

پیش‌بینی رفتار مشتریان بیمه از طریق ترکیب تکنیک‌های داده کاوی

احسان مختاری سید ابوالقاسم میرروشندل

دانشکده فنی، دانشگاه گیلان، رشت، ایران

چکیده

امروزه مهمترین اقدام شرکت‌های بیمه در بحث بازاریابی و تبلیغات، بخش‌بندی و تفکیک مشتریان براساس رفتار و نیاز آن‌ها است. از این‌رو، این شرکت‌ها برای شناسایی و تحریک کردن مخاطبان خود، بازاریابی و تبلیغات را به‌طور گسترده و هدفمند در تمام محیط‌های ارتباطی به انجام می‌رسانند. برای اثربخشی هرچه بهتر این رویکرد، مشتریان براساس معیارها و اهداف خاصی تفکیک و بخش‌بندی می‌شوند. خوشه‌بندی روشی تحلیلی برای کشف عملکرد و رفتار مخاطبان از طریق اطلاعات آن‌ها است. این امر باعث می‌شود تا شرکت‌ها بتوانند از طریق همین عملکرد مخاطبان، دست به اتخاذ تصمیم و تبلیغات هدفمند نسبت به آن‌ها بزنند. هدف اصلی این پژوهش، ارائه راهکاری برای شناخت و پیش‌بینی عملکرد و رفتار مشتریان جدید در انتخاب نوع بیمه برای حفاظت مسکن خود در برابر مخاطرات، از طریق ترکیب روش K-medoids با شبکه‌های عصبی در جهت تعیین خوشه مشتریان جدید برای ارائه تبلیغ محصولات بیمه‌ای است. در این راستا، بدلیل زیاد بودن مشخصه‌ها در اکثر مجموعه داده‌ها و پراکندگی آن‌ها، ابتدا از طریق تکنیک‌های K-means و K-medoids به کشف الگوهای مفهومی رسیده و با استفاده از همین الگوها بعد از مشخص شدن خوشه مشتریان، فقط با داشتن اطلاعات جمعیت شناختی از سوی مشتریان جدید، خوشه آن‌ها پیش‌بینی و اقدامات لازم صورت می‌گیرد. ویژگی متمایز این پژوهش، ترکیب روش‌های خوشه‌بندی با روش‌های دسته‌بندی در کشف الگو است. آزمایش‌های انجام شده، موفقیت روش پیشنهادی در شناخت و کشف نیازها، همچنین رفتار و عملکرد مشتریان را نشان می‌دهد که براساس آن تبلیغات صورت می‌گیرد.

کلمات کلیدی: بازاریابی و تبلیغات، خوشه‌بندی، شبکه‌های عصبی، K-medoids، K-means.

۱- مقدمه

سازمان‌ها این امکان را می‌دهد که بتوانند استراتژی‌های بازاریابی خود را براساس این بخش‌ها تنظیم کنند که تبلیغات یکی از مهمترین ابزار بر روی این بخش‌ها است.

همچنین با برتری رقابتی در این بخش‌ها، بقای خود را در محیط کنونی تضمین کنند [۵]. هدف از خوشه‌بندی^۴ داده‌های مشتریان این است که داده‌ها را به خوشه‌هایی تقسیم کنیم تا داده‌های درون یک خوشه دارای بیشترین شباهت و داده‌های خوشه‌های مختلف دارای کمترین شباهت باشند. مشتریان براساس ویژگی‌های رفتاری^۵، دموگرافیک^۶، جغرافیایی^۷ و روانشناختی^۸ در دسته‌های مجزا بخش‌بندی می‌شوند [۵]. فاکتورهای روان شناختی تأثیرگذار بر رفتار خرید مشتری به عنوان مدلی محاسباتی برای قصد خرید مشتری، در بخش‌بندی بازار مورد بررسی قرار گرفته‌اند [۶]. لذا در دنیای امروز استفاده از سیستم‌هایی همچون مدیریت ارتباط با مشتری تنها یک مزیت رقابتی نیست بلکه یک ضرورت برای سازمان محسوب می‌شود. با کندوکاو داده‌های مربوط به مشتریان، به رکوردهای

بسیاری از صاحب‌نظران عرصه بازار، تقسیم بازار را نوشداروی بازاریابی مدرن دانسته، علت آن را کمبود منابع ذکر می‌کنند [۱]. منطق نهفته در تقسیم بازار، ناهمگنی ترجیح محصولات و رفتار خرید مشتریان بوده، و این تفاوت‌ها معمولاً با اختلافات در محصولات یا مصرف‌کنندگان توضیح داده می‌شود [۲]. افزایش تعداد سازمان‌ها و در نتیجه، تشدید هر چه بیشتر رقابت بین آن‌ها، سازمان‌ها را مجبور کرده است که برای بهبود عملیات تبلیغاتی خود، سرمایه‌گذاری بیشتری انجام دهند. یکی از روش‌هایی که سازمان‌ها برای بهبود عملیات تبلیغاتی خود به کار بسته‌اند، بازاریابی هدف‌دار^۱ است [۳، ۴]. در بازاریابی هدف‌دار که در مقابل بازاریابی انبوه^۲ قرار می‌گیرد، سازمان‌ها فعالیت تبلیغاتی خود را (به جای تمامی مشتریان) بر روی گروه‌های خاصی از آن‌ها متمرکز می‌نمایند. بخش‌بندی بازار^۳ به

پیش‌بینی انتخاب گردشگران، پیش‌بینی رفتار گردشگران و پیش‌بینی تقاضا و تحلیل وفاداری میهمان‌های خارجی استفاده گردیده است [۱۳].

