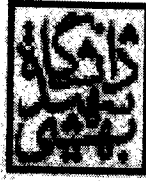


بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



دانشگاه شهید بهشتی
دانشکده مدیریت و حسابداری
جهت اخذ درجه کارشناسی ارشد مدیریت فناوری اطلاعات

عنوان پایان نامه:

طبقه بندی ریسک بیمه گذاران بیمه بدنه اتومبیل با استفاده از
داده کاوی

استاد راهنما

دکتر امین عمید

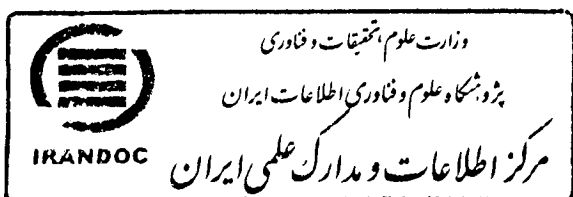
استاد مشاور

پروفسور غلامرضا نخعی زاده

نگارش

الهام عنبری

مهر ۸۹



۱۴۹۲۵۴

۱۳۸۹/۱۰/۱۹

تحت حمایت و همکاری

پژوهشکده بیمه

تقدیم به :

مادر عزیز و مهربانم

پدر خوب و صبورم

دو برادر دوست داشتیم

و خاله مهربانم بابت تمام حمایت‌هایش

سپاسگذاری:

از اساتید گرامی ام ، جناب آقای دکتر امین عمید و پروفیسور غلامرضا نخعی زاده که همواره مخلصانه در جهت به ثمر رسیدن این پایان نامه کمک رسان بنده بودند ، جناب آقای مهدی حائری مسئول پایگاه اطلاعات بیمه خودرو بیمه مرکزی ایران که همواره با کمال میل و رویی گشاده داده های مورد نیاز این پژوهش را در اختیارم می گذاردند ، دوستان ارجمندم جناب آقای احمد نادعلی و زهرا سهامی بابت حمایت های ارزنده و خانواده بسیار خوبم که همواره مشوق من بودند کمال تشکر را می نمایم.

چکیده:

شرکتهای بیمه ایرانی هم اکنون جهت تطبیق خود با محیط جدیدی که نظام آزادانه تعیین حق بیمه (نظام غیر تعرفه ای) برای آنها به وجود آورده است ، می بایستی ابزارها و امکانات فنی خود را ارتقا بخشند تا بتوانند در رقابت کاملی که به تبع تغییر این محیط برایشان به وجود می آید ، ادامه حیات دهند. یکی از کاربردهای داده کاوی می تواند کمک به شرکتهای برای قیمت گذاری محصولات و خدمات باشد ، که پایان نامه حاضر نیز از این کاربرد داده کاوی استفاده می کند. در این پایان نامه با استفاده از داده های مربوط به بیمه نامه بدنه از کل شرکتهای بیمه (بانک اطلاعاتی بیمه خودرو) سعی بر آن است تا بررسی شود که آیا میتوان بیمه گذاران بیمه بدنه اتومبیل را از نظر ریسک طبقه بندی کرد؟ و آیا درخت تصمیم برای طبقه بندی بیمه گذاران بهترین ابزار طبقه بندی می باشد؟ و آیا سن و جنسیت از موثرترین عوامل در ریسک بیمه گذار محسوب می شود؟ . نتایج این طبقه بندی به صورت درخت تصمیم و قوانین نشان داده شده است . و نتایج حاصل از صحت مدل درخت تصمیم با نتایج الگوریتم های شبکه عصبی و رگرسیون لجستیک مورد مقایسه قرار گرفته است.

