

## طراحی سیستم توصیه‌گر به منظور بهینه‌سازی و مدیریت تسهیلات بانکی بر مبنای الگوریتم‌های خوشه‌بندی و طبقه‌بندی تسهیلات

بابک سهرابی<sup>۱\*</sup>، ایمان رئیسی وانانی<sup>۲</sup>، فائزه زارع میرک‌آباد<sup>۳</sup>

- ۱- استاد مدیریت فناوری اطلاعات، دانشکده مدیریت، دانشگاه تهران، تهران، ایران
- ۲- استادیار مدیریت صنعتی، دانشکده مدیریت و حسابداری، دانشگاه علامه طباطبائی، تهران، ایران
- ۳- کارشناس ارشد، مدیریت فناوری اطلاعات، دانشکده مدیریت، دانشگاه تهران، تهران، ایران

پذیرش: ۱۳۹۵/۰۴/۱۲

دریافت: ۱۳۹۴/۱۰/۲۰

### چکیده

تسهیلات بانکی به عنوان یکی از پراهمیت‌ترین کارکرد بانک‌ها از اهمیت بالایی برخوردار است. اهمیت شناسایی مشتریان تسهیلاتی اگر بیشتر از مشتریانی سپرده‌گذار نباشد، به هیچ وجه کمتر نخواهد بود. در پژوهش‌های صورت گرفته، اهمیت این مشتریان به‌طور نسبی کمتر مورد توجه قرار گرفته است. هدف اصلی پژوهش حاضر، کاستن از مشکلات ارائه تسهیلات مناسب به مشتریان و طراحی سیستم توصیه‌گر تسهیلات بانکی با استفاده از داده‌های قبلی مربوط به تسهیلات ثبت شده در یک بانک منتخب و بزرگ کشور می‌باشد. راهکار ارائه شده، با دریافت داده‌های تسهیلات و اطلاعات حساب مشتری، گروه‌های مشتریان را در فرایند آموزش و اعتبارسنجی، فراگرفته و بر مبنای آن، امکان پیش‌بینی شرایط مشتری برای دریافت تسهیلات آتی را فراهم می‌کند. به منظور طراحی راهکار توصیه‌گر، روش‌های داده‌کاوی مناسب به همراه رویکرد اعتبارسنجی متناسب با هر روش انتخاب شد و سیستم نهایی با خطای اندک، جهت‌گیری مناسب بانک نسبت به متقاضیان دریافت تسهیلات را معین کرده است که به تصمیم‌گیری مدیران برای ارائه دقیق‌تر تسهیلات، یاری خواهد رساند. این سیستم به بانک‌ها کمک می‌کند تا با شناسایی انواع مشتریان

تسهیلاتی خود، تصمیم‌های مرتبط با هر گروه را به منظور ایجاد حداکثر سود و کاهش هزینه‌های وصول مطالبات اتخاذ نمایند و رویکرد متناسبی را برای هر مشتری در پیش گیرند. در نهایت، سیستم به سرپرستان بانک نیز ارائه شد و بازخوردهای مطلوبی برای بهبود و ارتقای راهکار دریافت شد.

واژه‌های کلیدی: تسهیلات بانکی، داده‌کاوی، خوشه‌بندی، طبقه‌بندی، سیستم توصیه‌گر.

## ۱- مقدمه

بهینه‌سازی تسهیلات بانکی برای بانک‌ها و مؤسسات مالی از اهمیت فوق‌العاده‌ای برخوردار است، زیرا از یک طرف رشد و توسعه سرمایه‌گذاری مهم‌ترین دلیل فعالیت بانک‌ها است و از طرفی دیگر اطمینان نداشتن از بازگشت سرمایه به بانک‌ها و مخاطرات بازگشت سرمایه مؤسسات مالی و اعتباری را در زمینه ارائه تسهیلات با محدودیت‌های فراوانی رو به رو کرده است و از طرف دیگر تسهیلات بانکی به‌عنوان پر اهمیت‌ترین اهرم اقتصادی نقش بسیار بااهمیتی را در رشد یک کشور از جنبه‌های گوناگون دارد. بنابراین تخصیص بهینه منابع تسهیلاتی از اهمیت فوق‌العاده‌ای برخوردار است. هم‌اکنون معیار تخصیص تسهیلات عواملی مانند موجودی حساب، شغل، تحصیلات، ضامن‌های معتبر و روند پرداخت اقساط در گذشته است که به صورت کاملاً تجربی و براساس نظر مدیران شعب بانک‌ها و یا مؤسسات مالی تخصیص داده می‌شود. همچنین نبود یک سیستم برای شناسایی مشتریان تسهیلات موجب می‌شود تا برنامه‌ریزی‌های مالی به‌خوبی صورت نگیرد.

افراد یا شرکت‌های گوناگون تسهیلات براساس ویژگی‌های فردی یا درآمدی و... باید شرایط مختص خود را داشته باشند ولی به دلیل نبود خوشه‌بندی مناسب برای همه مشتریان شرایط برابری قرار داده می‌شود که می‌تواند مشکلاتی از قبیل افزایش بدهی‌های جاری و غیر جاری به خاطر شرایط خاصی باشد که از قبل شناسایی نشده‌اند. برای مثال صنایعی که محصولات فصلی ممکن است به جای بازپرداخت ماهیانه با بازپرداخت با دوره‌های چندماهه بدهی جاری یا جرایم کمتری را متقبل شوند و غیره. از دیگر مسائل موجود، وجود تخلفات و سوء استفاده از تسهیلات است که در صورتی که این خوشه‌ها شناسایی شوند می‌توان با بروز علایمی خاص از مشتری

احتمال سوء استفاده را تعیین و اقدام‌های لازم پیش از دست رفتن منابع بانکی صورت پذیرد. ارائه سیستمی به منظور توصیه چگونگی مدیریت با مشتریان گوناگون تسهیلاتی برای ارائه تسهیلات راهگشا به نظر می‌رسد. در این سیستم با جمع‌آوری و تحلیل داده‌های مشتریان یکی از بانک‌های خصوصی کشور مشتمل بر داده‌های مانده حساب، بدهی جاری و غیر جاری، اصل و سود تسهیلات و داده‌های مشتریان در سیستم مدیریت ارتباط با مشتریان و پیاده‌سازی روش‌های داده‌کاوی<sup>۱</sup> در این سیستم یک سیستم توصیه‌گر<sup>۲</sup> منظور ارائه خدمات اعتباری ارائه خواهد شد و داده‌های مشتریان تسهیلاتی براساس داده‌های پایگاه داده بانک خوشه‌بندی<sup>۳</sup> شد. سپس با استفاده از طبقه‌بندی<sup>۴</sup> امکان پیش‌بینی داده‌های آینده محقق خواهد گردید. در این سیستم به این سؤال پاسخ داده می‌شود که چه خوشه‌هایی از مشتریان تسهیلاتی وجود دارند و چگونه می‌توان خوشه صحیح مشتریان را پیش‌بینی نمود، همچنین تغییرات عوامل را تحلیل حساسیت نمود.

پژوهش‌های صوت گرفته تاکنون در داخل و خارج از کشور به طراحی سیستم توصیه‌گر تسهیلاتی پرداخته‌اند. با این وجود الگوریتم‌های یادگیرنده ترکیبی که اقدام به بخش‌بندی داده‌های بزرگ یک بانک کرده و پس از تحلیل شرایط تسهیلاتی مشتریان اقدام به ارائه پیشنهاد کنند، کمتر مورد توجه قرار گرفته‌اند که ناشی از دسترسی نداشتن به داده‌های پالایش شده در مقیاس بالا و رایانه‌های توانمند در محاسبات بوده است. مدیران شعب و مدیران بالاتر همچنین مدیران بازاریابی بانک‌ها و سایر تصمیم‌گیرندگان در مؤسسات مالی و بانک‌ها با توجه به افزایش حجم داده‌ها و پیچیده شدن عوامل مؤثر بر تسهیلات و هزینه‌های وصول مطالبات و در بانک‌ها برای انتخاب روش‌های مدیریت مشتریان تسهیلاتی به سیستم‌های توصیه‌گر نیاز دارند تا گروه‌های مشتریان را به‌درستی پیش‌بینی کنند و تصمیم‌های متناسب را اتخاذ نمایند.

## ۲- پیشینه پژوهش

پژوهش‌های مرتبط با خدمات بانکی به‌طور عمده روش‌های بازاریابی و جذب و همچنین رضایت‌مندی مشتریان سپرده‌گذار بانک‌ها را مورد بررسی قرار می‌دهند و این در حالی است که مشتریان تسهیلاتی بانک علاوه بر اینکه با پرداخت سود جزو

منابع اصلی درآمد بانک هستند، در عین حال می‌توانند با خارج کردن منابع بانک به بانک خسارت وارد کنند.

سیستم‌های توصیه‌گر به منظور ارائه پیشنهاد به مشتریان زمان انتخاب کالا یا خدمت مورد نظر ایجاد و توسعه یافته‌اند که روش‌های مختلف را شامل می‌شود. در این پژوهش قصد داریم با استفاده از این سیستم مشتریان تسهیلاتی و گروه‌های آنها را شناسایی کرده و با تعیین وضعیت و رتبه‌بندی مشتریان تسهیلاتی برای مدیران شعب و مدیران ارشد یک سیستم توصیه‌گر مبتنی بر داده‌های تسهیلاتی ایجاد کنیم. این سیستم به منظور پیشنهاد اقدام‌های عملیاتی یا بازاریابی مورد نیاز متناسب با وضعیت هر مشتری ایجاد می‌شود. در این بخش تلاش بر این است تا به مرور مفاهیم سیستم توصیه‌گر، روش‌های داده‌کاوی، تسهیلات بانکی، پژوهش‌های انجام شده در بانک‌ها همچنین سیستم‌های توصیه‌گر مشابه بررسی شده و پیشینه این پژوهش نیز مورد بررسی قرار گیرد.

### ۳- پیشینه نظری

در این بخش به بررسی مفاهیم در حوزه سیستم توصیه‌گر تسهیلات پرداخته می‌شود.

#### ۳-۱- داده‌کاوی

تسهیلات بانکی، خروجی اصلی بانک‌ها هستند که از طریق آنها نقدینگی‌های سرگردان جامعه، به مبادی تعریف شده و هدفمند اقتصادی تزریق می‌شود؛ به این معنا که یک بانک با تجهیز منابع (شامل سرمایه و حقوق صاحبان سهام و انواع سپرده‌ها و یا سایر بدهی‌ها) آنها را در راستای اهداف از قبل تعیین شده به مصرف می‌رساند. با این تفسیر تسهیلات بانکی به عنوان چرخ دنده مؤسسات پولی و مالی نقش گسترده‌ای در اقتصاد کشورها دارد. داده‌کاوی به استخراج دانش از مقادیر زیادی از داده‌ها اشاره می‌کند.

داده‌ها ممکن است شامل داده‌های مکانی، داده‌های چند رسانه‌ای، داده‌های سری زمانی، داده متنی و داده‌های وب باشند. داده‌کاوی شامل فرایند استخراج الگوهای جالب، کوچک اما با اهمیت، ضمنی، ناشناخته و بالقوه مفید و یا دانش از مقدار بسیار

عظیمی از داده‌ها می‌باشد. این فرایند دربردارنده مجموعه‌ای از فعالیت‌ها برای پیدا کردن الگوهای جدید، ضمنی و یا غیر منتظره در داده‌ها یا الگوهای غیر معمول در آنها است. داده‌کاوی می‌تواند به سازمان‌ها در راستای درک بهتر کسب‌وکار خود، ارائه خدمات بهتر به مشتریان و افزایش اثربخشی سازمان در بلندمدت کمک کند [۱] در تعریفی دیگر وظیفه کشف الگوهایی در مجموعه داده‌های بزرگ است که شامل روش‌های هوش مصنوعی، یادگیری ماشینی، آمار و پایگاه داده می‌باشد. در این بخش، شرح مختصری از برخی روش‌های مورد استفاده برای تجزیه و تحلیل خوشه‌بندی داده و طبقه‌بندی ارائه شده است [۲]. از دیدگاه فنی، داده‌کاوی یک روش جمع‌آوری و فهرست‌بندی اطلاعات را ترکیب می‌کند و سپس به تولید دانش شبیه به قانون<sup>۵</sup> از بین مقادیر زیادی داده می‌پردازد. رایج‌ترین توابع مدلسازی در الگوریتم‌های داده‌کاوی موجود عبارتند از طبقه‌بندی، رگرسیون<sup>۶</sup>، خوشه‌بندی، قوانین انجمنی<sup>۷</sup>، ایجاد قانون<sup>۸</sup>، خلاصه‌سازی<sup>۹</sup>، مدلسازی وابستگی<sup>۱۰</sup> (تحلیل لینک) و تحلیل دنباله (کشف توالی)<sup>۱۱</sup>. مفهوم داده‌کاوی در پایگاه‌های داده، در اغلب تعاریف به صورت عمومی به شرح زیر تعریف می‌شود: «فرایند شناسایی الگوهای معتبر، جدید، مفید و درنهایت قابل فهم در بین داده‌ها». در دیدگاه آنها عبارت «داده‌کاوی در پایگاه داده» به معنای کل فرایند تبدیل داده‌های نا منظم به دانشی ارزشمند است. آنها همچنین نشان دادند که کل فرایند اکتشاف داده شامل چندین مرحله کلیدی است که عبارتند از پاکسازی داده<sup>۱۲</sup>، کاهش داده<sup>۱۳</sup>، تبدیل (تجمیع داده)<sup>۱۴</sup>، داده‌کاوی، ارزیابی الگوها<sup>۱۵</sup> که پس از طی این مراحل اکتشاف دانش روی می‌دهد. بدون شک داده‌کاوی به‌عنوان یک گام اساسی در فرایند اکتشاف دانش در نظر گرفته می‌شود و شامل استخراج الگوهایی از داده استبرای حصول اطمینان از اینکه الگوهای استخراج شده از داده، دانشی مفید هستند، گام‌های اضافی دیگری نیز لازم هستند [۳]. در مقالات متعددی نیز به تحلیل و پیش‌بینی با استفاده از روش‌های استخراج دانش و داده‌کاوی پرداخته شده است [۴؛ ۵].

### ۳-۲- خوشه‌بندی داده‌ها

تکنیک‌های خوشه‌بندی از راه شناسایی گروه‌هایی از کاربران که به نظر می‌رسد، اولویت‌های مشابهی داشته باشند و تقسیم‌بندی گروه‌هایی که اولویت‌های بسیار

متفاوتی دارند، کار می‌کنند. به‌طور خلاصه، یک روش خوشه‌بندی خوب، خوشه‌هایی را با کیفیت بالا تولید می‌کند، به‌طوری که درون هر کلاس بیشترین شباهت و بین کلاس‌های متفاوت کمترین شباهت وجود داشته باشد. با این حال، خوب بودن یک خوشه، درنهایت وابسته به نظر کاربر است. خوشه‌بندی، اقدامی است که سبب تقسیم‌بندی یک جمعیت همگن به تعدادی خوشه‌های ناهمگن می‌شود [۶]. خوشه‌بندی یک روش بدون ناظر نامیده می‌شود. این روش معمولاً برای آن دسته از بخش‌ها و خوشه‌هایی مورد استفاده قرار می‌گیرد که در تجزیه و تحلیل‌های آینده به کار گرفته خواهند شد، مانند تقسیم‌بندی مشتریان به بخش‌های مشابه که هر بخش دارای ویژگی‌ها و طرح بازاریابی منحصر به فرد خود است [۷]. یکی از روش‌های خوشه‌بندی که در این پژوهش مورد بررسی قرار گرفت، عبارت از خوشه‌بندی میانگین خوشه<sup>۱۶</sup> که یکی از الگوریتم‌های خوشه‌بندی کلاسیک بوده و به‌طور گسترده مورد استفاده واقع می‌شود و توسط مک‌کوئین در سال ۱۹۶۷ معرفی شد. این رویکرد، یک الگوریتم خوشه‌بندی نسبی است که کل مجموعه داده‌ها را در  $K$  خوشه مجزا دسته‌بندی می‌کند [۸].

الگوریتم‌های سلسله مراتبی به ایجاد یک تجزیه سلسله مراتبی از یک مجموعه از اشیای داده می‌پردازد. الگوریتم سلسله مراتبی می‌تواند براساس چگونگی تشکیل تجزیه سلسله مراتبی مجتمع یا متفرق باشد. روش مجتمع با هر شیء آغاز شده و به تشکیل یک خوشه جداگانه می‌پردازد و در هر تکرار پی در پی، اشیا و یا خوشه‌هایی را که نزدیک به یکدیگر هستند، در هم ادغام می‌کند تا زمانی که همه خوشه‌ها ادغام شده و یا شرط خاتمه برآورده شود. رویکرد متفرق، با تمام اشیا در خوشه مشابه شروع می‌شود و در هر تکرار پی در پی، یک خوشه به خوشه‌های کوچک‌تر تقسیم می‌شود تا درنهایت هر شیء در یک خوشه قرار گیرد و یا تا زمانی که شرط خاتمه برآورده شود. لینک منفرد، لینک متوسط و لینک کامل، نمونه‌هایی از الگوریتم‌های مجتمع می‌باشند، روش سلسله مراتبی با الگوی وارد<sup>۱۷</sup> می‌باشد. خوشه‌بندی در این روش خوشه‌بندی برای کاهش تلفات ناشی داده‌های دور افتاده از معیاری جدید برای محاسبه عدم شباهت بین خوشه‌ها استفاده می‌کند. در روش وارد از مجموع مربعات تفاضل هر داده از یک خوشه با بردار میانگین آن خوشه به‌عنوان معیاری برای سنجش

یک خوشه استفاده می‌شود. الگوریتم زیر را می‌توان برای روش وارد در نظر گرفت.

۱. نخست هر داده به‌عنوان یک خوشه در نظر گرفته می‌شود؛
۲. به ازای تمام جفت خوشه‌های ممکن از مجموعه خوشه‌ها آن دو خوشه‌ای که مجموع مربعات تفاضل داده‌های خوشه حاصل از اجتماع آنها با بردار میانگین خوشه حاصل کمینه باشد، انتخاب می‌شوند؛
۳. دو خوشه انتخاب شده با هم ترکیب می‌شوند؛
۴. تا زمانی که تعداد خوشه‌ها به تعداد مورد نظر نرسیده است، مراحل یک و دو و سه تکرار می‌شوند [۹].

### ۳-۳- طبقه‌بندی داده‌ها

طبقه‌بندی یکی از رایج‌ترین روش‌های کاربردی داده کاوی است که مجموعه‌ای از نمونه‌های از پیش طبقه‌بندی شده را برای توسعه مدلی که قادر به طبقه‌بندی جامعه سوابع بزرگ باشد، مورد استفاده قرار می‌دهد. برنامه‌های کاربردی تشخیص تقلب و مخاطرات اعتباری، به‌خوبی با این نوع از تجزیه و تحلیل تناسب دارند [۱].

الگوریتم درخت طبقه‌بندی یادگیری یک ساختار درختی است که تمام گره‌های غیر برگ با یک ویژگی و هر گره برگ با یک طبقه برچسب گذاری شده‌اند. شروع در گره ریشه انجام می‌شود، در هر گره، داده‌ها به صورت پی در پی و با شناسایی ویژگی‌ها و آستانه تقسیم، بخش‌بندی شده و تبعیض طبقات بهتر در زیرمجموعه‌های مربوطه داده‌ها انجام می‌شود [۱۰]. روش نزدیک‌ترین همسایه<sup>۸</sup> از دیگر روش‌های طبقه‌بندی است که با دریافت ویژگی هر خوشه به‌عنوان ورودی، داده جدید را با تک‌تک خوشه‌ها بررسی کرده و در خوشه‌ای که بیشترین مشابهت را داشته باشد، طبقه‌بندی می‌کند.

### ۴- سیستم‌های توصیه‌گر

سیستم‌های توصیه‌گر از دهه ۹۰ به‌طور گسترده‌ای برای ارائه خودکار پیشنهادهاى شخصی‌سازی‌شده کالاها و خدمات به کار گرفته می‌شوند [۹]. سیستم‌های توصیه‌گر

زمانی که حجم عظیمی از داده‌ها یا پیچیدگی محاسبات آنها مانع از انتخاب صحیح کاربر شود، به کاربران انتخاب‌های مناسب را معرفی می‌کند. سیستم‌های توصیه‌گر از توانایی فیلترکردن اطلاعات ضمنی برای پیش‌بینی ترجیح یا عدم ترجیح یک قلم خاص به وسیله کاربر در زمان تصمیم‌گیری برای انتخاب برخوردار می‌باشند هنگامی که تجربه شخصی ناکافی از یک دامنه خاص سبب ناممکن شدن انتخاب می‌شود، نیاز به توصیه از منابع قابل اعتماد وجود دارد از این سیستم‌ها استفاده می‌شود. سیستم‌های توصیه‌گر را هم اکنون می‌توان در بسیاری از برنامه‌های کاربردی جدید که کاربر را به سمت مجموعه‌های بزرگ از انتخاب‌ها سوق می‌دهند، دو گونه اصلی از سیستم‌های توصیه‌گر وجود دارد: سیستم‌های توصیه‌گر پالایشگر مشترک<sup>۱۹</sup> و سیستم‌های توصیه‌گر پالایشگر مبتنی بر محتوا<sup>۲۰</sup> که به شرح زیر می‌باشند، سیستم‌های توصیه‌گر پالایشگر مشارکتی منافع کاربران مشابه را مورد ملاحظه قرار داده و فرض می‌کنند کاربران فعال به مواردی علاقه‌مند هستند که کاربران مشابه آنها، به آنان امتیاز داده‌اند. سیستم پالایشگر مشارکتی کاربران را به عنوان عامل اصلی در نظر می‌گیرد. مبنای آن کاربرانی هستند که موارد مشابه را انتخاب کرده‌اند و براساس آن کاربران را مقایسه می‌نماید. بنابراین سیستم پالایشگر مشارکتی بر مبنای نظرات کاربران عمل می‌کند [۱۲]. سیستم پالایشگر مبتنی بر محتوا از خصوصیات متنی (محتوا) اقلام به منظور ایجاد توصیه‌های مناسب، استفاده می‌کنند. این رویکرد به آموزش طبقه‌بندی یادگیرندگان ماشینی در برابر پروفایل های کاربران و اقلام می‌پردازد.

## ۵- ارزیابی سیستم توصیه‌گر

رویکردهای توصیه‌گر مبتنی بر پالایش مشارکتی مبتنی بر میانگین خوشه، برای حل مسئله توصیه یا پیشنهادها ارائه شده‌اند که بر اساس معیارهای زیر ارزیابی می‌شوند. دقت: دقت، یکی از مهم‌ترین اهداف طراحی در جامعه سیستم‌های توصیه‌گر یا توصیه‌گر می‌باشد. اگر یک کاربر، سیستم توصیه‌گر را مورد استفاده قرار داده و پس از آن کشف کند که در حال دریافت توصیه‌های غلط می‌باشد، بعید است به ادامه استفاده از سیستم بپردازد.



**مقیاس‌پذیری:** یک الگوریتم توصیه‌گر باید به آرامی و با داده‌های تدریجی، مقیاس‌بندی شود. سیستم‌های توصیه‌گر پالایشگر مشترک، اغلب از مقیاس‌پذیری مناسب برخوردار نمی‌باشند. لازم به ذکر است که استحکام سیستم‌های توصیه‌گر توسط دو عامل دقت و مقیاس‌پذیری اندازه‌گیری می‌شود. این عوامل در تضاد هستند، هر چه زمان کمتری برای ساخت خوشه‌ها به وسیله الگوریتم صرف شود، مقیاس‌پذیری بیشتر خواهد بود؛ با این حال ممکن است توصیه‌ها با توجه به خوش‌های محدود از کیفیت پایین‌تری برخوردار باشند. کیفیت خوشه: ایده خوشه‌بندی الگوریتم، شامل دسته‌بندی کاربران در گروه‌های مختلف و براساس شباهت‌های میان آنها می‌باشد [۱۳].

**پوشش:** از آن جایی که در بسیاری از موارد و به خصوص در سیستم‌های پالایشگر مشارکتی، دقت پیش‌بینی سیستم توصیه با حجم داده افزایش پیدا می‌کند، بعضی از الگوریتم‌ها تنها بتوانند برای مقداری کمی از موارد توصیه با کیفیت ارائه دهند و این در حالی است که حجم عظیمی از داده وجود دارد. عبارت پوشش می‌تواند به چندین ویژگی متفاوت از یک سیستم اشاره کند:

- **پوشش فضای موارد:** به‌طور کلی، پوشش به نسبتی از موارد اشاره دارد که یک سیستم توصیه‌گر می‌تواند توصیه کند.

- **پوشش کاربر:** همچنین می‌تواند به درصدی از کاربران یا تعاملات کاربری اشاره کند که سیستم برای آنها موردی را توصیه می‌کند. در بسیاری از برنامه‌ها ممکن است توصیه‌گر پیشنهادی را فراهم نکند.

**شروع سرد (استارت سرد):** دسته‌ای دیگر از مسائل به هم مرتبط شروع سرد یا استارت سرد نام دارند: کارایی سیستم بر موارد و کاربران جدید، به شروع سرد تعبیر می‌شود. شروع سرد را می‌توان یک زیرمجموعه پوشش قلمداد کرد، زیرا که می‌تواند پوشش سیستم را بر دسته‌ای خاص از موارد و کاربران اندازه‌گیری کند. به غیر از اندازه‌گیری حجم کاربران و موارد، شروع سرد می‌تواند دقت سیستم را برای این کاربران و موارد اندازه‌گیری کند.

**اطمینان:** اطمینان در توصیه را می‌توان اعتماد سیستم به توصیه‌ها و پیش‌بینی‌های خودش تعریف کرد. همان طور که پیش از این اشاره شد، با رشد اطلاعات در مورد موارد، دقت توصیه‌گرهای پالایشگر مشارکتی نیز افزایش پیدا

می‌کند. به‌طور مشابه، اعتماد در ویژگی پیش‌بینی شده نیز با حجم داده افزایش پیدا می‌کند.

**اعتماد:** اعتماد مشابه اطمینان است، با این تفاوت که اعتماد به ایمان کاربر به توصیه سیستم اشاره دارد، برای مثال بهتر است سیستم نخست مواردی را پیش‌بینی کند که کاربر آن را از قبل می‌داند. اگر چه در این روش کاربر منفعتی از این توصیه‌ها نمی‌برد، اما مشاهده می‌کند که سیستم توصیه‌هایی منطقی ارائه می‌دهد و اعتماد وی به توصیه‌های سیستم در آیتم‌های نامعلوم افزایش پیدا می‌کند.

**تازگی:** توصیه‌های تازه، توصیه‌هایی هستند که کاربر آن را برای آیتم در نظر نداشته است. یک راهبرد ساده و کاربردی در برنامه‌هایی که به پیشنهاد تازه احتیاج است، حذف کردن مواردی است که کاربر پیش از این به آنها امتیاز داده یا از آنها استفاده کرده است.

**تعجب‌آوری:** مقیاسی است که نشان می‌دهد یک توصیه موفق چه قدر خارج از ذهن بوده است. برای مثال اگر کاربر امتیاز بالایی به تعداد زیادی از فیلم‌هایی که یک بازیگر خاص در آن بازی کرده است داده باشد، توصیه کردن فیلم جدید آن بازیگر می‌تواند تازه باشد، زیرا که ممکن است کاربر از آن اطلاع نداشته باشد، اما نمی‌توان آن را تعجب‌آور تلقی کرد. به‌طور مشخص ارائه توصیه به شکل تصادفی می‌تواند به طور شدیدی دور از ذهن باشد و به همین دلیل تعجب‌آوری را باید با مقیاس دقت متوازن کرد.

**گوناگونی:** به‌طور کلی گوناگونی متضاد شباهت در نظر گرفته می‌شود. در بسیاری از موارد، پیشنهاد کردن دسته‌ای مشابه از موارد برای کاربر مفید نیست، زیرا که کاوش گستردگی آیتم زمان بیشتری می‌طلبد. برای مثال توصیه برای مسافرت را در نظر بگیرید. بسیاری از وبسایت‌های تجارت الکترونیک از یک سیستم توصیه استفاده می‌کنند تا درآمد خود را به‌طور مثال از طریق فروش چندگانه تقویت کنند. در چنین مواردی موتور توصیه‌گر را می‌توان با معیار درآمدی که برای وبسایت به‌دست می‌آورد، ارزیابی کرد.

**ریسک:** در بعضی از موارد یک توصیه ممکن است با ریسک همراه باشد، برای مثال در توصیه خرید سهام، کاربران ممکن است ریسک‌گریز باشند و

سهام‌هایی را ترجیح می‌دهند که رشد احتمالی کمتری دارند ولی در عین حال خطر فروپاشی کمتری نیز دارند. استحکام درواقع پایداری یک توصیه در حضور اطلاعات غلط است که به‌طور عمد قرار داده شده است تا توصیه‌ها را مختل کند. با افزایش تعداد افرادی که برای راهنمایی به سیستم‌های توصیه‌گر اتکا می‌کنند.

**حریم خصوصی:** کاربر در یک سیستم پالایشگر مشارکتی تمایلات خود درباره موارد را به سیستم ارائه می‌کند تا بتواند توصیه‌های مفیدی دریافت کند. با این حال برای بسیاری از کاربران، خصوصی ماندن تمایلات آنها اهمیت بسیاری دارد؛ یعنی نمی‌خواهند که شخص ثالثی بتواند از سیستم توصیه، اطلاعاتی در مورد تمایلات شخصی او به‌دست آورد.

**انطباق‌پذیری سیستم‌های حقیقی:** توصیه ممکن است در یک فضای فعالیت کنند که دسته موارد به سرعت می‌کند یا روندهای تمایل به مواردی تغییر می‌کنند. **مقیاس‌پذیری:** از آن جایی که سیستم‌های توصیه‌گر به‌گونه‌ای طراحی شده‌اند که بتواند کاربران را در امر کاوش دسته‌هایی عظیم از موارد یاری کنند، یکی از اهداف اصلی طراحان این سیستم‌ها بالا بردن مقیاس سیستم تا حد حقیقی اطلاعات است. به همین دلیل بسیاری از الگوریتم‌ها ویژگی‌هایی مثل دقت یا پوشش را فدای به دست آوردن سریع نتایج در مقیاس دسته داده‌هایی شامل میلیون‌ها قلم از داده می‌کنند [۱۴].

در این پژوهش براساس مطالعه روش‌های پیشین و در نظر گرفتن داده‌های در دسترس، سیستم توصیه‌گر بر مبنای روش‌های داده‌کاوی مدنظر قرار گرفت. در همین راستا، سیستم ترکیبی توصیه‌گر مشتریان که در پژوهش‌های پیشین کمتر به آن اشاره شده است برای اجرای این پژوهش انتخاب گردید. همچنین از روش‌های اعتبارسنجی متناسب با هر روش داده‌کاوی برای بررسی دقت و صحت نتایج استفاده گردید.

## ۶- پیشینه تجربی

پژوهش‌های مرتبط با این موضوع در سه گروه انجام شده است سیستم‌های توصیه‌گر مبتنی بر پالایش مشارکتی با استفاده از خوشه‌بندی و طبقه‌بندی،

داده‌کاوی در حوزه بانکی و بخش‌بندی مشتریان که این دو گروه پژوهش به صورت خلاصه در جدول‌های ۱ و ۲ درج شده‌اند [۱۵؛ ۱۶؛ ۱۷؛ ۱۸؛ ۱۹؛ ۲۰؛ ۲۱؛ ۲۲؛ ۲۳].

جدول ۱ پژوهش‌های مرتبط با سیستم توصیه‌گر بر مبنای روش‌های داده‌کاوی

ردیف	سال	موضوع	مدل/ متغیرهای اصلی	جامعه آماری و نمونه	یافته‌های تحقیق
۱	۱۳۹۳	ارائه سیستم پیشنهاددهنده رشته تحصیلی در حوزه آموزش الکترونیکی مبتنی بر تکنیک‌های داده‌کاوی (مورد مطالعه دانشکده مدیریت پردیس مرکزی و دانشکده مدیریت پردیس قم دانشگاه تهران)	سیستم توصیه‌گر پالایشگر مشارکتی با استفاده از ترکیب خوشه‌بندی و طبقه‌بندی	داده‌های مرکز آموزش الکترونیکی دانشگاه تهران	اطلاعات فراگیران و رشته تحصیلی در هر خوشه نگاشت ایجاد کرده و در نتیجه به ارائه پیشنهاد رشته تحصیلی به فراگیران پرداخت
۲	۱۳۹۱	ارائه مدلی برای طراحی و توسعه یک سیستم پیشنهاددهنده محصول بر مبنای بخش‌بندی مشتریان و بهره‌گیری از تکنیک‌های داده‌کاوی	سیستم توصیه‌گر مبتنی بر خوشه بندی و سپس درک قواعد انجمنی	داده‌های سایت فروشگاه‌های آنلاین	ارائه سیستم توصیه‌گر برای خرید مشتریان در فروشگاه‌های آنلاین
۳	۱۳۹۳	طراحی سیستم پیشنهاددهنده عطر براساس سلیقه مشتریان بر مبنای ترکیبی از روش‌های داده‌کاوی	سیستم توصیه‌گر مبتنی بر خوشه‌بندی و سپس طبقه‌بندی با روش آموزش شبکه عصبی	جمع‌آوری داده‌های ترکیبات ساختاری شماری از عطرهای موجود در بازار دنیا	ارائه سیستم توصیه‌گر عطر به مشتریان
۴	۲۰۱۴	سیستم توصیه گر مبتنی بر طبقه بندی بر اساس تئوری رای گیری	ایجاد سیستم توصیه‌گر براساس تئوری انتخابات با استفاده از روش خوشه‌بندی	داده‌های مرتبط با فیلم	ارائه سیستم توصیه‌گر فیلم
۵	۲۰۱۵	شبکه نورو فازی تطبیقی برای سیستم های توصیه گر	یجاد سیستم توصیه گر بر مبنای روش نورو فازی برای سریا های ویدئویی	داده های مشتریان سایت سریال های ویدئویی	ارائه سیستم توصیه گر سریال

جدول ۲ پژوهش‌های صورت گرفته در مؤسسات مالی و بانک‌ها با استفاده از داده‌کاوی

ردیف	سال	موضوع	مدل/متغیرهای اصلی	یافته‌های تحقیق
۱	۱۳۹۱	خوشه‌بندی مشتریان در بانکداری خرد براساس وفاداری: موردکاوی بانک تجارت	خوشه‌بندی فازی مشتریان بانک تجارت	خوشه‌بندی مشتریان بانک تجارت براساس مدل RFM
۲	۱۳۹۰	بخش‌بندی مشتریان بانک و تعیین استراتژی ارتباط با مشتری با استفاده از داده‌کاوی	خوشه‌بندی براساس خوشه‌بندی k-means	خوشه‌بندی مشتریان بانکی براساس
۳	۱۳۸۳	طراحی و پیاده‌سازی سیستم هوشمند پرداخت تسهیلات بانکی	الگوریتم ژنتیک و شبکه‌های عصبی	پیاده‌سازی سیستم
۴	۲۰۰۸	مدل گسترده شده پذیرفته شده انتخاب سیستم توصیه‌گر نمونه یکی از بانک‌های ایران	مدل توسعه‌یافته	تأیید آماری مدل توصیه‌گر گسترش یافته
۵	۱۳۹۲	طبقه‌بندی مشتریان اینترنت بانک با کمک الگوریتم‌های داده‌کاوی	درخت تصمیم	طبقه‌بندی مشتریان براساس ویژگی‌های فردی
۶	۲۰۰۸	به‌کارگیری داده‌کاوی برای مدیریت ارتباط با مشتریان با تمرکز بر مشتریان کارت اعتباری و اینترنت بانک	الگوریتم ژنتیک و شبکه عصبی	استفاده از روش‌های داده‌کاوی برای مدیریت ارتباط مشتریان به‌ویژه کارت اعتباری
۷	۲۰۰۹	داده‌کاوی و کاربرد آن در مدیریت ارتباط با مشتریان بانک‌های تجاری	شبکه عصبی	بررسی ریسک در ارتباط با مشتری
۸	۱۳۸۷	به‌کارگیری داده‌کاوی برای کشف مدل امتیازبندی و تحلیل رفتار مشتریان بانک	شبکه عصبی	استخراج عوامل مؤثر در بازپرداخت تسهیلات
۹	۱۳۸۹	به‌کارگیری الگوریتم ژنتیک در بهینه‌سازی درختان تصمیم‌گیری برای اعتبارسجی مشتریان بانک‌ها	الگوریتم ژنتیک	استخراج طبقات اعتباری با مشتریان از روش الگوریتم ژنتیک
۱۰	۱۳۸۹	طراحی مدل انتخاب نیروی انسانی با رویکرد داده‌کاوی (مورد: استخدام داوطلبان آزمون‌های ورودی یک بانک تجاری در ایران)	درخت تصمیم	ایجاد سیستم رتبه‌بندی داوطلبان استخدام
۱۱	۱۳۹۱	پژوهشی در مشترکان تلفن همراه استان یزد	شبکه عصبی	بخش‌بندی مشتریان تلفن همراه
۱۲	۱۳۹۲	طبقه‌بندی مشتریان اینترنت بانک با کمک الگوریتم‌های داده‌کاوی	درخت تصمیم	ارائه سیستم پیش‌بینی‌کننده طبقات مشتریان اینترنت بانک

ردیف	سال	موضوع	مدل/متغیرهای اصلی	یافته‌های تحقیق
۱۳	۱۳۹۳	تحلیل مشکلات شعب بانک آینده در سراسر کشور با استفاده از روش داده‌کاوی	خوشه‌بندی با روش میانگین همسایه	ارائه مدل RFM براساس تماس‌های مشتریان
۱۴	۲۰۱۳	یک مدل جدید برای مدیریت شهری (مطالعه موردی: سیستم شهری در ایران)	خوشه‌بندی برای استخراج شاخص RFM	سیستم مدیریت روابط شهروندان، به‌منظور شناسایی مشکلات عمده و اصلی و خوشه‌بندی
۱۵	۱۳۹۳	بخش‌بندی کاربران بانکداری اینترنتی بر مبنای انتظارات: با رویکرد داده‌کاوی	خوشه‌بندی با روش میانگین همسایه	استخراج خوشه‌های مشتریان اینترنت بانک
۱۶	۲۰۱۵	مدل توسعه یافته برای خوشه‌بندی و طبقه‌بندی مشتریان بانکی	خوشه‌بندی با روش میانگین نزدیک‌ترین همسایه	ارائه مدلی ترکیبی از RFM و LD
۱۷	۲۰۰۴	مدل یکپارچه داده‌کاوی و رفتاری برای تجزیه و تحلیل مشتریان بانک [16]	الگوریتم ژنتیک	ارائه خوشه‌بندی رفتاری مشتریان
۱۸	۱۳۹۰	ارائه الگوریتمی برای طبقه‌بندی مشتریان بانک پارسیان و ارائه خدمات مناسب به گروه‌های مختلف مشتریان با استفاده از داده‌کاوی و شبکه‌های عصبی	ترکیب روش‌های طبقه و روش عصبی	طبقه‌بندی براساس ویژگی‌های مشتریان
۱۹	۱۳۹۳	شناسایی و رتبه‌بندی عوامل مؤثر بر سودآوری شعب بانکها با استفاده از روش داده‌کاوی - موردکاوی یک بانک	روش‌های خوشه‌بندی و وابستگی انجمنی	خوشه‌بندی مشتریان بانکی و تحلیل سودآوری هر گروه
۲۰	۱۳۸۶	ارزیابی کارایی شعب بانک پارسیان با استفاده از روش تلفیقی داده‌کاوی و تحلیل پوششی داده	روش خوشه‌بندی و تحلیل وابستگی با استفاده از شبکه‌های عصبی	ارائه سیستمی به منظور سنجش کارایی شعب
۲۱	۱۳۸۸	دسته‌بندی مشتریان در مدیریت ارتباط با مشتری با استفاده از داده‌کاوی	درخت تصمیم	ارائه سیستم پیش‌بینی وضعیت مشتری‌بگیران

## ۷- روش‌شناسی پژوهش

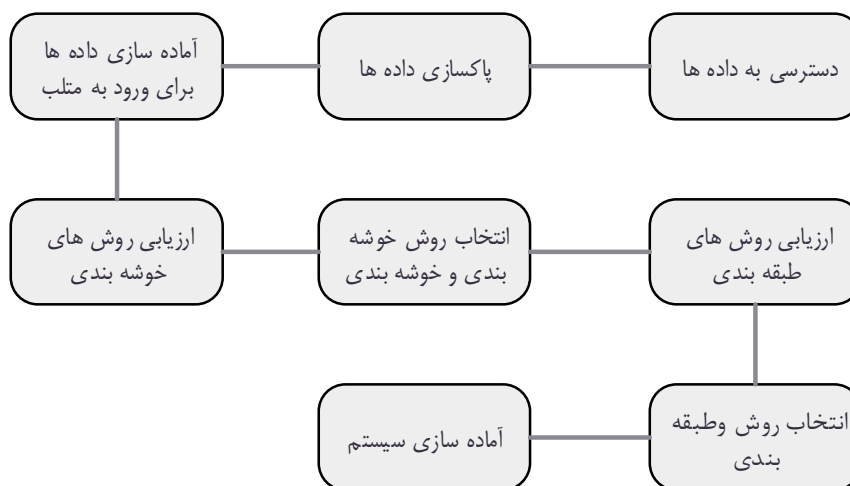
این پژوهش از منظور هدف کاربردی و از منظر روش به دلیل بررسی وضعیت موجود و استفاده از داده‌های حقیقی و توصیف آنها توصیفی و از منظر نوع داده‌ها کمی می‌باشد. جامعه هدف این پژوهش داده‌های مشتریان تسهیلاتی یکی از

بانک‌های کشور می‌باشد. پس از بررسی مطالعات و تجربیات صورت گرفته داده‌های مشتریان بانکی گردآوری و پاکسازی شدند. سپس تعداد خوشه‌ها و روش مناسب برای خوشه‌بندی انتخاب شد و روش‌های مختلف طبقه‌بندی اجرا و ارزیابی شدند. سپس براساس نتایج سیستم توصیه‌گر آماده و پیشخوان کاربران آماده‌سازی شد و در پرتال شیرپوینت بانک نصب گردید.

به‌منظور اجرای این پژوهش داده‌های مشتریان تسهیلاتی یکی از بانک‌های کشور شامل داده‌های پاکسازی‌شده مشتمل بر 85530 رکورد در نرم‌افزار متلب<sup>۲۱</sup> وارد شد. لازم به ذکر است که هر رکورد از این داده‌ها شامل مجموع مانده حساب و مبالغ وام به ازای یک مشتری است و تعداد وام هر مشتری نیز مورد بررسی قرار گرفته است سپس فیلدهای مناسب جهت اجرای خوشه‌بندی انتخاب شد، فیلدهایی که ویژگی‌های لازم برای استفاده در خوشه‌بندی از جمله یکتا بودن را داشتند، به شرح زیر انتخاب شد.

• مانده حساب، نوع وام، اصل وام، سود وام، تعداد وام هر مشتری، مانده جاری، مانده غیر جاری.

سپس فیلدهای انتخاب‌شده با روش Z-Score نرمال‌سازی شده و برای پیاده‌سازی سیستم توصیه‌گر آماده شدند. گام‌های اجرای تحقیق به شرح نمودار ۱ است:



نمودار ۱ گام‌های پژوهش

### ۸- طراحی سیستم توصیه‌گر

طراحی سیستم توصیه‌گر نخست نیازمند شناسایی خوشه‌های موجود و سپس ایجاد امکان پیش‌بینی داده‌های جدید است [15]. در این پژوهش به این منظور ابتدا داده‌ها با استفاده از روش‌هایی که در ادامه خواهد آمد، خوشه‌بندی و سپس طبقه‌بندی شده‌اند.

### ۹- خوشه‌بندی داده‌ها

به منظور خوشه‌بندی داده‌ها نخست باید بهترین روش خوشه‌بندی و تعداد خوشه‌های مناسب انتخاب می‌شود. بنابراین با استفاده از ابزار ارزیابی خوشه‌ها در متلب، روش‌های مختلف خوشه‌بندی شامل میانگین نزدیک‌ترین همسایه، روش پیوندی<sup>۲۲</sup> با استفاده از شاخص سیلوئت<sup>۲۳</sup> ارزیابی و مقایسه شد که روش خوشه‌بندی سلسله مراتبی وارد<sup>۲۴</sup> با هشت خوشه و عدد سیلوئت ۰,۹۹ برای خوشه‌بندی داده‌ها انتخاب گردید.

جدول ۳ بهترین تعداد خوشه براساس روش سیلوئت

تعداد خوشه	پیوندی	میانگین همسایه
۸	۰,۹۹۹۰۹	۰,۹۹۹۰۰

### ۱۰- طبقه‌بندی داده‌ها

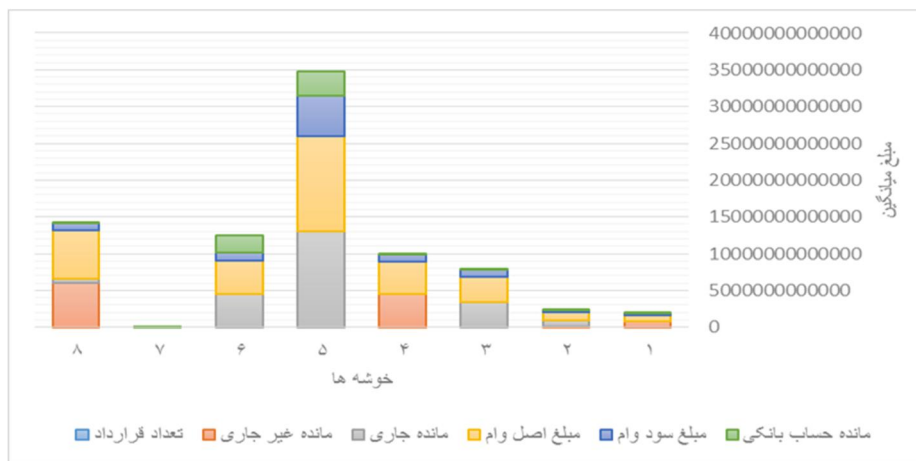
پس از خوشه‌بندی داده‌ها می‌توان برای ایجاد امکان پیش‌بینی به طبقه داده‌ها اقدام کرد. در این صورت سیستم توصیه‌گر می‌تواند داده‌های جدید را در طبقه متناسب با خوشه‌ها قرار دهد. نخست روش‌های درخت طبقه‌بندی<sup>۲۵</sup> و نزدیک‌ترین همسایه<sup>۲۶</sup> با استفاده از الگوریتم‌های ریزابلاس، لاس<sup>۲۷</sup> کافولد<sup>۲۸</sup> ارزیابی شدند که نتیجه ارزیابی در جدول ۴ قابل مشاهده است

جدول ۴ ارزیابی نتایج الگوریتم طبقه‌بندی

روش طبقه‌بندی	نزدیک ترین همسایه	درخت طبقه‌بندی	
خطای جایگزینی (ریزابلاس)	۰/۰۰۱	۰/۰۵۸	۱
خطای طبقه‌بندی	۰/۰۰۰۱	۰/۰۵۸	۲
ارزیابی دو طرفه کافولد	۰/۰۷	۰/۱۱	۳



براساس جدول بالا، روش نزدیک‌ترین همسایه برای طبقه‌بندی داده‌ها انتخاب شد. براین اساس سیستم توصیه‌گر تسهیلات بانکی با استفاده از خوشه‌بندی پیوندی و طبقه‌بندی نزدیک‌ترین همسایه طراحی گردید و نرم‌افزار در پورتال بانک نصب شد. پس از طبقه‌بندی داده‌ها براساس طبقات گروه‌بندی شدند تا ویژگی‌های هر خوشه و شاخص‌های دموگرافیک هر طبقه تحلیل شود. نمودار ۲ طبقه‌ها را براساس ویژگی‌هایی که با آنها آموزش دیده بودند، نمایش می‌دهد.



نمودار ۲ بررسی وضعیت خوشه‌ها

براساس خروجی نرم افزار چنانچه مشتری با اطلاعات مربوط به آن جستجو شود، عدد خوشه براساس طبقه‌بندی انجام شده پیش‌بینی می‌شود. بر مبنای عدد خوشه به دست آمده، امکان انجام تحلیل‌های بسیاری وجود دارد که روی هر یک از فیلدهای ورودی داده و با استفاده از روش‌هایی همچون محاسبه میانگین، میانه، واریانس و انحراف معیار آن فیلد در مقایسه با فیلدهای دیگر قابل بررسی و تحلیل است. در ادامه تحلیل خوشه‌های تسهیلاتی ارائه شده است.

## ۱۱- یافته‌های پژوهش

در این بخش به تشریح ویژگی‌های خوشه‌ها براساس یافته‌های تحقیق در گام پرداخته می‌شود. هدف هر سازمان مالی حفظ مشتریان فعلی و جذب مشتریان جدید

است. [24] و تحلیل خوشه‌هایی که در ادامه بیان خواهند شد می‌توانند در راستای تحقق این هدف برای مدیران موسسات بانکی راهگشا باشند.

## تحلیل خوشه‌ها

۱. خوشه وام‌های کم مبلغ بدون بازگشت سرمایه: این داده‌ها مشتریانی با تعداد وام کمتر و کم مبلغی هستند که بدهی غیر جاری بسیار بیشتر از بدهی جاری و مانده حساب بانکی می‌باشد. این گروه از مشتریان بدهی‌های خود را به‌طور مناسب پرداخت نمی‌کنند و با توجه به اینکه مبالغ وام‌ها کمتر است.

۲. خوشه وام‌های کم مبلغ با بازگشت سرمایه متوسط: مشتریانی هستند که تعداد وام کم و مبالغ کمتری وام گرفته‌اند. در این گروه بدهی غیر جاری نسبت به بدهی جاری کمتر است و نکته قابل توجه آن است که مانده حساب این مشتریان از بدهی غیر جاری آنها بیشتر است. در این صورت، این خوشه از مشتریان علی‌رغم آنکه بدهی غیر جاری دارند، با توجه به توانایی پرداخت از گروه‌های پر خطر به شمار نمی‌آیند اما به‌عنوان یک گروه در معرض آسیب به منابع نیازمند مراقبت دائمی جهت جلوگیری از خروج منابع از بانک هستند.

۳. خوشه مشتریان تازه‌وارد با مانده حساب کم: مشتریان، مشتریانی هستند که به‌تازگی وام دریافت کرده‌اند و تمام مبلغ اصل و سود آنها بدهی جاری این گروه است. این گروه با توجه به مانده حساب بسیار کمتر از مبلغ وام، نیز باید بسیار مراقبت شوند تا به گروه‌های پر خطر تبدیل نشوند.

۴. خوشه مشتریان مخاطره‌آمیز: مشتریانی هستند که مبالغ وام متوسط به بالا و مانده حساب صفر دارند و تمام بدهی ایشان به بدهی غیر جاری تبدیل شده است. این گروه از مشتریان پر خطرترین گروه مشتریان هستند که پس از دریافت وام علاوه بر عدم پرداخت بدهی منابع خود را نیز از بانک خارج کرده‌اند. این گروه را می‌توان گروه منابع از دست رفته بانک یا مشتریان تبه‌کار دانست.

۵. خوشه مشتریان تازه‌وارد با مانده حساب بالا: شامل مشتریان با مانده حساب بالا و مبلغ بدهی غیر جاری صفر و بدهی جاری برابر با مبلغ اصل وام است. این گروه مشتریان خاص وام می‌باشند که تشویق این گروه به افزایش همکاری می‌تواند منافع بانک را افزایش دهد.

۶. خوشه مشتریان تازه‌وارد رده متوسط: این خوشه مشابه خوشه پنجم می‌باشد با این تفاوت که مبالغ وام و مانده حساب کمتر از مشتریان گروه پنجم است. این گروه مشتریان ویژه با سطح درآمد متوسط هستند.

۷. خوشه مشتریان کم درآمد با بازگشت سرمایه کم: این خوشه شامل مشتریان دارای تعداد کم و یا مبالغ کم وام است و بدهی جاری آنها کمتر از بدهی غیر جاری مشتریان است و از سوی دیگر، مانده حساب آنها متوسط بوده و کمتر از بدهی غیر جاری است. این گروه برای جلوگیری از زیان به بانک باید تحت مراقبت بانک باشند.

۸. خوشه مشتریان در حال خارج کردن منابع: هشتم مشتریانی هستند که تعداد وام بالا و بدهی غیر جاری بالایی دارند و نسبت بدهی غیر جاری به جاری در این گروه بالاست، اما با توجه به مانده حساب متوسط این گروه با اقدام‌های مناسب می‌توان از خروج کامل منابع بانک جلوگیری کرد.

پس از پایان مراحل طراحی سیستم، این سیستم در وب سایت بانک پیاده‌سازی و به سرپرستان بانک ارائه گردید که شکل ۱، نمونه‌ای از نتیجه جستجو در این سامانه را نشان می‌دهد. سرپرستان نیز با تأیید کارکرد و خروجی سیستم، نکاتی را جهت بهبود و ارتقای آن در اختیار محققان قرار دادند.



شکل ۱ ارزیابی مشتری با استفاده از سامانه توصیه‌گر

در این سامانه با وارد کردن کد مشتری، خوشه مربوط به آن مشتری و تحلیل مرتبط با خوشه به مدیر شعبه ارائه می‌شود تا بر مبنای آن تصمیم‌گیری لازم را انجام دهد.

## ۱۲- نتیجه‌گیری

در این پژوهش مشتریان تسهیلاتی بانک مورد نظر خوشه‌بندی شده و خوشه‌های آن شناسایی شدند. سپس با طبقه‌بندی امکان پیش‌بینی داده‌های جدید به وجود آمد. این سیستم آمادگی پیش‌بینی خوشه و کمک به مدیران مربوط جهت اتخاذ تصمیم‌های فوری برای مشتریان تسهیلاتی را دارد. این سیستم می‌تواند نوع مشتری تسهیلاتی را پیش‌بینی کند، بنابراین می‌توان سایر روش‌های داده‌کاوی را نیز مورد آزمون و بررسی قرار داد. همچنین با وجود فیلدهای اطلاعاتی مانند زمان و تعداد اقساط معوق و دیگر فیلدهای مفید و کاربردی، می‌توان از این سیستم توصیه‌گر به منظور شناسایی مشکلات بالقوه استفاده نمود. همچنین این سیستم در بانک‌ها و مؤسسات دیگر قابلیت پیاده‌سازی و تحلیل نتایج را دارد. از این سیستم می‌توان برای اتخاذ برنامه‌های بازاریابی و وصول مطالبات استفاده کرد. همچنین ارائه سیستم توصیه‌گر مبلغ وثیقه‌ها می‌تواند فرایند اعطای تسهیلات و وصول مطالبات را تسریع کند. در پژوهش حاضر ارقام اطلاعاتی از مشتریان و اطلاعات فردی آنها به‌طور کامل در دسترس نبود. همچنین سیستم حاضر تنها می‌تواند وضعیت مشتریانی را بررسی کند که سابقه دریافت وام داشته باشند؛ به عبارتی مشتری فعلی بانک باشند. اگر ارقام اطلاعاتی افزایش پیدا کند، نمونه توسعه‌یافته سیستم با ارقام جدید داده می‌تواند مشتریانی را که به‌تازگی به بانک آمده‌اند، نیز ارزیابی کند که پیشنهاد می‌شود در پژوهش‌های آینده به طراحی این سیستم نیز مبادرت شود.

## ۱۳- پی‌نوشت‌ها

1. Data mining
2. Recommender system
3. Clustering
4. Classification
5. Rule-like knowledge
6. Regression

7. Association rules
8. Rule generation
9. Summarization
10. Dependency modeling
11. Sequence analysis
12. Data cleaning
13. Data reduction
14. Transformation (data integration)
15. Pattern evaluation
16. K-Means
17. Ward
18. KNN
19. Collaborative filtering
20. Content-based filtering
21. MATLAB
22. Linkage
23. Silhouette
24. Ward
25. Classification tree
26. Nearest Neighbors (KNN)
27. Loss
28. KFold

#### ۱۴- منابع

- [1] Mehta J. S., Gawande A. (2015) "A purpose of data mining in banking sector", *International Journal of Advance Research in Computer Science and Management Studies*, 3(3).
- [2] Ramos S., Duarte J. M., Duarte F. J., Vale Z. (2015) "A data-mining-based methodology to support MV electricity customers' characterization", *Energy and Buildings*, 91": 16-25.
- [3] Chen C. C., Chen A. P. (2007)"Using data mining technology to provide a recommendation service in the digital library", *The Electronic Library*, 25(6): 711-724. doi: doi:10.1108/02640470710837137.
- [4] Afsar A., Houshdar Mahjoub R., Minaie Bidgoli B. (2014) "Customer credit clustering for presenting appropriate facilities", *Management Researches in Iran*, 17(4):1-24.

- [5] Mahdavi K., Horri M.S., (2015) "Designing a model for predicting the credit ranking of bank customers using meta-heuristic algorithm and multi-criteria hybrid neuro-fuzzy and ant colony (Case Study of Post Bank Branches of Tehran), *Management Researches in Iran*,
- [6] Sajjadi K., Khatami-Firuzabadi M. A., Amiri M., Sadaghiani J. S. (2015) "A developing model for clustering and ranking bank customers", *International Journal of Electronic Customer Relationship Management*, 9(1): 73-86. doi: doi:10.1504/IJECRM.2015.070701.
- [7] Ngai E., Hu Y., Wong Y., Chen Y., Sun X. (2011) "The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature", *Decision Support Systems*, 50(3): 559-569.
- [8] Ramos S., Duarte J. M., Duarte F. J., Vale Z. (2015) "A data-mining-based methodology to support MV electricity customers' characterization", *Energy and Buildings*, 91": 16-25.
- [9] Q. He, (1999) "A review of clustering algorithms as applied in IR", Graduate School of Library and Information Science University of Illinois, at Urbana-Champaign.
- [10] Jianying M., Yongjian F., Yanguang S. (2009) A neural networks-based clustering collaborative filtering algorithm in E-commerce recommendation system, *Paper presented at the Web Information Systems and Mining, WISM 2009*, International Conference.
- [11] Das J., Mukherjee P., Majumder S., Gupta P. (2014), "Clustering-based recommender system using principles of voting theory", *Paper Presented at the Contemporary Computing and Informatics (IC3I) Conference*, Amity University, Noida, India.
- [12] Asosheha A., Bagherpour S., Yahyapour N. (2008) "Extended acceptance models for recommender system adaption, Case of retail and banking service in Iran", *WSEAS Transactions on Business and Economics*, 5(5):189-200.

- [13] Zahra S., Ghazanfar M. A., Khalid A., Azam M. A., Naeem U., Prugel-Bennett A. (2015) "Novel centroid selection approaches for KMeans-clustering based recommender systems", *Information Sciences*, 320, 156-189. doi: <http://dx.doi.org/10.1016/j.ins.2015.03.062>.
- [14] Shani G., Gunawardana A. (2011) "Evaluating recommendation Systems", In F. Ricci, L. Rokach, B. Shapira, & P. B. Kantor (Eds.), *Recommender Systems Handbook* (pp. 257-297): Springer, US.
- [15] Guo, G., et al. (2015). "Leveraging multiviews of trust and similarity to enhance clustering-based recommender systems." *Knowledge-Based Systems* 74: 14-27.
- [16] Hsieh N.C. (2004) An integrated data mining and behavioral scoring model for analyzing bank customers. *Expert Systems with Applications*, 27(4): 623-633.
- [17] Kim J.B., Song B. Y., Zhang Y. (2015) "Earnings performance of major customers and bank loan contracting with suppliers", *Journal of Banking & Finance*, 59:384-398. doi: <http://dx.doi.org/10.1016/j.jbankfin.2015.06.020>
- [18] Anh N. D. (2015) *Adaptive neuro-Fuzzy network for recommendation*, International University HCMC, Vietnam.
- [19] de Oña R., de Oña J. (2015) "Analysis of transit quality of service through segmentation and classification tree techniques", *Transportmetrica A: Transport Science*, 11(5): 365-387, doi: 10.1080/23249935.2014.1003111.
- [20] Hsieh N.C. (2004) An integrated data mining and behavioral scoring model for analyzing bank customers. *Expert Systems with Applications*, 27(4): 623-633.
- [21] Schetin V., Jakaite L., Jakaitis J., Krzanowski W. (2013) "Bayesian decision trees for predicting survival of patients: A study on the US national trauma data bank", *Computer Methods and Programs in Biomedicine*, 111(3): 602-612. doi: <http://dx.doi.org/10.1016/j.cmpb.2013.05.015>
- [22] Seret A., Vanden Broucke S. K. L. M., Baesens B., Vanthienen J. (2014) "A dynamic understanding of customer behavior processes based on clustering and sequence mining", *Expert Systems with Applications*, 41(10), 4648-4657. doi: <http://dx.doi.org/10.1016/j.eswa.2014.01.022>.

- [23] Yang X., Chen J., Hao P., Wang Y. J. (2015) *Application of clustering for customer segmentation in private banking*, Seventh International Conference on Digital Image Processing (ICDIP 2015), 96311Z (July 6, 2015); doi:10.1117/12.2197182.
- [24] Islam M., Habib M. (2015) A data mining approach to predict prospective business sectors for lending in Retail banking using decision tree, arXiv preprint arXiv:1504.02018.