

ارایه روش ترکیبی در خوشه بندی مشتریان بانک با استفاده از الگوریتم دو مرحله ای k-means و تحلیل عاملی

مهدی قنبری^۱

^۱ کارشناسی ارشد مهندسی نرم افزار دانشگاه آزاد اسلامی واحد مبارکه اصفهان

چکیده

با توجه به کاربردهای وسیع شناخته شده از داده کاوی در صنعت بانکداری این ابزار در سایر زمینه ها مانند خوشه بندی و رتبه بندی مشتریان در بازپرداخت اقساط تسهیلات مورد بررسی قرار می گیرد. لذا در این مقاله با اتخاذ به رویکرد مدیریت اعطای تسهیلات توسط بانک ها، به خوشه بندی و رتبه بندی ۱۰۰۲۲۴ نفر از مشتریان بر اساس بازدهی اعطای تسهیلات در موسسه امام رضا پرداخته شده است. بدین منظور از متغیرهای داده های جمعیت آماری مشتریان، پرداخت بدهی تسهیلات توسط مشتریان و دانش ضمنی خبرگان بانکی، هفت متغیر جهت خوشه بندی و رتبه بندی مشتریان استخراج شده است. در این مقاله در بخش اول از روش تحلیل عاملی ROBPCA برای طبقه بندی مشتریان استفاده شده است به طوری که در این روش مشتریان به چهار اهرم منظم به عنوان لایه داخلی و اهرم های عمود، خوب و بد افراز می شوند. همچنین بردار ویژه به دست آمده از ROBPCA به عنوان ورودی خوشه بندی الگوریتم دو مرحله ای K-means استفاده می شود. مشتریان در این مرحله به چهار خوشه طبقه بندی می شوند. نتایج به دست آمده نشان می دهد که ۷۳٪ مشتریان عضو مشتریان منظم بوده که میانگین مبلغ مصوب (وام) آن ها، ۷٪ میانگین مبلغ مصوب کل مشتریان بوده در حالی که بیشترین مقدار پرداخت را در بر دارند. با توجه به نتایج مشتریان اهرم عمود که ۱۷٪ مشتریان شامل می شود رفتار متفاوتی از مشتریان عادی (منظم) داشته و می تواند یکی از محورهای اصلی اعطای تسهیلات محسوب شوند. همچنین ۰۷٪ مشتریان عضو اهرم بد هستند که بیشترین مبلغ مصوب را دریافت کردند در حالی که کمترین مقدار برگشت را شامل می شوند. در پایان نیز پس از مجزا کردن مشتریان هم از طریق ROBPCA و هم از طریق خوشه بندی، مشتریان به شانزده قسمت طبقه بندی می شوند که سیاست های اجرایی مجزایی را در بر می گیرند.

واژه های کلیدی: خوشه بندی و رتبه بندی مشتریان، موسسه امام رضا، تحلیل عاملی ROBPCA، الگوریتم دو مرحله ای K-means.

۱. مقدمه

به‌طور ویژه بانک‌ها، سازمان‌هایی هستند که با مشتریان تعامل مستقیم دارند و عنصر مشتری در این سازمان‌ها اهمیت ویژه‌ای دارد. بدیهی است جهت پیشرو بودن در عرصه رقابت توجه به مشتری و اختصاص خدمات ویژه با تشخیص نیازمندی‌های آن‌ها و ارائه خدمات درست به آن‌ها محقق خواهد شد [۱]. تحلیل منابع با ارزش داده‌ای در رابطه با مشتریان کنونی بانک‌ها و مشتریان بالقوه همچنین شناخت مشتریان، ایجاد تمایز بین گروه‌های مختلف مشتریان و رتبه‌بندی آن‌ها یکی از مسائل مهم در بانک‌ها و سازمان‌های مشتری محور است برای رسیدن به درکی صحیح از مشتریان، سازمان‌ها نیازمند استفاده از مقیاسی هستند که از طریق آن‌ها بتوان میزان اهمیت مشتریان مختلف را سنجش نمود [۲]. از این رو یکی از ابزارهای اصلی نظام پولی و بانکی برای توسعه و ایجاد رونق اقتصادی در هر کشور اعطای تسهیلات هدفمند از طریق سیستم بانکی است [۳]. همچنین این تسهیلات در بخش‌های مختلف اقتصادی شامل تولیدی - صنعتی - کشاورزی - بازرگانی و خدماتی به متقاضیان عرضه می‌گردد. در حقیقت عدم شناسایی دقیق و نظارت ناکافی در حین و بعد از فرآیند اعطای تسهیلات می‌تواند ضمن هدر رفت منابع، آثار نامناسبی مانند رشد نامتوازن اقتصادی و همچنین نقدینگی در کشور ایجاد نماید. لذا بانک‌ها می‌کوشند از طریق شناسایی دقیق جامعه هدف و طبقه‌بندی آن‌ها به اعطای تسهیلات با بازدهی بالا اقدام نمایند. گسترش بهره‌گیری از فناوری‌های بروز در زمینه اطلاعات و ارتباطات خصوصاً در بانک‌ها باعث بر جای ماندن حجم عظیمی از داده‌ها شده که تحلیل و تصمیم‌گیری بر اساس آن‌ها با روش‌های معمول گزارش‌گیری و روش‌های آماری امکان‌پذیر نیست. لذا بهره‌گیری از روش‌های علمی و سامانمند نظیر خوشه‌بندی و رتبه‌بندی مشتریان از ضروریات این حوزه است. به‌طور کلی تسهیلات (وام) قراردادی است که بین مشتری و بانک بسته می‌شود، بانک‌های جمهوری اسلامی در اعطای تسهیلات بانکی از یازده راه عقد اسلامی مجاز انجام می‌شود. این عقود مجاز عبارت‌اند از: مشارکت مدنی - مشارکت حقوقی - فروش اقساطی - خرید دین - اجاره به شرط تملیک - جعاله - مساقات - مضاربه - قرض‌الحسنه - مزارعه. موسسه امام رضا یک نهاد حمایتی دولتی است که مسئولیت تأمین مالی کسب‌وکارهای خرد، کوچک و خانگی را در کشور بر عهده دارد. اعطای بالغ بر ۱۰۰۰ میلیارد تومان تسهیلات سالیانه در بخش‌های مختلف اقتصادی نشانگر تأثیر قابل توجه اعطای تسهیلات هدفمند توسط این صندوق به مشتریان است.