۲-۱- روش شناسی پژوهش

در این پژوهش نیز یک مدل بازاریابی مناسب برای انجام تبلیغات هدفمند بر روی مشتریان بیمه از طریق ترکیب روش K-medoids با شبکه‌های عصبی ارائه می‌شود که برای شرکت بیمه سودآور باشد. مجموعه داده شرکت بیمه سامان مربوط به شهر تهران، توسط ۲۵ نفر از نیروهای بازاریابی این شرکت طی ۵ ماه جمع‌آوری شده است. این مجموعه داده حاوی اطلاعاتی در رابطه با بیمه کردن واحدهای مسکونی و تجاری در برابر حوادث طبیعی و غیرطبیعی بکار رفته است. این مجموعه داده شامل ۳۰ ستون (مشخصه) و ۵۰۰ سطر (رکورد) که از سه بخش اصلی در قسمت مشخصه‌ها تشکیل یافته است. براساس این مشخصه‌ها و ارتباط آنها با یکدیگر خوشه‌بندی انجام گرفته است. این سه بخش شامل:

- مشخصه‌های مربوط به اطلاعات جمعیت شناختی مشتریان (سن، میزان درآمد، تحصیلات، نوع اشتغال، تعداد خانوار، میزان استفاده از اینترنت و میزان مطالعه روزنامه)
- مشخصه‌های مربوط به نوع بیمه‌نامه‌های خریداری شده در قبال بیمه مسکن موردنظر (سرقت، ترکیدگی، سیل، زلزله، آتش‌سوزی، رعد و برق)
- مشخصه‌های مربوط به اطلاعات محل شعبات و جایگاه‌های اختصاصی شرکت بیمه سامان در کل مراکز خرید شهر تهران (گلدیس، علاءالدین، اتکا، تیراژه، ایران زمین)

۳- تعیین پرتکرارترین نوع بیمه‌نامه‌ها از طریق روش FP-Growth

برای مشخص کردن پرتکرارترین نوع بیمه‌نامه‌های خریداری شده از روش‌های قواعد وابستگی^{۱۵} و تکنیک FP-Growth استفاده شده است. اطلاعات حاصل شده به‌عنوان کشف الگویی در رفتار خرید مشتریان است که همراه با اطلاعات جمعیت شناختی مشتریان وارد مرحله خوشه‌بندی می‌شوند. الگوریتم FP-Growth بر پایه یک ساختار جدید درخت الگوی تکرار شونده است که ساختار توسعه‌یافته‌ای از prefix-tree برای ذخیره اطلاعات فشرده و حساس درباره الگوهای تکرار شونده است. این الگوریتم، روش موثری برای یافتن مجموعه کاملی از الگوهای تکرار شونده با رشد الگو است [۱۴]. سپس از روش رشد الگو استفاده می‌شود تا از تولید پرهزینه تعداد زیاد مجموعه‌های کاندید جلوگیری شود. بعد از خوشه‌بندی، خوشه مشتریان براساس میزان ارزشی که برای شرکت بیمه دارند، نام‌گذاری شده‌اند.

بدین معنی که مشتریانی که بیشترین نوع بیمه را خریداری می‌کنند مشتریان وفادار، و مشتریان دیگر براساس ارزش‌شان نسبت به مشتریان وفادار مشخص شده و خوشه آن‌ها نام‌گذاری شده‌اند. مانند مشتریان فصلی، مشتریان حساس و مشتریان ضعیف. سپس با مشخص شدن وضعیت هر خوشه، همچنین با تعیین نام برای هر خوشه، نام خوشه‌ها این بار خود به عنوان ستون مشخصه هدف در مجموعه داده جدید اضافه شده و برچسب‌گذاری می‌گردند. سپس این مجموعه داده برای مدل‌سازی و پیش‌بینی نام خوشه مشتریان جدیدالورود وارد الگوریتم شبکه‌های عصبی می‌شود. به این صورت که با ورود مشتری جدید و تنها با داشتن اطلاعات مربوط به مشخصه جمعیت شناختی از سوی مشتریان، از طریق شبکه‌های عصبی نام خوشه برای مشتری جدید پیش‌بینی خواهد شد. نهایتاً تبلیغات و نوع بیمه‌های جدید براساس معیارها و متناسب با وضعیت هر خوشه به مشتریان جدید ارائه خواهد

اطلاعاتی مشتریان ساختار داده می‌شود. جریان تشخیص مشتریان با اهمیت به صورت خودکار صورت می‌گیرد که باعث تغییر در شیوه تشخیص مشتریان خاص و با ارزش از لیست کلیه مشتریان و در نهایت کشف مشتریان وفادار خواهد شد [۶].

۲- پیشینه پژوهش

تخمین درست از رفتار خرید مشتری یکی از مهمترین چالش‌ها در بستر بازاریابی است به گونه‌ای که با ترکیب روش‌های k-means و SOM^۱ مشتریان به چند بخش مختلف بخش‌بندی شده و با تخصیص مشتریان جدید به یکی از این بخش‌ها با استفاده از روش K-NN^{۱۱} دقت^{۱۱} نزدیک به ۹۰ درصد، قصد خرید مشتریان را به درستی تخمین زده است [۶].

همچنین با استفاده از CBR^{۱۲} روشی برای بخش‌بندی مشتریان ارائه و از الگوریتم ژنتیک برای بهبود دقت تخصیص یک مشتری جدید به بخش مربوط به خودش استفاده کرده‌اند [۷]. در این روش، ترکیبی از اطلاعات دموگرافیک مشتری و اطلاعات محصول خریداری شده، به عنوان ویژگی‌های توصیف‌کننده یک نمونه استفاده شده است. با استفاده از الگوریتم ژنتیک، نمونه‌ها و ویژگی‌هایی که کمتر معرف ویژگی‌های کلی مجموعه داده هستند، از پایگاه نمونه‌ها حذف شده‌اند. بنابراین الگوریتم با دقت بیشتری عمل بخش‌بندی مشتریان را انجام می‌دهد. با توجه به اطلاعاتی که در خصوص رفتار مشتریان از حیث میزان هزینه کردن بدست می‌آید، مشتریان در سه دسته مشتریان پرخرج، مشتریان کم خرج و مشتریان متوسط قرار می‌گیرند. هنگام ورود یک مشتری جدید، با استفاده از الگوریتم نزدیک‌ترین همسایه و پایگاه حاوی نمونه‌ها، رفتار این مشتری از نظر هزینه‌ای که احتمالاً در این شرکت خواهد کرد، تخمین زده می‌شود.

سازمان با شناخت این مشتریان، می‌تواند تمرکز خود را بر روی مشتریان با ارزش (مشتریانی که بیشتر خرج می‌کنند) قرار دهد و یا برای هر دسته، استراتژی بازاریابی متفاوتی را به کار گیرد [۸].

برخلاف بیشتر کارهای صورت گرفته در زمینه بخش‌بندی مشتریان، که از اطلاعات دموگرافیک مشتریان به عنوان متغیرهای بخش‌بندی استفاده می‌کنند، اطلاعات مربوط به اقلام خریداری شده و قیمت هر قلم که به صورت رکوردهای تراکنشی برای هر مشتری در دسترس را به عنوان متغیرهای بخش‌بندی مورد استفاده قرار می‌دهند و با ترکیب الگوریتم ژنتیک و الگوریتم خوشه‌بندی K-means، مشتریان را از لحاظ رفتار خرید، خوشه‌بندی و در نهایت برای تحلیل میزان سودبخشی مشتریان هر خوشه از RFM^{۱۳} استفاده می‌کنند. با داشتن میزان سودبخشی هر خوشه، سازمان می‌تواند مشتریان هدف را به سرعت شناسایی کرده و محصولات و خدمات، و منابع مناسبی را به خوشه‌های هدف تخصیص دهد [۹]. در اطلاعات مربوط به بانکداری برای بازاریابی بانکی، نحوه به‌کارگیری روش‌های داده‌کاوی را به منظور انتخاب مجموعه مناسبی از مشتریان بررسی کرده‌اند که یک بانک با در اختیار داشتن اطلاعات مشتریانش، می‌خواهد تعیین کند که در صورت تماس گرفتن با آن‌ها و پیشنهاد افتتاح سپرده مدت‌دار با نرخ سود بالا در آن بانک، چه مشتریانی اقدام به افتتاح حساب خواهند نمود. به عبارت دیگر، بانک برای پیش‌بینی نتیجه تماس با مشتریانش مدلی می‌سازد. این مدل می‌تواند از طریق مدیریت بهتر منابع سازمان (نظیر منابع انسانی، تماس‌های تلفنی، و زمان) کارایی عملیات تبلیغاتی را بهبود بخشد [۱۰]. همچنین یکی از روش‌های قابل استفاده در تقسیم بازار، استفاده از شبکه‌های عصبی مصنوعی^{۱۴} است. علت استفاده از آن، انعطاف‌پذیری در ساختن مدل و توانایی‌اش در استفاده از اطلاعات جدید است [۱۱]. از نمونه‌های مطالعات تقسیم بازار که از شبکه‌های عصبی با نقشه‌های خودسازمانده استفاده کرده‌اند می‌توان تقسیم بازار گردشگران در اتریش و تقسیم گردشگران برتر در استرالیا را نام برد [۱۲]. البته از سایر روشهای شبکه عصبی در

کرده‌اند مشخص شده است. از سه بخش اصلی در مجموعه داده، تنها مشخصه‌های بخش دوم که مربوط به نوع بیمه‌نامه واحدهای مسکونی و تجاری در برابر حوادث است و مشتریان آن‌ها را برای خانه‌های خود خریداری کرده‌اند وارد الگوریتم FP-Growth می‌شود. با این کار آن دسته از بیمه‌نامه‌هایی که کمتر خریداری شده‌اند و از دید مشتریان کمتر حائز اهمیت هستند مشخص شده و از بین مابقی بیمه‌نامه‌ها جدا می‌شوند.

قواعد استخراج شده با درجه پشتیبان و درجه اطمینان بالا و به صورت تجربی انتخاب شده است. که به درجه پشتیبان ۶۰ درصد و درجه اطمینان ۷۰ درصد حاصل شده‌اند. همچنین مقدار لیفت برای هر کدام از قواعد بیان شده است قواعد استخراج شده عبارتند از:

Association Rules

[سرقت] --> [ترکیدگی] (confidence:0.787) (lift 100%)

[آتش سوزی] --> [سرقت] (confidence:0.819) (lift 94%)

[زلزله] --> [سرقت] (confidence:0.853) (lift 98%)

[ترکیدگی] --> [سرقت] (confidence:0.904) (lift 100%)

در قواعد استخراج شده چهار قانون استخراج شده است:

قانون اول: کسانی که خانه خود را در برابر سرقت بیمه کرده‌اند، آنگاه با درجه اطمینان ۷۸ درصد خانه خود را در برابر خطر ترکیدگی لوله نیز بیمه کرده‌اند (نوع بیمه ترکیدگی را نیز خریداری کرده‌اند).

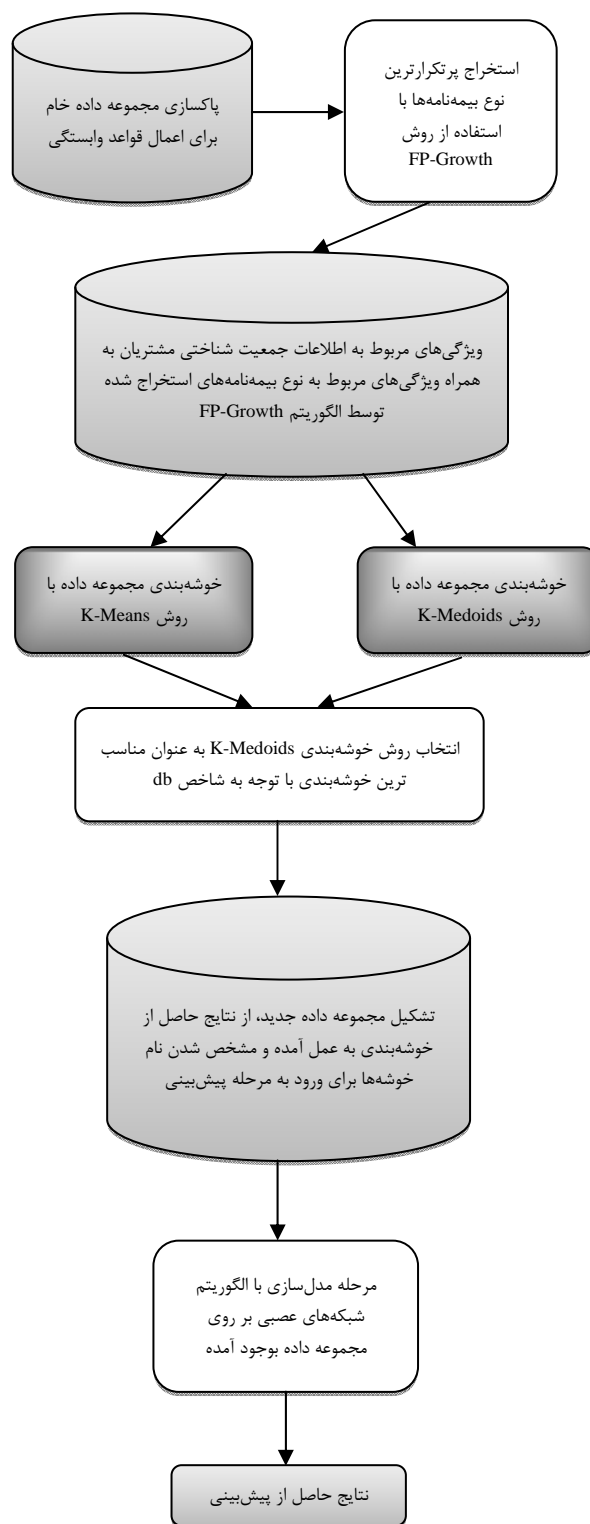
قانون دوم: کسانی که خانه خود را در برابر آتش‌سوزی بیمه کرده‌اند، آنگاه با درجه اطمینان ۸۱ درصد خانه خود را در برابر خطر سرقت نیز بیمه کرده‌اند (نوع بیمه سرقت را نیز خریداری کرده‌اند).

قانون سوم: کسانی که خانه خود را در برابر زلزله بیمه کرده‌اند، آنگاه با درجه اطمینان ۸۵ درصد خانه خود را در برابر خطر سرقت نیز بیمه کرده‌اند (نوع بیمه سرقت را نیز خریداری کرده‌اند).

قانون چهارم: کسانی که خانه خود را در برابر ترکیدگی لوله بیمه کرده‌اند، آنگاه با درجه اطمینان ۹۰ درصد خانه خود را در برابر خطر سرقت نیز بیمه کرده‌اند (نوع بیمه سرقت را نیز خریداری کرده‌اند).

با مشخص شدن پرتکرارترین نوع بیمه‌های خریداری شده، برای آشکار و تشریح دقیق‌تر خوشه‌ها نسبت به رفتار مشتریان در مرحله خوشه‌بندی، مشخصه‌های نوع بیمه‌نامه‌ها با توجه به نتایج حاصل از قواعد وابستگی در مرحله قبل، همراه با مشخصه‌های جمعیت شناختی وارد مرحله خوشه‌بندی می‌شوند. در این مرحله، عمل خوشه‌بندی با استفاده از الگوریتم K-means و K-medoids انجام می‌شود تا با توجه به معیارهای دقت و صحت خوشه، الگوریتمی که بهترین عملکرد و بالاترین دقت را در خوشه‌بندی انجام می‌دهد انتخاب شود. براساس نتایج حاصل از دقت و صحت خوشه‌بندی‌های صورت گرفته، روش K-medoids انتخاب می‌شود. این انتخاب براساس شاخص دیویس بولدین انجام شده است. این معیار از شباهت بین دو خوشه (Rij) استفاده می‌کند که براساس پراکندگی یک خوشه (si) و عدم شباهت بین دو خوشه (dij) تعریف می‌شود. این شاخص در واقع میانگین شباهت بین هر خوشه با شبیه‌ترین خوشه به آن را محاسبه می‌کند. می‌توان دریافت که هر چه مقدار این شاخص بیشتر باشد، خوشه‌های بهتری تولید شده است [۱۵]. با مقایسه الگوریتم K-means با K-medoids با توجه به شاخص دیویس بولدین دقت و صحت هر کدام از الگوریتم‌ها مبنی بر تعیین تعداد خوشه (K#) برای خوشه‌بندی مجموعه داده موردنظر مشخص شده است. بنابراین با توجه به معیار و شاخص دیویس بولدین الگوریتم K-medoids با عملکردی دقیق‌تر و مناسب‌تر نسبت به الگوریتم K-means، به عنوان الگوریتم خوشه‌بندی با تعیین ۴ خوشه انتخاب شده است. زیرا با توجه به شاخص دیویس بولدین هر چقدر شاخص به عدد ۱ نزدیکتر باشد، خوشه‌بندی با کیفیت‌تر است. جدول ۱ نتایج بررسی‌های

شد. نمودار کلی استخراج ویژگی و مراحل روش پیشنهادی در شکل ۱ نشان داده شده است.



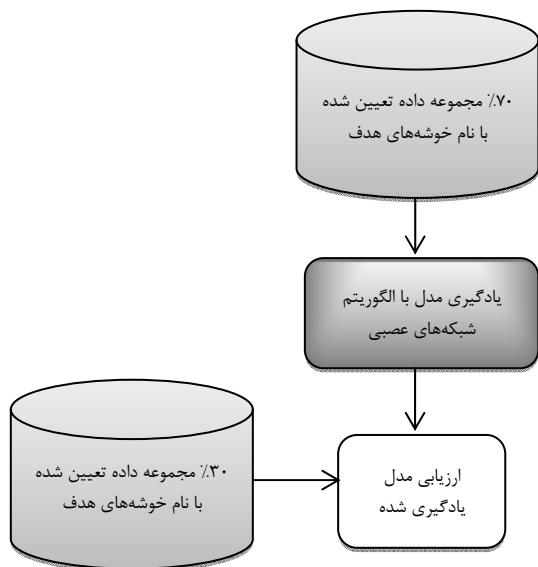
شکل ۱- نمودار کلی استخراج ویژگی‌ها و مراحل روش پیشنهادی

۳-۱- فرضیه‌ها و مدل مفهومی پژوهش

در راهکار پیشنهادی ترکیبی، پس از پاکسازی داده‌ها برای یافتن و پی‌بردن به پرتکرارترین نوع بیمه‌نامه‌های خریداری شده ابتدا از طریق قواعد وابستگی، نوع بیمه‌نامه‌هایی که افراد برای خانه‌های خود در برابر حوادث و مخاطرات انتخاب

خوشه شماره ۳ اکثراً سنی بین ۵۰ تا ۶۰ سال دارند، شغلشان مدیر است، درآمد آن‌ها بیش از ۳ میلیون تومان در ماه است، مدرک تحصیلی آن‌ها فوق‌لیسانس است، خانه خود را علاوه بر پرتکرارترین نوع بیمه‌نامه خریداری شده، بر تمامی بیمه‌نامه‌ها نیز بیمه کرده‌اند و معیارهای دیگری که با توجه به اطلاعات مربوط به هر خوشه، کارشناسان این خوشه را بنام خوشه مشتریان طلایی نام‌گذاری کرده‌اند.

همچنین خوشه‌های دیگر نیز با توجه به تجزیه و تحلیل هر خوشه توسط کارشناسان شرکت بیمه بنام مشتریان وفادار، مشتریان حساس و مشتریان ضعیف نام‌گذاری شده‌اند. پس از نام‌گذاری هر خوشه و مشخص شدن محل قرار گرفتن هر رکورد در خوشه‌بندی انجام شده، این بار همین مجموعه داده با اضافه شدن ستون نام خوشه به عنوان مشخصه یا متغیر هدف، برجسب‌گذاری شده و برای پیش‌بینی خوشه برای مشتریان جدید وارد الگوریتم شبکه‌های عصبی می‌شوند. نهایتاً با استفاده از ۷۰ درصد مجموعه داده به عنوان مجموعه داده آموزش جهت یادگیری مدل و انتخاب ۳۰ درصد از مجموعه داده به عنوان مجموعه داده آزمایش، الگوریتم شبکه‌های عصبی بر مجموعه داده حاصل شده اجرا می‌شود. شبکه‌های عصبی تا حد زیادی به عنوان جعبه سیاهی دیده شده‌اند که الگوی پیچیده در داده‌ها را مشخص می‌کنند و یادگیری از طریق آموزش از ویژگی‌های اساسی آن‌ها است [۱۶]. پس از اجرای الگوریتم بر روی مجموعه داده، خوشه‌ها برای مشتریان جدید با دقت ۸۷/۱۰ درصد پیش‌بینی می‌شوند. هر چند شبکه‌های عصبی مصنوعی محدودیت‌های خاص خود را دارند، اما آن‌ها دارای محاسن ویژه‌ای، همچون قدرت یادگیری، انعطاف‌پذیری، انطباق و کشف دانش هستند [۱۷]. نمودار کلی نحوه اعمال مدل پیشنهادی بر روی مجموعه داده در شکل ۳ ارائه شده است.



شکل ۳- نحوه اعمال مدل بر روی مجموعه داده (مرحله یادگیری)

۴- یافته‌ها و ارزیابی پژوهش

در راه کار پیشنهادی، برای بررسی شاخص دقت در نتیجه مدل‌سازی شبکه‌های عصبی از معیار دقت یا نرخ دسته‌بندی استفاده می‌شود:

$$\text{precision} = \frac{TP}{TP+FP} \quad (۱)$$

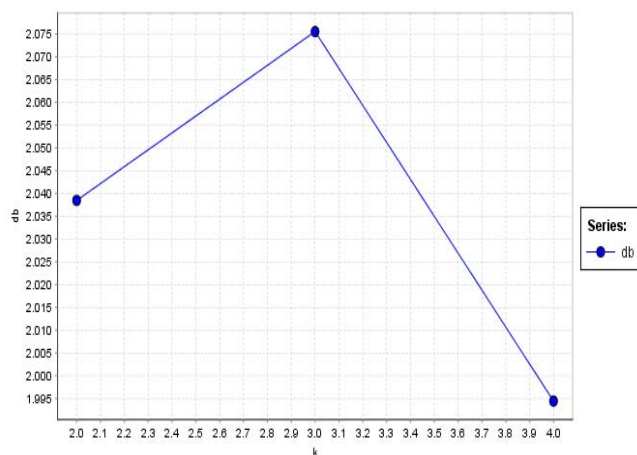
$$\text{recall} = \frac{TP}{FN+TP} \quad (۲)$$

حاصل از انتخاب الگوریتم K-medoids را بجای الگوریتم K-means نشان می‌دهد.

جدول ۱- نتایج مقایسه عملکرد الگوریتم K-medoids و K-means از لحاظ تعیین تعداد بهترین خوشه

K-medoids	
# Generated by Log[com.rapidminer.datatable.SimpleDataTable]	
# kdb	
2	2.038455301456895
3	2.0755253687899238
4	1.9144175531970258
5	1.9424645897881654
6	1.9514199199455575

K-means	
# Generated by Log[com.rapidminer.datatable.SimpleDataTable]	
# k db	
2	2.533608807984115
3	3.068261052728064
4	3.140990808692302
5	3.189741241276476
6	3.021981090479843



شکل ۲- نمودار شاخص db در الگوریتم K-medoids

شکل ۲ نمودار تعیین، بهترین تعداد خوشه را با توجه به شاخص دیویس بولدین در الگوریتم K-medoids نشان می‌دهد. با مشخص شدن روش خوشه‌بندی K-medoids به عنوان مناسب‌ترین روش خوشه‌بندی برای مجموعه داده موردنظر، نتایج خوشه‌بندی به این صورت حاصل شده:

Cluster Model
 Cluster 0: 147 items
 Cluster 1: 130 items
 Cluster 2: 64 items
 Cluster 3: 73 items
 Total number of items: 414

با توجه به بهترین حالت خوشه‌بندی مطابق با شاخص دیویس بولدین، داده‌ها در بهترین وضعیت در ۴ خوشه، خوشه‌بندی می‌شوند. سپس بعد از بررسی‌های به عمل آمده توسط کارشناسان شرکت بیمه بر روی محتویات هر خوشه، برای هر خوشه نامی تعیین می‌شود تا معرف آن خوشه باشد. به عنوان مثال: مشتریان

پرتکرارترین نوع بیمه‌نامه‌های خریداری شده در هر خوشه، همان بیمه‌نامه‌ها و بیمه‌نامه‌های جدید بدون آگاهی از نیاز و انتخاب مشتری، متناسب با وضعیت خوشه‌های پیش‌بینی شده برای مشتریان ارائه می‌شود.

این کار فقط با داشتن اطلاعات جمعیت شناختی مشتریان انجام می‌شود و بهترین راه تبلیغ محصولات و نوع بیمه‌نامه‌های جدید برای مشتریان و مشتریان جدید است. برای پیاده‌سازی مدل پیشنهاد شده از نرم‌افزار RapidMiner استفاده کرده‌ایم که یک نرم‌افزار مجتمع و مطلوب برای انجام تکنیک‌های خوشه‌بندی و دسته‌بندی است.

۴-۱- تحلیل خوشه‌ها بر روی مجموعه داده‌های هدف

واژه تحلیل خوشه‌ای برای اولین بار توسط ترایون استفاده شد. تحلیل خوشه‌ای، شامل مجموعه‌ای از الگوریتم‌ها و روش‌ها است که جهت گروه‌بندی موضوعات یا اشیای مشابه در طبقه‌های مرتبط استفاده می‌شود [۲۰].

بر اساس اهداف و معیارهای مختلف، تحلیل‌های خاصی بر روی خوشه‌ها انجام می‌گیرد که در نهایت با توجه به نتایج و خروجی خوشه‌ها، کارشناسان و متخصصان دست به اتخاذ تصمیم می‌زنند و برنامه عملیاتی خود را انجام می‌دهند. با توجه به خوشه‌بندی‌های انجام شده بر روی مجموعه داده‌های این پژوهش می‌توان بر اساس نیاز، تصمیمات مهمی را در امر بازاریابی و تبلیغات اتخاذ کرد:

- طبق خوشه‌بندی‌هایی که بر پایه اطلاعات جمعیت شناختی حاصل شده بهترین بستر تبلیغات بیمه را تشخیص و آن راه را برای انجام تبلیغ انجام داد و یا از طریق همین خوشه‌بندی از میزان سطح سواد و درآمد افراد آگاهی پیدا کرد و نوع بیمه و شرایط آن را جهت گرفتن بیمه طی تبلیغات هدفمندی به افراد پیشنهاد کرد و موارد گوناگونی که به هدف شرکت و یا سازمان مرتبط است.

- با استفاده از خوشه‌بندی‌هایی که بر پایه اطلاعات نوع و انتخاب بیمه توسط افراد مختلف انجام شده به کمترین و بیشترین انتخاب نوع بیمه دست پیدا کرد و بسته به شرایط افراد و موقعیت‌شان نوع بیمه را به مشتریان قدیمی و یا مشتریان جدید که با محصولات بیمه آشنایی ندارند از طریق تبلیغات طبقه‌بندی شده‌ای پیشنهاد کرد.

- مطابق با خوشه‌بندی‌هایی که بر پایه اطلاعات شعبات مختلف بیمه که در سطح شهر پراکنده شده‌اند به میزان دسترسی و مراجعه افراد به شعبات مختلف می‌توان پی برد و تبلیغات مختص با نقاط پر رفت و آمد و نقاط خلوت انجام داد.

این‌ها تنها مواردی از ارائه تبلیغ بر پایه خوشه‌بندی بود، بر اساس نیازهای مختلف و تحلیل‌های خاصی که بر روی خوشه‌ها انجام می‌شود هدف‌های راهبردی خاصی انجام می‌شود. همچنین از تکنیک‌های مختلف داده کاوی می‌توان در قسمت دسته‌بندی برای انجام پیش‌بینی‌ها گوناگون استفاده کرد که منجر به نتایج جالب توجه و مفیدی در بحث بازاریابی و ارائه تبلیغات می‌شود.

۵- نتیجه‌گیری و پیشنهادها

به دلیل افزایش رقابت بین سازمان‌ها و فعالیت‌های تبلیغاتی، بخش‌بندی بازار اهمیت ویژه‌ای برای تمرکز روی گروه خاصی از مشتریان و لذا کاهش هزینه‌ها پیدا کرده است. روش پیشنهادی ترکیب روش K-medoids با روش شبکه‌های عصبی در بازاریابی شرکت بیمه برای پیش‌بینی خوشه مشتریان استفاده می‌شود. نتایج آزمایش انجام شده بر روی داده‌های یک شرکت بیمه نشان می‌دهد که تکنیک

TN: این مقدار بیانگر تعداد رکوردهایی است که دسته واقعی آن‌ها منفی بوده و الگوریتم دسته‌بندی نیز دسته آن‌ها را به درستی منفی تشخیص داده است. FP: این مقدار بیانگر تعداد رکوردهایی است که دسته واقعی آن‌ها منفی بوده و الگوریتم دسته‌بندی، دسته آن‌ها را به اشتباه مثبت تشخیص داده است. FN: این مقدار بیانگر تعداد رکوردهایی است که دسته واقعی آن‌ها مثبت بوده و الگوریتم دسته‌بندی، دسته آن‌ها را به اشتباه منفی تشخیص داده است. TP: این مقدار بیانگر تعداد رکوردهایی است که دسته واقعی آن‌ها مثبت بوده و الگوریتم دسته‌بندی نیز دسته آن‌ها را به درستی مثبت تشخیص داده است.

معیار Recall دقت دسته‌بندی دسته x را با توجه به کل رکوردهای با برچسب x نشان می‌دهد. معیار Precision دقت دسته‌بندی دسته x را با توجه به مواردی نشان می‌دهد که برچسب x برای رکورد مورد بررسی توسط دسته‌بندی پیشنهاد شده است [۱۸]. جدول ۲ میزان دقت در مدل شبکه‌های عصبی را نشان می‌دهد که برابر است با:

جدول ۲- میزان دقت در مدل شبکه‌های عصبی

	Weighted mean precision	Weighted mean recall	Accuracy
مدل شبکه عصبی	۹۵.۷۳ درصد	۹۳.۵۲ درصد	۹۷.۱۰ درصد

همچنین جدول ۳ نتایج حاصل از میزان دقت در پیش‌بینی خوشه هر رکورد از مجموعه داده آزمایش را نشان می‌دهد.

جدول ۳- میزان دقت در پیش‌بینی هر خوشه

	True cluster 1 (مشتریان وفادار)	True cluster 0 (مشتریان ضعیف)	True cluster 2 (مشتریان حساس)	True cluster 3 (مشتریان طلایی)	Class precision
پیش‌بینی خوشه ۱	۳۳	۱	۴	۳	۹۰.۴۹ درصد
پیش‌بینی خوشه ۲	۱	۴۵	۱	۰	۹۷.۷۴ درصد
پیش‌بینی خوشه ۳	۱	۱	۱۴	۱	۹۲.۳۵ درصد
پیش‌بینی خوشه ۴	۱	۰	۲	۱۶	۹۴.۲۱ درصد
Class recall	۹۸.۶۷ درصد	۹۹.۷۴ درصد	۷۶.۶۷ درصد	۹۰.۰۰ درصد	

ارزیابی خوشه‌ای اندازه‌گیری، میزان برتری یک خوشه‌بندی نسبت به خوشه بندی‌های دیگر به وسیله الگوریتم‌های متفاوت خوشه‌بندی یا الگوریتم‌های مشابه ولی با مقدار پارامترهای متفاوت است [۱۹]. با توجه به نتایج و معیارهای حاصل شده، تنها با داشتن برخی اطلاعات جمعیت شناختی از قبیل سن، میزان درآمد و میزان تحصیلات یا میزان استفاده از اینترنت در روز با استفاده از مدل‌سازی شبکه‌های عصبی، خوشه مورد نظر مطابق با محتوای اطلاعاتی درون خوشه برای مشتریان جدید پیش‌بینی و بر اساس سیاست‌ها و استراتژی‌های بازاریابی صورت می‌گیرد.

همچنین تبلیغات مورد نظر برای ارائه محصولات و بیمه‌نامه‌های جدید متناسب با وضعیت مشتریان هر خوشه به آنها ارائه می‌شود. همچنین با بررسی

[10] W. Kuang-Wei, and K. Peng. "Market segmentation via structured click stream analysis," *Industrial Management & Data Systems*, vol. 102, no. 9, pp. 493-502, 2002.

[11] J. A. Mazanec, "Classifying tourists into market segments: A neural network approach," *Journal of Travel & Tourism Marketing*, vol. 1, no.1, pp. 39-60, 1992.

[12] J. Kim, SH. Wei, and H. Ruys, "Segmenting the market of West Australian senior tourists using an artificial neural network," *Tourism Management*, vol. 24, no. 1, pp. 25-34, 2003.

[13] R. Law, and N. Au, "A neural network model to forecast Japanese demand for travel to Hong Kong," *Tourism Management*, vol. 20, no. 1, pp. 89-97, 1997.

[14] H. Lijun, and et. al., "Comparison and Analysis of algorithms for association rules," *Database Technology and Applications, First International Workshop on, IEEE*, 2009.

[15] D. L. Davies, and D. W. Bouldin, "A cluster separation measure," *IEEE transactions on pattern analysis and machine intelligence*, vol. 1, pp. 224-227, 1979.

[16] S. Walczak, and N. Cerpa, "Heuristic principles for the design of artificial neural networks," *Information and software technology*, vol. 41, no. 2, pp. 107-117, 1999.

[17] S. Goonatilake, and C. P. Treleaven, *Intelligent systems for finance and business*, New York, USA: John Wiley & Sons, Inc, 1995.

[18] D. Powers, and D. Martin, "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation," 2011.

[19] P. K. Roy, and et. al., "Automated Segmentation of White Matter Lesions Using Global Neighbourhood Given Contrast Feature-Based Random Forest and Markov Random Field," *Healthcare Informatics (ICHI), IEEE International Conference on, IEEE*, 2014.

[20] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural networks*, vol. 2, no. 5, pp. 359-366, 1989.

احسان مختاری فارغ‌التحصیل رشته فناوری اطلاعات گرایش تجارت الکترونیکی از دانشگاه گیلان هستند و زمینه‌های مورد علاقه ایشان داده‌کاوی، متن‌کاوی و هوش تجاری هستند.
آدرس پست‌الکترونیکی ایشان عبارت است از:
ehsan_m63000@yahoo.com



سید ابوالقاسم میرروشندل فارغ‌التحصیل از دانشکده فنی دانشگاه تهران در مقطع کارشناسی در رشته مهندسی کامپیوتر با گرایش نرم‌افزار و دانشگاه صنعتی شریف در مقاطع کارشناسی‌ارشد و دکتری در رشته مهندسی کامپیوتر گرایش هوش مصنوعی. از سال ۱۳۹۱ عضو هیات



پیشنهادی راهکاری جدید با دقت نسبتاً خوب برای پیش‌بینی عملکرد رفتاری مشتریان برای ارائه تصمیمات مختلف است.

همچنین با این تکنیک دقت در پیش‌بینی به مراتب بالاتر از مواقعی است که متغیر هدفی در مجموعه داده نیست. لذا، استفاده از روش K-medoids تنها برای تشریح برخی مسائل یا شبکه‌های عصبی برای پیش‌بینی یک عمل به تنهایی دقت بالایی ندارد و تکنیک‌های اختصاصی‌تری نظیر ترکیب K-medoids با الگوریتم درخت تصمیم و یا تکنیک‌های قواعد وابستگی بسیار مناسب‌تر از الگوریتم‌های عام منظوره هستند. در آینده می‌توان کارایی این روش پیشنهادی را در کاربردهای دیگر بازاریابی مورد ارزیابی قرار داد. به علاوه، ترکیب روش پیشنهادی با سایر روش‌های موجود، با هدف افزایش کارایی و یا دقت آن نیز می‌تواند موضوع پژوهشی مناسبی برای کار در آینده باشد. می‌توان شکل‌های مختلف ترکیب تکنیک را نیز به مجموعه قواعد مورد استفاده اضافه کرد و شکل پیچیده‌تری را نیز برای فرضیه‌ها تعریف کرد. بدین صورت فضای فرضیات بزرگتری را مورد جستجو قرار داده و لذا احتمال یافتن فرضیه‌های بهتر با دقت پیش‌بینی بالاتر افزایش می‌یابد.

مراجع

[1] Y. Wind, "Issues and advances in segmentation research," *Journal of marketing research*, vol. 15, no. 3, pp. 317-337, 1978.

[2] V. Mahajan, and A. K. Jain, "An approach to normative segmentation," *Journal of Marketing Research*, vol. 15, no. 3, pp. 338-345, 1978.

[3] D. Chaffey, and P. R. Smith, "eMarketingXcellence: Planning and optimizing your digital marketing," Routledge, 2013.

[4] D. Chaffey, and et. al., "Internet marketing: strategy, implementation and practice," Pearson Education, 2009.

[5] S. Goyat, "The basis of market segmentation: a critical review of literature," 2011.

[6] T. Hong, and E. Kim, "Segmenting customers in online stores based on factors that affect the customer's intention to purchase," *Expert Systems with Applications*, vol. 39, no. 2, pp. 2127-2131, 2012.

[7] Y. Chen, CH. Yi Wang, and Y. Feng, "Application of a 3NN+ 1 based CBR system to segmentation of the notebook computers market," *Expert Systems with Applications*, vol. 37, no. 1, pp. 276-281, 2010.

[8] C-Y. Tsai, and C-C. Chiu, "A purchase-based market segmentation methodology," *Expert Systems with Applications*, vol. 27, no. 2, pp. 265-276, 2004.

[9] M. Sergio, R. Laureano, and P. Cortez, "Using data mining for bank direct marketing: An application of the crisp-dm methodology," *Proceedings of European Simulation and Modelling Conference-ESM, Eurosis*, 2011.

علمی گروه مهندسی کامپیوتر دانشگاه گیلان بوده و زمینه‌های مورد علاقه ایشان پردازش زبان‌های طبیعی، داده‌کاوی، یادگیری ماشینی و پردازش تصویر هستند. آدرس پست‌الکترونیکی ایشان عبارت است از:

mirroshandel@guilan.ac.ir

اطلاعات بررسی مقاله:

تاریخ ارسال: ۱۳۹۵/۱۰/۰۱

تاریخ اصلاح: ۱۳۹۵/۱۱/۰۸

تاریخ قبول شدن: ۱۳۹۵/۱۱/۳۰

نویسنده مرتبط: دکتر سید ابوالقاسم میرروشندل، دانشکده فنی، دانشگاه گیلان، رشت، ایران.

-
- ¹ Target Marketing
 - ² Mas Marketing
 - ³ Market Segmentation
 - ⁴ Clustering
 - ⁵ Behavioral
 - ⁶ Demographic
 - ⁷ Geographic
 - ⁸ Psychographic
 - ⁹ Self-Organizing Map
 - ¹⁰ K-Nearest Neighbors
 - ¹¹ Accuracy
 - ¹² Case-Based Reasoning
 - ¹³ Recency, Frequency, Monetary
 - ¹⁴ Artificial Neural Networks
 - ¹⁵ Association Rule