واژه های کلیدی: داده کاوی ، طبقه بندی ، درخت تصمیم

فهرست مطالب

فصل ۱.....	ک
کلیات تحقیق.....	۱
۱-۱ بیان مسأله تحقیق.....	۲
۱-۱-۱ تشریح و معرفی مسأله.....	۲
۱-۱-۲ جنبه های مجهول و مبهم.....	۳
۱-۱-۳ متغیرهای مسأله.....	۳
۱-۱-۴ منظور تحقیق.....	۴
۱-۲ اهمیت و ضرورت انجام تحقیق.....	۴
۱-۲-۱ خلاء های تحقیقاتی.....	۴
۱-۲-۱-۳ فواید احتمالی نظری و عملی تحقیق.....	۵
۱-۲-۱-۴ جنبه جدید بودن و نوآوری در تحقیق.....	۵
۱-۳ اهداف تحقیق.....	۵
۱-۳-۱ هدف آرمانی تحقیق.....	۶
۱-۳-۲ هدف کلی تحقیق.....	۶
۱-۳-۳ اهداف ویژه تحقیق.....	۶
۱-۳-۴ اهداف کاربردی تحقیق.....	۶
۱-۴ سؤالات تحقیق.....	۷
۱-۴-۱ سؤال اصلی تحقیق.....	۷
۱-۴-۲ سؤالات فرعی تحقیق.....	۷
۱-۵ فرضیه تحقیق.....	۸
مدل مفهومی که این پژوهش قصد یافتن آن را دارد ، مجموعه ای از الگوهایی است که پیش بینی ریسک بیمه گذاران بیمه اتومبیل را ممکن می سازد.....	۸
۶-۱ قلمرو تحقیق (زمانی، مکانی و موضوعی).....	۸
۷-۱ روش شناسی تحقیق.....	۹
۱-۷-۱ شرح روش تحقیق.....	۹

۱۰	۲-۷-۱ مدل مفهومی متغیرهای مورد بررسی
۱۱	۳-۷-۱ مدل اجرایی
۱۳	۴-۷-۱ شرح روش و ابزار گردآوری داده‌ها
۱۳	۵-۷-۱ جامعه آماری، روش نمونه‌گیری و حجم نمونه
۱۳	۶-۷-۱ روش‌ها و ابزار تجزیه و تحلیل داده‌ها
۱۴	۸-۱ محدودیت‌های تحقیق
۱۵	۹-۱ تعریف واژه‌ها و اصطلاحات فنی و تخصصی (تعاریف مفهومی و عملیاتی)
۱۵	۱-۹-۱ قراردادها
۱۶	۱۰-۱ جمع بندی
۱۷	فصل ۲
۱۷	مروری بر ادبیات و پیشینه تحقیق
۱۸	مقدمه
۱۸	۱-۲ اصول و مفاهیم ریسک در بیمه
۱۸	۱-۱-۲ مشخصات ریسک های قابل بیمه شدن
۱۹	۱-۲-۲ انواع ریسک در صنعت بیمه
۲۰	۲-۲ داده کاوی و کشف دانش از پایگاه داده
۲۱	۱-۲-۲ مفهوم داده کاوی
۲۲	۲-۲-۲ فرایند کلی داده کاوی
۲۵	۳-۲-۲ روشها و تکنیک های داده‌کاوی
۲۷	۳-۲ طبقه بندی
۲۸	۱-۳-۲ الگوریتم های مختلف طبقه بندی
۲۹	۲-۳-۲ یادگیری ماشین و طبقه بندی
۲۹	۴-۲ درخت تصمیم
۳۲	۱-۴-۲ استخراج قاعده از یک درخت تصمیم
۳۲	۲-۴-۲ درخت تصمیم و انواع الگوریتم های آن
۳۳	۳-۴-۲ نقاط قوت و ضعف الگوریتم های درخت تصمیم
۴۰	۶-۲ مروری بر تحقیقات پیشین

۴۱ ۱-۶-۲ مطالعات داخلی
۴۴ ۲-۶-۲ مطالعات خارجی
۴۶ ۷-۲ جمع بندی
۴۷ فصل ۳
۴۷ روش شناسی تحقیق
۴۸ مقدمه
۴۸ ۲-۳ الگوریتم ها و مدل های مورد استفاده در تحقیق
۴۹ ۱-۲-۳ مدل فرایندی داده کاوی بر اساس استاندارد CRISP-DM
۵۰ شکل ۱-۳ مراحل مدل فرایندی داده کاوی بر اساس استاندارد [۲۹] CRISP-DM
۵۲ ۲-۲-۳ داده های تحقیق
۵۷ ۳-۲-۳ شیوه و ابزار جمع آوری داده ها
۵۷ ۴-۲-۳ نوع داده ها و مقیاس آنها
۵۸ ۳-۳ اعتبار و کارایی روش ارائه شده
۵۹ ۴-۳ ساختار اجرایی تحقیق
۶۰ ۱-۴-۳ درک مساله کسب و کار
۶۱ ۲-۴-۳ درک داده ها
۶۲ ۳-۴-۳ آماده سازی داده ها
۶۵ ۴-۴-۳ مدلسازی داده ها
۶۶ ۵-۴-۳ ارزیابی نتایج
۶۶ ۶-۴-۳ بکارگیری مدل
۶۷ ۵-۳ جمع بندی
۶۹ فصل ۴
۶۹ تجزیه تحلیل
۶۹ داده های تحقیق

