



## استفاده از داده کاوی در سیستم مدیریت ارتباط با مشتری شرکت های بیمه (مطالعه موردی در زمینه شناسایی مشتری)

رسول نادری<sup>۱</sup>، احسان شکاری<sup>۲</sup>

۱- کارشناس ارشد مهندسی صنایع - سیستم های اقتصادی و اجتماعی، دانشگاه علوم اقتصادی

۲- کارشناس ارشد علوم تصمیم و مهندسی دانش، دانشگاه علوم اقتصادی

Email: Rasul.naderi@gmail.com

### خلاصه

بازار بیمه هر روز پیش از پیش رفاقتی می شود بیمه نامه های متعددی نیز مرتبط با جنبه های مختلف زندگی به وجود آمده اند. شرکت های بیمه به منظور موفقیت در ارائه خدمات بهتر و توسعه کسب کار خود باید توصیف بهتری از مشتریان، و نیازهای آن ها داشته باشند. شرکت های بیمه حجم عظیمی از داده ها در مورد بیمه شوندگان را ثبت می کنند ولی از آنها نمی توانند بدین منظور استفاده مناسبی بکنند. داده کاوی علمی جدیدی است که توانایی کار با حجم عظیم داده ها و کشف دانش پنهان موجود در آن ها را دارد و استفاده از آن در این مورد می تواند مفید باشد. در این مقاله طرحی کاربردی برای استفاده از داده کاوی در جنبه های مختلف مدیریت ارتباط با مشتری ارائه می شود تا شرکت های بیمه بتوانند خدمات بهتری به مشتریان خود ارائه دهند که نتیجه آن رضایت بیشتر مشتری و موفقیت بیشتر شرکت ها بیمه است.

کلمات کلیدی: داده کاوی<sup>۱</sup>، مدیریت ارتباط با مشتری<sup>۲</sup>، شناسایی مشتری، بیمه.

### ۱. مقدمه

بازار امروز در تمام تجارت ها بسیار رقابتی شده است. هر تجارتی به منظور بقا در این عرصه باید بتواند با ارائه خدمات و کالای مطلوب رضایت مشتریان خود را جلب و سهم مناسبی از بازار را به دست آورد. در صنعت بیمه نیز از زمان پیدایش تا کنون تغییرات زیادی رخ داده است و بازار شرکت های بیمه نیز رقابتی شده است. مدیریت ارتباط با مشتری استراتژی پیش رو در بازار رقابتی کنونی است که می تواند مواردی همچون افزایش حفظ و وفاداری مشتریان، سودآوری بیشتر مشتریان و ارائه خدمات بهتر و مناسب تر به مشتریان را به همراه داشته باشد. تعریف دقیقی برای مدیریت ارتباط با مشتری وجود ندارد و نویسندگان مختلف تعاریف مختلفی را در مقالات خود ارائه می دهند [1] اما کلیت آنها تقریباً یکسان است. در اینجا تعریفی که کیم و همکارانش در مقاله خود ارائه داده اند [2] آورده شده است که بدین صورت است. مدیریت ارتباط با مشتری را می توان به تمام فعالیت های مدیریتی به منظور مدیریت تراکنش

<sup>1</sup>. Data Mining

<sup>2</sup>. Customer Relationship management (CRM)



های تجاری با مشتریان، با استفاده از فرآیندهای تجاری و تکنولوژی‌هایی که سعی در فهم مشتریان شرکت دارند، اطلاق کرد. در تعریف بالا استفاده از تکنولوژی یا همان روشی فنی و علمی برای فهم و شناخت مشتریان شرکت به چشم می‌خورد. داده کاوی می‌تواند روش و ابزاری مناسب و کارا به منظور به دست آوردن این شناخت از مشتریان باشد. داده کاوی علم تقریباً جدیدی با کاربرد های بسیار متنوع است که توانایی کار با حجم عظیمی از داده‌ها و استخراج دانش پنهان موجود در آنها را دارد. استفاده از ابزار های داده کاوی در مدیریت ارتباط با مشتری نیز پدیده‌ای در حال ظهور است [1]. اساس مدیریت ارتباط با مشتری بررسی و تحلیل مشتریان و شناخت رفتارها و خصوصیات آنها است، که بدین منظور می‌توان از تکنیک‌های مناسب داده کاوی استفاده کرد.

شرکت‌های بیمه حجم عظیمی از داده‌ها را در مورد مشتریان خود ثبت می‌کنند که تکنیک‌های داده کاوی را می‌توان بر روی این داده‌ها به منظور به دست آوردن شناخت از مشتریان و استفاده از این شناخت در استراتژی مدیریت ارتباط با مشتری به کار برد. در این مقاله به کارهایی که توسط ابزارهای داده کاوی در بیمه می‌توان انجام داد پرداخته می‌شود و آنها را در قالب مدیریت ارتباط با مشتری برای یک شرکت بیمه دسته‌بندی می‌کنیم. در قسمت ۲ این مقاله به معرفی داده کاوی می‌پردازیم. سپس در قسمت ۳ مدیریت ارتباط با مشتری در چهار بخش شناسایی، جذب، حفظ و توسعه ارتباط با مشتری معرفی می‌شود. در قسمت ۵ کارهایی که با داده کاوی در بیمه می‌توان انجام داد و اینکه هر یک از این کارها در کدام یک از بخش‌های مدیریت ارتباط با مشتری قرار می‌گیرد آورده شده است. در بخش ۶ به عنوان مطالعه موردی به شناسایی مشتریان علاقه‌مند به یک نوع بیمه نامه خاص پرداخته می‌شود و در نهایت نتیجه‌گیری آورده شده است.

## ۲. داده کاوی

داده کاوی علم نوینی است که به بررسی و تحلیل حجم عظیمی از داده‌ها می‌پردازد. دانش پنهان در این داده‌ها را کشف می‌کند. این دانش به دست آمده می‌تواند در فرایند تصمیم‌گیری به ما کمک کند. با پیشرفت تکنولوژی و امکان ذخیره‌سازی حجم زیادی از داده‌ها در پایگاه‌های داده و انبارهای داده، بالا رفتن توان محاسباتی، در دسترس قرار گرفتن نرم‌افزارهای داده کاوی و همچنین رقابتی شدن بازار گرایش استفاده از داده کاوی در موارد مختلف افزایش یافته و این علم بیش از پیش مورد توجه قرار گرفته است.

داده کاوی شامل چندین مدل اصلی است که می‌توان بسیاری از مسائل را در قالب یکی از این مدل‌ها گنجانند. در حل هر مسئله با داده کاوی ابتدا باید تعیین کنیم که این مسئله جز کدامیک از این مدل‌ها قرار می‌گیرد و با توجه به آن تکنیک داده کاوی مناسب را انتخاب کنیم. این وظایف عبارتند از:

۱. **قوانین وابستگی**<sup>۳</sup>: در این مورد به دنبال روابط بین اشیایی هستیم که با هم در یک رکورد وجود دارند. به عبارت دیگر به دنبال احتمال وقوع یا عدم وقوع همزمان ویژگی‌ها می‌گردیم. تحلیل سبد خرید و فروش متقابل مثال‌های رایجی در این مورد هستند.
۲. **دسته بندی**<sup>۴</sup>: یکی نوع از رایج‌ترین مدل‌هایی که در داده کاوی ساخته می‌شوند مدل‌هایی برای دسته‌بندی هستند در این وظیفه داده کاوی مدلی ساخته می‌شود که یک شی جدید را با توجه به خصوصیاتش به یکی از چند کلاس از پیش تعریف شده نسبت می‌دهد. به عنوان مثال می‌توان مدلی برای پیش‌بینی رفتار آینده مشتریان ساخت که بر مبنای معیارهای خاصی آنها را به چند گروه از پیش تعریف شده نسبت دهد.
۳. **خوشه بندی**<sup>۵</sup>: در این وظیفه جمعیت‌ها همگن به چند گروه همگن تقسیم می‌شود. فرق آن با دسته بندی در این است که گروه‌ها از قبل معلوم نیستند.
۴. **پیشگویی**<sup>۶</sup>: پیشگویی مقدار آینده را بر مبنای الگوهای یک رکورد تعیین می‌کند. در این مدل با متغیرهای پیوسته خروجی سر و کار داریم. پیش‌بینی تقاضا مثال رایج در این نوع مدل سازی است.

3. Association  
4. Classification  
5. Clustering  
6. Forecasting



۵. **رگرسیون<sup>۷</sup>**: رگرسیون نوعی تکنیک تخمین آماری است که از آن برای نسبت دادن هر شی داده ای به یک عدد حقیقی که همان مقدار تخمین زده شده است استفاده می شود.
۶. **کشف ترتیب<sup>۸</sup>**: که در این مدل به شناسایی وابستگی ها و الگوها در طول زمان هستیم. هدف آن مدل کردن حالات یک فرآیند است به نحوی که ترتیب اعمال یا انحرافات و گرایش ها را در طول زمان مشخص کند.
۷. **بصری سازی<sup>۹</sup>**: بصری سازی به ارائه ای تصویری از داده ها به منظور نمایش الگوهای پیچیده به کاربران گفته می شود. این روش در کنار سایر تکنیک های داده کاوی به منظور درک بهتری از الگوهای پیدا شده توسط تکنیک های داده کاوی استفاده می شود.
- در استفاده از داده کاوی در حل یک مسئله تجاری ابتدا باید مسئله ای را که با آن رو به رو هستیم به خوبی درک کنیم و دقیقاً بفهمیم به دنبال چه چیزی هستیم. سپس باید تعیین کنیم این مسئله جز کدامیک از وظایف داده کاوی قرار می گیرد و با توجه به آن تکنیک مناسب داده کاوی برای ساخت مدل را انتخاب کنیم. سپس نتایج به دست آمده از داده کاوی بررسی می شوند این نتایج می توانند در تصمیم گیری استفاده شوند و باعث تصمیم گیری بهتر و بر مبنای دانش بیشتر شوند. در نهایت باید نتایج استفاده از داده کاوی را بسنجیم و ببینیم آیا استفاده از داده کاوی در مسئله مورد نظر واقعاً مفید بوده و در این صورت میزان تاثیر آن چقدر بوده است.
- نکته دیگر در استفاده از داده کاوی وجود داده های مورد نیاز برای حل یک مسئله و صحت داده های موجود است. در صورتی که داده ها ناقص یا صحیح نباشند نمی توان انتظار جوابی قابل اعتماد را داشت.

### ۳. مدیریت ارتباط با مشتری

dataacademy.ir

در بازار رقابتی امروز محصول مداری جای خود را به مشتری مداری داده است. در این نوع بازار مدیریت ارتباط با مشتری استراتژی پیش رو برای به دست آوردن مزیت رقابتی است. مدیریت ارتباط با مشتری شامل مجموعه ای از فرآیندها برای برقراری ارتباط طولانی مدت و سودآور با مشتریان است. اهداف کلی آن پیدا کردن، جذب کردن و کسب مشتریان جدید، گسترش ارتباط با مشتریان و حفظ مشتریان کنونی است. تعاریف مختلفی برای مدیریت ارتباط با مشتری ارائه شده است. در اینجا تعریفی که Kim و همکارانش در مقاله خود ارائه داده اند [2] آورده شده است که بدین صورت است: مدیریت ارتباط با مشتری را می توان به تمام فعالیت های مدیریتی به منظور مدیریت تراکنش های تجاری با مشتریان، با استفاده از فرآیندهای تجاری و تکنولوژی هایی که سعی در فهم مشتریان شرکت دارند، اطلاق کرد.

واضح است در این استراتژی به تحلیل خصوصیات و رفتار های مشتریان نیازمندیم. تکنیک های داده کاوی روش های متداول تجزیه و تحلیل اطلاعات مشتری در چارچوب مدیریت ارتباط با مشتری هستند. که سازمان ها می توانند از این روش ها برای به دست آوردن دانش و استفاده از این دانش در تصمیم گیری استفاده کنند.

مدیریت ارتباط با مشتری از ۴ بخش اصلی تشکیل می شود که این ۴ بخش عبارتند از [1]:

۱. شناسایی مشتری<sup>۱۰</sup>
۲. جذب مشتری<sup>۱۱</sup>
۳. حفظ مشتری<sup>۱۲</sup>
۴. گسترش ارتباط با مشتری<sup>۱۳</sup>

7. Regression

8. Sequence Discovery

9. Visualization

10. Customer Identification

11. Customer Attraction

12. Customer Retention



**شناسایی مشتری:** مدیریت ارتباط با مشتری با شناسایی مشتری آغاز می شود. این مرحله شامل هدف قرار دادن جمعیتی است که احتمال اینکه مشتری شرکت شوند یا برای شرکت سودآور باشند، زیاد است. علاوه بر این، این مرحله شامل تحلیل مشتریانی است که در رقابت با دیگر شرکت ها جذب یکی از شرکت دیگری شده اند و اینکه برای به دست آوردن دوباره آنها چه کار می توان کرد. عناصر شناسایی مشتری عبارتند از: تحلیل مشتریان هدف و بخش بندی مشتریان. تحلیل مشتریان هدف شامل جستجو برای پیدا کردن بخش سودآور مشتریان با استفاده از تحلیل خصوصیات مشتریان است و بخش بندی مشتریان شامل تقسیم کردن کل مشتریان به گروه های کوچکتر به نحوی که مشتریان موجود در هر گروه خصوصیات تقریباً مشابهی داشته باشند.

**جذب مشتری:** بعد از شناسایی مشتری نوبت به جذب مشتریان می رسد. بعد از شناسایی گروه های مختلف مشتریان بالقوه شرکت می تواند تلاش و منابع خود را صرف جذب گروه مشتریان هدف کند. یکی از عناصر جذب مشتری بازاریابی مستقیم<sup>۱۴</sup> است. بازاریابی مستقیم امکان سفارشی کردن تبلیغات را برای تک تک مشتریان فراهم می کند. ارسال پست الکترونیک و توزیع کوپن نمونه های متداول بازاریابی مستقیم هستند.

**حفظ مشتری:** نگرانی اصلی در مدیریت ارتباط با مشتری، حفظ مشتری است. رضایت مشتری که به مقایسه انتظارات او و میزان برآورده شدن آن ها مربوط می شود عامل اصلی در حفظ مشتری به حساب می آید. عناصر حفظ مشتری شامل بازاریابی فرد به فرد<sup>۱۵</sup>، برنامه های وفاداری<sup>۱۶</sup> و مدیریت شکایات است. بازاریابی فرد به فرد فعالیت های بازاریابی شخصی شده برای یک فرد خواست گفته می شود که با تحلیل، شناسایی و پیش بینی رفتار مشتریان امکان پذیر است. برنامه های وفاداری فعالیت های جنبی هستند که به منظور ایجاد یک رابطه طولانی مدت با مشتری انجام می شوند. تخمین تعداد مشتریانی که تمایل به ترک شرکت را دارند، رتبه دهی اعتباری مشتریان و توزیع فرم های نظرسنجی میزان رضایت از خدمات جز برنامه های وفاداری به حساب می آید.

**گسترش ارتباط با مشتری:** این مرحله شامل افزایش تعداد و ارزش تراکنش ها و سودآوری مشتری است. عناصر گسترش ارتباط با مشتری شامل سنجش ارزش مشتری در طول مدتی که در شرکت حضور دارد، فروش متقابل و تحلیل سبد خرید است. ارزش مشتری در طول مدت حضورش در شرکت به عنوان پیش بینی از کل در آمد خالصی که شرکت می تواند از آن مشتری کسب کند تعریف می شود. فروش متقابل به پیشنهاد خرید خدمات دیگر شرکت به مشتری گفته می شود که این خدمات معمولاً مکمل یا مرتبط با خدمات یا کالایی است که مشتری از آن استفاده می کند و تحلیل سبد خرید نیز با هدف افزایش تراکنش با مشتری به بررسی رفتار خرید مشتری برای پی بردن به عادات خرید مشتری انجام می گیرد.

#### ۴. داده کاوی در بیمه

همانطور که گفته شد شرکت های بیمه داده های بسیار زیادی در مورد مشتریان و سوابق بیمه آن ها را ثبت کرده و در اختیار دارند و این پایگاه داده عظیم فرصت بسیار خوبی برای استفاده از ابزار های داده کاوی را فراهم کرده است. داده کاوی در صنعت بیمه می تواند در موارد زیر استفاده شود که آن را به دو گروه داده کاوی برای مدیریت ارتباط با مشتری در بیمه و سایر کاربردها تقسیم شده است.

#### ۴-۱ داده کاوی برای مدیریت ارتباط با مشتری در بیمه

**شناسایی و جذب مشتری:** همانطور که گفته شد شناسایی و جذب مشتری دو مرحله اول در مدیریت ارتباط با مشتری است. بدین منظور شرکت های بیمه می توانند با استفاده از تکنیک های خوشه بندی یا دسته بندی مشتریان خود را گروه بندی کنند. این گروه بندی می تواند بر اساس میزان ریسک یا ارزش مشتری در طول مدت حضورش در شرکت یا معیار دیگری باشد. سپس شرکت می تواند با توجه به میزان سودآوری هر کدام از این گروه ها اولویت بندی انجام دهد. به عنوان مثال شرکت می تواند تمرکز بازاریابی خود را بر روی گروهی با ریسک کمتر جلب کند. همچنین شرکت می تواند به این موضوع پی

<sup>13</sup>. Customer Development

<sup>14</sup>. Direct Marketing

<sup>15</sup>. One-to-One Marketing

<sup>16</sup>. Loyalty Programs



ببرد که چه بیمه نامه های برای یک شخص خاص مناسب تر است و همان بیمه نامه ها را به آن شخص پیشنهاد دهد تا درصد پاسخ بیشتری به دست آید. کار دیگری که در این مورد می توان انجام داد انتخاب کانال ارتباطی مناسب و پیام تبلیغاتی مناسب برای هر شخص با توجه به خصوصیاتش است. پس با استفاده از ابزار های داده کاوی می توان از محصول گرایی به سمت مشتری گرایی حرکت کرد و درصد پاسخ بیشتری به دست آورد.

**فروش متقابل:** فروش متقابل به پیشنهاد خرید خدمات دیگر شرکت به مشتری گفته می شود که این خدمات معمولاً مکمل یا مرتبط با خدمات یا کالایی است که مشتری از آن استفاده می کند. احتمال ترک شرکت توسط شخص بیمه شونده ای با دو یا تعدادی بیشتر بیمه نامه کمتر از شخصی با یک بیمه نامه است. می توان دو یا تعداد بیشتری بیمه نامه را به صورت یک بسته یا تخفیف برای فروش ارائه کرد. اینکه چه بیمه نامه هایی با هم انتخاب شوند و به چه گروهی از مشتریان ارائه شوند را می توان با استفاده از تکنیک های داده کاوی با توجه به خصوصیات مشتریان و سوابق آن ها پیدا کرد. فروش متقابل را در مدیریت ارتباط با مشتری در قسمت های حفظ و گسترش ارتباط با مشتری تاثیر گذار است.

**شناسایی مشتریانی که می خواهند شرکت را ترک کنند:** با استفاده از تکنیک های داده کاوی می توان مدل های پیش بینی کننده ای طراحی کرد و افرادی را که احتمال دارد بیمه نامه های خود را تمدید نکنند شناسایی کرد. سپس با خصوصیات این افراد را تحلیل کرد و علل این امر را کشف کرد و اقدامات باز دارنده ای را انجام داد. این کاربرد داده کاوی در قسمت حفظ مشتری قرار می گیرد.

**تعیین ارزش مشتری در مدت حضورش در شرکت:** تاکید مدیریت ارتباط با مشتری بر برقراری روابط بلند مدت با مشتریان پر منفعت برای شرکت است. این ارزش مقدار عایدی شرکت از یک مشتری بعد از کم کردن هزینه های جذب، فروش، ارائه خدمات با احتساب ارزش زمانی پول است. بدین منظور می توان مدل های داده کاوی مختلفی ساخت. سپس می توان با توجه به این ارزش مشتریان را گروه بندی کرد و مشتریان هدف و میزان اهمیت هر کدام برای شرکت را تعیین کرد.

## dataacademy.ir

### ۴-۲ سایر کاربرد های داده کاوی در بیمه

**شناسایی عوامل ریسک و تعیین نرخ:** قیمت مهمترین عامل در جذب مشتری است. برای تعیین قیمت، شرکت های بیمه باید عوامل یا فاکتور های ریسک در احتمال وقوع و میزان ادعاها را پیش بینی کنند. به عنوان مثال در بیمه خودرو برای شخصی که در ناحیه پر تردد شهر رفت آمد می کند باید احتمال تصادف بیشتری در نظر گرفته شود. بعضی از فاکتورهای ریسک بسیار واضح هستند ولی ممکن است روابط غیر آشکاری ما بین متغیرهای تاثیر گذار وجود داشته باشد که نیاز به بررسی و تحلیل پیچیده تری داشته باشد. ابزارهای داده کاوی به صورت دقیق تری می توانند فاکتورهای ریسک و میزان تاثیر هر یک را مشخص کنند که این امر در نهایت می تواند به کاهش قیمت، جذب مشتریان بیشتر و سود بیشتر منتهی شود.

**کشف تقلب:** شناسایی ادعاهای تقلبی در موفقیت شرکت های بیمه مهم است. ابزار های داده کاوی در تعیین موفق تر این ادعاها می توانند استفاده شوند. در تعیین تقلب تکنیک های داده کاوی به دنبال رفتار های غیر طبیعی و الگوهای غیر عادی هستند. به عنوان مثال اگر تعداد بیمارانی که توسط یک پزشک خاص معاینه شده اند و ادعای خسارت دارند از متوسط تعداد بیمارانی که توسط سایر پزشکان معاینه شده اند بیشتر باشد می توان این امر را به عنوان یک مورد محتمل تقلب بررسی کرد.

**بیمه اتکایی<sup>۱۷</sup>:** بیمه اتکایی یعنی اینکه قسمتی از ریسک بیمه را شرکت دیگری در عوض مقداری از حق بیمه بر عهده بگیرد. با انجام بیمه اتکایی ریسک کمتری شرکت را تهدید می کند اما سود کمتری نیز عاید شرکت می شود. این که در مورد کدامیک از بیمه نامه ها و به چه مقداری این کار را انجام دهیم می تواند توسط داده کاوی انجام شود.

**تخمین میزان خسارت های پرداخت نشده:** شرکت های بیمه با مبالغ دریافتی از مشتریان خود سرمایه گذاری انجام می دهند و سود خود را از این سرمایه گذاری ها به دست می آورند. اما سالیانه مقداری از مبلغ دریافتی خود را به منظور جبران خسارت به مشتریان خود بر می گردانند. تخمین این

<sup>17</sup>. Reinsurance



میزان سرمایه که به عنوان جبران خسارت پرداخت می شود اهمیت دارد زیرا شرکت های بیمه نمی توانند این مبلغ را در سرمایه گذاری های بلند مدت استفاده کنند. تکنیک های داده کاوی به منظور تخمین این مبلغ می توانند استفاده شوند. شرکت های بیمه می توانند از داده کاوی حتی برای شناسایی فرصت های سرمایه گذاری با سود بیشتر نیز استفاده کنند.

**مدیریت منابع انسانی:** شرکت های بیمه نه تنها از داده کاوی در تحلیل رفتار مشتریان خود می توانند استفاده کنند بلکه با داده کاوی می توانند خصوصیات کارمندان خود را نیز بررسی کنند از جمله میزان فروش هر کدام و اینکه کدام کارمند بیشتر چه بیمه نامه ای را فروخته است. سپس می توان کارمند های پر فروش را در آموزش سایر کارمندان به کار گرفت یا یک کارمند را به طور اختصاصی مسئول فروش یک نوع بیمه نامه کرد. اختصاص تعداد مناسب و کارکنان مناسب به قسمت های مختلف نیز از جمله کارهایی است که می توان انجام داد.

## ۵. مطالعه موردی در شناسایی مشتری:

با تحلیل بازار هدف و پیدا کردن افرادی که با احتمال بیشتری به مشتریان شرکت تبدیل می شوند. می توان تبلیغات هدفمند تری انجام داد و در ضمن کاهش هزینه های تبلیغات، مشتریان بیشتری را به دست آورد. داده کاوی در این زمینه می تواند بدین منظور استفاده شود. ابزار های داده کاوی در بخش بندی مشتریان، تحلیل بازار هدف، تخمین علاقه مندی افراد مختلف از استفاده از خدمات و خرید کالاها مختلف مورد استفاده قرار گیرند و بعد با بررسی و استفاده از نتایج به دست آمده توسط داده کاوی می توان تصمیم های بهتری اتخاذ کرد.

در اینجا می خواهیم به تعدادی از مشتریان یک شرکت بیمه پیشنهادی در مورد یکی بیمه نامه خاص دیگر<sup>۱۸</sup>، ارائه کنیم. تعداد افرادی که می خواهیم برای آن ها تبلیغات را ارسال کنیم محدود است و تنها ۸۰۰ نفر از مشتریان برای ارسال تبلیغ انتخاب می شوند. با استفاده از داده کاوی می خواهیم میزان علاقه مشتریان این شرکت بیمه را به این بیمه نامه تخمین بزنیم و این افراد را طوری انتخاب بکنیم که با ارسال تبلیغ برای این تعداد افراد پاسخ بیشتری دریافت کنیم. سپس می خواهیم خصیصه های مهم در علاقه مندی مشتریان به این بیمه نامه را مشخص کنیم تا توصیف بهتری از مشتریان علاقه مند به این بیمه نامه داشته باشیم. چون هر رکورد در این مجموعه داده حاوی ۸۵ خصیصه است که نسبتاً زیاد است. برای به دست آوردن خصیصه های مهم تر در تعیین علاقه مندی و به دست آوردن توصیف بهتری از مشتریانی که به این بیمه خاص علاقه مند هستند سعی در کم کردن تعداد خصیصه ها داریم. علاوه بر آن این کار هم مدل سازی را سریع تر کرده و ممکن است نتایج بهتری حاصل شود.

داده های استفاده شده در این کار مربوط به یک شرکت بیمه است که از سایت UCI دانلود شده است [8] و مربوط به مسابقه ای است که با همین هدف در سال ۲۰۰۰ برگزار شده است [9]. این مجموعه داده از دو بخش آموزش و تست تشکیل شده است. هر رکورد دارای ۸۵ خصیصه است و به یکی از دو کلاس خرید بیمه نامه و عدم خرید بیمه نامه تعلق دارد. در جدول ۱ تعداد رکورد های موجود در دو مجموعه آموزشی و تست دیده می شود. در جدول ۲ نیز خصیصه های هر رکورد لیست شده است.

نرم افزار های داده کاوی استفاده شده در این مقاله Weka و Clementine هستند که جز نرم افزارهای رایج داده کاوی هستند. در ساخت مدل-های داده کاوی برای تخمین علاقه مندی و تهیه لیست ۸۰۰ نفری سه مدل داده کاوی با ماشین پشتیبان بردار (SVM)، رگرسیون خطی (Regression) و رگرسیون غیر خطی (Generalized Regression) در Clementine ساخته شده است. در رتبه بندی خصیصه های مشتریان نیز ابتدا از Clementine استفاده شد که نتایج خوبی حاصل نشد سپس به سراغ Weka رفتیم و از الگوریتم های زیادی که برای انتخاب و رتبه بندی خصیصه ها در این الگوریتم پیاده سازی شده است استفاده کردیم که نتایج آن ها در ادامه آمده است.

<sup>18</sup>. Caravan



جدول ۱: تعداد رکورد های موجود در مجموعه داده

کلاس ۱: خرید	کلاس ۰: عدم خرید	
۳۴۸	۵۸۲۲	مجموعه داده آموزشی
۲۳۸	۴۰۰۰	مجموعه داده تست

جدول ۲: خصیصه های مجموعه داده

No	Variable name	Variable Description	No	Variable name	Variable Description
1	MOSTYPE	Customer Subtype	44	PWAPART	Contribution private third party insurance
2	MAANTHUI	Number of houses	45	PWABEDR	Contribution third party insurance (firms)
3	MGEMOMV	Avg size household	46	PWAAND	Contribution third party insurance (agriculture)
4	MGEMLEEF	Avg age	47	PPERSAUT	Contribution car policies
5	MOSHOOFD	Customer main type	48	PBESAUT	Contribution delivery van policies
6	MGODRK	Roman catholic	49	PMOTSCO	Contribution motorcycle/scooter policies
7	MGODPR	Protestant	50	PVRAAUT	Contribution lorry policies
8	MGODOV	Other religion	51	PAANHANG	Contribution trailer policies
9	MGODGE	No religion	52	PTRACTOR	Contribution tractor policies
10	MRELGE	Married	53	PWERKT	Contribution agricultural machines policies
11	MRELSA	Living together	54	PBROM	Contribution moped policies
12	MRELOV	Other relation	55	PLEVEN	Contribution life insurances
13	MFALLEEN	Singles	56	PPERSONG	Contribution private accident insurance policies
14	MFGEKIND	Household without children	57	PGEZONG	Contribution family accidents insurance policies
15	MFWEKIND	Household with children	58	PWAOREG	Contribution disability insurance policies
16	MOPLHOOG	High level education	59	PBRAND	Contribution fire policies
17	MOPLMIDD	Medium level education	60	PZEILPL	Contribution surfboard policies
18	MOPLLAAG	Lower level education	61	PPLEZIER	Contribution boat policies
19	MBERHOOG	High status	62	PFIETS	Contribution bicycle policies
20	MBERZELF	Entrepreneur	63	PINBOED	Contribution property insurance policies
21	MBERBOER	Farmer	64	PBYSTAND	Contribution social security insurance policies
22	MBERMIDD	Middle management	65	AWAPART	Number of private third party insurance
23	MBERARBG	Skilled labourers	66	AWABEDR	Number of third party insurance (firms)
24	MBERARBO	Unskilled labourers	67	AWALAND	Number of third party insurance (agriculture)
25	MSKA	Social class A	68	APERSAUT	Number of car policies
26	MSKB1	Social class B1	69	ABESAUT	Number of delivery van policies
27	MSKB2	Social class B2	70	AMOTSCO	Number of motorcycle/scooter policies
28	MSKC	Social class C	71	AVRAAUT	Number of lorry policies
29	MSKD	Social class D	72	AAANHANG	Number of trailer policies
30	MHHUUR	Rented house	73	ATRACTOR	Number of tractor policies
31	MHKOOP	Home owners	74	AWERKT	Number of agricultural machines policies



32	MAUT1	1 car	75	ABROM	Number of moped policies
33	MAUT2	2 cars	76	ALEVEN	Number of life insurances
34	MAUT0	No car	77	APERSONG	Number of private accident insurance policies
35	MZFONDS	National Health Service	78	AGEZONG	Number of family accidents insurance policies
36	MZPART	Private health insurance	79	AWAOREG	Number of disability insurance policies
37	MINKM30	Income < 30.000	80	ABRAND	Number of fire policies
38	MINK3045	Income 30-45.000	81	AZEILPL	Number of surfboard policies
39	MINK4575	Income 45-75.000	82	APLEZIER	Number of boat policies
40	MINK7512	Income 75-122.000	83	AFIETS	Number of bicycle policies
41	MINK123M	Income >123.000	84	AINBOED	Number of property insurance policies
42	MINKGEM	Average income	85	ABYSTAND	Number of social security insurance policies
43	MKOOKLA	Purchasing power class	86	CARAVAN	Number of mobile home policie

سه مدل داده کاوی ذکر شده در بالا با استفاده از تمام خصیصه ها ساخته شد و بر مبنای هر یک از این مدل ها ۸۰۰ نفر را بر اساس مرتب سازی امتیاز های حاصل از این مدل ها انتخاب کردیم. نتایج حاصل از این کار در جدول ۳ آورده شده است.

جدول ۳: نتایج

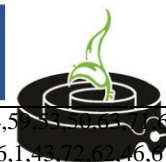
روش	تعداد پاسخ مثبت دریافتی
SVM	۹۳
Generalized Regression	۱۰۴
Regression	۱۱۸

برای به دست آوردن خصیصه های مهم تر در تعیین علاقه مندی و به دست آوردن توصیف بهتری از مشتریانی که به این بیمه خاص علاقه مند هستند سعی در کم کردن تعداد خصیصه ها داریم. علاوه بر آن این کار هم مدل سازی را سریع تر کرده و ممکن است نتایج بهتری حاصل شود. در ابتدا از نود feature selection در نرم افزار Clementine استفاده شد که نتایج خوبی حاصل نشد. بعد از آن از الگوریتم هایی که در Weka بدین منظور پیاده سازی شده است استفاده شد که نتایج مربوط به اهمیت خصیصه ها در جدول ۴ آورده شده است.

جدول ۴: نتایج مربوط به اهمیت خصیصه ها

No	Attribute Evaluator	Search method	ترتیب اهمیت خصیصه ها
1	ChiSquaredAttributeEval	Ranker	47,59,68,1,5,43,61,42,82,44,18,37,25,30,31,65,16,34,19,39,32,28,80,23,10,64,85,12,40,35,29,22,36,49,24,7,26,76,21,9,17,13,57,54,15,75,8,83,55,3,6,58,79,11,60,14,78,38,63,62,20,27,51,81,52,45,41,2,4,46,67,84,72,73,33,74,53,70,56,48,69,50,71,77,66
2	FilteredAttributeEval InfoGainAttributeEval	Ranker	47,59,68,1,5,42,43,37,18,44,34,65,30,31,25,16,32,39,61,10,80,12,19,28,82,23,29,40,35,36,24,21,22,7,9,13,54,64,26,85,17,75,15,76,8,49,55,11,6,3,57,2,14,83,38,52,41,58,79,45,27,63,78,4,20,62,46,51,67,60,53,74,73,72,81,84,56,48,33,50,71,69,70,77,66





3	GainRatioAttributeEval	Ranker	82,61,60,81,85,64,47,79,58,57,68,78,74,56,55,56,1,4,3,2,9,1,6,7,48,3 4,75,49,45,51,76,83,42,80,5,84,52,55,56,1,4,3,2,9,1,6,7,48,3 7,34,25,21,16,18,32,29,40,30,31,69,39,73,10,2,12,19,23,28,77, 24,35,36,22,9,13,7,26,41,6,17,8,11,3,15,20,70,4,14,38,27,66,33
4	ReliefFAttributeEval	Ranker	9,38,30,31,7,15,18,23,39,32,28,17,22,10,14,24,37,1,34,12,27,3 6,35,33,13,26,42,8,19,5,43,16,25,29,11,40,4,6,3,21,20,41,2,68, 47,82,61,64,85,58,79,62,83,84,63,57,78,60,81,71,50,48,69,74,5 3,77,56,51,72,76,55,66,45,73,52,67,46,70,49,75,54,59,44,65,80
5	SymmetricalUncertAttributeEval	Ranker	47,61,68,82,59,85,64,44,65,42,5,54,1,80,75,43,57,49,37,76,34, 25,18,16,55,79,58,21,32,30,31,29,40,39,83,78,63,10,12,45,19,5 2,23,28,60,51,35,24,62,36,46,2,22,9,7,13,74,53,67,26,81,17,72, 6,8,11,3,84,41,73,56,48,15,50,71,20,69,4,14,38,27,70,77,33,66

سپس با استفاده از روش های انتخاب خصیصه در Weka زیر مجموعه هایی از خصیصه ها را انتخاب و ۳ مدل داده کاوی را با این مجموعه خصیصه ها آزمایش کردیم که نتایج حاصل از این کار نیز در جدول ۵ آورده شده است. سه ستون آخر این جدول حاوی نتایج مدل های داده کاوی بالا را که ستون S نتایج مدل SVM ستون G نتایج مدل Generalized Regression و ستون R نتایج مدل Regression است. ستون تعداد نشان دهنده تعداد خصیصه هایی است که با آن روش انتخاب شده است دو ستون ابتدایی نیز روش استفاده شده را مشخص می کنند.

جدول ۵: استفاده از چند الگوریتم انتخاب خصیصه در Weka و نتایج مدل ها با آن ها

No	Attribute Evaluator	Search method	تعداد	S	G	R
1	CfsSubsetEval	GeneticSearch	36	86	111	113
2	CfsSubsetEval	RankSearch	30	80	109	116
3	CfsSubsetEval	GreedyStepwise	8	42	111	116
4	ConsistencySubsetEval	GeneticSearch	45	91	106	112
5	ConsistencySubsetEval	BestFirst GreedyStepwise	23	83	109	102

با نگاه به جدول ۵ می بینیم که کاهش خصیصه های انتخاب شده باعث کاهش دقت مدل SVM شده است اما در مورد رگرسیون غیر خطی Generalized Regression باعث افزایش کارایی مدل گشته است. در مورد مدل رگرسیون خطی Regression به عدد به دست آمده هنگامیکه از تمام خصیصه ها استفاده کرده ایم نرسیده ایم ولی اختلاف با این عدد ناچیز است. در ردیف ۲ و ۳ این اختلاف به ۲ رسیده است و این در حالی است که از



۸۵ خصیصه به ۳۰ و به ۸ خصیصه کاهش بعد داشته ایم که این کاهش بعد چشمگیر است. با مقایسه خصیصه های ردیف ۳ از جدول ۷ که تنها با ۸ متغیر جواب های خیلی خوبی به دست آمده با ردیف های جدول ۶ می بینیم که خصیصه های موجود در این ردیف جز اولین خصیصه های پر اهمیت قرار دارند.

## ۶. نتیجه گیری

شرکت های بیمه داده های بسیار زیادی در مورد مشتریان و سوابق بیمه آن ها را ثبت کرده و در اختیار دارند و این پایگاه داده عظیم فرصت بسیار خوبی برای استفاده از ابزار های داده کاوی را فراهم کرده است. بازار شرکت های بیمه پیش از پیش در حال رقابتی شدن است و مدیریت ارتباط با مشتری استراتژی پیش رو در بازارهای رقابتی است. شرکت های بیمه از داده های خود در هر ۴ بخش مدیریت ارتباط با مشتری یعنی شناسایی، جذب، حفظ و گسترش ارتباط می توانند استفاده کنند. در اینجا به کارهایی که شرکت های بیمه با داده کاوی می توانند انجام دهند اشاره و با استفاده از داده کاوی مدلی برای شناسایی مشتریانی بالقوه یک بیمه نامه خاص ارائه شد تا تنها با انتخاب مشتریانی که به این بیمه نامه علاقه مند و این بیمه نامه برایشان مناسب است و تمرکز بر روی این گروه، با صرف هزینه مالی و وقت کمتر به تعداد بیشتری پاسخ مثبت برسیم. در آینده می توان بر روی هر یک از کاربردهای داده کاوی در بیمه با تمرکز بیشتری کار کرد و نرم افزار جامعی شامل مدیریت ارتباط با مشتری در بیمه و سایر کاربرد های داده کاوی در بیمه طراحی کرد.

dataacademy.ir

مراجع

1. E.W.T. Ngai, Li Xiu, D.C.K. Chau, Application of data mining techniques in customer relationship management: A literature review and classification, Expert Systems with Applications 36, 2009.
2. Kim, J., Suh, E., Hwang, H., A model for evaluating the effectiveness of CRM using the balanced scorecard, Journal of Interactive Marketing 17, 2003.
3. A SAS Institute White Paper Data Mining in the Insurance Industry SOLVING BUSINESS PROBLEMS USING SAS ENTERPRISE MINER SOFTWARE, 2000.
4. Su-Yeon Kim, Tae-Soo Jung, Eui-Ho Suh, Hyun-Seok Hwang, Customer segmentation and strategy development based on customer lifetime value A case study, Expert Systems with Applications 31, 2006
5. K. A. Smith, R. J. Willis, M. Brooks, An Analysis of Customer Retention and Insurance Claim Patterns Using Data Mining A Case Study, 2000
6. S. Sumathi, S.N. Sivanandam, Introduction to Data Mining and its Applications, Springer, 2006
۷. داده کاوی چاپ دوم، دکتر جمال شهرابی، جهاد دانشگاهی واحد صنعتی امیر کبیر، ۱۳۹۰
8. [http://archive.ics.uci.edu/ml/datasets/Insurance+Company+Benchmark+\(COIL+2000\)](http://archive.ics.uci.edu/ml/datasets/Insurance+Company+Benchmark+(COIL+2000))
9. <http://www.liacs.nl/~putten/library/cc2000/report2.html>
10. Charles Elkan, Magical Thinking in Data Mining: Lessons From CoIL Challenge 2000, 2001