شناسایی الگوی رفتاری مشتریان در بیمه عمر و تشکیل سرمایه با استفاده از داده‌کاوی[`]

امیر بروفر[۱]، علی رضائیان[۲*]، سجاد شکوهیار[۳]

۱. دانشجوی کارشناسی ارشد مدیریت فناوری اطلاعات، دانشگاه شهید بهشتی، تهران، ایران
۲. استاد، ریاست دانشکده مدیریت و حسابداری، دانشگاه شهید بهشتی، تهران، ایران
۳. استادیار گروه مدیریت فناوری اطلاعات، دانشکده مدیریت و حسابداری، دانشگاه شهید بهشتی، تهران، ایران

دریافت: ۹۴/۰۸/۰۹ پذیرش: ۹۵/۰۷/۱۴

چکیده

امروزه تعامل شرکت‌ها با مشتریان در قالب مدیریت ارتباط با مشتری به طور قابل توجهی تغییر یافته است. شناسایی ویژگی‌های مشتریان مختلف و تخصیص بهینه منابع به آنها با توجه به ارزشی که برای شرکت‌ها دارند، به یکی از دغدغه‌های اصلی در حوزه مدیریت ارتباط با مشتری تبدیل شده است. هدف این مقاله ارائه مدل مناسبی جهت بخش‌بندی مشتریان براساس برخی از مهم‌ترین ویژگی‌های مالی، جمعیت شناختی در قالب عوامل مؤثر بر شاخص‌های ارزش دوره عمر مشتری (آر.اف.ام) می‌باشد. در فرایند پیشنهادی این تحقیق که در شرکت بیمه سامان اجرا شده است، پس از تعیین مقادیر شاخص‌های مدل آر.اف.ام (RFM) شامل تازگی مبادله، تعداد دفعات مبادله و ارزش پولی مبادله در ۱۸۰۰۰۰ مشتری و وزن‌دهی آنها با استفاده از فرایند تحلیل سلسله مراتبی، تعداد خوشه بهینه براساس شاخص سیلوئت و نرخ تأثیر شاخص‌های آر.اف، ام با استفاده از الگوریتم Two-step انجام شد و در مرحله بعد به خوشه‌بندی مشتریان با استفاده از روش K-means پرداخته شده است. نتایج مطالعه حاضر، زمینه را برای تحلیل ویژگی‌های مشتریان شرکت در سه بخش اصلی فراهم نمود. همچنین با اولویت‌بندی خوشه‌ها براساس شاخص‌های آر.اف.ام، مشتریان کلیدی و با ارزش شرکت مشخص شدند. درنهایت نیز پیشنهادهایی به شرکت برای بهبود سیستم مدیریت ارتباط با مشتری ارائه گردید.

واژه‌های کلیدی: مدیریت ارتباط با مشتری، بخش‌بندی مشتریان، ارزش دوره عمر مشتری، مدل آر.اف.ام (RFM)، خوشه‌بندی کا- میانگین.

---

*نویسندهٔ مسؤول مقاله: E-mail: A-rezaeian@sbu.ac.ir

۱- مقدمه

در دهه‌های اخیر توانایی بشر برای تولید و جمع‌آوری داده‌ها به‌سرعت افزایش پیدا کرده است. عواملی نظیر استفاده گسترده از توانایی فناوری اطلاعات، تجهیزات آزمایشگاهی، پیشرفت درجمع‌آوری داده‌ها و سیستم‌های سنجش از راه دور ماهواره‌ای، در این تغییرات نقش مهمی داشته‌اند. این رشد انفجاری در داده‌های ذخیره شده باعث پیدایش فناوری جدیدی شده تا این حجم داده را به اطلاعات و دانش تبدیل کند. داده‌کاوی به عنوان یک راه‌حل برای این مسائل مطرح می‌باشد [۱]. داده‌کاوی به بهره‌گیری از ابزارهای تجزیه و تحلیل داده‌ها به منظور کشف الگوها و روابط معتبری که تاکنون ناشناخته بوده‌اند، اطلاق می‌شود [۲]. درصنعت بیمه، مراحل تشخیص مشتری و جذب وی، فعالیت‌هایی زمانگیر و پرهزینه‌ای هستند که معمولاً شرکت‌ها این وظایف را به عهده نمایندگان می‌گذارند. اما می‌توان به راحتی و از طریق تکنیک‌های داده‌کاوی مراحل حفظ، جذب مشتریان جدید و حتی گسترش ارتباط با مشتری را به راحتی شناسایی کرد. در این مقاله بعد حفظ مشتری از توجه خاصی برخوردار است. بعد جذب مشتری که روزی اولین برگ برنده در فعالیت‌های بازاریابی به شمار می‌آمد، جای خود را به حفظ مشتری و تلاش برای گسترش ارتباط با او داده است. از جمله تجارب ارزشمند سازمان‌ها که به صورت آماری بیان شده‌اند تا اهمیت حفظ مشتری را بیان کنند، از این قرارند :

-هزینه فروش کالا به یک مشتری جدید، شش برابر هزینه فروش به مشتری قدیمی است.

-معمولاً هر مشتری ناراضی، عدم رضایت خود را با ۸ الی ۱۰ نفر در میان می‌گذارد.

-اگر شرکتی بتواند میزان نگهداری و حفظ مشتری سالیانه خود را ۵ درصد افزایش دهد، می‌تواند منافع و سود خود را بین ۳۰ تا ۱۲۵ درصد افزایش دهد [۳].

بر اساس پژوهش‌های [٤] می‌توان مشتریان را براساس رضایتی که از شرکت و محصول کسب می‌کنند در چهار دسته زیر قرار داد: دسته اول مشتریانی هستند که هم از شرکت و هم از محصول راضی هستند و مشتریان وفادار نامیده می‌شوند. دسته دوم کسانی هستند که از محصول راضی هستند اما از نحوه ارتباط خود با شرکت ناراضی هستند، این مشتریان از دیدگاه شرکت در معرض خطر هستند و هر لحظه امکان دارد جذب سازمان دیگری شوند که با آنها رابطه خوبی برقرار کند. سومین دسته، مشتریان خطرناک هستند، آنها نه از محصول رضایت دارند و نه از سازمان، پس هر لحظه امکان دارد با ارتباط مستقیم خود با دیگران وجهه

سازمان را در نظر مردم خراب کنند و چهارمین دسته مشتریان امیدوار نامیده می‌شوند که با وجود نارضایتی از محصول از ارتباط خود با سازمان راضی هستند و انتظار دارند در خرید بعدی رضایتشان جلب شود. هدف این است مشتریان در حال ریزش را تشخیص داده و برای تداوم ارتباط با آنها کوشید. براساس تحقیقاتی که دو محقق به نام‌های [٥] انجام دادند با ٥ درصد افزایش حفظ مشتری در سازمان می‌توان بین ۳۵ تا ۹۵ درصد ارزش مشتری فعلی در سازمان را افزایش داد، برای مثال با ٥ درصد افزایش در حفظ مشتری، در شرکت‌های بیمه‌ای به طور کلی ٨٤ درصد و به طور اختصاصی ۹۰ درصد ارزش مشتری فعلی را افزایش داد. برای آنکه بتوان مشتریان وفادار را در سازمان حفظ کرد و به مشتریان در معرض خطر، خدمات بهتری ارائه داد باید در اولین قدم آنها را شناخت، با شناخت هر دسته از مشتریان می توان متناسب با نیاز آنها خدمات شرکت را بهبود داد و به این صورت مشتریان را حفظ کرد.

در بازارهای رقابتی امروزی، با گرایش شرکت ها به سمت مشتری‌مداری، مدیریت ارتباط با مشتری نیز به سمت پیچیدگی‌های خاصی گرایش پیدا کرده است. براساس مطالعات گذشته تخمین زده شده است که هزینه‌های جذب مشتریان جدید، پنج برابر هزینه‌های حفظ مشتریان موجود خواهد بود [٦]. از طرفی بسیاری از مدیران معتقدند که شرکت نباید برای به دست آوردن هر مشتری در هر سطح از سودآوری، هزینه پرداخت کند، بلکه باید منابع محدود خود را در جهت کسب و نگهداری مشتریان کلیدی شرکت به صورت بهینه صرف نماید [٧]. از سویی دیگر، تمرکز شرکت‌های امروزی تنها بر فروش کالاهایشان نیست، آنها در پی خلق و حفظ مشتریان سودآور هستند .اما سؤال اصلی این است که چگونه می‌توان مشتریان کلیدی و سودآور شرکت را شناسایی کرد؟ شرکت‌ها می‌توانند با بخش‌بندی مشتریان به گروه‌های مختلف برساس معیارهایی خاص به شناسایی و تحلیل ویژگی‌های رفتاری آنها بپردازند. این کار، زمینه تخصیص بهینه منابع محدود، به‌کارگیری استراتژی‌های مناسب بازاریابی و درنهایت مدیریت سودآوری در کنار مدیریت ارتباط با مشتری را فراهم می‌آورد. ارزش دوره عمر مشتری (CLV)، مفهومی است که می‌تواند به شرکت‌ها در این راستا کمک فراوانی کند. این مفهوم بیانگر ارزشی است که مشتری در طول دوره عمرش برای سازمان ایجاد می‌کند و هدف اصلی از محاسبه آن، ایجاد یک برداشت وزنی از مشتریان بر براساس ارزشی فعلی و بالقوه‌هایی است که برای شرکت دارند که با استفاده از مدل‌های مختلف تعیین می‌شود. نتایج

حاصل از داده‌کاوی کمک می‌کند تا سازمان‌هایی نظیر بیمه مرکزی و شرکت‌های بیمه‌ای با شناخت هر دسته از مشتریان متناسب با نیاز آنها خدمات شرکت را بهبود داده و به این صورت مشتریان را حفظ کنند. پس از تشخیص مشتریان، مشخصات آنها، خواسته‌ها، نیازها و انتظارهای هر طبقه و سپس یافتن فرصت‌ها و تهدیدها سودآوری شرکت را بالا ببرند.

