Archive of SID

پیش بینی رفتار مشتریان بیمه با استفاده از ترکیب روش های خوشه بندی و دسته بندی

احسان مختاری[1*]، سید ابوالقاسم میرروشندل[2]،
[1*] دانشجوی کارشناسی ارشد مهندسی فناوری اطلاعات، دانشگاه گیلان ، رشت، 09123410507
ehsan_m63000@yahoo.com
[2] استادیار، دانشکده فنی دانشگاه گیلان، mirroshandel@gmail.com

چکیده

امروزه مهمترین اقدام شرکت های بیمه در بحث بازاریابی و تبلیغات، بخش بندی و تفکیک مشتریان بر اساس رفتار و نیاز آن ها است. این اقدام برای ارائه و معرفی بیمه نامه ها و محصولات جدید، همچنین افزایش فروش بیمه نامه ها و ارتباطات موثر با مخاطبان به صورت بازاریابی هدف دار انجام می گیرد. از این رو، این شرکت ها برای شناسایی و تحریک کردن مخاطبان خود، بازاریابی و تبلیغات را به طور گسترده و هدفمند در تمام محیط های ارتباطی به انجام می رسانند. برای اثر بخشی هرچه بهتر این رویکرد، مشتریان بر اساس معیارها و اهداف خاصی تفکیک و بخش بندی می شوند. خوشه بندی روشی تحلیلی برای کشف عملکرد و رفتار مخاطبان از طریق اطلاعات آن ها است. این امر باعث می شود تا شرکت ها بتوانند از طریق همین عملکرد مخاطبان، دست به اتخاذ تصمیم و تبلیغات هدفمند نسبت به آن‌ها بزنند. هدف اصلی این پژوهش، ارائه راه کاری برای شناخت و پیش بینی عملکرد و رفتار مشتریان جدید در انتخاب نوع بیمه برای حفاظت مسکن خود در برابر مخاطرات، از طریق ترکیب روش K-medoids با درخت تصمیم در جهت تعیین خوشه مشتریان جدید برای ارائه تبلیغ محصولات بیمه ای است. دراین راستا، بدلیل زیاد بودن مشخصه ها در اکثر مجموعه داده ها و پراکندگی آنها، ابتدا از طریق تکنیک های 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 پیش بینی و اقدامات لازم صورت می گیرد. ویژگی متمایز این پژوهش، ترکیب روش های خوشه بندی با روش های دسته بندی در کشف الگو است. آزمایش های انجام شده شناخت و کشف نیاز ها، رفتار و عملکرد مشتریان را نشان می دهد که بر اساس آن تبلیغات صورت می گیرد.

واژگان کلیدی: بازاریابی، تبلیغات هدفمند، تفکیک مشتریان، خوشه بندی، K-medoids، K-means، درخت تصمیم، کشف الگو.

www.SID.ir

Archive of SID

1- مقدمه

افزایش تعداد سازمان‌ها و در نتیجه، تشدید هرچه بیشتر رقابت بین آن‌ها، سازمان ها را مجبور کرده است که برای بهبود عملیات تبلیغاتی خود، سرمایه گذاری بیشتری انجام دهند. یکی از روش‌هایی که سازمان‌ها برای بهبود عملیات تبلیغاتی خود به کار بسته‌اند، بازاریابی هدف دار[1] است. در بازاریابی هدف دارکه در مقابل بازاریابی انبوه[2] قرار می گیرد، سازمان‌ها فعالیت تبلیغاتی خود را (به جای تمامی مشتریان) بر روی گروه‌های خاصی از آن‌ها متمرکز می‌نمایند. بخش بندی بازار[3] به سازمان‌ها این امکان را می‌دهد که بتوانند استراتژی های بازاریابی خود را بر اساس 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. هدف از خوشه بندی داده‌های مشتریان این است که داده ها را به خوشه‌هایی تقسیم کنیم تا داده‌های درون یک خوشه دارای بیشترین شباهت و داده های خوشه‌های مختلف دارای کمترین شباهت باشند. مشتریان بر اساس ویژگی‌های رفتاری[4] ، دموگرافیک[5] ، جغرافیایی[6] و روان شناختی[7] در دسته‌های مجزا بخش می‌شوند. فاکتورهای روان شناختی تأثیرگذار بر رفتار خرید مشتری به عنوان مدلی محاسباتی برای قصد خرید مشتری، در بخش بندی بازار مورد بررسی قرار گرفته اند. 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 کند وکاو داده‌های مربوط به مشتریان، به رکوردهای اطلاعاتی مشتریان ساختار داده می‌شود. جریان تشخیص مشتریان با اهمیت به صورت خودکار صورت می گیرد که باعث تغییر در شیوه تشخیص مشتریان خاص و با ارزش از لیست کلیه مشتریان و در نهایت کشف مشتریان وفادار خواهد شد. در این پژوهش نیز یک مدل بازاریابی مناسب برای انجام تبلیغات هدفمند بر روی مشتریان بیمه از طریق ترکیب روش K-medoids با درخت تصمیم ارائه

---

[1] Target marketing
[2] Mas marketing
[3] Market segmentation
[4] Behavioral
[5] Demographic
[6] Geographic
[7] Psycographic

www.SID.ir

Archive of SID

می شود که در عین حال برای شرکت بیمه سودآور باشد. در ادامه ساختار این مقاله به شکل زیر است: در بخش بعد، تاریخچه‌ای از پژوهش مورد مطالعه و تحقیقاتی که در این زمینه انجام شده، بررسی و در بخش 2-1 به تشریح و توصیف مجموعه داده استفاده شده پرداخته می شود و در قسمت 2-2 به توصیف روش FP-Growth پرداخته سپس در بخش 3 روش پیشنهادی را مطرح و در بخش های 4 ارزیابی از روش ارائه شده، محیط و ابزار پیاده سازی بیان می‌شوند. در پایان، در بخش های 5 و6 تحلیل خوشه‌ها و نتیجه گیری روش انجام شده را در بر خواهیم داشت.

2- تاریخچه پژوهش

در پژوهشی، با ترکیب روش‌های k-means و SOM[8] ، مشتریان را به چند بخش مختلف بخش بندی کرده و با تخصیص مشتریان جدید به یکی از این بخش‌ها با استفاده از روش K-NN[9] با دقت نزدیک به 90 درصد قصد خرید مشتریان به درستی تخمین زده است. در پژوهشی دیگر، با استفاده از CBR[10] روشی برای بخش بندی مشتریان ارائه و از الگوریتم ژنتیک برای بهبود دقت تخصیص یک مشتری جدید به بخش مربوط به خودش استفاده کرده اند. در این روش، ترکیبی از اطلاعات دموگرافیک مشتری و اطلاعات محصول خریداری شده، به عنوان ویژگی های توصیف کننده ی یک نمونه استفاده شده است. با استفاده از الگوریتم ژنتیک، نمونه ها و ویژگی‌هایی که کمتر معرف ویژگی‌های کلی مجموعه داده هستند، از پایگاه نمونه ها حذف می کند. 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 می گیرند. هنگام ورود یک مشتری جدید، با استفاده از الگوریتم نزدیک ترین همسایه و پایگاه حاوی نمونه‌ها، رفتار این مشتری از نظر هزینه‌ای که احتمالا 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 می کنند)قرار دهد و یا برای هر دسته، استراتژی بازاریابی متفاوتی را به کار گیرد. بر خلاف بیشتر کارهای صورت گرفته در زمینه

[8] Self-Organizing Map
[9] K-Nearest Neighbors
[10] Case-Based Reasoning

www.SID.ir

Archive of SID

بخش بندی مشتریان که از اطلاعات دموگرافیک مشتریان به عنوان متغیرهای بخش بندی استفاده می‌کنند، در پژوهشی اطلاعات مربوط به اقلام خریداری شده و قیمت هر قلم که به صورت رکوردهای تراکنشی برای هر مشتری در دسترس است، را به عنوان متغیرهای بخش بندی مورد استفاده قرار می دهند. آن‌ها با ترکیب الگوریتم ژنتیک و الگوریتم خوشه بندی K-means، مشتریان را از لحاظ رفتار خرید، خوشه بندی کرده و در نهایت برای تحلیل میزان سودبخشی مشتریان هر خوشه از RFM[11] استفاده می‌کنند. با داشتن میزان سودبخشی هر خوشه، سازمان می تواند مشتریان هدف را به سرعت شناسایی کرده و محصولات و خدمات، و منابع مناسبی را به خوشه‌های هدف تخصیص دهد. در پژوهشی مربوط به بانکداری برای بازاریابی بانکی، نحوه به کارگیری روش‌های داده کاوی را به منظور انتخاب مجموعه‌ی مناسبی از مشتریان بررسی کرده اند. آن‌ها فرض کرده اند که یک بانک با در اختیار داشتن اطلاعات مشتریانش، می خواهد تعیین کند که در صورت تماس گرفتن با آن‌ها و پیشنهاد افتتاح سپرده مدت دار با نرخ سود بالا در آن بانک، چه مشتریانی اقدام به افتتاح حساب خواهند نمود. به عبارت دیگر، بانک برای پیش بینی نتیجه تماس با مشتریانش مدلی می‌سازد. این مدل می‌تواند از طریق مدیریت بهتر منابع سازمان(نظیر منابع انسانی، تماس های تلفنی، و زمان) کارایی عملیات تبلیغاتی را بهبود ببخشد.

2-1- تشریح مجموعه داده به کار گرفته شده بر اساس مشخصه های ساختاری

مجموعه داده شرکت بیمه سامان مربوط به شهر تهران، توسط 25 نفر از نیروهای بازاریابی این شرکت طی 5 ماه جمع آوری شده است. این مجموعه داده حاوی اطلاعاتی در رابطه با بیمه کردن واحدهای مسکونی و تجاری در برابر حوادث طبیعی و غیر طبیعی بکار رفته است . این مجموعه داده شامل 30 ستون(مشخصه) و 500 سطر(رکورد) که از سه بخش اصلی در قسمت مشخصه‌ها تشکیل یافته است. بر اساس این مشخصه ها و ارتباط آن‌ها با یکدیگر خوشه بندی انجام گرفته است. این سه بخش شامل:

[11] Recency, Frequency, Monetary

www.SID.ir

Archive of SID

- مشخصه‌های مربوط به اطلاعات جمعیت شناختی مشتریان(سن، میزان در آمد، تحصیلات، نوع اشتغال، تعداد خانوار، میزان استفاده از اینترنت، میزان مطالعه روزنامه و...)
- مشخصه‌های مربوط به نوع بیمه نامه های خریداری شده در قبال بیمه مسکن مورد نظر(سرقت، ترکیدگی، سیل، زلزله، آتش سوزی، رعد و برق)
- مشخصه‌های مربوط به اطلاعات محل شعبات و جایگاه های اختصاصی شرکت بیمه سامان در کل مراکز خرید شهر تهران(گلدیس، علاءالدین، اتکا، تیراژه، ایران زمین)

2-2- تعیین پرتکرارترین نوع بیمه نامه‌ها از طریق روش FP-Growth

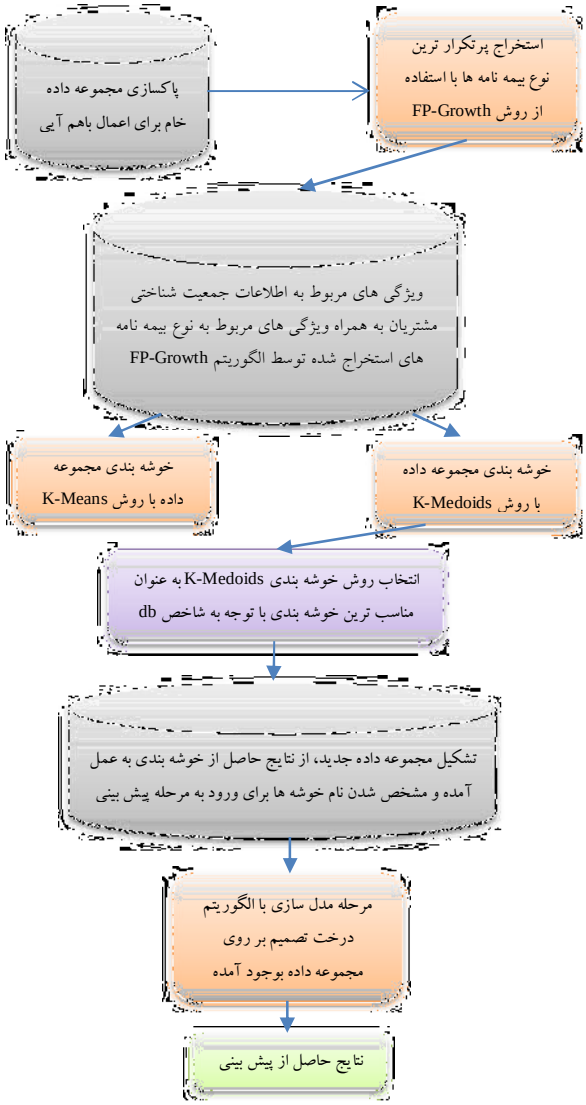
برای مشخص کردن پرترکرارترین نوع بیمه نامه‌های خریداری شده از روش های باهم آیی[12] و تکنیک FP-Growth اشتفاده شده است. اطلاعات حاصل شده بعنوان کشف الگویی در رفتار خرید مشتریان است که همراه با اطلاعات جمعیت شناختی مشتریان وارد مرحله خوشه بندی می شوند. الگوریتم FP-Growth بر پایه یک ساختار جدید درخت الگوی تکرار شونده است که ساختار توسعه یافته ای 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 می‌باشد. سپس از روش رشد الگو استفاده می‌شود تا از تولید پرهزینه تعداد زیاد مجموعه‌های کاندید جلوگیری شود. بعد از خوشه بندی، خوشه مشتریان بر اساس میزان ارزشی که برای شرکت بیمه دارند، نام گذاری شده‌اند. بدین معنی که مشتریانی که بیشترین نوع بیمه را خریده اند مشتریان وفادار، و مشتریان دیگر بر اساس ارزششان نسبت به مشتریان وفادار مشخص شده و خوشه آن‌ها نام گذاری شده اند. مانند مشتریان فصلی، مشتریان حساس، مشتریان ضعیف. سپس با مشخص شدن وضعیت هر خوشه، همچنین با تعیین نام برای هر خوشه، نام خوشه ها این بار خود به عنوان ستون مشخصه هدف در مجموعه داده جدید اضافه شده و برچسب گذاری می گردند. سپس این مجموعه داده برای مدل سازی و پیش بینی نام خوشه مشتریان جدید الورود وارد الگوریتم درخت تصمیم می شود. به این صورت که با ورود مشتری جدید و تنها با

[12] Association rules

www.SID.ir

Archive of SID

داشتن اطلاعات مربوط به مشخصه جمعیت شناختی از سوی مشتریان، از طریق درخت تصمیم نام خوشه برای مشتری جدید پیش بینی خواهد شد. نهایتاً تبلیغات و نوع بیمه های جدید بر اساس معیارها و متناسب با وضعیت هر خوشه به مشتریان جدید ارائه می شود. نمودار کلی استخراج ویژگی و مراحل روش پیشنهادی در شکل1 نشان داده شده است.



www.SID.ir

Archive of SID

شکل 1- نمودار کلی استخراج ویژگی ها و مراحل روش پیشنهادی

3- راهکار ترکیبی K-medoids با درخت تصمیم

در راهکار پیشنهادی ترکیبی، پس از پاکسازی داده‌ها برای یافتن و پی بردن به پرتکرارترین نوع بیمه نامه‌های خریداری شده ابتدا از طریق باهم آیی، نوع بیمه نامه‌هایی که افراد برای خانه‌های خود در برابر حوادث و مخاطرات انتخاب کرده اند مشخص شده است. از سه بخش اصلی در مجموعه داده، تنها مشخصه‌های بخش دوم که مربوط به نوع بیمه نامه واحدهای مسکونی و تجاری در برابر حوادث است و مشتریان آن‌ها را برای خانه های خود خریداری کرده اند وارد الگوریتم FP-Growth می شود. با این کار آن دسته از بیمه نامه‌هایی که کمتر خریداری شده اند و از دید مشتریان کمتر حائز اهمیت هستند مشخص شده و از بین مابقی بیمه نامه‌ها جدا می‌شوند. قواعد استخراج شده با درجه پشتیبان و درجه اطمینان بالا و به صورت تجربی انتخاب شده است. که به درجه پشتیبان 60٪ و درجه اطمینان 70٪ حاصل شده اند. قواعد استخراج شده عبارتند از:

Association Rules
[serghat] -- > [terekidegi] (confidence: 0.787)

[asa] -- > [serghat] (confidence: 0.819)

[earth] -- > [serghat] (confidence: 0.853)

[terekidegi] -- > [serghat] (confidence: 0.904)

در قواعد استخراج شده چهار قانون استخراج شده است:

- قانون اول: کسانی که خانه خود را در برابر سرقت بیمه کرده اند، آنگاه با درجه اطمینان 78٪ خانه خود را در برابر خطر ترکیدگی لوله نیز بیمه کرده اند(نوع بیمه ترکیدگی را نیز خریداری کرده اند).
- قانون دوم: کسانی که خانه خود را در برابر آتش سوزی بیمه کرده اند، آنگاه با درجه اطمینان 81٪ خانه خود را در برابر خطر سرقت نیز بیمه کرده اند(نوع بیمه سرقت را نیز خریداری کرده اند).

www.SID.ir

Archive of SID

- قانون سوم: کسانی که خانه خود را در برابر زلزله بیمه کرده اند، آنگاه با درجه اطمینان 85٪ خانه خود را در برابر خطر سرقت نیز بیمه کرده اند(نوع بیمه سرقت را نیز خریداری کرده اند).
- قانون چهارم: کسانی که خانه خود را در برابر ترکیدگی لوله بیمه کرده اند، آنگاه با درجه اطمینان 90٪ خانه خود را در برابر خطر سرقت نیز بیمه کرده اند(نوع بیمه سرقت را نیز خریداری کرده اند).

با مشخص شدن پرتکرارترین نوع بیمه‌های خریداری شده، برای آشکار و تشریح دقیق تر خوشه ها نسبت به رفتار مشتریان در مرحله خوشه بندی، مشخصه‌های نوع بیمه نامه‌ها با توجه به نتایج حاصل از باهم آیی در مرحله قبل، همراه با مشخصه‌های جمعیت شناختی وارد مرحله خوشه بندی می شوند. در این مرحله، عمل خوشه بندی با استفاده از الگوریتم K-means و K-medoids انجام می شود تا با توجه به معیارهای دقت و صحت خوشه، الگوریتمی که بهترین عملکرد و بالاترین دقت را در خوشه بندی انجام می‌دهد انتخاب شود. بر اساس نتایج حاصل از دقت و صحت خوشه بندی‌های صورت گرفته، روش K-medoids انتخاب می شود. این انتخاب بر اساس شاخص دیویس بولدین[13] انجام شده است. این معیار از شباهت بین دو خوشه (Rij) استفاده می‌کند که بر اساس 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 هرچه مقدار این شاخص بیشتر باشد، خوشه‌های بهتری تولید شده است. با مقایسه الگوریتم K-means با K-medoids با توجه به شاخص دیویس بولدین دقت و صحت هر کدام از الگوریتم ها مبنی بر تعیین تعداد خوشه(#K) برای خوشه بندی مجموعه داده مورد نظر مشخص شده است. بنابراین با توجه به معیار و شاخص دیویس بولدین الگوریتم K-medoids با عملکردی دقیق تر و مناسب تر نسبت به الگوریتم K-means، به عنوان الگوریتم خوشه بندی با تعیین 4 خوشه انتخاب شده است. زیرا با توجه به شاخص دیویس بولدین هر چقدر

[13] Davis Bouldin(db)

www.SID.ir

Archive of SID

شاخص به عدد 1 نزدیکتر باشد، خوشه بندی با کیفیت تر است. جدول 1 نتایج بررسی های حاصل از انتخاب الگوریتم K-medoids را بجای الگوریتم K-means نشان می دهد.

جدول 1- نتایج مقایسه عملکرد الگوریتم K-medoids و K-means از لحاظ تعیین تعداد بهترین خوشه

K-medoids

Generated by Log[com.rapidminer.datatable.SimpleDataTable]

# k	db
2	2.038455301456895
3	2.0755253687899238
4	1.9144175531970258
5	1.9424645897881654
6	1.9514199199455575

K-means

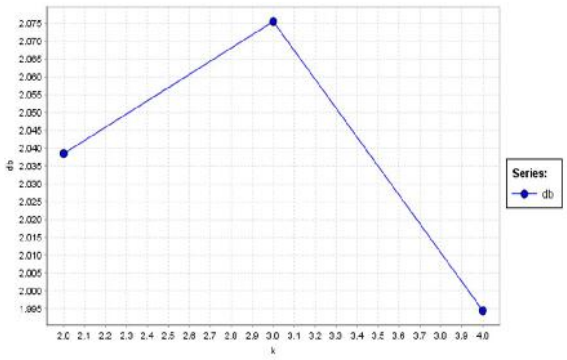
Generated by Log[com.rapidminer.datatable.SimpleDataTable]

# k	db
2	2.533608807984115
3	3.068261052728064
4	3.140990808692302
5	3.189741241276476
6	3.021981090479843

شکل 2 نمودار تعیین، بهترین تعداد خوشه را با توجه به شاخص دیویس بولدین در الگوریتم K-medoids نشان می دهد.

www.SID.ir

Archive of SID



شکل2- نمودار شاخص db در الگوریتم K-medoids

با مشخص شدن روش خوشه بندی K-medoids به عنوان مناسب ترین روش خوشه بندی برای مجموعه داده مورد نظر، نتایج خوشه بندی به این صورت حاصل شده:

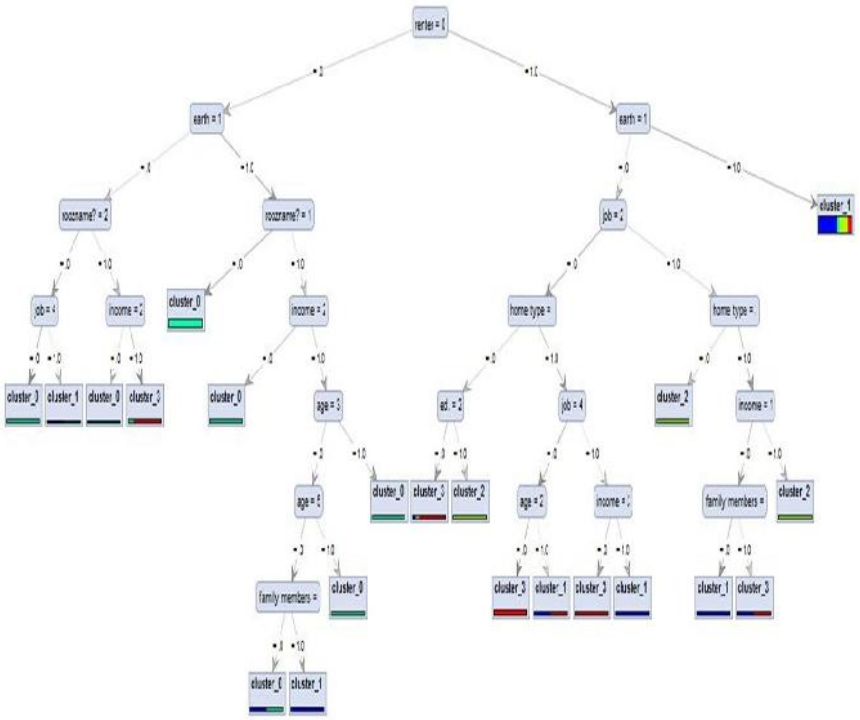
Cluster Model
Cluster 0: 147 items
Cluster 1: 130 items
Cluster 2: 64 items
Cluster 3: 73 items
Total number of items: 414

با توجه به بهترین حالت خوشه بندی مطابق با شاخص دیویس بولدین، داده ها در بهترین وضعیت در 4 خوشه، خوشه بندی می شوند. سپس بعد از بررسی های به عمل آمده توسط کارشناسان شرکت بیمه بر روی محتویات هر خوشه، برای هر خوشه نامی تعیین می شود تا معرف آن خوشه باشد. به عنوان مثال: مشتریان خوشه شماره 3 اکثراً سنی بین 50 تا 60 سال دارند، شغل شان مدیر است، درآمد آن‌ها بیش از 3 میلیون تومان در ماه است، مدرک تحصیلی آن‌ها فوق لیسانس است، خانه خود را علاوه بر پرتکرار ترین نوع بیمه نامه خریداری شده، بر تمامی بیمه نامه‌ها نیز بیمه کرده اند و معیار های دیگری که با توجه به اطلاعات مربوط به هر خوشه، کارشناسان این خوشه را بنام خوشه مشتریان طلایی نام گذاری کرده اند. همچنین خوشه های دیگر نیز با توجه به تجزیه

www.SID.ir

Archive of SID

و تحلیل هر خوشه توسط کارشناسان شرکت بیمه بنام مشتریان وفادار، مشتریان حساس و مشتریان ضعیف نام گذاری شده اند. شکل3 نمای کلی اجرای الگوریتم درخت تصمیم بر روی مجموعه داده و تعیین نام خوشه ها برای مشتریان جدید را نشان می دهد.



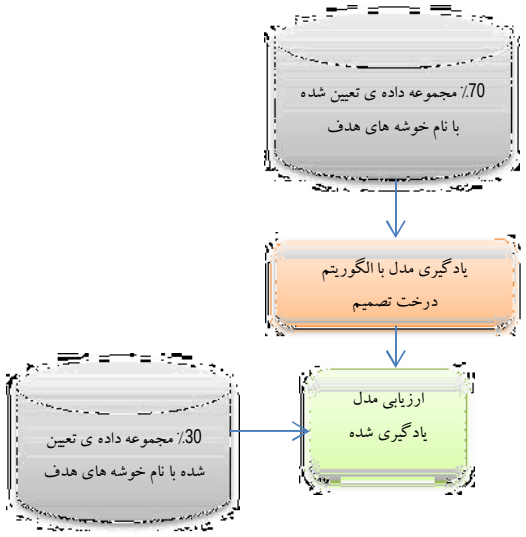
شکل 3 - نمای کلی اجرای الگوریتم درخت تصمیم

پس از نام گذاری هر خوشه و مشخص شدن محل قرار گرفتن هر رکورد در خوشه بندی انجام شده، این بار همین مجموعه داده با اضافه شدن ستون نام خوشه به عنوان مشخصه یا متغیر هدف، برچسب گذاری شده و برای پیش بینی خوشه برای مشتریان جدید وارد الگوریتم درخت تصمیم می شوند. نهایتاً با استفاده از 70٪ مجموعه داده به عنوان مجموعه داده آموزش جهت یادگیری مدل و انتخاب 30٪ از مجموعه داده به عنوان مجموعه داده آزمایش، الگوریتم درخت تصمیم بر

www.SID.ir

Archive of SID

مجموعه داده حاصل شده اجرا می شود. پس از اجرای الگوریتم درخت تصمیم بر روی مجموعه داده، خوشه ها برای مشتریان جدید با دقت ٪87,10 پیش بینی می شوند. همچنین نمودار کلی نحوه اعمال مدل پیشنهادی بر روی مجموعه داده در شکل 4 ارائه شده است.



شکل 4- نحوه اعمال مدل بر روی مجموعه داده (مرحله یادگیری)

4- ارزیابی

در راه کار پیشنهادی، برای بررسی شاخص دقت در نتیجه مدل سازی درخت تصمیم از معیار دقت[14] یا نرخ دسته بندی استفاده می شود:

$$Precision(+) = \frac{TP}{TP + FP}$$

$$Recall(+) = \frac{TP}{FN + TP}$$

[13] Accuracy

www.SID.ir

Archive of SID

TN: این مقدار بیانگر تعداد رکوردهایی است که دسته واقعی آن ها منفی بوده و الگوریتم دسته بندی نیز دسته آن ها را به درستی منفی تشخیص داده است.

FP: این مقدار بیانگر تعداد رکوردهایی است که دسته واقعی آن ها منفی بوده و الگوریتم دسته بندی، دسته آن ها را به اشتباه مثبت تشخیص داده است.

FN: این مقدار بیانگر تعداد رکوردهایی است که دسته واقعی آن ها مثبت بوده و الگوریتم دسته بندی، دسته آن ها را به اشتباه منفی تشخیص داده است.

TP: این مقدار بیانگر تعداد رکوردهایی است که دسته واقعی آن ها مثبت بوده و الگوریتم دسته بندی نیز دسته آن ها را به درستی مثبت تشخیص داده است.

معیار Recall(+) دقت دسته بندی دسته x را با توجه به کل رکوردهای با برچسب x نشان می دهد. معیارPrecision(+) دقت دسته بندی دسته x را با توجه به مواردی نشان می دهد که برچسب x برای رکورد مورد بررسی توسط دسته بند پیشنهاد شده است. جدول 2 میزان دقت در مدل درخت تصمیم را نشان می دهد که برابر است با:

جدول 2- میزان دقت در مدل درخت تصمیم

	Weighted mean precision	Weighted mean recall	Accuracy
مدل درخت تصمیم	**85,70%**	**83,52%**	**87,10%**

همچنین جدول 3 نتایج حاصل از میزان دقت در پیش بینی خوشه هر رکورد از مجموعه داده آزمایش را نشان می دهد.

www.SID.ir

Archive of SID

جدول3 - میزان دقت در پیش بینی هر خوشه

	True cluster 1 (مشتریان وفادار)	True cluster 0 (مشتریان ضعیف)	True cluster 2 (مشتریان حساس)	True cluster 3 (مشتریان طلایی)	Class precision
پیش بینی خوشه 1	33	1	4	3	80,49%
پیش بینی خوشه 2	1	45	1	0	95,74%
پیش بینی خوشه 3	1	1	14	1	82,35%
پیش بینی خوشه 4	1	0	2	16	84,21%
Class recall	91,67%	95,74%	66,67%	80,00%	

با توجه به نتایج و معیارهای حاصل شده، تنها با داشتن برخی اطلاعات جمعیت شناختی از قبیل سن، میزان درآمد، میزان تحصیلات یا میزان استفاده از اینترنت در روز و .. ، با استفاده از مدل سازی درخت تصمیم، خوشه مورد نظر مطابق با محتوای اطلاعاتی درون خوشه برای مشتریان جدید پیش بینی شده و بر اساس سیاست‌ها و استراتژی‌های بازاریابی، همچنین تبلیغات مورد نظر برای ارائه محصولات و بیمه نامه‌های جدید متناسب با وضعیت مشتریان هر خوشه به آن‌ها ارائه می شود. همچنین با بررسی پر تکرار ترین نوع بیمه نامه‌های خریداری شده در هر خوشه، همان بیمه نامه ها و بیمه نامه‌های جدید بدون آگاهی از نیاز و انتخاب مشتری، متناسب با وضعیت خوشه های پیش بینی شده برای مشتریان ارائه می‌شود. این کار فقط با داشتن اطلاعات جمعیت شناختی مشتریان انجام می‌شود. و بهترین راه تبلیغ محصولات و نوع بیمه نامه‌های جدید برای مشتریان و

www.SID.ir

Archive of SID

مشتریان جدید است. برای پیاده سازی مدل استفاده شده از نرم افزار Rapid Miner استفاده کرده ایم که یک نرم افزار مجتمع و مطلوب برای انجام تکنیک های خوشه بندی و دسته بندی است.

5- تحلیل خوشه‌ها بر روی مجموعه داده‌های هدف

بر اساس اهداف و معیار های مختلف، تحلیل های خاصی بر روی خوشه ها انجام می‌گیرد که در نهایت با توجه به نتایج و خروجی خوشه‌ها، کارشناسان و متخصصان دست به اتخاذ تصمیم می زنند و برنامه عملیاتی خود را انجام می‌دهند. با توجه به خوشه بندی‌های انجام شده بر روی مجموعه داده‌های این پژوهش می‌توان بر اساس نیاز، تصمیمات مهمی را در امر بازاریابی و تبلیغات اتخاذ کرد:

- می‌توان طبق خوشه بندی‌هایی 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 می‌توان از طریق همین خوشه بندی 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.

- می‌توان طبق خوشه بندی‌هایی 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 طبقه بندی شده ای پیشنهاد کرد.

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

این ها تنها مواردی از ارائه تبلیغ بر پایه خوشه بندی بود، بر اساس نیازهای مختلف و تحلیل‌های خاصی که بر روی خوشه‌ها اتجام می‌شود هدف های راهبردی خاصی انجام می‌شود. همچنین از

www.SID.ir

Archive of SID

تکنیک های مختلف داده کاوی می توان در قسمت دسته بندی برای انجام پیش بینی ها گوناگون استفاده کرد که منجر به نتایج جالب توجه و مفیدی در بحث بازاریابی و ارائه تبلیغات می شود.

6- نتیجه گیری و راهکارهای آینده

به دلیل افزایش رقابت بین سازمان‌ها و فعالیت های تبلیغاتی، بخش بندی بازار اهمیت ویژه ای برای تمرکز روی گروه خاصی از مشتریان و لذا کاهش هزینه ها پیدا کرده است. روش پیشنهادی ترکیب روش K-medoids با روش درخت تصمیم در بازاریابی شرکت بیمه برای پیش بینی خوشه مشتریان استفاده می‌شود. نتایج آزمایش انجام شده بر روی داده های یک شرکت بیمه نشان می دهدکه تکنیک پیشنهادی راهکاری جدید با دقت نسبتاً خوب برای پیش بینی عملکرد رفتاری مشتریان برای ارائه تصمیمات مختلف است. همچنین با این تکنیک دقت در پیش بینی 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 درخت تصمیم برای پیش بینی یک عمل به تنهایی دقت بالایی ندارد و تکنیک‌های اختصاصی تری نظیر ترکیب K-medoids با الگوریتم شبکه های عصبی و یا تکنیک‌های باهم آیی بسیار مناسب تر از الگوریتم های عام منظوره هستند. در آینده می‌توان کارایی این روش پیشنهادی را در کاربردهای دیگر بازاریابی مورد ازریابی قرار داد. به علاوه، ترکیب روش پیشنهادی با سایر روش‌های موجود، با هدف افزایش کارایی و یا دقت آن نیز می تواند موضوع پژوهشی مناسبی برای کار در آینده باشد. می توان شکل‌های مختلف ترکیب تکنیک را نیز به مجموعه قواعد مورد استفاده اضافه کرد و شکل پیچیده تری را نیز برای فرضیه ها تعریف کرد. بدین صورت فضای فرضیات بزرگتری را مورد جستجو قرار داده و لذا احتمال یافتن فرضیه های بهتر با دقت پیش بینی بالاتر افزایش می‌یابد.

7- مراجع

Chaffey, D., Smith, P. R., & Smith, P. R. (2012). *"eMarketing eXcellence: Planning and optimizing your digital marketing"*. Routledge.

www.SID.ir

Archive of SID

Chaffey, D., Ellis-Chadwick, F., Mayer, R., & Johnston, K. (2009). *"Internet marketing: strategy, implementation and practice"*. Pearson Education.

Goyat, S. (2011). "The basis of market segmentation: a critical review of literature". *European Journal of Business and Management, 3(9),* 45-54.

Hong, T., & Kim, E. (2012). "Segmenting customers in online stores based on factors that affect the customer's intention to purchase". *Expert Systems with Applications, 39(2),* 2127-2131.

Chen, Y. K., Wang, C. Y., & Feng, Y. Y. (2010)." Application of a 3NN+ 1 based CBR system to segmentation of the notebook computers market". *Expert Systems with Applications, 37(1),* 276-281.

Tsai, C. Y., & Chiu, C. C. (2004). "A purchase-based market segmentation methodology". *Expert Systems with Applications, 27(2),* 265-276.

Moro, S., Laureano, R., & Cortez, P. (2011). "Using data mining for bank direct marketing: An application of the crisp-dm methodology". In *Proceedings of European Simulation and Modelling Conference-ESM'2011* (pp. 117-121). Eurosis.

He, L., Li, L., Li, X., Wang, D.,.(2009) "Comparison and Analysis of algorithms for association rules". *First International Workshop on Database Technology and Applications*, pp. 196 – 198.

Davies, D. L., & Bouldin, D. W. (1979). "A cluster separation measure". *IEEE transactions on pattern analysis and machine intelligence,* (2), 224-227.

Powers, D. M. (2011). "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation".

www.SID.ir

Archive of SID

www.SID.ir