به‌رغم بسط و توسعه کارایی و روش‌های محاسبه آن در دو دهه اخیر، تاکنون کمتر مطالعه‌ای در مورد رتبه‌بندی بازدهی اعطای تسهیلات مشتریان و خوشه‌بندی مشتریان بر اساس بازدهی آن‌ها با استفاده از روش تحلیل داده‌های موسسه امام رضا پرداخته است. در این تحقیق بر اساس شاخص‌های بازدهی که در طرح‌های پشتیبانی شده و خدمات ارائه‌شده توسط امام رضا صورت می‌گیرد مشتریان خوشه‌بندی و رتبه‌بندی می‌شوند که در نتیجه امام رضا نامبرده بر اساس آن خوشه‌بندی، تصمیمات مؤثرتری رو مبنی بر اعطای تسهیلات بر روی بخش موردنظر می‌گیرند. در حال حاضر مشتریان صندوق کارآفرین امید توسط نظر خبرگان این بانک و صرفاً با استفاده از چند معیار محدود انتخاب می‌گردند. استفاده از الگوریتم‌های داده‌کاوی مانند، قواعد همبستگی و استخراج قوانینی جهت شناسایی مشتریان کلیدی بانک می‌تواند معیار خوبی جهت تکمیل نظر خبرگان در جهت انتخاب این دسته از مشتریان باشد. لذا در این پژوهش قصد داریم با استفاده از داده‌های صندوق کارآفرینی امید، الگویی جهت خوشه‌بندی و رتبه‌بندی مشتریان بانک با استفاده از روش تحلیل عاملی قدرتمند (ROBPCA) و الگوریتم خوشه‌بندی دو مرحله‌ای K-means پرداخته می‌شود.

۲. مبانی نظری

سرمایه‌گذاری و انباشت سرمایه در تحول اقتصادی کشورها نقش بسزایی دارد. اهمیت این عامل تا حدی است که یکی از مهم‌ترین عوامل رشد اقتصادی محسوب می‌شود. بدون تردید بانک به‌عنوان جایگاهی برای جذب و انتقال سرمایه‌ها می‌تواند نقش مهمی را ایفا کند، در مطالعه نقش بانک‌ها و به‌طور کلی بازارهای مالی دستیابی به سخت‌افزار و نرم‌افزارهای پیشرفته روند فزاینده‌ای داشته است. جذب منابع در بانک با کمک سپرده‌های مردم انجام می‌شود و سرمایه‌گذاری‌ها و یا مصارف بانک نیز در غالب اعطای تسهیلات مختلفی است که در اختیار متقاضیان قرار می‌گیرد ولی اینکه چه کسی صلاحیت دریافت این تسهیلات

را دارد جای بحث دارد؟ روش‌های مختلفی از طریق بررسی پیشینه مشتریان، نوع و کارکرد حساب‌هایشان وجود دارد ولی اغلب در نهایت بر اساس نظر مدیران رفتار می‌شود و یک قانون کلی وجود ندارد و بیشتر از زمینه‌های ذهنی کمک گرفته می‌شود. به‌طور کلی می‌توان گفت که جایگاه مشتری در فضای رقابتی کسب‌وکار بر کسی پوشیده نیست. به‌ویژه در نظام بانکی مشتری، نقش بسیار مؤثری در مسیر نیل به اهداف سازمان ایفا می‌کنند. این نقش در بانک‌های خصوصی بسیار مورد توجه قرار گرفته و این بانک‌ها مطالعات و فعالیت‌های ویژه‌ای را در این زمینه اختصاص داده‌اند.

بانک‌ها برای شناخت مشتریان، ارضای نیازمندی‌ها و ارائه خدمات مالی مناسب نیازمند شناسایی دقیق ویژگی‌های اعتباری آن‌ها هستند. یکی از خدمات مالی در بانک، اعطای تسهیلات مالی از جمله وام به متقاضیان اعتباری است. بانک‌ها همچون کسب‌وکارهای دیگر در طول حیات خود با ریسک‌هایی مواجه می‌شوند. یکی از مهم‌ترین آن‌ها ریسک اعتباری است که باید با آن مقابله کنند. اعتبارسنجی به‌عنوان یک تکنیک مؤثر به شناخت مشتریان خوب و بد پرداخته و با این کار می‌تواند ریسک اعتباری آن‌ها را تعیین کند. تحقیقات زیادی بر روی مدل‌های اعتبارسنجی در بانک‌ها صورت گرفته است. در ابتدا مدل‌های اعتبارسنجی به‌صورت قضاوتی بودند. سپس روش‌های پارامتریک در اعتبارسنجی مطرح شدند. اخیراً از روش‌های ناپارامتریک در اعتبارسنجی مشتریان بانک‌ها استفاده می‌شود. فیشر در سال ۱۹۶۳ برای اولین بار ایده متمایز کردن گروه‌ها را مطرح کرد. از جمله مطالعاتی که در زمینه بانکداری انجام شده است، مطالعه‌ای است که در سال ۲۰۱۵ توسط اسکار و همکارانش انجام شده است که در آن مدلی برای خوشه بندی و رتبه بندی مشتریان ارائه نمودند. آنان با استفاده از تکنیک تحلیل عاملی PCA و تحلیل پوششی داده‌ها DEA به خوشه بندی و رتبه بندی مشتریان ۹۶۶ شعب بانک در کانادا پرداختند. آنان در این پژوهش از سه متغیر ورودی و چهار متغیر خروجی جهت تحلیل شعبه‌های مربوط به هر یک از خوشه‌های به دست آمده استفاده کردند. نتایج این مطالعه نشان می‌دهد که شعب بانک‌ها در چهار خوشه مستقل قرار گرفته‌اند و هر خوشه نیز بر اساس اهرم‌های خوب و بد، منظم و عمود مورد بررسی قرار گرفتند [۴]. مطالعه‌ای دیگری در زمینه داده کاوی در سال ۲۰۱۴ توسط هنلی، کاپلن و بیسر انجام شده است که در آن مدلی برای محاسبه ارزش چرخه عمر مشتری ارائه نمودند. [۵]. هاشمی نژاد و همکارانش (۲۰۱۸)، ایجاد یک تکنیک‌های داده کاوی برای تحلیل مشتریان بانک پرداختند. هدف این مطالعه بررسی تکنیک‌های داده کاوی مورد استفاده برای تحلیل مشتریان بانک به منظور کمک به بانک‌ها برای شناسایی بهتر مشتریان و طراحی استراتژی‌های بازاریابی کارآمدتر است. [۶]. پور میر ارسلانی و خلیلیان (۲۰۱۷)، طی مطالعه‌ای تحت عنوان تشخیص تقلب در بانکداری الکترونیک با استفاده از انتخاب ویژگی ترکیبی و الگوریتم‌های تکاملی پرداخته‌اند. مطالعه حاضر روش جدیدی را برای تشخیص تقلب در بانکداری الکترونیک فراهم می‌کند که بر پایه انتخاب ویژگی ترکیبی و الگوریتم ژنتیک است. با توجه به نتایج به دست آمده، می‌توان گفت که سیستم پیشنهادی برای شناسایی تقلب در بیکن بسیار مفید است [۷]. رامون و همکاران (۲۰۱۷)، یک روش ترکیبی جدید برای بهبود مدیریت مشتریان بانکی معرفی کرده‌اند. هدف این مقاله فراهم نمودن یک مدل ارزش عمر مشتری CLV برای ارزیابی دقیق و طبقه‌بندی مشتریان بانکی با استفاده از مقیاس‌های فردی و پوشش روابط مشتریان با مجموعه محصولات شرکت می‌باشد. این تحقیق نشان می‌دهد که کدام یک از مشتریان ارزش مشتری را بهتر پیش‌بینی می‌کنند و استفاده از محصول برای هر یک از محصولات مورد نظر برای رسیدن به سنجه CLV را پیش‌بینی می‌کنند. [۸]. آیانو و همکارانش (۲۰۱۷)، یک روش طبقه‌بندی داده کاوی چند الگوریتم برای تراکنش‌های تقلبی بانک پرداختند. این مقاله یک استراتژی چند الگوریتم برای تشخیص تقلب در کارت پیشنهاد می‌کند. نتایج نشان می‌دهد الگوریتم dbscan با استفاده از الگوریتم پایه قاعده، دقت پیش‌بینی تقلب را نسبت به الگوریتم dbscan موجود در زمانی که تنها مورد استفاده قرار می‌گیرد، ارزیابی می‌دهد. [۹]. ایکنینی و همکارانش (۲۰۱۷)، در پژوهشی به ارزیابی یک مدل ارزش عمر مشتری برای صنعت بانکداری: یک راهنما برای فعالیت‌های بازاریابی پرداختند. هدف این مطالعه توسعه یک مدل کاربردی و دقیق برای ارزش عمر مشتری CLV و برجسته کردن مهم‌ترین شاخصه‌های مرتبط با یک صنعت خاص یعنی بخش بانکداری است. نتایج در مقایسه با مطالعات قبلی به شدت رقابتی هستند. این مطالعه نشان می‌دهد که در کنار شاخص‌هایی که عمدتاً در مقالات در اندازه‌گیری CLV استفاده می‌شوند، دو گروه دیگر، یعنی ارزش پولی و ریسک خدمات بانکی خاص، و همچنین