در این تحقیق قصد داریم به سؤال‌های زیر پاسخ دهیم :

۲- سؤال اصلی

۱-نتایج به دست آمده از داده‌کاوی بر روی داده‌های مشتریان بیمه عمر چه تأثیری بر روی بهبود سیاست‌های مدیریت و ارتباط با مشتری دارد؟

۳- سؤال فرعی

۱-آیا استفاده از داده‌کاوی به تصمیم‌سازی مدیران شرکت بیمه سامان کمک می‌نماید ؟

خوشه‌بندی کاربران در سازمان‌ها به‌وسیله داده‌کاوی باعث می‌شود که مدیران بتوانند اطلاعات آنها را تحلیل نمایند؛ الگوی کاربران را پیدا کنند؛ ارتباط بین سازمان و کاربران بهبود پیدا کند تا درنهایت کاربران احساس رضایت نمایند [۸]. اطلاع از این موضوع نیز از آنجا اهمیت دارد که خوشه‌بندی کاربران براساس تقاضای آنها و شناخت نیاز کاربران، موجب خواهد شد تا مدیریت در جهت جلب رضایت کاربران گام برداشته شود و سرمایه‌های سازمانی به‌درستی استفاده شود. به این ترتیب از تمرکز روی جمع‌آوری داده‌های کم اهمیت جلوگیری می‌شود و برای داده‌های با اهمیت‌تر، سرمایه‌گذاری بیشتری خواهد شد. عدم وجود این شناخت باعث خواهد شد تا سرمایه‌های سازمان در پی داده‌های ناکارآمد اتلاف شود.

۴- پیشینه پژوهش

پیشینه پژوهش در دو بخش انجام گرفته است. در پیشینه نظری به تعاریف پایه، مباحث مدیریت ارتباط با مشتری، داده‌کاوی، ارزش دوره عمر مشتری (CLV) و مدل RFM پرداخته شده است و در پیشینه تجربی، پژوهش‌های مرتبط انجام شده با استفاده از تکنیک داده‌کاوی و

الگوریتم‌های خوشه‌بندی مرور شده است.

۵- پیشینه نظری

۵-۱- بیمه عمر و تشکیل سرمایه

بیمه عمر تشکیل سرمایه براساس مدل شناخته شده جهانی Flexible Universal Life طراحی شده است و در سراسر جهان یکی از پرطرفدارترین بیمه‌های عمر می‌باشد. بیمه عمر تشکیل سرمایه قرارداد بیمه‌ای است که همزمان با بیمه کردن عمر شما امکانات متنوع سرمایه‌گذاری ذخایر حق بیمه پرداختی شما را در سبد سرمایه‌گذاری شرکت بیمه، متشکل از مطمئن‌ترین بازارهای سرمایه‌گذاری، در دسترس شما قرار می‌دهد تا در صورت تحقق خطر فوت و یا پس از اتمام قرارداد شما و خانواده از تأمین مالی مورد نیاز برخوردار شوید.

۵-۲- مدیریت ارتباط با مشتری

[۹] مدیریت ارتباط با مشتری را فرایند راهبردی انتخاب مشتریان با سودآوری بالا و برقراری تعامل با آنها با هدف بهینه‌سازی ارزش جاری و آینده مشتریان برای شرکت تعریف کرده‌اند. [۱۰] بیان می‌کنند که تعاریف جدید بر اهمیت مدیریت ارتباط با مشتری به عنوان یک فرایند جامع و استراتژیک برای حداکثرسازی ارزش مشتری برای سازمان تأکید بسیاری دارند. [۱۱] بیان می‌کنند که در مدیریت ارتباط با مشتری تحلیلی که به تحلیل ویژگی‌های رفتاری مشتری برای حمایت از استراتژی‌های مدیریت ارتباط بامشتری می‌پردازد، معمولاً از ابزارهای داده‌کاوی برای انجام تحلیل‌های مختلف استفاده می‌شود.

۵-۳- داده‌کاوی

تعاریف مختلفی از داده‌کاوی ارائه شده است که در ادامه به چند مورد خواهیم پرداخت:

-داده‌کاوی فعالیتی در راستای جستجوی ذخایر داده‌ای بزرگ در جهت کشف الگوها و شیوه‌هاست که فراتر از تحلیل‌های اولیه و پایه هستند. داده‌کاوی از الگوریتم‌های پیچیده ریاضی جهت دسته‌بندی داده‌ها و ارزش‌گذاری احتمال وقایع آنی، استفاده می‌کند.

مشخصات کلیدی داده‌کاوی به شرح زیر است:

۱-کشف خودکار الگوها

۲-پیش‌بینی خروجی‌های محتمل

۳-ساخت اطلاعات قابل استناد

۴-تمرکز بر مجموعه‌های داده‌ای و پایگاه داده‌های بزرگ

با توجه به اینکه تمرکز اصلی این پروژه در بین روش‌های داده‌کاوی بر خوشه‌بندی است، ادامه مفاهیم آن را به تفضیل بیشتر شرح خواهیم داد.

۵-۴- خوشه‌بندی

خوشه‌بندی را می‌توان به عنوان مهم ترین مسئله در یادگیری بدون نظارت در نظر گرفت. خوشه‌بندی یافتن یک ساختار درون یک مجموعه از داده‌های بدون برچسب است. خوشه به مجموعه‌ای از داده‌ها گفته می‌شود که به هم شباهت داشته باشند. در خوشه‌بندی سعی می‌شود تا داده‌ها به خوشه‌هایی تقسیم شوند که شباهت بین داده‌های درون هر خوشه حداکثر و شباهت بین داده‌های درون خوشه‌های متفاوت حداقل شود.

در ادامه به توضیح چند الگوریتم خوشه بندی که در این تحقیق مورد استفاده قرار گرفته‌اند، پرداخته می‌شود:

۵-۵- روش خوشه‌بندی K-Means

این روش، یک روش پایه برای بسیاری از روش‌های خوشه‌بندی دیگر (مانند خوشه‌بندی فای) محسوب می‌شود. این روش، روشی انحصاری و مسطح محسوب می‌شود [۱۲].

برای این الگوریتم شکل‌های مختلفی بیان شده است. ولی همه آنها دارای روالی تکراری هستند که برای تعدادی ثابت از خوشه‌ها سعی در تخمین موارد زیر دارند.:

۱- به دست آوردن نقاط مرکزی دسته ها، این نقاط در واقع همان میانگین نقاط متعلق به هر دسته هستند؛

۲- نسبت دادن هر نمونه داده به یک خوشه که آن داده کمترین فاصله تا مرکز آن خوشه را دارا باشد؛

در نوع ساده‌ای از این روش نخست به تعداد خوشه‌های مورد نیاز، نقاطی به صورت تصادفی انتخاب می‌شوند. سپس داده‌ها با توجه با میزان نزدیکی (شباهت) به یکی از این خوشه‌ها نسبت داده شده و به این ترتیب خوشه‌های جدیدی حاصل می‌شود. با تکرار همین روال می‌توان در هر تکرار با میانگین‌گیری از داده‌ها مراکز جدیدی برای آنها محاسبه کرد و دوباره داده‌ها را به خوشه‌های جدید نسبت داد. این روند تا زمانی ادامه پیدا می‌کند که دیگر تغییری در داده‌ها حاصل نشود. تابع زیر به عنوان تابع هدف مطرح است. رابطه (۱) الگوریتم پایه برای این روش محسوب می‌شود:

$$J = \sum_{j-1}^{k} \sum_{1-1}^{x} \left\| x_i^{(j)} c_j \right\|^2$$ رابطه(۱)

۱- در آغاز k نقطه به عنوان نقاط مراکز خوشه‌ها انتخاب می‌شوند.

۲- هر نمونه داده به خوشه‌ای که مرکز آن خوشه کمترین فاصله تا آن داده را داراست، نسبت داده می‌شود.

۳- پس از تعلق تمام داده‌ها به یکی از خوشه‌ها برای هر خوشه یک نقطه جدید به عنوان مرکز محاسبه می‌شود (میانگین نقاط متعلق به هر خوشه).

۴- مراحل ۲و۳ تکرار می‌شوند تا زمانی که دیگر هیچ تغییری در مراکز خوشه‌ها حاصل نشود.

مشکلات روش خوشه‌بندی K-Means

علی‌رغم اینکه خاتمه‌پذیری الگوریتم بالا تضمین شده است ولی جواب نهایی آن واحد نبوده و همواره جوابی بهینه نمی‌باشد. به طور کلی روش ساده بالا دارای مشکلات زیر است:

۱- جواب نهایی به انتخاب خوشه‌های اولیه وابستگی دارد.

۲- روالی مشخص برای محاسبه اولیه مراکز خوشه‌ها وجود ندارد.

۳- اگر در تکراری از الگوریتم تعداد داده‌های متعلق به خوشه‌ای صفر شد، راهی برای تغییر و بهبود ادامه روش وجود ندارد.

٤- مدل حاصل بستگی به ترتیب داده‌های آموزشی دارد. مرتب‌سازی مجدد داده‌ها و ساخت مجدد مدل ممکن است به نتیجه نهایی متفاوتی منجر شود.

۵-۶- خوشه‌بندی براساس چگالی

این روش‌های خوشه‌بندی بر این اصل استوارند که خوشه‌ها، ناحیه‌هایی از فضای داده با چگالی زیادی هستند که به‌وسیله نواحی با چگالی کمتر از همدیگر جدا شده‌اند [۱۳].

مزایای خوشه‌بندی براساس چگالی

۱-خوشه‌ها می‌توانند دارای اشکال دلخواه باشند.

۲-تعداد خوشه‌ها به صورت اتوماتیک همزمان با عمل خوشه‌بندی تعیین می‌شود.

۵-در تشخیص نویزها بسیار کارا هستند.

۵-۷- خوشه‌بندی دو مرحله‌ای

این نوع از خوشه‌بندی همچون سایر انواع آن در راستای افراز مجموعه‌های داده‌ای به گروه‌های مجزا است در وضعیتی که این گروه‌ها و ویژگی‌های آن از آغاز مشخص نیستند. درست مانند کوهنن و K-means در این مدل نیز فیلد هدف مشخص نمی‌شود.