۷۰ مقدمه
۷۸ فصل ۵
۷۸ نتیجه گیری و پیشنهادات
۷۹ مقدمه
۷۹ ۱-۵ نتایج و یافته های تحقیق
۷۹ ۱-۱-۵ پاسخ به سوالهای تحقیق
۸۰ ۲-۱-۵ دستاوردهای پژوهش
۸۹ ضمائم
۹۰ قوانین تولید شده
۱۱۱ چکیده انگلیسی

فهرست جداول:

جدول ۱-۳ روش تحقیق	۴۸
جدول ۲-۳ فیله‌های مورد استفاده در مدل نهایی	۵۶
جدول ۱-۴ نتایج مدل درخت تصمیم C5.0	۷۵
جدول ۳-۴ نتایج حاصل از صحت طبقه بندی مدل NEURAL NETWORK	۷۵
جدول ۴-۴ نتایج حاصل از صحت طبقه بندی مدل LOGISTIC REGRESION	۷۶
جدول ۵-۴ نتایج حاصل از الگوریتم FEATURE SELLECTION	۷۷
جدول ۱-۵ سهم استانهای مختلف در میزان ریسک بیمه گذاران	۸۱
جدول ۲-۵ مقایسه های یافته های این پژوهش با یافته های پژوهش دیگر	۸۳

فهرست اشکال:

- شکل ۱-۱ مدل مفهومی متغیرهای مورد بررسی تحقیق ۱۰
- شکل ۱-۲ مدل اجرایی تحقیق ۱۲
- شکل ۱-۲ گامهای اصلی فرایند اکتشاف دانش از پایگاه داده ۲۵
- شکل ۲-۲ مدل‌های شبکه عصبی ۳۷
- شکل ۲-۳ کاربردهای داده کاوی در بیمه ۳۹
- شکل ۱-۳ مراحل مدل فرایندی داده کاوی بر اساس استاندارد CRISP-DM ۵۰
- شکل ۱-۴ درخت تصمیم حاصل شده ۷۱

فصل ۱

کلیات تحقیق

۱-۱ بیان مسأله تحقیق

این بخش شامل تشریح مسأله و معرفی آن، بیان جنبه‌های مجهول و مبهم، بیان متغیرهای مربوطه و منظور از تحقیق می باشد؛ که در ادامه تشریح می گردد.

۱-۱-۱ تشریح و معرفی مسأله

صنعت بیمه در هر کشوری به عنوان یکی از نهادهای مالی تاثیر گذار محسوب می شود که نقش موثری در کاهش ریسک در ابعاد مختلف اقتصادی جامعه دارد. از اینرو شرکتهای بیمه با قبول ریسک باعث ایجاد آرامش در جامعه می گردند. به همین خاطر لازم است که شرکتهای بیمه به ابزارهای تحلیل ریسک قدرتمندی دسترسی داشته باشند تا بتوانند ریسک دریافتی را بخوبی مدیریت کنند. صنعت بیمه برای تحلیل ریسک وابسته به تجارب خسارتی می باشد. به همین خاطر تکنیک های داده کاوی به علت متکی بودن به داده می تواند از کاربرد زیادی در صنعت بیمه برخوردار باشند.

به دلیل نفوذ زیاد بیمه های اتومبیل و نقشی که این نوع بیمه در بین دیگر رشته های بیمه در ایران دارد^۱، بیمه بدنه اتومبیل برای این مطالعه انتخاب شد. ولی متأسفانه بکارگیری ابزارهای جدیدی مانند داده کاوی در صنعت بیمه ایران تا کنون جایگاه خود را پیدا نکرده است. اما به علت تغییر محیط به لحاظ قانونی (تبدیل نظام تعرفه ای حق بیمه به نظام نرخ گذاری آزاد) استفاده از فناوری های جدید برای رقابت در یک بازار کاملاً رقابتی از سوی شرکتهای بیمه به

^۱ طبق آمار سال ۸۸ بیمه مرکزی ج.ا. ایران سهم بیمه بدنه اتومبیل از نظر میزان حق بیمه ۱۱،۱۵ درصد است که در بین سایر رشته های بیمه ای جایگاه سوم را داراست