شاخص‌های مرتبط با مالکیت محصولات / خدمات نیز فاکتورهای مهمی هستند. [۱۰]. جوواویچ و همکارانش (۲۰۱۷)، به و تحلیل مقایسه‌ای نتایج رضایت مشتری و وفاداری در خدمات بانکی پرداختند. در این مقاله آنان از آمارگیری از مشتریان خدمات بانکداری نه تنها برای نشان دادن سطوح رضایت و وفاداری آن‌ها، بلکه برای تجزیه و تحلیل مقایسه‌ای داده‌ها حاصل از تحقیقات آنلاین و آفلاین استفاده کردند. [۱۱]. ماگوترا و همکارانش (۲۰۱۸)، طی مطالعه‌ای تحت عنوان بررسی ارتباط بین ارزش مشتری و رفتار اتخاذ تکنولوژی: مطالعه بخش بانکداری در هند پرداختند. مطالعه حاضر به منظور بررسی رابطه بین ادراک ارزش مشتری و رفتار اتخاذ فن‌آوری آن‌ها با توجه به مشتریان بانکی در هند تهیه شده است. نتایج این مطالعه نشان‌دهنده نیت رفتاری در جهت اتخاذ فن‌آوری به عنوان قوی‌ترین پیش‌بینی‌کننده درک ارزش از مشتریان است که با گرایش فردی آن‌ها به اتخاذ فن‌آوری، سودمندی درک شده و سهولت استفاده درک شده مورد استفاده قرار می‌گیرد [۱۲]. یانینگ و همکارانش (۲۰۱۲)، در پژوهشی به‌کارگیری رویکرد داده‌کاوی در پیاده‌سازی سیستم مدیریت با مشتریان را مورد توجه قرار داده‌اند و مزایای بهره‌گیری از داده‌کاوی را از دیدگاه کسب‌وکار در پذیرش مشتریان بانکداری اینترنتی، در قالب یک مدل مفهومی بیان کردند. [۱۳]. مطالعه‌ای نیز توسط کومار و همکارانش در سال ۲۰۰۶ انجام شده است که در آن شناسایی مشتریان وفادار بر اساس مفاهیمی چون طول دوره ارتباط مشتری با سازمان، فراوانی خرید مشتریان و متغیرهای RFM مربوط به مشتریان انجام شده است. بر اساس نتایج حاصل از این مطالعه ۲۰٪ مشتریان برتر از لحاظ 95% CLV سود را خلق کرده‌اند. [۱۴]. به‌رغم بسط و توسعه کارایی و روش‌های محاسبه آن در دو دهه اخیر، تاکنون کمتر مطالعه‌ای در مورد پیش‌خوشه‌بندی و رتبه‌بندی مشتریان بازدهی اعطای تسهیلات با بر اساس تکنیک‌های تحلیل عاملی انجام گرفته است. لذا در این مقاله به بررسی خوشه‌بندی و رتبه‌بندی مشتریان بر اساس بازدهی اعطای تسهیلات با استفاده از روش الگوریتم‌های PCA، ROBPCA و الگوریتم دو مرحله‌ای K-means بر اساس شاخص‌های بازدهی که در طرح‌های پشتیبانی شده و خدمات ارائه‌شده توسط موسسه امام رضا صورت گرفته است. در این مقاله سعی شده تا با استفاده از روش داده‌کاوی دانش نهفته در اطلاعات موجود در پایگاه داده‌های حاصل از تعاملات بانک و مشتری را استخراج نموده و از آن در جهت اخذ راهبردهای مدیریت ارتباط با مشتری بهره‌گیری نماییم. در ادامه به شناسایی متغیرها و روش‌های مورد استفاده در این مقاله خواهیم پرداخت.

۳. روش تحقیق

۱.۳. تجزیه مؤلفه‌های اصلی (ROBPCA)

بررسی مؤلفه‌های اصلی یکی از ارزشمندترین نتایج جبر خطی است زیرا روش ساده و غیرپارامتری استخراج اطلاعات مربوطه از مجموعه‌های گنجاننده است. در PCA داده‌ها از فضای با بعد بالا به فضای با بعد پایین برده می‌شوند و در واقع PCA یکی از روش‌های انتخاب ویژگی (Feature selection) است و به منظور کاهش بعد بکاربرده می‌شود تا بررسی ویژگی‌ها (Feature) در فضای با بعد پایین تر که ساده تر است انجام گیرد. کاری که PCA انجام می‌دهد این است که یک تبدیل خطی به گونه‌ای به دست می‌آورد که بردار ویژگی‌های h -بعدی را به بردارهای ویژگی-d-بعدی که (d, h) تبدیل می‌کند به طوری که اطلاعات به طور تقریباً کاملی (بیشینه اطلاعات ممکن) نگه داشته می‌شوند و کمینه خطای میانگین مربعات را به دست می‌دهد. تبدیل PCA می‌تواند به وسیله کمینه کردن خطای کمترین مربعات به دست آید. با فرض این که بردار ویژگی‌ها $X \in R^d$ (فضای d بعدی) باشد بعد فضای ویژگی‌های کاهش یافته $Y \in R^h$. اگر بردار ویژگی‌ها $X \in R^d$ باشد آنگاه خطای کمترین مربعات می‌تواند به صورت رابطه (۳) بیان شود.

$$MSE = E \left\| x - \hat{x} \right\|^2 \quad (3)$$

در PCA به دنبال تبدیلی هستیم که کمترین مقدار خطای مربعات را بدست آوریم یعنی که تبدیل ذیل مورد نظر است.

$$T_D = \arg \min_r \left\{ E \left\| x - \hat{x} \right\|^2 \right\} \quad (4)$$

همان بزرگ منطبق با مقدار ویژه بزرگ بایستی برای کاهش بعد و یا به هدف دسته بندی برای کمینه MSE گرفته شوند. در اینجا لازم است به برخی خواص ماتریس گونه که بیان شد $T^T \text{cov}_{x-\hat{x}} T$ PCA سعی دارد یک تبدیل خطی T را پیدا کند که کمترین خطای مربعات را به دست دهد و این تبدیل خطی در واقع به گونه ای است که را ماکزیمم کند که در آن ماتریس کوواریانس داده هایی با میانگین صفر از X است. بنابراین PCA مساله ویژه را حل کرده و بردار ویژه های ستونهای تبدیل خطی، ماتریس T را تشکیل می دهند. داده ها در بعد پایین یعنی Y نشان داده می شوند. با توجه به مطالب ذکر شده، MSE حداقل دارای d-h مقدار ویژه است یعنی بردار ویژه کوواریانس پرداخته شود چون در واقع مساله اصلی پیدا کردن مقادیر ویژه ماتریس کوواریانس است. تعریف کوواریانس به صورت

که در واقع ارتباط بین دو بعد که در اینجا X و Y است را بیان می کند با محاسبه $\text{COV}(X, X)$ تعریف واریانس حاصل می شود. پس واریانس که معیاری از پراکندگی مجموعه داده هاست در یک بعد به بررسی این پراکندگی می پردازد ولی کوواریانس در دو بعد به بررسی پراکندگی می پردازد. حال برای بعدهای بالاتر باز هم با کوواریانس مواجه می شویم با این تفاوت که در اینجا ماتریس کوواریانس وجود دارد. مثلاً در سه بعد ماتریس کوواریانس به صورت زیر است:

$$C = \begin{bmatrix} \text{cov}(x, x) & \text{cov}(x, y) & \text{cov}(x, z) \\ \text{cov}(y, x) & \text{cov}(y, y) & \text{cov}(y, z) \\ \text{cov}(z, x) & \text{cov}(z, y) & \text{cov}(z, z) \end{bmatrix} \quad (5)$$

همانگونه که از ماتریس کوواریانس مشاهده می شود این ماتریس متقارن است و این ماتریس واریانس داده ها را به دست می آورد و همچنین این ماتریس یک ماتریس مربعی است. بنابراین می توان بردارهای مشخصه را برای ماتریس کوواریانس به دست آورد با توجه به این که شرط لازم برای وجود بردارهای مشخصه مربعی بودن ماتریس است. بردارهای ویژه های ماتریس کوواریانس برهم عمود هستند در اینجا و در مورد PCA بردار ویژه هایی مد نظر هستند که مطابق با بزرگترین مقادیر ویژه باشند که در واقع مؤلفه های اصلی بردار ویژه های با بزرگترین مقادیر ویژه هستند که در امتداد این مؤلفه های اصلی بیشترین همبستگی را دارد [۱۵].

۲,۳. انتخاب متغیرها

شناسایی متغیرها: از گامهای اساسی در مسیر تحقیق شناسایی متغیرها است. با در نظر گرفتن اهداف تحقیق مطالعه تحقیقات انجام شده در حوزه بخش بندی رفتار مشتریان بانک و بررسی رویکردهای سازمانی امام رضا، سه دسته متغیر زیر انتخاب شده است:

داد

ه های جمعیت آماری مشتریان: این متغیرها شامل جنسیت، سن، تحصیلات، وضعیت تأهل، شغل و تاریخ تولد،

محل منطقه بانکی و .. است. که در این بخش متغیرهای مورد استفاده در مقاله عبارتند از: سطح سود مشتریان و نام بخش اقتصادی (شغل) مشتریان است.

پردا

خت بدهی تسهیلات توسط مشتریان: پرداخت بدهی تسهیلات مشتری می تواند شاخص مناسبی برای بررسی الگوهای رفتاری مشتری است و تحقیقات زیادی در این زمینه صورت گرفته است. در این میان متغیرهایی چون تکرار، متغیر پولی، ارزش پولی، در مطالعات داده کاوی کاربر گسترده ای یافته اند. این متغیرها در پایگاه اطلاعاتی بانک وجود دارند و از طریق پردازش داده های تراکنشی مشتریان محاسبه می شوند از این رو متغیرهای مورد استفاده در این بخش عبارتند از : مقدار مبلغ مصوب برای مشتری ، تعداد پرداختی های مشتری از شروع قرارداد تا تاریخ مطالعه، سود بانکی و ارزش تهریمی وثایق پرونده می باشند.

داز

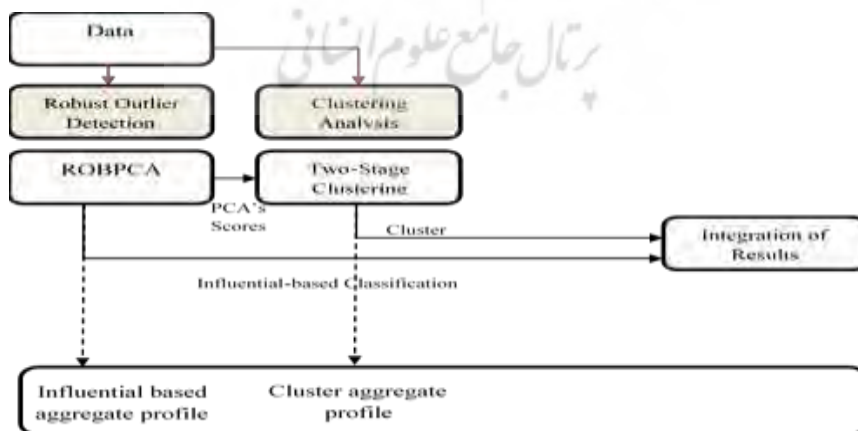
ش ضمنی خبرگان بانکی: امام رضا برخی از مشتریان خود را به عنوان مشتریان کلیدی می شناسد. مشتریان کلیدی بانکی در هر استان بنا بر نظر خبرگان بانکی با ویژگی های خاصی تعریف می شوند. منشا این ویژگی ها دانش ضمنی خبرگان بانکی در هر استان است. که در این بخش متغیر وضعیت حساب (فعال یا غیر فعال) استفاده شده است.

۴. تحلیل و نتایج

در این بخش مدل مفهومی مورد استفاده در این مقاله در شکل (۱)، نشان داده شده است. دیاگرام ارائه شده شامل سه نوع ارزیابی خوشه بندی و رتبه بندی مشتریان است که هر یک به صورت زیر توصیف می شود:

خوشه بندی با ROBPCA: در این مشتریان به روش تحلیل عاملی ROBPCA از اثرات آن در خروجی ها مورد ارزیابی قرار می گیرد. روش ROBPCA ارائه شده وزن هایی بین صفر و یک برای هر مشاهده ارائه می دهد که به صورت مستقیم نشان دهنده اثرات مشاهدات روی بردار میانگین و ماتریس کواریانس دارد. همچنین ROBPCA اطلاعات مفیدی در رابطه به کلاس بندی داده ها ارائه می دهد .

خوشه بندی K-means: در این روش امتیازها یا مقادیر ویژه ارائه شده توسط ROBPCA با ترکیب با الگوریتم خوشه بندی K-means به صورت دومرحله ای خوشه بندی می شود.



شکل ۱: مدل مفهومی تحقیق

برای خوشه بندی و شناسایی مشتریان بانک از دیاگرام ترکیبی نشان داده در شکل (۳) استفاده شده است. دیاگرام شامل کلاس بندی تک مرحله ای از طریق تحلیل عامل قدرتمند ROBPCA و خوشه بندی دو مرحله ای است. در روش ROBPCA اطلاعات مشتریان با ابعاد بالا به متغیرهای کوچکتری تجزیه می شوند که این مؤلفه ها ترکیب خطی از متغیرهای اصلی هستند. هدف اصلی PCA قدرتمند یا ROBPCA بدست آوردن مؤلفه هایی است که از لایه خارجی داده ها تأثیر نمی پذیرند که ماتریس کواریانس با ماتریس تخمین کواریانس قدرتمند جایگزین شده است. به این ترتیب در تمام روش های ROBPCA هدف پیدا کردن ترکیب خطی از متغیرها حتی برای داده های لایه خارجی و نوع و تعداد داده های لایه خارجی است. در این مقاله از روش ROBPCA به در نظر گرفتن هفت متغیر اصلی مسئله شامل (۱) مقدار مبلغ مصوب برای مشتری (۲) تعداد پرداختی های مشتری از شروع قرار داد تا تاریخ مطالعه (۳) سود بانکی (۴) ارزش تهرینی وثایق پرونده (۵) سطح سواد مشتریان (۶) وضعیت حساب (۷) نام بخش اقتصادی استفاده شده است. از بین متغیرهای اصلی مسئله سطح سواد مشتریان، وضعیت حساب و نام بخش اقتصادی متغیرهای کیفی هستند که به صورت کمی به شرح زیر کد گذاری شده اند:

جدول ۱: کدگذاری سطح سواد مشتریان

| بخش اقتصادی | کشاورزی | تعمیر مسکن | ساخت مسکن | صنعتی |
|-------------|---------|----------------|-----------|-------|
| کد | ۱ | ۲ | ۳ | ۴ |
| بخش اقتصادی | دیگر | نمایندگان مجلس | خدمات | رفاهی |
| کد | ۵ | ۶ | ۷ | ۸ |

جدول ۲: کدگذاری وضعیت حساب مشتریان

| وضعیت | فعال | معوق | مشکوک الوصول |
|-------|-----------|---------------|-----------------|
| کد | ۱ | ۲ | ۳ |
| وضعیت | تسویه شده | سر رسید گذشته | آماده آزاد سازی |
| کد | ۴ | ۵ | ۶ |

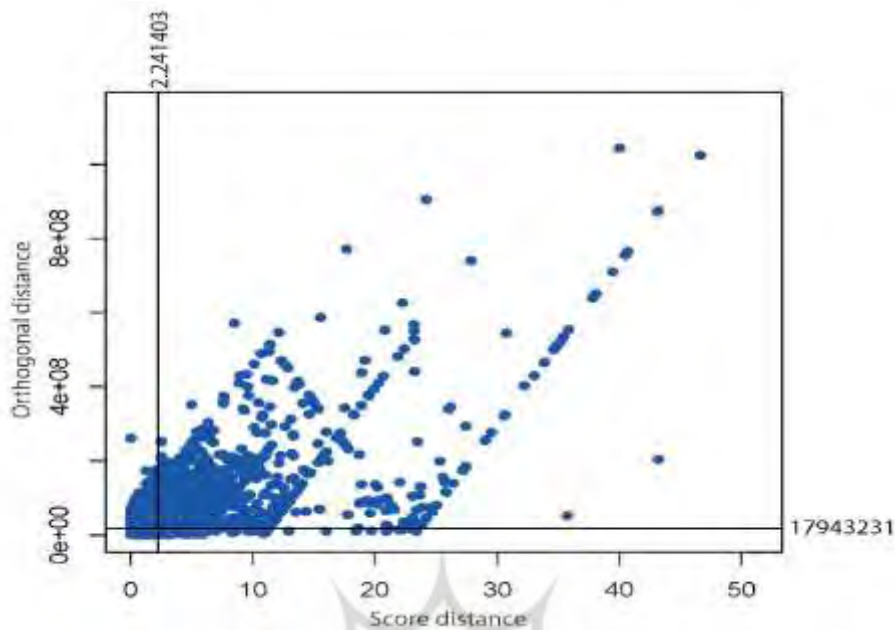
جدول ۳: کدگذاری بخش اقتصادی مشتریان

| سطح سواد | بی سواد | زیر دیپلم | دیپلم | لیسانس | فوق لیسانس |
|----------|---------|-----------|---------|--------|------------|
| کد | ۱ | ۲ | ۳ | ۴ | ۵ |
| سطح سواد | دکتری | فوق دکتری | کاردانی | فقه | نامعلوم |
| کد | ۶ | ۷ | ۸ | ۹ | ۱۰ |

تعداد مؤلفه های k مورد استفاده در ROBPCA با استفاده از معیار هابرت^۱ ۲۰۰۵ با توجه به فاصله مقادیر ویژه اول و دوم تعیین می شوند. در این مقاله با توجه به فاصله زیاد بین مقدار ویژه اول و دوم $k = 1$ انتخاب شده است. که نتیجه آن تبدیل فضای هفت بعدی از متغیرهای مسئله به فضای تک بعدی است که برابر بردار ویژه مشاهدات است. از بردار ویژه مشاهدات به

۱. Hubert

عنوان رتبه برای خوشه بندی در مرحله بعد استفاده می شود. با توجه به دو معیار فاصله رتبه SD و فاصله عمود OD نمودار تشخیص در شکل (۲) نشان داده شده است.