به جای تلاش برای پیش‌بینی خروجی، این مدل خوشه‌بندی سعی در یافتن الگوهای موجود مجموعه داده‌های ورودی را دارد.

این مدل یک روش دو مرحله‌ای است، در مرحله اول یک پیمایش کلی روی داده‌ها انجام می‌شود که در طی آن داده‌های خام ورودی در قالب زیر خوشه‌های تقسیم می‌شوند. در مرحله دوم روش خوشه‌بندی سلسله مراتبی بر روی داده‌ها اعمال می‌شود که زیرخوشه‌ها را با هم ادغام می‌کند تا بدون نیاز به پیمایش مجدد، به خوشه‌های بزرگ‌تری برسد. مزیت خوشه‌بندی سلسله مراتبی در این است که در ابتدا نیازی به مشخص کردن تعداد خوشه‌ها وجود ندارد. اگرچه اعمال این فرضیه برای مجموعه‌های داده ای بزرگ دشوار است، خوشه‌بندی اولیه مدل دو مرحله‌ای، خوشه‌بندی سلسله مراتبی را حتی برای مجموعه‌های داده‌ای بزرگ سریع‌تر می‌کند.

از نقاط ضعف این نوع خوشه‌بندی این است که مقادیر مفقود را در نظر نمی‌گیرد. رکوردهای با مقادیر تهی، به عنوان ورودی در زمان ساخت مدل، رد می‌شوند.

نقاط قوت : خوشه‌بندی دو مرحله‌ای می‌تواند از داده‌های با نوع ترکیبی نیز استفاده کند و قابل استفاده بر مجموعه‌های داده‌ای بزرگ نیز می‌باشد. همچنین قادر به بررسی روش‌های

خوشه‌بندی مختلف و انتخاب بهترین آنها نیز هست، بنابراین نیازی به دانستن و مشخص کردن تعداد خوشه‌های هدف وجود نخواهد داشت. این الگوریتم می‌تواند طوری تنظیم شود که به صورت خودکار داده‌های پرت یا غیرمعمول را که ممکن است نتیجه را مغشوش کنند، حذف نماید.

۵-۸- ارزش دوره عمر مشتری

ارزش دوره عمر مشتری با عناوینی نظیر ارزش مشتری، ارزش دوره عمر، حقوق مشتری و سودآوری مشتری در مطالعات متعدد مورد بررسی قرار گرفته است. به طور کلی می‌توان گفت ارزش دوره عمر مشتری، ارزشی است که مشتری در طول دوره عمر خود برای سازمان ایجاد می‌کند. این مفهوم علاوه بر ارزش فعلی مشتریان، به ارزش بالقوه و آتی آنها برای شرکت نیز اشاره دارد و هدف اصلی از محاسبه آن، ایجاد یک برداشت وزنی از مشتریان به‌منظور تخصیص بهینه منابع به آنها می باشد. وجود تعاریف مختلف از ارزش دوره عمر مشتری، بیانگر دیدگاه‌ها و روش‌های متفاوتی است که نسبت به این موضوع وجود دارد. متداول‌ترین روش‌های مطرح شده برای تعیین ارزش دوره عمر مشتری عبارتند از روش ارزش فعلی خالص، روش سهم کیف پول، روش زنجیره مارکوف، روش ارزش گذشته مشتری، روش بازگشت سرمایه و روش آر.اف.ام. در میان روش‌های ذکر شده مدل آر.اف.ام یکی از روش‌های متداول و پرکاربردی است که در تعیین ارزش مشتری، سه معیار (شاخص) را در نظر می گیرد، از این رو دیدگاهی چند بعدی در این راستا ارائه می‌نماید، در صورتی که بسیاری از روش‌های دیگر دیدگاهی تک بعدی داشته و معمولاً از یک معیار (شاخص) برای تعیین ارزش دوره عمر مشتری استفاده می‌کنند. از طرفی، در مدل آر.اف.ام، فقط نگرش‌های مالی مطرح نمی‌باشند و گرایش اصلی روش در تحلیل ویژگی‌های مشتریان به سمت مسائل غیرمالی است، در صورتی که بسیاری از روش‌های دیگر، بیشتر از جنبه مالی بر این مسئله تمرکز دارند. با توجه به توضیحات اشاره شده، مدل آر.اف.ام که مورد استفاده تحقیق حاضر نیز می‌باشد، در ادامه مورد بحث قرار می‌گیرد.

۵-۹- مدل آر.اف.ام

مدل آر.اف.ام اولین بار توسط [١٤] معرفی گردید. وی برای تحلیل آر.اف.ام از رفتار گذشته مشتری که به آسانی قابل پیگیری و دسترسی است، استفاده نمود. این مدل از سه بعد مربوط به داده‌های مبادلاتی مشتریان برای تحلیل رفتار آنها استفاده می‌کند. شاخص‌های این مدل به صورت زیر تعریف می‌شوند:

۱- تازگی مبادله : این شاخص اشاره دارد بر فاصله زمانی بین آخرین خرید صورت گرفته به‌وسیله مشتری تا پایان دوره خاص (پایان محدوده زمانی مورد بررسی). کمتر بودن این فاصله نشانگر بالا بودن ارزش این شاخص در مدل می‌باشد.

۲- تعداد تکرار مبادله : این شاخص بیانگر تعداد مبادلاتی است که یک مشتری در یک دوره زمانی خاص انجام داده است. بیشتر بودن تعداد مبادلات، نشانگر بالا یودن ارزش این شاخص در مدل می‌باشد.

۳- ارزش پولی مبادله : این شاخص نشان‌دهنده مقدار پولی است که یک مشتری در یک دوره زمانی خاص برای مبادلات، صرف کرده است. بیشتر بودن مقدار پول صرف شده، بیانگر بالا بودن ارزش این شاخص در مدل می‌باشد.

۴- در مدل آر.اف.ام.، ارزش دوره عمر هر مشتری مجموع ارزش‌های حاصل از شاخص‌های آر.اف.ام. به دست می‌آید، از این رو در این مدل، فرض بر این است که مشتریان دارای ارزش بالای هریک از شاخص‌های مدل، بهترین مشتریان هستند، البته تا زمانی که در آینده همانند گذشته رفتار نمایند.

۵- نظرات مختلفی پیرامون اهمیت شاخص‌های مدل آر.اف.ام. وجود دارند. [١٤] بیان می‌کند که این سه شاخص اهمیت یکسانی دارند، بنابراین وزن هر سه آنها مشخص و یکسان است. از سویی دیگر، استون بر این عقیده است که به‌واسطه ویژگی‌های متفاوت در هر صنعت، سه شاخص اهمیت متفاوتی خواهند داشت. وی در تحقیق خود، وزن‌های شاخص‌های آر.اف.ام. را به صورت قضاوت ذهنی تعیین نمود. بعدها، لیو و شیه از فرایند تحلیل سلسله مراتبی برای تصمیم‌گیری بهتر جهت تعیین وزن‌های نسبی شاخص‌های آر.اف.ام. در ارزیابی ارزش دوره عمر مشتری بهره گرفتند.

۶- پیشینه تجربی

در مورد داده کاوی و کاربردهای آن پژوهش‌های فراوانی در دنیا انجام شده است. این تحقیقات علاوه بر مفاهیم کشف دانش و داده‌کاوی، روشها و الگوریتم‌های آن به بررسی کاربردهای آن در حوزه‌های مختلف بانکداری، بیمه، مدیریت ارتباط با مشتری، پزشکی، خرده‌فروشی، فضا و سفرهای فضایی و غیره پرداخته‌اند [۱۵] در داده‌کاوی معمولاً به کشف الگوهای مفید از میان داده‌ها اشاره می شود. منظور از الگوی مفید، مدلی در داده‌ها است که ارتباط میان یک زیر مجموعه از داده‌ها را توصیف می‌کند و معتبر، ساده، قابل فهم و جدید است [١٦]. تحلیل داده‌های تصادفات با استفاده از فنون داده‌کاوی مورد توجه بسیار زیادی از سوی محققان قرارگرفته است. تکنیک‌های استفاده شده در بیشتر این پژوهش‌ها با استفاده از تحلیل پیش‌بینی کننده، یعنی نگاشت مجموعه‌ای از ورودی‌ها به خروجی خاصی اشاره دارند. از سوی دیگر تحلیل پیش‌بینی کننده برای کشف مجموعه‌ای از اشیا داده (مشاهدات یا متغیرها) مبتنی بر شباهت یا عدم شباهت در میان این اشیا انجام شده‌اند. در ساختار پیشینه تحقیق به پژوهش‌هایی که درزمینه کاربردهای داده‌کاوی، تکنیک‌ها و الگوریتم‌ها، تکنیک خوشه‌بندی و به‌کارگیری الگوریتم‌های داده‌کاوی در مباحث مختلف پییش‌بینی الگوها در صنعت بیمه انجام شده است، پرداخته می‌شود.

