

معرفی روش استخراج اطلاعات^۱ در صنعت بیمه

مترجم: نازنین حاتمی^۲

جایگاه استخراج اطلاعات در صنعت بیمه

استخراج داده ها شامل مراحل انتخاب، جست و جو و مدل سازی مقدار زیادی از داده ها در جهت کشف الگوها و طرح های ناشناخته تعریف شده است. در صنعت بیمه استخراج داده ها به شرکت های بیمه در جهت کسب سود بازرگانی بیشتر کمک می نماید. برای مثال، با بهره برداری از تکنیک استخراج داده ها، شرکت های بیمه می توانند با بهره برداری کامل از الگوهای رفتاری مشتریان خود و با شناخت بیشتری که از انگیزه های ایشان به دست می آورند به کاهش میزان تقلب، افزایش امکان پیش بینی منابع تقاضا، افزایش فراگیری و نیز جلوگیری از کاهش مشتریان کمک نمایند.

مقاله حاضر به این موضوع می پردازد که شرکت های بیمه چگونه می توانند با بهره گیری از تکنیک استخراج داده ها موجبات کاهش هزینه ها، افزایش سود، حفظ مشتریان، جلب مشتریان تازه و نیز توسعه و ارائه خدمات جدید را فراهم آورند.

۱. تغییرات فناوری اطلاعات

همانند سایر بخش های اقتصادی، صنعت بیمه نیز شاهد تغییرات بسیاری در عرصه فناوری اطلاعات در طی سال های گذشته بوده است. پیشرفت در زمینه های سخت افزار، نرم افزار و شبکه های فناوری

۱. این روش به عنوان داده کاوی نیز شناخته می شود. [مسئول شورای سردبیری]

۲. کارشناس مدیریت بیمه

اطلاعات فواید بسیاری را از قبیل کاهش هزینه ها، کاهش زمان پردازش داده ها، افزایش قابلیت های سودآوری همزمان با رویارویی با چالش های جدید در فضایی که هر روز شاهد افزایش میزان رقابت در آن بوده ایم به همراه داشته است.

ابداعات و نوآوری های فنی از قبیل استخراج و نیز نگهداری داده ها به طور قابل ملاحظه ای موجبات کاهش هزینه های نگهداری، دستیابی و نیز پردازش داده ها را فراهم آورده است. بسیاری از سؤال های تجاری که سابقاً به علت کمبود داده ها و امکانات پردازشگری غیر ممکن، غیر عملی و یا غیر قابل حل می نمودند، در حال حاضر با استفاده از تکنیک استخراج داده ها قابل بررسی و پاسخگویی هستند. برای مثال یکی از سؤال های معمول تجاری این بود که چگونه شرکت های بیمه می توانند مشتریان خود را حفظ نمایند؟ در حال حاضر این سؤال با بهره گیری از تکنولوژی استخراج داده ها جواب داده می شود و شرکت های بیمه قادرند نرخ و خدمات مناسب را با توجه به نیاز مشتریان خود ارائه نمایند و درعین حال تقارن صحیحی میان حق بیمه و رفتار مشتریان نیز برقرار نمایند.

در عین حال با بهره برداری از تکنولوژی جدید استخراج داده ها می توان طرح های صحیح تر و عملی تری را در زمانی کمتر در مقایسه با طرح های ایجاد شده با تکنولوژی های قبلی ارائه داد. میانجی نگاره ای کاربر (GUI) در حال حاضر قادر به تهیه مدل های بسیار پیچیده و نیز مقایسه ماهرانه میان روش های مختلف است. با ایجاد مدل های خوب، کامل و آزمایش شده در زمانی کوتاه تر نسبت به قبل این امکان برای شرکت های بیمه فراهم شده است تا نظارت صحیح تری به مسائلی از قبیل اشتباهات در زمان صدور بیمه نامه و نیز انتخاب نامساعد در زمان بازاریابی داشته باشند.

۲. استفاده از روش استخراج داده ها در صنعت بیمه

روش استخراج داده ها اغلب می تواند امکان اصلاح دستاوردهای آماری سنتی را در جهت حل مسائل بازرگانی فراهم آورد. برای مثال، استفاده از رگرسیون خطی امکان حل مسائل بازرگانی را در راستای نیاز بیمه گران و به همراه مدل ها و پارامترهای قابل تفسیر مهیا می نماید. امکان اصلاح مدل های حاضر در اغلب موارد با استفاده داده ها و از طریق کشف متغیرهای جدید، شناخت روابط و نیز

۱. GUI (میانجی گرافیکی کاربر) : روشی جهت بهره گیری از کامپیوتر با استفاده از محیطی گرافیکی علاوه بر متن می باشد. این عبارت به عنوان "رابط گرافیکی کاربر" نیز شناخته می شود. [مستول شورای سردبیری]

کشف روابط غیر خطی فراهم می شود. مدل‌هایی که قادر به تخمین و پیش بینی روابط و رفتارها باشند می توانند سود بیشتری را با هزینه های کمتر حاصل نمایند.

استخراج داده ها شرکت های بیمه را در انجام امور ذیل کمک می کنند :

- ارائه نرخ
- جذب مشتریان جدید
- حفظ مشتریان سابق
- ارائه خدمات جدید بیمه ای
- ارائه گزارش منطقه ای از ریسک
- کشف ادعاهای جعلی
- اجرای مدیریت ماهرانه طرح‌های بازاریابی
- تخمین میزان خسارت معوق
- کمک به بیمه‌گران در جهت شناخت نرخ‌ها و مدل‌های مناسب
- هماهنگی واحدهای محاسبات آماری و بازاریابی

۱-۲. ارائه نرخ

مهم‌ترین مشکل در علوم آماری نحوه نرخ گذاری یا قیمت گذاری بیمه نامه هاست. هدف از نرخ‌گذاری این است که میزان ریسک بیمه‌گذار از طریق ارائه نرخ مناسب نشان داده شود به طوری که هر چه ریسک کمتر باشد نرخ مورد عمل نیز پایین تر باشد.

۱-۱-۲. شناسایی مؤلفه های پیش بینی‌کننده خسارت، سود و زیان

سؤال مهمی که در زمینه نرخ گذاری مطرح می شود این است که از چه مولفه هایی برای تعریف ریسک استفاده شود که نشان دهنده میزان احتمال وقوع خسارت و بزرگی آن باشد؟ برای مثال در بیمه اتومبیل، ارتباط مستقیمی میان احتمال وقوع خسارت و نزدیکی بیمه گذار به مناطق شهری وجود دارد که به همین ترتیب یک آکچوئر با علم به این مطلب به ارائه نرخ بر اساس آدرس پستی اعلامی توسط بیمه‌گذار خواهد پرداخت، در نتیجه به مزدی سی ساله ساکن در مناطق شهری با وجود دریافت یک

خسارت در سه سال گذشته نرخ بالاتری تعلق خواهد گرفت. به همین ترتیب در بیمه‌های درمان نیز حق‌بیمه بالاتری به فرد سیگاری تعلق می‌گیرد، لذا برخی از شرکت‌های بیمه میان مواردی از قبیل تنباکو و محصولات آن که سلامتی را در معرض خطر قرار می‌دهند و سایر موارد تفاوت قایل هستند. اگرچه بسیاری از مولفه‌های ریسک که بر روی میزان نرخ تاثیرگذارند مشخص می‌باشند اما روابط غیر مشهود و پیچیده‌ای میان متغیرهای مختلف وجود دارند که شناسایی آنها مشکل بوده و بررسی آنها نیازمند تجزیه و تحلیل‌های پیچیده است. مدل‌های پیشرفته استخراج اطلاعات می‌توانند به شناخت بهتر ریسک بیانجامند، بنابراین شرکت‌های بیمه می‌توانند نرخ بهتر و مناسبتری را ارائه داده و سود بیشتری را هم‌زمان با هزینه‌های کمتر به دست آورند.

۲-۱-۲. افزایش میزان صحت پایگاه داده‌های اطلاعاتی

جهت افزایش میزان صحت و درستی داده‌ها، می‌توان آنها را به گروه‌های یکنواختی تقسیم کرد که بنابراین داده‌های هرگروه را جداگانه تحت بررسی، تجزیه و تحلیل و مدل‌گذاری قرارداد و همین امر می‌تواند در تفکیک متغیرهای مرتبط با مولفه‌های ریسک به کار گرفته شود. چنانچه تفکیک داده‌ها با توجه به گروه‌های خاص خود انجام شده باشد، می‌تواند قابلیت بقایسه سریع‌تر و تفسیر راحت‌تر را به همراه داشته باشد که در نتیجه آکچوئرها نیز می‌توانند پیش‌بینی صحیح‌تری را از احتمال وقوع خسارت و میزان آن داشته باشند. برای مثال، یک شرکت بیمه با این روش به این نتیجه رسید که گروهی از رانندگان مرد در رده سنی ۱۸ تا ۲۰ سال از میزان خسارت کمتری نسبت به کل جامعه آماری برخوردار می‌باشند. حال چه عاملی موجب تفاوت بارز میان خسارت گروه مزبور نسبت به کل شده بود؟ بنابر تحقیقات انجام گرفته افراد مزبور از ماشین‌هایی که به نسبت قدیمی‌تر از سایر افراد کل گروه بود، استفاده می‌کردند، لذا ایشان زمانی را نیز صرف تعمیرات و نگهداری از وسیله نقلیه خود می‌کردند که همین امر موجب گشته بود تا این گروه در راندن وسیله نقلیه خود بسیار محتاط‌تر از سایر افراد هم سن خود باشند.

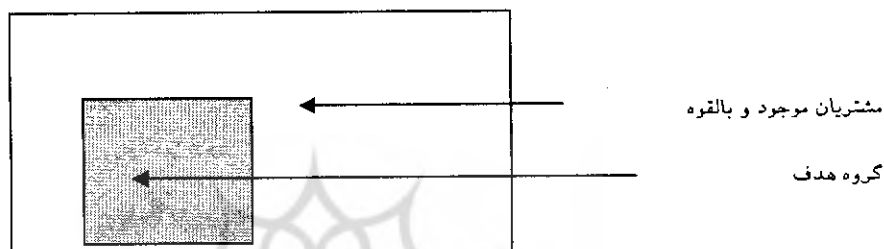
۲-۲. جذب مشتریان جدید

یکی دیگر از مسائل مهم مرتبط با میزان جذب مشتریان جدید نحوه ارائه نرخ است. اگر چه روش‌های سنتی افزایش جذب مشتریان را با افزایش تلاش واحد فروش یکسان می‌داند، اما چنانچه این

تلاش به وسیله دستاوردهای حاصل از استخراج کمی داده ها هدایت شده باشد، می تواند نتایج بهتری را به ارمغان بیاورد.

۲-۲-۱. تمرکز استراتژی بازاریابی بر دستیابی به گروه های هدف

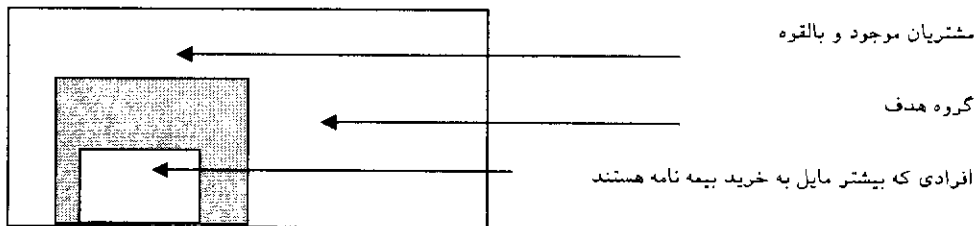
در روش بازاریابی سنتی، افزایش تعداد بیمه گذاران با افزایش دسترسی به افرادی که در محدوده گروه های هدف قرار داشتند صورت می گرفت (نمودار شماره ۱). اشکال چنین خط مشی بازدهی کم در قبال تلاش بسیار جهت بازاریابی بوده و در بعضی مواقع علی رغم صرف بودجه بسیار، میزان درآمد حاصله از روند رو به کاهشی برخوردار بوده است.



نمودار ۱. بازاریابی تمامی افراد مشمول بیمه نامه

۲-۲-۲. افزایش بازگشت سرمایه با به کارگیری طرح های بازاریابی

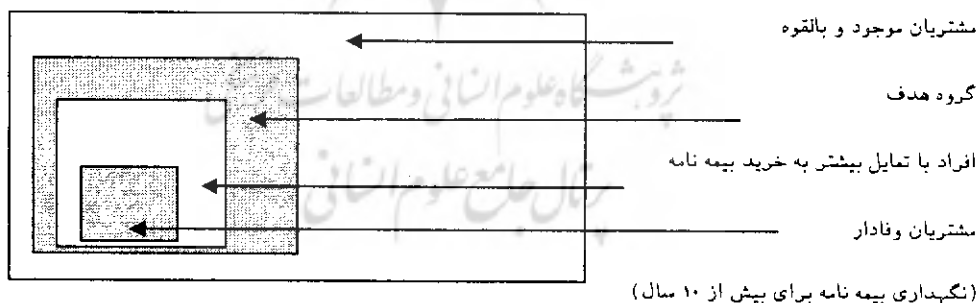
در مقایسه با خط مشی سنتی فروش، استراتژی استخراج داده ها بسیاری از تحلیل گران را به اصلاح دیدگاه های بازاریابی رهنمون ساخته است، برای مثال با افزایش طول عمر بیمه گذاران می توان به حق بیمه بیشتری با توجه به افزایش طول زمان دست یافت بنابراین همان طور که از نمودار ۲ برمی آید یکی از مهم ترین سوالات بازاریابان این است که از میان افرادی که می توانند مشمول شرایط بیمه نامه باشند چه افرادی بیشتر مایل به خرید بیمه نامه هستند چرا که در این صورت تنها تعدادی از افراد مشخص که به احتمال بیشتر خرید خواهند کرد هدف قرار گرفته و در نتیجه میزان درآمد حاصله در قبال بازاریابی انجام شده افزایش می یابد.



نمودار ۲. بازاریابی افرادی که تمایل بیشتر به تهیه بیمه نامه دارند

آیا با این روش ممکن است نتیجه بهتری حاصل شود؟ یا به عبارت دیگر آیا در صورت جمع آوری داده های بیشتر می توان مدل های بهتر و در نتیجه بازاریابی متمرکزتری انجام داد؟ جهت تمرکز بیشتر برخی از تحلیل گران بیمه با بهره گیری از تکنیک استخراج داده ها و نیز مجزا کردن گروهی از مشتریان که از عمر مفید بالاتری برخوردار هستند اقدام به ارائه مدل هایی برای تشخیص گروهی کردند تمایل بیشتری به خرید بیمه نامه دارند.

برای مثال، در ابتدا طرح بازاریابی شامل افرادی می شود که در محدوده قرارداد گنجیده و با توجه به وفاداریشان به شرکت مزبور اقدام به تهیه بیمه نامه می نمایند. (نمودار ۳). لذا تفکیک یا مجزا کردن مشتریان بالقوه جهت تمرکز بیشتر بر روی ایشان در طرح های بازاریابی موجب افزایش کارایی و نیز افزایش بازدهی می شود.



(نگهداری بیمه نامه برای بیش از ۱۰ سال)

نمودار ۳. بازاریابی افرادی که بیشتر مایل به حفظ بیمه نامه های خود هستند

۳-۲. حفظ بیمه گذاران

مسئله حفظ بیمه گذاران کنونی مرتبط با تمرکز عمل بازاریابی بر اشخاصی است که احتمال

خرید مجدد بیمه نامه در ایشان بسیار زیاد است.

۲-۳-۱. پیشنهاد مجموعه ای از بیمه نامه ها

تجربه نشان داده است که احتمال مراجعه برای تمدید بیمه نامه در بیمه‌گذارانی که دارای دو بیمه نامه از یک شرکت بیمه هستند نسبت به بیمه‌گذارانی که دارای یک بیمه نامه‌اند، بیشتر است و به طریق اولی بیمه‌گذارانی که دارای سه بیمه نامه از یک شرکت بیمه‌اند تمایل بسیار کمتری به تغییر بیمه‌گر خود خواهند داشت. بنابراین با ارائه تخفیف مناسب در صورت تهیه مجموعه ای از بیمه نامه ها به بیمه‌گذار، بیمه‌گر علاوه بر افزایش میزان وفاداری بیمه‌گذار خود امکان کاهش مراجعه ایشان به شرکت‌های رقیب را فراهم می‌نماید.

۲-۳-۲. تجزیه و تحلیل سطح نیاز مشتریان

حفظ موفقیت‌آمیز بیمه‌گذاران نیازمند تجزیه و تحلیل داده‌ها با توجه به هر یک از سطوح مشخص مشتریان و نه تمامی سطوح است. صنعت بیمه مدت‌های مدیدی را صرف تجزیه و تحلیل سوددهی با توجه به هر یک از سطوح مشتریان و همگام با پاسخ به این پرسش که متقاضیان بیشتر مایل به تهیه چه نوع بیمه نامه‌هایی به صورت جامع هستند، کرده است و در همین راستا با استفاده از ابزار داده‌ها، گسترش فن آوری کامپیوتر و استخراج اطلاعات امکان پاسخگویی فراهم شده است. با استفاده از تکنیکی به نام تحلیل پیوندها^۱، شرکت‌های بیمه می‌توانند به اموری به قرار ذیل و نیز ارائه خدمات متناسب با نیاز مشتریان پردازند:

- تهیه خلاصه اطلاعات مشتریان به تفکیک اطلاعات ایشان
- تحلیل میزان نرخ و خسارت با توجه به یک بیمه‌گذار و یک نوع خاص از خدمات، برای مثال بررسی ارائه خدمات جدید توسط بیمه‌گر جهت گروه خاصی از بیمه‌گذاران
- تحلیل رفتار دسته ای از بیمه‌گذاران با توجه به مولفه‌های متعدد و ارائه بیمه نامه‌های متفاوت، برای مثال تحلیل میزان سود رسانی ارائه مجموعه ای از بیمه نامه‌ها (عمر، منازل مسکونی، اتومبیل و...)
- به گروه خاصی از بیمه‌گذاران
- تحلیل بازار از نقطه نظر بیمه‌گذاران، برای مثال چه درصدی از بیمه‌گذاران کنونی که از بیمه اتومبیل استفاده می‌کنند در طی ۵ سال آینده مایل به خرید بیمه نامه‌های عمر نیز خواهند بود؟

۲-۳-۳. حفظ بیمه گذارانی که مایل به تغییر بیمه گران خویش می باشند

با استفاده از تفکیک اطلاعات و بهره گیری از روش مدرن مدل سازی، تحلیل گران قادر خواهند بود تا گروه هدف را تعیین نمایند. بیمه گذاران حاضر که بیشتر مایل به تغییر بیمه گر خویش می باشند می توانند با استفاده از مدل پیش بینی مشخص شوند. مدل رگرسیون منطقی دستاوردی سنتی است که افرادی را که احتمالاً بیمه گر خود را تغییر می دهند به عنوان گروه هدف قرار می دهد.

• شناسایی گروه هدف توسط مدل سازی بر اساس رفتار بیمه گذاران بهبود می یابد. با استفاده از معادلات غیر خطی، روابط و مدل های شبکه عصبی امکان ارائه داده های صحیح تر در راستای شناسایی بیمه گذارانی را که احتمالاً بیمه گر خود را تغییر خواهند داد فراهم می شود. علاوه بر این، مدل تصمیم گیری درختی می تواند با تفکیک بیمه گذاران در گروه هایی متناسب و همجنس، موجبات کاهش هزینه ها و نیز افزایش حفظ بیمه گذاران را فراهم نماید.

۲-۴. ارائه محصولات جدید بیمه ای

تغییرات مداوم بازار، جستجوی مداوم بیمه گذاران را جهت دستیابی به خدمات مناسب ایجاب می نماید. لذا این امر که یک بیمه گر بتواند با مشاهده و تشخیص تغییر نیازها به تغییر و اصلاح خدمات پیشنهادی و بیمه نامه های خود بپردازد، می تواند تعیین کننده تداوم سود یک شرکت باشد.

۲-۴-۱. شناسایی بیمه گذاران سودده

شرکت های بیمه می توانند با شناسایی بیمه گذاران پردرآمد و نیز اولویت بندی عملیات بازاریابی، سود حاصله خویش را افزایش دهند. برای مثال، لویدز لندن از شهرت خاصی در امر بیمه کردن هر موردی با توجه به نرخ های اعلام شده از سوی پروکرهای لویدز برخوردار است، مثلاً در یک مورد نادر، قدرت چشایی فردی به نام اگون رونای - متخصص و کارشناس تغذیه - به میزان ۲۵۰,۰۰۰ پوند توسط لویدز لندن در سال ۱۹۹۹ بیمه شد.

مشکلاتی مرتبط با میزان سود زمانی به وجود می آید که شرکت های بیمه، بیمه نامه مناسب را به همراه نرخ مناسب و در زمان مناسب، به گروه مناسبی از بیمه گذاران تخصیص نمی دهند. برای مثال پرسودترین گروه بیمه گذاران ممکن است پرخطرترین گروه باشند که تنها نرخ بالاتری در خصوصشان اعمال می شود و یا محتمل است که پرسودترین بیمه گذاران، آنانی باشند که از پوشش عمر زمانی

برخوردار بوده و در طول دوره بیمه خویش، بیمه نامه ها و خدمات دیگری را نیز خریداری می نمایند. هر گروه از بیمه گذاران فوق نیازمند تحلیل و بررسی جداگانه جهت پیش بینی میزان سود مورد انتظار می باشند.

۲-۴-۲. تطبیق با تغییرات بازار

جهت تطبیق با تغییرات، شرکت های بیمه نیاز دارند بدانند که کدامیک از بیمه نامه های جدید سود آور خواهند بود. برای مثال، آکچوئرها به طور ذهنی به ارائه نرخ جهت بیمه عمر مسافران سفینه شاتل می پردازند. واحد بازاریابی می تواند در جهت توسعه بیمه فوق تلاش کند اما آیا درآمد حاصل از فروش می تواند جوابگوی تلاش های انجام شده باشد؟ چه بسا نه، زیرا فروش چنین بیمه نامه هایی در مقیاس زیاد تقریباً غیر ممکن است.

مشخصاً این سوال که چه سرویس و یا بیمه نامه ای می باید پیشنهاد شود مرتبط با مسائلی از قبیل جذب، عدم امکان اشتباه و نیز حفظ بیمه گذار است. تکنیک استخراج داده ها شرکت های بیمه را قادر می سازد که از پاسخ یک مسئله به عنوان داده ای جهت ایجاد سوالات دیگر استفاده نمایند. تحلیل گران در جستجوی پاسخ به پرسش های متعدد مرتبط با تفکیک مشتریان، معرفی محصولات جدید و نیز میزان سودآوری از طریق پاسخ به سوالات ذیل می باشند:

۱. پس از انجام تحلیل های مختلف بر خدمات موجود کدام پوشش برای گروه های مختلف مشتریان پرسود قلمداد می شود؟

۲. کدامیک از خدمات مزبور، بازار بزرگتری را در بر خواهد داشت؟

۳. کدامیک از خدمات مزبور، به راحتی در بازار به فروش می رسند؟

۴. شرکت های بیمه در حال حاضر می توانند با استفاده از تمامی اطلاعات در دست خود به بازاریابی و ارائه پوشش بهتر بپردازند.

۲-۴-۳. اولویت بندی معرفی خدمات جدید

خدمات جدیدی که به عنوان خدمات سودآور به بازار معرفی می شوند می توانند جهت ورود به بازار اولویت بندی شوند. لذا با توجه به اهداف شرکت خدمات جدید بر اساس سود مورد انتظار، تعداد بیمه گذاران و یا سرعت مورد انتظار در پذیرش آن اولویت بندی می شوند. تکنولوژی استخراج

داده ها به تشخیص گروه هدف جهت مدل سازی رفتارها و محاسبه میزان سود حاصل از عملیات بازاریابی پرداخته و در نهایت با مقایسه استراتژی های مختلف به اتخاذ راهکارهای مناسب می پردازد.

۵-۲. گزارشی از میزان ریسک در مناطق مختلف

شرکت های بیمه همچنین می توانند داده های آماری خود را با استفاده از داده های جغرافیای اجتماعی که با توجه به طول و عرض جغرافیایی و مسافت ها به دست می آید تکمیل نمایند. تکمیل داده های حاضر با توجه به اطلاعات جغرافیای اجتماعی به این علت است که منطقه سکونت متقاضیان یکی از مولفه های مهم در مدل نرخ گذاری است. برای مثال در مناطق مختلف نحوه رانندگی و احتمال وقوع تصادف و سرقت ماشین متفاوت است در نتیجه تفاوت در مولفه های خطر در احتمال خسارت، میزان خسارت و در نهایت میزان نرخ اثرگذار خواهد بود.

با برخورداری از داده های جغرافیایی بانک داده ها، شرکت های بیمه می توانند نقشه ای ایجاد نمایند به این صورت که تحلیل گران با کنار هم قرار دادن داده ها به ارزیابی و نیز نشان دادن ریسک در مناطق مختلف پردازند. لازم به ذکر است چنین توانایی برای تحلیل و ترسیم داده ها تنها برای ترسیم جغرافیایی داده ها و نشان دادن آنان نبوده بلکه برای نرخ گذاری و تجزیه و تحلیل مدل ها نیز استفاده می شود و چنانچه مناطق پرریسک مشخص شوند می توان با استفاده از تعدیل نرخ و یا استفاده از پوشش اتکایی به کاهش میزان ریسک پرداخت.

۶-۲. کشف خسارات جعلی

به طور آشکار، خسارات جعلی از مشکلات همیشگی شرکت های بیمه محسوب می شوند و استفاده از روش هایی جهت شناسایی و کشف آنان می تواند در دراز مدت نقش مهمی را در موفقیت شرکت های بیمه به همراه داشته باشد. در اغلب موارد استفاده از پروژه استخراج داده ها جهت کشف موفقیت آمیز اعمال کلاهبردارانه به افزایش میزان درآمد منجر می شود.

۶-۲-۱. کاهش میزان خسارت با کشف هر چه بیشتر خسارات جعلی

کلاهبرداری در حال حاضر امری رو به رشد است، برای مثال سالانه در آمریکا بیش از ۱ تریلیون دلار به امر بهداشت اختصاص داده می شود که از این میزان تقریباً بیش از ۱۰۰ میلیارد دلار به علت کلاهبرداری و جعل از بین می رود که به آن می توان ۲۰ میلیارد دلار در قسمت اموال را نیز افزود. در

عین حال لازم به ذکر است که کلاهبرداری در صنعت بیمه رتبه ۲۵ را در میان ۵۰۰ صنعت دیگر دارا می باشد. حذف مشکل بزرگی چون کلاهبرداری این امکان را برای شرکت های بیمه فراهم می آورد که بتوانند به ارائه محصولاتی رقابتی اقدام نمایند که ضمن حفظ طولانی مدت، سوددهی را نیز دربرداشته باشد.

۲-۶-۲. کدامیک؛ وجود الگو یا تصادفی بودن

خسارات جعلی غالباً خسارات بزرگ را شامل نمی شود چراکه مجرمان می دانند موارد بزرگ نسبت به موارد متوسط بیشتر تحت نظارت و موشکافی هستند. در نتیجه مرتکبان چنین اعمالی نگاه دقیق تری به این مسئله خواهند داشت. به همین دلیل تحلیل گران می باید در دستیابی به خسارات جعلی، در داده های مرتبط با خود به گروه های غیر معمول، غیر متعارف و نیز الگوهای سرمایه گذاری توجه نمایند. از جمله روش های تجزیه و تحلیل مخصوصی که برای یافتن چنین مواردی وجود دارد روش تحلیل سبد بازار، روش خوشه ای و مدل پیش گوینه است. با مقایسه متسفیهرای مورد انتظار و واقعی، میزان انحراف زیادی را می توان به دست آورد که کاملاً قابل پیگیری است. برای مثال طبق گزارش Empire Blue Cross and Blue Shield ایشان با استفاده از پایگاه داده ها به جمع آوری تعداد برونکوسکوپی هایی که توسط پزشکان گوش و حلق و بینی انجام می شد پرداختند و در گزارشی که اخیراً اعلام شده است این میزان را با تعداد ۴/۱ میلیون نفر بیمه شده ای که از این خدمات استفاده کردند مقایسه نمودند که همین امر بالا بودن تعداد برونکوسکوپی های یکی از پزشکان را از میانگین موجود نشان داد، لذا تحقیقات بیشتری صورت گرفت که مؤید این مطلب بود که پزشک مزبور اقدام به ارائه نسخه های جعلی نموده است. بنابراین وی محکوم به ۱۰ ماه حبس، ضبط حکم اشتغال به کار و نیز جریمه نقدی به میزان ۵۸۶,۰۰۰ دلار شد. به همین ترتیب با استفاده از روش استخراج داده ها و کشف پرداخت های جعلی به تنهایی موجب ذخیره ۴ میلیون دلار در سال ۱۹۹۷ گردید.

۲-۷. تهیه پوشش اتکایی

پوشش اتکایی در واقع پذیرش ریسک از بیمه گر در قبال دریافت بخشی از حق بیمه پرداختی توسط بیمه گذار است. پوشش اتکایی از طریق پراکندگی ریسک این امکان را برای یک شرکت فراهم

آورد که ریسک‌های بزرگی را که به تنهایی توسط یک بیمه گر قابل قبول نیستند، پذیرند. اگر چه ابتیاع پوشش اتکایی موجب کاهش میزان خسارت می گردد اما در عین حال، میزان در آمد شرکت‌های بیمه را نیز کاهش می دهد. در واقع هدف دست یابی به سطح مناسبی از پوشش اتکایی است، زیرا همان‌طور که اخذ پوشش اتکایی کم، توسط شرکت‌ها آنها را در معرض خطر بزرگ قرار می دهد، اخذ پوشش اتکایی بسیار نیز موجب کسب در آمد کم می‌شود. علاوه بر این، بخشی از تناسب خطر/ در آمد نیازمند این است که بیمه گر اتکایی از اعتبار خوبی به جهت نگهداری ریسک و نیز پرداخت خسارت برخوردار باشد.

۲-۷-۱. محدودیت های حاصل از روش‌های تجزیه و تحلیل سنتی

روش‌های سنتی در کسب پوشش اتکایی می تواند منجر به اتکایی کردن بیمه نامه‌هایی شود که از احتمال خسارت کمتری در طول دوره بیمه نامه برخوردارند. برای مثال، در بسیاری از انواع بیمه های عمومی توزیع خسارت از چولگی برخوردار بوده و به سمت راست مایل است. چنانچه بیمه‌گر از بیمه نامه‌های بسیاری در یک بخش خاص برخوردار باشد انتظار می رود که نمودار پرداخت خسارت آن به دلیل تعداد زیادی از خسارت کوچک از منحنی نرمال برخوردار شود. از سوی دیگر چنانچه میزان داده‌ها محدود باشد، دنباله احتمالات نقش مهمی را ایفا می نماید. در چنین شرایطی سطح اطمینان بیمه‌گران به نسبت تخمین میزان ریسک کاهش یافته است که همین کاهش سطح اطمینان، نقش مهمی را در ایجاد انگیزه جهت اخذ پوشش اتکایی ایفا می نماید. در چنین شرایطی بیمه گر اغلب اقدام به اخذ پوشش اتکایی برای مواردی می نماید که احتمال وقوع خسارت در آن بالا باشد.

۲-۷-۲. فواید تکنیک استخراج اطلاعات در عملیات اتکایی

تکنولوژی استخراج داده ها برای تفکیک اطلاعاتی استفاده می شود که مدل‌سازی میزان وقوع خسارات آن در پوشش اتکایی با توجه به خسارات تسویه شده صورت می گیرد. در نتیجه با تفکیک بیشتر، تحلیل‌گران می توانند نسبت به نتیجه حاصله مطمئن تر باشند. لذا انتخاب بیمه نامه ها برای اخذ پوشش اتکایی براساس مدل ارزیابی ریسک به جای مدل سنتی صورت می گیرد.

۲-۷-۳. کشف نحوه پراکندگی خسارات

معمولاً بیمه گران در بسیاری از موارد برای یافتن توزیع خسارت از توزیع نرمال استفاده می‌نمایند

اما در مواردی که داده ها به میزان مناسب وجود نداشته و یا امکان انحراف و یا احتمال وقوع خسارات سنگین بسیار باشد، استفاده از توزیع پارتو (Pareto) به علت در بر گرفتن دامنه وسیعی از داده ها و احتمالات مناسب است. در واقع، برای بیمه گر و بیمه گر اتکایی استفاده از توزیع نرمال برای نرخ گذاری در زمانی که استفاده از توزیع پرتو مناسب تر بوده می تواند اشتباه بزرگی قلمداد شود چراکه در این قبیل موارد در دست داشتن ابزار صحیح جهت شناخت و پیش بینی توزیع خسارت مهم است.

۴-۷-۴. نقش مدل ساز

مدل ساز می تواند با جستجوی سوابق خسارات به تفکیک بیشتر اطلاعات در یک زمینه خاص پردازد که تجربه فوق می تواند با مدل گذاری صحیح در شناخت بیمه نامه های پر ریسک، به پیش بینی های صحیح تری نیز بیانجامد. به همین ترتیب شرکت های بیمه با اجتناب از تقسیم حق بیمه دریافتی در زمانی که نیاز به اخذ پوشش اتکایی وجود ندارد و نیز اخذ پوشش اتکایی در مواردی که بیمه نامه ها و موارد پرریسک وجود دارد می توانند به کاهش هزینه ها توأم با افزایش سود دست یابند. با استفاده از توزیع خسارات می توان پاسخ این پرسش را که کدامیک از بیمه نامه ها نیازمند اخذ پوشش اتکایی هستند با شناسایی بیمه نامه های پر ریسک انجام داد. لذا بر همین اساس بیمه نامه هایی که احتمال بروز خسارت در آنها بالا بوده در قبال پرداخت بخشی از حق بیمه، اتکایی می شوند.

۴-۸. اجرای مدیریت ماهرانه عملیات بازاریابی

توسعه ارتباطات با مشتریان سابقه طولانی در عرصه تجارت داشته است. شرکت های کوچک و خرده فروش ها قادر به برقراری ارتباط شخصی و رو در رو با مشتریان خود می باشند. هرچند با رشد سازمان ها در حال حاضر اغلب واحدهای بازاریابی به فکر برقراری شرایط لازم جهت توسعه محصول خود به جای توسعه ارتباطات و حفظ مشتریان خود می باشند که همین امر موجب شده تا تمرکز واحد فروش و بازاریابی بر روی شرایط و سرعت هر چه بیشتر جهت ارائه انبوهی از محصولات به بازار به جای تمرکز بر روی ارائه محصولاتی مطابق با خواسته های تک تک مشتریان امری غیر معمول نباشد. نهایتاً مشکل آنجاست که وقتی بازارها اشباع می شوند، تاثیر بازاریابی تولید انبوه کاهش یافته و یا به طور کامل متوقف می شود. خوشبختانه تکنولوژی پیشرفته استخراج اطلاعات، شرکت های بیمه را قادر می سازد تا تمرکز بازاریابی خود را به سوی خصوصیات مختلف مشتریان وفادار، بازگردانند.

استراتژی‌های بازاریابی خلاق، داده محور و علمی، سنگ بنای بازگشت به مدیریت ارتباطات با مشتریان، در اقتصادی ساده و کارآمد اما در مقیاسی جامع و بزرگ است.

۲-۸-۱. تمرکز مشتری محور

بسیاری از شرکتهای بیمه پیشرو تلاش می کنند تا از طراحان محصول محور گذشته دوری گزیده و تمرکز خود را به سوی طراحان مشتری محور، جهت ارائه سرویس بهتر به مشتریان جلب نمایند. تکنولوژی استخراج اطلاعات می تواند برای درک بهتر از خواسته ها و نیازهای مشتریان مورد استفاده قرار گیرد. تجزیه و تحلیل عملیات بازاریابی بازخورد عمیقی را مهیا نموده و زیرساخت توسعه عملیات آتی را مهیا می نماید.

۲-۸-۲. بازاریابی؛ گامی دیگر به سوی اتوماسیون

شناخت عمیق تر نیازها، خواسته ها و پاسخهای مشتریان هم اکنون از طریق قدرت متدولوژی استخراج اطلاعات و حمایت تکنولوژیهای کامپیوتری میسر است. راه حل‌های مدیریتی شامل ابزارهای گرافیکی است که بازاریابان را قادر به تجزیه و تحلیل اطلاعات مشتریان، تجزیه و تحلیل عملیات بازاریابی پیشین، طراحی عملیات جدید، ارزیابی آنان قبل از اجرا و در حال اجرا و نیز به ارزیابی کارایی طرحهای تکمیل شده می پردازد.

بازاریابی مشتری محور از طریق تلفیق مدیریت استخراج اطلاعات و مدیریت عملیات بازاریابی میسر می شود. این تلفیق یک ارتباط دایره وار را به وجود می آورد که از طریق آن آنالیزورهای مدیریت استخراج اطلاعات می توانند رفتارهای مشتریان را بر اساس خواسته های بازاریابان ایجاد و آزمایش نمایند، سپس بازاریابان می توانند این مدل را جهت پیش بینی رفتار مشتریان به کار ببرند. با ارزیابی رفتارهای پیش بینی شده مشتریان، بازاریابان حرفه ای می توانند اشکالات عملیات بازاریابی را بیشتر مورد شناسایی و تصحیح قرار دهند.

۲-۸-۳. راه حل کلی

ابزار مدیریت عملیات بازاریابی می بایست حامی اجرای تمام و نه فقط قسمتی از مراحل دوره‌های بازاریابی - شامل تجزیه و تحلیل، برنامه ریزی، اجرا و ارزیابی- باشد. متأسفانه بسیاری از محصولات مدیریت عملیات بازاریابی شامل ابزارهای جداگانه در سیستم جهت استفسار، به روز رسانی، تجزیه و

تحلیل و استخراج اطلاعات طراحی شده اند. اغلب این سیستمها نیاز به تغییرات بسیار برای انطباق با نیازهای مختص هر شغل دارند. لذا بهترین راه حل آن است که بتوان با تلفیق عملیات اصلی ارزیابی، ذخیره سازی و استخراج داده ها و مدیریت عملیات بازاریابی سیستمی کارا و منسجم را ایجاد نمود در حالی که اگر این ابزار به طور واضح و شفاف جهت کار با یکدیگر طراحی و هماهنگ نشده باشند، مشکلاتی در قابلیت انتقال این طرحها و در توان عملیاتی شان به وجود خواهد آمد. چنانچه در بسیاری از مواقع یک راه حل تدریجی کاربردی ممکن است متفاوت از یک راه حل کلی خوب باشد.

۲-۸-۴ مدیریت مؤثر طرح های بازاریابی

هماهنگی و یکپارچگی تکنولوژی های ارزیابی، ذخیره سازی و استخراج داده ها و مدیریت عملیات بازاریابی، بازاریابان حرفه ای را قادر می سازد مدل های اولیه استخراج اطلاعات را در سیستم مدیریت طرح های بازاریابی خود به کار برند، سپس بازاریابان می توانند از میان لیست این مدل ها، مدل مورد نظر را انتخاب نموده و آن را به همراه سیستم مدیریت عملیات بازاریابی شناسایی شده، جهت اهداف زیر مجموعه مشخص خود به کار برند. کد امتیازبندی معمولاً بر روی زیرمجموعه های انتخاب شده ای از محصولات حاصل از استخراج اطلاعات به کار رفته و سپس فایل های امتیازبندی شده جهت استخراج و بهره برداری بر مبنای عملیات بازاریابی هدفمند، به بازاریابی حرفه ای عودت داده می شود. این روش هماهنگ سازی معمولاً به عنوان امتیازبندی پویا شناخته شده است.

۲-۹ تخمین میزان خسارات معوق

پرداخت خسارات همواره با تأخیر بوده است، برای مثال در بیمه مسئولیت، شدت و یا مقدار خسارت بعد از گذشت سالها از وقوع خسارت قابل ارزیابی است. در موضوعاتی مانند مسئولیت کارمندان حتی تأخیر در اعلام خسارات نیز وجود دارد. این تأخیرها باعث عدم توزیع استاندارد خسارات، مخصوصاً ایجاد چولگی توزیع ها (در توزیع نرمال) در خلال زمان و در سطوح مختلف بیمه ای می گردد. کارهای تجاری شرکت می باید ادامه داشته باشد لذا اغلب تا زمان شناسایی ارزش واقعی خسارات از روش تخمین میزان خسارات وارده استفاده می شود. این تخمین به موارد ذیل بستگی خواهد داشت:

شدت خسارت

زمان محتمل برای تسویه نهایی خسارت

تأثیرات متغیرهای مالی همانند تورم، نرخ سود و غیره

تغییرات در عرف اجتماعی. برای مثال صنعت تنباکو بسیار تحت تأثیر تغییر تفکرات اجتماعی

در مورد سیگار کشیدن بوده است.

۲-۹-۱. پیش بینی ارزش واقعی خسارات

تدارک تسویه خسارات، که لازمه ادامه فعالیت یک واحد تجاری است، با تخمین میزان خسارات همراه است. دقت و صحت این بخش بسیار مهم است چراکه بودجه ای که جهت تسویه این خسارات در نظر گرفته می شود نمی تواند در بخش سرمایه گذاری بلند مدت و یا خرید دارایی های با بازده خوب، به کار گرفته شود. اگر پیش بینی خسارات کمتر از حد واقع در نظر گرفته شود ممکن است شرکت را با مشکلات مالی روبه رو نماید. در مقابل هم، اگر خسارات بیشتر از حد واقع در نظر گرفته شوند ممکن است شرکت با مشکل عدم سودآوری مواجه شود. بنابراین هر چه این تخمین به واقعیت نزدیکتر باشد، فرصت سودآوری برای شرکت بیشتر فراهم خواهد شد. تجزیه و تحلیل توزیع خسارات براساس منطقه جغرافیایی و زمان به تخمین بهتر خسارات کمک می نماید.

فناوری استخراج اطلاعات می تواند برای ایجاد توزیع خسارات و نیز تسویه خسارات گذشته به کار گرفته شود. داده ها تجزیه و تحلیل شده و مدل گذاری می شود و زمانی که یک مدل پیش بینی طراحی می شود، خسارات معوق موجود نیز بر این اساس طبقه بندی می گردد، به خصوص پارامترهای مدل و اطلاعات خسارتی برای تخمین میزان واقعی خسارت در خسارات معوق به کار می رود، لذا تخمین خسارت واقعی پشتیبانی کننده بخش خسارت است.

۲-۹-۲. به روز کردن مدل پیش بینی کننده

تخمین میزان خسارات بر اساس مدل پیش بینی گر براساس فرضیه "آینده بسیار شبیه گذشته خواهد بود" طراحی می شود. اگر این مدل به روز نشود، به مرور زمان این فرضیه تبدیل به "آینده بسیار شبیه به گذشته های دور نخواهد بود" می شود، اما چنانچه اطلاعات بیشتری در دسترس باشد، مدل پیشگویانه استخراج اطلاعات به روز شده و فرضیه به "آینده بسیار شبیه به گذشته نزدیک است" تبدیل خواهد شد.

فناوری استخراج اطلاعات آنالیزورهای بیمه ای را قادر می سازد مدل های مختلف را مقایسه نموده و آنها را بر اساس عملکردشان ارزیابی نمایند. زمانی که مدل به روز شده دارای عملکرد بهتری نسبت به مدل قدیمی تر می گردد، آن زمان برای تغییر مدل مناسب خواهد بود. با وجود تکنولوژیهای جدیدتر آنالیزورها می توانند این مدل پیشگویانه را بررسی کرده و بر اساس نیازهایشان آن را به روز نمایند.

۲-۱۰. کمک به بیمه گران

اگر چه تجزیه و تحلیل های آماری توسط متخصصان کیفی جهت دستیابی به نرخهایی بر اساس مدل های غیرخطی و سایر تجزیه و تحلیل های مرتبط محرمانه است، ولی تغییرات نهایی نرخها نیاز به بررسی بیمه گران دارد. توضیح این مدل ها و نتیجه آنها برای دیگران امری چالش زا است. خوشبختانه با به کار گیری نمودارهای گرافیکی به همراه مهارت فن بیان ارائه کنندگان آن و بیان مکتوب تجزیه و تحلیل های به کار رفته، تجزیه و تحلیل های پیچیده برای عموم ساده تر و قابل فهم تر خواهد شد. برای مثال، با افزودن نمودارهای گرافیکی مانند مدل درختی به قوانین منطقی در یک متن ساده، به تجزیه و تحلیل گران در انتقال اساس آماری تغییرات نرخ و دیگر شرایط به بیمه گران و دیگر مخاطبان صنعت بیمه کمک می نماید.

۲-۱۰-۱. هماهنگی واحدهای محاسبات آماری و بازاریابی

هماهنگی اقدامات واحدهای محاسبات آماری و بازاریابی می تواند درآمدزایی را تقویت نموده و سود آوری را افزایش دهد. به طور معمول و سنتی، مشکل محاسبات آماری و یا نرخ گذاری از مشکل بازار هدف جدا شده است و واحدهای محاسبات آماری و بازاریابی به طور مستقل از یکدیگر عمل می نمایند اما در عمل این واحدها به یکدیگر وابسته اند و اقدامات یکی بر دیگری تاثیر گذار است.

شناخت وابستگی این دو واحد به هم توسط مدیریت به همکاری ایشان خواهد انجامید. درعین حال هماهنگی لازم از طریق به کارگیری فناوری استخراج اطلاعات به دست خواهد آمد. لذا واحد بازاریابی می تواند یافته های واحد محاسبات آماری را به کار بسته و نتایج عملیات بازاریابی نیز می تواند اطلاعات لازم برای تحقیقات بیشتر در واحد محاسبات آماری را فراهم آورد.

۳. اجرای پروژه استخراج اطلاعات

مطالب زیادی در مورد نحوه اجرای پروژه استخراج اطلاعات نوشته شده است. بسیاری از کتب خودآموز و عملی که توسط متخصصان فناوری اطلاعات به رشته تحریر در آمده این موضوع را جزء به جزء مورد بررسی قرار داده‌اند. در عین حال لازم به ذکر است که پیام اصلی مشترکی که در تمامی این نوشته ها آمده آن است که اجرای این پروژه می باید در مدل واقعی و چالش زا مورد نظر قرار گیرد. مخصوصاً به کارگیری یک روش داده محور بیشتر موثر بوده و می تواند به حوزه های عملی زیر تقسیم شود:

- دستیابی به داده ها
- ذخیره سازی داده ها
- تجزیه و تحلیل داده ها
- گزارش نتایج حاصل از تجزیه و تحلیل ها
- اجرا و به کار گیری نتایج جهت کسب مزایای تجاری

۳-۱. دستیابی به داده ها

- داده های صحیح و قابل اطمینان از نیازهای اولیه سیستم استخراج اطلاعات است. در عین حال، استراتژی دستیابی به اطلاعات می باید شامل اصول کلیدی ذیل باشد:
- دسترسی به هر و یا همه نوع منابع اطلاعاتی
 - دسترسی به منابع اطلاعاتی بدون توجه به خواستگاه آن
 - حفظ منابع اطلاعاتی از طریق استفاده از سیستمهای امنیتی رایج
 - استفاده از نرم افزار گرافیکی با قابلیت اجرای استفاده آسان جهت پاسخگویی هر چه بهتر نیازها در مواقعی که احتیاجی به دانش گسترده در مورد هر داده نباشد.
 - هماهنگی با فناوری موجود به جای دسترسی به اطلاعات ریشه ای که نیازمند جایگزینی ابزار سخت افزاری و نرم افزاری جدید و یا آموزش های اضافی به کاربران است.

۳-۲. ذخیره سازی اطلاعات

منبع ذخیره سازی اطلاعات محققان صنعت بیمه را قادر می سازد تا به آسانی به اطلاعاتی که

احتمالاً در مجموعه‌های اطلاعاتی و در مکان‌های مختلف نگهداری شده باشد دست پیدا نمایند. از طریق ذخیره سازی اطلاعات محققان می‌توانند داده‌های مختلف را به منظور نیل به هدف خود در کنار یکدیگر قرار دهند. در عین حال نیاز است قبل از انجام تجزیه و تحلیل، اطلاعات نادرست، ناقص و یا سایر مشکلات موجود دیگر برطرف شوند. یک روش برای دستیابی به این امر پاک نمودن اطلاعات ناقص است، هرچند این امر خود می‌تواند باعث انحراف نتیجه حاصله گردد چراکه ممکن است اطلاعات مزبور بخش مهمی از زنجیره ارتباطی اطلاعات باشد. بخش اصلی پاک‌سازی اطلاعات مانند جای‌گذاری اطلاعات ناقص و از بین رفته، شناخت و تصحیح اطلاعات نادرست و شناسایی اطلاعات غیرمرتبط می‌تواند به راحتی از طریق فناوری استخراج اطلاعات انجام پذیرد.

ابزار خوب ذخیره سازی اطلاعات به‌طور وسیعی می‌تواند کارایی تیم استخراج اطلاعات را بهبود بخشد که همین امر موجب دستیابی به نتایج مهم، سریع‌تر و اغلب با هزینه بسیار کمتر می‌شود. نرم افزار مجری ذخیره سازی (SAS)^۱ محیط بصری را جهت مدیریت ذخیره سازی اطلاعات مهیا می‌سازد. با به‌کارگیری برنامه ذخیره سازی، آنالیزورها می‌توانند اطلاعات مهمی که خود مورد مصرف منابع اطلاعاتی، پایگاه‌های اطلاعاتی، کتابخانه‌های طبقه‌بندی شده و دیگر انبارهای اطلاعاتی می‌باشند مشخص نمایند. سپس برنامه مزبور با استفاده از اطلاعات، به ایجاد و بارگذاری کدهای مخصوصی می‌پردازد که از طریق آن اطلاعات جداسازی و تغییر شکل داده و بارگذاری اطلاعات نیز در جداول و انبارهای اطلاعاتی صورت پذیرد.

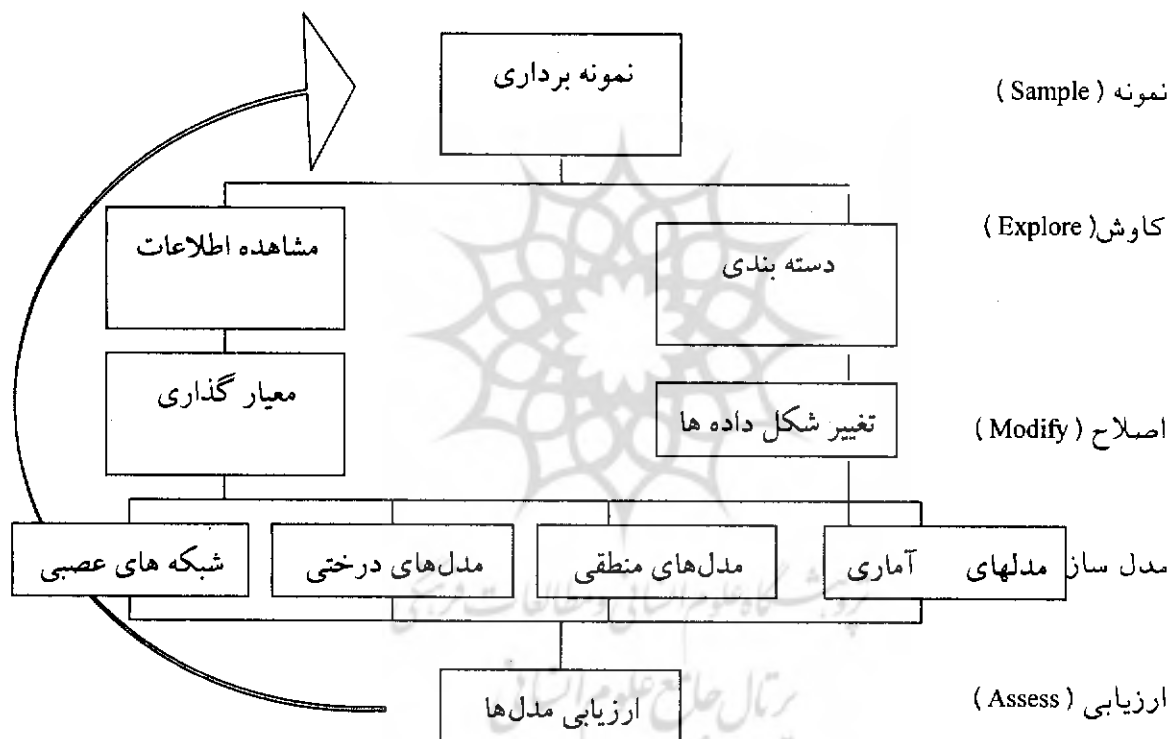
۴. تجزیه و تحلیل اطلاعات و به‌کارگیری متدولوژی SEMMA

حتی بعد از این‌که اطلاعات درخصوص موضوع خاصی ادغام و جمع آوری شد باز هم در اغلب مواقع نمی‌توان بسیاری از ارتباطات پیچیده را بر اساس داده‌های خام اطلاعاتی و یا تجزیه و تحلیل‌های آماری مشاهده کرد. در این صورت ممکن است خصوصیات واقعی به‌طور تصادفی شناخته شده و یا همچنان ناشناخته باقی بمانند. علاوه بر این، مدل‌های خطی به آسانی نمی‌توانند ارتباطات غیر خطی را پایه ریزی نمایند. فناوری پیشرفته استخراج اطلاعات می‌تواند با به‌کارگیری روش‌های ذیل بر این

۱. SAS: نرم‌افزار تحلیل آماری و برنامه ریزی تولید شده توسط شرکت نرم‌افزاری SAS است.

محدودیت‌ها غلبه کند:

اکتشاف اطلاعات GUI پیشرفته و ابزارهای نقشه کشی برای تشریح بهتر ارتباطات میان متغیرها
 متدولوژی‌های انتخاب متغیرها برای شناسایی مهم‌ترین متغیرها در جهت به کار گیری آنها در مدل‌ها
 تکنیک‌های پیشرفته مدلسازی مانند مدل‌های خطی با اثر متقابل
 شبکه های غیرخطی عصبی و مدل‌های درختی
 تکنیک‌های ارزیابی برای کمک به تجزیه و تحلیل کننده ها در انتخاب بهترین مدل اجرایی بر
 اساس معیارهای سود و زیان.



نمودار ۴: چرخه تحلیل SEMMA

با استفاده از GUI که استفاده از الگوریتم‌های پیچیده استخراج اطلاعات را تسهیل می نماید، می‌توان به ارتباطات پیچیده داده ها دست یافت. برای مثال اطلاعات زیر مجموعه ای می تواند ارتباطات مهم را برای عملیات بازاریابی آشکار نماید. جداسازی براساس منطقه جغرافیایی و شرکت بازرگانی همان‌طور که می‌تواند باعث عملیات غیر متعارف و پر هزینه برای شرکت بیمه گردد، می‌تواند باعث از دست دادن فرصت‌های سودآور به دلیل فرورفتن بیش از حد در میان داده ها نیز شود.

تجزیه و تحلیل واقعی جهت استخراج داده ها شامل مراحل انتخاب، اکتشاف و مدل سازی مقادیر زیادی از داده ها برای مشخص نمودن اطلاعات مخفی می باشد که ممکن است در مسائل تجاری مورد استفاده قرار گیرد. اغلب در این بخش پاسخ به یک سؤال ما را به سوی سؤال های جدید و سؤال های خاص بیشتری هدایت می نماید از این رو استخراج اطلاعات یک فرایند تکراری است و بهره گیری از آن نیز منجر به کشف دستاوردهای جدید و تکراری می شود.

به منظور تهیه یک مسیر قابل پیش بینی منعطف در آنالیز استخراج اطلاعات، SAS یک سیکل تجزیه و تحلیل استخراج اطلاعات معروف به SEMMA ارائه نموده است. این کلمه اختصاری نشانگر پنج مرحله آنالیزی است که به طور معمول در داخل پروژه استخراج اطلاعات وجود دارد. این پنج مرحله عبارتند از:

- نمونه: Sample
- کاوش: Explore
- اصلاح: Modify
- مدل سازی: Model
- ارزیابی: Assess

نمودار ۴ فرایند استخراج اطلاعات و طرح موقعیت پنج مرحله متدولوژی SEMMA را نشان می دهد.

در آغاز و با انتخاب نمونه، سیکل آنالیزی SEMMA تجزیه و تحلیل گران را به مرحله کشف اطلاعات با استفاده از تکنیک های بصری و آماری، آشکار سازی متغیرهای بسیار مهم قابل پیش بینی از طریق تغییر داده ها، مدل سازی متغیرها برای پیش بینی نتایج و ارزیابی مدل از طریق آزمایش آن با داده های جدید رهنمون می سازد. بنابراین سیکل آنالیز SEMMA نسخه ای پیشرفته از متدهای علمی است.

۴-۱. نمونه

اولین مرحله در روش آنالیز استخراج اطلاعات ایجاد یک یا چند جدول اطلاعاتی با نمونه برداری از مخزن اطلاعات است. نمونه ها می باید به اندازه کافی بزرگ باشد تا در برگیرنده

اطلاعات مهم باشند ولی نه آنقدر بزرگ و زیاد که از سرعت کار بکاهند. این روش از نقطه نظر هزینه با بکارگیری نمونه های قابل اطمینان و آماری از تمام اطلاعات موجود، بسیار مقرون به صرفه خواهد بود. استخراج نمونه به جای به کارگیری کل اطلاعات موجود، به طرز چشمگیری زمان کسب اطلاعات اساسی و تجاری مورد نیاز را کاهش می دهد.

به طور کلی خصوصیات کل داده ها در نمونه های آن نمایان می گردد. لذا چنانچه داده خاص و مهمی در میان داده ها وجود داشته باشد که از روش نمونه برداری قابل دستیابی نباشد از طریق خلاصه برداری می توان آنها را به دست آورد.

۴-۲. کاوش

بعد از نمونه برداری اطلاعات، مرحله بعدی کاوش به صورت بصری و یا دستیابی به داده هایی برای تشخیص تمایل و سمت و سوی ذاتی اطلاعات و یا گروه بندی آنهاست. کاوش به پالایش فرآیند کسب اطلاعات کمک می کند. چنانچه بررسی بصری اطلاعات تمایل و سمت و سوی خاصی را در اطلاعات مشخص نموده باشد. آنالیزورها می توانند از طریق تکنیک های آماری شامل آنالیز عوامل، آنالیز متناظر و دسته بندی، اطلاعات را کاوش نمایند. برای مثال اغلب خانواده ها به نیازهای واقعی شان برای داشتن بیمه های عمر آگاه هستند اما ممکن است در پی بهترین پوشش با کمترین هزینه ممکن نیز باشند. لذا به نیاز این گروه از طریق ارسال نامه و اعلام شرایط بیمه پاسخ داده خواهد شد.

۴-۳. اصلاح

اصلاح داده ها اشاره به ایجاد، انتخاب و تغییر یک یا چند متغیر برای تمرکز بر مدل فرآیند انتخاب در مسیری قابل اجرا و یا اضافه نمودن داده ها برای شفافیت و ثبات بیشتر دارد. بر اساس اکتشافات به عمل آمده در بخش کاوش ممکن است لازم باشد تا آنالیزورها داده ها را جهت لحاظ نمودن اطلاعاتی مانند گروه بندی مشتریان، زیر گروه های مهم و یا معرفی متغیرهای جدید مثل ضرایب به دست آمده از طریق مقایسه متغیرهای مشخص شده گذشته، اصلاح نمایند. آنالیزورها همچنین ممکن است نیاز داشته باشند داده های خارج از محدوده را به منظور کاهش تعداد متغیرها و دستیابی به مهم ترین آنها، شناسایی نمایند. علاوه بر این، از آنجایی که استخراج اطلاعات یک فرآیند پویا و تکرار شدنی است، در اغلب موارد وقتی اطلاعات استخراج شده قبلی به نوعی تغییر

می کنند، نیاز به اصلاح داده ها نیز احساس می شود.

۴-۴. مدل سازی

ایجاد مدل داده ها شامل به کارگیری نرم افزار استخراج اطلاعات برای جست و جوی اتوماتیک داده های ترکیبی است که نتایج قابل اطمینانی را پیش بینی می نماید.

بعد از این که داده ها ارزیابی و اصلاح شد، آنالیزورها می توانند تکنیک های مدل سازی داده ها را برای ایجاد مدل هایی که خصوصیات داده ها را تشریح می کنند، به کارگیرند. تکنیک های مدل سازی در استخراج اطلاعات شامل شبکه های عصبی، مدل های درختی، مدل های منطقی و دیگر مدل های آماری مانند آنالیز سری های زمانی است. در سیستم استخراج اطلاعات هر مدل توانایی های مخصوص خود را داشته و مناسب برای داده های خاص و در موقعیت های خاص می باشد.

۴-۵. ارزیابی

مرحله بعدی در استخراج اطلاعات ارزیابی مدل برای تشخیص عملکرد مناسب آن می باشد. یک راه معمول و مرسوم ارزیابی مدل، اجرای مدل بر روی بخشی از داده هایی است که در مرحله نمونه برداری انتخاب شده است. اگر یک مدل به خوبی طراحی شده باشد، می باید به خوبی بر روی این نمونه ها جوابگو باشد، دقیقاً همانند اطلاعاتی که در هنگام ایجاد مدل به کار گرفته شده و مدل به آنها پاسخ داده بود.

آنالیزورها می توانند مدل را در مقابل اطلاعات شناخته شده تست نمایند. برای مثال اگر کسی بداند کدام مشتری در پرونده خود بیشترین سهم نگهداری را داشته است و مدل میزان نگهداری را پیش بینی نماید، آنالیزورها می توانند پاسخ های مدل را بررسی نمایند تا مشخص شود که آیا این مشتریان به درستی انتخاب شده اند یا خیر.

۴-۶. تکرار

اگرچه ارزیابی مدل های داده ها آخرین مرحله در متدولوژی SEMMA است، اما ارزیابی در واقع آخرین مرحله از این متدولوژی نیست، چراکه SEMMA یک سیکل است که مراحل داخلی آن در اغلب مواقع، همان طور که در پروژه استخراج اطلاعات نیاز است تکرار می شوند.

۵. گزارش نتایج

گزارش قابلیت‌ها و توانمندی‌ها برای ایجاد داده های مناسب از طریق اجباری نمودن ارائه گزارش کاربران از انبار داده‌ها و سایر منابع اطلاعاتی میسر می‌شود.

خصوصیات اساسی سیستم گزارش دهی شامل موارد ذیل است:

- تکمیل دوره تکرار سیستم
- ساده سازی گزارش انبار اطلاعات
- کاربرد آسان از طریق استفاده از میانجی نگاره ای کاربر (GUI)
- انتشار سریع گزارشات
- توانمندی‌های اینترنتی

ابزارهای پیشرفته گزارش‌دهی مانند ابزاری که در SAS Enterprise Miner به آن اشاره شد، کاربران تجاری را قادر به ایجاد، انتشار و چاپ گزارش‌های چارچوب دار از اطلاعات ذخیره شده در انبار اطلاعات می‌نماید. با استفاده از میانجی نگاره ای کاربر (GUI)، کاربران توانایی ایجاد نمودارها، جداول، چارت‌ها و متون در یک گزارش و از طریق کامپیوتر خود را خواهند داشت.

۶. بهره برداری از نتایج در جهت کسب مزیت‌های تجاری

اطلاعات جدید به دست آمده از طریق استخراج اطلاعات می‌تواند با اطلاعات اجرایی و یا فرآیندهای آنالیز کننده آنلاین و سیستم گزارش دهی هماهنگ شده و سپس در سازمان با توجه به نیازمندی‌های موجود منتشر گردد. تصمیم گیران شرکت می‌توانند نتایج استخراج اطلاعات را به منظور پاسخگویی به سؤال های مرتبط تجاری مانند "چطور می‌توان برگشت حاصل از سرمایه گذاری در بخش طرح‌های بازاریابی را افزایش داد؟" و به منظور برنامه ریزی‌های استراتژیک و اقدامات مربوطه به کار گیرند. با بهره برداری از نتایج استخراج اطلاعات، شرکت‌ها می‌توانند خود را بهتر برای رشد و پیشرفت و بهبود در استفاده از موقعیت‌ها و ایجاد موفقیت‌های بلند مدت آماده کنند.

خلاصه

عامل اصلی در کسب مزیت تجاری در صنعت بیمه، تشخیص منابع اطلاعاتی مشتریان است که اگر به خوبی مدیریت، تجزیه و تحلیل و مورد بهره برداری قرار گیرد، خود به عنوان یک دارایی ارزشمند و

بی همتا برای یک سازمان قلمداد می شود. شرکتهای بیمه می توانند اطلاعات بسیار مفید موجود در منابع اطلاعاتی مشتریان خود را از طریق فناوری استخراج اطلاعات به دست آورند. سیستم استخراج اطلاعات می تواند با استفاده از مدل سازی پیش بینی کننده، تفکیک داده ها، آنالیز سبد بازار و ترکیبهای آنها، به سؤالهای اساسی تجاری با بیشترین دقت و صحت پاسخ گوید. محصولات جدید را می توان توسعه داد و استراتژیهای بازاریابی را می توان اجرا نمود تا شرکتهای بیمه قادر شوند ثروت نهفته در اطلاعات را به ثبات، ثروت قابل پیش بینی و سود تبدیل نمایند.

واژگان کلیدی:

داده کاوی، فناوری اطلاعات، استخراج اطلاعات، بیمه

منبع:

" Data Mining in the Insurance Industry ", Sase- Intelligence , 2000