شکل ۲: نمودار تشخیص برای مشتریان بانک

از آنجا که در فرمت موجود محل مستقلی برای فهرست و تعریف همه متغیرها پیش‌بینی نشده است، لازم است که کلیه متغیرها بلافاصله پس از طرح در مقاله به شکل کامل تعریف گردند، به تعریف متغیرها بعد از رابطه (۱) رجوع شود.

جدول ۴: میانگین و انحراف معیار متغیرها

| | مبلغ مصوب | تعداد پرداخت | سود | رهن |
|--------------|-----------|--------------|------------|-------------|
| میانگین | ۸۹۶۴۵۸۴۲ | ۲۰/۸۵۱۶۲۲۳۷ | ۷۴۶۹۸۰۴/۷۶ | ۱۹۳۴۳۱۵۳۴/۶ |
| انحراف معیار | ۹۲۹۴۰۵۳۶ | ۱۵/۲۵۴ | ۱۲۷۲۷۲۳۸ | ۲۴۰۸۱۱۳۵۰ |
| میانگین | سواد | حساب | اقتصادی | |
| انحراف معیار | ۶/۴۴۷۴۱۸ | ۲/۶۹۷۰۳۹ | ۶/۷۴۶۲۸۸ | |
| میانگین | ۳/۴۵۵۸ | ۱/۷۷۹۷ | ۰/۰۰۶۱۹ | |

با توجه به شکل (۲) مشاهدات مسئله شامل اطلاعات ۱۰۰۲۲۴ مشتری است که به دو متغیر فاصله رتبه SD و فاصله عمود OD رسم شده است. این نمودار مشتریان را به چهار لایه افراز کرده است که سه لایه شامل اهرم بد (راست بالا)، اهرم خوب (راست پایین) و اهرم عمود (چپ بالا) به عنوان لایه های خارجی و افراز مشاهدات منظم (چپ پایین) به عنوان لایه داخلی شناخته شده اند. دو خط برش افقی و عمودی به ترتیب با مقادیر ۲/۲۴۱۴۰۳ و ۱۷۹۴۳۲۳۱ این چهار بخش را از هم مجزا کرده است.

جدول (۴) و (۵) میانگین متغیرهای مسئله برای هر اهرم نشان می‌دهد. با توجه به جدول ۷۳۲۲۹ نفر از مشتریان شامل ۷۳٪ در لایه داخلی قرار دارند که به عنوان مشتری منظم محسوب می‌شوند. همچنین ۲۷٪ از مشتریان در لایه خارجی قرار دارند که ۱۷٪ از آن‌ها به صورت عمود بر فضای ROBPCA قرار دارند که به صورت مستقیم با مشاهده داده‌های قابل تشخیص نیستند. همچنین ۲٪ و ۷٪ از مشاهدات به ترتیب در دو لایه خارجی اهرم خوب و بد قرار دارند. میانگین مبلغ مصوب برای اهرم منظم یا مشتریان عادی بر اساس جدول (۶) با ۷۳٪ فراوانی ۶۵٪ میانگین مصوب کل وام دریافت شده بانک است. ۷۴٪ درصد از مشتریان به طور تقریباً مساوی دارای حساب فعال یا تسویه شده هستند. از لحاظ اقتصادی اکثر (۵۹/۸۴٪) وام در مراکز رفاهی دریافت شده و همچنین سطح سواد اکثر مشتریان ثبت نشده است.

شاهدات اهرم بد همان طور که در جدول (۷) نشان داده شده است مشاهداتی هستند که میانگین مبلغ مصوبشان ۳/۲ برابر میانگین مبلغ مصوب کل است. همچنین نتایج کلاس بندی نشان می‌دهد که ۷٪ از مشتریان ۵ برابر بیشتر از مبلغ وام مشتریان عادی یا منظم شامل ۷۳٪ در کل مشتریان را شامل می‌شود.

همچنین با توجه به جدول میانگین پرداختی‌های مشتریان اهرم بد کمترین مقدار را نسبت به دیگر اهرم‌ها دارند به طوری که میانگین پرداختی اهرم بد ۳۸٪ میانگین پرداختی مشتریان عادی را دارند. بیشتر مشتریان اهرم بد از لحاظ اقتصادی در بیشتر (۵۱٪) شامل مراکز خدماتی می‌شود. ۵۸/۱۵ درصد مشتریان اهرم بد حساب فعال داشته و سطح سواد نامعلومی دارند.

جدول ۵: درصد میانگین متغیرها برای هر اهرم

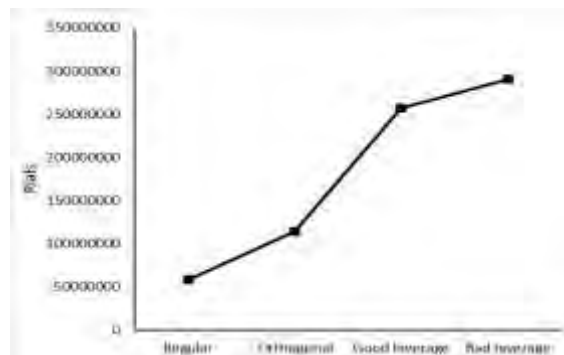
| انواع هرم | تعداد | مبلغ مصوب | تعداد پرداخت | سود |
|-----------|-------|-----------|--------------|-------|
| منظم | ۰/۷۳۱ | ۰/۶۵۴ | ۱/۱۵۷ | ۰/۶۶۹ |
| عمود | ۰/۱۷۴ | ۱/۲۷۹ | ۰/۶۱۶ | ۱/۰۸۹ |
| خوب | ۰/۰۲۵ | ۲/۸۷۱ | ۰/۶۲۸ | ۳/۵۵۸ |
| بد | ۰/۰۷۰ | ۳/۲۴۰ | ۰/۴۴۹ | ۳/۳۱۸ |
| منظم | رهن | سواد | حساب | |
| عمود | ۰/۶۸۹ | ۰/۹۹۵ | ۱/۱۰۵ | |
| خوب | ۰/۸۹۳ | ۱/۰۳۱ | ۰/۶۹۸ | |
| بد | ۳/۴۱۳ | ۰/۹۴۲ | ۰/۶۹۰ | |
| منظم | ۳/۶۴۸ | ۰/۹۹۹ | ۰/۷۶۲ | |

با توجه به نمودارهای مبالغ دریافتی و برگشتی مطابق با شکل‌های (۳) و (۴) به ترتیب اهرم منظم، اهرم عمود، اهرم خوب و اهرم بد بیشترین وام را دریافت کرده‌اند همچنین با افزایش مبلغ مصوب و حرکت به سمت اهرم بد تعداد قسط‌های برگشتی کاهش می‌یابد. اگر مشتریان خوش حساب مشتریانی تعریف شود که بیشترین برگشت پول و کمترین وام را دریافت کرده‌اند

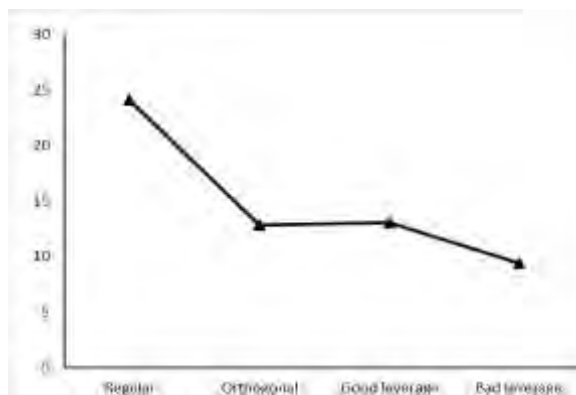
در این صورت می توان رتبه بندی به ترتیب اهرم منظم، اهرم عمود، اهرم خوب و اهرم بد برای خوش حسابی مشتریان انجام داد. در این قسمت از تحقیق از الگوریتم دو مرحله ای K-means برای خوشه بندی مشتریان استفاده شده است .
جدول ۶: سطح سواد، وضعیت حساب و بخش اقتصادی مشتریان منظم

| سطح سواد منظم | | | | |
|---------------|--------------|-----------------|---------------|-------------|
| لیسانس | دیپلم | زیر دیپلم | بی سواد | سطح سواد |
| ۱۱/۴۴ | ۲۱/۴۸ | ۱۴/۶۹ | ۰/۰۹ | درصد افراد |
| فقه | کاردانی | فوق دکتری | دکتری | سطح سواد |
| ۰/۰۱ | ۵/۱۰ | ۰/۰۰۴ | ۰/۰۱ | درصد افراد |
| وضعیت حساب | | | | |
| تسویه شده | مشکوک الوصول | معوق | فعال | وضعیت |
| ۳۵/۰۴ | ۲/۰۹ | ۵/۱۲ | ۳۹/۰۸ | درصد افراد |
| | گمرک غرب | آماده آزاد سازی | سر رسید گذشته | وضعیت |
| | ۰ | ۹ | ۹/۶۶ | درصد افراد |
| بخش اقتصادی | | | | |
| صنعتی | ساخت مسکن | تعمیر مسکن | کشاورزی | بخش اقتصادی |
| ۵/۲۲ | ۱/۱۲ | ۰/۳۸ | ۵/۶۱ | درصد افراد |
| رفاهی | خدمات | نمایندگان مجلس | دیگر | بخش اقتصادی |
| ۵۹/۸۴ | ۲۹/۱۰ | ۱/۴۲ | ۱/۳۳ | درصد افراد |

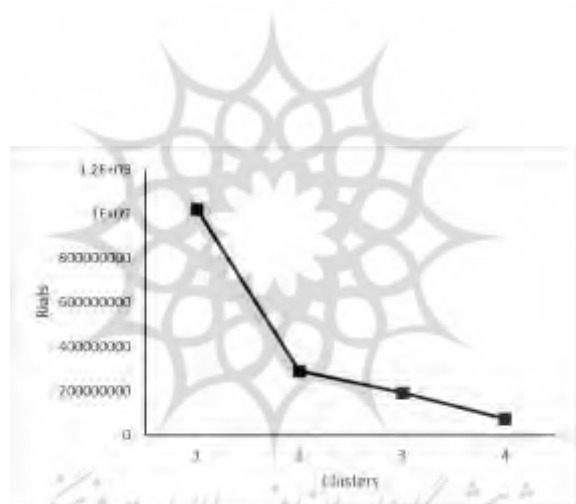
برای این منظور ضریبی از رتبه های ROBPCA به عنوان ورودی خوشه بندی به جای داده های اصلی استفاده شده است. در این مسئله با توجه به اینکه از یک مؤلفه در ROBPCA استفاده شده است، مسئله شامل یک بردار ویژه و یک مقدار ویژه است. پس از خوشه بندی به روش K-means به چهار خوشه نتایج به شرح زیر است. با توجه به جدول (۹)، خوشه اول شامل ۳۸ (۰/۰۳٪) نفر از مشتریان است که ۱۱/۳۹ برابر میانگین کل وام (۱ میلیارد ریال) دریافت کرده اند که می توانند از مشتریان با وام بالا محسوب شوند. این خوشه از مشتریان از بین خوشه های دیگر کمترین پرداخت را داشته اند. ۷۹٪ از این مشتریان حسابی فعال دارند که ۶۸٪ از این خوشه در بخش صنعتی سرمایه گذاری شده اند. خوشه چهارم شامل ۹۲/۸٪ از فراوانی مشتریان است که میانگین وام ۷۴/۲ میلیون ریالی دریافت کرده اند میزان برگشت پول برای این خوشه از مشتریان برابر میانگین تعداد پرداختی ها است. با توجه به شکل های (۵) و (۶) به تدریج از خوشه اول به چهارم با کاهش وام مصوب، تعداد قسط های برگشتی پول افزایش یافته است.



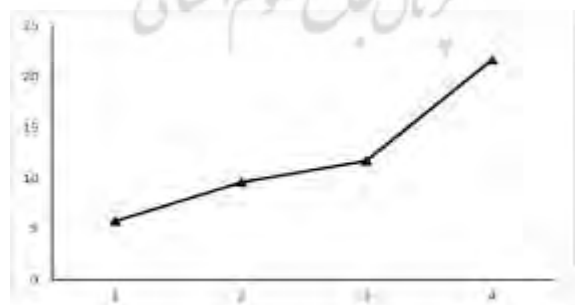
شکل ۳: تعداد پرداختی های هر اهرم



شکل ۴: مبلغ مصوب در هر اهرم



شکل ۵: مبلغ مصوب در هر خوشه



شکل ۶: تعداد پرداختی های هر خوشه

جدول ۷: سطح سواد، وضعیت حساب و بخش اقتصادی مشتریان اهرم بد

| سطح سواد بد | | | | |
|-------------|--------------|-----------------|---------------|------------|
| لیسانس | دیپلم | زیر دیپلم | بی سواد | سطح سواد |
| ۱۲/۹۸ | ۲۱/۵۲ | ۱۲/۴۲ | ۰/۰۱ | درصد افراد |
| فقه | کاردانی | فوق دکتری | دکتری | سطح سواد |
| ۰ | ۴/۷۸ | ۰ | ۰/۵۵ | درصد افراد |
| وضعیت حساب | | | | |
| تسویه شده | مشکوک الوصول | معوق | فعال | وضعیت |
| ۲۰/۲۲ | ۰/۵۲ | ۲/۵۲ | ۵۸/۱۵ | درصد افراد |
| | گمرک غرب | آماده آزاد سازی | سر رسید گذشته | وضعیت |
| | ۰/۰۸ | ۰/۸۴ | ۹/۱۹ | درصد افراد |
| بخش اقتصادی | | | | |
| صنعتی | ساخت مسکن | تعمیر مسکن | کشاورزی | اقتصادی |
| ۲۸/۴۸ | ۰ | ۰ | ۱۰/۱۹ | درصد افراد |
| رفاهی | خدمات | نمایندگان | دیگر | اقتصادی |
| ۹/۳۳ | ۵۱/۰۸ | ۰/۰۱ | ۰/۹۲ | درصد افراد |

| سطح سواد خوب | | | | |
|--------------|--------------|-----------------|---------------|------------|
| لیسانس | دیپلم | زیر دیپلم | بی سواد | سطح سواد |
| ۱۲/۵۷ | ۲۴/۸۵ | ۱۴/۹۲ | ۰/۰۸ | درصد افراد |
| فقه | کاردانی | فوق دکتری | دکتری | سطح سواد |
| ۰ | ۵/۸۸ | ۰ | ۰/۵۷ | درصد افراد |
| وضعیت حساب | | | | |
| تسویه شده | مشکوک الوصول | معوق | فعال | وضعیت |
| ۴/۵۰ | ۰/۶۹ | ۳/۷۳ | ۷۴/۴۶ | درصد افراد |
| | گمرک غرب | آماده آزاد سازی | سر رسید گذشته | وضعیت |

| | | | | |
|-------------|---------|------------|-----------|-------|
| درصد افراد | ۱۵/۱۲ | ۱/۵۰ | ۰ | |
| بخش اقتصادی | | | | |
| اقتصادی | کشاورزی | تعمیر مسکن | ساخت مسکن | صنعتی |
| درصد افراد | ۹/۴۹ | ۰ | ۰/۰۴ | ۱۵/۱۲ |
| اقتصادی | دیگر | نمایندگان | خدمات | رفاهی |
| درصد افراد | ۰/۲۸ | ۰ | ۷۲/۷۶ | ۲/۳۱ |

جدول ۸: سطح سواد، وضعیت حساب و بخش اقتصادی مشتریان اهرم

جدول ۹: میانگین متغیرها برای هر خوشه

| خوشه ها | اهرم منظم | اهرم عمود | اهرم خوب | اهرم بد | جمع |
|---------|-----------|-----------|----------|---------|--------|
| ۱ | ۰ | ۰ | ۰ | ۳۸ | ۳۸ |
| ۲ | ۰ | ۰ | ۱۸۵۶ | ۴۹۲۱ | ۶۷۷۷ |
| ۳ | ۰ | ۰ | ۹۰ | ۲۹۲ | ۳۸۲ |
| ۴ | ۷۳۲۲۹ | ۱۷۴۶۴ | ۵۲۱ | ۱۸۱۳ | ۹۳۰۲۷ |
| جمع | ۷۳۲۲۹ | ۱۷۴۶۴ | ۲۴۶۷ | ۷۰۶۴ | ۱۰۰۲۲۴ |

مراجع

- نام اسلامی علیشاه، ح. (۱۳۸۷). مهارت‌های بازاریابی در شعب بانک‌ها، بانک مسکن، ماهنامه اقتصادی، تبری و اجتماعی، شماره ۹۲.
- ۲- هاشمیان، م.؛ عیسایی، م. ت.؛ میکائیلی، ف.؛ طباطبائی، م. (۱۳۹۱). عوامل مؤثر بر پذیرش ابزارهای بانکداری الکترونیک از سوی مشتریان مدیریت فناوری اطلاعات، ۴ (۱۱)، صص ۱۵۵-۱۷۴.
- اداره آمدیریت بانک ملی، (۱۳۸۷). بانک ملی، تهران، اداره آموزش بانک ملی.
- Herrera-Restrepo, O., Triantis, K., Seaver, W. L., Paradi, J. C., & Zhu, H. (2016). Bank branch operational performance: A robust multivariate and clustering approach. *Expert Systems*, 50, ۱۰۷-۱۱۹.
- Haenlein, M., Kaplan, A. M., & Beeser, A. J. (2014). *A model to determine customer lifetime value in a retail banking context*. *Management Journal*, 25, 21-2۹.
- Hasheminejad, S. M. H., & Khorrami, M. Data mining techniques for analyzing bank customers: A survey. *Intelligent Decision Technologies*, (Preprint), 1-1۹.
- Pouramirarsalani, A., Khalilian, M., & Nikravanshalmani, A. (2017). Fraud detection in E-banking by using the hybrid feature selection and algorithms., 17(8), 271.
- Estrella-Ramon, Antonia, Manuel Sánchez-Pérez, Gilbert Swinnen, and Koen VanHoof. (2017): "A model to improve management of banking customers." *Industrial Management & Data Systems* pp 250-266.
- Ayano, O., & Akinola, S. O. (2017). A multi-algorithm data mining classification approach for bank fraudulent transactions. *African Journal of Mathematics and Computer Science Research*, ۱۰(۱), ۵-۱۳.
- Ekinci, Y., Uray, N., & Ülengin, F. (2014). A customer lifetime value model for the banking industry: a guide to marketing actions. *European Journal*, 48, 761-784.
- Jovović, M., Femić-Radosavović, B., & Lipovina-Božović, M. (۲۰۱۷). results of online and offline customer satisfaction & loyalty surveys in banking services in Montenegro. *Journal of Central Banking Theory and Practice*, 6(2), 65-76.
- Magotra, I., Sharma, J., & Sharma, S. K. (2018). Investigating linkage between customer value and technology adoption behaviour: A study of banking sector in India. *European Research on Management and Business Economics*, 24(1), 17-2۶.
- Li, Y., Li, K., & GAO, F. (2008, October). *Data mining approach and application in CRM project for internet-focused banking*. In 2008 4th International Conference on Wireless Computing (pp. ۱-۴).
- Kumar, V., Shah, D., & Venkatesan, R. (2006). Managing retailer profitability—one customer at a time! *Journal of Retailing*, ۸۲(۴), ۲۷۷-۲۹۴.
- ROBPCA: a new approach to robust principal component analysis. *Technometrics*, 47(1), pp.64-۷۹.