نظر اجتناب ناپذیر می باشد. در حال حاضر از ابتدا تمامی بیمه گذاران حق بیمه واحدی پرداخت می کنند که این باعث می شود که بیمه گذاران با ریسک پایین تر حق بیمه بیشتری نسبت به بیمه گذاران با ریسک بالا پرداخت کنند که به نظر منصفانه نمی رسد. با دسترسی به ابزار های تحلیل و طبقه بندی ریسک می توان این مشکل را برطرف ساخت. با استفاده از داده کاوی و تحلیل بر روی داده های تجربی به جامانده از بیمه گذاران ، می توان این بیمه گذاران را به لحاظ ریسک طبقه بندی کرد. با استفاده از این طبقه بندی می توان طبقه بیمه گذاران آتی را پیش بینی کرد و در میزان حق بیمه دریافتی از بیمه گذاران مختلف تعدیل ایجاد کرد و با ایجاد سیستم نرخ گذاری مبتنی بر ریسک بیمه گذاران ، میزان رضایت بیمه گذاران را افزایش داد و از طرفی بر سودآوری شرکتهای بیمه تاثیری مثبت گذاشت.

۲-۱-۱ جنبه های مجهول و مبهم

- ۱- داده کاوی ابزار مناسبی برای طبقه بندی بیمه گذاران بیمه بدنه اتومبیل است یا خیر؟
- ۲- در صورتی که داده کاوی ابزار مناسبی برای طبقه بندی بیمه گذاران بیمه بدنه اتومبیل باشد، آیا تکنیک درخت تصمیم برای این امر مناسب خواهد بود؟

۳-۱-۱ متغیرهای مسئله

متغیر مستقل: " عوامل تاثیر گذار بر رفتار بیمه گذاران بیمه بدنه اتومبیل " که از بررسی پایگاه داده بیمه مرکزی استخراج می گردد.

متغیر وابسته : طبقه های بیمه گذاران از نظر ریسک(ریسک بالا، متوسط ، پایین)

۱-۱-۴ منظور تحقیق

۱. طبقه بندی بیمه گذاران بیمه بدنه اتومبیل به ریسک بالا ، متوسط ، پایین در صنعت بیمه اتومبیل کشور بر اساس ویژگی های شخصی و مشخصات خودرو.
۲. پیش بینی وضعیت بیمه گذاران آینده بیمه بدنه اتومبیل با استفاده از شناسایی عوامل تاثیرگذار بر ریسک.
۳. کمک به افزایش سودآوری شرکتهای بیمه از طریق ارائه طبقه بندی مناسب از ریسک بیمه گذاران بیمه بدنه اتومبیل.
۴. شناسایی عوامل تاثیر گذار بر رفتار بیمه گذاران بیمه بدنه اتومبیل با توجه پایگاه داده های موجود.

۱-۲ اهمیت و ضرورت انجام تحقیق

این بخش شامل خلاءهای تحقیقاتی موجود، فواید احتمالی نظری و عملی آن ، همچنین روش و یا فرآیند تحقیقی جدیدی که در این تحقیق مورد استفاده قرار می گیرد، می باشد که در ادامه بیان شده است.

۱-۲-۲ خلاء های تحقیقاتی

با بررسی تحقیقات گذشته ، تحقیق داخلی با این موضوع خاص یعنی طبقه بندی ریسک در صنعت بیمه انجام نشده است .

۱-۲-۳ فواید احتمالی نظری و عملی تحقیق

- ۱- افزایش سودآوری صنعت بیمه در زمینه طبقه بندی بیمه گذاران برای تعیین حق بیمه.
- ۲- کاهش خطاهای انسانی و جلوگیری از فرایند قضاوتی کارشناسان خبره بوسیله بکارگیری الگوهای کشف شده این تحقیق در یک سیستم هوشمند (یادگیری ماشین).
- ۳- تعیین چارچوب ریسک برای طبقات مختلف بیمه گذاران و در نتیجه پرداخت میزان عادلانه حق بیمه بر مبنای میزان ریسک هر بیمه گذار.
- ۴- مدیریت بهینه بیمه گذاران آینده و پذیرش ریسک بصورت آگاهانه و متعادل کردن طبقات ریسک انتخابی.
- ۵- افزایش قابلیت دسترسی، زیرا افراد خبره جهت تعیین ریسک همیشه در دسترس نیستند؛ اما یک سیستم هوشمند همواره در دسترس است.

۱-۲-۴ جنبه جدید بودن و نوآوری در تحقیق

این تحقیق با ارائه یک چارچوب پیشنهادی جدید، بیمه گذاران بیمه بدنه اتومبیل را با استفاده از "درخت تصمیم" بعنوان یکی از تکنیک های داده کاوی، طبقه بندی می نماید؛ که سوابق تحقیقاتی بررسی شده، بیانگر جدید بودن و نوآوری طرح می باشد.

۱-۳ اهداف تحقیق

این بخش شامل اهداف آرمانی، کلی، ویژه، کاربردی و همچنین بهره وران از نتایج تحقیق می باشد؛ که در ادامه بیان می گردند.

۱-۳-۱ هدف آرمانی تحقیق

جایگزینی سیستم ارزیابی ریسک بیمه گذاران بیمه بدنه اتومبیل کشور با طبقات ریسک بدست آمده از داده کاوی.

۱-۳-۲ هدف کلی تحقیق

طبقه بندی ریسک بیمه گذاران بیمه بدنه اتومبیل با استفاده از داده کاوی .

۱-۳-۳ اهداف ویژه تحقیق

- ۱- تعیین مهمترین متغیر های دخیل در پیش بینی رفتار بیمه گذاران بیمه بدنه اتومبیل
- ۲- جستجوی الگوهای موجود بین داده های بیمه گذاران سابق و ارائه چارچوب مناسبی برای طبقه بندی بیمه گذاران بیمه بدنه اتومبیل با استفاده از داده کاوی .

۱-۳-۴ اهداف کاربردی تحقیق

۱. طبقه بندی بیمه گذاران بیمه بدنه اتومبیل به ریسک بالا ، متوسط ، پایین در صنعت بیمه اتومبیل کشور بر اساس ویژگی های شخصی و مشخصات خودرو.
۲. پیش بینی وضعیت بیمه گذاران آینده بیمه بدنه اتومبیل با استفاده از شناسایی عوامل تاثیرگذار بر ریسک.
۳. افزایش سودآوری شرکتهای بیمه از طریق ارائه طبقه بندی مناسب از ریسک بیمه گذاران بیمه بدنه اتومبیل.

۴. شناسایی عوامل تاثیر گذار بر رفتار بیمه گذاران بیمه بدنه اتومبیل با توجه پایگاه داده های موجود .

۱-۳-۴-۱ بهره وران از نتایج تحقیق

تحقیق پیش رو یک تحقیق کاربردی است که با استفاده از پایگاه داده بیمه مرکزی ج.ا.ا. ، مورد بررسی قرار گرفته است و الگوگیری از فرایند انجام تحقیق و نتایج آن می تواند برای کلیه شرکتهای بیمه مفید واقع شود.

۱-۴-۱ سوالات تحقیق

سوالات این تحقیق در دو بخش سوال اصلی و سوال فرعی تحقیق بیان شده است.

۱-۴-۱-۱ سؤال اصلی تحقیق

۱. چگونه می توان در صنعت بیمه ، بیمه گذاران بیمه بدنه اتومبیل را با استفاده از داده کاوی ، طبقه بندی نمود؟

۲. آیا استفاده از تکنیک درخت تصمیم طبقه بندی صحیح تری را نسبت به سایر تکنیکها حاصل خواهد کرد؟

۱-۴-۲ سوالات فرعی تحقیق

۱. آیا سن و جنسیت موثرترین عوامل در میزان ریسک بیمه گذاران می باشند؟

۱-۵ فرضیه تحقیق

با توجه به اینکه هدف داده کاوی یافتن الگوها و دانستی‌هایی در داده هاست که ساختار آنها از قبل مشخص نمی باشد، بنابراین نمی توانیم فرضیه پژوهشی از قبل برای این تحقیق تعیین نماییم.

سئوالات تحقیق:

۱. چگونه می توان بیمه گذاران بیمه اتومبیل را از نظر ریسک با استفاده از داده کاوی طبقه بندی کرد؟

۲. آیا تکنیک درخت تصمیم بهترین تکنیک برای طبقه بندی بیمه اتومبیل است؟

سئوال فرعی: ۳. آیا سن و جنسیت موثرترین عوامل در ریسک بیمه گذاران می باشد؟

مدل مفهومی که این پژوهش قصد یافتن آن را دارد ، مجموعه ای از الگوهایی است که پیش بینی ریسک بیمه گذاران بیمه اتومبیل را ممکن می سازد.

۱-۶ قلمرو تحقیق (زمانی، مکانی و موضوعی)

قلمرو زمانی تحقیق:

با توجه به اطلاعات قابل دسترسی، داده های بیمه گذاران مربوط به سالهای ۱۳۸۸-۱۳۸۴ می باشد. ؛ که به عنوان قلمرو زمانی تحقیق مد نظر قرار گرفته است.

قلمرو مکانی داده های تحقیق:

بانک اطلاعاتی بیمه خودرو بیمه مرکزی ج.ا.

قلمرو موضوعی تحقیق: