

پیاده‌سازی سیستم هوشمند برای پیش‌بینی مشتریان پر ریسک بیمه بدنه اتومبیل

آیدا رشید مهرآبادی^۱

الهام رشید مهرآبادی^۲

چکیده

مدیریت ریسک در صنعت بیمه یکی از مباحث چالش‌برانگیز است، چراکه مدیریت ریسک صحیح می‌تواند هزینه‌های شرکت‌های بیمه‌گر را کاهش و درآمدهای آنها را افزایش دهد. از آنجایی که کسب‌وکار شرکت‌های صنعت بیمه، قبول ریسک است؛ آن‌ها برای تأمین امنیت سرمایه خود و کاهش خطرات بالقوه به دنبال راهکارهای مناسب و جدید هستند، از این رو، تلاش برای به‌کارگیری ابزارهای فناوری اطلاعات در راستای تحقق مدیریت ریسک بهینه بسیار حائز اهمیت است. در پژوهش پیش‌رو با توجه به ضرورت به‌کارگیری فناوری اطلاعات، سیستم‌های یادگیری ماشین در هوش مصنوعی به‌منظور مدیریت ریسک هوشمند محور معرفی می‌شود.

این پژوهش ضمن بررسی ارتباط هوش مصنوعی با مدیریت ریسک، با پیاده‌سازی درخت تصمیم به‌عنوان یک ماشین یادگیر هوشمند، شرکت‌های بیمه را با مدیریت ریسک هوشمند محور آشنا می‌سازد. هدف سیستم هوشمند مدل‌سازی شده در این پژوهش، دسته‌بندی بیمه‌گذاران به لحاظ ریسک داشتن خسارت یا عدم خسارت در رشته بیمه بدنه اتومبیل است، تا درک بهتری از چگونگی استفاده از این علم جدید در مدیریت ریسک شرکت‌های صنعت بیمه ارائه شود.

واژگان کلیدی: مدیریت ریسک هوشمند، هوش مصنوعی، یادگیری ماشین، درخت تصمیم

۱. دانشجوی کارشناسی ارشد مدیریت بازرگانی، دانشگاه آزاد اسلامی واحد تهران مرکز، کارشناس بیمه شرکت بیمه نوین Aida.mehrabadi@hotmail.com

em.mehrabadi@gmail.com

۲. کارشناس ارشد هوش مصنوعی، دانشگاه آزاد اسلامی واحد علوم تحقیقات تهران

مقدمه

امروزه پیشرفت‌هایی که در زمینه فناوری اطلاعات و علوم مرتبط صورت گرفته، بسیاری از موضوعات را تحت تاثیر خود قرار داده است. از این رو سعی بر آن است که از برنامه‌های کامپیوتری علاوه بر ذخیره و بازیابی و بررسی داده‌ها، در پشتیبانی تصمیم‌گیری هم استفاده شود که یکی از زمینه‌های تحقیقاتی در این راستا "هوش مصنوعی" است. کاربرد ابزارها و الگوریتم‌های هوش مصنوعی در زمینه‌های مختلف رو به افزایش است و یکی از این زمینه‌ها حوزه "کسب‌وکار" می‌باشد. از دلایلی که کسب‌وکار به هوش مصنوعی علاقه دارد افزایش دانش و هوش سازمانی از طریق ایجاد تکنیک‌هایی برای حل مشکلات پیچیده و خاص است که انسان نمی‌تواند آن‌ها را در یک مقطع کوتاه زمانی تحلیل کند.

باتوجه به آنکه کسب‌وکار بیمه با فرض ریسک سروکار دارد و ریسک را نمی‌توان به‌طور کامل حذف کرد؛ بنابراین نگرش علمی به مسئله ریسک چیزی جز مدیریت آن نیست و می‌توان با به کارگیری تکنیک‌های هوش مصنوعی در راستای پیاده‌سازی مدیریت ریسک هوشمند محور گام برداشت.

هدف اصلی مقاله حاضر بررسی چگونگی استفاده از هوش مصنوعی و تکنیک‌های آن برای مدیریت ریسک در صنعت بیمه است. در ادامه و در بخش اول با مرور برخی تحقیقات مرتبط، سابقه این پژوهش را مورد بررسی قرار خواهیم داد، در بخش دوم مفاهیم مورد نیاز ادبیات موضوع پژوهش توضیح داده خواهد شد، در بخش سوم، یک درخت تصمیم به‌عنوان مثالی از ماشین یادگیرها^۲ در صنعت بیمه برای طبقه‌بندی مشتریان پریسک و کم‌ریسک، پیاده‌سازی می‌شود.

تاریخچه و مفاهیم مورد نیاز

تاریخچه

ضرورت مدیریت و کنترل بهینه ریسک در صنعت باعث شده است که مطالعه‌های گسترده‌ای در این باره صورت گیرد که در حال رشد و شکوفایی است. در ادامه برخی از تحقیق‌های صورت گرفته در این باره را بررسی خواهیم کرد: نورعلیزاده و همکاران در تحقیق مروری بر ابزارهای کاربردی در مدیریت ریسک؛ الزامات و کارکردها، ابتدا مراحل مدیریت ریسک را بررسی و سپس ابزارهای مناسب در حوزه مدیریت ریسک که منطبق با فرایند مدیریت ریسک هست را معرفی کرده‌اند (نورعلیزاده، ماجدی، ۱۳۹۱). زهرا ماجدی در تحقیق مدیریت ریسک در صنعت بیمه، به بررسی انواع ریسک‌ها در صنعت بیمه می‌پردازد (ماجدی، ۱۳۹۲). رشیدی در ترجمه تحقیق خود باعنوان از مدیریت ریسک تا مدیریت ریسک شرکتی، سه وجه ممیزه مدیریت ریسک شرکتی را به‌عنوان یک مفهوم جدید از مدیریت ریسک بررسی می‌کند (رشیدی). محسن قره‌خانی و همکاران در تحقیق کاربرد رویکرد کل‌نگر در تعیین حدود ریسک براساس مدیریت ریسک سازمان، از مدل عام تحلیل مالی پویا برای مدل‌سازی حدود ریسک فعلی یک شرکت بیمه تعاونی چند رشته فرضی استفاده کرده‌اند (قره‌خانی، ماجدی).

مدیریت ریسک

مدیریت ریسک یک فرایند پیشگیرانه است که سعی دارد در قبال استفاده از فرصت‌ها و پیشگیری از وقوع تهدیدها از منافع سازمان محافظت نماید، پس به‌طور کلی، مدیریت ریسک فرایند سنجش یا ارزیابی ریسک و سپس طرح

استراتژی‌هایی برای اداره ریسک است. مدیریت ریسک شامل پنج فعالیت اصلی است: شناسایی، ارزیابی، پایش و کاهش تاثیر ریسک‌ها بر یک کسب‌وکار (نیاکان). بنا به نظر بوهوم، مدیریت ریسک فرایندی دو مرحله‌ای است:

- تخمین ریسک: این مرحله شامل شناسایی، تحلیل و اولویت‌بندی است،
 - کنترل ریسک: این مرحله شامل برنامه‌ریزی مدیریت ریسک، برنامه‌ریزی نظارت ریسک و اقدامات اصلاحی است. از آنجاییکه هدف اصلی مدیریت ریسک، مدیریت نااطمینانی است؛ یک برنامه مدیریت ریسک صحیح با داشتن استراتژی‌های مناسب می‌تواند مشکلات هزینه‌بر را به حداقل رسانده و ادعای خسارت و حق بیمه را کاهش دهد (نیاکان).
- شرکت‌های صنعت بیمه با توجه به نوع فعالیت خود در معرض محدوده وسیعی از ریسک‌ها هستند. معمولاً ریسک‌ها در صنعت بیمه به پنج طبقه اصلی تقسیم می‌شوند. این طبقات ریسک عبارتند از ریسک راهبردی، ریسک عملیاتی، ریسک بیمه‌گری، ریسک بازار و ریسک اعتبار (ماجدی، ۱۳۹۲) در زیر هر طبقه ریسک به اختصار توضیح داده می‌شود:
- ریسک راهبردی: ریسک‌های راهبردی در نتیجه اجرای استراتژی‌ها پدید می‌آیند. این دسته از ریسک‌ها، مجموعه نیروهای بالقوه‌ای هستند که می‌توانند نظام استراتژی سازمان را بی‌تعادل کنند. چرا که ریسک‌های راهبردی ناشی از تصمیم‌گیری‌های نادرست و اجرای نامطلوب آنها، تغییرات صنعت و اثر آن بر تداوم کسب‌وکار و رتبه اعتباری است. مدیریت ریسک‌های راهبردی رویکردی از مدیریت ریسک است که حمایت از هدف‌های راهبردی سازمان را مورد توجه قرار داده و به دنبال آن است که با اتخاذ اقدام‌های پیش‌گیرانه زمینه‌های مساعد را برای دستیابی به این هدف‌ها پدید آورد (فرتوک زاده و الهی، ۱۳۸۹).

- ریسک عملیاتی: ریسک‌های عملیاتی به کنترل و اداره نامناسب کلی فعالیت‌های سازمان مانند ضعف در مدیریت و یا اختلال در سیستم کنترل‌های داخلی مربوط می‌شوند. (خوش سیما و شهیکی تاش، ۱۳۹۱) اگرچه، تعریف واحدی برای ریسک‌هایی که در این طبقه قرار می‌گیرند وجود ندارد، در بسیاری از سیستم‌های موجود، این ریسک را به‌عنوان مجموعه‌ای از، خطاهای انسانی، ریسک‌های ناشی از ناتوانی در کنترل و مدیریت و غیره در نظر گرفته می‌شود (ماجدی، ۱۳۹۲).

- ریسک بیمه‌گری: شرکت بیمه به علت صدور قراردادهای بیمه با ریسک‌هایی مواجه می‌شود این ریسک‌ها شامل خطرات تحت پوشش هر رشته‌ی خاص بیمه‌ای و همچنین شامل فرآیندهای خاص مربوط به مدیریت و هدایت کسب و کار می‌شود. برخی ریسک‌های که در این طبقه قرار می‌گیرند شامل ریسک فرآیند بیمه‌گری، ریسک قیمت‌گذاری، ریسک خسارات، ریسک ظرفیت نگهداری و ریسک تعیین ذخایر می‌شوند. لازم به ذکر است بیمه اتکایی در این طبقه مورد بررسی قرار نمی‌گیرد، این ریسک در طبقه ریسک اعتبار در نظر گرفته می‌شود (ماجدی، ۱۳۹۲).

- ریسک بازار: ریسک بازار به‌عنوان زیان‌های احتمالی ناشی از تغییر در فاکتورهای مؤثر بازار تعریف می‌شود (حبیبی ثمر ۱۳۹۴). این ریسک از تغییر سطوح یا نوسانات قیمت بازار بر دارایی‌ها به‌وجود می‌آید و شامل قرار گرفتن در معرض خطر تغییرات متغیرهای مالی نظیر نرخ‌های بهره، قیمت‌های اوراق قرضه، سهام و املاک و مستغلات و نرخ است (بینا، ۱۳۹۵).

- ریسک اعتبار: ریسک اعتبار، ریسک زیان مالی ناشی از نکول یا تغییر در صلاحیت اعتباری ناشرین اوراق بهادار، بدهکاران، یا طرفین معامله (به‌عنوان مثال، قراردادهای اتکایی و مشتقات و واسطه‌هایی است که شرکت با آنها طرف معامله است (ماجدی، ۱۳۹۲) (نوروزی، ۱۳۹۳).

مدیریت ریسک هوشمند محور

رویکرد مدیریت ریسک هوشمند محور، نشانگر تصمیم‌گیری مبتنی بر ریسک با به کارگیری امکانات فناوری اطلاعات است. از آنجاییکه یکی از پایه‌های موسسات بیمه‌ای مدرن، اطلاعات است؛ اگر موسسات بیمه بتوانند از داده‌های سیستم‌های اطلاعاتی موجود و یا داده‌های سازمان‌های خارجی و یا داده‌های بازار مرتبط با تجارت استفاده کنند، سرعت تصمیم‌گیری‌های سازمانی افزایش خواهد یافت. در این راستا سیستم‌های هوش مصنوعی، مانند ماشین یادگیری با عنوان ابزار کشف دانش و شناسایی الگوهای پنهان می‌توانند نقش مهمی در تحلیل ریسک ایفا کند، از این رو استفاده از این تکنیک به عنوان یک تکنولوژی قابل اتکاء در هوشمندسازی موسسات بیمه باید مورد توجه قرار گیرد.

هوش مصنوعی و یادگیری ماشین

بدیهی است که فرایند "تصمیم‌گیری" با "استدلال" رابطه مستقیم دارد و از آنجاییکه یکی از تعریف‌ها و مشخصات هوش مصنوعی، به فرایندهای شناختی و استدلال اشاره دارد و افراد نیز قبل از گرفتن هر تصمیمی استدلال می‌کنند پس طبیعی است که به دنبال اکتشاف ارتباطی بین تصمیم‌گیری و هوش مصنوعی باشیم (Pomerol, 1997). در ابتدا منظور از عبارت "هوش مصنوعی" بیان می‌گردد. اینکه هوش مصنوعی چیست و چه تعریفی می‌توان از آن بیان نمود، مبحثی است که تاکنون دانشمندان به یک تعریف جامع از آن نرسیده‌اند. حداقل دو دیدگاه راجع به هوش مصنوعی وجود دارد، اولین دیدگاه هوش مصنوعی را "علوم مصنوعی" می‌داند یا "علوم طراحی و ساخت محصولات مصنوعی مبتنی بر کامپیوتر که بتوانند وظایف مختلف انسانی را انجام دهند"، پذیرفتن این دیدگاه راجع به هوش مصنوعی این فایده را دارد که بسیاری از بحث‌های فلسفی مربوط به ذات هوش و امکان‌سنجی فعالیت‌های هوش مصنوعی کنار گذاشته می‌شود. این دیدگاه راجع به هوش مصنوعی ارتباط نسبتاً کمی با تصمیم‌گیری دارد و اعتقاد دارد که اتخاذ تصمیم توسط یک محصول مصنوعی نمی‌تواند به دلیل تصمیم‌گیری خود محصول باشد و اگر تصمیم‌گیری وجود داشته باشد توسط طراح سیستم است که در آن سیستم، مدل و تعبیه شده، یعنی اگر سیستم توانایی تشخیص و پی بردن به هر مجموعه داده ورودی را دارد به خاطر دستورات راه‌اندازی سیستم است؛ در واقع در این دیدگاه مفهوم "تصمیم" ناسازگار با مفهوم "برنامه‌ریزی" قرار خواهد گرفت؛ هنگامی که یک کار برنامه‌ریزی می‌شود دیگر تصمیم وجود ندارد و باتوجه به امکان وقوع هر یک از شرایط ممکن تعریف شده برای سیستم، اقدام مناسب توسط سیستم برنامه‌ریزی شده صورت می‌گیرد (Pomerol, 1997).

دومین دیدگاهی که راجع به هوش مصنوعی وجود دارد به جنبه "شناختی" آن توجه دارد و آن را اغلب "علوم بازنمایی دانش و استدلال" در نظر می‌گیرد، از این رو اگر هوش مصنوعی را به عنوان یک علم به منظور تقلید بشر فرض کنیم آنگاه رابطه تهنی بین هوش مصنوعی با تصمیم وجود خواهد داشت، اما در این دیدگاه نیز؛ هوش مصنوعی لزوماً علم طراحی و توسعه سیستم‌های تقلید از انسان نیست بلکه سوژه اصلی در آن انسان است و چون هوش مصنوعی توجه ویژه‌ای به مسایل تشخیص، نگه‌داری و بازنمایی دانش بشری دارد، به موجب آن تکنولوژی‌های مبتنی بر هوش مصنوعی توانایی نگه‌داری، بازیابی و انتقال داده و اطلاعات را به صورت موثرتر و سریعتر داشته و با بکارگیری داده‌های خام و تولید اطلاعات از آنها نقش بالقوه‌ای در گسترش دانش دارند (Pomerol, 1997).

یادگیری ماشین، به عنوان یکی از شاخه‌های وسیع و پرکاربرد هوش مصنوعی، به تنظیم و اکتشاف شیوه‌ها و الگوریتم‌هایی می‌پردازد که بر اساس آنها رایانه‌ها توانایی تعلم و یادگیری پیدا می‌کنند. در ادامه درخت تصمیم به عنوان یک ماشین یادگیر معرفی و مثالی از کاربرد آن در صنعت بیمه مدلسازی می‌شود.

درخت تصمیم

ساختار درخت تصمیم در یادگیری ماشین، یک مدل پیش‌بینی کننده می‌باشد که حقایق مشاهده شده در مورد یک پدیده را به استنتاج‌هایی در مورد مقدار هدف آن پدیده نقش می‌کند. تکنیک یادگیری ماشین برای استنتاج یک درخت تصمیم از داده‌ها، یادگیری درخت تصمیم نامیده می‌شود. یک درخت تصمیم ساختاری را نشان می‌دهد که برگ‌ها نشان دهنده دسته‌بندی و شاخه‌ها ترکیبات فصلی صفاتی که منتج به این دسته‌بندی‌ها را بازنمایی می‌کنند. یادگیری یک درخت می‌تواند با تفکیک کردن یک مجموعه منع به زیرمجموعه‌هایی براساس یک تست مقدار صفت انجام شود. این فرآیند به شکل بازگشتی در هر زیرمجموعه حاصل از تفکیک تکرار می‌شود. عمل بازگشت زمانی کامل می‌شود که تفکیک بیشتر سودمند نباشد یا بتوان یک دسته‌بندی را به همه نمونه‌های موجود در زیرمجموعه بدست آمده اعمال کرد. درختان تصمیم قادر به تولید توصیفات قابل درک برای انسان، از روابط موجود در یک مجموعه داده‌ای هستند و می‌توانند برای وظایف دسته‌بندی و پیش‌بینی بکار روند.

یادگیری درخت تصمیم یکی از رایج‌ترین تکنیک‌های یادگیری ماشین می‌باشد که به دلیل سادگی و کارآمدی باعث شده است علی‌رغم مشکلاتی که در استفاده از آن همچون صفات دارای نویز و یا صفات فاقد مقدار یا ... وجود دارد به شکل گسترده‌ای در مسائل مربوط به یادگیری ماشین استفاده شود (Chau, 2009 & Ngai, Xiu).

روش پیشنهادی: ارائه مدل درخت تصمیم برای پیش‌بینی مشتریان پر ریسک بیمه بدنه اتومبیل

به منظور مطالعه تجربی یکی از بزرگترین شرکت‌های بیمه، داده‌های مورد نیاز را فراهم کرده است.

جامعه آماری

جامعه آماری این پژوهش شامل داده‌های مربوط به بیمه‌نامه‌های بیمه‌گذاران بیمه بدنه اتومبیل است که بخشی از آنها دچار حادثه شده‌اند و خسارت دیده‌اند. حجم نمونه آماری متشکل از ۷۷۰۴۳۵ داده مربوط به بیمه‌نامه‌های بیمه بدنه اتومبیل است که طی تاریخ ۸۹/۱/۲ تا ۹۱/۳/۳۱ توسط شرکت صادر شده است که شامل بیمه‌نامه‌های خسارتی و غیرخسارتی است.

ابزار جمع آوری داده

داده‌های این مجموعه در یک رکورد بانک اطلاعاتی بیمه اتومبیل در سیستم اطلاعاتی شرکت مورد مطالعه ثبت بوده و از آن طریق و به کمک جستجوهای نسبتاً پیشرفته SQL اخذ شده است. اجرای این مرحله از فرایند مدلسازی که در گام پیش‌پردازش قرار می‌گیرد، شامل تبدیل فرمت داده‌ها و ساماندهی آن در قالب جداول اطلاعاتی قابل بازخوانی برای نرم افزار مدلسازی است. با طی این مرحله ۸ متغیر مجزا حاوی ویژگی‌هایی از بیمه‌گذار و کالای مورد بیمه در هر رکورد مجموعه داده شناسایی شدند.

آماده‌سازی داده شامل انتخاب، تمیز کردن و تبدیل داده‌ها

در بین این حجم داده، تمام داده‌ها از کیفیت لازم برخوردار نبودند و صلاحیت ورود به مدل نهایی را نداشتند، بدین صورت که در بعضی فیلدها مقادیر نامرتبط وارد شده بود. بر همین اساس تا حد امکان سعی شد تا رکوردهای معیوب اصلاح و یا حذف شود.

خصیصه‌های بیمه‌گذاران و کالای مورد بیمه (متغیرهای مستقل پژوهش) هستند؛ به منظور تعقیب هدف پژوهش ایجاد تغییراتی در ساختار هر رکورد مجموعه داده اخذ شده، ضروری بود در نتیجه با استفاده از جستجوهای نسبتاً پیشرفته SQL، کد "وضعیت خسارت مشتری" استخراج و به هر رکورد مجموعه داده به‌عنوان متغیر وابسته پروژ، الحاق شد. واضح است که این کد وضعیت سابقه خسارت بیمه‌گذار را در دوره خدمات نشان می‌دهد. عناوین این فیلدها در پایگاه داده نهایی و پالایش شده، شامل ۸ ویژگی (ورودی مدل) و ۱ فیلد طبقه مشتریان (خروجی مدل) است. داده‌های اولیه قبل از مدل‌سازی، با کمک نرم افزار Excel پیش پردازش شده‌اند تا از کیفیت قابل قبولی برخوردار باشند (جدول ۱ و ۲).

متغیرهای پیش‌بینی کننده

در این مطالعه با استفاده از رفتار مشاهده شده مشتری در گذشته و اطلاعات دموگرافی او تعداد ۸ متغیر به منظور پیش‌بینی ریسک خسارت ایجاد شده است. جدول ۱ متغیرهای پیش‌بینی کننده مورد استفاده در این مطالعه را نشان می‌دهد. در پاراگراف‌های بعدی به بررسی هر یک از این متغیرها پرداخته شده است.

متغیرهای دموگرافی: متغیرهای رفتاری مانند سن، جنسیت، موقعیت جغرافیایی و غیره به‌طور گسترده‌ای در ادبیات موضوع برای پیش‌بینی ریسک خسارت بیمه‌گذاران به کار گرفته شده‌اند. در مطالعه مورد بررسی، با توجه به اینکه مشتریان سازمان شامل مشتریان حقیقی و حقوقی می‌شوند، این ویژگی نیز به‌عنوان پیش‌بینی کننده دموگرافی مورد استفاده قرار گرفته است. شاخص پولی ۱: این مشخصه نشان‌دهنده مقدار پولی است که بیمه‌گذار در سازمان خرج کرده است. در این مطالعه ارزش مورد بیمه به‌عنوان شاخص پولی در نظر گرفته شده است.

سایر متغیرها: برخی ویژگی‌های محصول فرد بیمه‌گذار نیز برای آزمایش اینکه آیا در ریسک خسارت او تاثیر دارد با خیر به‌عنوان ورودی مدل در نظر گرفته شده است.

جدول (۱) - متغیرهای ورودی سیستم هوشمند

نام	ویژگی	نوع
نوع بیمه گذار	حقیقی: ۱ حقوقی: ۲	flag
جنسیت بیمه گذار	مرد: ۱ زن: ۲	Flag
تحصیلات بیمه گذار	زیر دیپلم، دیپلم، بالای دیپلم	set
نوع پلاک	دولتی: ۱ سیاسی: ۲ شخصی: ۳ عمومی: ۴ فاقد پلاک: ۵ موقت: ۶ نظامی: ۷ یدک: ۸	set
نوع تخفیف ها	خسارت دارد، جدید، تخفیف ندارد: ۰ تخفیف عدم خسارت سال اول: ۱ تخفیف عدم خسارت سال دوم: ۲ تخفیف عدم خسارت سال سوم: ۳ تخفیف عدم خسارت سال چهارم: ۴ تخفیف عدم خسارت سال پنجم: ۵	Set
تعداد سیلندر	3,4,6,8	set
نرمال ارزش مورد بیمه	[0 1]	Rang
کاربری اتومبیل	آمبولانس: ۱ آژانس: ۲ اشتعال زا و سوختی: ۳ حمل زباله: ۴ کرایه: ۵ سرویس کارکنان بیمه گذار: ۶ شخصی: ۷ تعلیم رانندگی: ۸ تاکسی: ۹ یدک: ۱۰	set

- از آنجایی که ساختار پایگاه داده شامل داده‌های عددی و غیر عددی به صورت توأم است؛ ابتدا کلیه داده‌های غیر عددی را با کدگذاری به داده‌های عددی تبدیل نمودیم. در جدول ۱ عددهایی که در مقابل برخی ویژگی‌ها قرار گرفته است؛ کدگذاری مورد استفاده را نشان می‌دهد.

- برای همگن کردن داده‌ها برای فیلد "ارزش مورد بیمه" از فرمول (۱) به جهت نرمال کردن اعداد بین ۰ و ۱ استفاده نمودیم:

$$X = (X - X_{min}) / (X_{max} - X_{min}) \quad (1)$$

- باتوجه با اینکه در الگوریتم مورد استفاده لازم است سن اتومبیل وارد شود ابتدا سال‌های ساخت اتومبیل‌ها به سال شمسی تبدیل شده و سپس سن هر اتومبیل محاسبه شده است.

جدول (۲) - متغیر خروجی سیستم هوشمند

نام	ویژگی	نوع
وضعیت مشتری	خسارت ندارد: ۰ خسارت دارد: ۱	Flag

در جدول ۲؛ از آنجایی که در این مرحله از پروژه تصمیم بر این است که یک طبقه‌بندی به لحاظ تشخیص ریسک خسارت و یا عدم خسارت بیمه‌گذار صورت گیرد در نتیجه در فیلد طبقه مشتری، برچسب "۱" و برچسب "۰" وجود دارد. در الگوریتم پیاده‌سازی شده در این پژوهش از ۶۰٪ داده‌ها به‌عنوان داده‌های آموزش و از ۴۰٪ از آن‌ها به‌عنوان داده‌های آزمایش الگوریتم استفاده می‌شود.

قبل از مدلسازی، با کمک تکنیک‌های نمونه‌گیری؛ ۶۵۵۳۶ از جامعه آماری برای ورود به مدل آماده شد.

مدلسازی

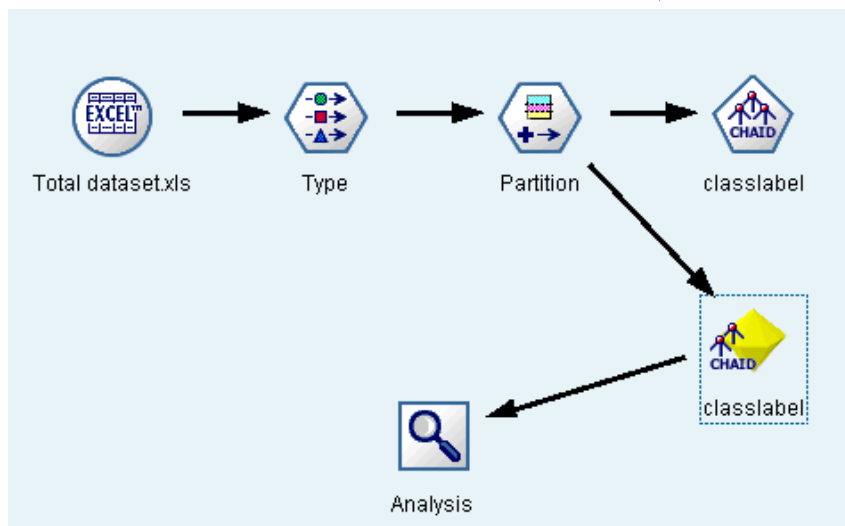
از آنجایی که هدف نهایی این پژوهش نگاشت مشتریان به دو گروه "پرریسک و کم ریسک" است؛ کارکرد متناسب با این هدف "دسته‌بندی" خواهد بود. کارکرد دسته‌بندی که از جمله کارکردهای دارای ماهیت پیش‌بینی‌گری داده‌کاوی است، به ساختن مدل‌هایی برای پیشگویی کلاس یک شی براساس ویژگی‌ها یا صفات آن معطوف می‌شود. واژه‌های عمومی شی و کلاس به تناسب موضوع پژوهش تفسیر می‌شود. در این مطالعه منظور از شی "بیمه‌گذار" و منظور از کلاس رفتار دوجهی "خسارت و عدم خسارت" است.

در ایجاد یک مدل دسته‌بندی، مراحل انتخاب تکنیک دسته‌بندی، انتخاب مجموعه داده مدل، تعیین متغیر هدف و معرفی صفات؛ طی می‌شود.

ابزار مدلسازی: ابزاری که برای پیاده‌سازی این الگوریتم از پروژه مورد استفاده قرار گرفت، IBM SPSS Modeler است. تا به امروز نرم افزارهای تجاری و آموزشی فراوانی برای مدلسازی در حوزه‌های مختلف داده‌ها به دنیای علم و فناوری عرضه شده‌اند. هر یک از آنها باتوجه به نوع اصلی داده‌هایی که مورد کاوش قرار می‌دهند، روی الگوریتم‌های خاصی متمرکز شده‌اند. این نرم‌افزار به گونه‌ای طراحی شده است که می‌توان به سرعت، روش‌های موجود را به صورت انعطاف‌پذیری روی مجموعه‌های جدید داده، آزمایش کرد.

پیاده سازی مدل

ابتدا ورودی یا خروجی بودن داده‌ها در نرم افزار مشخص گردید. سپس ۶۰٪ از این داده‌ها برای آموزش الگوریتم درخت تصمیم و ۴۰٪ از داده‌ها برای آزمایش صحت الگوریتم مورد استفاده قرار گرفت و بعد الگوریتم پیاده‌سازی شده است؛ شکل ۱ ساخت مدل درخت تصمیم chaid را نشان می‌دهد.



شکل (۱) - سیستم هوشمند مدل‌سازی شده

درخت تصمیم پیاده‌سازی شده با نرم افزار IBM SPSS Modeler به ما این توانایی را می‌دهد که پیش‌بینی‌های خود را در قالب یکسری قوانین ارائه دهیم؛ همچنین به ما نشان می‌دهد که کدام فیلد یا متغیرها تأثیرات مهمی در پیش‌بینی ما دارند.

ارزیابی مدل

هنگام استفاده از کارکردهایی مثل دسته‌بندی که ماهیت پیش‌بینی دارند، ارزیابی مدل‌های استخراج شده ضروری است. در این تحقیق ارزیابی مدل‌ها با استفاده از ماتریس انطباقی (جدول شماره ۳) صورت گرفته است. در این روش کل مجموعه داده مدل به دو قسمت داده آموزش و داده آزمایش یا آزمون تقسیم می‌شود. نسبت این تقسیم در این مطالعه با توجه به مطالعات پیشین ۶۰ به ۴۰ در نظر گرفته شده است. در روش ارزیابی ماتریس انطباقی، مجموعه داده آزمون به‌عنوان داده‌های ورودی، به خروجی مدل حاصل از داده آموزش، وارد شده و ضریب صحت مدل از فرمول‌های زیر محاسبه می‌شود:

$$\text{Accuracy Rate} = \frac{N1+N2}{N1+N2+M1+M2} \quad (2)$$

$$\text{Error Rate} = \frac{M1+M2}{N1+N2+M1+M2} \quad (3)$$

که در آن:

N1: تعداد مشاهداتی از مجموعه داده آزمون که در کلاس ۱ قرار دارند و مدل هم کلاس آنها را به درستی پیش‌بینی کرده است.

N2: تعداد مشاهداتی از مجموعه داده آزمون که در کلاس ۲ قرار دارند و مدل هم کلاس آنها را به درستی پیش‌بینی کرده است.

M1: تعداد مشاهداتی از مجموعه داده آزمون که در کلاس ۱ قرار دارند و مدل کلاس آنها را به غلط ۲ پیش‌بینی کرده است.

M2: تعداد مشاهداتی از مجموعه داده آزمون که در کلاس ۲ قرار دارند و مدل کلاس آنها را به غلط ۱ پیش‌بینی کرده است.

جدول (۳) ماتریس ارزیابی مدل

پیش بینی مدل			
کلاس ۲	کلاس ۱		
M1	N1	کلاس ۱	موقعیت واقعی مشاهده
N2	M2	کلاس ۲	

نتایج و تحلیل یافته‌ها

خروجی فرایند یادگیری ماشین در پیش‌بینی ریسک بیمه‌گذاران بیمه بدنه اتومبیل در این تحقیق، شامل استخراج صفات موثر در پیش‌بینی ارزش متغیر هدف، تعیین ضریب اثر نسبی هر یک در این پیش‌بینی و ماتریس ارزیابی ضریب صحت پیش‌بینی مدل می‌شود. جدول ۴ و ۵ این خروجی‌ها را نشان می‌دهد.

جدول ۱- ضرایب اهمیت در پیش‌بینی تغییرات متغیر هدف

ضریب اهمیت در پیش‌بینی تغییرات متغیر هدف	صفت
* ۰/۵۹	ارزش مورد بیمه
* ۰/۲۴	مشخصه فنی خودرو(سیلندر)
* ۰/۱۵	کاربری خودرو
* ۰/۰۰۴	جنسیت راننده
- ۰	نوع پلاک
- ۰	نوع مشتری
- ۰	تحصیلات راننده

جدول (۵) نتیجه ارزیابی سیستم هوشمند پیاده سازی شده

نتیجه پیش‌بینی	تعداد مشاهدات (داده آموزش)	درصد	تعداد مشاهدات (داده آزمون)	درصد
درست	۲۶۸۸۵	۶۸/۴۹	۱۸۰۱۱	۶۸/۵۳
نادرست	۱۲۳۶۲	۳۱/۵۱	۸۲۷۲	۳۱/۴۷
جمع	۳۹۲۵۲		۲۶۲۸۳	

در جدول بالا دیده می‌شود که مدل از میان هفت صفت پیشگوی پیش فرض، چهار صفت نوع پلاک، نوع مشتری، تحصیلات راننده را بی اهمیت تشخیص داده است.

• بررسی اولین خروجی مدل (جدول ۴)، اولویت نسبی هفت عامل پیش فرض را در تشخیص ریسک بیمه‌گذاران نشان می‌دهد. مشاهده می‌شود که "ارزش مورد بیمه" عامل اصلی پیش‌بینی رفتار پریسک است. متغیرهای مشخصه فنی خودرو و کاربری خودرو در مراتب بعد و متغیر جنسیت راننده نیز با تاثیر نه چندان قابل توجه پیش‌بینی کننده هستند.

- وارد نشدن سه متغیر دیگر یعنی نوع مشتری، نوع پلاک و تحصیلات راننده به مدل نشان می‌دهد که فراوانی رفتار پرریسک در گروه‌های مختلف مشتریان در این سه متغیر تفاوت معناداری با یکدیگر ندارند.
- ذکر این نکته لازم است که کاربرد این نتایج در فرایند تصمیم‌گیری مدیران باید با توجه به میزان مطلوبیت ضریب صحت به دست آمده برای مدل در ماتریس ارزیابی آن (۶۸٪)، در نگاه مدیران این سازمان صورت بگیرد.

نتیجه‌گیری

با گسترش استفاده روز افزون از ابزارهای فناوری اطلاعات و ارتباطات در صنعت بورس، پیش‌بینی‌های مالی و حسابداری، تجارت الکترونیک و پزشکی لازم است که این ابزار در مدیریت ریسک سازمان‌ها و صنایع مختلف نقش آفرینی کند، از جمله در صنعتی مانند بیمه که به پیش‌بینی بازار، مشتری، مواد اولیه و ریسک و خسارت نیاز دارند. استفاده از روش‌های آماری و گزارش‌گیری به تنهایی اطلاعات لازم را در اختیار مدیران قرار نمی‌دهند تا بهترین تصمیم را بگیرند. در جهان پر از تغییر امروزی، لازم است تا از ابزارهای نوین برای کشف دانش استفاده کرد. با اختصاص بودجه‌ای برای نوین‌سازی کشف دانش و اطلاعات می‌توان بازگشت سرمایه قابل توجهی برای شرکتهای صنعت بیمه ایجاد کرد. امروزه بقای سازمان‌ها در گرو هوشمندی آنهاست و یادگیری ماشین در هوش مصنوعی یکی از ابزارهای کسب این هوشمندی در دنیای در حال تغییر است.

منابع

۱. حبیبی ثمر، ج.، تهرانی، ر. & انصاری، ک.، ۱۳۹۴. بررسی رابطه بین ریسک نقدشوندگی و ریسک بازار با بازده سهام رشدی و ارزشی در بورس اوراق بهادار تهران. *AHP مجله مهندسی مالی و مدیریت اوراق بهادار*, جلد ۲۳, pp. 39-58.
۲. ۴. بینا، م.، ۱۳۹۵. گزارشگری ریسک بازار در بانک های ایرانی. سی و هشتمین همایش انجمن حسابداران خبره ایران.
۳. ۵. خوش سیما، ر. & شهیکی تاش، م. ن.، ۱۳۹۱. تاثیر ریسک های اعتباری، عملیاتی و نقدینگی بر کارایی نظام بانکی ایران. فصلنامه علمی پژوهشی برنامه ریزی و بودجه، جلد ۴, pp. 69-95.
۴. ۶. رشیدی، ر.، بدون تاریخ از مدیریت ریسک تا مدیریت ریسک شرکتی. تازه های جهان بیمه، جلد ۱۴۶ و ۱۴۷, pp. 27-36.
۵. ۷. فرتوک زاده، ح. ر. & الهی، م.، ۱۳۸۹. شناسایی و اولویت بندی ریسک های راهبردی در بانکداری بدون ربا. فصلنامه علمی پژوهشی اقتصاد اسلامی، جلد ۲۷, pp. 147-172.
۶. ۸. قره خانی، م. & ماجدی، ز.، بدون تاریخ کاربرد رویکرد کل نگر در تعیین حدود ریسک براساس مدیریت ریسک سازمان. جلد ۱۷۴, pp. 42-48.
۷. ۹. ماجدی، ز.، ۱۳۹۲. مدیریت ریسک در صنعت بیمه. بینش بیمه، نوآوری، شکوفایی، جلد ۴۹, pp. 1-9.
۸. ۱۰. نورعلیزاده، ح. & ماجدی، ز.، ۱۳۹۱. مروری بر ابزارهای کاربردی در مدیریت ریسک؛ الزامات و کارکردها. بینش, pp. 1-29.
۹. ۱۱. نوروزی، پ.، ۱۳۹۳. تأثیر متغیرهای کلان بر ریسک اعتباری بانک ها در ایران. فصلنامه پژوهش های پولی - بانکی، جلد ۲۰, pp. 237-257.
۱۰. ۱۲. نیاکان، ل.، مدیریت ریسک چیست؟. تازه های جهان بیمه، جلد ۱۸۶, pp. 16-26.
11. Ngai, E., Xiu, L. & Chau, D., 2009. Application of data mining techniques in customer relationship management: A literature review and classification. *Expert Systems with Applications*, p. 2592-2602.
12. Pomerol, J.-C., 1997. . Artificial intelligence and human decision making. *European Journal of Operational Research*, pp. 3-25.