در تحقیقی توسط [۱۷] سعی شده است تا با به‌کارگیری تکنیک‌های داده کاوی و مدل DFMT ارزش‌گذاری مشتریان مشخص شود. داده‌ها با استفاده از الگوریتم K-means خوشه‌بندی می‌شوند و پس از آن با استفاده از شاخص SSE تعداد بهینه خوشه تعیین می‌گردد. در ادامه هرم ارزش مشتری تشکیل می‌شود که مشتریان را در ٤ طیف ارزشی دسته‌بندی می‌کند. در تحقیقی توسط [۱۸] هدف، خوشه‌بندی و کشف الگوهای خسارت مشتریان بیمه شخص ثالث به منظور شناسایی مشتریان با ارزش بیمه شخص ثالث – که خسارت کمتری دارند – می‌باشد. با استفاده از دو شاخص دان و دیویس بولدین تعداد بهینه خوشه تعیین می‌شود و پس از آن داده ها با استفاده از الگوریتم K-means خوشه بندی می شوند. سه متغیر سال‌های عدم خسارت، تعداد ادعای خسارت و سابقه رانندگی به عنوان متغیرهای تعیین ارزش مشتری در نظر گرفته شده و وزن آنها با استفاده از روش تحلیل سلسله مراتبی تعیین می‌گردد. پس از آن ارزش مشتریان هر خوشه محاسبه شده و خوشه‌ها به‌ترتیب ارزش مشتریان رتبه‌بندی می‌شوند. در

پایان نیز با استفاده از قوانین انجمنی روابط و الگوهای رفتار مشتریان کشف و شناسایی می‌شوند. با استفاده از نتایج این پژوهش شرکت‌های بیمه‌ای می‌توانند ضمن بخش‌بندی و شناسایی مشتریان باارزش که خسارت کمتری دارند، میزان خسارت مشتریان جدید را نیز پیش‌بینی کنند. در تحقیقی توسط [۱۹] هدف، دستیابی به قوانین تصمیم‌گیری و مدل پیش‌بینی مشتریان آتی در یک شرکت بیمه‌ای می‌باشد. در تحقیق فوق از دوتکنیک درخت تصمیم و شبکه‌های عصبی در دسته‌بندی داده‌ها استفاده شده است که بررسی‌ها نشان داده است که درخت تصمیم نتایج بهتری را حاصل کرده‌اند. در تحقیقی توسط [۲۰]، هدف شناسایی و دسته‌بندی مشتریان بیمه با استفاده از داده‌کاوی می‌باشد. از آن جایی که اغلب بیمه‌نامه‌های آتش‌سوزی اماکن مسکونی، صنعتی و غیر صنعتی در بازه زمانی یکساله صادر و پس از طی دوره اعتبار منقضی می‌گردد، پایان مدت بیمه‌نامه می‌تواند پایان عمر مشتری نیز باشد، از این رو شناخت و شناسایی مشتریانی که احتمال خرید مجدد یا تمدید بیمه‌نامه آنها بیشتر است، برای شرکت‌های بیمه حایز اهمیت و آنها را در بازاریابی هدفمند، تبلیغات و حفظ مشتریان فعلی و شناخت مشتریان آتی یاری می‌کند. این پژوهش براساس اطلاعات موجود یک شرکت بیمه خصوصی، مشتریان بیمه آتش‌سوزی را در دسته‌های مجزایی تفکیک کرده و مشتریان مستعد خرید را شناسایی کرده و براساس درخت تصمیم و با استفاده از نرم‌افزار Rapidminer مدل کرده است. در تحقیقی توسط [۱۵] هدف، به‌کارگیری فرایند داده‌کاوی برای پیش‌بینی رویگردانی مشتری در بیمه می‌باشد. پژوهش فوق با هدف تبیین قابلیت‌های داده‌کاوی در مدیریت رویگردانی مشتری و با بهره‌گیری از متدولوژی استاندارد داده‌کاوی CRISP-DM به کاوش در پایگاه داده‌های یکی از شرکت‌های بیمه‌ای در رشته بیمه آتش‌سوزی پرداخته است. نتایج نشان می‌دهد کانال جذب مشتری عامل اصلی پیش‌بینی کننده رویگردانی مشتری در شرکت بوده و در مراتب بعد سابقه خرید و کاربری مکان بیمه شده به‌عنوان عوامل پیش‌بینی کننده رویگردانی قرار می‌گیرند. در این تحقیق از ابزار درخت تصمیم الگوریتم CHAID در جهت دسته‌بندی مشتریان استفاده شده است. در تحقیقی توسط [۲۱] هدف، مطالعه یک شرکت بیمه‌ای در ارتباط با تجزیه و تحلیل دو مشکل حفظ مشتری و ادعای حق بیمه و حل آنها با استفاده از ابزارهای داده‌کاوی در جهت رسیدن به یک قیمت‌گذاری بهینه برای رسیدن به یک تعادل میان افزایش سودآوری و حفظ مشتری می‌باشد. در پژوهش فوق در ارتباط با مشکل اول از تکنیک های طبقه‌بندی در جهت

شناسایی مشتریانی که قصد تمدید و یا فسخ قرارداد خود را دارند، استفاده شده است. در ارتباط با ادعای حق بیمه نیز به شناسایی دو عامل مهم و تأثیرگذار مشتریان در معرض خطر پذیری بیشتر و الگوهای ادعای حق بیمه پرداخته شده است. در تحقیق فوق از ابزار شبکه عصبی در جهت پیش‌بینی سیاست‌هایی که منجر به فسخ یا تمدید قرارداد می‌شود و از ابزار خوشه‌بندی در جهت شناسایی گروه‌های مختلف هزینه‌ای و در نهایت از درخت تصمیم‌گیری برای انتخاب تصمیم‌های متعادل استفاده شده است. در تحقیقی توسط [۲۲] هدف، یافتن رابطه بین تعداد خرید و درآمد حاصل از مشتری می‌باشد. در تحقیق فوق از ابزار داده‌کاوی در مطالعه برای پیدا کردن آن دسته از مشتریان که بیشتر از یک بیمه عمر خریداری کرده‌اند و رضایت خاطر از پرداخت خود داشته‌اند، استفاده شده است. از الگوریتم K-means برای طبقه‌بندی و در ارتباط با پیش‌بینی از نرم‌افزار اکسل استفاده شده است. متدولوژی مورد استفاده CRISP-DM و از نرم‌افزار SPSS استفاده شده است. خروجی تحقیق فوق یک‌سری قوانین ساده در جهت پیش‌بینی خوشه‌هایی از مشتریان می‌باشد که بیش از یک بیمه‌نامه خرید خواهند کرد و سودآوری بیشتری خواهند داشت. در تحقیقی توسط [۲۳]، پژوهش فوق به تحلیل نرخ بازگشت مشتریان بیمه عمر به‌عنوان یکی از مهم‌ترین فاکتورهای فرسایش مشتریان بیمه عمر پرداخته شده است. هدف از این تحقیق طبقه‌بندی بیمه‌شدگانی می‌باشد که تمایل به فسخ قرارداد دارند و از میان آنها مشتریانی که خطر پذیری بالاتری را شامل می‌شوند، گروه هدف را تشکیل می‌دهند. در این پژوهش از درخت‌های تصمیم و شبکه‌های عصبی بهره گرفته شده است. در انتها نیز مدل ایجاد شده با استفاده از نمودار ROC و مقدار AUC زیر نمودار ارزیابی شده است. در تحقیقی توسط [۲٤] به بررسی کاربردها و وظایف داده‌کاوی، انواع سیستم‌های داده کاوی، چرخه حیات، روش‌ها و گستره داده‌کاوی پرداخته شده است. در این تحقیق یکی از وظایف داده‌کاوی کشف الگوها و قوانین، ذکر شده است که از این وظیفه در یافتن الگوی پنهان در یک خوشه استفاده می‌شود. برای این کار از الگوریتم‌های k-mean و k-medoids استفاده شده است.

۷- روش تحقیق

روش مورد استفاده در این تحقیق داده‌کاوی می‌باشد. این تکنیک الگوهای حاوی اطلاعات را در داده‌های موجود جستجو می‌کند. تکنیک‌های داده‌کاوی از جمله تکنیک های نوین علمی هستند

که در توصیف، تشریح، پیش‌بینی وکنترل پدیده‌ها به کار می‌روند. این تکنیک‌ها به اندازه‌گیری، تشریح و پیش‌بینی درجه وابستگی میان متغیرها می‌پردازند. روش‌های داده‌کاوی نه تنها بر جنبه‌های تحلیلی مطالعات، بلکه در طراحی و ابزارهای جمع‌آوری داده برای تصمیم‌گیری و حل مسائل نیز تأثیر می‌گذارند. انواع متدولوژی های داده‌کاوی شامل ابزارهای پرس و جو، فنون آماری، مصورسازی، پردازش تحلیلی پیوسته، یادگیری مبتنی بر مورد، درختان تصمیم‌گیری، قوانین وابستگی، شبکه‌های عصبی و الگوریتم ژنتیک می‌باشند.

جامعه آماری مورد مطالعه شامل اطلاعات خرید بیمه‌نامه عمر و تشکیل سرمایه مشتریان شرکت بیمه سامان در سال‌های ۱۳۹۰ تا ۱۳۹۳ می باشدکه فیلدهای مختلفی دارند، از جمله :

مشخصات بیمه شده: تاریخ تولد، جنسیت، وضعیت تأهل، شغل، درآمد، میزان تحصیلات

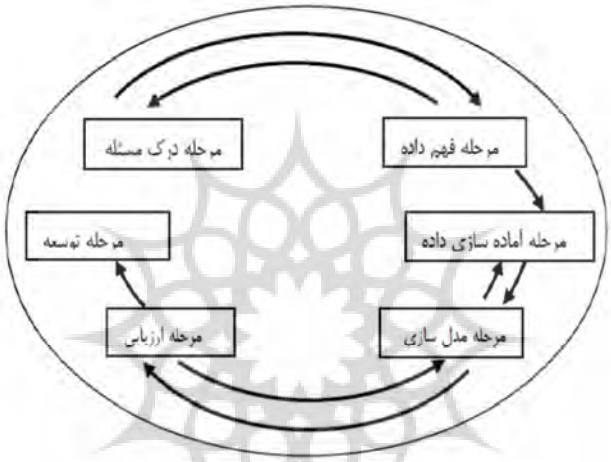
مشخصات بیمه‌نامه : حق بیمه سالیانه، سرمایه فوت، پوشش‌های تکمیلی، میزان خطر پذیری، روش پرداخت

پوشش‌های تکمیلی: فوت بر اثر حادثه، از کار افتادگی، امراض صعب العلاج.

۸- ابزار مورد استفاده

در این تحقیق برای تجزیه و تحلیل‌ها داده از نرم‌افزار داده‌کاوی SPSS MODELER ۱٤.۲ استفاده شده است که در حال حاضر یکی از نرم‌افزارهای قدرتمند در زمینه داده‌کاوی به شمار می‌آید، این محصول یک برنامه نرم‌افزاری تحلیل متن و داده‌کاوی است که به‌وسیله شرکت IBM ساخته شده است و از آن برای ساخت مدل‌های پیش‌بینی و انجام سایر کارهای تحلیلی استفاده می‌شود و به کاربران این امکان را می‌دهد که تصمیم‌های درست تری بگیرند. از قابلیت‌های بارز این نرم‌افزار، می‌توان به رابط کاربری منعطف، پشتیبانی از استاندارد CRISP-DM و وجود راهنمای قوی در محیط آن اشاره کرد. این عوامل در کنار پشتیبانی از طیف وسیعی از الگوریتم‌های شناخته شده و معتبر باعث شده است تا این نرم‌افزار به یکی از محبوب‌ترین و پر استفاده‌ترین نرم‌افزارهای تجاری داده‌کاوی در سطح جهانی تبدیل شود. این نرم‌افزار به صورت گسترده‌ای در طرح های داده‌کاوی کاربرد دارد و به‌وسیله امکانات و الگوریتم‌هایی که در اختیار کاربران خود قرار می‌دهد، مدلسازی و انجام امور داده‌کاوی را به بهترین شکل انجام می‌دهد.

در این تحقیق از روش شناسی CRISP-DM استفاده شده است. این روش اولین بار در سال ۱۹۹٦ ارائه شد. در تحقیقی که در سال ۲۰۱۰ انجام شد، نشان می‌دهد که این متدولوژی بیش از ۵۰ درصد نسبت به سایر متدولوژی‌ها در سال ۲۰۰۷ به کار گرفته شده است [۲۵]. این متدولوژی از گام‌های شناخت سیستم، درک داده، آماده‌سازی داده، مدلسازی، ارزیابی و توسعه سیستم براساس شکل ۱ تشکیل شده است.



شکل ۱ مراحل متدولوژی [۲٦]

مرحله شناخت و درک کسب‌وکار: به منظور داشتن درکی موفق از مسئله، معمولاً تجربه و توانمندی در حیطه یک دانش خاص لازم است. [۲۷] در این مرحله باید متخصص داده‌کاوی از توان و تجربه متخصص کسب‌وکار بهره مند شود. البته در یک طرح موفق داده‌کاوی این مشارکت در مرحله اولیه متوقف نخواهد شد. با توجه به بررسی‌های انجام شده، یکی از اهداف اصلی که از اولویت بالای سازمانی برخوردار است، شناسایی الگوی رفتاری مشتریان در مدیریت بیمه‌های عمر و تشکیل سرمایه در جهت شناخت و حفظ مشتریان وفادار در سازمان و ارائه خدمات بهتر به مشتریان در معرض خطر می‌باشد. پس از تدوین اهداف مورد نظر کسب‌وکار، در جلسه‌های متعددی با خبرگان و کارشناسان مدیریت بیمه‌های عمر و تشکیل سرمایه، نسبت به شناسایی روال‌ها و سیاست‌های کاری اقدام شد. به لحاظ فنی برای شناخت

سامانه ثبت و صدور بیمه‌نامه جلساتی با خبرگان سامانه مذکور در مدیریت فناوری ارتباطات و اطلاعات برگزار شد. در جلسه‌های برگزار شده مواردی از قبیل شناسایی جداول مرتبط با بیمه‌نامه‌های صادر شده و ارتباط بین جداول در پایگاه داده، مشخصه‌های اطلاعاتی و نحوه دسترسی به داده‌ها مورد بررسی قرار گرفت. از دیگر بخش‌های این مرحله ترجمه اهداف و محدودیت آن در تعریف مسئله داده‌کاوی و مهیا کردن استراتژی اولیه برای نائل شدن به اهداف تعریف می‌شود. در این مرحله اهداف تحقیق از منظر کسب‌وکار که شناسایی الگوی رفتاری مشتریان بوده است، به همراه جزئیات مشخص شدند. از آن جایی که تنها مدیریت عمر و تشکیل سرمایه مورد هدف می‌باشد، اطلاعات سامانه ثبت و صدور بیمه‌نامه‌های مدیریت عمر و تشکیل سرمایه به‌عنوان بخش کاربردی تحقیق مورد بررسی قرار گرفت. به منظور انجام تحقیق از منظر فنی و به‌کارگیری تکنیک داده‌کاوی اهمیت فراهم بودن شرایط مناسب مورد نظر قرار گرفته شد. برای این منظور تمام داده‌های بیمه‌ناهه‌های مربوط به سال‌های ۱۳۹۰ تا ۱۳۹۳ انتخاب شدند که به لحاظ حجم داده، فراوانی بالای (۱۸۰۰۰۰ رکورد) را شامل می‌شود.

مراحل CRISP-DM به شرح زیر می‌باشد:

- مرحله شناخت و درک داده‌ها : شناخت داده‌ها عبارت است از جمع‌آوری داده‌های اولیه، توصیف‌ها داده‌ها ، بازرسی و بررسی داده‌ها و اعتبارسنجی کیفیت داده‌ها. کارایی داده‌کاوی به طور مستقیم مرتبط با داده‌های مورد استفاده است. هر اندازه داده‌ها دقیق‌تر، جامع‌تر و با کیفیت‌تر باشند، خروجی داده‌کاوی کاراتر خواهد بود. بنابراین انتخاب و جمع‌آوری داده‌های درست، توصیف آنها، یکپارچه‌سازی قالب آنها به منظور استفاده در داده‌کاوی، از اهمیت بسیار بالایی برخوردار می‌باشد. علاوه بر این بازرسی و بررسی داده‌ها به منظور تعیین میزان کیفیت آنها بسیار مهم می‌باشد. در این تحقیق داده‌های مربوط به سال‌های ۱۳۹۰ تا ۱۳۹۳ جمع‌آوری شده است. داده‌های جمع‌آوری شده مربوط به بیمه‌نامه‌های عمر و تشکیل سرمایه است. پس از تحلیل کسب‌وکار از پایگاه داده مشتریان بیمه‌های عمر و تشکیل سرمایه، ۳٦ فیلد اطلاعاتی شامل متغیرهای مالی و جمعیت‌شناختی برای نیل به اهداف تحقیق جمع‌آوری شد. به منظور بررسی کیفی داده‌های موجود، انتقال داده‌ها به نرم‌افزار SPSS MODELER ۱٤.۲ انجام شد. برای آنکه بتوان کیفیت داده‌های موجود را بررسی کرد، در این نرم‌افزار با استفاده از ابزار ممیزی داده (Data Audit) این کار انجام می‌شد. بر این اساس، در این مرحله مواردی همچون

مقادیر پرت و گم شده شناسایی شد که به عنوان ورودی مرحله بعدی ارائه می‌شود تا مشکلات موجود رفع گردد. در این مرحله مشخص شد که اطلاعات اولیه موجود به‌تنهایی تمامی بخش‌های تحقیق را تحت پوشش قرار نمی‌دهند اما با استفاده از اطلاعات پایه موجود می‌توان مشخصه‌های جدید اطلاعاتی مورد نیاز را ایجاد کرد.

- مرحله آماده‌سازی‌ها و پیش‌پردازش داده: پس از گردآوری داده‌ها باید خطاهای احتمالی موجود در آنها را از بین برده و تمیز کرد. [۲۸] این خطاهای احتمالی عبارتند از مقادیر خارج از رفتار یا حدی، ارزش‌های گم شده، صفات تکراری، داده‌هایی که در قالب مناسب برای مدلسازی نیستند. برای رفع خطاهای ذکر شده در این پژوهش مراحل زیر انجام شده است:

۱-حذف مقادیر گمشده در ارتباط با هر یک از متغیرها، درصورتی که کمتر از ۱ درصد کل رکوردهای موجود را شامل شوند.

۲- برگذاری جلسه با خبرگان در جهت شناسایی رنج مقادیر داده‌ای قابل قبول برای هر یک از متغیرهای استخراج شده و به‌کارگیری تکنیک حذف جهت برخورد با خطای داده‌های پرت.

۳- جایگزین کردن مقادیر داده‌ای گمشده با استفاده از الگوریتم های C&R Tree و تصادفی به‌ترتیب برای متغیرهای عددی و اسمی.

۴- استفاده از نود transformer برای نرمال‌سازی برخی از متغیرها جهت برخورد با خطای داده‌های پرت.

۵- به‌کارگیری نود Anomaly برای شناسایی رکوردهای پرت و حذف آنها .

۶- تقسیم مجموعه داده‌ای به دو دسته داده‌های آموزشی و آزمایشی به نسبت ۷۰ به ۳۰ در مرحله مدلسازی.

پس از به‌کارگیری تکنیک‌های ذکر شده در بالا در راستای حذف مقادیر پرت و گمشده در نهایت ۱۷۱۱۹۲ رکورد جهت خوشه‌بندی مورد استفاده قرار گرفت. همچنین برای داده‌های پرتی که پس از طی مرحله قبل باقی مانده بود، به دلیل نوع ماهیت داده و ماهیت مسئله مقادیر خارج از رده در نوع حل مسئله تأثیرگذار می‌بود و با توجه به قلمرو مسئله نمی‌توان مواردی را حذف کرد. یکی دیگر از اقدام‌های انجام شده در این مرحله، انتخاب ویژگی‌های مؤثر و حذف متغیرهایی که اهمیت زیادی در انجام تحقیق ندارند، می‌باشد. پس از جلسه‌های متعدد با خبرگان، ۵ ویژگی جدید (نسبت سرمایه اولیه به حق بیمه اولیه، نسبت سرمایه فوت حادثه به

حق بیمه اولیه، نسبت سرمایه از کار افتادگی حادثه به حق بیمه اولیه، نسبت سرمایه امراض به حق بیمه اولیه، ماه شروع بیمه نامه) از ترکیب مشخصه‌های اطلاعاتی موجود استخراج شد. همچنین با الهام گرفتن از شاخص های مدل آر.اف.ام (تازدگی مبادله، تعداد دفعات مبادله و ارزش پولی مبادله) سه متغیر زیر با تعاریف تطبیق داده شده با بستر مدیریت بیمه‌های عمر و تشکیل سرمایه به مجموعه متغیرهای موجود اضافه گردید.

-تازگی عمل مبادله (R) : بازه زمانی بین آخرین پرداخت و موعد آخرین پرداخت (با درنظر گرفتن تأخیر و تقدم بر مبنای روز)

-تعداد دفعات مبادله (F) : تعداد اقساط پرداختی بر تعداد کل اقساط

-ارزش مالی مبادله (M) : نسبت مبلغ پرداختی بر کل ارزش پولی که باید پرداخت می‌شد.

لازم به توضیح می‌باشد که به دلیل وجود روش‌های پرداختی متفاوت (ماهانه – سالیانه) در مدیریت بیمه‌های عمر و تشکیل سرمایه، جهت ایجاد یک ارزش وزنی متناسب با روش پرداخت، رابطه (۲) جایگزین متغیر ارزش مالی مبادله (M) شد که $\bar{m}$ متوسط حق بیمه پرداختی در کل مجموعه داده‌ای می‌باشد..

$$M \times (\frac{M}{\bar{m}})$$ رابطه(۲)

در مرحله بعد برای ارزیابی میزان اهمیت متغیرهای ورودی و انتخاب متغیرهای پر اهمیت‌تر از الگوریتم‌های شبکه‌های عصبی و درخت تصمیم (CHAID ،CRT ،C5.0) با پارامترهای مختلف در ورودی بهره گرفته شد که در مجموع از میان ١٤٤ مدل ایجاد شده، ۲۱ ویژگی برای ورودی مرحله بعد انتخاب شدند .

• مرحله مدلسازی: پس از شناخت‌ها داده‌ها و آماده‌سازی آنها می‌توان به مدلسازی پرداخت. در اولین قدم از مدل باید روش مناسب را انتخاب کرد، انتخاب روش مناسب بسیار تعیین کننده است. پارامترهای مورد نیاز مدل نیز پس از تعیین روش مورد استفاده مشخص می‌شوند. پس از انتخاب مدل و تعیین پارامترها، بخش‌های کوچکی از پروژه تعریف شده و پس از اجرا شدن، در هر مرحله به دقت بررسی می‌شوند تا کیفیت مدل ایجاد شده تضمین شود. در این مرحله با توجه به هدف شناسایی الگوی رفتاری مشتریان از طریق دسته‌بندی مشتریان، روش‌ها و ابزارهای مختلفی به کار گرفته می‌شود و نتایج و دانش‌های گوناگونی نیز استخراج می‌گردد که در ادامه تشریح خواهد شد. در این تحقیق با انتخاب و به‌کارگیری تکنیک‌های

مدلسازی مناسب و روش داده‌کاوی معین نتایج مدلسازی را بهینه کرده و در صورت نیاز می‌توان با برگشت به عقب تحلیل مدلسازی را بهینه‌تر کرد.

۱- خوشه‌بندی و تعیین تعداد بهینه خوشه‌ها:

در این مرحله به منظور مدلسازی از نرم‌افزارIBM SPSS MODELER۱٤.۲ (که یکی از نرم‌افزارهای مشهور در زمینه داده‌کاوی است) استفاده شده است. در این تحقیق خوشه‌بندی مشتریان از طریق سه شاخص آر.اف.ام و با به‌کارگیری دو الگوریتم K-Means و Two-step صورت پذیرفت. هریک از این دو الگوریتم روی داده‌ها اعمال و با توجه به شاخص سیلوئت بهترین الگوریتم انتخاب شد. الگوریتم Two-step از یک روش خوشه‌بندی دو مرحله‌ای استفاده می‌کند. مرحله اول با یک بار گذر از داده‌ها، آنها را در مجموعه قابل قبولی از زیر خوشه‌ها فشرده می‌کند. قدم دوم از یک روش خوشه‌بندی سلسله مراتبی به منظور ادغام تکاملی این زیر خوشه‌ها به خوشه‌های بزرگ‌تر بهره می‌برد [۲٦]. یکی از مزایای این الگوریتم اجرا بر روی مجموعه داده‌های بزرگ و تعیین تعداد بهینه خوشه ها است [۲٦]. ارزیابی کیفیت خوشه‌بندی، میزان برتری یک خوشه نسبت به خوشه‌های دیگر به‌وسیله الگوریتم‌های متفاوت خوشه‌بندی یا الگوریتم‌های مشابه ولی با مقدار پارامترهای متفاوت می‌باشد [۲۹]. برای تعیین تعداد بهینه خوشه‌ها از الگوریتم Two-step و جهت ارزیابی مدل از شاخص سیلوئت و میزان اهمیت متغیرهای ورودی مدل استفاده شد. در مرحله اول پس از ارزیابی ۳۲ مدل ایجاد شده، تعداد خوشه بهینه براساس خروجی نمایش داده شده در جدول ۱ با بالاترین سیلوئت و اهمیت متغیرهای ورودی، دو خوشه تعیین گردید. در ادامه براساس نظر خبرگان، خروجی الگوریتم Two-step با خروجی الگوریتم k-means با تعداد ورودی دو خوشه انطباق داده شد، از آن جایی که اهمیت شاخص R در الگوریتم K-means ، ۰.۰۱ می‌باشد، سه به‌عنوان تعداد خوشه بهینه لحاظ شد.

شاخص سیلوئت: شاخص تراکم و جدایی سیلوئت با مقادیر ضعیف، متوسط و خوب نشان داده می‌شود. میانگین مقدار شاخص سیلوئت برای ارزیابی اعتبار خوشه‌بندی و همچنین برای تصمیم‌گیری در مورد انتخاب تعداد کلاس‌های بهینه مورد استفاده قرار می‌گیرد که این میزان براساس دوری و نزدیکی مشاهدات و خوشه‌ها به یکدیگر محاسبه می‌شود. مقدار S(i) با استفاده از رابطه (۳) قابل محاسبه است:

رابطه(۳)

$$s(i)=\frac{(b(i)-a(i))}{\max(a(i),b(i))}$$

a(i) میانگین فاصله بین مشاهده i با سایر مشاهدات در یک خوشه مشابه و b(i) میانگین فاصله مشاهده i به تمام مشاهدات در خوشه‌های دیگر می‌باشد. براساس فرمول بالا مقدار S(i) بین ،-۱ و +۱ قرار دارد. اگر S(i) به +۱ نزدیک‌تر باشد، به این معنا است که خوشه‌بندی نمونه خوب صورت گرفته است و خوشه پیشنهاد شده برای نمونه مورد نظر مناسب می‌باشد. ولی اگر S(i) به -۱ نزدیک‌تر باشد، به این معناست که خوشه‌بندی نمونه به‌خوبی انجام نشده و خوشه پیشنهاد شده برای داده مورد نظر نامناسب می‌باشد.

جدول ۱ تعیین تعداد بهینه خوشه‌ها با استفاده از شاخص سیلوئت و میزان اهمیت متغیرهای ورودی

الگوریتم	تعداد خوشه	شاخص سیلوئت	میزان اهمیت شاخص‌های آر.اف.ام
Two-step	٢	٠.٧٠	R=١, F=١, M=١
	٣	٠.٦٤	R=١, F=١, M=١
	٤	٠.٦٠	R=١, F=١, M=١
	٥	٠.٥٥	R=١, F=١, M=١
K-Means	٢	٠.٧٠	R=٠.٠١, F=١, M=١
	٣	٠.٧٠	R=١, F=١, M=١

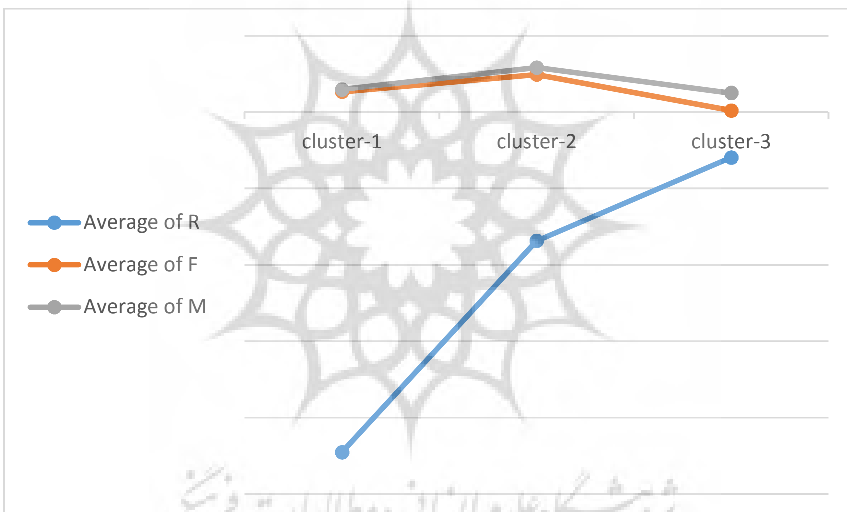
پژوهشگاه علوم انسانی و مطالعات فرهنگی

۹- نتایج خوشه‌بندی

با استفاده از خروجی به دست آمده در مرحله قبل، تعداد سه خوشه به همراه شاخص‌های آر.اف.ام به عنوان ورودی الگوریتم‌های K-Means و Two-step برای خوشه‌بندی مشتریان بیمه‌های عمر و تشکیل سرمایه در نظر گرفته شد. نتایج حاصل شده در جدول ۲ و شکل ۲ ارائه شد.

