ریسک، متغیرهای ریسک از فضای خوشه بندی حذف و در لایه پسینی کنترل ریسک نگهداری شدند. برای انتخاب تعداد خوشه، مقادیر k از ۲ تا ۸ با معیارهای Calinski-Harabasz، Davies-Bouldin، Silhouette برابر ۰،۵۵۱۹، برای Silhouette برابر ۰،۹۹۷۶ و برای Calinski-Harabasz برابر ۰،۹۹۸۶ انتخاب شد. از ۱۱۹۹۹،۱۳ میانگین ARI برابر ۰،۹۹۷۶ و میانگین جاکارد برابر ۰،۹۹۸۶ انتخاب شد. از ۳۷۸۰۳۱۲ مشتری مدل پذیر، ۶۶،۷۴ درصد در گروه کم دیجیتال تر و کم ارزش تر و ۳۳،۲۶ درصد در گروه دیجیتال تر و با ارزش مالی بالاتر قرار گرفتند. همچنین ۹۰۳۰۷ رکورد با پایه مالی هم زمان صفر به جای تشکیل یک خوشه مصنوعی، در صف کنترل کیفیت نگهداری شدند. یافته ها نشان می دهد MiniBatchKMeans می تواند برای بخش بندی مقیاس پذیر و قابل تفسیر مشتریان بانکی مفید باشد، مشروط بر آنکه نام گذاری خوشه ها توصیفی باقی بماند و به پیش بینی آینده، علیت یا ریسک اعتباری تعبیر نشود.

نوع مقاله: پژوهشی



© نویسندگان

استناد: شاهی شاوون، زهرا و رستمی، محمد رضا و نهاوندی، بیژن. (۱۴۰۵). بخش بندی مشتریان حقیقی بانک با شاخص های توسعه یافته RFM و الگوریتم. مدیریت مالی هوشمند، ۳(۳)، ۱۱-۳۲.

نشریه مدیریت مالی هوشمند، ۱۴۰۵، دوره ۲، شماره ۳، صفحه ۱۱-۳۲.

ناشر: موسسه علمی طلوع بینش آینده

<https://doi.org/10.25843/JIFM.2025.8563.78453>

مقدمه

تحول دیجیتال در صنعت بانکداری، منطق رقابت میان بانک‌ها را از ارائه خدمات عمومی و محصول محور به سمت شناخت دقیق، داده محور و شخصی سازی شده مشتریان تغییر داده است. گسترش بانکداری الکترونیک، افزایش استفاده از کانال‌های غیر حضوری، رشد فین تک‌ها و انباشت حجم عظیمی از داده‌های تراکشی و رفتاری موجب شده است که بانک‌ها برای حفظ مشتریان ارزشمند، کاهش هزینه خدمت‌رسانی و طراحی پیشنهادهای متناسب، به روش‌های تحلیلی پیشرفته نیازمند شوند (باروسو و لایورد، ۲۰۲۲؛ نوفل، ۲۰۲۴). در چنین فضایی، بخش‌بندی سنتی مشتریان بر پایه شاخص‌های محدود مانند مانده حساب، سن، درآمد یا طبقه‌بندی‌های عمومی دیگر پاسخگوی پیچیدگی رفتار مشتریان بانکی نیست؛ زیرا مشتریان ممکن است از نظر ارزش مالی، سودآوری، سطح ریسک، الگوی تراکنش، میزان استفاده از خدمات دیجیتال و هزینه خدمت تفاوت‌های معناداری داشته باشند (یان و همکاران، ۲۰۲۵؛ فالکون و همکاران، ۲۰۲۵).

یکی از رویکردهای رایج در تحلیل رفتار مشتری، مدل RFM^۱ است که بر سه بعد تازگی تعامل، فراوانی تعامل و ارزش مالی مشتری تمرکز دارد (هندجو و همکاران، ۲۰۲۳؛ ویلبرت و همکاران، ۲۰۲۳). این مدل به دلیل سادگی، قابلیت تفسیر و ارتباط مستقیم با رفتار مشتری، در مطالعات بازاریابی و مدیریت ارتباط با مشتری کاربرد گسترده‌ای یافته است. با این حال، در محیط بانکداری، اتکا به RFM کلاسیک می‌تواند تصویری ناقص از مشتری ارائه دهد؛ زیرا ارزش مشتری بانکی صرفاً از تعداد یا مبلغ تراکنش‌ها حاصل نمی‌شود، بلکه شاخص‌هایی مانند سودآوری خالص، هزینه خدمت، ریسک اعتباری، رفتار دیجیتال، تنوع کانال و پایداری رابطه مالی نیز در تصمیم‌گیری بانکی اهمیت دارند (هسیه، ۲۰۰۴؛ کاروالیو و ماسیل، ۲۰۲۶). بنابراین، توسعه مدل RFM با شاخص‌های مالی، رفتاری، دیجیتال و ریسکی می‌تواند مبنای دقیق‌تری برای شناخت مشتریان حقیقی بانک فراهم سازد.

خوشه‌بندی به‌عنوان یکی از مهم‌ترین روش‌های یادگیری بدون نظارت، امکان کشف الگوهای پنهان در داده‌های مشتریان را بدون نیاز به برچسب از پیش تعیین‌شده فراهم می‌کند. مطالعات جدید در حوزه بانکداری نشان می‌دهند که ترکیب شاخص‌های رفتاری و الگوریتم‌های خوشه‌بندی می‌تواند به شناسایی گروه‌های متمایز مشتریان و طراحی راهبردهای بازاریابی هدفمند کمک کند (یان و همکاران، ۲۰۲۵). با این حال، داده‌های بانکی معمولاً حجیم، نامتوازن، دارای مقادیر پرت، توزیع‌های چوله و متغیرهای ناهمگن هستند؛ از این‌رو، انتخاب الگوریتمی مقیاس‌پذیر و قابل استفاده در محیط‌های عملیاتی اهمیت ویژه‌ای دارد. الگوریتم MiniBatchKMeans با استفاده از نمونه‌های کوچک در هر تکرار، هزینه محاسباتی خوشه‌بندی را کاهش داده و برای داده‌های بزرگ بانکی گزینه‌ای مناسب‌تر از اجرای کامل K-Means^۲ محسوب می‌شود (اسکالی، ۲۰۱۰).

از سوی دیگر، پروژه‌های داده‌کاوی در صنعت مالی نباید صرفاً از منظر دقت الگوریتمی ارزیابی شوند، بلکه لازم است کیفیت داده، قابلیت تفسیر نتایج، ملاحظات حاکمیت داده و امکان استفاده مدیریتی از خروجی‌ها نیز مورد توجه قرار گیرد. چارچوب FIN-DM تأکید می‌کند که داده‌کاوی در خدمات مالی باید با الزامات ریسک، انطباق، حاکمیت و کیفیت داده همراه باشد (پلاتنیکووا و همکاران، ۲۰۲۳). همچنین نهادهای بین‌المللی مالی بر این نکته تأکید دارند که استفاده از هوش مصنوعی و تحلیل داده در خدمات مالی، در کنار مزایایی مانند شخصی‌سازی خدمات و تحلیل پیشرفته، نیازمند کنترل ریسک مدل، شفافیت و مدیریت کیفیت داده است.

با وجود کاربرد گسترده مدل RFM در بخش‌بندی مشتریان، اتکای این مدل به سه مؤلفه تازگی، فراوانی و ارزش مالی، در محیط بانکداری تصویری کامل از ماهیت رابطه مشتری با بانک ارائه نمی‌کند. ممکن است مشتری از نظر حجم یا تعداد تراکنش‌ها ارزشمند به نظر برسد، اما هم‌زمان از سودآوری پایین، هزینه خدمت بالا یا الگوی متفاوتی از تعامل دیجیتال برخوردار باشد. از

¹ Barroso & Laborda

² Nofal

³ Yan

⁴ Falcón

⁵ recency, frequency, and monetary

⁶ Handojo

⁷ Wilbert

⁸ Hsieh

⁹ Carvalho & Maciel

¹⁰ Sculley

¹¹ Plotnikova

سوی دیگر، بخشی از مطالعات پیشین توسعه مدل RFM را صرفاً از طریق افزودن یک یا چند متغیر تکمیلی دنبال کرده‌اند، در حالی که مسئله بانکداری تنها افزایش تعداد متغیرها نیست، بلکه ایجاد یک فضای ویژگی منسجم، مقاوم و قابل تفسیر است که ابعاد مالی و رفتاری مشتری را بدون ایجاد افزونگی، نسبت‌های نامعین یا غلبه متغیرهای کم‌رخداد در ساختار خوشه‌ها بازنمایی کند. بر همین اساس، پژوهش حاضر به جای افزودن مکانیکی متغیرها به RFM، پس از بررسی کیفیت، واریانس، همبستگی، چولگی و قابلیت تفسیر ویژگی‌ها، مجموعه‌ای از متغیرهای مقاوم مالی، رفتاری و دیجیتال را برای خوشه‌بندی انتخاب می‌کند و متغیرهای نامناسب، از جمله شاخص‌های بسیار کم‌رخداد، را در لایه‌ای مستقل از خوشه‌بندی مدیریت می‌کند. این منطق موجب می‌شود چارچوب پیشنهادی برخی محدودیت‌های مدل‌های ساده RFM، از جمله تک‌بعدی بودن ارزش مشتری، حساسیت به مقادیر خام و ناتوانی در بازنمایی رفتار چندکاناله را کاهش دهد. در رساله نیز وجه تمایز پژوهش صرفاً افزودن متغیرها تعریف نشده، بلکه بر غریب ویژگی‌ها، کنترل کیفیت داده، ارزیابی پایداری خوشه‌ها و حرکت از بخش‌بندی صرف به یک چارچوب قابل ممیزی برای تصمیم‌گیری تأکید شده است.

از منظر الگوریتمی نیز انتخاب MiniBatchKMeans بر مبنای ادعای «برتری مطلق» آن نسبت به همه روش‌های خوشه‌بندی صورت نگرفته، بلکه بر اساس تناسب آن با مسئله پژوهش انجام شده است. روش‌های سلسله‌مراتبی اگرچه ساختار درختی مناسبی ارائه می‌کنند، در داده‌های بزرگ از هزینه محاسباتی بالاتری برخوردارند؛ روش‌هایی نظیر DBSCAN در شناسایی نویز و خوشه‌های نامنظم مفیدند، اما نسبت به پارامترهای چگالی حساس‌اند؛ روش‌های مدل‌محور مانند GMM امکان عضویت احتمالی را فراهم می‌کنند، ولی به مفروضات توزیعی و پیچیدگی بیشتری نیاز دارند؛ و روش‌های عمیق نیز عموماً هزینه محاسباتی و دشواری تفسیر بیشتری دارند. در مقابل، MiniBatchKMeans ضمن حفظ سادگی و تفسیرپذیری خانواده K-Means، با استفاده از دسته‌های کوچک تصادفی در هر تکرار، هزینه پردازش را برای داده‌های حجیم کاهش می‌دهد و از این جهت برای محیط عملیاتی بانک مناسب‌تر است. داده‌های پژوهش در مقیاس چندصد هزار مشتری قرار دارند و اجرای مکرر مدل برای ارزیابی مقادیر مختلف k و پایداری چندپذیری ضرورت داشته است؛ بنابراین سرعت اجرا، مقیاس‌پذیری و امکان تفسیر روشن مراکز خوشه از معیارهای اصلی انتخاب الگوریتم بوده‌اند.

البته حساسیت K-Means و MiniBatchKMeans به مقیاس متغیرها، مقادیر پرت و ساختارهای غیرکروی نادیده گرفته نشده است. به همین دلیل، در این پژوهش پیش از خوشه‌بندی، تبدیل متغیرهای چوله، کنترل مقادیر افراطی، استانداردسازی و بررسی افزونگی متغیرها انجام شده است. بنابراین انتخاب MiniBatchKMeans باید به عنوان یک موازنه میان مقیاس‌پذیری، هزینه محاسباتی، تفسیرپذیری و تناسب با هدف مدیریتی پژوهش تفسیر شود، نه ادعای اینکه این روش در تمام انواع داده‌ها از سایر الگوریتم‌ها برتر است.

بر این اساس، شکاف پژوهشی مطالعه حاضر فقدان چارچوبی صرفاً با «متغیرهای بیشتر» نیست، بلکه کمبود رویکردی است که در مقیاس واقعی داده‌های بانکی، توسعه RFM را با غریب نظام‌مند ویژگی‌ها، خوشه‌بندی مقیاس‌پذیر، ارزیابی هم‌زمان کیفیت و پایداری خوشه‌ها و تفسیر مدیریتی نتایج تلفیق کند. سهم روش‌شناختی پژوهش، توسعه عملیاتی چارچوب RFM از طریق طراحی و غریب ویژگی‌های بانکی مقاوم، تلفیق ابعاد مالی، رفتاری و دیجیتال، کنترل متغیرهای کم‌رخداد و ارزیابی ساختار خوشه‌ها بر مبنای هم‌زمان کیفیت، پایداری و قابلیت تفسیر مدیریتی است.

بر این اساس، پژوهش حاضر در پی توسعه رویکردی است که ضمن بهره‌گیری از شاخص‌های توسعه‌یافته RFM و ویژگی‌های مالی، رفتاری و دیجیتال مشتریان، امکان خوشه‌بندی مقیاس‌پذیر، پایدار و قابل تفسیر مشتریان حقیقی بانک را فراهم سازد. از این رو، سؤال اصلی پژوهش چنین مطرح می‌شود: مدل مناسب برای خوشه‌بندی مشتریان حقیقی بانک بر پایه شاخص‌های توسعه‌یافته RFM و الگوریتم MiniBatchKMeans چگونه طراحی و ارزیابی می‌شود؟

ادامه پژوهش به این صورت سازمان‌دهی شده است: در بخش دوم، مبانی نظری و پیشینه پژوهش در زمینه بانکداری داده‌محور، خوشه‌بندی مشتریان، مدل RFM و توسعه آن در محیط بانکی، و الگوریتم‌های خوشه‌بندی مقیاس‌پذیر مرور می‌شود و در پایان این بخش، شکاف پژوهشی مطالعه حاضر تبیین می‌گردد. بخش سوم به روش‌شناسی پژوهش اختصاص دارد و در آن منبع داده‌ها، فرآیند آماده‌سازی و پاک‌سازی داده، مهندسی ویژگی، ساخت شاخص‌های توسعه‌یافته RFM، نرمال‌سازی متغیرها، اجرای الگوریتم MiniBatchKMeans و معیارهای ارزیابی تعداد خوشه‌ها تشریح می‌شود. در بخش چهارم، یافته‌های تجربی پژوهش شامل نتایج ارزیابی تعداد خوشه‌ها، انتخاب مدل نهایی، پروفایل‌سازی و نام‌گذاری خوشه‌های مشتریان ارائه می‌شود. بخش پنجم به بحث و

تحلیل نتایج، مقایسه یافته‌ها با مطالعات پیشین و تبیین دلالت‌های مدیریتی خوشه‌های شناسایی شده می‌پردازد. در نهایت، بخش ششم نتیجه‌گیری، نوآوری‌ها، پیشنهادهای کاربردی، محدودیت‌های پژوهش و مسیرهای پیشنهادی برای مطالعات آینده را ارائه می‌کند.

مبانی نظری و پیشینه پژوهش

بانکداری داده‌محور و ضرورت شناخت هوشمند مشتریان

بانکداری معاصر بیش از گذشته بر داده، تحلیل و تصمیم‌گیری هوشمند متکی است. گسترش بانکداری دیجیتال، کانال‌های غیرحضور، فین‌تک‌ها و رشد داده‌های تراکنشی، رفتاری و اعتباری، بانک‌ها را به فاصله‌گرفتن از رویکردهای سنتی و حرکت به سمت بانکداری داده‌محور سوق داده است (گل و الفرین، ۲۰۲۳؛ یونگ‌جی و شانیه‌ث، ۲۰۲۵). در این رویکرد، تصمیم‌های مربوط به شناخت مشتری، طراحی محصول، تخصیص خدمت، مدیریت ریسک و بازاریابی بر تحلیل داده‌های واقعی استوار است. با توجه به تفاوت مشتریان از نظر تراکنش، مانده، رفتار دیجیتال، رابطه اعتباری، سودآوری، هزینه خدمت و ریسک، طبقه‌بندی بر پایه چند شاخص محدود نمی‌تواند تصویر جامعی از ارزش و رفتار آنان ارائه دهد (نوفل، ۲۰۲۴؛ گل و الفرین، ۲۰۲۳). از این رو، شناخت چندبعدی مشتری و تحلیل یکپارچه اطلاعات مالی، رفتاری، دیجیتال و ریسکی اهمیت یافته است.

مطالعات خدمات مالی نشان می‌دهند هوش مصنوعی، یادگیری ماشین و داده‌کاوی می‌توانند در شخصی‌سازی خدمات، بهبود تجربه مشتری، افزایش کارایی و مدیریت ریسک مؤثر باشند. در عین حال، استفاده از این فناوری‌ها مستلزم توجه به کیفیت داده، ریسک مدل و کنترل‌های نظارتی است (دوان^۳ و همکاران، ۲۰۱۹). همچنین، در بانکداری و بیمه بر ضرورت همراهی کاربرد هوش مصنوعی با حاکمیت مناسب، شفافیت و ملاحظات نظارتی تأکید شده است (دیویدی^۴ و همکاران، ۲۰۱۹). در سطح فرآیندی نیز، پلاتینیکووا^۵ و همکاران (۲۰۲۲) در چارچوب FIN-DM نشان می‌دهند که فرآیندهای عمومی داده‌کاوی مانند CRISP-DM باید در خدمات مالی با الزامات حاکمیت داده، انطباق، مدیریت ریسک و تضمین کیفیت تکمیل شوند.

از منظر مدیریت ارتباط با مشتری، شناخت هوشمند مشتریان پیش‌شرط شخصی‌سازی خدمات و تخصیص مؤثر منابع است. مطالعات بازاریابی بانکی نیز نشان می‌دهند که هوش مصنوعی و تحلیل داده می‌توانند جذب، نگهداشت، تعامل شخصی‌سازی شده و کارایی عملیاتی را بهبود دهند (گرلینگ و لسمن^۶، ۲۰۲۴). در این میان، خوشه‌بندی به‌عنوان یکی از روش‌های یادگیری بدون نظارت، ابزار مهمی برای کشف گروه‌های پنهان مشتریان و تبدیل داده‌های خام به بینش مدیریتی است. با این حال، داده‌های بانکی معمولاً حجیم، چوله، ناهمگون و دارای مقادیر پرت‌اند و ممکن است مشتریان از نظر ارزش مالی، ریسک و سودآوری الگوهای متفاوتی داشته باشند. بنابراین، مدل‌های تحلیلی باید فراتر از شاخص‌های تک‌بعدی، مجموعه‌ای از ابعاد رفتاری، مالی، دیجیتال، سودآوری، هزینه و ریسک را در نظر گیرند. بر این اساس، بانکداری داده‌محور نه فقط یک تحول فناورانه، بلکه ضرورتی راهبردی برای شناخت دقیق‌تر مشتریان و پشتیبانی از تصمیم‌گیری مدیریتی است.

خوشه‌بندی مشتریان در صنعت بانکداری

خوشه‌بندی مشتریان یکی از کاربردهای مهم یادگیری بدون نظارت در بانکداری است که بدون نیاز به برچسب‌های از پیش تعیین‌شده، مشتریان را بر اساس شباهت‌های رفتاری، مالی، اعتباری و دیجیتال گروه‌بندی می‌کند. برخلاف بخش‌بندی سنتی مبتنی بر قواعد ثابت یا متغیرهای محدودی مانند سن، درآمد و مانده حساب، این رویکرد ساختارهای پنهان در داده‌های واقعی را شناسایی کرده و می‌تواند مبنایی برای تصمیم‌های بازاریابی، مدیریت ارتباط با مشتری و تخصیص خدمات باشد. مطالعات پیشین نیز کاربرد خوشه‌بندی در طراحی خدمات هدفمند و شناخت گروه‌های متمایز مشتریان را تأیید کرده‌اند. علی‌اف^۷ و همکاران (۲۰۲۰) با ترکیب RFM و الگوریتم‌های خوشه‌بندی در داده‌های واقعی بانکی نشان دادند که ارائه یکسان خدمات به همه مشتریان رویکردی کارآمد نیست و شناسایی خدمت مناسب برای گروه مناسب می‌تواند اثربخشی فروش خدمات را افزایش دهد. در مطالعات

¹ Gul & Al-Faryan

² Yongjie & Shanyue

³ Duan

⁴ Dwivedi

⁵ Plotnikova

⁶ Gerling & Lessmann

⁷ Aliyev

جدیدتر، تمرکز از شاخص‌های محدود به سمت ترکیب داده‌های رفتاری، زمانی و چندمنبعی حرکت کرده است. برخوردار و همکاران (۲۰۲۱) نیز با تلفیق LSTM Encoder-Decoder، فاصله‌های Dynamic Time Warping، شاخص‌های RFM و داده‌های جمعیت‌شناختی نشان دادند که کیفیت و نحوه بازنمایی ویژگی‌های مشتری نقش مهمی در ایجاد خوشه‌های معنادار دارد. از منظر الگوریتمی، روش‌هایی مانند K-Means، خوشه‌بندی سلسله‌مراتبی، DBSCAN، Gaussian Mixture Models، Fuzzy C-Means و شبکه‌های عصبی در بخش‌بندی مشتریان به کار رفته‌اند (داوود و همکاران، ۲۰۱۹). K-Means و نسخه‌های توسعه‌یافته آن به دلیل سادگی، سرعت و تفسیرپذیری همچنان پرکاربردند؛ در مقابل، DBSCAN قابلیت شناسایی خوشه‌های غیرکروی و نقاط پرت را دارد، اما به پارامترهای چگالی حساس است. یان و همکاران (۲۰۲۵) نیز نشان دادند که ترکیب روش‌های مرکزگرا و چگالی‌محور در الگوریتم KM-DBSCAN می‌تواند کیفیت بخش‌بندی مشتریان بانکی را بهبود دهد. با این حال، در محیط عملیاتی بانک، انتخاب الگوریتم علاوه بر کیفیت آماری باید بر اساس سرعت، مقیاس‌پذیری، تفسیرپذیری و قابلیت استقرار انجام شود. از این رو، MiniBatchKMeans برای داده‌های حجیم گزینه‌ای عملی است؛ زیرا به جای پردازش کل داده در هر تکرار، از زیرمجموعه‌های کوچک تصادفی استفاده کرده و هزینه محاسباتی را کاهش می‌دهد. بنابراین، پژوهش حاضر با تلفیق شاخص‌های توسعه‌یافته RFM و MiniBatchKMeans در پی ارائه مدلی مقیاس‌پذیر و قابل تفسیر برای شناخت مشتریان حقیقی بانک است.

مدل RFM و کاربرد آن در تحلیل مشتریان بانکی

مدل RFM یکی از پرکاربردترین چارچوب‌های کمی در تحلیل رفتار مشتری است و بر سه شاخص تازگی تعامل، فراوانی تعامل و ارزش مالی استوار است. این مدل به دلیل سادگی، تفسیرپذیری و اتکا به داده‌های واقعی رفتاری، در مدیریت ارتباط با مشتری و شناسایی مشتریان ارزشمند کاربرد گسترده‌ای دارد. بولت و وانسیک (۱۹۹۵) نیز نشان دادند که استفاده از این شاخص‌ها می‌تواند در انتخاب مشتریان هدف و تخصیص مؤثر منابع بازاریابی نقش داشته باشد.

در بانکداری، مفاهیم RFM باید متناسب با ماهیت خدمات مالی بازتعریف شوند. Recency می‌تواند فاصله از آخرین تراکنش یا استفاده از خدمت، Frequency تعداد تراکنش‌ها یا تعاملات، و Monetary مانده حساب، حجم تراکنش یا ارزش مالی رابطه مشتری با بانک را نشان دهد. مطالعات بانکی نیز نشان داده‌اند که ترکیب RFM با روش‌های خوشه‌بندی می‌تواند به شناخت بهتر مشتریان و طراحی خدمات هدفمند کمک کند. علی‌اف و همکاران (۲۰۲۰) نشان دادند ارائه یکسان خدمات به همه مشتریان کارآمد نیست، و برخوردار و همکاران (۲۰۲۱) نیز نشان دادند ترکیب RFM با ویژگی‌های زمانی تراکنش و روش‌های یادگیری عمیق می‌تواند خوشه‌های معنادارتری ایجاد کند.

با این حال، RFM کلاسیک تنها سه بعد اصلی را پوشش می‌دهد و متغیرهایی مانند سودآوری، هزینه خدمت، ریسک اعتباری، رفتار تسهیلاتی، تنوع کانال و استفاده از خدمات دیجیتال را مستقیماً در نظر نمی‌گیرد. از این رو، ادبیات جدید به سمت توسعه این چارچوب حرکت کرده است. مدل‌هایی مانند RFMB با افزودن Balance، RFME با افزودن Engagement و RFMI با افزودن Interval تلاش کرده‌اند رفتار مشتری را چندبعدی‌تر تحلیل کنند. گان و همکاران (۲۰۲۴) نشان دادند افزودن Interval می‌تواند ارزیابی ارزش مشتری را بهبود دهد و پاندی و همکاران (۲۰۲۴) نیز نقش Engagement را در بخش‌بندی رفتاری و طراحی راهبردهای بازاریابی تأیید کردند.

بر این اساس، در پژوهش حاضر RFM به عنوان نقطه آغاز تحلیل در نظر گرفته شده و با شاخص‌های مالی، رفتاری، دیجیتال، سودآوری و هزینه‌ای توسعه یافته است. هدف از این توسعه، کاهش محدودیت‌های RFM کلاسیک و ایجاد مبنایی جامع‌تر برای خوشه‌بندی مشتریان حقیقی بانک است.

¹ Yan

² Bult and Wansbeek

³ Gan

⁴ Pandey

الگوریتم MiniBatchKMeans و مزیت آن در خوشه‌بندی مقیاس‌پذیر

الگوریتم K-Means یکی از روش‌های رایج خوشه‌بندی در یادگیری بدون نظارت است که به دلیل سادگی، سرعت و تفسیرپذیری در تحلیل مشتریان کاربرد گسترده‌ای دارد. با این حال، در داده‌های حجیم، محاسبه فاصله همه نمونه‌ها از مراکز خوشه در هر تکرار می‌تواند هزینه محاسباتی بالایی ایجاد کند؛ مسئله‌ای که در تحلیل داده‌های گسترده بانکی اهمیت بیشتری دارد (سیناگا و یانگ، ۲۰۲۰؛ وانگ، ۲۰۲۵).

MiniBatchKMeans نسخه‌ای مقیاس‌پذیر از K-Means است که به جای پردازش کل داده‌ها در هر تکرار، از زیرمجموعه‌های کوچک و تصادفی (Mini-batch) برای به‌روزرسانی مراکز خوشه استفاده می‌کند. این سازوکار موجب کاهش زمان پردازش و مصرف حافظه می‌شود. اسکالی (۲۰۱۰) نیز نشان داد که این روش می‌تواند هزینه محاسباتی K-Means را در داده‌های بزرگ کاهش دهد، درحالی‌که کیفیت خوشه‌بندی در کاربردهای عملی حفظ می‌شود. مزیت MiniBatchKMeans در بانکداری، ایجاد تعادل میان مقیاس‌پذیری، سرعت و تفسیرپذیری است. مراکز خوشه همچنان قابل تحلیل‌اند و امکان توصیف ویژگی‌های غالب هر گروه را فراهم می‌کنند؛ موضوعی که برای تفسیر مدیریتی بخش‌های مشتریان اهمیت دارد (هیکس^۳ و همکاران، ۲۰۲۱). با این حال، این الگوریتم مانند K-Means به مقیاس متغیرها، مقادیر پرت و توزیع‌های چوله حساس است؛ بنابراین پاک‌سازی، کنترل مقادیر افراطی، تبدیل متغیرهای چوله و استانداردسازی پیش از خوشه‌بندی ضروری است. در پژوهش حاضر، MiniBatchKMeans به دلیل حجم بالای داده‌ها و نیاز به اجرای مکرر مدل برای ارزیابی تعداد و پایداری خوشه‌ها انتخاب شد. تعداد مناسب خوشه‌ها نیز با استفاده از معیارهایی مانند Davies-Bouldin، Silhouette و Calinski-Harabasz و شاخص‌های پایداری بررسی شد. از این رو، انتخاب این الگوریتم نه بر مبنای برتری مطلق نسبت به سایر روش‌ها، بلکه بر اساس تناسب آن با مقیاس داده، هزینه محاسباتی و نیاز به تفسیر مدیریتی انجام گرفته است.

مرور مطالعات پیشین

برای تبیین جایگاه پژوهش حاضر در ادبیات موجود، در جدول زیر مهم‌ترین مطالعات مرتبط با خوشه‌بندی مشتریان، مدل RFM، داده‌کاوی در بانکداری و روش‌های مقیاس‌پذیر خوشه‌بندی مرور شده‌اند. این مطالعات از نظر حوزه کاربرد، نوع داده، روش تحلیلی، شاخص‌های مورد استفاده، یافته‌های اصلی و شکاف‌های پژوهشی مقایسه شده‌اند. هدف از این مقایسه آن است که نشان داده شود اگرچه پژوهش‌های پیشین نقش مهمی در توسعه بخش‌بندی و خوشه‌بندی مشتریان داشته‌اند، اما همچنان نیاز به مدلی وجود دارد که شاخص‌های توسعه‌یافته RFM را با ابعاد مالی، رفتاری، دیجیتالی و ریسکی در محیط واقعی بانکداری ترکیب کرده و با استفاده از الگوریتمی مقیاس‌پذیر، خوشه‌هایی قابل تفسیر برای مشتریان حقیقی بانک ارائه دهد.

¹ Sinaga & Yang

² Wang

³ Hicks

جدول ۱. مرور مطالعات پیشین مرتبط با خوشه‌بندی مشتریان، مدل RFM و داده‌کاوی در بانکداری

ردیف	پژوهش	بستر بانکی	RFM	خوشه‌بندی	RFM توسعه‌یافته / ویژگی تکمیلی	داده حجیم / مقیاس‌پذیری	ابعاد ریسک / سود/هزینه	رفتار دیجیتال / کانال
۱	بولت و وانسیک، ۱۹۹۵	—	—	—	—	—	✓	—
۲	اسکالی، ۲۰۱۰	—	—	✓	—	✓	—	—
۳	داوود و همکاران، ۲۰۱۹	✓	—	✓	✓	—	—	—
۴	مانسیسیدور و همکاران، ۲۰۱۹	✓	—	✓	✓	—	✓	—
۵	علی‌اف و همکاران، ۲۰۲۰	✓	✓	✓	—	—	—	—
۶	موسایی‌راد، ۲۰۲۰	✓	—	✓	✓	—	—	—
۷	برخوردار و همکاران، ۲۰۲۱	✓	✓	✓	✓	—	—	—
۸	پلوتنیکووا و همکاران، ۲۰۲۳	✓	—	—	—	—	✓	—
۹	جان و همکاران، ۲۰۲۴	—	✓	✓	—	✓	—	—

۱۰	گان و همکاران، ۲۰۲۴	—	✓	✓	✓	—	—	—
۱۱	پاندی و همکاران، ۲۰۲۴	—	✓	✓	✓	—	—	✓
۱۲	یان و همکاران، ۲۰۲۵	✓	✓	✓	—	✓	—	—
۱۳	ویانا فیلو و همکاران، ۲۰۲۵	—	✓	✓	✓	—	—	—
۱۴	عباسی و ایمانی، ۱۳۹۶	—	✓	✓	✓	—	—	—
۱۵	غریب و همکاران، ۱۳۹۷	✓	✓	✓	✓	✓	—	✓
۱۶	یوسفی زاد و ثریایی، ۱۳۹۷	✓	✓	✓	✓	✓	—	—
۱۷	خاتمی فیروزآبادی و همکاران، ۱۳۹۸	✓	✓	✓	✓	✓	—	—
۱۸	تقوی فرد و همکاران، ۱۴۰۰	✓	✓	✓	✓	✓	—	✓
۱۹	زارعی و همکاران، ۱۴۰۴	✓	✓	✓	✓	✓	—	—
۲۰	مطالعه حاضر	✓*	✓	✓	✓	✓	✓	✓

جمع‌بندی پیشینه و شکاف پژوهشی

مرور مطالعات پیشین نشان می‌دهد که خوشه‌بندی مشتریان و استفاده از مدل RFM در حوزه‌های بازاریابی، خرده‌فروشی، بانکداری و خدمات مالی سابقه قابل توجهی دارد. بخش مهمی از پژوهش‌ها، از RFM کلاسیک برای شناسایی مشتریان ارزشمند، رتبه‌بندی مشتریان، محاسبه ارزش دوره عمر مشتری و طراحی راهبردهای بازاریابی استفاده کرده‌اند. در این مطالعات، شاخص‌های تازگی، فراوانی و ارزش مالی به‌عنوان مبنای اصلی تحلیل رفتار مشتریان به کار رفته‌اند و الگوریتم‌هایی مانند Fuzzy C-Means، DBSCAN، Means خوشه‌بندی سلسله‌مراتبی و روش‌های ترکیبی برای گروه‌بندی مشتریان استفاده شده‌اند. همچنین برخی پژوهش‌های جدیدتر تلاش کرده‌اند با افزودن ابعادی مانند Interval، Engagement، کارمزد، تسهیلات، معوقات یا الگوهای زمانی تراکنش، محدودیت‌های RFM کلاسیک را کاهش دهند. خلاصه مطالعات پیشین در جدول ۱ نشان می‌دهد که ادبیات موجود از نظر روش‌های خوشه‌بندی، توسعه RFM و کاربردهای بانکی تنوع قابل توجهی دارد. با وجود این، بررسی مطالعات نشان می‌دهد که بخش قابل توجهی از پژوهش‌های پیشین همچنان بر RFM کلاسیک یا نسخه‌های محدود توسعه‌یافته آن متمرکز بوده‌اند. در این مطالعات، مشتریان معمولاً براساس سه شاخص تازگی، فراوانی و ارزش مالی یا تعداد محدودی شاخص تکمیلی بخش‌بندی شده‌اند. درحالی‌که در محیط بانکداری، رفتار و ارزش مشتری فقط با مبلغ تراکنش یا تعداد تعاملات قابل تبیین نیست. مشتری بانکی ممکن است از نظر حجم تراکنش فعال باشد، اما به دلیل هزینه خدمت بالا، ریسک اعتباری زیاد یا سودآوری پایین برای بانک ارزش راهبردی محدودی داشته باشد. برعکس، ممکن است مشتری از نظر تراکنشی کم‌فعال باشد، اما به دلیل مانده سپرده، رفتار اعتباری مطلوب یا ظرفیت توسعه رابطه مالی، برای بانک ارزشمند تلقی شود. بنابراین، یکی از شکاف‌های اصلی ادبیات، کم‌توجهی به ترکیب همزمان شاخص‌های مالی، رفتاری، دیجیتالی، سودآوری، هزینه‌ای و ریسکی در قالب یک مدل خوشه‌بندی جامع برای مشتریان حقیقی بانک است.

از منظر روش‌شناختی نیز بیشتر مطالعات از الگوریتم‌های رایج خوشه‌بندی مانند Fuzzy C-Means، DBSCAN یا K-Means استفاده کرده‌اند. این الگوریتم‌ها از نظر کاربردی و تفسیری ارزشمندند، اما در داده‌های بسیار حجیم بانکی ممکن است با محدودیت‌هایی مانند افزایش زمان اجرا، حساسیت به مقیاس متغیرها، وابستگی به پارامترها یا دشواری استقرار عملیاتی مواجه شوند. برخی مطالعات جدید به روش‌های پیچیده‌تری مانند LSTM، Autoencoder، گراف، مدل آمیخته گاوسی یا جنگل تصادفی پرداخته‌اند؛ اما این روش‌ها نیز گاهی از نظر تفسیر مدیریتی یا پیاده‌سازی عملیاتی در بانک چالش‌برانگیز هستند. از این رو، هنوز نیاز به الگوریتمی وجود دارد که هم مقیاس‌پذیر باشد، هم قابلیت اجرا روی داده‌های بزرگ بانکی را داشته باشد و هم خروجی آن برای مدیران بانک قابل فهم و قابل استفاده باشد. مطالعات داخلی مرتبط با بانکداری نیز هرکدام بخشی از مسئله را پوشش داده‌اند. برخی پژوهش‌ها بر مشتریان دارای POS، تراکنش‌های شتابی، کارمزد، تخصیص خدمت، الگوهای پویای رفتار مشتری یا پیش‌بینی سطح اعتباری آینده تمرکز کرده‌اند. این مطالعات برای توسعه ادبیات بانکداری داده‌محور در ایران ارزشمندند، اما معمولاً یکی از سه محدودیت را دارند: یا برپایه RFM و تعداد محدودی شاخص بانکی متکی هستند، یا تمرکز اصلی آنها بر مرحله پس از خوشه‌بندی مانند تخصیص خدمت و پیش‌بینی اعتبار است، یا مسئله مقیاس‌پذیری و طراحی یک مدل خوشه‌بندی قابل استقرار برای داده‌های حجیم مشتریان حقیقی بانک را به‌صورت مستقیم دنبال نکرده‌اند. بنابراین، مقاله حاضر تلاش می‌کند میان این حوزه‌ها پیوند برقرار کند و تمرکز خود را به‌طور مشخص بر طراحی و ارزیابی یک مدل خوشه‌بندی هوشمند قرار دهد. بر این اساس، شکاف پژوهشی مطالعه حاضر را می‌توان در سه محور خلاصه کرد. نخست، در ادبیات موجود هنوز مدلی که RFM را به‌صورت نظام‌مند با شاخص‌های خاص بانکداری از جمله سودآوری، هزینه خدمت، ریسک، رفتار دیجیتال و تنوع کانال ترکیب کند، کمتر دیده می‌شود. دوم، بسیاری از مطالعات از الگوریتم‌های خوشه‌بندی رایج استفاده کرده‌اند، اما مسئله مقیاس‌پذیری در داده‌های بزرگ بانکی و قابلیت اجرای عملیاتی مدل کمتر مورد توجه قرار گرفته است. سوم، در برخی مطالعات، خروجی خوشه‌بندی صرفاً به‌عنوان نتیجه آماری گزارش شده و پیوند آن با تفسیر مدیریتی، نام‌گذاری خوشه‌ها و کاربرد در تصمیم‌گیری بانکی به اندازه کافی برجسته نشده است. در پاسخ به این شکاف‌ها، پژوهش حاضر با استفاده از شاخص‌های توسعه‌یافته RFM و الگوریتم MiniBatchKMeans، مدلی مقیاس‌پذیر و قابل تفسیر برای خوشه‌بندی مشتریان حقیقی بانک ارائه می‌کند. سهم نظری پژوهش، توسعه نگاه کلاسیک RFM به چارچوبی چندبعدی متناسب با منطق بانکداری است. سهم روش‌شناختی آن، به‌کارگیری الگوریتمی سریع‌تر و مناسب‌تر برای داده‌های حجیم بانکی و ارزیابی کیفیت خوشه‌ها با معیارهای چندگانه است. سهم کاربردی پژوهش نیز

ارائه خوشه‌هایی قابل تفسیر برای مدیران بانک است تا بتوانند از آن در شناخت دقیق‌تر مشتریان، طراحی راهبردهای مدیریت ارتباط با مشتری، اولویت‌بندی خدمات و تخصیص هدفمند منابع استفاده کنند.

روش‌شناسی پژوهش

نوع پژوهش و رویکرد کلی تحقیق

پژوهش حاضر در چارچوب پارادایم اثبات‌گرایی و با رویکردی کمی و داده‌محور انجام شده است. انتخاب این پارادایم با توجه به اتکای پژوهش بر داده‌های عینی و ثبتی مشتریان، سنجش کمی متغیرها و استفاده از روش‌های آماری و یادگیری ماشین برای شناسایی الگوهای موجود در داده‌ها صورت گرفته است. از نظر هدف، پژوهش کاربردی-توسعه‌ای است؛ زیرا علاوه بر تحلیل داده‌های واقعی مشتریان، در پی طراحی و ارزیابی مدلی قابل استفاده برای خوشه‌بندی مشتریان حقیقی بانک و پشتیبانی از تصمیم‌گیری مدیریتی است. از نظر استراتژی اجرا، پژوهش مبتنی بر داده‌کاوی و مدل‌سازی تحلیلی-توصیفی بوده و فرایند آن با الهام از منطق چارچوب‌های CRISP-DM و FIN-DM طراحی شده است؛ به گونه‌ای که ابتدا داده‌های موجود شناسایی و آماده‌سازی شده، سپس مهندسی ویژگی انجام گرفته و در ادامه، مدل خوشه‌بندی و سایر مراحل تحلیلی اجرا شده‌اند.

داده‌های مورد استفاده در پژوهش از نوع کمی، ثبتی، عملیاتی و ثانویه بوده و از سامانه‌های اطلاعاتی و پایگاه‌های داده یکی از بانک‌های خصوصی کشور استخراج شده‌اند. واحد تحلیل «مشتری حقیقی» است و اطلاعات هر مشتری پس از تجمیع در سطح مشتری، شامل داده‌های تراکنشی، مانده حساب، تسهیلات، درآمد، هزینه، سودآوری، رفتار دیجیتال و سایر شاخص‌های مورد نیاز پژوهش بوده است. داده‌ها مربوط به یک بازه زمانی یک‌ساله هستند و به دلیل استفاده از داده‌های واقعی سامانه‌های بانکی، جمع‌آوری داده‌ها از طریق پرسشنامه یا مصاحبه انجام نشده است.

برای تجزیه و تحلیل داده‌ها، ابتدا عملیات پاک‌سازی، کنترل مقادیر گمشده و پرت، تجمیع داده‌ها در سطح مشتری، مهندسی ویژگی، تبدیل متغیرهای چوله و استانداردسازی انجام شد. سپس الگوریتم MiniBatchKMeans به دلیل قابلیت مقیاس‌پذیری و تناسب آن با حجم بالای داده‌های بانکی برای خوشه‌بندی مشتریان به کار گرفته شد. به منظور تعیین تعداد مناسب خوشه‌ها، مقادیر مختلف k با استفاده از شاخص‌های سیلوئت، دیویس-بولدین، کالینسکی-هاراباز، رند تعدیل شده (ARI) و جاکارد و همچنین حداقل سهم هر خوشه ارزیابی شدند. علاوه بر این، پایداری خوشه‌ها با استفاده از چند بذر تصادفی مورد بررسی قرار گرفت تا انتخاب نهایی تعداد خوشه‌ها تنها بر یک معیار منفرد متکی نباشد و از نظر کیفیت آماری و پایداری قابل دفاع باشد.

جامعه آماری و منبع داده‌های پژوهش

جامعه آماری پژوهش حاضر شامل مشتریان حقیقی یکی از بانک‌های خصوصی کشور است که در بازه زمانی مورد بررسی، دارای سابقه فعالیت در سامانه‌های عملیاتی بانک بوده‌اند. با توجه به ملاحظات محرمانگی داده‌های مشتریان و الزامات امنیت اطلاعات در صنعت بانکداری، نام بانک در مقاله ذکر نمی‌شود و تحلیل‌ها به صورت ناشناس و در سطح داده‌های تجمیع شده انجام شده است. داده‌های پژوهش از سامانه‌های اطلاعاتی و پایگاه‌های داده بانک استخراج شده و پس از پردازش، پاک‌سازی و تجمیع، در محیط تحلیلی پژوهش مورد استفاده قرار گرفته‌اند. واحد تحلیل در این پژوهش، مشتری حقیقی بانک است؛ بنابراین، هر رکورد نهایی بیانگر پروفایل یک مشتری در دوره زمانی مطالعه است، نه یک تراکنش منفرد. به عبارت دیگر، داده‌های تراکنشی، مالی، رفتاری، دیجیتال و ریسکی ابتدا از سامانه‌های مختلف بانک استخراج شده و سپس در سطح مشتری تجمیع شده‌اند تا برای هر مشتری، مجموعه‌ای از شاخص‌های تحلیلی قابل استفاده در مدل خوشه‌بندی ساخته شود. این رویکرد با هدف پژوهش، یعنی خوشه‌بندی مشتریان براساس الگوهای رفتاری-مالی، سازگار است؛ زیرا هدف، شناخت وضعیت کلی مشتری در تعامل با بانک است، نه تحلیل یک تراکنش خاص.

داده‌های اولیه پژوهش در سطح مشتری از سامانه‌های عملیاتی بانک استخراج و در هشت حوزه اصلی شامل سپرده و مانده، تسهیلات و تعهدات، تراکنش‌های حضوری، تراکنش‌های دیجیتال، تراکنش‌های کارتی و پایانه‌ای، انتقال وجه و اسناد، سودآوری و هزینه و شاخص‌های RFM سازمان‌دهی شدند. پس از کنترل کیفیت، مهندسی ویژگی و بررسی معیارهایی نظیر میزان داده گمشده، فراوانی مقادیر صفر، چولگی، واریانس، همبستگی شدید، پایداری محاسباتی و قابلیت تفسیر مدیریتی، ۱۰ ویژگی نهایی برای خوشه‌بندی انتخاب شد: لگاریتم علامت‌دار ارزش مالی، سودآوری و میانگین مانده، تأخر، تناوب، نسبت دیجیتال، شدت دیجیتال، تنوع کانال، تنوع خدمات و حاشیه سود مقاوم.

در پالایش رکوردها، ۲۳ رکورد فاقد متغیرهای کلیدی از تحلیل حذف شدند و ۹۳۰۷ رکورد دارای امضای مالی همزمان صفر به جای تفسیر به عنوان یک بخش مشتری، در صف کنترل کیفیت قرار گرفته و از خوشه بندی کنار گذاشته شدند. همچنین متغیرهای دارای واریانس بسیار پایین، نسبت های نامعین یا ناپایدار و متغیرهای دارای افزونگی شدید بررسی و در صورت نداشتن ارزش تفکیکی از فضای نهایی حذف شدند. مقادیر پرت معتبر به صورت گسترده حذف نشدند و با تبدیل های مناسب و کران بندی صدک های ۰.۵ و ۹۹.۵ کنترل شدند.

منبع اصلی داده های پژوهش، داده های ثانویه و عملیاتی بانک است. داده های خام شامل اطلاعات تراکنشی، مانده ها، درآمد و هزینه، سودآوری، تعاملات دیجیتال، تنوع کانال و خدمت و متغیرهای ریسک بود. با این حال، ورود به مدل خوشه بندی منوط به کیفیت داده، واریانس کافی، نبود ناپایداری عددی و معنای کسب و کاری روشن بود. رخدادهای ریسک مثبت تنها در ۲۵۷ رکورد مشاهده شد؛ بنابراین متغیرهای ریسک برای جلوگیری از تحریف فاصله ها از خوشه بندی کنار گذاشته و به لایه مستقل بازبینی ریسک منتقل شدند. با توجه به استفاده از داده های واقعی مشتریان بانک، ملاحظات محرمانگی و حفاظت از حریم خصوصی در فرایند پژوهش مورد توجه قرار گرفت. داده ها در سطح مشتری و با استفاده از شناسه های پژوهشی پردازش شدند و اطلاعات هویتی مستقیم مشتریان در تحلیل ها و گزارش نتایج مورد استفاده قرار نگرفت. همچنین نام بانک به دلیل الزامات محرمانگی ذکر نشده و نتایج صرفاً به صورت تجمیعی گزارش شده اند. دسترسی به داده ها محدود به محیط پژوهش بوده و از انتشار شناسه واقعی مشتریان جلوگیری شده است. افزون بر این، خروجی مدل مبنای تصمیم اعتباری خودکار قرار نگرفته و موارد حساس، از جمله رخدادهای ریسک، نیازمند بازبینی انسانی هستند.

پیش پردازش داده ها و ساخت شاخص های توسعه یافته RFM

پس از استخراج داده های خام از سامانه های عملیاتی بانک، فرایند پیش پردازش با هدف افزایش کیفیت داده، کاهش خطاهای تحلیلی و آماده سازی داده ها برای اجرای الگوریتم خوشه بندی انجام شد. داده های اولیه شامل اطلاعات مالی، تراکنشی، رفتاری، دیجیتال و ریسکی مشتریان حقیقی بانک بود که پس از تجمیع در سطح مشتری مورد تحلیل قرار گرفت. با توجه به وجود مقادیر گمشده، مقادیر پرت، تفاوت شدید مقیاس متغیرها و چولگی متغیرهای مالی، کنترل کیفیت داده بخش مهمی از فرایند مدل سازی را تشکیل داد.

در گام نخست، رکوردهای فاقد متغیرهای اصلی، ناسازگاری های ساختاری و مقادیر نامعتبر شناسایی شدند. در این مرحله میان «مقدار گمشده» و «صفر رفتاری» تمایز قائل شد؛ زیرا برای مثال، صفر بودن تعداد تراکنش می تواند نشان دهنده عدم استفاده واقعی مشتری از یک خدمت باشد و نباید به عنوان داده گمشده تلقی شود. رکوردهای فاقد متغیرهای کلیدی از تحلیل کنار گذاشته شدند و در موارد قابل جایگزینی، مقدار میانه برای متغیرهای عددی مورد استفاده قرار گرفت. همچنین ۹۳۰۷ رکورد دارای امضای مالی همزمان صفر به عنوان مسئله کیفیت داده در صف کنترل قرار گرفتند و به عنوان یک گروه رفتاری وارد خوشه بندی و تفسیر مشتریان نشدند.

در گام دوم، ویژگی های نهایی بر مبنای منطق RFM توسعه یافته ساخته شدند. در کنار تأخر و تناوب، متغیرهایی مانند ارزش مالی، سودآوری، میانگین مانده، رفتار دیجیتال، تنوع کانال، تنوع خدمات و حاشیه سود مقاوم وارد فضای ویژگی شدند. برای متغیرهای مالی دارای مقادیر مثبت و منفی از تبدیل لگاریتمی علامت دار به صورت $\text{sign}(x) \times \log(1+|x|)$ استفاده شد تا ضمن کاهش چولگی، جهت سود یا زیان و مانده های منفی حفظ شود. شدت دیجیتال از طریق لگاریتم یک به علاوه مجموع مبلغ تراکنش های انجام شده از طریق کانال های دیجیتال محاسبه شد؛ به این صورت که برای هر مشتری، مجموع مبالغ تراکنش های دیجیتال محاسبه و سپس تبدیل لگاریتمی $\log(1+\text{مجموع مبلغ تراکنش های کانال های دیجیتال})$ بر آن اعمال شد.

در مورد ریسک، با توجه به اینکه تنها ۲۵۷ رکورد دارای رخداد مثبت بودند، این متغیر به دلیل کمیابی شدید و خطر ایجاد یک محور تقریباً ثابت، از فضای خوشه بندی کنار گذاشته شد. ریسک در مرحله پس از خوشه بندی به صورت وقوع رخداد و شدت مشروط آن در موارد مثبت مورد استفاده قرار گرفت و مشتریان دارای رخداد مثبت برای بازبینی انسانی علامت گذاری شدند. بنابراین، خوشه های نهایی ماهیت رفتاری-مالی-دیجیتال دارند و نباید به عنوان خوشه های ریسکی یا برآوردکننده احتمال نکول تفسیر شوند.

برای کنترل مقادیر پرت، حذف گسترده رکوردها انجام نشد؛ زیرا مقادیر افراطی در داده‌های بانکی ممکن است علاوه بر خطای ثبت، نشان‌دهنده مشتریان بزرگ و واقعی باشند. پس از اعمال تبدیل‌های مناسب، مقادیر هر متغیر با استفاده از کران‌های صدک ۵/۰ و ۵/۹۹ کنترل شدند. این کران‌ها فقط از داده‌های آموزشی برآورد و بدون تغییر بر داده‌های اعتبارسنجی و آزمون اعمال شدند تا از نشت اطلاعات جلوگیری شود. استفاده از تبدیل لگاریتمی نیز با هدف کاهش چولگی متغیرهای مالی و جلوگیری از غلبه مشتریان دارای مقادیر بسیار بزرگ بر فاصله‌های خوشه‌بندی انجام گرفت.

در نهایت، به دلیل مبتنی‌بودن MiniBatchKMeans بر فاصله اقلیدسی، ویژگی‌ها با روش StandardScaler استاندارد شدند تا متغیرهای دارای مقیاس بزرگ‌تر به‌صورت ناخواسته بر ساختار خوشه‌ها مسلط نشوند. میانه‌های جایگزینی، کران‌های صدکی و پارامترهای استانداردسازی صرفاً بر داده آموزشی برازش شدند. در مرحله نهایی، پیش‌پردازش و الگوریتم MiniBatchKMeans بر ۸۰ درصد داده‌های مدل‌سازی برازش شد و ۲۰ درصد داده‌ها تا زمان ارزیابی نهایی به‌صورت دست‌نخورده نگه داشته شدند. نتایج مهندسی ویژگی و شاخص‌های نهایی مدل

پس از پالایش اولیه، تنها ده ویژگی مقاوم و قابل ردیابی وارد خوشه‌بندی شدند: لگاریتم علامت‌دار ارزش مالی، سودآوری و میانگین مانده؛ لگاریتم تأخر و تناوب؛ نسبت دیجیتال؛ شدت دیجیتال؛ تنوع کانال؛ تنوع خدمات؛ و حاشیه سود مقاوم. این انتخاب، فضای خوشه‌بندی را از RFM کلاسیک به یک بازنمایی رفتاری-مالی توسعه داد، بدون آنکه متغیرهای بسیار کم‌رخداد یا نسبت‌های عددی ناپایدار وزن مصنوعی در فاصله‌ها ایجاد کنند.

تأخر، فاصله از آخرین تعامل معنادار مشتری با بانک را بازنمایی می‌کند و پس از کران صفر به‌صورت $\log(1+x)$ تبدیل شد. تناوب نیز شدت تکرار تعامل را نشان می‌دهد و با همان منطق تبدیل شد. این دو متغیر، هسته رفتاری RFM را در مدل نهایی حفظ می‌کنند.

بعد مالی با سه متغیر اصلی نمایش داده شد: ارزش مالی، سودآوری و میانگین مانده. هر سه متغیر به دلیل چولگی و امکان مقادیر منفی، در صورت لزوم با تبدیل لگاریتمی علامت‌دار وارد مدل شدند. این طراحی موجب می‌شود مشتریان بسیار بزرگ یا مقادیر افراطی نتوانند به‌صورت نامتناسب مراکز خوشه را جابه‌جا کنند و در عین حال علامت سود/زیان یا مانده منفی حفظ شود.

برای جلوگیری از نسبت‌های انفجاری، حاشیه سود مقاوم به‌جای نسبت‌های ساده تقسیم بر درآمد استفاده شد. این شاخص سود را نسبت به مجموع قدرمطلق درآمد، هزینه و سود با یک ثابت کوچک مقیاس می‌کند و در حالت درآمد صفر پایدار می‌ماند. متغیرهای ریسک در این مرحله وارد خوشه‌بندی نشدند؛ زیرا تنها ۲۵۷ رخداد مثبت وجود داشت و ورود یک محور تقریباً ثابت می‌توانست ساختار گروه‌ها را مخدوش کند. ریسک به‌صورت وقوع و شدت مشروط به وقوع در لایه پسینی کنترل شد.

رفتار دیجیتال با دو ویژگی نسبت دیجیتال و شدت دیجیتال و عمق رابطه با دو ویژگی تنوع کانال و تنوع خدمات سنجیده شد. نسبت دیجیتال سهم تعاملات دیجیتال از تعاملات دیجیتال و شعبه را در دامنه صفر تا یک نشان می‌دهد و شدت دیجیتال با تبدیل لگاریتمی مبلغ کانال‌های دیجیتال ساخته شد. در نهایت، ماتریس ده‌ویژگی نهایی پس از کنترل واریانس، متناهی بودن مقادیر و تفسیرپذیری، ورودی MiniBatchKMeans قرار گرفت.

جدول ۲. جمع‌بندی شاخص‌های نهایی مدل خوشه‌بندی مشتریان

بُعد/ویژگی تحلیلی	تعریف نهایی	هدف/کنترل	تفسیر مدیریتی
ارزش مالی	logعلامت‌دار ارزش مالی	کنترل چولگی و حفظ علامت	سنجش ظرفیت/ارزش مالی بدون سلطه مقادیر افراطی
سودآوری	logعلامت‌دار سودآوری	حفظ تمایز سود و زیان	تفکیک ارزش مالی از سودآوری واقعی
میانگین مانده	logعلامت‌دار میانگین مانده	کنترل مقیاس و حفظ مانده منفی	بازنمایی پایه مالی مشتری
تاخر	$\log(1+\text{Recency})$ پس از کران صفر	توصیف فاصله از آخرین تعامل	تمایز الگوی زمانی تعامل
تناوب	$\log(1+\text{Frequency})$ پس از کران صفر	کنترل چولگی شدت تعامل	تمایز مشتریان کم‌تعامل و پرتکرار
نسبت دیجیتال	تعداد دیجیتال ÷ مجموع دیجیتال و شعبه	دامنه صفر تا یک	سنجش سهم تعامل غیرحضوری
شدت دیجیتال	$\log(1 + \text{مجموع مبلغ تراکنش‌های کانال‌های دیجیتال})$	کنترل چولگی مبالغ دیجیتال	سنجش شدت استفاده از کانال دیجیتال
تنوع کانال و خدمت	دو ویژگی: تعداد کانال‌های فعال و تعداد خدمات فعال	نمایش عمق و تنوع رابطه	تفکیک مشتریان تک کاناله/چندکاناله و کم خدمت/چندخدمت
حاشیه سود مقاوم	$\text{سود} \div (1 + \text{درآمد} + \text{هزینه} + \text{سود})$	پرهیز از انفجار نسبت در درآمد صفر	مقایسه سودآوری نسبی با پایداری عددی

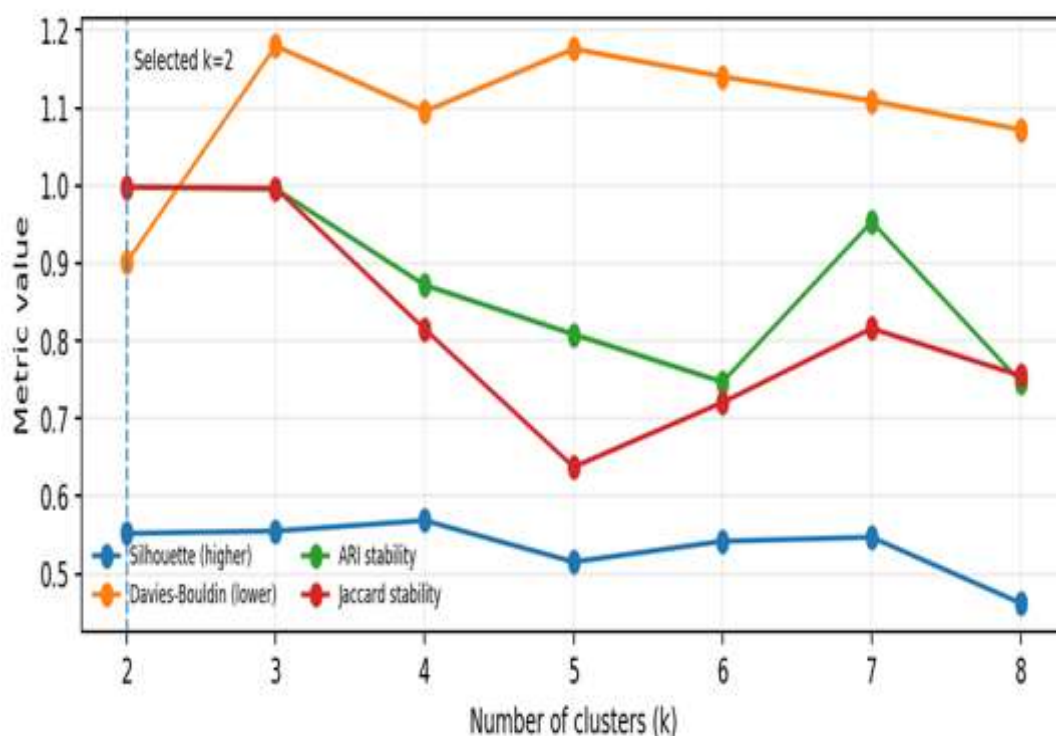
ارزیابی تعداد خوشه‌ها با معیارهای اعتبار درونی و پایداری چندبذری

برای تعیین تعداد مناسب خوشه‌ها، MiniBatchKMeans برای k های ۲ تا ۸ اجرا شد. علاوه بر Davies-Bouldin، Calinski-Harabasz، پایداری برچسب‌ها با ARI چندبذری و جاکارد هم‌تراز ارزیابی شد. پنج بذر ۱۱، ۲۹، ۴۲، ۷۱ و ۱۰۱ در آزمون پایداری به کار رفت. قاعده انتخاب فقط بیشینه یا کمینه یک معیار نبود؛ گزینه نهایی باید هم کیفیت درونی مناسب، هم پایداری قابل قبول، هم حداقل سهم گروه بیش از آستانه ازپیش‌تعریف‌شده و هم قابلیت تفسیر بانکی داشته باشد. نتایج نشان داد $k=2$ کمترین مجموع رتبه را میان گزینه‌های واجد شرایط به دست آورد. این ساختار Silhouette برابر ۰٫۵۵۱۹، Davies-Bouldin برابر ۰٫۹۰۰۹ و Calinski-Harabasz برابر ۱۱۹۹۹٫۱۳ داشت و از نظر پایداری نیز میانگین ARI برابر ۰٫۹۹۷۶ و جاکارد برابر ۰٫۹۹۸۶ ثبت شد. سهم کوچک‌ترین گروه در ارزیابی ۳۳٫۰۹ درصد بود؛ بنابراین ساختار نهایی نه به یک بذر تصادفی وابسته بود و نه خوشه بسیار کوچک و مصنوعی ایجاد می‌کرد.

جدول ۳. نتایج اعتبارسنجی تعداد خوشه‌ها

تعداد خوشه‌ها (k)	ضریب سیلوئت	شاخص دیویس-بولدین	شاخص کالینسکی-هاراباز	شاخص رند تعدیل شده (ARI)	ضریب جاکارد	حداقل سهم کوچک ترین خوشه	وضعیت
۲	۰,۵۵۱۹	۰,۹۰۰۹	۱۱۹۹۹,۱	۰,۹۹۷۶	۰,۹۹۸۶	٪۳۳,۰۹	منتخب
۳	۰,۵۵۵۰	۱,۱۷۹۸	۹۸۳۸,۸	۰,۹۹۵۸	۰,۹۹۶۷	٪۱۷,۵۹	واجد شرایط
۴	۰,۵۶۸۵	۱,۰۹۵۵	۸۷۶۹,۷	۰,۸۷۱۸	۰,۸۱۵۳	٪۸,۳۳	واجد شرایط
۵	۰,۵۱۵۲	۱,۱۷۶۱	۷۴۴۵,۷	۰,۸۰۸۴	۰,۶۳۷۵	٪۵,۳۷	رد به دلیل پایداری ناکافی
۶	۰,۵۴۱۹	۱,۱۴۰۴	۷۶۹۱,۲	۰,۷۴۷۱	۰,۷۲۰۸	٪۷,۱۲	رد به دلیل پایداری ناکافی
۷	۰,۵۴۷۰	۱,۱۰۹۳	۷۵۱۷,۵	۰,۹۵۳۴	۰,۸۱۵۹	٪۵,۱۵	واجد شرایط
۸	۰,۴۶۳۱	۱,۰۷۳۰	۷۱۷۸,۲	۰,۷۴۶۹	۰,۷۵۵۲	٪۳,۹۳	رد به دلیل پایداری ناکافی

بر پایه قاعده چندمعیاره، ساختار دوگروهی به عنوان مدل نهایی انتخاب شد. برخی k های بزرگ تر در یک معیار درونی عملکرد نزدیک یا حتی بهتری داشتند، اما پایداری یا حداقل اندازه گروه در همه آنها به اندازه ساختار دوگروهی مناسب نبود. بنابراین انتخاب $k=2$ بر مبنای تعادل میان کیفیت درونی، پایداری تکرار و پرهیز از گروه‌های مصنوعی انجام شد، نه صرفاً سادگی مدیریتی یا یک شاخص منفرد.



شکل ۱. جمع‌بندی معیارهای اعتبار درونی و پایداری برای k های ۲ تا ۸

انتخاب مدل نهایی خوشه‌بندی

پس از ارزیابی k های ۲ تا ۸، مدل نهایی MiniBatchKMeans با $k=2$ انتخاب شد. معیارهای پایداری چندبذری نشان دادند این ساختار در تکرارها بسیار پایدار است و سهم کوچک‌ترین گروه نیز در محدوده قابل قبول قرار دارد. این تصمیم با نسخه نهایی رساله سازگار است و ساختار دوگروهی نسخه‌های قبلی به دلیل ناپایداری نسبی و تشکیل گروه‌های کم‌حجم از استنباط نهایی کنار گذاشته شد.

مدل دوگروهی امکان تفسیر مقاوم‌تری از تفاوت‌های اصلی جامعه مشتریان فراهم کرد. گروه نخست عمدتاً کم‌دیجیتال‌تر و دارای ارزش مالی پایین‌تر بود، در حالی که گروه دوم فعالیت دیجیتال و ارزش مالی بالاتری داشت. با این حال، ارزش مالی بالاتر گروه دوم به معنای سودآوری قطعی بیشتر نیست؛ زیرا حاشیه سود مقاوم این گروه ضعیف‌تر بود. همچنین ریسک به دلیل خروج از فضای خوشه‌بندی، نباید برای نام‌گذاری هیچ گروهی مستقیماً از برچسب خوشه استنباط شود.

خروجی نهایی خوشه‌بندی شامل دو گروه آماری برای ۳۷۸،۳۱۲ مشتری مدلی‌پذیر بود: ۲۵۲،۴۷۴ مشتری (۶۶،۷۴ درصد) در گروه اول و ۱۲۵،۸۳۸ مشتری (۳۳،۲۶ درصد) در گروه دوم. ۹،۳۰۷ رکورد با پایه مالی همزمان صفر خارج از خوشه‌بندی و در صف کنترل کیفیت باقی ماندند. بنابراین این رکوردها نه «خوشه خاص» و نه «خوشه پرریسک» تلقی شدند.

جدول ۴. خروجی نهایی مدل خوشه‌بندی مشتریان

جهت‌گیری مدیریتی اولیه	ویژگی غالب	نام تفسیری موقت	گروه/وضعیت
فعال‌سازی کم‌هزینه و آزمون محدود دیجیتال	سهم ۶۶٫۷۴٪؛ نسبت دیجیتال و ارزش مالی پایین تر؛ حاشیه سود مقاوم بهتر	کم‌دیجیتال تر و کم‌ارزش تر	۱
بازبینی سودآوری و پیشنهاد مشروط؛ ارزش بالا $\neq$ سود قطعی	سهم ۳۳٫۲۶٪؛ تناوب و نسبت دیجیتال بالاتر؛ حاشیه سود مقاوم ضعیف تر	دیجیتال تر و با ارزش مالی بالاتر	۲
عدم پیشنهاد تا تأیید منشأ داده توسط مالک داده	۹۰۳۰۷ رکورد با پایه مالی هم‌زمان صفر	گروه مشتری نیست	صف کنترل کیفیت

پروفایل‌سازی و تفسیر خوشه‌های مشتریان

پس از انتخاب ساختار دوگروهی، پروفایل هر گروه بر پایه شاخص‌های مقاوم و مقادیر میانه تفسیر شد تا برجسته‌های عددی به توصیف‌هایی روشن و قابل‌دفاع تبدیل شوند. گروه اول ۷۴/۶۶ درصد مشتریان مدل‌پذیر را شامل می‌شود و میانه تأخر آن ۲۴، تناوب ۰، نسبت دیجیتال ۰/۰ و ارزش مالی ۱۳۱،۸۲۰ است؛ بنابراین این گروه «کم‌دیجیتال تر و کم‌ارزش تر» توصیف شد. بالین حال، میانه حاشیه سود مقاوم این گروه ۰٫۵۰۰ است که در مقایسه با گروه دوم وضعیت بهتری را نشان می‌دهد. گروه دوم ۳۳٫۲۶ درصد جامعه تحلیلی را شامل می‌شود و میانه تناوب ۳۱، نسبت دیجیتال ۱،۰۰۰ و ارزش مالی ۵۰،۵۶۰،۵۱۷ دارد؛ بنابراین از نظر فعالیت دیجیتال و ارزش مالی بالاتر است. با این حال، میانه سودآوری آن ۱،۱۳۴،۰۵۳- و حاشیه سود مقاوم ۰٫۲۵۱- است. این نتیجه نشان می‌دهد «ارزش مالی بالا «نباید با» سودآوری قطعی بالاتر «یکسان تلقی شود. رکوردهای پایه مالی صفر نیز جدا از این دو گروه و در صف کنترل کیفیت باقی ماندند.

جدول ۵. پروفایل‌سازی و تفسیر خوشه‌های نهایی مشتریان

گروه/وضعیت	سهم/تعداد	و تأخر تناوب (میانه)	ارزش مالی/سودآوری	رفتار دیجیتال	حاشیه سود مقاوم	برداشت مجاز
۱	۲۵۲،۴۷۴ ۶۶٫۷۴٪	تأخر ۲۴؛ تناوب ۰	ارزش مالی ۱۳۱،۸۲۰	نسبت دیجیتال ۰٫۰۰۰	۰٫۵۰۰	کم‌دیجیتال تر و کم ارزش تر؛ نام‌گذاری موقت
۲	۱۲۵،۸۳۸ ۳۳٫۲۶٪	تأخر ۱۹؛ تناوب ۳۱	ارزش مالی ۵۰،۵۶۰،۵۱۷ سودآوری میانه ۱،۱۳۴،۰۵۳-	نسبت دیجیتال ۱،۰۰۰	-۰٫۲۵۱	دیجیتال تر و با ارزش مالی بالاتر؛ ارزش بالا معادل سود قطعی نیست
صف کیفیت	۹،۳۰۷	—	پایه مالی هم زمان صفر	—	—	گروه مشتری نیست؛ ارجاع به مالک داده

گروه اول را می‌توان به‌طور موقت «کم‌دیجیتال‌تر و کم‌ارزش‌تر» نامید. سیاست مدیریتی اولیه برای این گروه باید بر فعال‌سازی کم‌هزینه، آزمون محدود خدمات دیجیتال و شناسایی نشانه ظرفیت تمرکز کند؛ زیرا اندازه بالای گروه به‌تنهایی دلیل تخصیص منابع گران نیست و نام‌گذاری نهایی باید با تأیید خبره و مالک داده انجام شود.

گروه دوم به‌طور موقت «دیجیتال‌تر و با ارزش مالی بالاتر» نام‌گذاری می‌شود. این گروه برای پیشنهادهای انتخابی و حفظ رابطه ظرفیت بیشتری نشان می‌دهد، اما حاشیه سود مقاوم ضعیف‌تر آن ایجاب می‌کند سودآوری و هزینه خدمت پیش از هر تخصیص پرهزینه بازبینی شود. همچنین هر رخداد ریسک مثبت مستقل از عضویت گروه باید به بازبینی انسانی ارجاع شود.

رکوردهای دارای پایه مالی هم‌زمان صفر به‌عنوان یک گروه واقعی از مشتریان تفسیر نشدند و تا زمان تأیید منشأ داده، از فرایند پیشنهاد خدمت کنار گذاشته شدند. این تفکیک میان «گروه رفتاری واقعی» و «صف کنترل کیفیت»، از اصلاحات مهم نسخه نهایی مدل است و مانع از آن می‌شود که مسئله کیفیت داده به‌اشتباه به‌عنوان یک خوشه مدیریتی تعبیر شود.

بحث و نتیجه‌گیری

هدف پژوهش حاضر، طراحی و ارزیابی یک مدل مقیاس‌پذیر برای خوشه‌بندی مشتریان حقیقی بانک بر پایه RFM توسعه‌یافته و الگوریتم MiniBatchKMeans بود. نتایج نهایی وجود دو الگوی عمده و پایدار در میان مشتریان مدل‌پذیر را نشان داد؛ به‌گونه‌ای که ناهمگنی مشتریان را می‌توان با ساختاری ساده، پایدار و قابل‌تفسیر بازنمایی کرد، مشروط بر آنکه کیفیت داده و حدود استنباط مدل رعایت شود. در نسخه نهایی، توسعه RFM به معنای افزودن همه متغیرهای در دسترس نبود، بلکه تنها ۱۰ ویژگی مقاوم مالی، رفتاری و دیجیتال وارد خوشه‌بندی شدند و متغیر ریسک به دلیل کمیابی، نسبت‌های ناپایدار و رکوردهای دارای پایه مالی نامطمئن از فضای خوشه‌بندی کنار گذاشته شدند.

از منظر مدیریتی، گروه اول با سهم ۷۴/۶۶ درصد، فعالیت دیجیتال و ارزش مالی پایین‌تری دارد و برای فعال‌سازی کم‌هزینه و آزمون‌های محدود مناسب‌تر است. گروه دوم با سهم ۲۶/۳۳ درصد، دیجیتال‌تر و از نظر ارزش مالی بالاتر است، اما حاشیه سود مقاوم ضعیف‌تری دارد؛ بنابراین ارائه خدمات پرهزینه به این گروه باید با بازبینی سودآوری و هزینه همراه باشد. رخداد ریسک و کیفیت داده نیز مستقل از عضویت خوشه بررسی می‌شوند و نباید در نام‌گذاری خوشه‌ها ادغام شوند.

یافته‌ها با مطالعات پیشین درباره RFM و بخش‌بندی مشتریان همسو است. بولت و وانسیک (۱۹۹۵) بر اهمیت ارزش اقتصادی مشتری در هدف‌گذاری تأکید کردند و پژوهش حاضر این منطق را با خوشه‌بندی مشتریان بر اساس ویژگی‌های مالی، رفتاری و دیجیتال در محیط بانکداری دنبال می‌کند. از نظر روش‌شناختی نیز با اسکالی (۲۰۱۰) همسو است؛ زیرا Mini-batch K-Means امکان کاهش هزینه محاسباتی در داده‌های بزرگ را فراهم می‌کند. در پژوهش حاضر، این مزیت با پیش‌پردازش بدون نشت، کنترل کیفیت داده و ارزیابی پایداری چندبذری ترکیب شده است.

در مقایسه با مطالعات بین‌المللی، داوود و همکاران (۲۰۱۹) نقش روش‌های یادگیری ماشین را در پروفایل‌سازی مشتریان بانکی نشان دادند، در حالی که پژوهش حاضر بیشتر بر کیفیت ویژگی‌ها، کنترل داده و پایداری خوشه‌ها متمرکز است. برخلاف مانسیسیدور و همکاران (۲۰۱۹) که از خودرمزگذار متغیر (Variational Autoencoder) استفاده کردند، رویکرد حاضر ساده‌تر و از نظر مدیریتی قابل‌تفسیرتر است. همچنین علی‌اف و همکاران (۲۰۲۰) ترکیب RFM و خوشه‌بندی را در بانکداری به کار گرفتند، اما پژوهش حاضر RFM را با سودآوری، میانگین مانده، رفتار دیجیتال، تنوع کانال و خدمت و حاشیه سود مقاوم توسعه داده و ریسک کم‌رخداد را خارج از فضای خوشه‌بندی نگه داشته است.

یافته‌ها با تأکید موسایی‌راد (۲۰۲۰) بر اهمیت مهندسی ویژگی و نیز پژوهش برخورداری و همکاران (۲۰۲۱) درباره نقش نمایش مناسب رفتار مشتری همخوان است؛ با این تفاوت که مطالعه حاضر بر سادگی عملیاتی، قابلیت تمیزی و استفاده از ۱۰ ویژگی مقاوم تأکید دارد. همچنین با پلوتنیکو و همکاران (۲۰۲۳) همسو است که پروژه‌های داده‌کاوی مالی را فراتر از اجرای صرف الگوریتم دانسته و بر کیفیت و حاکمیت داده تأکید کرده‌اند. در همین مسیر، مطالعات جان و همکاران (۲۰۲۴)، گان و همکاران (۲۰۲۴) و پاندی و همکاران (۲۰۲۴) نیز توسعه RFM را دنبال کرده‌اند، اما پژوهش حاضر به‌جای افزودن یک بعد منفرد، مجموعه‌ای از ویژگی‌های مالی و دیجیتال مقاوم را به کار گرفته است.

در مقایسه با یان و همکاران (۲۰۲۵) و ویانا فیلو و همکاران (۲۰۲۵)، پژوهش حاضر از نظر توجه به مقیاس‌پذیری همسو است، اما راهکاری ساده‌تر و قابل‌تفسیرتر با MiniBatchKMeans و ارزیابی پایداری چندبذری ارائه می‌کند. در مطالعات داخلی نیز

غریب و همکاران (۱۳۹۷) بر تغییرات زمانی مشتریان تمرکز داشتند، در حالی که پژوهش حاضر بر غنای شاخص‌ها و مقیاس‌پذیری مدل متمرکز است. یوسفی‌زاد و ثریایی (۱۳۹۷) نیز کارایی الگوریتم‌های مرکز‌گرا را در خوشه‌بندی مشتریان بانک نشان دادند که با انتخاب MiniBatchKMeans در مطالعه حاضر همسو است.

مطالعه خاتمی فیروزآبادی و همکاران (۱۳۹۸) خوشه‌بندی را به تخصیص خدمت پیوند داد و تقوی‌فرد و همکاران (۱۴۰۰) نیز RFM را در بستر بانکداری ایران توسعه دادند. پژوهش حاضر این مسیر را با تأکید بیشتر بر کنترل کیفیت داده، ویژگی‌های مالی و دیجیتال، حاشیه سود مقاوم و مقیاس‌پذیری ادامه می‌دهد. همچنین در مقایسه با زارعی و همکاران (۱۴۰۴) که بر پیش‌بینی رفتار و اعتبارسنجی آینده تمرکز داشتند، هدف این پژوهش پیش‌بینی آینده نیست، بلکه شناسایی و تفسیر گروه‌های فعلی مشتریان و پشتیبانی از تصمیم‌گیری مدیریتی است.

در مجموع، سهم پژوهش حاضر را می‌توان در سه محور خلاصه کرد: توسعه عملیاتی RFM با ویژگی‌های بانکی مقاوم، استفاده از MiniBatchKMeans برای خوشه‌بندی مقیاس‌پذیر و پایدار، و تبدیل نتایج خوشه‌بندی به گروه‌هایی قابل تفسیر و قابل استفاده در تصمیم‌گیری بانکی. بنابراین، سهم اصلی پژوهش بیش از آنکه صرفاً افزودن متغیرهای جدید باشد، در تلفیق مهندسی ویژگی، مقیاس‌پذیری، کنترل کیفیت و تفسیر مدیریتی در یک چارچوب منسجم قرار دارد.

از منظر کاربرد مدیریتی، نتایج خوشه‌بندی می‌تواند به اقدامات مشخص برای هر گروه منجر شود. برای مشتریان گروه اول که فعالیت دیجیتال و ارزش مالی پایین‌تری دارند، اما از حاشیه سود مقاوم بهتری برخوردارند، اقداماتی مانند آموزش و تشویق استفاده از کانال‌های دیجیتال، ارائه پیشنهادهای کم‌هزینه و آزمایشی، فعال‌سازی تدریجی خدمات غیرحضوری و پایش تغییر در سطح تعامل پیشنهاد می‌شود. در مقابل، برای مشتریان گروه دوم که فعالیت دیجیتال و ارزش مالی بالاتری دارند اما حاشیه سود مقاوم آنان ضعیف‌تر است، پیشنهاد می‌شود پیش از ارائه خدمات پرهزینه، سودآوری رابطه با مشتری بازبینی و پیشنهادهای به‌صورت انتخابی و مشروط طراحی شوند؛ برای مثال، تمرکز بر خدمات دیجیتال، فروش متقاطع محصولات متناسب و پیشنهادهایی که هزینه خدمت را افزایش ندهند. ارزش مالی بالاتر این گروه نباید به‌تنهایی معادل سودآوری بیشتر تلقی شود. همچنین در هر دو گروه، وجود رخداد ریسک باید مستقل از خوشه‌بندی بررسی شود و مشتریان دارای رخداد مثبت ریسک پیش از هر پیشنهاد اعتباری به بازبینی انسانی ارجاع شوند. رکوردهای قرارگرفته در صف کنترل کیفیت نیز تا زمان تأیید منشأ داده، نباید مبنای اقدام خودکار قرار گیرند.

محدودیت‌های پژوهش

یکی از محدودیت‌های پژوهش، عدم اجرای مقایسه تجربی هم‌سناریو میان MiniBatchKMeans و الگوریتم‌هایی مانند K-Means استاندارد، DBSCAN و خوشه‌بندی سلسله‌مراتبی بر روی کل مجموعه داده است؛ بنابراین انتخاب الگوریتم بر مبنای تناسب روش‌شناختی، مقیاس‌پذیری، تفسیرپذیری و پایداری داخلی مدل انجام شده است. مقایسه تجربی نظام‌مند این روش‌ها می‌تواند در پژوهش‌های آینده انجام شود. با وجود تلاش برای طراحی و ارزیابی یک مدل مقیاس‌پذیر و قابل تفسیر برای خوشه‌بندی مشتریان حقیقی بانک، پژوهش حاضر با محدودیت‌هایی همراه است. نخست، داده‌های مورد استفاده مربوط به یک بانک مشخص است؛ از این رو تعمیم نتایج به سایر بانک‌ها باید با احتیاط انجام شود. همچنین برخی متغیرهای کیفی نظیر رضایت مشتری، اعتماد، ترجیحات فردی و تجربه خدمت در داده‌های موجود در دسترس نبودند.

همچنین، تحلیل حساسیت مستقل نسبت به ابرپارامترهای اجرایی MiniBatchKMeans، از جمله اندازه mini-batch و تعداد تکرارها، در پژوهش حاضر انجام نشده است. ارزیابی پایداری مدل صرفاً از طریق اجرای چندبذری و شاخص‌های ARI و جاکارد صورت گرفته است؛ بنابراین بررسی نظام‌مند حساسیت مدل نسبت به این پارامترها می‌تواند در پژوهش‌های آینده انجام شود.

از سوی دیگر، کیفیت خروجی مدل به کیفیت داده‌های ثبتی بانک وابسته است و خطاهای ثبت، داده‌های ناقص یا تغییر در سامانه‌های اطلاعاتی می‌تواند بر نتایج اثرگذار باشد. علاوه بر این، تحلیل SHAP در این پژوهش صرفاً برای تبیین تصمیم مدل استفاده شده و نباید به‌عنوان اثبات رابطه علی میان متغیرها تفسیر شود. رفتار مشتریان نیز ماهیتی پویا دارد؛ بنابراین ساختار خوشه‌ها ممکن است در طول زمان تغییر کند و مدل نیازمند بازآموزی دوره‌ای باشد. در نهایت، داده واقعی مربوط به پذیرش یا رد پیشنهادهای خدماتی در دسترس نبوده و نرخ‌های پذیرش، نکول و برخی مفروضات مالی در بخش پایلوت ماهیت سناریویی داشته‌اند؛ در نتیجه، سودهای گزارش‌شده در این بخش باید «سود مورد انتظار» تلقی شوند و نه سود تحقق‌یافته. با توجه به اینکه داده‌های پژوهش در یک بازه زمانی یک‌ساله تجمیع و تحلیل شده‌اند، پایداری ساختار خوشه‌ها در زیردوره‌های زمانی مانند فصل‌ها

یا ماه‌های مختلف به صورت مستقل بررسی نشده است. از آنجاکه رفتار مالی و دیجیتال مشتریان می‌تواند در طول زمان تغییر کند، تعمیم ساختار خوشه‌های شناسایی شده به دوره‌های بعد نیازمند پایش و بازآموزی دوره‌ای مدل است.

پیشنهادهایی برای پژوهش‌های آینده

پژوهش‌های آینده می‌توانند با استفاده از داده‌های طولی و چنددوره‌ای، تغییر رفتار مشتریان و جابه‌جایی آنان میان خوشه‌ها را بررسی کنند. همچنین در صورت دسترسی به داده‌های واقعی مربوط به پیشنهادهای ارائه شده و پذیرش یا رد آنها، امکان توسعه بخش پیشنهاددهنده از یک چارچوب قاعده‌محور به مدل‌های یادگیری نظارت شده یا یادگیری تقویتی فراهم خواهد شد. استفاده از آزمایش‌های کنترل شده و روش‌های استنباط علی نیز می‌تواند به بررسی اثر واقعی پیشنهادهای بانکی بر سودآوری، نگهداشت یا کاهش ریسک کمک کند. مقایسه MiniBatchKMeans با الگوریتم‌های پیشرفته‌تر نظیر LightGBM ، CatBoost ، AutoML و مدل‌های گرافی، اجرای مدل بر داده‌های چند بانک برای افزایش اعتبار بیرونی، و تلفیق داده‌های متنی مانند شکایات، تماس‌های مرکز ارتباط و پیام‌های دیجیتال از دیگر مسیرهای پیشنهادی است. افزون بر این، توسعه داشبوردهای مدیریتی و طراحی چارچوب‌های جامع‌تر برای پایش سوگیری، مستندسازی و حسابرسی مدل‌های هوش مصنوعی می‌تواند زمینه استقرار عملی‌تر چنین مدل‌هایی را در صنعت بانکداری فراهم کند.

منابع:

- ایمانی، ع.، و عباسی، م. (۱۳۹۶). خوشه‌بندی مشتریان بر مبنای مدل RFM با استفاده از الگوریتم C-means فازی: مورد مطالعه فروشگاه زنجیره‌ای رفاه شهر زاهدان. پژوهش‌های مدیریت عمومی، ۱۰۳۷، ۲۵۱-۳۶۸. <https://doi.org/10.22111/jmr.2017.36862>
- تقوی‌فرد، م. ت.، خانی، ا. م.، و لطفی، ف. (۱۴۰۰). خوشه‌بندی مشتریان در حوزه بانکداری الکترونیک با بهره‌گیری از تراکنش‌های الکترونیکی و اطلاعات دموگرافیک: مورد مطالعه بانک رفاه. مدیریت تبلیغات و فروش، ۲۳، ۴۸-۶۴. <https://doi.org/10.52547/JABM.2.3.48>
- خاتمی فیروزآبادی، س. م.، ع.، تقوی‌فرد، م. ت.، سجادی، س. خ. ا.، و بامدادصوفی، ج. (۱۳۹۸). ارائه مدل بهینه‌سازی چندهدفه تخصیص خدمت به مشتریان بانک به کمک داده‌کاوی و شبیه‌سازی. پژوهش در مدیریت تولید و عملیات، ۱۰۲، ۱۶۱-۱۸۰. <https://doi.org/10.22108/jpom.2019.107082.1082>
- زارعی، س.، مهدیان، ع.، مصلاح، ز.، و صحبت‌زاده، ر. (۱۴۰۴). اعتبارسنجی هوشمند و پیش‌بینی رفتار آینده مشتری بانک: رویکرد مبتنی بر داده با مدل RFM و خوشه‌بندی. فصل‌نامه مطالعاتی در مدیریت بانکی و بانکداری اسلامی، ۱۱ (تابستان)، ۳۱-۷۹. <https://doi.org/10.22034/jifb.2025.544969.1635>
- غریب، ا.، طلوعی، ع.، حیدرزاده، ک.، و رادفر، ر. (۱۳۹۷). شناسایی الگوهای غالب رفتار مشتریان بانک در طول زمان بخش‌های مختلف: مطالعه موردی بانک انصار. تحقیقات بازاریابی نوین، ۸۴، ۱۴۹-۱۶۵. <https://doi.org/10.22108/nmrj.2019.109137.1575>
- یوسفی‌زاد، ا.، و ثریایی، ع. (۱۳۹۷). بررسی و خوشه‌بندی مشتریان براساس مدل RFM و طراحی الگویی برای ارائه خدمات به مشتریان کلیدی. پژوهشنامه مدیریت اجرایی، ۱۰۲۰، ۱۷۵-۱۹۸. <https://doi.org/10.22080/jem.2019.15814.2834>
- Aliyev, M., Ahmadov, E., Gadirli, H., Mammadova, A., & Alasgarov, E. (2020). Segmenting bank customers via RFM model and unsupervised machine learning. arXiv (Cornell University). <https://doi.org/10.48550/arxiv.2008.08662>
- Barkhordar, E., Shirali-Shahreza, M. H., & Sadeghi, H. R. (2021). Clustering of Bank Customers using LSTM-based encoder-decoder and Dynamic Time Warping. arXiv (Cornell University). <https://doi.org/10.48550/arxiv.2110.11769>
- Barroso, M., & Laborda, J. (2022). Digital transformation and the emergence of the Fintech sector: Systematic literature review. Digital Business, 2(2), 100028. <https://doi.org/10.1016/j.digbus.2022.100028>
- Bult, J. R., & Wansbeek, T. (1995). Optimal selection for direct mail. Marketing Science, 14(4), 378-394. <https://doi.org/10.1287/mksc.14.4.378>
- Carvalho, H., & Maciel, F. (2026). RFM-Based Customer Segmentation: A Pedagogical case study for Marketing Analytics Education. Journal for Advancement of Marketing Education, 34(1). <https://doi.org/10.63963/001c.157562>
- Dawood, E. a. E., Elfakhrany, E., & Maghraby, F. A. (2019). Improve profiling bank customer's behavior using machine learning. IEEE Access, 7, 109320-109327. <https://doi.org/10.1109/access.2019.2934644>
- Duan, Y., Edwards, J. S., & Dwivedi, Y. K. (2019). Artificial intelligence for decision making in the era of Big Data – evolution, challenges and research agenda. International Journal of Information Management, 48, 63-71. <https://doi.org/10.1016/j.ijinfomgt.2019.01.021>
- Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Duan, Y., Dwivedi, R., Edwards, J., Eirug, A., Galanos, V., Ilavarasan, P. V., Janssen, M., Jones, P., Kar, A. K., Kizgin, H., Kronemann, B., Lal, B., Lucini, B., . . . Williams, M. D. (2019). Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. International Journal of Information Management, 57, 101994. <https://doi.org/10.1016/j.ijinfomgt.2019.08.002>
- Falcón, A. W. C., Sánchez, R. V., Periche, I. C. Z., & Rumiche, V. L. R. (2025). The influence of security perception on banking profitability: A systematic study based on digital customer segmentation. Multidisciplinary Reviews, 9(7), 2026358. <https://doi.org/10.31893/multirev.2026358>
- Filho, A. L. C. V., Leonardo, D. L., & Kleina, M. (2025). A graph-based approach to customer segmentation using the RFM model. arXiv (Cornell University). <https://doi.org/10.48550/arxiv.2505.08136>
- Gan, W., Zhou, P., Wan, S., Zeng, J., & Qi, Z. (2024). RFMI-based Customer Segmentation with K-means. 2024 IEEE International Conference on Big Data (BigData), Washington, DC, USA, 8286-8295. <https://doi.org/10.1109/bigdata62323.2024.10825251>
- Gerling, C., & Lessmann, S. (2024). Leveraging AI and NLP for Bank Marketing: A Systematic Review and Gap analysis. arXiv (Cornell University). <https://doi.org/10.48550/arxiv.2411.14463>

- Gul, R., & Al-Faryan, M. a. S. (2023). From insights to impact: leveraging data analytics for data-driven decision-making and productivity in banking sector. *Humanities and Social Sciences Communications*, 10(1). <https://doi.org/10.1057/s41599-023-02122-x>
- Handojo, A., Pujawan, N., Santosa, B., & Singgih, M. L. (2023). A multi layer recency frequency monetary method for customer priority segmentation in online transaction. *Cogent Engineering*, 10(1). <https://doi.org/10.1080/23311916.2022.2162679>
- Hicks, S. C., Liu, R., Ni, Y., Purdom, E., & Risso, D. (2021). mbkmeans: Fast clustering for single cell data using mini-batch k-means. *PLoS Computational Biology*, 17(1), e1008625. <https://doi.org/10.1371/journal.pcbi.1008625>
- Hsieh, N. (2004). An integrated data mining and behavioral scoring model for analyzing bank customers. *Expert Systems With Applications*, 27(4), 623–633. <https://doi.org/10.1016/j.eswa.2004.06.007>
- John, J. M., Shobayo, O., & Ogunleye, B. (2024). An exploration of clustering algorithms for customer segmentation in the UK retail market. *arXiv (Cornell University)*. <https://doi.org/10.48550/arxiv.2402.04103>
- Mancisidor, R. A., Kampffmeyer, M., Aas, K., & Jenssen, R. (2019). Learning Latent Representations of Bank Customers With The Variational Autoencoder. *arXiv (Cornell University)*. <https://doi.org/10.48550/arxiv.1903.06580>
- Mousaeirad, S. (2020). Intelligent vector-based customer segmentation in the banking industry. *arXiv (Cornell University)*. <https://doi.org/10.48550/arxiv.2012.11876>
- Nofal, S. (2024). Identifying highly-valued bank customers with current accounts based on the frequency and amount of transactions. *Heliyon*, 10(13), e33490. <https://doi.org/10.1016/j.heliyon.2024.e33490>
- Pandey, A. K., Goyal, A., & Sikka, N. (2024). RE-RFME: Real-Estate RFME Model for customer segmentation. *arXiv (Cornell University)*. <https://doi.org/10.48550/arxiv.2404.17177>
- Plotnikova, V., Dumas, M., Nolte, A., & Milani, F. (2022). Designing a data mining process for the financial services domain. *Journal of Business Analytics*, 6(2), 140–166. <https://doi.org/10.1080/2573234x.2022.2088412>
- Sculley, D. (2010). Web-scale k-means clustering. *Proceedings of the 19th International Conference on World Wide Web*, 1177–1178. <https://doi.org/10.1145/1772690.1772862>
- Sinaga, K. P., & Yang, M. (2020). Unsupervised K-Means clustering algorithm. *IEEE Access*, 8, 80716–80727. <https://doi.org/10.1109/access.2020.2988796>
- Wang, D. (2025). Parallelization of K-Means and spatial join algorithms on heterogeneous platforms using Apache Spark and GPU integration for enhanced AI information management. *Informatica*, 49(30). <https://doi.org/10.31449/inf.v49i30.10302>
- Wilbert, H. J., Hoppe, A. F., Sartori, A., Stefenon, S. F., & Silva, L. A. (2023). Recency, Frequency, Monetary Value, Clustering, and Internal and External Indices for Customer Segmentation from Retail Data. *Algorithms*, 16(9), 396. <https://doi.org/10.3390/a16090396>
- Yan, X., Li, Y., Nie, F., & Li, R. (2025). Bank customer segmentation and marketing strategies based on improved DBSCAN algorithm. *Applied Sciences*, 15(6), 3138. <https://doi.org/10.3390/app15063138>
- Yongjie, Z., & Shanyue, J. (2025). The impact of bank digital transformation on the financial performance of commercial banks. *SAGE Open*, 15(3). <https://doi.org/10.1177/21582440251365342>