جدول ۲ نتایج حاصل از خوشه‌بندی براساس دو الگوریتم K-Means و Two-step

الگوریتم	شاخص سیلوئت	شماره خوشه	میانگین R	میانگین F	میانگین M
Two-step	۰.٦٤	۱	-۱۰.٦۹	۰.٥٦	۰.٤۱
		۲	-۲.٤۰	۰.۹۸	۱.٥۹
		۳	-۰.۲۷	۰.۰۳	۰.٥٤
K-Means	۰.۷۰	۱	-۸.۹۱	۰.٥۳	۰.٥۹
		۲	-۳.۳۷	۰.۹۸	۱.۱۷
		۳	-۱.۱۹	۰.۰٤	۰.٥۰



شکل ۲ میانگین شاخص‌های آر.اف.ام برای هر خوشه در الگوریتم K-Means

۲- اولویت‌بندی خوشه‌ها

درنهایت برای ارزیابی بهتر نتایج خوشه‌بندی، خوشه‌ها براساس شاخص‌های آر.اف.ام منطبق بر ساختار تحقیقات [۳۰؛ ۳۱] رتبه‌بندی شد. این فرایند شامل مراحل زیر است:

-نرمال‌سازی شاخص‌های آر.اف.ام :

امیر بروفر و همکاران ـــ

براساس تعریف، پارامترهای آر.اف.ام همگن نیستند. مقدار این پارامترها باید نرمالیزه شود. فرایند نرمال‌سازی به دو طریق برحسب میزان تأثیر هر پارامتر بر ارزش مشتری صورت می‌گیرد. فرم‌های مختلف نرمال‌سازی، حالت سود و هزینه می‌باشند که به‌ترتیب در روابط ٤ و ٥ نشان داده شده‌اند.

رابطه (٤) $$x' = (x - x^S) / (x^L - x^S)$$

رابطه (٥) $$x' = (x^L - x) / (x^L - x^S)$$

در روابط بالا x' ، x^L و x^S به‌ترتیب، نرمالیزه شده، بیشترین و کمترین مقدار x می‌باشند. در این تحقیق برای هر سه شاخص آر.اف.ام با توجه به اینکه تأثیر مثبت بر ارزش مشتری دارند، از فرم سود استفاده شد. درنهایت R' ، F' و M' نیز نشان‌دهنده مقادیر نرمال شده شاخص‌ها می‌باشند.

- وزن‌دهی شاخص‌های آر.اف.ام:

از آنجا که ارزش هر خوشه براساس سه شاخص آر.اف.ام سنجش می‌شود، لذا نخست باید وزن‌های این سه شاخص محاسبه شود. وزن‌های این سه شاخص براساس نظر خبرگان و از طریق پرسشنامه به روش تحلیل سلسله مراتبی در نرم‌افزار Expert Choice تعیین شدند. در این پرسشنامه از طیف لیکرت ۹ گزینه‌ای استفاده شده است و ۱۹ خبره صنعت بیمه به آن پاسخ داده‌اند. درنهایت وزن هر یک از متغیرهای آر.اف.ام به‌ترتیب با WF=۰.۳، WR=۰.۲، WM=۰.۵ و تعیین شد که مجموع این وزن‌های نسبی برابر با یک است.

- تعیین متوسط ارزش شاخص‌های آر.اف.ام در هر خوشه:

متوسط ارزش هر یک از شاخص‌ها در هر خوشه با تقسیم مجموع ارزش شاخص در آن خوشه به تعداد مشتریان آن خوشه (n) تعیین شده است. متوسط ارزش شاخص‌ها در هر خوشه را با علایم R''، F''، M'' نمایش می‌دهیم.

رابطه(٦) $$R'' = \frac{\sum R'}{n}$$

رابطه(٧) $$F'' = \frac{\sum F'}{n}$$

رابطه(۸)

$$M'' = \frac{\sum M'}{n}$$

-اولویت‌بندی هر خوشه :

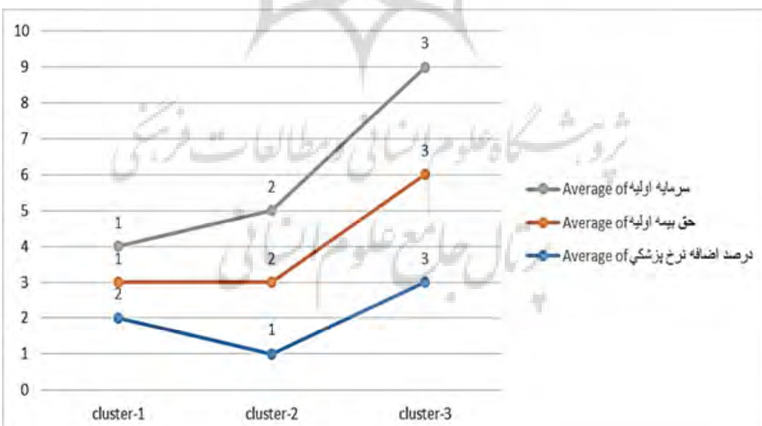
درنهایت ارزش هر خوشه از جمع مقادیر به‌دست آمده از ضرب ارزش متوسط هر شاخص در وزن آن از طریق رابطه ۹ تعیین شده است.

رابطه(۹)

$$Q = WR.\ R'' + WF.\ F'' + WM.\ M''$$

ارزیابی : پس از مدلسازی باید به ارزیابی نتایج حاصل از مدل پرداخت. نتایج مرحله ارزیابی باعث بهبود مدل شده و مدل را قابل استفاده می‌کند. در این گام اعتبار مدل بررسی شده و گزارشی از کل فرایند تهیه می‌شود. در انتها نیز فهرستی از اقدام‌های اصلاحی قابل انجام، تهیه و به عنوان راهکار ارائه شده است و تصمیم‌گیری ها بر این اساس انجام می‌شود.

براساس جدول ۲ بین دو الگوریتم K-means و Two-step الگوریتم K-means شاخص سیلوئت بیشتری دارد، بنابراین در این تحقیق الگوریتم K-means ، الگوریتم انتخابی برای خوشه‌بندی مشتریان بیه عمر و تشکیل سرمایه می‌باشد. در ادامه برخی از اطلاعات مهم مالی و جمعیت‌شناختی در هر خوشه و در جدول ۳ و شکل ۳ نمایش داده شده است.



شکل ۳ اطلاعات مالی خوشه‌های مشتریان

جدول ۳ اطلاعات جمعیت‌شناختی خوشه‌های مشتریان

شماره خوشه		خوشه ۱	خوشه ۲	خوشه ۳
درصد تعداد اعضای خوشه		۲۰	٦۰	۲۰
میانگین سن (سال)		۲٥.٤	۲٤.۹	۲٥.۹
در صد جنسیت بیمه شده	مرد	٦۳	٦۰	٦۸
	زن	۳۷	٤۰	۳۲
روش پرداخت		شش‌ماهه	سالیانه	شش‌ماهه

با توجه به خروجی‌های مراحل قبل، نتایج رتبه‌بندی خوشه‌ها با توجه به مقدار متوسط ارزش شاخص‌های آر.اف.ام و وزن های استخراج شده در قالب جدول ٤ انجام شد.

جدول ۴ مشخصات مراکز خوشه‌ها و رتبه آنها براساس معادلات ذکر شده

شماره خوشه	WR. R″	WF. F″	WM. M″	Q	رتبه خوشه
۱	۰.٥۳	۰.٥۳	۰.۱٦	۰.۳٤٥	۲
۲	۰.٥۹	۰.۹۸	۰.۳۲	۰.٥۷۲	۱
۳	۰.٦۱	۰.۰٤	۰.۱٤	۰.۲۰٤	۳

- مرحله تحکیم و گسترش: استفاده کردن از مدل ایجاد شده، برای مثال می‌توان تولید یک گزارش ساده از خروجی‌ها را نام برد و برای یک مثال پیچیده تکمیل کردن پردازش داده‌کاوی موازی در سایر حوزه‌ها می‌باشد که این الگوها به یک دانش مفید و قابل استفاده تبدیل می‌شوند و پس از بهبود آنها، الگوهایی که کارا محسوب می‌شوند، در یک سیستم اجرایی به کارگرفته خواهند شد.

۱۰- نتیجه‌گیری

با استفاده از الگوهای خوشه‌بندی مشتریان بیمه عمر و تشکیل سرمایه خوشه‌بندی شدند و با مقایسه صورت گرفته میان دو الگوریتم خوشه‌بندی K-Means و Two-step مشخص گردید

که الگوریتم K-Means از نظر معیار شاخص سیلوئت با مقدار ۰.۷۰ برتری دارد که درنهایت این الگوریتم به عنوان بهترین الگوریتم برای خوشه بندی مشتریان انتخاب شد. پس از خوشه‌بندی مشتریان با استفاده از الگوریتم K-Means، با در نظر گرفتن متغیرهای خوشه‌بندی که شامل شاخص‌های آر.اف.ام است، سه خوشه از مشتریان مشخص شد. براساس نتایج و تجربیاتی که در طول این تحقیق حاصل شده است، خوشه ۱ با تخصیص دادن ۲۰ درصد از جمعیت مشتریان بیمه عمر و تشکیل سرمایه به خود رتبه دوم را کسب کرد. بیشتر افراد این خوشه را آقایان با میانگین سنی ۲۵.٤ تشکیل می‌دهند. از خصیصه‌های بارز این گروه در ارتباط با شاخص‌های آر.اف.ام ، نرخ پایین تازگی مبادله می‌باشد. همچنین در ارتباط با شاخص‌های مالی، در میان خوشه‌های دیگر کمترین مقدار متوسط حق بیمه اولیه و سرمایه اولیه را دارا می‌باشد. خوشه ۲ با تخصیص دادن ٦۰ درصد از جمعیت مشتریان بیمه عمر و تشکیل سرمایه به خود رتبه اول را کسب کرد. بیشتر افراد این خوشه را آقایان با میانگین سنی ۲٤.۹ تشکیل می‌دهند. از خصیصه‌های بارز این گروه در ارتباط با شاخص‌های آر.اف.ام ، کسب بالاترین مقدار متوسط در ارزش پولی مبادله شده و تعداد دفعات مبادله می‌باشد. همچنین در ارتباط با شاخص‌های مالی در میان خوشه‌های دیگر کمترین مقدار متوسط اضافه نرخ پزشکی را دارا می‌باشد. خوشه ۳ با تخصیص دادن ۲۰ درصد از جمعیت مشتریان بیمه عمر و تشکیل سرمایه به خود رتبه سوم را کسب کرد. بیشتر افراد این خوشه را آقایان با میانگین سنی ۲۵.۹ تشکیل می‌دهند. از خصیصه‌های بارز این گروه در ارتباط با شاخص‌های آر.اف.ام ، نرخ بالای تازگی مبادله در میان خوشه‌های دیگر می‌باشد. همچنین در ارتباط با شاخص‌های مالی در میان ای دیگر بالاترین مقدار متوسط اضافه نرخ پزشکی، سرمایه اولیه و حق بیمه اولیه را دارا می‌باشد.

در این قسمت با توجه به تجزیه و تحلیل داده‌ها و نتایج حاصل از طرح به سؤال‌های تحقیق پاسخ داده می‌شود.

سؤال اصلی : نتایج به دست آمده از داده‌کاوی بر داده‌های مشتریان بیمه عمر و تشکیل سرمایه چه تأثیری بر بهبود سیاست‌های مدیریت و ارتباط با مشتری دارد ؟

پاسخ سؤال اصلی : براساس تعریف، یک مشتری با داشتن مقدار آر.اف.ام بالاتر، مشتری ارزشمندتری برای شرکت بیمه می‌باشد. با توجه به اینکه مشتریان خوشه ۲، از نظر

شـاخص‌هـای آر.اف.ام. در وضعیت مطلوبی قرار دارند، لذا پیشنهاد مـی‌شود کـه بـه منظـور حفـظ ایـن مشتریان، شرکت با برقراری ارتباطات و تعاملات بیشتر بـا آنهـا، سـعی در تبـدیل وفـاداری رفتاری این مشتریان به وفاداری نگرشی نماید. اگرچه مشتریان خوشـه ۳ رتبه پایینی از نظـر شـاخص‌هـای آر.اف.ام دارند، اما تازگی مبادله آنها در مقایسه با خوشه های دیگر، بسیار بـالا می‌باشد. شرکت می‌تواند با در نظر گرفتن تخفیف‌های خـاص بـرای مـشتریان ایـن خوشه‌ها، ارزش مالی مبادله مشتریان را افزایش دهد. با توجه به وضـعیت تقریباً مطلـوب مشتریان خوشه ۲ از نظـر شـاخص‌هـای آر.اف.ام. در مقایسه با خوشه ۳، احتمال ارتقای این مشتریان به سطوح بـالای ارزش مشتری، بسیار بالا می‌باشد و شرکت می‌تواند با در نظر گرفتن مزایایی خـاص بـرای خریـد آنها، وفاداری رفتاری را در بین آنها ترویج دهد. در اینجا پیشنهاد می‌شود کـه در زمینـه مـشتریان خوشه ۳ شرکت با انجام مطالعات دقیق‌تر به علـت‌یـابی در زمینـه تعداد دفعات کم مبادله و ارزش مالی پایین مبادلات مـشتریان ایـن خوشـه بپردازد تا با کشف علل به وجود آمدن این وضـعیت، در صـورت امکـان از رویگردانـی‌هـای بعدی مشتریان جدید خود بکاهد.

سؤال فرعی : آیا استفاده از داده‌کاوی به تصمیم‌سازی مدیران شرکت بیمه سامان کمک می‌کند ؟

پاسخ سؤال فرعی: باتوجه به ضرایب اطمینان و مقادیر پشتیبان مدل و نیز تأیید خروجی‌های آن در تکرارهای متوالی اجرای الگوریتم‌ها، استفاده از داده‌کاوی برای یافتن دانش نهفته در حجم داده زیاد باعث رسیدن به راهکارهایی شده است که درصورت به‌کارگیری در شناسایی الگوی رفتاری مشتریان می‌تواند تأثیر زیادی در تخصیص بهینه منابع داشته باشد. در این تحقیق با توجه به شناسایی مشکلات و کسب دانش به‌کارگیری داده‌کاوی به‌عنوان یکی از ابزارهای مطرح مدیریت‌ها داده، بستر اولیه شناسایی الگوی رفتاری مشتریان مهیا شده است. همان طور که پیش از این ذکر شده بود، براساس راهبردهای شرکت و با توجه به اینکه مدیران ارشد شرکت توجه ویژه‌ای به استفاده از ابزارهای فناوری اطلاعات و ارتباطات در راستای تحقق بخشیدن به راهبردهای شرکت دارند، استفاده از دانش کسب شده در این تحقیق، تخصیص بهینه منابع را محقق خواهد کرد.

۱۱- منابع

[1] Mahmoudi K., Rostami H., Sayebani M., Moradi A. (2013) "An overview of data mining and its applications in the offshore industry", The 5th National Conference on Offshore Industries, Tehran, Sharif University of Technology, May and June 23.

[2] Vafaei B., Chegini V., Saqaei A., Ezam M. (2013) "Categorizing water masses in Chabahar Bay using clustering method", *Oceanography*, 4(13):1-19.

[3] Jahanian S., Ebrahimi B. (2010) Customer relationship management a competitive advantage in the chain of today's business value, Industry Engineering Information Center, < http://www.iie.ir/index.php? option= com_content&tas.

[4] Gurau C., Ranchhod A., Hackney R. (2003) "Customer-centric strategic planning: Integrating CRM in online business systems", *Information Technology and Management*, The Netherlands: Springer Netherlands, (4):199-214.

[5] Ryals L., Knox S. (2001) "Cross-FunctionalIssues in the Implementation of Relationship Marketing Trogh Customer Relationship Management", *European Management Journal*, 19(4): 534-542.

[6] Kotler Philip (1994) "Marketing management, analysis, planning, implementation, and control", Prentice Hall. New Jersey, US.

[7] Blattberg R.C, Gary G., Jacquelyn S.T. (2001) *Customer equity: Building and managing relationships as valued assets*, Boston, Massachusetts: Harvard Business School Press.

[8] Lai X. (2009) "Segmentation study on enterprise customers based on data mining technology", *Paper presented at the Database Technology and Applications*, pp. 247-250.

[9] Kumar L., Reinartz W.J. (2006) *Customer relationship management: A data based approach*, New York: John Wiley.

[10] Mishar Alok, Mishar Deepti (2009) "Customer relationship management: Implementation process perspective", Acta Polytechnic Hungarica, 6(4): 83-99.

[11] Ngai E.W.T, Xiu Li, Chau D.C.k. (2009) "Application of data mining techniques in customer relationship management: A literature review and classification", *Expert Systems with Applications*, 36(1):2592-2602.

[12] Alpaydin E. (2004) "Introduction to machine learning", The MIT Press.

[13] T.L Oshini Goonetilleke and H.A Caldera. (2013)" Mining Life Insurance Data for Customer Attrition Analysis ", *Journal of Industrial and Intelligent Information*, 1(1).

[14] Hughes Arthur M. (1994) *Strategic database marketing*, Chicago: Probus Publishing Company.

[15] Akhundzade Noghabi E., Al-Badawi A., Aghdasi M. (2014) "Exploring the dynamics of the customer in the design of segmentation using data mining techniques", *Journal of Information Technology Management*, 6(1): 1-30.

[16] Ghazanfari M., Alizadeh S., Teimourpour B. (2014) *Data mining and knowledge discovery*, Publications of the University of Science and Technology of Iran.

[17] Bashiri mousavi S.A, Afsar A., Mahjubifard A. (2015) "Customer value analysis in bank with data mining technique and fuzzy analytic hierarchy process", *Management Researches in Iran*, 19(1):23-43, (in Persian).

[18] Mozaffari A., Bonyadi A., Alizadeh F. (2013) "Clustering and discovering the damage patterns of the third party insurance customers by using data mining techniques", *The 10th International Conference on Industrial Engineering*.

[19] Izadparast S.M. (2011) "Classifying insurance customers by data mining", The 1st Conference on Intelligent Computer Systems and Their Application.

[20] Mousavi F., Khademi M., Ahmadifar H. (2013) "Identification and classification of insurance customers using data mining", The 1st Regional Conference on Information Technology.

[21] K. A., Smith R.J Willis , Brooks M. (2000) "An analysis of customer retention

and insurance claim patterns using data mining: A case study", *The Journal of the Operational Research Society*, 51(5):532-541.

[22] Mehregan S., Samizadeh R. (2012) "Customer Retention Based on the Number of Purchase: A Data Mining Approach", Inte. J. Manag. Bus. Res, 2(1):41- 50.

[23] Zaïane O.R. (1999) "Data Clustering"*, In Principles of Knowledge Discovery in Databases (Internet)*.. <http://webdocs.cs.ualberta.ca/~zaiane/courses/cmput690/slides/Chapter8/index.html>.

[24] Neelamadhab Padhy, Rasmita Panigrahi. (2012) "Survey of data mining application and Feature scope", *International Journal of Computer Science*, Engineering and Information Technology (IJCSEIT), 2(3).

[25] Mariscal G., Marbán Ó., Fernández C. (2010) "A survey of data mining and knowledge discovery process models and methodologies", *The Knowledge Engineering Review*, 25(02):137-166.

[26] Hasanzadeh A., Hesam M., Ghanbari (2012) "Classification of mobile banking users by data mining approach: Comparison between artificial neural networks and naïve bayes techniques", Management Researches in Iran, 16(2):57-71, (in Persian).

[27] Hand D.J., Mannila H., Smyth P. (2001) "Principles of data mining", and MIT Press.

[28] Forcht K. A., Cochran K. (1999) "Using data mining and data warehousing techniques", *Industrial Management & Data Systems*, pp. 189-196.

[29] Sefidari E., Kadkhodaei A., Sharifi M. (2013) "Comparison of self-building neural network and cluster analysis for the assessment of organic carbon in formations containing hydrocarbons using intelligent systems", *Journal of Oil Research*, 23(75): 117-130.

[30] D.C Li, Wen-Li Dai, Wan-Ting Tseng (2011) "A two-stage clustering method to analyze customer characteristics to build discriminative customer management":

A case of textile manufacturing business", *expert systems with applications*, 14(1):1-6.

[31] Hu Wang and Jing Zhang (2008) "Study of segmentation for auto services companies based on RFM model [online]".

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی