
بسم الله الرحمن الرحيم

دانشکده مهندسی صنایع و مدیریت

پایان‌نامه کارشناسی ارشد مدیریت صنعتی

ارزیابی و تشخیص صحت تقاضاهای خسارت بیمه ای با استفاده از تکنیک‌های داده‌کاوی مبتنی بر رویکرد یادگیری هدایت شده

نگارنده: نگار صادقیان بروجنی

استاد راهنما

دکتر سید محمد حسن حسینی

استاد مشاور

دکتر علی‌اکبر حسنی

آبان ۱۳۹۵

شماره:
تاریخ: ۱۳۹۵/۹/۱۷
ویرایش:

مدیریت تحصیلات تکمیلی

باسمه تعالی

فرم شماره ۶: صورتجلسه نهایی دفاع از پایان نامه دوره کارشناسی ارشد

با تأییدات خداوند متعال و با استعانت از حضرت ولی عصر (عج) ارزیابی جلسه دفاع از پایان نامه کارشناسی ارشد خانم / آقای **صادقیان بروجنی نگار** به شماره **دانشجویی ۹۳۱۰۸۹۴ رشته مدیریت صنعتی** تحت عنوان ارزیابی و تشخیص صحت تقاضاهای خسارت بیمه ای با استفاده از تکنیک های داده کاوی مبتنی بر رویکرد یادگیری هدایت شده با حضور هیأت محترم داوران در دانشگاه صنعتی شاهرود برگزار گردید به شرح ذیل اعلام می‌گردد:

قبول (با درجه : عالی) امتیاز ۱۸/۷۶ ☐ دفاع مجدد ☐ مردود ☐

نوع تحقیق: نظری ☐ عملی ☐

۱ـ عالی (۲۰ ـ ۱۹)
۲ـ بسیار خوب (۱۸/۹۹ ـ ۱۸)
۳ـ خوب (۱۷/۹۹ ـ ۱۶)
۴ـ قابل قبول (۱۵/۹۹ ـ ۱۴)
۵- نمره کمتر از ۱۴ غیر قابل قبول

اعضاء هیئت داوران	نام و نام خانوادگی	مرتبهٔ علمی	امضاء
۱ـ استادراهنمای اول	دکتر سید محمد حسینی	استادیار	
۲- استادراهنمای دوم	ـــــــ	ـــــــ	
۳- استاد مشاور	دکتر علی اکبر حسنی	استادیار	
۴- نماینده شورای تحصیلات تکمیلی	آقای مجید عامری	مربی	
۵- استاد ممتحن اول	دکتر رضا شیخ	دانشیار	
۶- استاد ممتحن دوم	دکتر محمد فتاحی حسن آباد	استادیار	

نام و نام خانوادگی رئیس دانشکده:

تاریخ و امضاء و مُهر دانشکده:

تقدیم:

این پایان‌نامه را ضمن تشکر و سپاس بیکران و در کمال افتخار و امتنان تقدیم می‌نمایم به محضر ارزشمند پدر و مادر عزیزم به خاطر همه‌ی تلاش‌های محبت‌آمیزی که در دوران مختلف زندگی‌ام انجام داده‌اند و با مهربانی چگونه زیستن را به من آموخته‌اند و به آنان که درراه کسب دانش راهنمایم بودند و نفس خیرشان و دعای روح‌پرورشان بدرقه‌ی راهم بود.

تقدیر تشکر:

خداوند بزرگ را شاکرم که لطف خود را شامل حال من نمود تا بتوانم تحقیق خود را به پایان برسانم و بتوانم سهمی هرچند اندک، در راه توسعه علمی ایران عزیز بردارم که چو ایران نباشد، تن من مباد.

برخودلازم می‌دانم از کلیه کسانی که بنده را در تدوین و نگارش این پایان‌نامه یاری نمودند صمیمانه تشکر و قدردانی نمایم. به‌خصوص از استاد فرزانه جناب آقای دکتر سید محمدحسن حسینی (استاد راهنما) که در کلیه مراحل انجام این پژوهش با خوش‌رویی، یاری و راهنمائی ام نمودند و همچنین از استاد فرهیخته جناب آقای دکتر علی‌اکبر حسنی (استاد مشاور) که وقت خود را بی‌شائبه در اختیار من گذاشته و با دقت نظر خاصی مشاوره لازم در این خصوص ارائه نمودند صمیمانه تشکر و قدردانی می‌نمایم.

تعهدنامه

اینجانب نگار صادقیان بروجنی دانشجوی دوره کارشناسی ارشد رشته **مدیریت صنعتی** دانشکده **مهندسی صنایع و مدیریت** دانشگاه صنعتی شاهرود نویسنده پایان‌نامه **ارزیابی و تشخیص صحت تقاضاهای خسارت بیمه ای با استفاده از تکنیک‌های داده‌کاوی مبتنی بر رویکرد یادگیری هدایت شده** تحت راهنمایی دکتر محمد حسن حسینی متعهد می‌شوم.

- تحقیقات در این پایان‌نامه توسط اینجانب انجام‌شده است و از صحت و اصالت برخوردار است.
- در استفاده از نتایج پژوهش‌های محققان دیگر به مرجع مورداستفاده استناد شده است.
- مطالب مندرج در پایان‌نامه تاکنون توسط خود یا فرد دیگری برای دریافت هیچ نوع مدرک یا امتیازی در هیچ جا ارائه نشده است.
- کلیه حقوق معنوی این اثر متعلق به دانشگاه شاهرود می‌باشد و مقالات مستخرج با نام «دانشگاه شاهرود» و یا « **University of Shahrood**» به چاپ خواهد رسید.
- حقوق معنوی تمام افرادی که در به دست آمدن نتایج اصلی پایان‌نامه تأثیرگذار بوده‌اند در مقالات مستخرج از پایان‌نامه رعایت می‌گردد.
- در کلیه مراحل انجام این پایان‌نامه، در مواردی که از موجود زنده (بافتهای آن‌ها) استفاده‌شده است ضوابط و اصول اخلاقی رعایت شده است.

در کلیه مراحل انجام این پایان‌نامه، در مواردی که به حوزه اطلاعات شخصی افراد دسترسی یافته یا استفاده‌شده است اصل رازداری، ضوابط و اصول اخلاق انسانی رعایت شده است.

تاریخ:

امضای دانشجو

چکیده

در دنیایی که حوادث متفاوتی جان و مال انسان‌ها را تهدید می‌کند، بیمه راه‌کار مناسبی است تا ریسک این خطرات را از فرد به شرکت‌های بیمه منتقل کند شرکت‌های بیمه‌ای محصولات مختلفی ارائه می‌دهند که هر کدام می‌تواند بخشی از این خطرات را پوشش دهد؛ اما متاسفانه گاها فلسفه وجودی بیمه فراموش می‌شود و افرادی با ترفندهای مختلف به قصد تقلب و سوءاستفاده از بیمه شخص ثالث برمی‌آیند. تقلب های بیمه ای از مسائل مهم و خسارت زا برای شرکت های بیمه و بیمه گذاران، در تمام رشته های بیمه ای است. یکی از راه های شناسایی تقلب در خسارت های اعلام شده، استفاده از اطلاعات تقلب های کشف شده در گذشته است. امروزه روش های داده کاوی به طور گسترده در کشف الگوها در داده ها استفاده می شوند. استفاده از این روش ها می تواند در شناسایی خسارت های تقلبی در صنعت بیمه مفید باشد. در این مقاله ضمن بررسی روش های رایج برای شناسایی تقلب در بیمه اتومبیل از سه روش داده کاوی ماشین بردارپشتیبان، جنگل تصادفی و ترکیب این دو روش استفاده شده است که به شرکت های بیمه در شناسایی تقلب ها در بیمه اتومبیل کمک می کنند. همچنین در یک مطالعه تجربی این روش ها بر روی داده های واقعی (شامل اطلاعات ۱۰۰۰ پرونده خسارت بیمه نامه های شخص ثالث) آزمایش و کارایی هر روش سنجیده شد. نتایج نشان داد که دقت الگوریتم

ترکیبی در آموزش ۹۹.۷۹٪ و در تست ۸۵.۰۷٪ است که از دو الگوریتم پایه بالاتر است.

کلمات کلیدی: خسارت بیمه ای، یادگیری هدایت شده، داده کاوی، تقلب

فهرست مطالب

فهرست اشکال

فهرست جداول

فصل اول

کلیات تحقیق

فصل ۱: کلیات تحقیق

۱-۱ مقدمه

در زندگی امروزی به دلیل افزایش احتمال خطراتی همچون تصادف، بیماری، از دست دادن شغل و ...، موضوع بیمه تبدیل به یک پدیده عمومی و فراگیر شده و کمتر شخص یا سازمانی را می‌توان یافت که با بیمه و مسائل مربوط به آن سروکار نداشته باشد. در این میان حوادث و تصادفات معمول همراه با مصدومیت‌های جسمی و خسارات مالی زیاد است و لذا افزایش تمایل مردم به استفاده از خدمات بیمه‌ای وسایل حمل‌ونقل امری عادی به نظرمی رسد. از طرف دیگر توسعه زندگی ماشینی و افزایش روزافزون تمایل به استفاده از وسایل نقلیه شخصی، اهمیت جایگاه شرکت‌های بیمه‌ای در این حوزه را بیشتر نمایان می‌کند. با توجه به هزینه‌های زیاد حوزه‌ی بیمه به دلیل تقلب پذیر بودن و همچنین منعطف بودن جهت تعامل با همه نوع قشر اجتماعی، همواره درخطر تقلب و سوءاستفاده است. تقلب در صنعت بیمه یکی از منابع عمده ریسک عملیاتی شرکت‌های بیمه است و یک بخش قابل‌توجهی از ضررها را تشکیل می‌دهد.

یکی از شاخه‌های بیمه‌ای، بیمه شخص ثالث است که بیمه‌ای اجباری و بسیار خطرپذیر است. این نوع بیمه خسارت‌های راننده‌ی مقصری که تحت پوشش این بیمه باشد را با پرداخت نقدی جبران می‌کند. اجباری بودن و حادثه‌خیز بودن این نوع بیمه باعث شده که هزینه‌های وارده بر این بیمه بسیار زیاد باشد. ازاین‌رو تقلب‌های متفاوتی در این زمینه اتفاق می‌افتد.

به‌طورکلی، بیمه‌گذاران در دو موقعیت مرتکب تقلب می‌شوند: مورد اول، شرایطی است که در آن، فرد آگاهانه سعی در ایجاد خسارت یا اغراق در میزان و نوع خسارت دارد. به‌عنوان‌مثال، در یک سانحه‌ی تصادف ممکن است فرد بیمه‌گذار با توجه به‌حق بیمه‌ای که برای سالیان متمادی به شرکت بیمه پرداخت‌نموده است درصدد بهره‌برداری از فرصت برآید و با تجمیع کلیه‌ی زیان‌های پیشین با خسارت

فعلی سعی در کسب موقعیت مالی بهتر کند. مورد دوم که ممکن است منجر به خسارت‌های جعلی گردد، مواردی است که بیمه‌گذار به‌صرف داشتن بیمه‌نامه، احتیاط کمتری می‌کند. بدین معنی که گرچه ممکن است شخص قصد ایجاد خسارت یا اغراق در میزان آن را نداشته باشد، بااین‌حال اقدام به انجام فعالیت‌هایی می‌کند که در صورت نداشتن بیمه‌نامه، این فعالیت را انجام نمی‌داد.

هزینه‌های بالای حل‌وفصل ادعاهای دروغین بر حق بیمه تأثیر می‌گذارد و باعث کاهش مزیت رقابتی شرکت‌های بیمه می‌شود؛ بنابراین، تقلب‌های بیمه یک مشکل جدی در کل صنعت بیمه هستند و با توجه به هزینه‌های زیاد ناشی از تقلب‌ها و کلاه‌برداری‌های بیمه‌ای، موضوع بررسی تقلب‌های بیمه‌ای امروزه به یکی از حوزه‌های مهم تحقیقاتی تبدیل‌شده است.

۱-۲ تعریف مسئله

با توجه به حجم روزافزون تصادفات و تعهد شرکت‌های بیمه‌ای به پرداخت هزینه‌های وارده به اتومبیل و دیه اشخاص آسیب‌دیده (به شرطی که فرد مقصر دارای گواهینامه و بیمه‌نامه باشد)، همواره موارد قابل‌توجهی از تصادف‌ها در جامعه در حال رخ دادن است که از طریق بیمه‌نامه شخص ثالث خساراتشان را جبران می‌کنند که این هزینه‌ها خارج از تعهدات شرکت‌های بیمه‌ای است؛ و یا مبالغی را ادعا می‌کنند که بیشتر از خسارت‌های واقعی هستند و یا به‌عبارت‌دیگر اغراق در خسارت می‌کنند. ازاین‌رو برای شرکت‌های بیمه‌ای شناسایی این موارد بسیار حائز اهمیت است تا ضمن برخورد با آن‌ها، این عمل ناشایست سوءاستفاده از بیمه در جامعه را ریشه‌کن کنند. شرکت‌های بیمه برای دستیابی به این هدف ابزارهای مختلفی را مورداستفاده قرار می‌دهند. ازجمله مرسوم‌ترین و قوی‌ترین ابزار مورداستفاده در این زمینه، داده‌کاوی می‌باشد که تکنیک های مختلفی را شامل می‌شود.[۱]

تکنیک‌های داده‌کاوی می‌تواند با تحلیل بانک‌های اطلاعاتی شرکت‌های بیمه، این شرکت‌ها را در راستای تحقق هدفشان یاری نماید و هرچه کیفیت داده‌ای موجود بهتر باشد نتایج حاصل از داده‌کاوی دقیق‌تر و

قابل‌اعتمادتر است. ازجمله تکنیک های داده کاوی خوشه بندی و دسته‌بندی اطلاعات می‌باشد که دید بسیار مناسبی از تصادفات را ارائه می‌دهد. داده‌های موجود در هر خوشه دارای ویژگی‌های مشترکی هستند که به شناسایی نوع تقلباتی که در هر خوشه می‌تواند وجود داشته باشد کمک می‌کنند. البته باید اشاره کرد که تعداد زیادی از داده‌های موجود در بعضی خوشه‌ها هیچ مورد مشکوکی ندارند و به‌راحتی می‌توان از بررسی دقیق آن‌ها صرف‌نظر کرد و اقداماتی که در راستای پرداخت است را هرچه سریع‌تر انجام داد تا موجبات رضایت هرچه بیشتر بیمه‌گزاران قانونمند فراهم شود.

با توجه به این‌که در حال حاضر مسئولیت شناسایی تقلب در شرکت‌های بیمه‌ای به عهده کارشناس خسارت است، تحقیقات انجام‌شده در این راستا به دنبال جایگزین کردن فرآیند سیستمی برای کشف تقلب به‌جای نیروی انسانی متخصص نیستند بلکه ابزاری را در جهت کمک به کارشناسان برای تسریع بخشیدن به کشف و شناسایی تقلب معرفی می‌نماید.

هدف از این تحقیق ارائه یک مدل پیشنهادی جدید جهت ارزیابی و تشخیص صحت تقاضاهای خسارت بیمه‌ای با استفاده از تکنیک‌های داده‌کاوی مبتنی بر رویکرد یادگیری هدایت‌شده می‌باشد. با توجه به اینکه در استفاده از تکنیک‌های داده‌کاوی همواره نیازمند انبار داده و اطلاعات اولیه به‌عنوان داده‌های آموزشی می‌باشیم لذا از داده‌ها و سوابق موجود دریکی از بیمه‌های خودرو به‌عنوان داده‌های آموزشی استفاده می‌شود. پس از توسعه مدل مبتنی بر داده‌های موجود، این مدل مورد ارزیابی و تحلیل قرار می‌گیرد تا از عملکرد مناسب آن اطمینان حاصل شود. همچنین به‌منظور انجام تجزیه‌وتحلیل نتایج و ارزیابی مدل، از نرم‌افزارهای داده‌کاوی مانند Clementine و یا سایر نرم‌افزارهای مناسب نیز استفاده خواهد شد.

۱-۳- اهداف تحقیق

هدف اصلی تحقیق حاضر به شرح زیر می‌باشد:

- توسعه یک مدل پیشنهادی جهت ارزیابی و تشخیص صحت تقاضاهای خسارت بیمه‌ای با استفاده از تکنیک‌های داده‌کاوی مبتنی بر رویکرد یادگیری هدایت‌شده.

اهداف فرعی این تحقیق نیز عبارت‌اند از:

- ✓ شناسایی انواع مختلف تقلبات بیمه‌ای و دلایل ارتکاب آن‌ها
- ✓ بررسی تکنیک‌های مختلف داده‌کاوی قابل‌استفاده در ارزیابی و شناسایی صحت تقاضاهای بیمه‌ای
- ✓ به‌کارگیری الگوریتم‌های مختلف داده‌کاوی جهت انجام پیش‌بینی صحت تقاضاهای خسارت بیمه‌ای
- ✓ ارزیابی دقت انواع الگوریتم‌های داده‌کاوی در پیش‌بینی صحت تقاضاهای خسارت بیمه‌ای و معرفی برترین الگوریتم

۱-۴ - اهمیت موضوع

افراد مختلفی با اتومبیل‌های متفاوت سعی می‌کنند که به‌نوعی از شرکت‌های بیمه سوءاستفاده کنند. یکی از مهم‌ترین دلایل این پدیده، درک نادرست از فلسفه وجودی بیمه است. [۲]دلیل مهم دیگر موفقیت آن‌ها در اعمال متقلبانه بوده است که موارد بسیاری از آن‌ها به نتیجه رسیده و هزینه‌ای سنگینی را برای شرکت‌های بیمه‌ای به بار آورده‌اند. از طرفی شگردهای تقلب همواره در حال تغییر هستند و شناسایی آن‌ها بسیار وقت‌گیر و هزینه‌بر است. کارکرد هم‌زمان مقابله با موارد تقلبی و فرهنگ‌سازی

استفاده درست از بیمه‌نامه و معرفی فلسفه وجودی بیمه می‌تواند تأثیر قابل‌توجهی بر تعداد تقلبات آتی داشته باشد و اگر راهکاری برای مقابله با این پدیده اندیشه نشود، این عمل ناشایست در جامعه بیش‌ازپیش رواج پیدا خواهد کرد.

ازآنجایی‌که روند ارزیابی خسارت‌ها معمول به‌صورت دستی انجام می‌گیرد و کمتر از دستگاه‌های کامپیوتری استفاده می‌شود، ادعاهای تقلبی معمولاً شناسایی نمی‌شوند. همچنین با توجه به این‌که همواره شیوه‌های جدیدی در کلاه‌برداری‌ها به کار گرفته می‌شود، روش‌های مورداستفاده در کشف تقلب باید قابلیت کافی برای شناسایی کشف تقلب را داشته باشند. هرچند کشف کاملاً خودکار کلاه‌برداری‌ها در عمل ممکن نیست، اما استفاده از اطلاعات تقلب‌های کشف‌شده درگذشته و بهره‌گیری از تکنیک‌های آماری می‌تواند به کارشناسان خسارت در شناسایی خسارت‌های جعلی کمک کند.[۳]

بنابراین باوجود خسارت‌های جبران‌ناپذیری که سازمان‌ها و افراد دچار آن شده‌اند، کشف تقلب به یک موضوع بااهمیت تبدیل‌شده است. با افزایش حجم اطلاعات و پیچیده شدن آن در تجزیه‌وتحلیل اطلاعات به‌دست‌آمده، تنها روش‌های داده‌کاوی توانای پردازش این اطلاعات را دارد. علم داده‌کاوی با استفاده از روش‌های متعددی که برخوردار است، الگوهای مناسبی را برای تشخیص تقلب ارائه می‌دهند. از طرفی باوجود ضررهای مالی هنگفت تقلب، اهمیت این موضوع برای سازمان‌ها دوچندان شده است. امروزه با پیشرفت تکنولوژی و با استفاده از روش‌های پیشرفته آماری، کشف تقلب مؤثر گشته است، بر این اساس ضرورت دانستن روش‌های داده‌کاوی که اساس آن یادگیری آماری است بیشتر از قبل شده است. با توجه به توضیحات فوق، بررسی علمی موضوع تقلب و به‌کارگیری مدلی منطقی و بدون تأثیرپذیری از سلیقه‌ها و نظرات شخصی جهت بررسی ادعاهای مشتریان کاملاً ضروری به نظر می‌رسد تا از این طریق بتوان تا حدودی از تخلفات و ضررهای هنگفت ناشی از آن‌ها جلوگیری نمود.

تقلب در بیمه اتومبیل از روش‌های مختلفی مثل اغراق در میزان خسارت وارده، تصادفات ساختگی، اسناد جعلی و ارائه اطلاعات نادرست می تواند رخ دهد و ازآنجایی‌که بیمه شخص ثالث بسیار در معرض انواع تقلبات است و سهم زیادی در پرتفوی حق بیمه دریافتی شرکت‌های بیمه در بسیاری از کشورها ازجمله ایران دارد و حجم ادعای خسارت دریافتی بالا و بالطبع ضریب خسارت بالایی را نسبت به سایر بیمه‌ها به خود اختصاص می‌دهد، بررسی این رشته و کشف و شناسایی تقلب در آن ضروری هست. ازاین‌رو در تحقیق حاضر به بررسی این شاخه از بیمه و کشف و شناسایی تقلبات بیمه شخص ثالث می‌پردازیم.

۱-۵ کاربرد نتایج تحقیق

با استفاده از نتایج این تحقیق می‌توان توسط نرم‌افزار و الگوریتم‌های جدید روند کشف و پیش‌بینی تقلبات بیمه‌ای را با دقت و سرعت بیشتری ارتقا داد. همچنین می‌توان به شناسایی عوامل تأثیرگذار بر تقلب پرداخت و از اعمال سلیقه‌های شخصی و غیرقابل‌دفاع در شرکت‌های بیمه جلوگیری کرد. نتایج این تحقیق با دربرداشتن فرآیند خوشه‌بندی خسارات در خوشه‌هایی با ریسک‌های مختلف تقلب و همچنین با دسته‌بندی و پیش‌بینی تقلبات آتی کمک شایانی به شرکت‌های بیمه خواهد نمود تا از این طریق زمان رسیدگی به خسارات را کاهش داده و با دقت بالاتری به جعلی بودن یا نبودن ادعاهای خسارات بپردازد.

به‌طور خلاصه نهادهای زیر می‌توانند از نتایج این تحقیق استفاده نمایند:

- دانشگاه‌ها
- مراکز تحقیقاتی و پژوهشی
- شرکت‌های بیمه‌گذار
- مراکز آموزشی و مشاوره‌ای
- اداره پلیس

- قوه قضاییه و دادگستری

۱-۶- سؤالات تحقیق

۱- چگونه می‌توان تکنیک‌های داده‌کاوی را جهت مقابله با پدیده‌ی تقلب در صنعت بیمه بکار گرفت؟

۲- کدام‌یک از الگوریتم‌های موردبررسی در تکنیک‌های فوق دقت بیشتری در پیش‌بینی تقلب دارند؟

۳- با توجه به میزان دقت و صحت الگوریتم‌های بررسی‌شده برای کشف تقلب؛ در چه مرحله‌ای می‌توان از این الگوریتم‌ها استفاده نمود؟

۱-۷- محدودیت‌های تحقیق

- اکثر فرم‌ها و پرونده‌ها به‌صورت غیر سیستمی و دستی ذخیره‌شده‌اند و تعداد کمی از آن‌ها به‌صورت الکترونیکی ذخیره و نگهداری می‌شوند و جستجوی دستی در این اسناد بسیار زمان‌بر است.
- شرکت‌های بیمه اطلاعات و پرونده‌های مشتریان خود را به‌آسانی در اختیار افراد خارج از سازمان قرار نمی‌دهند و اخذ مجوز برای دسترسی به این اطلاعات فرآیندی زمان‌بر و مشکل می‌باشد.
- دسترسی به تعداد اندکی پرونده که وقوع تقلب در آن‌ها محرز شده است بسیار مشکل است زیرا این پرونده‌ها اغلب محرمانه هستند و هرکسی حق دسترسی به آن‌ها را ندارد.

- کسب دانش موردنیاز از افراد خبره و کارشناسان کشف تقلبات بیمه‌ای مشکل است زیرا اکثر آن‌ها مدیران بخش بیمه اتومبیل هستند و به‌ندرت وقت آزادی برای قبول و انجام مصاحبه دارند.

۱-۸- ساختار پایان‌نامه

این پایان‌نامه شامل ۵ فصل می‌باشد. در فصل اول کلیات تحقیق شامل بیان مسئله؛ ضرورت تحقیق؛ اهداف تحقیق و ... تشریح شد. در فصل دوم به شرح مبانی نظری تحقیق پرداخته و سپس خلاصه‌ای از

تحقیقات انجام‌شده در این حوزه تشریح می‌گردد. در فصل سوم روش تحقیقی که قصد انجام آن راداریم و الگوریتم‌ها و نرم‌افزار مورداستفاده را تشریح می‌شود و در فصول چهارم و پنجم نتایج به‌دست‌آمده از پژوهش تحلیل خواهد شد و نتیجه نهایی وپیشنهادات برای تحقیقات آتی مطرح می‌شود.

۱-۹-مراحل تحقیق

گام اول: مطالعات کتابخانه‌ای و مرور ادبیات و بررسی تحقیقات پیشین.

گام دوم: انتخاب تکنیک داده‌کاوی و انتخاب روش مناسب با مساله تعریف شده.

گام سوم: انتخاب جامعه آماری مناسب و جمع‌آوری داده‌ها به‌صورت کتابخانه‌ای و میدانی.

گام چهارم: انتخاب نرم‌افزار مناسب جهت ساخت و آنالیز مدل.

گام پنجم: آماده‌سازی داده‌ها در نرم‌افزار و ساخت مدل به‌وسیله‌ی آن.

گام ششم: تجزیه‌وتحلیل مدل‌های ساخته‌شده و مقایسه‌ی آن‌ها ازنظر دقت و انتخاب بهترین آن‌ها.

۱-۱۰- جمع‌بندی

در این فصل ابتدا به بیان مقدمه‌ای پیرامون داده‌کاوی و کاربردهای آن پرداخته شد. از اهمیت موضوع موردبررسی و کاربرد نتایج تحقیق و محدودیت‌های آن سخن به میان آورده شد و همچنین اهدافی که در طی این پژوهش به دنبال آن‌ها هستیم و سوا لاتی که به دنبال پاسخ آن‌ها هستیم معرفی شدند.

فصل دوم

ادبیات موضوع

فصل ۲: ادبیات موضوع و پیشینه‌ی تحقیق

۲-۱ - مقدمه

در این بخش ابتدا به معرفی تقلب‌های مالی و تقلب‌های بیمه‌ای که یکی از شاخه‌های تقلب مالی است و توضیحات پیرامون آن می‌پردازیم، سپس در ادامه به معرفی تکنیک‌های تشخیص تقلب پرداخته و به‌تفصیل به شرح تکنیک داده‌کاوی و الگوریتم‌های اساسی آن در بخش کشف تقلب می‌پردازیم و در آخر مروری به تحقیقات انجام‌شده در این زمینه می‌کنیم. در این میان یکی از روش‌های نوین ارزیابی و کشف تقلب بیمه‌ای؛ داده‌کاوی هست. داده‌کاوی هم‌زمان با ایجاد و استفاده از پایگاه داده‌ها در اوایل دهه ۸۰ برای جستجوی دانش در داده‌ها شکل گرفت. شاید بتوان لوول (۱۹۸۳) را اولین شخصی دانست که مقاله‌ای در مورد داده‌کاوی تحت عنوان "شبیه‌سازی فعالیت داده‌کاوی " ارائه نمود. هم‌زمان با او پژوهشگران و متخصصان علوم رایانه، آمار، هوش مصنوعی، یادگیری ماشین و ... نیز به پژوهش در این زمینه و زمینه‌های مرتبط با آن پرداخته‌اند. البته پژوهش جدی روی موضوع داده کاوی از اوایل دهه ۱۹۹۰ شروع شد و تا آن زمان این واژه به شکل و معنای امروزی بکار برده نمیشد.

پیشرفت در ذخیره‌سازی داده‌ها سبب وجود پایگاه داده‌های بزرگ شد. با توجه به وجود اطلاعات ارزشمند در این پایگاه داده‌ها تلاش برای استخراج اطلاعات شروع شد. با نگاهی به کلمه کاو[1] به این نکته پی می‌بریم که به معنای استخراج از منابع نهفته و باارزش زمین می‌باشد. پیوند این کلمه با داده[2] به معنی جستجوی عمیق جهت پیدا کردن اطلاعات مفید که قبلاً نهفته بودند را می‌دهد.

1. Mine
2. Data

ایده‌ای که مبنای داده‌کاوی است، یک فرآیند بااهمیت از شناخت الگوهای بالقوه، مفید، بدیع و نهایتاً قابل‌درک از داده‌هاست. واژه کشف دانش " کشف دانش در پایگاه داده‌ها"[1] که به معنای جستجوی دانش در اطلاعات است، در اوایل دهه ۸۰ شکل‌گرفته است. کشف دانش و داده‌کاوی یک حوزه میان‌رشته‌ای و در حال رشد است که حوزه‌های مختلف همچون پایگاه داده، آمار، یادگیری ماشین[2]، مصورسازی[3]، هوش مصنوعی[4]، بازشناسی الگو[5] و سایر زمینه‌های مرتبط را باهم تلفیق کرده است تا اطلاعات و دانش ارزشمند نهفته در حجم بزرگی از داده‌ها را استخراج نماید.[۴]

سیستم‌های پایگاه داده با فراهم کردن ابزارها و محیط‌های لازم، بستر لازم برای مدیریت و دسترسی سیستماتیک و مؤثر به این حجم از داده را تسهیل کرده‌اند؛ اما استخراج دانش از پایگاه‌های داده بدون استفاده از کامپیوتر و به‌کارگیری ابزارهای تحلیلی قدرتمند و خودکار کاری بسیار دشوار و شاید غیرممکن است. وجود چنین ابزارهایی فاصله‌ی چشمگیر موجود میان تولید داده و فهم آن را کاهش داده و راه‌های کشف الگوهای مفید از پایگاه‌های داده را که در عرصه‌های گوناگون علمی، تحقیقی و تجاری موردتوجه است، هموارتر می‌سازد. در حال حاضر ابزارها و تکنیک‌های زیادی برای رسیدن به این هدف پیشنهادشده‌اند که همگی تحت عنوان کلی داده‌کاوی مطرح می‌شوند.[۵]

1. Knowledge Discovery in Database
2 .Machine Learning
3 . Virtualization
4 . Artificial Intelligence
5 . Pattern Recognition

۲-۲-مبانی نظری

۲-۲-۱-تقلب

اگر بخواهیم تقلب را با تعریف لغت‌نامه‌ای آن در آکسفورد معرفی کنیم باید بگوییم:

" فریب غیرقانونی و مجرمانه که باهدف دستیابی به سود مالی و یا سود شخصی صورت می‌پذیرد."

طبق گفته‌ی نگی[1] هیچ تعریف پذیرفته‌شده‌ای از تقلب مالی وجود ندارد. ولی چهار رکن در همه‌ی تعریف‌ها مشترک است[۶]:

- عملی آشکار و عمدی
- عملی که ضد قانون باشد.
- عملی که سود و منفعتی مالی داشته باشد.
- ارائه اطلاعات به‌اشتباه صورت گرفته باشد.

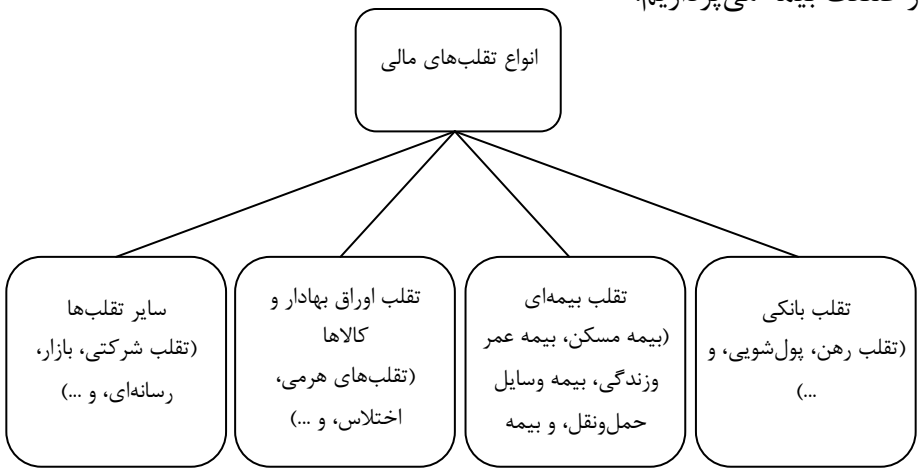
بنابراین فرد در صورتی متقلب شناخته می‌شود و مورد پیگرد قانونی قرار می‌گیرد که هر ۴رکن نام‌برده در عملش ثابت‌شده باشد. ازاین‌رو تقلب به دودسته اصلی تقسیم می‌شود، دسته اول تقلبات کیفری است و دسته دوم که از فراوانی بیشتری برخوردارند تقلباتی هستند که کیفری بودن آن‌ها ثابت نشده و باید بررسی شوند.

تقلب انواع متفاوتی دارد و در همه‌ی صنایع رخ می‌دهد و می‌تواند طیف گسترده‌ای را شامل شود. در یک دسته‌بندی کلی تقلب را به چهار گروه اصلی تقسیم کرده‌اند: تقلب‌های بانکی، تقلب‌های بیمه‌ای، تقلب‌های امنیتی و تجاری و سایر تقلب‌های مربوطه.

[1]. Ngai

۲-۲-۲- تقلب‌های مالی

در سال‌های اخیر تقلبات مالی زیادی در حوزه‌های مختلف رخ‌داده است و جامعه را متحمل ضررهای زیادی نموده است. طبق تحقیقات نگی وهمکاران انواع تقلب مالی را می‌توان مطابق شکل (۲-۱) درچهاردسته تقسیم کرد. با توجه به هزینه‌های زیاد ناشی از تقلب‌ها و کلاه‌برداری‌های بیمه‌ای، موضوع بررسی تقلب‌های بیمه‌ای امروزه به یکی از چهار حوزه تحقیقاتی درزمینه ی تقلب و کلاه‌برداری تبدیل‌شده است. و در این تحقیق به بررسی تقلبات بیمه‌ای و ارائه روشی جهت پیشگیری از این نوع تقلب در صنعت بیمه می‌پردازیم.



شکل (۲-۱) دسته‌بندی تقلب‌های مالی

گیل و همکارانش، تقلب بیمه‌ای را این‌گونه تعریف کرده‌اند:

" اعلام عمدی خسارت‌های جعلی ، اعلام خسارت بیش از مقدار واقعی آن، یا هر روش دیگر برای به دست آوردن مبلغی بیش ازآنچه بیمه گزارقانونا مستحق دریافت آن باشد."

تقلب‌های بیمه‌ای به شیوه‌های مختلفی قابل رخ دادن هستند که می‌توان آن‌ها را به دودسته‌ی تقلب نرم[1] و تقلب سخت[2] تقسیم کرد.

بعضی از کلاه‌برداری‌ها در صنعت بیمه کاملاً آگاهانه و عمدی است مثلاً بیمه‌گزار موجبات بروز خسارتی را فراهم می‌کند که از این طریق از بیمه‌نامه خود منفعتی کسب کند (تقلب سخت). در این شرایط فرد آگاهانه سعی در ایجاد خسارت یا اغراق در میزان و نوع خسارت دارد. مانند: تصادفات ساختگی، آتش‌سوزی‌های ساختگی، اسناد جعلی و ارائه اطلاعت نادرست.

تقلب نرم و فرصت طلبانه زمانی رخ می‌دهد که بیمه‌گزار به‌صرف داشتن بیمه‌نامه، احتیاط کمتری می‌کند. تقلب در بیمه ماهیت فرصت طلبانه دارد. وجود بیمه و تعهد به جبران خسارت باعث می‌شود بیمه‌گزاران ریسک‌هایی را متحمل شوند که در صورت عدم وجود بیمه از این ریسک‌ها دوری می‌جستند. بدین معنی که ممکن است خسارت سهوا رخ‌داده باش اما اشخاص و یا شرکت عمدا سعی در اغراق میزان خسارت کنند[۷].لازم به ذکر است که شاید همه‌ی مردم مرتکب تقلب سخت نشوند اما اکثریت مردم در هنگام بروز خسارت اقدام به اغراق در میزان خسارت و به‌اصطلاح تقلب نرم می‌شوند.

از بزرگ‌ترین تقلب‌ها در صنعت بیمه ارائه اطلاعات نادرست است. برخی از بیمه‌گزاران اطلاعات غیرواقعی به بیمه‌گر می‌دهند، این دوحالتی است که در صورت ارائه اطلاعات صحیح، بیمه‌گر تصمیمی متفاوت اتخاذ می‌کند.

غفلت شرکت‌های بیمه از مطالعه پدیده تقلب بیمه‌ای دست‌کم، دو دلیل دارد. دلیل اول این است که این پدیده یک جرم و طبیعتاً، فی‌نفسه مذموم است و آشکار کردن این عمل مجرمانه معدودی از بیمه‌گزاران، ضمن آن‌که به لحاظ نمادین، کار ناخوشایندی است این پیام تلویحی را برای سایر بیمه گزاران دارد که

1.Hard Fraud
2 .Soft Fraud

با اقدام کم هزینه فریب شرکت بیمه، به درآمد بادآورده می توان دست‌یافت. دلیل دوم این است که درگذشته، پیامدهای مالی تقلب‌های بیمه‌ای در آن حد و اندازه نبوده که ارزش بررسی و تلاش برای یافتن راه‌حل‌های ممکن را داشته باشد. اما در چند سال اخیر این وضع دگرگون‌شده و شرکت‌های بیمه به اهمیت مطالعه عوامل تقلب بیمه‌ای و کشف آن توجه کرده‌اند.[۸]

تقلب و کلاه‌برداری امروزه به یکی از سرسخت‌ترین دشمنان بیمه تبدیل‌شده است. تقلبات در بخش اموال و حوادث در آمریکا سالانه ۳۰ میلیارد دلارو در انگلیس ۱ میلیارد پوند به شرکت‌های بیمه ضرر وارد می‌کند. کلاه‌برداری جرمی پیچیده و دشواری است که فقط افراد ماهر و آگاه یا گروه‌های سازمان‌یافته می‌توانند مرتکب آن شوند. اما کلاه‌برداری و تقلب بیمه‌ای برخلاف سایر جرم‌ها بین تمام گروه‌های قومی، سطوح درآمدی مختلف، همه‌ی سطوح تحصیلی و تمامی مناطق جغرافیایی وجود دارد.

تحقیقاتی که انجمن بیمه گران بریتانیا انجام داده است نشان می‌دهد تقلب و سوءاستفاده از بیمه بخشی از ناآگاهی و عدم شناخت مردم درباره آنچه "درست" است می‌باشد. هدف اصلی این تحقیق سنجش دیدگاه مردم در خصوص ادعاهای تقلبی در صنعت بیمه بود. هدف دیگری که از طراحی این تحقیق دنبال می‌شد این بود که تقلب و سوءاستفاده را جز اقدامات خلاف قانون در جامعه مطرح کند.

یکی از یافته‌های تحقیق این بود که از میان شرکت‌کنندگان در این تحقیق ۶ درصد از آن‌ها به تقلب در بیمه اذعان کرده بودند.۲ درصد از شرکت‌کنندگان نیز به ادعای خسارت ساختگی اذعان کرده بودند. این نتایج بدان معنا است که حدود ۳/۳ میلیون نفر در انگلیس مرتکب تقلب بیمه‌ای شده‌اند. به دلیل این‌که برخی از افراد ممکن است به‌درستی عملکرد خود اذعان نکنند، آمار مربوط به تقلب ممکن است بیشتر باشد. این تحقیق همچنان مشخص کرد که بخش عمده‌ای از گروه‌های بالای اجتماعی-اقتصادی جامعه مرتکب تقلب بیمه‌ای می‌شوند. این یافته برای بیمه گران اهمیت بسیار زیادی دارد زیرا چنین گروه‌هایی خریداران اصلی بیمه‌نامه‌ها بشمار می‌آیند.[۹]

برای دستیابی به درک بهتری از دیدگاه‌ها و رفتارهای کسانی که مرتکب تقلب می‌شوند، انجمن بیمه گران بریتانیا مطالعات گسترده‌تری انجام داد. آنان افراد موردمطالعه را با توجه به نوع تقلب و دیدگاه‌ها و انگیزه‌ی آن‌ها به ۵ دسته تقسیم کرد.[۹]

سوءاستفاده‌چی‌ها: این گروه بیشترین آمادگی را برای سوءاستفاده و تقلب‌های بیمه‌ای و حتی طرح ادعاهای ساختگی داشتندوآنان سعی می‌کردند تا آنجا که می‌توانند از شرکت‌های بیمه سوءاستفاده کنند. سوءاستفاده‌چی‌ها بسیار گستاخ و بی‌پروا، خودرأی و در برخی موارد ناآرام و بی‌قرارند.

بازیگران: این افراد آمادگی مبالغه و گزاف‌گویی درباره‌ی میزان خسارت وارده را دارنده شرایط بیمه‌نامه‌ها را به‌درستی رعایت نمی‌کنند. البته باید گفت که این گروه ادعای دروغین و ساختگی طرح نمی‌کنند. این افراد بیمه را نوعی بازی می‌دانند که باید در آن پیروز شد. این افراد را مردان برون‌گرا و بانشاط تشکیل می‌دهند ولی همانند گروه نخست (سوءاستفاده‌چی‌ها) عصبانی و متقلب نیستند.

انتقام جویان: این افراد در خصوص بیمه تجربه‌های منفی داشته‌اند و یا ادعای خسارت آن‌ها مردود شده ویا کاهش پیداکرده است.

بنابراین آنان از هر فرصتی استفاده می‌کنند تا آنچه را که می‌پندارند که حق آن‌هاست از بیمه بازستانند. البته این گروه ادعای تقلبی مطرح نمی‌کنند اما در پرونده‌های صحیح در اعلام میزان خسارت اغراق می‌کنند. انتقام جویان معمولاً محتاط‌تر از دو گروه قبلی (سوءاستفاده‌چی‌ها و بازیگران) هستند ولی تندخوتر و برآشفته‌تر از آنان بوده و احساس می‌کنند که ازنظر مالی و اجتماعی تحت‌فشار هستند.

محتاطان: این گروه در مبالغه و بیشتر از مقدار واقعی اعلام کردن خسارت مردد هستند (البته آن را تقلب یا سوءاستفاده نمی‌دانند مگر این‌که به آن‌ها گفته شود) دیگران (مانند همکاران و شرکا) ممکن است آنان را تشویق به مبالغه در اعلام میزان خسارت وارده نمایند. افراد این گروه معمولاً زنان هستند و درباره آنچه انجام داده‌اند احساس گناه می‌کنند اما دیگران آن‌ها را گمراه کرده‌اند.

۲-۲-۳- انواع تقلبات بیمه‌ای

در شرکت‌های بیمه انواع تقلب‌ها به‌صورت زیر دسته‌بندی می‌شوند[۹]:

- تقلب بیمه‌گزار: تقلب علیه شرکت بیمه در خرید یا ابطال خدمات بیمه توسط یک شخص ویا افرادی که باهم برای به دست آوردن پوشش غیرقانونی ویا ادعای خسارت ساختگی تبانی می‌کنند.
- تقلب در نمایندگی‌ها: تقلب توسط واسطه‌ها علیه شرکت بیمه ، بیمه‌گزاران، مشتریان ویا سایر ذینفعان.
- تقلب داخلی: تقلب علیه شرکت بیمه توسط یک عضو هیئت‌مدیره، مدیر ارشد و یا عضو دیگر از کارکنان خود شرکت

۲-۲-۳ -۱-تقلب بیمه‌گزار

این نوع تقلب که دردسته ی تقلب‌های خارجی قرار دارد گاها به شکل بیمه‌گزار احتمالی و یا یک بیمه‌گزار بالفعل ظاهر می‌شوند. بیمه‌گزار می‌تواند در همه‌ی مراحل قرارداد چه در مرحله عقد بیمه‌نامه و چه در طول قرارداد و چه در زمان ادعای خسارت مرتکب تقلب شود شرکت‌های بیمه باید بتوانند ریسک تقلب در خدمات موجود و حتی خدمات جدید خود را ارزیابی کنند؛ و حتی ممکن است با اشخاص

دیگری هم تبانی کند مثلاً با پزشکان، افسران ترسیم‌کننده کروکی، رانندگان جرثقیل و دیگر افرادی که به‌نوعی درگیر با وقوع و صحنه حادثه هستند.

نمونه‌هایی از تقلبات بیمه‌گزاران:

- صحنه‌سازی تصادف
- گزارش و ادعای خسارت ساختگی
- اغراق در خسارات وارده
- خرید بیمه‌نامه بدنه از چندین شرکت بیمه
- جابه‌جایی راننده حادثه با فردی دیگر
- فریب نمایندگی و یا تبانی با آن‌ها برای خرید بیمه‌نامه بدنه برای ماشین تصادفی
- و ...
-

۲-۲-۳-۲ -تقلب در نمایندگی‌ها:

نمایندگان بیمه بین خریداران و فروشندگان بیمه قرار می‌گیرند. نمایندگان نقش حساسی را درفروش خدمات بیمه ایفا می‌کنند چراکه آنان دقیقا بین طرفین یعنی دقیقا جایی که اعتماد نقش مهمی بازی می‌کند قرار دارند. در هر معامله‌ای اعتماد حرف اول را می‌زند و از عناصر اصلی هر معامله است، درست در همین‌جاست که شرکت‌های بیمه احساس خطر می‌کنند و آسیب‌پذیرند. چراکه ممکن است نمایندگی‌ها از این اعتماد سوءاستفاده کرده و به نفع خود و متضرر شدن بیمه، عملی انجام دهد. مثل:

- نمایندگی با علم به این‌که مورد بیمه‌ای وجود ندارد، بیمه‌نامه صادر می‌کند درصورتی‌که بیمه‌گزار حق بیمه اولیه خود را پرداخت کرده است، حق کمیسیون خود را می‌گیرند و سپس بیمه‌نامه را ابطال می‌کنند.

- تبانی با بیمه‌گزاران متقلب در طرح ادعاهای جعلی مانند این‌که تاریخ صدور بیمه‌نامه را دست‌کاری کنند ویا ارائه اطلاعات نادرست به بیمه‌گر.
- نگه‌داشتن حقبیمه‌های بیمه‌گزاران تا زمانی که ادعای خسارت کنند.
- تبانی با بیمه‌گزار درصدور بیمه‌نامه بدنه برای خودروهای تصادفی.
- و...

برای کنترل و نظارت بر تقلب‌های نمایندگی‌ها به چند موضوع باید توجه داشت:

- رابطه فامیلی و آشنایی نزدیک‌بین بیمه‌گزار و نمایندگی و جود داشته باشد.
- نمایندگی با سبد سرمایه‌گذاری کوچک تعداد زیادی بیمه‌نامه صادر کرده و فروخته باشد.
- بیمه‌گزاران خارج از حوزه کاری نمایندگان زندگی کنند.
- تغییرات متوالی در مالکیت و مدیریت نمایندگی‌ها
- نمایندگی‌ها اغلب شماره تلفن و نشانی خود را تغییر دهند.

برای مقابله با این دسته از تقلب‌ها باید سیاست‌ها و قوانینی جدیدی وضع شود، برگزاری آموزش‌های دوره‌ای، فرستادن دوره‌ای ممیزان به نمایندگی‌ها جهت بررسی و کنترل فرآیندهای کسب‌وکاری آن‌ها.

۲ -۲-۳-۳- تقلب داخلی

به‌طورمعمول هر کسب‌وکاری همیشه با تقلب داخلی چه در سطوح بالای مدیریتی و چه در سطوح پایین مواجه‌اند. تقلب داخلی بخشی از مدیریت ریسک عملیاتی است و خطری برای شهرت شرکت‌های بیمه محسوب می‌شود و حتی در مواردی می‌تواند تباهی اقتصادی شرکت را تسریع بخشد. فاکتورهایی که بر آسیب‌پذیری بیمه‌ها از جهت تقلب داخلی تأثیر می‌گذارد شامل موارد زیر هستند:

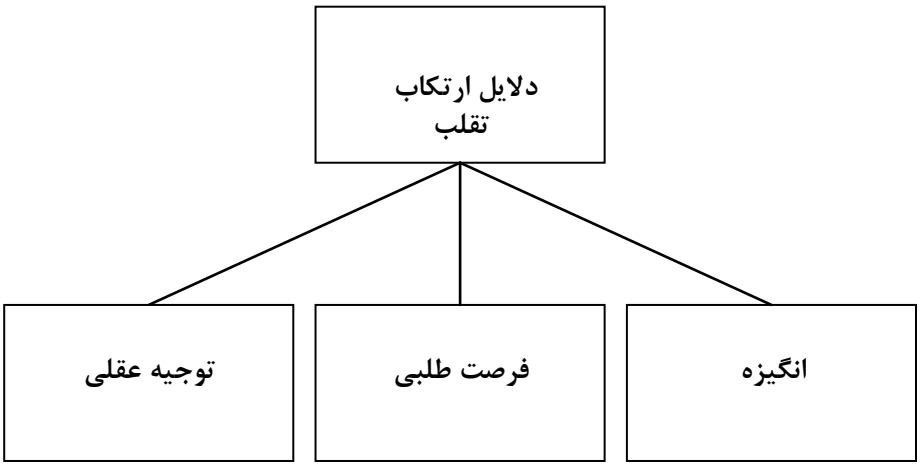
- پیچیدگی در ساختار سازمانی سازمان‌ها
- سرعت زیاد در نوآوری، توسعه محصولات و کامپیوتری شدن
- ایجاد جو همکاری و هماهنگی بین مدیران ارشد و هیئت‌مدیره و کارکنان.
- حفظ نظارت کافی بر مدیران و کارکنان
- نوشتن شرح شغل و مسئولیت‌ها به‌طور صریح
- صدور دستورالعمل‌های رفتار اخلاقی برای مدیران و کارکنان
- انجام غربال‌گری مدیران و کارکنان دائمی و موقت از قبل و در حین استخدام
- تفکیک وظایفی که مستعد تضاد منافع در شرکت‌ها هستند.
- و...

۲-۲-۴-دلایل ارتکاب تقلب

به‌طورکلی جهت شناسایی انگیزه‌ها و محرک‌های ارتکاب تقلب، بهتر است اجزای مثلث تقلب را بررسی کنیم.

در تقلب به فرصت برای تقلب یا مقدار زیادی سختی و فشار و توجیه عقلی نیاز است تا فرد انگیزه ارتکاب تقلب را پیدا کند.[۱۰]

شکل (۲-۲) سه جزء اصلی که در وقوع تقلب نقش حیاتی دارند، نشان می‌دهد. عناصر تقلب دارای ارتباط متقابل هستند.



شکل (۲-۲). دلایل ارتکاب تقلب

انگیزه: ارتکاب تقلب نیازمند انگیزه است و در بسیاری از موارد این سختی‌ها و مشکلات مالی انگیزه ساز تقلب می‌شوند. ازاین‌رو شرکت‌های بیمه می‌بایست شرایط و پتانسیل‌های زمینه‌ساز تقلبات را شناسایی کنند و به دنبال نشانه‌ای از بروز تقلب احتمالی باشند.

فرصت‌طلبی: اگر فردی بداند که احتمال کشف و شناسایی تقلب او بسیار کم است، حتی بدون شرایط فشار و سختی از فرصت‌های به‌دست‌آمده برای ارتکاب تقلب استفاده خواهد کرد. متقلبان به‌احتمال‌زیاد درزمانی و به شکلی عمل می‌کنند که احتمال تشخیص و شناسایی تقلب را اندک بدانند؛ بنابراین شرکت‌های بیمه باید سیاست‌ها و کنترل‌های ویژه‌ای را اعمال کنند تا علاوه بر اینکه از وقوع تقلب در ابتدای امر جلوگیری می‌کنند همچنین بتوانند در صورت بروز تقلب آن را شناسایی کنند.

توجیه عقلی: افراد متقلب معمولاً به دنبال یک توجیه عقلی برای عملشان هستند؛ و ممکن است به بسیاری از دلایل که ریشه در اصول و باورهای آن‌ها دارد، کلاه‌برداری از بیمه را امری غیراخلاقی و غیرقانونی تلقی نکنند. مثل باورهای زیر[۱۱]:

- اغلب افراد خدمات بیمه‌ای را به جبران احتمال ناشناخته‌ای که امکان بروز حادثه در آینده وجود دارد، خریداری می‌کنند. بااین‌حال حق بیمه پرداختی اتومبیل اکثراً بیش‌ازحدی است که بیمه‌گزار متصور شده است و پرداخت آن در سال‌های متمادی درصورتی‌که طی این سال‌ها از بیمه ادعای خسارتی نداشته باشد، نوعی موقعیت برد_ باخت را در ذهن بیمه‌گزار متصور می‌کنند؛ بنابراین نگرش " من آن چیزی را می‌گیرم که بابتش هزینه پرداخت کرده‌ام " در ذهن بیمه‌گزار نقش می‌بندد .پس در جستجوی موقعیتی برمی‌آید تا بتواند آنچه را طی سالیان تحت عنوان حق بیمه پرداخت کرده است از طریق ادعایی دروغین و متقلبانه باز پس گیرد.
- نگرش " اگر می‌توانید من را بگیرید" ویا "من برنده‌ام اگر شما نتوانید من را بگیرید" از موارد دیگری برای توجیه عمل متقلبانه ازنظر متقلبان است. شایان توجه است که حجم پرونده‌های خسارت دریافتی بسیار زیاد است و بررسی پرونده‌های مشکوک زمان‌بر و هزینه‌بر هستند و از طرفی اثبات تقلب نیازمند شواهدی است که در برخی موارد اطلاعاتش در دسترس نیست ویا قابلیت پیگیری ندارند. همچنین بسیاری از پرونده‌های مشکوک نیازمند همکاری نزدیک با سایر ارگان‌ها مثل نیروی انتظامی، مراکز درمانی، ثبت‌احوال و سایر شرکت‌های بیمه هستند و ازاین‌رو شناسایی و کشف همه‌ی موارد مشکوک به تقلب در عمل امکان‌پذیر نیست و در هر برهه‌ی زمانی درصدی از تقلب‌ها کشف نمی‌شوند.
- یک باور دیگر در این زمینه به‌صورت "اگر هم مرا بگیرید من چیزی را از دست نمی‌دهم"است و این تصور شاید به این دلیل است که افرادی که در بیمه تقلب می‌کنند معمولاً در بین مردم مجرم شناخته نمی‌شوند. همان‌طور که طی یک بررسی در انگلستان مشخص شد که، مردم دزدیدن یک بسته شکلات را عملی تبهکارانه‌تر از صحنه‌سازی برای خسارت بیمه‌ای دروغین می‌دانند. از طرفی چون مجازات تعریف‌شده در قانون برای کلاه‌برداران بیمه‌ای به‌اندازه کافی بازدارنده نیست و همچنین پیگیری‌های قانونی هزینه‌بر و زمان‌بر هستند و حتی ممکن است شهرت شرکت را خدشه‌دار کنند، شرکت‌های بیمه

راغب می‌شوند که تقلبات نرم رابین خود و بیمه‌گزار حل‌وفصل کنند و یا حتی در بعضی موارد تقلبات سخت را نیز با مذاکره برطرف می‌کنند و به مراجع قانونی ارجاع نمی‌دهند؛ و همین امور نیز فرصت را برای افراد متقلب و سودجو بیش‌ازپیش فراهم می‌نماید.

- این نگرش که " تا می‌توانید بخورید" ازجمله باورهای ناشایستی است در بین مردم نقش بسته است. این باور افرادی است که معتقدند تا زمانی که پول‌دارند باید به هر نحوی که شده آن را خرج کنند مخصوصاً اگر آن پول مربوط به دولت یا شخص دیگری باشد که این باور و اعتقاد مفهوم بیمه‌ای " در هر سرمایه‌گذاری در بیمه باید ریسک به‌طور مساوی بین دو طرف تقسیم شود " را زیر سؤال می‌برد؛ بنابراین این افراد اگر در موقعیتی قرار بگیرند که بتوانند ادعای خسارت کنند (چه قانونی و تقلبی و چه در یک خسارت واقعی جزئی یا هنگفت) تمام سعی خود را می‌کنند تا از حداکثر فرصت به وجود آمده استفاده کنند و حق بیمه‌ای که طی سالیان پرداخت کرده‌اند را یکجا از بیمه دریافت کنند.

با توجه به آنچه گفته شد مشخص می‌شود که تمامی این باورهای غلط ناشی از فقر فرهنگی جامعه است و درک اشتباه عموم مردم و سیستم عدالت کیفری از ماهیت بیمه باعث می‌شود که برچسب غیراخلاقی و غیرقانونی به تقلب بیمه‌ای نزنند و میدان را برای افراد متقلب بازتر کنند.

۲-۲-۵- بیمه شخص ثالث

این نوع بیمه مربوط به سرنشینان خودرو بیمه‌شده هست و بیمه‌گزار (مالک خودرو) را شامل نمی‌شود. و کلیه خسارت‌هایی که به‌واسطه‌ی خودرو یا بار آن به سرنشینان وارد می‌شود را تحت پوشش قرار می‌دهد. بند ۱ از این بیمه که به قانون مبدل گشته به‌صورت زیر است:

" کلیه دارندگان وسایل نقلیه موتوری زمینی و ریلی اعم از این‌که اشخاص حقیقی یا حقوقی باشند مکلف‌اند وسایل نقلیه مذکور را در قبال خسارت بدنی و مالی که در اثر حوادث وسایل نقلیه مزبور و یا یدک و تریلر متصل به آن‌ها ویا محصولات آن‌ها به اشخاص ثالث وارد شود، نزد یکی از شرکت‌های بیمه که مجوز فعالیت در این رشته را از بیمه مرکزی ایران داشته باشد، بیمه نمایند. کلیه زیان دیدگان ناشی از وسایل نقلیه موتوری زمینی (اعم از جانی و مالی) خواه اشخاص حقیقی یا اشخاص حقوقی (شرکت‌ها) باشند، شخص ثالث نامیده می‌شوند و خسارت‌های مالی وارده به کلیه اموال منقول و غیرمنقول متعلق (تحت مالکیت) اشخاص ثالث جبران می‌گردد. هر زیان دیده‌ای که آسیب ببیند چه جانی و چه مالی شخص ثالث است به‌استثناء راننده مسئول حادثه که به‌عنوان سرنشین محسوب می‌گردد "[۱۲].

این بیمه از سال ۱۳۴۷ اجبار است و درصورت نداشتن بیمه‌نامه شخص ثالث ویا تأخیر در تمدید آن جریمه لحاظ خواهد شد. بیمه‌نامه شخص ثالث به‌طورکلی ۴ نوع تعهد دارد:

۱. خسارت مالی: جبران خسارت‌های مالی به اشخاص ثالث در اثر تصادف، آتش‌سوزی، انفجار و کلیه خسارت‌های مالی در اثر بار خودروهای مجاز به حمل بار.

۲. صدمات بدنی و فوت: منحصراً مربوط به آسیب جانی و صدمات وارده به جسم آدمی است نه موجودات دیگر بلکه فقط انسان. خسارت جانی شامل: دیه جرح، دیه فوت، دیه نقص و هزینه درمان می باشد.

۳. فوت راننده: در صورت فوت راننده مقصر

۴. هزینه پزشکی راننده: در صورت مصدویت و نقص عضو.(به راننده دیه تعلق نمی‌گیرد) لازم به توضیح است که سایر سرنشینان به‌جز راننده، شخص ثالث حساب می‌شوند.

شاید مناسب‌ترین رشته بیمه برای بررسی تقلب بیمه‌ای و ابعاد مختلف آن در کشور ما، بیمه شخص ثالث باشد. زیرا این رشته ازنظر سهم بری در پرتفوی حق بیمه دریافتی شرکت‌های بیمه در بسیاری از کشورها،

ازجمله ایران رتبه اول را دارد. همچنین به علت زیان ده بودن این رشته در کشور ما بسیاری از کشورهای دیگر، مقابله با تقلب‌های بیمه‌ای که باعث کاهش خسارت‌های پرداختی و درنتیجه، کاهش زیان عملیاتی در بیمه شخص ثالث می‌شود اهمیت زیادی پیداکرده است.

۲-۲-۶- انواع تقلبات بیمه شخص ثالث

تقلب در بیمه شخص ثالث در طیفی از ترفندها اتفاق می‌افتد که از زیاد جلوه دادن خسارت واقعی آغاز می‌شود و تا حد ادعای خسارتی کاملاً ساختگی و دروغین پیش می‌رود. تقلب نوع اول ازنظر مبلغ خسارت دریافتی کم‌ارزش است و بیشتر تقلب‌ها در بیمه شخص ثالث در این رده جا می‌گیرند. در مقابل، خسارت‌های دروغین ازنظر تعداد نسبتاً اندک‌اند اما خسارت‌های دریافتی در آن‌ها نسبتاً سنگین است[۱۳]. به نظر مارسد خصوصیت عمده تقلب نوع اول این باشد که بسیاری از بیمه‌گزارانی که مرتکب آن می‌شوند این تقلب را با این استنباط که حقبیمه‌های پرداختی آن‌ها در سال‌هایی که خسارت نداشته‌اند جبران شود یا خسارت دریافتی لطمه‌ای به خزانه‌ی پرپول شرکت بیمه وارد نمی‌کند، ازلحاظ اخلاقی این عمل خود را توجیه می‌کنند.

درحالی‌که خسارت‌های نوع اول را عامه بیمه‌گزاران از شرکت بیمه مطالبه می‌کنند، خسارت‌های نوع دوم را اشخاص یا گروه‌هایی مدعی می‌شوند که اخذ خسارت دروغین را به‌صورت سازمان‌یافته و ممر درآمد بادآورده، خود قرار داده‌اند.

نوعی دیگر تقلب در بیمه شخص ثالث خریداری بیمه‌نامه‌های مضاعف است. در این حالت بیمه‌گزار، به‌قصد کسب درآمد بادآورده هم‌زمان از چند شرکت بیمه برای اتومبیل خود بیمه شخص ثالث خریداری می‌کند. در نوع دیگر تقلب، هنگام ترسیم کروکی صحنه‌ی تصادف جای مقصر و خسارت‌دیده عوض می‌شود یا تاریخ حادثه در کروکی جلو برده می‌شود تا مقصر بتواند بیمه‌نامه شخص ثالث خریداری کند و از آن برای پوشش خسارت واقع‌شده استفاده کند.

نوع دیگر تقلب در رشته شخص ثالث آن است که بیمه‌گزاران با ترتیب دادن صحنه حوادث ساختگی، شرکت بیمه را ناچار به پرداخت غرامت بابت اشخاص جراحت‌دیده یا فوت‌شده براثر حوادث غیر رانندگی یا عوامل طبیعی می‌کنند. علاوه بر موارد فوق موارد ذیل نیز از تقلب‌هایی است که در بیمه شخص ثالث اتفاق می‌افتد.

- جابه‌جایی فرد دارای گواهینامه رانندگی با راننده عامل حادثه که فاقد گواهینامه رانندگی است.
- سرقت کوپن بیمه‌نامه شخص ثالث یک اتومبیل برای استفاده از آن در تصادفات اتومبیل‌های دیگر.
- مبادرت راننده اتومبیل دارای بیمه‌نامه بدنه و شخص ثالث به مقصر بودن خود از حادثه رانندگی باهدف دریافت خسارت بدنه‌ی اتومبیل خود از طریق بیمه‌نامه بدنه آن.

۲-۲-۷- استراتژی‌های شرکت بیمه جهت برخورد با تقلب

شرکت‌های مختلف بیمه استراتژی‌های گوناگونی در برخورد با تقلبات بیمه‌ای خوددارند. دسته‌ای از شرکت‌ها خسارت‌های جعلی را به‌عنوان بخشی از هزینه‌های عملیاتی خود تلقی می‌کنند و عملاً اقدام خاصی را جهت مقابله با تقلبات انجام نمی‌دهند. این شرکت‌ها با چشم‌پوشی از تقلبات و پرداخت سریع خسارت شاید در کوتاه‌مدت به جلب رضایت مشتریان و افزایش پرتفوی شرکت بپردازند اما در بلندمدت عواقب ناگواری را متحمل صنعت بیمه خواهند کرد.

در طرف دیگر دسته‌ی دیگری از شرکت‌های بیمه هستند که اساساً تقلب را یک پدیده ناشایست فرهنگی و اجتماعی می‌دانند که بر جامعه و صنعت بیمه اثرات مخربی می‌گذارد و ازاین‌رو جلوگیری از تقلبات بیمه را به‌عنوان یک استراتژی و فرهنگ‌سازمانی موردتوجه قرار می‌دهند و با رویکرد پیشگیری، شناسایی و کشف تقلب اقداماتی را در پیش می‌گیرند که به دیگر شرکت‌ها نیز می‌تواند در الگوبرداری از این فرآیندها برای جلوگیری از اقدامات متقلبانه کمک شایانی کند

همچنین در ذیل برخی از راهکارهای عملیاتی برای شرکت‌های بیمه‌ای که به‌عنوان متولی امر بررسی و پرداخت خسارت، می‌توانند برای پیشگیری و کشف تقلبات در پیش گیرند اشاره می‌شود.

- تشکیل اداره تحقیق ویژه در شرکت‌های بیمه: شرکت‌های بیمه‌ای باید برای مبارزه با کلاه‌برداری‌های بیمه‌ای، نسبت به راه‌اندازی این اداره در شرکت، اقدام نمایند. شرکت‌های بیمه‌ای باید با به‌کارگیری کارشناسان زبده، اقدام به تشکیل و راه‌اندازی این اداره نمایند. این اداره با بررسی‌ها و تحقیقات میدانی در خصوص پرونده‌هایی که توسط ارزیابان و کارشناسان خسارت، مشکوک به تقلب تشخیص داده‌شده‌اند، صحت‌وسقم حادثه و سایر موارد مشکوک به کلاه‌برداری را بررسی نموده و مدارک و شواهد لازم برای اقامه دعوا توسط واحدهای حقوقی شرکت‌ها در دادگاه را تهیه نموده تا از پرداخت خسارت‌های جعلی به کلاه‌برداران جلوگیری شود.
- استفاده از کارشناسان ارزیاب خسارت با داشتن اطلاعات و تخصص بیمه‌ای: یکی از مهم‌ترین مراحل فرآیند تسویه خسارت، ارزیابی پرونده توسط کارشناسان خسارت می‌باشد. متأسفانه شرکت‌های بیمه‌ای به علت نداشتن کارشناسان ارزیاب خسارت خبره و باتجربه نمی‌توانند به‌درستی، تمام موارد مشکوک به تقلب و تخلف را تشخیص دهند. نتیجه اینکه بسیاری از پرونده‌های جعلی خسارت، قابل پرداخت تشخیص داده‌شده و علاوه بر زیان وارده به شرکت‌های بیمه، کلاه‌برداران را برای تشکیل پرونده‌ای جعلی و دریافت خسارت‌های ساختگی ترغیب می‌کند. ازاین‌رو استفاده از کارشناسان ارزیاب پرونده خسارت که تخصص بیمه‌ای داشته باشند می‌تواند به‌خوبی منجر به تشخیص پرونده‌های جعلی شود.
- لزوم نظارت بر مراکز/شعب پرداخت خسارت برای ثبت دقیق پرونده‌ها: شرکت‌های بیمه‌ای باید با تدوین دستورالعمل‌ها و آیین‌نامه‌های لازم، نظارت کافی خود را بر مراکز/شعب پرداخت خسارت اعمال

نمایند تا این مراکز مدارک لازم و مستندات لازم را به‌درستی دریافت نمایند تا ضمن جلوگیری از ارجاعات متعدد برای تکمیل پرونده، امکان بررسی بهتر پرونده‌های مشکوک نیز فراهم شود.

- لزوم آموزش‌های لازم برای کارشناسان و ارزیابان خسارت: با توجه به اینکه برخی کارشناسان و ارزیابان خسارت افراد کم‌تجربه هستند و نیز با توجه به ماهیت پویای تقلب و روش‌های نوین که به‌مرورزمان ایجاد می‌شود، شرکت‌ها باید آموزش‌های لازم و کارگاه‌های آموزشی مستمر و منسجمی را برای کلیه کارشناسان خسارت برگزار نمایند.
- غربالگری واسطه‌های بیمه: واسطه‌ها مانند کارگزاران، نمایندگان و کارکنان نقش کلیدی در به دست آوردن و حفظ مشتری بر عهده‌دارند. این بدین معنی است که کارگزاران باید قبل از شروع فروش محصولات بیمه غربال شوند. آن‌ها همچنین باید دائماً تحت نظارت باشند و همانند تقلب از سوی بیمه‌گزاران، تقلب و رفتارهای غیراخلاقی آن‌ها نیز به‌هیچ‌وجه تحمل نشود؛ بنابراین انتخاب و غربال کارگزاران باید نه‌تنها بر روی مهارت‌های فنی و حرفه‌ای آن‌ها تمرکز یابد، بلکه باید شامل سوابق مالی و اخلاقی آن‌ها نیز باشد.[۱۴]

۲-۲-۸- پیامدهای تقلب و ضرورت شناسایی آن

بر اساس گزارش اداره تقلب بیمه‌ای[1] مبلغی حدود ۲/۱ بیلیون دلار بابت تقلب در بیمه در هرسال هزینه می‌شود که این مبلغ خود به ۳ بخش تقسیم می‌شود.۱/۷ بیلیون دلار به آن دسته از تقلب‌هایی که کشف نمی‌شوند و ۳۵۰ میلیون دلار به صحنه‌سازی‌های سازمان‌یافته در تصادفات اتومبیل و ۳۸ میلیون دلار بابت تقلب‌هایی که کشف‌شده‌اند، اختصاص می‌یابد. اثرات منفی تقلب به‌کارگیری سیستم‌های شناسایی تقلب را توجیه‌پذیر ساخته است. بااین‌وجود کشف تقلب برای پیشگیری از پیامدهای نابودکننده‌ی تقلب‌های مالی حیاتی هستند و رویکردهای سنتی برای این امر کافی نیستند. تحقیقات اخیر نشان داده

1. Insurance fraud bureau

است که بررسی ادعاهای مشکوک به تقلب باعث کاهش پرداخت به ادعاهای دروغین تا حدود ۱۸٪ می‌شود که خود صرفه‌جویی منابع شرکت را به دنبال دارد و از سویی تأثیر مستقیم بر سودآوری شرکت می‌گذارد.[۱۵]

۲-۲-۹- محدودیت‌های بیمه در راستای پیگیری تقلبات

شرکت‌های بیمه می‌توانند با آگاهی از انواع تقلبات و فرآیندهایی که احتمال بروز تقلب در آن‌ها وجود دارد سیستم هشداردهنده و پیشگیرانه‌ای را طراحی کنند و با آگاهی از میزان آسیب‌پذیری خود استراتژی‌های مؤثرتری را به‌کارگیرند؛ اما برای تحقق این امور شرکت‌های بیمه با محدودیت‌ها و پیچیدگی‌های زیادی مواجه‌اند مانند[۱۶]:

- پنهان بودن ماهیت تقلب
- پویایی و حساسیت به تغییر در تقلب (هنگامی‌که یک سبک تقلب شناسایی می‌شود، کلاه‌برداری با سبک دیگری درحالی‌که شکل‌گیری است.)
- عدم توافق اجماع برآن چه که واقعاً به‌منزله‌ی تقلب در بیمه است. و آن شکلی از تقلب که برآن متمرکز می‌شویم.
- نحوه‌ی عملکرد شوراهای حل اختلاف در مورد ترسیم کروکی فرضی بر اساس اظهارات افراد ذینفع (بعد از وقوع حادثه و به هم خوردن صحنه تصادف) توسط کارشناسان رسمی دادگستری (بدون تحقیق و تفحص)؛
- عدم پذیرش نماینده شرکت بیمه ذینفع در دادگاه‌های ذی‌ربط؛
- عدم تطبیق احکام صادره از سوی دادگاه‌ها با مفاد قانون مجازات اسلامی (دیات)؛

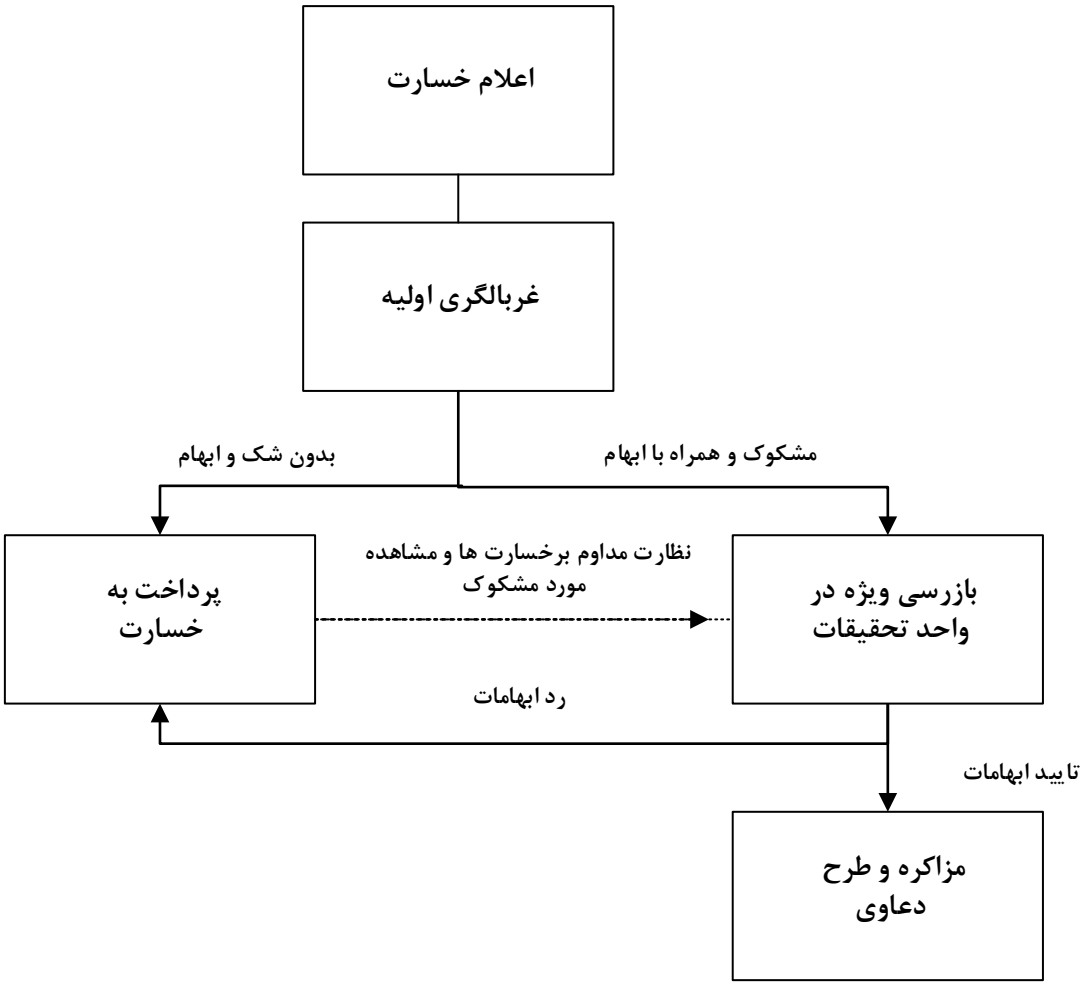
- برخی برآوردهای غیردقیق از میزان صدمات جسمی وارده به مصدوم یا مصدومین رانندگی (عدم تطابق با قانون مجازات اسلامی)؛
- عدم احزار هویت و نیز عدم معاینه مصدومین توسط نهاد ذی‌ربط و استناد صرف به مدارک پزشکی ارائه‌شده؛
- در برخی موارد اقدام به ترسیم کروکی نادرست با نظر طرفین؛
- عدم درج دقیق نحوه استقرار سرنشینان خودروی بارکش در گزارشات مقامات انتظامی؛
- عدم وجود بانک اطلاعاتی مشخص و به‌روز، از متقلبان و کلاه‌برداران بیمه‌ای؛
- ضعف سیستم‌های کنترلی مبتنی بر فناوری اطلاعات؛
- نحوه‌ی تنظیم گزارش‌ها نهادهای درگیر و عدم بیان واقعیت‌های حادثه و افراد درگیر در آن، به‌طور عمدی یا غیرعمدی؛
- عدم پاسخگویی یا پاسخگویی با تأخیر ارگان‌های ذی‌ربط به استعلامات شرکت‌های بیمه‌ای؛
- نبود دستورالعمل و آیین‌نامه مشخص برای پیشگیری و کشف تقلبات و کلاه‌برداری‌های بیمه‌ای؛
- نگاه جامعه به شرکت‌های بیمه‌ای به‌عنوان نهادهای حمایتی.

وجود این محدودیت‌ها باعث شده است که در بسیاری از موارد، شرکت‌های بیمه‌ای نتوانند موارد تقلب و کلاه‌برداری را کشف کنند و در بسیاری از موارد نیز، علی‌رغم کشف تقلب در پرونده و ارائه مستندات و شواهد لازم، این مدارک و مستندات موردقبول واقع نمی‌شود و علی‌رغم وجود جعل، تحریف و تقلب در پرونده بیمه، خسارت‌های جعلی به افراد پرداخت می‌دهد.

۲-۲-۱۰- فرآیند بررسی خسارات

طبق روند بررسی ادعای خسارت در شکل (۲-۳)، شرکت بیمه پس از دریافت گزارشی از بیمه‌گزار خود مبنی بر درخواست خسارت جهت بررسی صحت یا عدم صحت ادعای خسارت، پرونده را غربالگری می‌کند تا آن را دریکی از دودسته‌ی پرونده قانونی یا مشکوک به تقلبی جای دهد. پرونده‌هایی که قانونی تشخیص داده می‌شوند برای پرداخت خسارت به مرحله بعد می‌روند و حداقل هزینه‌ی بررسی و پردازش را شامل می‌شوند؛ اما پرونده‌هایی که در گروه مشکوک به تقلب قرار می‌گیرند، اگر به‌واسطه‌ی مبلغ خسارت ارزش تحقیق و بررسی و صرف هزینه‌های تحقیقات را داشته باشد به واحد تحقیقات ویژه[1] منتقل می‌شوند. در این مرحله پس از بازرسی‌های ویژه‌ی پرونده‌های مشکوک، اگر شک تائید شود و پرونده تقلبی باشد برای پیگیری‌های قانونی به مراکز قانونی سپرده می‌شود. و اگر پرونده‌ی مشکوک رد ابهام شود برای پرداخت خسارت به مرحله بعدی می‌رود.

1 . Specialized Investigation Unit (SIU)



شکل (۲-۳) فرآیند بررسی ادعای خسارت[۱۷]

لازم به ذکر است که معمولاً درصد زیادی از پرونده‌های متقلبانه به دلیل نبود دلایل کافی جهت اثبات تقلب، مورد پیگیری قانونی قرار نمی‌گیرند که دلیل این امر معمولاً پایین بودن ارزش خسارت اغلب پرونده‌های مشکوک به تقلب نسبت به هزینه‌های انجام تحقیقات کشف تقلب است و پرداخت هزینه‌های گزاف برای چنین پرونده‌هایی مقرون‌به‌صرفه و توجیه‌پذیر نیست.

تحقیق و بررسی ممیزان در مرحله‌ی غربالگری بسیار زمان‌بر و فاقد کارایی لازم است زیرا ممکن است بررسی پرونده‌ها متناسب با ماهیتشان آن‌قدر زمان ببرد که شرکت بیمه مجبور به پرداخت خسارت در

موعد مقررشده باشد و نتایج تحقیقات دیگر سودی نداشته باشد. از طرفی محقق در مرحله‌ی غربالگری پرونده‌ها عمدتاً از تجربه، مهارت و خلاقیت خود بهره می‌برد که ممکن است نظرش تحت تأثیر تعصبات ذهنی و شخصی‌اش قرار گیرد و نتواند به‌درستی تشخیصات لازم را انجام دهد. این‌رو رساله حاضر سعی بر این دارد که سیستم مکانیزه‌ی تصمیم‌گیری‌ای طراحی نماید تا از طریق آن تصمیم‌گیری مبنی بر تقلبی بودن یا نبودن پرونده‌ها در مرحله‌ی غربالگری با سرعت و دقت بالایی انجام گیرد و به‌دوراز هرگونه تعصب و سوء نگری باشد و همچنین به دلیل پایایی نتایج، این سیستم می‌تواند کمک بسیار زیادی به ممیزان در مرحله‌ی غربالگری نماید و به‌هیچ‌وجه جایگزین ممیزان در این مرحله نیست و تنها کمک‌کننده به آن‌هاست.

محققان در راستای بررسی پرونده‌های مشکوک به تقلب می‌توانند از نوع رفتار بیمه‌گزاران اطلاعات مفیدی جهت تشخیص صحیح پرونده استفاده نمایند. یک تحقیقی که روی مجموعه‌ای از تراکنش‌های واقعی صورت گرفته، مشخص کرد که اغلب تقلب‌ها دارای ویژگی‌های رفتاری مشابهی هستند. برای مثال، برخی از رفتارهایی که دال بر تقلب در بیمه بدنه هستند عبارت‌اند از:

- حادثه در مکانی خارج شهر و در نیمه‌های شب رخ می‌دهد.
- مدعی خسارت در مراحل تحقیق و بررسی همکاری نمی‌کند.
- شدت حادثه خودروهای درگیر باهم همخوانی ندارند.
- و...

این‌گونه رفتارهای بیمه‌گزاران می‌توانند به‌عنوان رفتارهای مشکوک در نظر گرفته شوند و در صورت بروز مجدد به‌عنوان رفتار متقلبانه لحاظ شوند.[۱۸]

۲-۲-۱۱- تکنیک های کشف تقلب

با استفاده از تکنیک کشف تقلب می‌توان تقلب‌های انجام‌شده را شناسایی کرد و از طریق تجزیه‌وتحلیل آن‌ها به‌پیش ینی رفتارهای آتی مشتریان پرداخت و ریسک انجام تقلبات را کاهش داد. از نتایج حاصل از بکار بستن تکنیک‌های ضد متقلبانه در شرکت‌های بیمه می‌توان به موارد زیر اشاره کرد:[۱۹]

بهبود امکانات و تسهیلات در انجام غربالگری ادعانامه‌ها، آموزش ویژه برای کارکنان و کارکنان که درگیر فرآیند تحقیق و بررسی هستند، سرمایه‌گذاری بر مهارت‌های تخصصی محققان و متخصصان حوزه، بهبود ارتباطات و همکاری بین صنعت بیمه و نیروهای آگاهی و پلیس برای پیگیری‌های قانونی و بهبود و توسعه‌ی ممیزان داخلی، ادغام اطلاعات در سراسر بخش‌ها و اطمینان از پروتکل‌های مناسب برای به اشتراک‌گذاری اطلاعات با پلیس و محققان.

در ادامه به معرفی برخی از تکنیک‌های مختلف شناسایی و کشف تقلب می‌پردازیم:

۲-۱۲-۱-سیستمهای خبره

منظور از سیستم‌های خبره، سیستمهای محاسباتی است که توانایی ارائه و استدلال در برخی از حوزه‌های غنی دانش با نگاه حل مشکلات و دادن راهکار داشته باشد[20]

آشکارسازهای سیستمهای خبره، دانش را در قالب قانون اگر ـ آنگاه رمزگذاری می‌کنند. به این معنی که به کمک قانون اگر ـ آنگاه مشخص می‌کنند در چه شرایطی چه اتفاقی می‌افتد

.به‌عنوان یک مثال سیستم NIDES که توسط شرکت SRI پیاده‌سازی شده است، از رویکرد سیستمهای خبره به‌منظور شناسایی حمالت به کمک مانیتورینگ برخط فعالیت‌های کاربران استفاده می‌کند. سیستم NIDES شامل اجزای ۳۳ تجزیه‌وتحلیل آماری به‌منظور تشخیص ناهنجاری و همچنین ابزار تجزیه‌وتحلیل قواعد ۳۴ به‌منظور تشخیص سوءاستفاده می‌باشد[۲۱].

۲-۱۲-۲-رویکرد مبتنی بر قواعد[1]

این روش ترکیبی است از کاربردهای تجزیه‌وتحلیل مطلق و تفاضلی .در تجزیه‌وتحلیل تفاضلی، یک سری معیارهای قابل‌انعطافی می‌توانند پیاده‌سازی شوند تا هرگونه تغییر در جزئیات سابقه‌ی رفتار یک کاربر را شناسایی نمایند. رویکردهای مبتنی بر قواعد عموماً با شناسه کاربرانی که شامل اطلاعات شفافی هستند و در آن‌ها معیارهای تقلب به قواعد اشاره می‌کنند، بهترین عملکرد رادارند .مدیریت کردن این روش کاری بسیار دشوار است و این مسئله به دلیل این است که پیکربندی مناسب قواعد، نیازمند برنامه‌نویسی زمان‌بر، دقیق و پرزحمتی برای هر امکان تقلب قابل‌تصور می‌باشد .یکی از ابزارهای تولیدشده با این رویکرد، PDATاست که توسط شرکت زیمنس ZFE تهیه‌شده و ابزاری کاملاً انعطاف‌پذیر با کاربردی وسیع، به‌منظور تشخیص تقلب در تلفن‌های همراه می‌باشد [۲۲].

۲-۱۲-۳-شبکه عصبی

شبکه عصبی تکنیک دیگری برای کشف تقلبات است که در آن سیستم به‌وسیله‌ی داده‌های گذشته مشتریان آموزش می‌بیند و می‌تواند برای تحلیل رفتار جاری مشتریان و کشف ناهنجاری‌ها بکار رود. در شبکه‌های عصبی، به‌صورت نرم‌افزاری، ساختار داده‌ای طراحی می‌شود که می‌تواند همانند نورون عمل نماید؛ به این ساختار داده‌ها، گره گفته می‌شود. سپس با ایجاد شبکه‌ای بین این گره‌ها و اعمال یک الگوریتم آموزشی به آن، شبکه را آموزش می‌دهند. در شبکه عصبی مجموعه‌ای از گره‌های به هم متصل از کارکرد مغز انسان تقلید می‌کنند و هر گره ارتباطات وزن‌داری با چندین گره دیگر در لایه‌های مجاور دارد.[۲۳]

1 . Rule-Based Approach

۲-۱۲-۴-الگوریتم ژنتیک[1]

الگوریتم ژنتیک جست‌وجوی اکتشافی است که فرآیند تکامل طبیعی را پیروی می‌کند، این اکتشاف برای ایجاد راه‌حل‌های مفید در بهینه‌سازی و جست‌وجوی مسائل به‌طور مداوم استفاده‌شده است. الگوریتم‌های ژنتیک متعلق به کلاس بزرگ‌تری از الگوریتم‌های تکاملی هستند، که برای ایجاد راه‌حل‌های بهینه‌سازی مسائل از روش‌های الهام گرفته از تکامل طبیعی، مانند وراثت، جهش، گزینش و تقاطع استفاده می‌کنند، و به این صورت است که طبیعت، افراد قوی‌تر را برای زندگی برمی‌گزیند[۲۴].

۲-۱۲-۵-تجزیه‌وتحلیل حالت گذار[2]

این روش نیز یک تکنیک شناسایی تقلب است که در آن، حملات به‌عنوان دنباله‌ای از حالت گذار سیستم مونیتور شده، نمایش داده می‌شود. فعالیت‌هایی که در یک حمله اتفاق می‌افتند، به‌عنوان یک گذار بین حالت‌ها تعریف می‌شوند.

سناریوهای حمله نیز در قالب نمودارهای گذار حالت تعریف می‌شوند .در این نمودارها، گره‌ها به‌منزله حالت‌های سیستم و کمان‌ها نیز به‌منزله اقدامات مرتبط می‌باشند .درهرصورت اگر به یک حالت نهایی برسیم، بدین معنی خواهد بود که یک حمله خواهیم داشت. سیستم STAT یک سیستم خبره قاعده مدار بسیار معروف

است که به‌منظور جستجوی نفوذهای شناخته‌شده در یک دنباله ممیزی از سیستم‌های رایانه‌ای چندکاربره طراحی‌شده است.[۲۵]

1 . Genetic Algorithm
2 . State Transition Analysis

همچنین USTAT نیز یک نمونه اولیه از STAT است که تحت سیستم‌عامل یونیکس طراحی گردیده است.[۲۶]

۲-۱۲-۶-داده‌کاوی

تکنیک دیگری که می‌توان از آن برای کشف تقلبات بهره برد داده‌کاوی است . یکی از قابلیت‌های فوق‌العاده داده‌کاوی در تشخیص تقلبات، امکان پیاده‌سازی مدل‌هایی است که تقلبات و کلاه‌برداری‌ها را قبل از تشخیص متخصصان، شناسایی کرده و ارائه می‌دهد.

داده‌کاوی بر تحلیل آماری و کشف رفتار مشتریان و استفاده از الگو برای شناسایی جرم تمرکز می‌کند و از بخشی از علم آمار به نام تحلیل اکتشافی داده‌ها[1] استفاده می‌کند که در آن بر کشف اطلاعات نهفته و ناشناخته از میان حجم انبوهی از داده‌ها تأکید می‌شود. علاوه بر این داده‌کاوی باهوش مصنوعی و یادگیری ماشین ارتباط تنگاتنگی دارد؛ بنابراین می‌توان گفت که درداده کاوی تئوری‌های پایگاه داده‌ها، هوش مصنوعی، یادگیری ماشین، علم آمار را در هم می آمیزندتا زمینه‌ی کاربردی فراهم شود.[۲۷]

داده‌کاوی یک تکنولوژی جدید و قدرتمند است که با پتانسیل بسیار بالای خود به یاری شرکت‌های بیمه آمده است. ازآنجاکه در صنعت بیمه با حجم زیادی از اطلاعات (داده‌های جمع‌آوری‌شده در رابطه با رفتار مشتریان و مشتریان بالقوه) سروکار داریم؛ داده‌کاوی با تمرکز بر روی مهم‌ترین اطلاعات از میان انبوه اطلاعات کمک شایانی به تحلیل‌گران شرکت‌های بیمه می‌کند. در حقیقت، داده‌کاوی با پیش‌بینی خسارت‌های تقلبی و پوشش‌های درمانی واهی و همچنین با پیش‌بینی نیازهای مشتریان کمک فراوانی به

1 . Exploratory Data Analysis

صنعت بیمه می‌کند. اگرچه داده‌کاوی به‌صورت گسترده در صنعت بیمه به‌کاربرده می‌شود؛ اما تنها شرکت‌هایی از مزایای رقابتی آن بهره‌مند می‌شوند که داده‌کاوی را به‌درستی اجرا کنند.

در متون آکادمیک تعاریف گوناگونی برای داده‌کاوی ارائه‌شده‌اند. در برخی از این تعاریف داده‌کاوی در حد ابزاری که کاربران را قادر به ارتباط مستقیم با حجم عظیم داده‌ها می‌سازد معرفی گردیده است و در برخی دیگر، تعاریف دقیق‌تر که در آن‌ها به کاوش درداده ها توجه می‌شود موجود است. برخی از این تعاریف عبارت‌اند از:

- اصطلاح داده‌کاوی به فرایند نیم خودکار تجزیه‌وتحلیل پایگاه داده‌های بزرگ به‌منظور یافتن الگوهای مفید اطلاق می‌شود.[۲۸]
- داده‌کاوی استخراج اطلاعات مفهومی، ناشناخته و به‌صورت بالقوه مفید از پایگاه داده می‌باشد . [۲۹]
- داده‌کاوی علم استخراج اطلاعات مفید از پایگاه‌های داده یا مجموعه داده‌ای می باشد..[۳۰]

همان‌گونه که در تعاریف گوناگون داده‌کاوی مشاهده می‌شود، تقریباً در تمامی تعاریف به مفاهیمی چون استخراج دانش، تحلیل و یافتن الگوی بین داده‌ها اشاره‌شده است.

۲-۱۲-۶-۱- انواع داده‌کاوی

دو هدف مهم داده‌کاوی پیش‌بینی[1] و تشریح[2] است. در پیش‌بینی، بعضی از متغیرها یا حوزه‌هایی از مجموعه‌های داده‌ای به‌منظور پیش‌بینی ارزش ناشناخته یا آینده‌ی داده‌های دیگر مورداستفاده قرار

1 .Predictive
2 .Descriptive

می‌گیرند. از سوی دیگر تشریح، بر یافتن الگوهای تشریحی داده‌ها که می‌توانند به‌وسیله انسان تعبیر شوند تمرکز می‌نمایند. درنتیجه داده‌کاوی را می‌توان دریکی از گروه های زیر جای داد.

- در داده‌کاوی پیش‌بینی کننده با استفاده از داده‌ها، مدل‌هایی برای پیش‌بینی مقادیر متغیرهای موردنظر تولید می‌گردد.
- داده‌کاوی تشریحی با استفاده از الگوهایی که در اعداد می‌یابد به تجزیه‌وتحلیل و علت‌یابی یک یا چند پدیده می‌پردازد.

ازنظر پیش‌بینی کننده، هدف از داده‌کاوی تولید مدلی است که با استفاده از یک کد اجرایی، وظایفی چون پیش‌بینی، دسته‌بندی، تخمین مقدار، تخمین عملکرد و غیره را انجام دهد.

ازنظر تشریح کننده، هدف حصول درکی کامل از سیستم تحلیل‌شده به‌وسیله‌ی الگوهای پنهان در آن و روابط درون مجموعه‌های داده‌ای است...

۲-۱۲-۶-۲- مراحل کشف دانش

در سال ۱۹۵۹ اصطلاح یادگیری ماشین برای اولین بار توسط ساموئل[1] مطرح شد که به معنی راه‌هایی که رایانه می‌تواند از داده‌ها به‌طور مستقیم دانش به دست آورد و توسط یادگیری ماشین بررسی می‌شود. برخی از مؤلفین داده‌کاوی را کشف دانش و معرفت از پایگاه داده‌ها[2] می‌دانند که طبق شکل (۲-۴) شامل مراحل زیر است:[۳۱]

۱- پاک‌سازی داده‌ها[3]: حذف داده‌های کاملاً متفاوت

۲- یکپارچه‌سازی داده‌ها[4]: ترکیب منابع متعدد، پراکنده و ناهمگن داده‌ها

1 .Samuel
2 .Kdd: Knowledge Discovery In Databases
3 .Data Cleaning
4 .Data Integration

۳-انتخاب داده‌ها[1]: بازیابی داده‌های مربوط به کشف دانش

۴- تبدیل داده‌ها[2]: برای به کار بردن در روش‌های مختلف

۵- داده‌کاوی[3]: ضروری‌ترین مرحله در کشف دانش و معرفت از پایگاه داده‌ها که در آنالیز روش‌های آماری مخصوصاً یادگیری ماشین برای استخراج الگوی مناسب استفاده می‌شود و شامل مراحل زیرمی‌باشد:

۵-۱ انتخاب عملیات داده‌کاوی (دسته‌بندی، خوشه‌بندی، پیش‌بینی، تعیین وابستگی و...)

۵-۲ انتخاب روش داده‌کاوی (شبکه‌های عصبی، درخت تصمیم‌گیری، الگوریتم ژنتیک و...)

۵-۳ داده‌کاوی برای یافتن الگوی مناسب است.

۶- ارزیابی الگوها: انتخاب بهترین الگوی مناسب

۷- ارائه دانش: ارائه دانش استخراج‌شده با استفاده از تکنیک‌های نمایش اطلاعات.

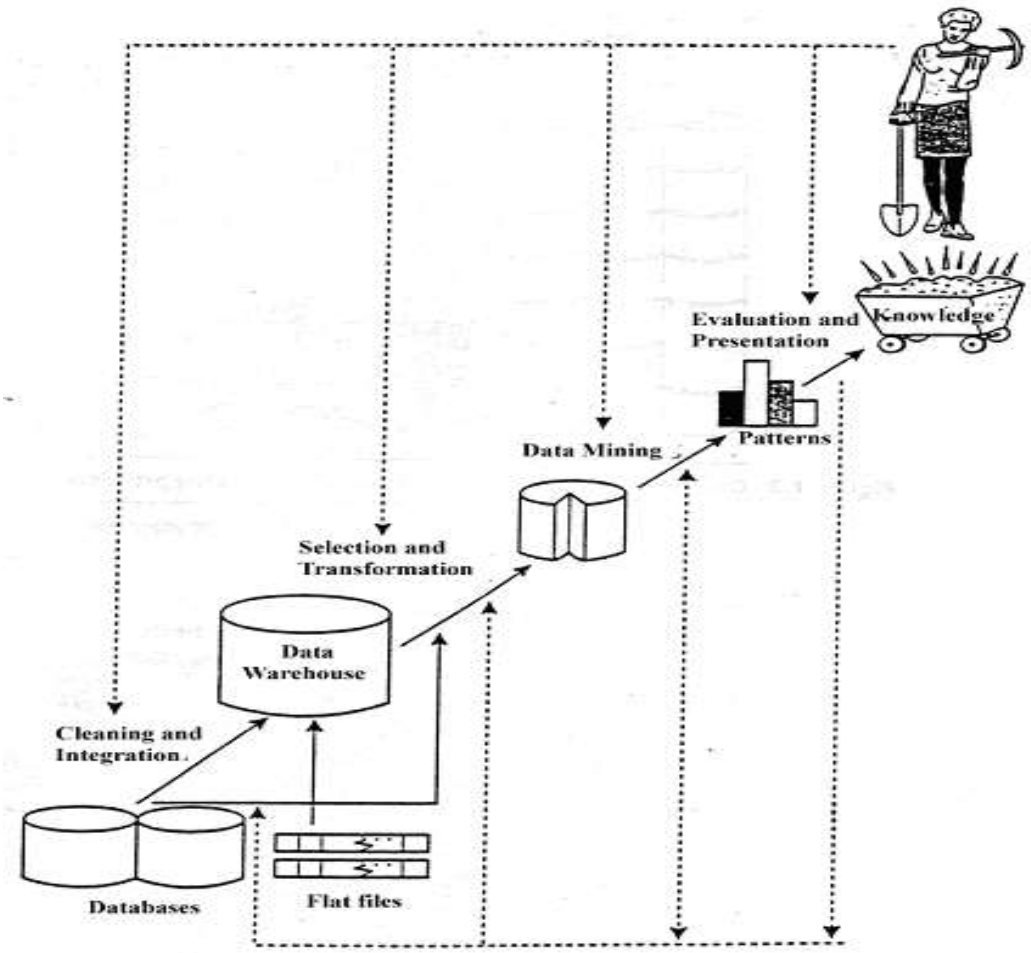
۸- کاوش در داده‌ها

در این مرحله است که داده‌کاوی انجام می‌شود. در این مرحله با استفاده از تکنیک‌های داده‌کاوی داده‌ها مورد کاوش قرارگرفته، دانش نهفته در آن‌ها استخراج‌شده و الگوسازی صورت می‌گیرد.

۹- تفسیر نتیجه

در این مرحله نتایج و الگوها توسط ابزار داده‌کاوی موردبررسی قرارگرفته و نتایج مفید معین می‌شود.

1 .Choose Data
2 .Data Transformation
3 .Data Mining



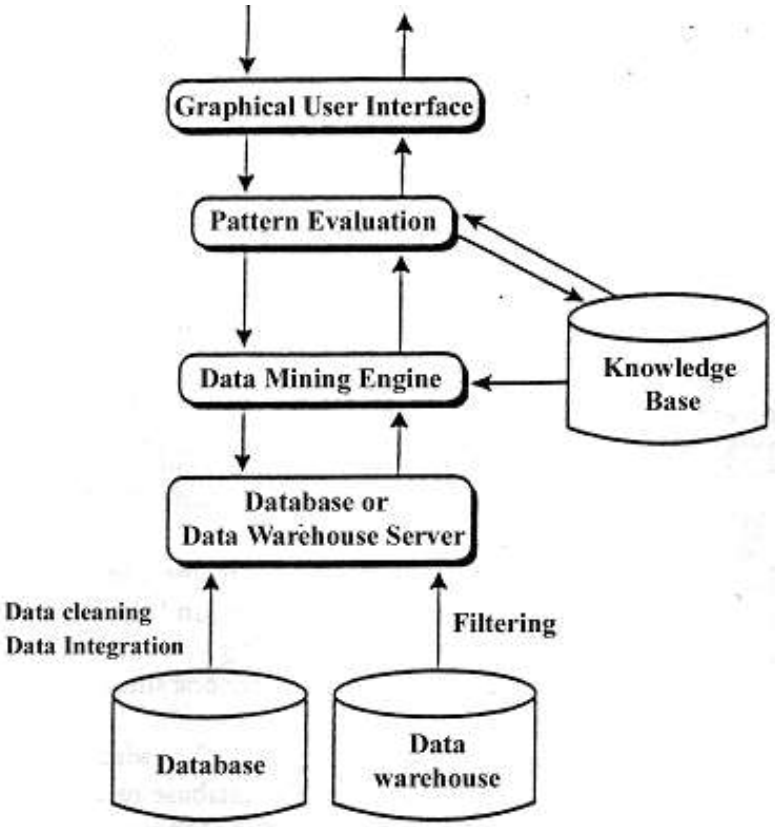
شکل (۲-۴): شماتیك مراحل کشف دانش[۳۲]

۲-۱۲-۶-۳- مراحل فرآیند داده‌کاوی

اغلب شرکت‌ها هنگامی‌که می‌خواهند از داده‌کاوی بهره بگیرند ترجیح می‌دهند از فرآیند استانداردی استفاده کنند. کریسپ[1] فرآیند استاندارد و مستقل از صنعت، ابزار و کاربرد است که در سال ۱۹۹۶ توسط اس پی اس اس[2]، ان سی آر[3] و ... توسعه داده شد. براساس این استاندارد یک پروژه داده‌کاوی چرخه

1 .Crisp Dm: Cross-Industry Standard Process For Data Mining
2 .Spss: Statistical Package For Social Science
3 .Ncr: National Cash Register

حیاتی شامل ۶ فاز است. رعایت توالی این فازها چندان اهمیت ندارد و معمولاً نیاز می‌شود که بین فازهای مختلف عقب و جلو شود. خروجی هر فاز تعیین می‌کند که آیا نیاز است به فاز قبلی برگردیم و یا می‌توان به فاز بعدی رفت. فازهای مدل کریسپ در شکل (۲-۵) آمده است [۳۳].



شکل (۲-۵): مراحل انجام یک فرآیند داده‌کاوی[۳۳].

✓ درک خواسته‌های تجاری[1]

در این مرحله اولیه بر روی درک اهداف و نیازمندی های پروژه از دیدگاه تجاری و تبدیل آن‌ها به تعاریف حوزه داده‌کاوی تمرکز می‌شود و سپس یک طرح اولیه برای به رسیدن به اهداف در نظر گرفته می‌شود.

1 .Business Understanding

✓ درک داده[1]

این فاز با گردآوری داده‌های اولیه آغاز می‌شود و با فعالیت‌هایی جهت آشنایی با داده، شناسایی مشکلات کیفیتی داده، کشف ماهیت داده‌ها و یافتن زیرمجموعه‌های جالب‌توجهی که فرضیات اطلاعات پنهان را شکل می‌دهند، ادامه می‌یابد.

✓ آماده‌سازی داده[2]

فاز آماده‌سازی داده شامل همه‌ی فعالیت‌هایی است که منجر به ساخت مجموعه داده نهایی از داده‌های خام اولیه می‌شود. وظایف آماده‌سازی داده معمولاً چندین بار انجام می‌گیرند. این وظایف شامل انتخاب جدول، رکورد و ویژگی‌هاست، همچنین تغییر شکل و خالص‌سازی داده با ابزارهای مدل‌سازی.

✓ مدل‌سازی[3]

در این فاز، تکنیک‌های مدل‌سازی مختلفی انتخاب و به‌کاربرده می‌شوند و تنظیمات مدل برای بهینه‌سازی نتایج به شکل اصول معینی درآورده می‌شود. معمولاً از تکنیک‌های مختلفی برای حل مسئله داده‌کاوی استفاده می‌شود. در برخی از این تکنیک‌ها، لازم است داده به‌صورت خاصی باشد؛ بنابراین در صورت نیاز به فاز قبلی یعنی فاز آماده‌سازی داده بازمی‌گردیم.

[1] .Data Understanding
[2] .Data Presentation
[3] .Modeling

- ✓ **ارزیابی**[1]

در این مرحله از پروژه، مدل یا مدل‌هایی ساخته‌شده و قبل از رفتن به فاز استقرار، باید این مدل‌ها و گام‌هایی که در ساخت آن‌ها به‌کاررفته، ارزیابی شود. هدف کلیدی در این فاز، بررسی وجود مسائلی است که به‌قدر کافی در نظر گرفته نشده‌اند. در انتهای این فاز، تصمیمی با توجه به کاربرد نتایج داده‌کاوی به‌دست‌آمده، گرفته می‌شود .[۳۳]

- ✓ **استقرار**[2]

ساخت مدل، انتهای پروژه نیست. حتی اگر هدف مدل افزایش دانش داده‌ها باشد، دانش به‌دست‌آمده باید به صورتی که برای مشتری قابل‌استفاده باشد، سازمان‌دهی و ارائه شود. با توجه به نیازمندی‌ها، فاز استقرار می‌تواند به‌سادگی تولید یک گزارش ساده و یا به پیچیدگی پیاده‌سازی فرآیند داده‌کاوی تکرارپذیر باشد. خیلی اوقات، این مشتری است که به‌جای تحلیل‌گر فاز استقرار را انجام می‌دهد.

۲-۱۲-۶-۴- تکنیک‌های داده‌کاوی

تکنیک‌های مختلف داده‌کاوی را می‌توان بر اساس نوع عملیاتی که انجام می‌دهند به دودسته پیش‌بینی کننده[3] و توصیفی[4] تقسیم کرد. تکنیک‌های پیش‌بینی کننده با ساخت مدلی برای پایگاه داده، وظیفه پیش‌بینی موارد ناشناخته را بر عهده‌دارند. درحالی‌که تکنیک‌های تشریح کننده، الگوهایی قابل‌فهم از داده‌ها را برای انسان کشف می‌کنند.

1. Evaluation
2. Deployment
3. Predictive
4. Descriptive

تکنیک‌های توصیفی عبارتنداز:

- خلاصه‌سازی[1]
- خوشه‌بندی[2]
- قوانین انجمنی[3]
- الگوهای متوالی[4]

تکنیک‌های پیشگویی عبارتنداز:

- طبقه‌بندی[5]
- پس‌گرایی[6]
- تحلیل سری زمانی[7]

داده‌کاوی با همه عظمت و بزرگی خود که امروزه در تمامی موضوعات جهان ورود پیداکرده است شامل شش عمل و وظیفه مهم است که می‌توان بسیاری از مسائل محیط اطراف خود را در قالب یکی از این شش عمل و وظیفه زیر گنجاند:

- ✓ دسته‌بندی[8]
- ✓ خوشه‌بندی[9]
- ✓ تخمین[10]

1. Summarization
2. Clustering
3. Association Rule
4. Sequential Pattern
5. Classification
6. Regression
7. Time Series Analysis
8. Classification
9. Clustering
10. Estimation

- ✓ پیش‌بینی[1]
- ✓ گروه‌بندی شباهت[2]
- ✓ توصیف و نمایه‌سازی[3]

سه مورد اول همگی داده‌کاوی هدایت‌شده هستند که هدف آن‌ها یافتن ارزش یک متغیر هدف خاص است. گروه‌بندی شباهت و خوشه‌بندی جزو داده‌کاوی غیر هدایت‌شده هستند که در آن هدف، یافتن ساختار پنهان درون داده‌ها بدون توجه به یک متغیر هدف خاص است. نمایه‌سازی عملی توصیفی است که می‌تواند هم هدایت‌شده و هم غیر هدایت‌شده باشد.[۳۴]

- ✓ **دسته‌بندی:**

به نظر می‌رسد دسته‌بندی که یکی از معمول‌ترین کارکردهای داده‌کاوی است، یکی از واجبات بشر باشد. تمامی خلقت خداوند بر پایه دسته‌بندی ایجاد گردیده است. ما برای شناخت و برقراری رابطه دربارهٔ دنیا، به‌طور مداوم دسته‌بندی، طبقه‌بندی[4] و درجه‌بندی[5] می‌کنیم. ما موجودات زنده را به شاخه‌ها وگونه ها، مواد را به عناصر و حیوانات و انسان را به نژادها تقسیم می‌کنیم.

دسته‌بندی شامل بررسی ویژگی‌های یک شی جدید و تخصیص آن به یکی از مجموعه‌های از قبل تعیین‌شده می‌باشد. عمل دسته‌بندی با تعریف درستی از دسته‌ها و مجموعه‌ای از ویژگی‌ها که حاوی موارد از پیش دسته‌بندی‌شده هستند مشخص می‌گردد؛ این عمل شامل ساختن مدلی است که بتوان از آن برای دسته‌بندی کردن داده‌های دسته‌بندی نشده، استفاده نمود. اشیایی که باید دسته‌بندی شوند،

1. Prediction
2. Affinity Grouping
3. Profiling
4. Categorization
5. Ranking

معمولاً به‌وسیلهٔ اطلاعاتی در جدول پایگاه داده‌ها یا یک فایل ارائه می‌شوند و عمل دسته‌بندی شامل افزودن ستون جدیدی با کد دسته‌بندی خاصی است. مثال‌هایی از دسته‌بندی در زیر ارائه‌شده است:

- دسته‌بندی متقاضیان وام و اعتبار به‌عنوان کم‌خطر، متوسط و پرخطر
- انتخاب محتویات یک صفحهٔ وب برای قرار دادن در شبکهٔ اینترنت
- تعیین شماره تلفن‌های متصل به دستگاه‌های فکس
- تشخیص مدعیان غیرواقعی دریافت خسارت از بیمه

در همهٔ این مثال‌ها تعداد محدود و از پیش تعیین‌شده‌ای از دسته‌ها وجود دارد و انتظار داریم بتوانیم هر اطلاعاتی را به یک یا دو مورد از آن‌ها تخصیص دهیم. تکنیک‌های درخت تصمیم[1] و نزدیک‌ترین همسایه[2] ازجمله تکنیک‌های دسته‌بندی می‌باشند؛ شبکه‌های عصبی[3] و تحلیل پیوند[4] نیز در شرایط خاصی عمل دسته‌بندی را انجام می‌دهند [۳۴].

۲-۳- پیشینه تحقیق

داده‌کاوی و کشف دانش در پایگاه داده‌ها ازجمله موضوع‌هایی هستند که هم‌زمان با ایجاد و استفاده از پایگاه داده‌ها در اوایل دهه ۸۰ برای جستجوی دانش در داده‌ها شکل گرفت. اخیر داده‌کاوی موضوع بسیاری از مقالات، کنفرانس‌های علمی و رساله‌ها شده است. ایجاد و توسعه مدل‌های داده‌ای برای پایگاه سلسله مراتبی، شبکه‌ای و بخصوص رابطه‌ای در دهه‌ی ۷۰، منجر به معرفی مفاهیمی همچون شاخص گذاری و سازمان‌دهی داده‌ها و درنهایت ایجاد زبان پرسش ساختاربندی۵ شده در اوایل دهه‌ی ۸۰ گردید

1 .Decision Tree
2 .Nearest Neighbor
3. Neural Network
4. Link Analysis
5 . SQL(Structured Query Language)

تا کاربران بتوانند گزارش‌ها و برگه‌های اطلاعاتی موردنظر خود را از این طریق ایجاد کنند و دستگاه‌های پایگاهی پیشرفته در دهه ۸۰ و ایجاد پایگاه‌های شی گرا، کاربرد گرا و فعال باعث توسعه همه‌جانبه و کاربردی شدن این سیستم‌ها در سراسر جهان گردید. بدین ترتیب سیستم‌های مدیریت بانک‌های اطلاعاتی همچون دی بی، اوراکل و ... ایجاد شدند و حجم زیادی پژوهش جدی روی موضوع داده‌کاوی از اوایل دهه نود شروع شد. شاید بتوان لوول را اولین شخصی دانست که گزارشی در مورد داده‌کاوی تحت عنوان " شبیه‌سازی فعالیت داده‌کاوی" ارائه نمود [[۳۵، اما پژوهش جدی روی موضوع داده‌کاوی از اوایل دهه ۹۰ شروع شد [۳۶]. هم‌زمان با او پژوهشگران و متخصصان علوم رایانه، آمار، هوش مصنوعی، یادگیری ماشین و ... نیز به پژوهش در این زمینه و زمینه‌های مرتبط با آن پرداختند. همچنین سمینارها، دوره‌های آموزشی و کنفرانس‌هایی نیز برگزار شد. برای مثال برای اولین بار مفهوم داده‌کاوی در کارگاه آموزشی ljcai درزمینهٔ کشف دانش توسط شاپر مطرح گردید. به دنبال آن در سال‌های ۱۹۹۱ تا ۱۹۹۴، کارگاه‌های کشف دانش مفاهیم جدیدی در این شاخه از علم ارائه کردند، به‌طوری‌که بسیاری از علوم و مفاهیم با آن مرتبط گریدند. در طی این سال‌ها هافمن و نش استفاده از داده‌کاوی و انبار داده توسط بانک‌های آمریکا را بررسی نموده و بیان کردند که چگونه این سیستم ها برای بانک‌های آمریکا قدرت بیشتری ایجاد می‌کنند. چت فیلد مشکلات ایجادشده توسط داده‌کاوی را بررسی نمود و همچنین مقاله‌ای تحت عنوان " مدل‌های خطی غیردقیق داده‌کاوی و استنباط آماری " ارائه نمود. هندری نیز دیدگاه اقتصادسنجی روی داده‌کاوی را ارائه داد. در سال ۱۹۹۵ انجمن داده‌کاوی هم‌زمان با اولین کنفرانس بین‌المللی " کشف دانش و داده‌کاوی " شروع به کارکرد. این کنفرانس توسعه‌یافته شش دوره آموزشی بین‌المللی در پایگاه‌های داده سال ۱۹۸۹ تا ۱۹۹۴ بود. انجمن مذکور ، یک سازمان علمی به نام ACM-SIGKDD را ایجاد نمود. امروزه داده‌کاوی به‌سرعت در حال رشد بوده و سازمان‌های زیادی در حال استفاده از داده‌کاوی برای کمک به مدیریت سازمان هستند [۳۷].

از دهه ۹۰ تاکنون تحقیقات بسیاری درزمینه ی کشف خسارت‌های تقلبی در رشته‌ی بیمه اتومبیل انجام‌شده است. برای مثال می‌توان به تحقیق ویسبرگ و دریگ اشاره کرد. بلهادجی و دیون تحقیقاتی را در این زمینه با استفاده از داده‌های خسارت بیمه‌ی اتومبیل کشور کانادا انجام داده‌اند. همچنین کیومینزو تنیسن تحقیقات مشابهی را در ایالات‌متحده انجام داده‌اند. محققان دیگری مانند دریگ و استاسزیوسکی، ویزبرگ و دریگ و براکت و همکارانش در مقالات خود تکنیک هایی برای شناسایی خسارت‌های تقلبی و دسته‌بندی کلاه‌برداری‌ها ارائه داده‌اند. آرتریس و همکارانش عملکرد انتخاب مدل های باینری را برای کشف تقلب در بازار بیمه‌ی اتومبیل اسپانیا برای سال‌های ۱۹۹۳ تا ۱۹۹۶ تجزیه‌وتحلیل کردند. آن‌ها روشی برای اصلاح طبقه‌بندی نوع خسارت معرفی کردند. پس از آم ها بلهادجی و دیون در مقاله خود در سال ۱۹۹۷، ابتدا با تحقیق و تفحص از خبرگان صنعت بیمه اتومبیل، به شناسایی عوامل کلیدی در تقلبات بیمه‌ای پرداختند سپس با محاسبه‌ی احتمال شرطی تقلب برای هر شاخص و به‌کارگیری الگویتم رگرسیون، مهم‌ترین شاخص‌ها را تعیین کردند. همچنین به کمک الگوریتم رگرسیون، به پیش‌بینی خسارت‌های تقلبی پرداختند. بروت و همکارانش در سال ۱۹۹۸، نیز در مقاله خود ابتدا، به کمک الگوریتم تحلیل مؤلفه‌های اصلی[1] به انتخاب ویژگی‌ها پرداختند سپس با ترکیب الگوریتم‌های خوشه‌بندی و شبکه‌های عصبی BP به کشف تقلبات بیمه اتومبیل پرداختند. پس‌ازآن آرتیس و همکارانش نیز در سال ۲۰۰۲، به مقایسه‌ی مدل‌های لاجیت چندجمله ای و مدل لاجیت چندجمله‌ای تودرتو در شناسایی تقلبات بیمه ی اتومبیل پرداختند. فوآ و همکارانش با ترکیب الگوریتم‌های شبکه‌های عصبی پس انتشاری[2]، بیز ساده و درخت تصمیم C4.۵ به کشف تقلب در بیمه‌های اتومبیل پرداختند. واین و ددن در مقاله ای در سال ۲۰۰۴، با به‌کارگیری روش بیز ساده۳، اقدام به کشف تقلب درداده های بیمه اتومبیل

1 .Backpropagation Neural Network
2 .Nested MultinominalLogit Model(NMLM)
3 .Naive Bayes

کردندودر این مقاله از الگوریتم‌های تقویت‌کننده[1] استفاده شد و نتایج آن با نتایج حاصل از اعمال الگوریتم بیز ساده بدون به‌کارگیری الگوریتم‌های تقویت‌کننده مقایسه شد و ثابت‌شده است که استفاده از الگوریتم‌های تقویت‌کننده، نتایج دقیق‌تری میدهد. [۳۸]

استنبرگ و رینالددر سال ۱۹۹۷ با استفاده از الگوریتم های فراابتکاری به کشف تقلب های بیمه اتومبیل تنسون و همکاران با استفاده از یک مدل رگرسیون لجستیک اقتصاد سنجی، نقش ممیزی و بازرسی ادعاهای خسارت را در صنعت بیمه اتومبیل بررسی کردند [۳۹] وسبرگ و همکاران در سال ۱۹۹۸ برای کاهش پرداخت های غیرقابل توجیه، از مدل رگرسیون لجستیک استفاده نموده است و مقدارمعمول پرداخت برای بیمه گر درمورد بیمه بدنه اتومبیل را پیش بینی کرده است. [۴۰]بلهاج و همکاران در سال ۲۰۰۰ با استفاده از مدل رگرسیون، یک سیستم تصمیم گیری برای مبارزه با تقلب بیمه اتومبیل ارائه نمودند.[۴۱]

ویلن و همکاران در سال ۲۰۰۲ برای کشف تقلب های مربوط به بیمه اتومبیل، از یک مدل رگرسیون لجستیک برای امتیازدهی به ادعاهای خسارت استفاده نموده است[۴۲]

آتریس و آیسودر سال ۲۰۰۲ از مدل های تصمیم گیری گسسته برای درک رفتارمتقلبانه در بیمه اتومبیل استفاده نموده و تاثیر مشخصه های بیمه گزارو ادعای خسارت را براحتمال تقلبی بودن پرونده بررسی نموده است [۴۳] پاتک و ویدیارتدر سال ۲۰۰۵ از یک سیستم تشخیص فازی برای ارزیابی مولفه های تقلب در پرونده های تقلبی بیمه اتومبیل استفاده کردند.[۴۴] ویلن و دریگ در سال ۲۰۰۵ در تحقیقی دیگر با استفاده از شبکه های عصبی و شبکه های بیز به بررسی تقلب های تصادفات جرحی پرداخت.[۴۵] آیسو و پنگوت در سال ۲۰۰۷ از روش رگرسیون برای بازرسی و کشف تقلب در بیمه اتومبیل استفاده کرده است [۴۶]برموز در سال ۲۰۰۸ ازمدل رگرسیون نامتقارن وشبکه بیز برای کشف

1 .Boosting Algorithms

تقلب بیمه اتومبیل استفاده نمودند.[۴۷] رخا در سال ۲۰۱۱ با استفاده از داده کاوی و استفاده از تکنیک های بیز ساده و درخت تصمیم و مصورسازی داده ها در جهت مقابله با تقلب های بیمه اتومبیل برآمد. وی با استفاده از یک بانک اطلاعاتی از تصادفات، به بررسی نوع تصادفات (تقلبی یا واقعی) پرداخته است. ویژگی های موجود در این بانک اطلاعاتی شامل داشتن بیمه نامه، نرخ رانندگی[1]، داشتن کروکی، قیمت وسیله و سن بوده است. سپس به بررسی تصادفات تقلبی با استفاده از تکنیک های درخت تصمیم C 4.5 وشبکه بیز پرداخته است.[۴۸]

علاوه بر تحقیقات بین المللی که در زمینه داده کاوی و کشف تقلب طی سل های مختلف به انجام رسیده، در ایران نیز تحقیقات داخلی ای در این راستا شکل گرفته است از جمله: ایزدپرست و همکاران که با استفاده از روش‌های داده‌کاوی، چهارچوبی را برای شناسایی مشتریان بیمه (با تمرکز به مشتریان بیمه بدنه اتومبیل) ارائه کرده‌اندبدین گونه که بوسیله روش k-means به خوشه بندی مشتریان پرداخته اند تا مشتریانی را که بیشتر به یکدیگر شبیه هستند مشخص کنند، و سپس بوسیله درخت تصمیم و نظر کارشناسان خوشه ها را برچسب گذاری و دسته بندی نموده تا با استفاده از این دسته‌ها و ویژگی‌های آن میزان خطرپذیری هر دسته را مشخص کنند.[۴۹]

بعلاوه فیروزی و شکوری از سه روش داده کاوی رگرسیون لجستیک، بیز ساده و درخت تصمیم برای پیدا کردن الگوهایی استفاده کرده اند که به شرکت های بیمه برای شناسایی تقلب ها در بیمه اتومبیل کمک می کند.[۵۰]

۲-۴- جمع بندی

در این فصل مبانی نظری و مفاهیم رایج در تقلبات بیمه‌ای و همچنین فرآیند و الگوریتم‌های داده‌کاوی و کاربرد آن در کشف تقلب تشریح شد. همان‌طور که عنوان شد این الگوریتم‌ها به دودسته‌ی توصیفی و

1 .Driver rate

پیش بینانه می‌شوند که در این تحقیق به‌منظور استفاده ارزیابی و تشخیص صحت ادعاهای خسارت بیمه‌ای استفاده از الگوریتم‌های پیش بینانه مفید و مؤثر خواهد بود. سپس خلاصه‌ای از مهم‌ترین تحقیقات انجام‌شده در این حوزه اعم از تحقیقات بین‌المللی و داخلی ارائه گردید. ازآنجایی‌که تقلب‌های بیمه یک مشکل جدی در کل صنعت بیمه هستند و هزینه‌های زیادی را متحمل بیمه می‌کنند، نیاز به روشی دقیق برای جلوگیری از این تقلبات هرروز بیشتر احساس می‌شود؛ و ازآنجایی‌که در هیچ‌کدام از تحقیقات داخلی و خارجی از الگوریتم‌های ماشین بردار پشتیبان و جنگل تصادفی که الگوریتم‌های قدرتمند و دقیقی هستند درکنار هم و به‌صورت ترکیبی استفاده‌نشده است، در این پژوهش از این الگوریتم‌ها استفاده خواهد شد تا روش نوینی در عرصه کشف تقلب معرفی شود

فصل سوم

روش تحقیق

فصل ۳: روش تحقیق

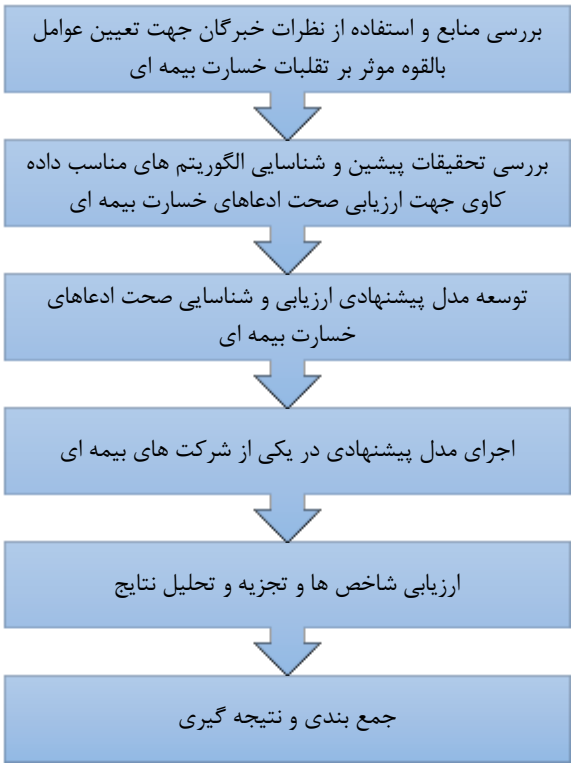
۳-۱- مقدمه

تحقیق را می‌توان ثبت عینی و نظامند مشاهدات کنترل‌شده و تجزیه‌وتحلیل آن‌ها که به پروراندن قوانین کلی، اصول نظریه‌ها و همچنین به‌پیش بینی ویا احتمالاً به کنترل نهایی رویدادها منجر می‌شود تعریف کرد.

ازآنجایی‌که بررسی تحقیقات گذشته در هر موضوع تحقیقاتی می‌تواند نتایج و دستاوردهای مهم و مفیدی برای محققان بعدی داشته باشد و آن‌ها را دریافتن مسیرهای شناخته‌شده و اهداف یافته شده یاری نماید در فصل گذشته به بررسی تحقیقات مشابه انجام‌شده و دستاوردهای آن‌ها پرداختیم و در این فصل نیز به معرفی روش تحقیق موردنظر و معرفی تکنیک‌های مفیدی که در این راستا به کمک ما می‌آیند می‌پردازیم. در این تحقیق به بررسی و مقایسه بین تکنیک ها و الگوریتم‌های داده‌کاوی مانند بردار ماشین پشتیبان، جنگل تصادفی و ترکیب این دو الگوریتم مختلف خواهیم پرداخت تا بتوان به پیش‌بینی و تشخیص تقلبات بیمه‌ای پرداخت. روش پژوهش در این تحقیق از نوع توسعه‌ای تحقیقاتی می‌باشد .در این تحقیقات به دنبال توسعه یک مدل یا روش برای پیش‌بینی بهتر تقلبات از طریق استفاده و ترکیب الگوریتم‌های جدید دسته‌بندی می‌باشیم؛ بنابراین تعدادی از الگوریتم‌های داده‌کاوی مناسب موردبررسی قرارگرفته و بامطالعه و ارزیابی معیارهای مربوط به الگوریتم‌های داده‌کاوی در زمینه بیمه بهترین الگوریتم با ترکیب الگوریتم‌های موجود و با توجه به معیارهای مشخص‌شده تعیین و ارائه خواهد شد.

۳-۲- روش تحقیق

ابتدا با بررسی منابع موجود به شناسایی عوامل مؤثر در تقلبات بیمه‌ای پرداخته و همچنین بامطالعه تکنیک‌ها و الگوریتم‌های مختلف داد کاوی، تکنیک های مناسب و مؤثر جهت ارزیابی صحت ادعاهای بیمه‌ای مشخص می‌گردد. این پژوهش از نوع کمی و باهدف کاربردی می‌باشد و نحوه جمع‌آوری داده‌ها در مراحل اولیه از نوع کتابخانه‌ای و در مراحل بعدی از نوع میدانی بوده و با استفاده از داده‌های واقعی ۱۰۰۰ مشتری بیمه شخص ثالث در بیمه دی تهران انجام‌شده است. و مدل ساخته می‌شود و درنهایت مدل توسعه داده‌شده توسط داده‌های موجود مورد تست و ارزیابی قرارگرفته و شاخص‌های من مورد ارزیابی و تجزیه‌وتحلیل قرار می‌گیرد. با توجه به توضیحات فوق مراحل تحقیق حاضر مطابق شکل (۳-۱) است.



شکل (۳-۱) فرآیند تحقیق

۳-۳ -جامعه آماری

با توجه به گستردگی صنعت بیمه در کشور جامعه‌ی آماری که در این پژوهش بکار گرفتیم یکی از شرکت‌های بیمه دی در تهران است که ۱۰۰۰ پرونده بیمه در حوزه‌ی " بیمه شخص ثالث" که در سال ۱۳۹۴ ادعای خسارت داشته‌اند و صحت ادعای خسارت آن‌ها در طی این‌یک سال بررسی‌شده و موارد تقلبی مشخص‌شده است را به‌عنوان نمونه برگزیدیم و با استفاده ازنظر خبرگان و کارشناسان معیارهای مهم تشخیص تقلب مانند: مدل ماشین مقصر، علت حادثه، مدت اعتبار بیمه‌نامه و... را تعیین نمودیم و پس از غربالگری و آماده‌سازی داده‌ها مدل خود را بر روی آن‌ها پیاده‌سازی کردیم. .

۱-۳-۳ متغیرهای مورداستفاده

در این پژوهش هر نوع اطلاعاتی در مورد شخص بیمه‌شده و اطلاعاتی درباره جزییات حادثه می‌تواند بر روی جواب نهایی تأثیر گزار باشد اما ازآنجایی‌که جمع‌آوری برخی اطلاعات به‌سختی صورت میپزیرد و در موارد زیادی با داده های گم‌شده روبرو می‌شویم تصمیم برآن گرفتیم که از متغیرهای زیر به‌عنوان متغیرهایی که می‌توانند صحت یا عدم صحت ادعای خسارت بیمه‌شدگان را به کمک مدل تشخیص دهند استفاده نماییم.

❖ **متغیرهای ورودی**

- ساعت وقوع
- تعداد نفرات آسیب‌دیده
- مبلغ خسارت
- جنسیت مالک

۰ معرف زن

۱ معرف مرد

- علت حادثه:

۱. تغییر مسیر ناگهانی

۲. عدم توجه به جلو

۳. عدم رعایت نکات ایمنی

۴. عدم رعایت حق تقدم

۵. باز شدن درب به‌طور ناگهانی

۶. حرکت غیرضروری بادنده عقب

۷. عدم توانایی در کنترل سرعت مطمئنه

۸. عدم رعایت فاصله طولی

۹. تجاوز از سرعت مجاز

۱۰. ترکیب و تقارن دو خطا

۱۱. دور زدن در محل ممنوع

۱۲. گردش به سمت چپ یا راست در محل ممنوع

۱۳. واژگونی

۱۴. انحراف به چپ

۱۵. برخورد با عابر پیاده

- بومی:

معرف این است که مقصر اهل شهرمورد نظر (تهران) است یا خیر

- نوع صدمه بدنی:

۱- فوت

۲- صدمات حیاتی و خطرناک: صدمات کمر، لگن، قفسه سینه، نخاع، شکستگی سر، چشم، دنده و شکم، صدمات سنگین پا و شکستگی ران پا، جمجمه، فک، زانو، گردن

۳- صدمات غیر حیاتی: خراش صورت، لب، شکستگی دست و انگشت و ساعد، آرنج، بینی، دندان، آسیب‌های جزئی سر و آسیب‌های کوچک و متوسط پا.

- اعتبار بیمه‌نامه:

تعداد ماه‌های باقی‌مانده از اعتبار بیمه‌نامه مقصرتا پایان اعتبار بیمه‌نامه.

- سابقه بیمه:

تعداد سال‌های بدون خسارت بیمه‌نامه شخص مقصر

- وسیله صاحب بیمه‌نامه:

۱. سواری داخلی

۲. سواری خارجی

۳. پیکان

۴. اتوبوس

۵. ماشین سنگین

۶. آمبولانس

۷. موتور

- مدل:

سال تولید ماشین مقصر و زیان‌دیده

- استعلام:

استعلام بیمه مرکزی از وسیله زیان‌دیده در مورد اعلام خسارات آن

❖ متغیر خروجی

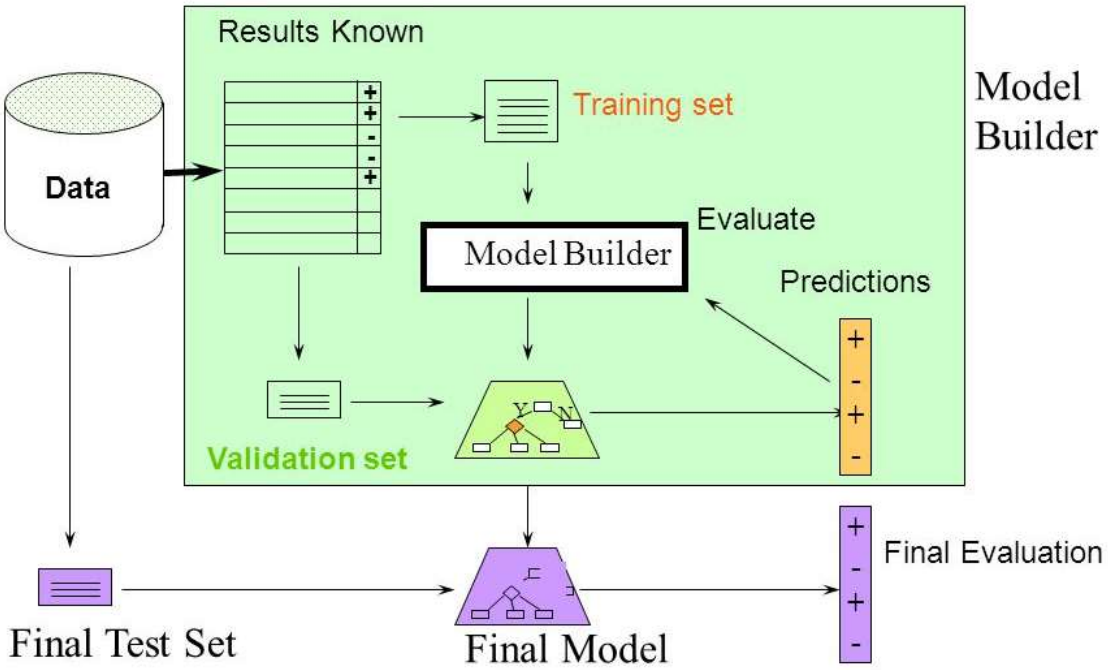
متغیر خروجی که در این پژوهش به‌عنوان متغیر هدف در نظر گرفته شد "تقلب " بود که به‌صورت باینری تعریف می‌شود و درصورتی‌که ادعای مشتری قبل از اعمال مدل و به‌وسیله نظر کارشناسان، تقلبی شناخته‌شده باشد آن را در دیتا بیس با عدد ۱ و در غیر این صورت با عدد ۰ نمایان کرده‌ایم.

۳-۴- مدل پیشنهادی

پس از شناخت داده‌ها و آماده‌سازی آن‌ها، حال می‌توان به مدل‌سازی پرداخت. یکی از مهم‌ترین تکنیک‌های پیش بینانه داده‌کاوی، دسته‌بندی است. دسته‌بندی شامل بررسی ویژگی‌های یک شی جدید و تخصیص آن به یکی از مجموعه‌های از قبل تعیین‌شده می‌باشد. عمل دسته‌بندی با تعریف درستی از دسته‌ها و مجموعه‌ای از ویژگی‌ها که حاوی موارد از پیش دسته‌بندی‌شده هستند مشخص می‌گردد؛ این عمل شامل ساختن مدلی است که بتوان از آن برای دسته‌بندی کردن داده‌های دسته‌بندی نشده، استفاده نمود. اشیایی که باید دسته‌بندی شوند، معمولاً به‌وسیلهٔ اطلاعاتی در جدول پایگاه داده‌ها یا یک فایل ارائه می شوند و عمل دسته بندی شامل افزودن ستون جدیدی با کد دسته‌بندی خاصی است.[۳۴].

فرآیند کلی دسته‌بندی در داده‌کاوی که در این تحقیق نیز جهت ساخت مدل موردنظر از آن استفاده‌شده مطابق شکل (۳-۲) می‌باشد. بر اساس آنچه در شکل دیده می‌شود ابتدا داده‌های آماده‌شده جهت پردازش به دودسته‌ی داده‌های آموزشی و داده‌های آزمایشی تقسیم می‌شوند. در هردو دسته داده‌ها متغیر خروجی که در این تحقیق تقلبی بودن یا نبودن ادعای خسارت است با استفاده از دو کد ۱ برای تقلب و ۰ برای عدم تقلب از قبل مشخص‌شده است. سپس با استفاده از داده‌های آموزشی مدل آموزش می‌بیند که چگونه پیش‌بینی کند و نمونه اولیه‌ای از مدل ساخته می‌شود سپس این مدل به‌وسیله‌ی داده‌های آزمایشی بررسی و تست می‌شود و این روند طی یک چرخه به‌وسیله‌ی نرم‌افزار به‌اندازه‌ای تکرار می‌شود تا بهترین مدلی که دقت بالایی در آموزش و تست داشته است تعیین گرد.

در این تحقیق الگوریتم‌های بردار ماشین پشتیبان و الگوریتم جنگل تصادفی استفاده می‌شود زیرا در تحقیقات دیگران درگذشته در حوزه‌های مختلف پزشکی و سایر حوزه‌های غیر مرتبط با داده‌کاوی بکار گرفته‌شده‌اند و ثابت کرده‌اند که دارای دقت زیادی در پیش‌بینی هستند و درصد خطای پایینی دارند و تاکنون از این دو الگوریتم قوی برکنار یکدیگر در حوزه‌ی داده‌کاوی استفاده‌نشده است، لذا از این دو الگوریتم قدرتمند با احتساب ۶۰ درصد داده‌ها به‌عنوان داده‌های آموزشی و ۴۰ درصد به‌عنوان داده‌های آزمایشی طبق پیش‌فرض و راهنمای نرم‌افزار کلمنتاین برای کسب نتایج بهتر استفاده شده است.



شکل (۳-۲): فرآیند ساخت مدل.[۳۴]

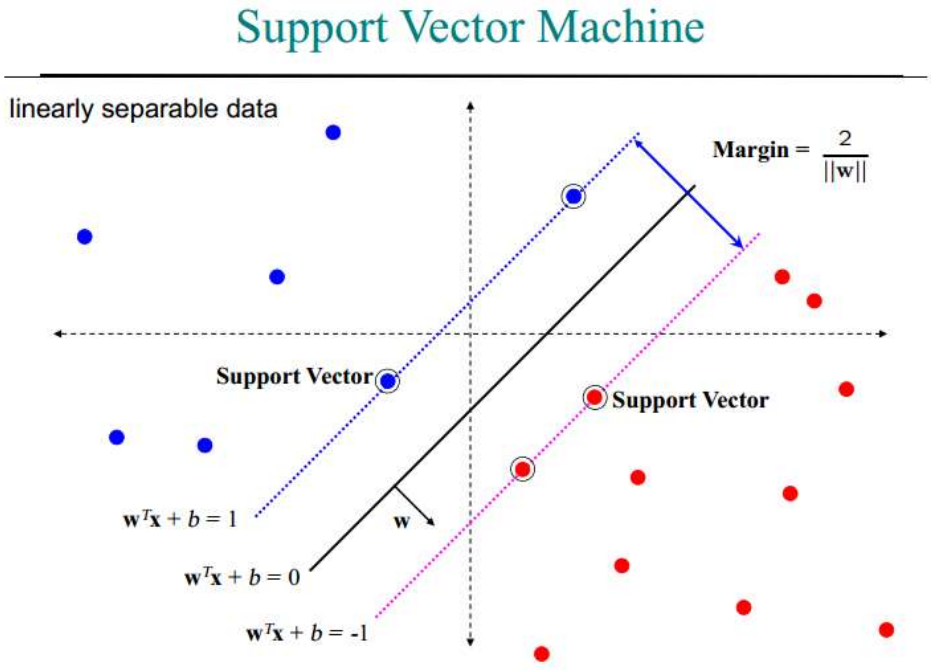
۳-۴-۱ الگوریتم ماشین بردار پشتیبان (SVM):

ماشین بردار پشتیبان (SVM)به‌عنوان یکی از قوی‌ترین و دقیق‌ترین متدها در میان الگوریتم‌های معروف شناخته می‌شود. ماشین بردار پشتیبان یکی از روش‌های یادگیری با ناظر است. از آن برای طبقه‌بندی و پیش‌بینی استفاده می‌شود. مبنای کار دسته‌بندی کننده SVM دسته‌بندی خطی داده‌هاست و در تقسیم خطی داده‌ها سعی بر آن است خطی انتخاب شود که حاشیه امنیت بیشتری داشته باشد. این کار کارایی ودقت طبقه‌بندی را افزایش می‌دهد و فضا را نیز برای طبقه‌بندی بهتر داده‌های آتی مهیا می‌کند.

این الگوریتم داده‌ها را با توجه به دسته‌های از پیش تعیین‌شده آن‌ها به یک فضای جدید می‌برد به‌گونه‌ای که داده‌ها به‌صورت خطی (ابر صفحه) قابل‌تفکیک و دسته‌بندی باشند و سپس با یافتن خطوط پشتیبان

(صفحات پشتیبان در فضای چندبعدی)، سعی در یافتن معادله خطی دارد که بیشترین فاصله رابین دودسته ایجاد می‌کند.

در شکل(۳-۳) که نمایی از عملکرد ماشین بردار پشتیبان است ،داده‌ها در دو دودسته آبی و قرمز نمایش داده‌شده‌اند و خطوط نقطه‌چین، بردارهای پشتیبان متناظر با هر دسته را نمایش می‌دهند که با دایره‌های دوخط مشخص شده‌اند و خط سیاه ممتد نیز همانSVM است. بردارهای پشتیبان هم هرکدام یک فرمول مشخصه دارند که خط مرزی هر دسته را توصیف می‌کند.



شکل (۳-۳): نمایی از عملکرد الگوریتم ماشین بردار پشتیبان

ماشین‌های بـردار پشتیبان برای حل مسائل غیرخطـی، ابعـاد مسئله را از طریـق توابـع کرنل تغییر می‌دهند. انتخاب کرنل برای SVM بـه حجـم داده‌های آموزشی و ابعاد بردار ویژگی بستگی دارد.

به‌عبارت‌دیگر، باید با توجه به این پارامترها تابع کرنلی را انتخاب نمود که توانایی آموزش برای ورودی‌های مسئله را داشته باشد. در عمل چهار نوع کرنل خطی[1] کرنل چندجمله‌ای[2]، کرنل سیگموییدی[3] کرنل گوسی[4] (RBF)بکار گرفته می‌شوند؛ که هرکدام دارای دقت متفاوتی هستند

۳-۴-۱-۱ مفهوم کرنل

ضرب داخلی ساده در یک فضای برداری می‌تواند نشان‌دهنده شباهت بین داده‌ها باشد. کرنل‌ها توسعه‌یافته ضرب داخلی داده‌ها می‌باشند که در یک فضای تبدیل یافته محاسبه می‌شود. معادل با هر هسته معتبر یک فضای هیلبرت فرآورده هسته (RKHS) موجود است:

$$K(x.y) = \langle \phi(x).\phi(y) \rangle \qquad (۱-۳)$$

$$\phi: x \rightarrow \phi(x)$$

بردارهای x و y تحت نگاشت $\varphi(x)$ قرارگرفته‌اند و در فضای جدید ضرب داخلی محاسبه‌شده است. کرنل‌ها خواص غیرخطی دارند، برای مثال فضای معادل با هسته گوسی دارای ویژگی‌های غیرخطی از فضای اولیه است. در مسئله‌ی با داده‌های چندگونه فرض می‌کنیم برای هرگونه یک هسته مجزا داریم، که نشان‌دهنده شباهت دو داده از منظر آن‌گونه است:

[1] .Linear kernel
[2] . Polynomial kernel
[3] . Sigmoid kernel
[4] . Radial Base Function kerne

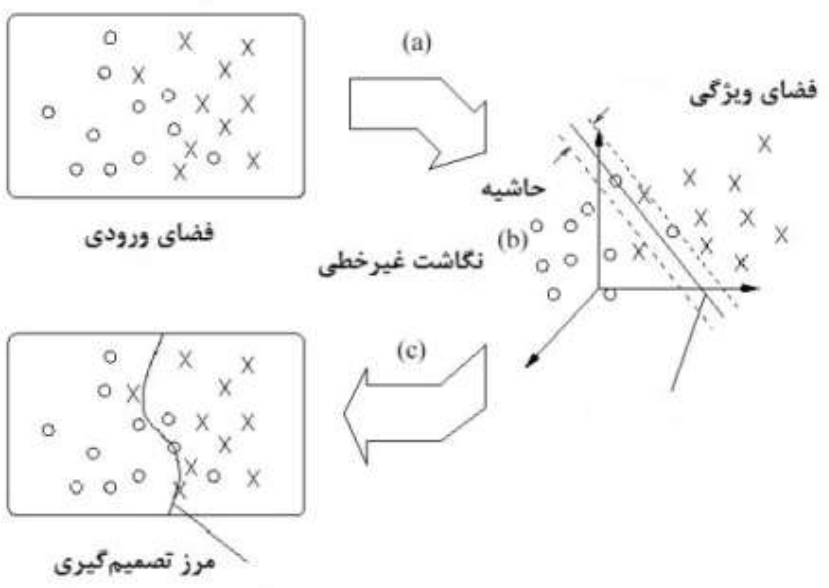
$$k_m\left(x_m^i . x_m^j\right) = \langle \phi_1\left(x_1^i\right) . \phi_1\left(x_1^j\right) \rangle . \quad m \in \{1. \ldots . \rho\} \qquad (۲-۳)$$

برای مسائلی که یک ابر صفحه تصمیم‌گیری غیرخطی دارند، یک تابع نگاشت $\emptyset(x)$ برای انتقال داده‌های اصلی به یک فضای ویژگی با ابعاد بالاتر استفاده می‌شود. دراین موارد می‌توان از یک تابع $k\left(x_i . x_j\right)$ که ضرب نقطه‌ای را در قضای ویژگی محاسبه می‌کند ،به‌عنوان یک عملیات مستقیم بر روی نمونه داده‌های اصلی استفاده کرد.[۵۱]

$$k(x_i . x_j) = \emptyset(x_i) \cdot \emptyset(x_j) \qquad (۳-۳)$$

تابع k یک کرنل نامیده می‌شود و SVMها یک عضو از کلاس گسترده از روش‌های کرنل هستند.[۵۲]

شکل (۳-۴) عملکرد یک تابع کرنل را در نگاشت داده‌های ورودی به یک فضای ویژگی و انجام عملیات کلاس‌بندی نشان می‌دهد [۵۱]



شکل (۳-۴): نمایش نقش کرنل دریک مسئله کلاس‌بندی[۵۱]

در یک تابع کرنل بجای ضرب نقطه‌ای، بردارهای تبدیل یافته جایگزین می‌شوند و شکل واضح و روشن تابع تبدیل $\emptyset(x)$ لزوماً شناخته‌شده نیست. بعلاوه استفاده از تابع کرنل به‌شدت به محاسبات کمتری نیاز دارد.

فرمول‌بندی تابع کرنل از ضرب نقطه‌ای یک مورد خاص از نظریه مرکز می‌باشد. مسئله بهینه‌سازی به‌صورت زیر درمی‌آید:

بیشینه:

$$\sum_{i=1}^{K} \alpha_i - \frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{k} \alpha_i \alpha_j y_i y_j K(x_i . x_j) \qquad (۴-۳)$$

با شرط:

$$\sum_{i=1}^{k} \alpha_i y_i = 0 \;\&\; 0 \leq \alpha_i \leq C \quad for\ i = 1. \ldots . k \tag{۵-۳}$$

پس تابع تصمیم‌گیری به‌صورت زیر می‌شود:

$$f(x) = sign\left(\sum_{support\ Vector} y_i \alpha_i^0\, k(x_i . x) + b^0 \right) \tag{۶-۳}$$

توابع کرنل مرسوم در جدول (۳-۱) لیست شده است.[۵۳],[۵۴]

جدول (۳-۱)توابع کرنل مرسوم

توابع کرنل	تعریف توابع
Linear	$\boldsymbol{X_i^T \cdot X_J}$
Polynomial	$(1 + \boldsymbol{X_i^T \cdot X_J})^P$
Sigmoid	$\boldsymbol{\tanh(\beta_o x_i^T x_j + \beta_1)}$
RBF	$\boldsymbol{exp(\frac{-\|x - x_i\|^2}{2\sigma^2})}$

۳-۴-۲- الگوریتم جنگل تصادفی (RF):

جنگل تصادفی درخت تصمیم‌های زیادی را تولید می‌کند. برای طبقه‌بندی یک شیء جدید برداری ورودی را در انتهای هر یک از درختان جنگل تصادفی قرار می‌دهد. هر درخت به ما یک طبقه‌بندی می‌دهد و میگوییم این درخت به آن کلاس "رای" می‌دهد. جنگل طبقه‌بندی‌ای که بیشترین رای را داشته باشد (بین همه درخت‌های جنگل)انتخاب می‌شود.

هر درخت به‌صورت زیر تشکیل می‌شود:

۱. اگر N تعداد حالت‌ها در مجموعه داده‌های train (مجموعه‌ی کار) باشد، N حالت را به‌صورت تصادفی با جایگذاری از داده‌های اصلی، نمونه‌گیری می‌کنیم. این نمونه مجموعه‌ی کار برای این درخت می‌باشد.

۲. اگر M متغیر داشته باشیم و m را کوچک‌تر از M در نظر بگیریم به‌طوری‌که در هر گره، m متغیر به‌صورت تصادفی از M انتخاب می‌شوند و بهترین جداسازی روی این m متغیر برای جداسازی گره استفاده می‌شود. مقدار m در طول ساخت جنگل ثابت در نظر گرفته می‌شود.

۳. هر درخت به‌اندازه‌ی ممکن بزرگ می‌شود. هیچ هرسی وجود ندارد.

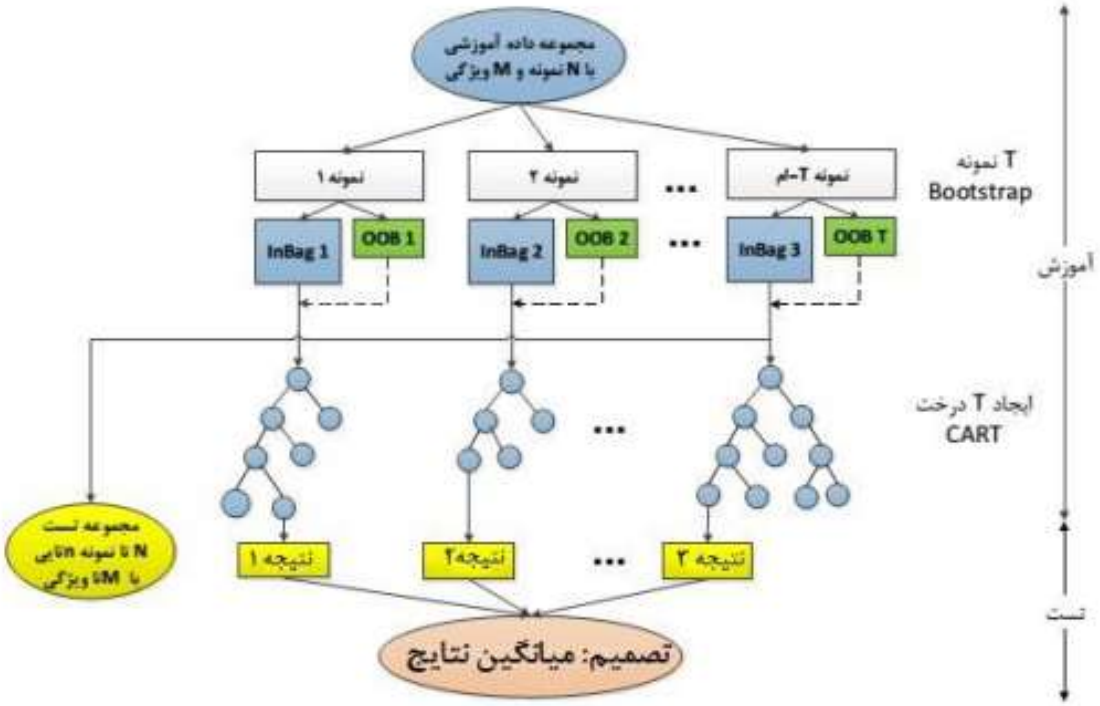
نرخ خطای جنگل به دو مورد زیر بستگی دارد:

–همبستگی بین هر دو درخت در جنگل. افزایش همبستگی نرخ خطای جنگل را افزایش می‌دهد.

–قدرت هر یک از درختان در جنگل. هر درخت با نرخ خطای کم یک طبقه بند قوی است. افزایش قدرت هر یک از درختان نرخ خطای جنگل را کاهش می‌دهد.

کاهش m هم همبستگی و هم قدرت را کاهش می‌دهد؛ و افزایشش هر دو را افزایش می‌دهد.

این الگوریتم در میان الگوریتم‌های فعلی ازنظر دقت بی‌نظیر است؛ و روی‌داده‌های بسیار بزرگ قابل‌اجراست و می‌تواند هزاران متغیر را بدون حذف متغیرها مدیریت کرد. همچنین برآوردی از مهم‌ترین متغیرها در طبقه‌بندی می‌دهد و راه کارایی برای برآورد داده‌های گم‌شده دارد. روند کلی الگوریتم RF به‌صورت ساده در شکل (۳-۴) نشان داده‌شده است.



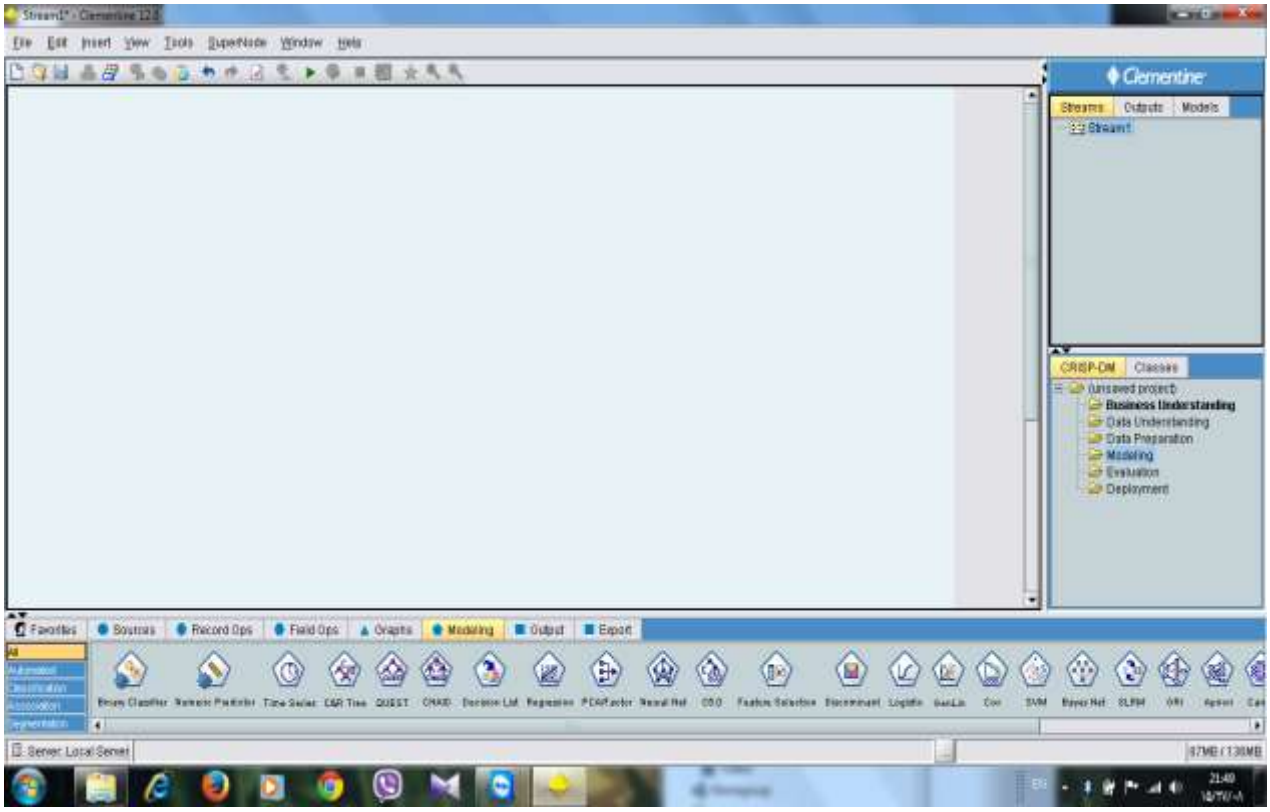
شکل (۳-۴): فرآیند الگوریتم RF

اصطلاح Bagging از مخففBootstrap Aggregating با دست آماده است.[۵۵] Baggingروشی است که از طریق نمونه‌برداری دوباره تصادفی از مجموعه‌ی داده‌های اصلی و همراه با جایگزینی، برای ایجاد داده‌های آموزشی بکار می‌رود و در این مرحله هیچ‌کدام از داده‌های انتخاب‌شده از نمونه‌های ورودی را برای تولید زیرمجموعه‌های بعدی حذف نمی‌کند، بدین ترتیب واریانس نیز کاهش می‌یابد. ازاین‌رو برخی از داده‌ها ممکن است بیش از یک‌بار در شاخه‌های آموزشی استفاده شوند، درحالی‌که برخی از داده‌های دیگر که در مدل‌سازی مؤثر نیستند، هرگز استفاده نمی‌شوند. بنابراین ثبات بیشتری برای مدل به دست می‌آید و مدل را در برابر تغییرات جزئی درداده های ورودی قابل‌اعتمادتر می‌کند و دقت پیش‌بینی آن را افزایش می‌دهد.[۵۶]

آن دسته از نمونه‌هایی که در آموزش درختان در فرآیند Bagging انتخاب نمی‌شوند شامل بخشی از زیرمجموعه‌هایی می‌شوند که الگوهای خارج از کیسه (OOB)نامیده می‌شوند. و این قسمت در روش RF می‌تواند برای ارزیابی عملکرد مدل استفاده شود. به‌این‌ترتیب RF می‌تواند تخمین غیر مرتبط داخلی از خطای تعمیم را محاسبه کند بدون اینکه از زیرمجموعه‌های داده‌های خارجی استفاده کند. [۵۷]

۳-۵ -نرم‌افزار پیشنهادی

نرم‌افزار پیشنهادی در این پژوهش نرم‌افزار کلمنتاین است. نرم‌افزار کلمنتاین[1] نرم‌افزاری بسیار قوی در حوزه داده‌کاوی هست. صفحه اصلی نرم‌افزار در شکل زیر نشان داده‌شده است.



شکل (۳-۵): صفحه اصلی نرم‌افزار کلمنتاین

1 Clementine

اجزای این صفحه عبارت‌اند از:

الف- صفحه‌نمایش جریان: بزرگ‌ترین فضای موجود در صفحه اصلی نرم‌افزار، صفحه‌نمایش جریان نامیده می‌شود که از آن می‌توان برای ساختن و تغییر جریان‌های داده استفاده نمود. هر عملیاتی به‌وسیله یک گره نشان داده می‌شود و برای نمایش جریان داده‌ها، گره‌ها به یکدیگر متصل می‌شوند.

ب- صفحه رنگی گره‌ها: این صفحه در پایین صفحه اصلی نرم‌افزار قرار داشته و شامل تمامی گره‌های موردنیاز برای ایجاد جریان‌ها می‌باشد. مجموعه گره‌ها عبارت‌اند از:

منبع[1]: گره‌هایی که داده‌ها را وارد نرم‌افزار می‌کند.

عملیات ثبت[2]: گره‌هایی که انجام فعالیت‌هایی چون انتخاب، ترکیب و افزودن بر روی رکوردهای داده را میسر می‌سازد.

عملیات رکورد[3]: گره‌هایی که انجام فعالیت‌هایی چون فیلترسازی، مشخص نمودن نوع داده‌ها و افزودن فیلد جدید را بر روی فیلدهای داده میسر می‌سازد.

گراف[4]: گره‌هایی که به‌طور گرافیکی داده‌ها را قبل و بعد از مدل‌ساز نمایش می‌دهند مثل هیستوگرام‌ها، نمودار ارزیابی و ...

مدل‌سازی[5]: گره‌هایی هستند که الگوریتم‌های قابل‌استفاده توسط کلمنتاین برای انجام مدل‌سازی استفاده می‌کنند.

خروجی[1]: گره‌هایی هستند که انواع مختلفی از خروجی‌ها را برای داده‌ها فراهم می‌سازد. این خروجی هارا می‌توان بر نرم‌افزاری‌های دیگری چون اکسل[2] فرستاد.

1 Source
2 Record Ops
3 Field Ops
4 Graphs
5 Modeling

۳-۵-۱ - مدیریت کلمنتاین:

این صفحه در سمت راست صفحه اصلی قرارداد. با استفاده از گزینه استریم۳ در این صفحه می‌توان جریان‌های ایجادشده را ذخیره، حذف و یا تغییر نام دارد.

گزینه خروجی: شامل تعداد زیادی فایل مانند گراف‌ها و جداول می‌باشد که با عملیات جریان داده به وجود آمده‌اند. می‌توان این فایل ها را ذخیره، حذف و یا تغییر نام داد.

گزینه مدل‌سازی: قوی‌ترین ابزار این صفحه است. این بخش شامل تمام اجزای مدلی که در نرم‌افزار ایجادشده‌اند، می‌باشد.

۳-۵-۲- پروژه‌های کلمنتاین:

این صفحه در قسمت پایین و سمت راست صفحه اصلی قرار دارد و برای ایجاد پروژه جدید و یا مدیریت نمودن پروژه‌های داده‌کاوی مورداستفاده قرار می‌گیرد. دو راه برای مشاهده پروژه‌های ایجادشده در نرم‌افزار وجود دارد. مشاهده در دو نمای کلاس۴ و کریسپ امکان‌پذیر است. گزینه کریسپ سازمان‌دهی پروژه‌ها را بر اساس فرآیندهای استاندارد ساخته‌شده در زمینه داده‌کاوی انجام می‌دهد. استفاده از این ابزار هم برای تازه‌کاران و هم افراد باتجربه بسیار مفید است. گزینه کلاس یک روش برای سازمان‌دهی پروژه‌ها بر اساس نوع اجزایی که ایجادشده‌اند مهیا می‌سازد. این نما برای زمانی که فهرستی از داده‌ها، جریان‌ها و مدل‌ها داریم، مؤثر است.

1Output

2 Excel

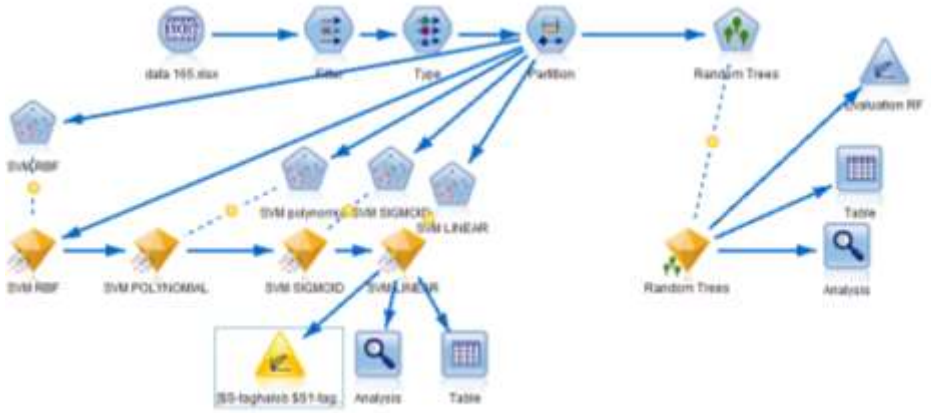
3 Stream

4 Classes

۳-۵-۳- جریان‌ها[1]:

داده‌کاوی، از نرم‌افزار کلمنتاین جهت تمرکز بر فرآیند اجرای مدل‌ها بر روی‌داده‌ها استفاده می‌نماید. این کار از طریق یک مجموعه از گره‌ها که موسوم به جریان می‌باشند، انجام می‌پذیرد. گره‌ها نشان‌دهنده مجموعه فعالیت‌هایی هستند که می‌بایست بر روی‌داده‌ها انجام پذیرد و ارتباطات بین آن‌ها نشان‌دهنده جهت انتقال اطلاعات می‌باشد. به‌عبارت‌دیگر جریان داده[2] جهت خواندن اطلاعات در کلمنتاین، اجرای داده‌ها از طریق یکسری از ابزارها و سپس انتقال آن‌ها به یک مقصد نهایی مانند فایل‌های اس پی اس اس[3] استفاده می‌نماییم.

آنچه ما در این پژوهش به کمک نرم‌افزار کلمنتاین انجام داده‌ایم استفاده از الگوریتم ماشین بردار پشتیبان با کرنل‌های مختلف آن و الگوریتم جنگل تصادفی و ترکیب دو الگوریتم ماشین بردار پشتیبان و جنگل تصادفی به‌طور هم‌زمان برای فرآیند دسته‌بندی می‌باشد که نمای کلی آن را در شکل (۳-۶) مشاهده می‌کنید.



شکل (۳-۶): شماتیک مدل

1 Stream

2 Data Stream

3 Spss

۳-۶ - جمع‌بندی

با توجه به توضیحات داده‌شده، تحقیق حاضر ازنظر هدف، تحقیقی کاربردی است. به دلیل این‌که ابزار پیشنهادی (مدل مورداستفاده) به‌صورت اجرایی در یک سازمان مورداستفاده قرار می‌گیرد. در این فصل به بیان روش تحقیق، فرآیند تحقیق، جامعه آماری و معرفی و شرح تکنیک و الگوریتم‌های مورداستفاده و نرم‌افزار بکار گرفته‌شده پرداختیم.

فصل ۴

تجزیه‌وتحلیل اطلاعات

فصل ۴ تجزیه و تحلیل اطلاعات

۴-۱- مقدمه

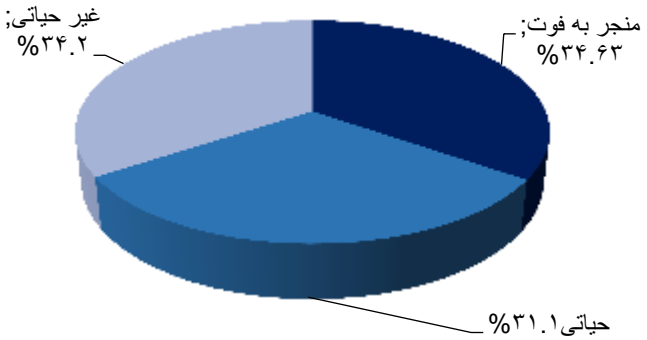
همان‌طور که در بخش‌های قبلی بیان شد در این تحقیق با بررسی ۱۰۰۰ پرونده از پرونده‌های تصادفی که از طریق بیمه‌نامه شخص ثالث ادعای خسارت نموده بودند سعی در کشف الگوهای تقلبی نموده است.

نوع ماشین صاحب بیمه، مدل آن، مدل ماشین مقصر، اعتبار بیمه‌نامه، سابقه بیمه، نوع صدمه، جنسیت مالک، ساعت وقوع حادثه، کروکی، مبلغ خسارت، تعداد نفرات آسیب‌دیده و علت حادثه و تقلبی بودن یا نبودن ادعای خسارت اطلاعاتی هستند که در این تحقیق از میان هر پرونده استخراج‌شده است. در این فصل روش پیشنهادی مطرح و به‌صورت کامل شرح و بسط داده می‌شوند. هدف ما از این فصل اعمال تکنیک ماشین بردار پشتیبان و جنگل تصادفی بر روی دیتا بیس یکی از شعب بیمه دی تهران است.

۴-۲- آمار توصیفی

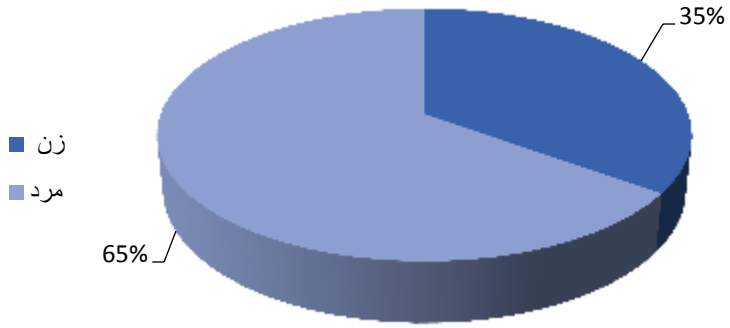
در این بخش به بررسی آمار توصیفی داده‌های مشتریان بیمه می‌پردازیم

بر اساس آنچه در شکل (۴-۱) مشاهده می‌شود مشتریان بیمه ازنظر نوع صدمه‌ای که به شخص ثالث آن‌ها واردشده است به سه دسته تفکیک‌شده‌اند که ۳۴.۶۳٪ افرادت است پوشش بیمه فوت‌شده‌اند،۳۴.۲٪ آسیب غیر حیاتی مثل شکستگی دست‌وپا دیده‌اند و ۳۱.۱٪ دچار آسیب حیاتی مثل صدمات لگن و نخاع شده‌اند.



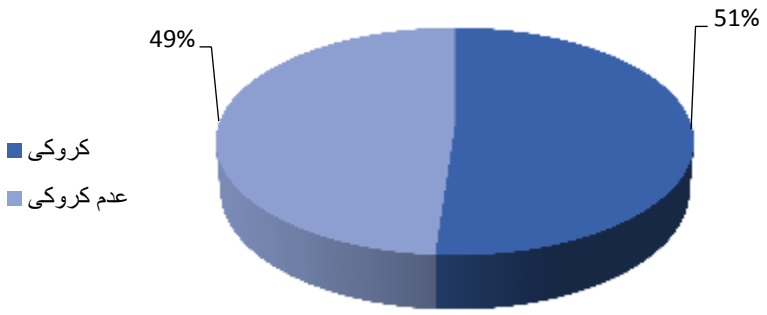
شکل (۴-۱): مشتریان بیمه به تفکیک نوع خسارت

در تفکیک مشتریان بر اساس جنسیت آن‌ها طبق شکل (۴-۲) ۳۵٪ مشتریان زن و ۶۵٪ مشتریان مرد بودند.



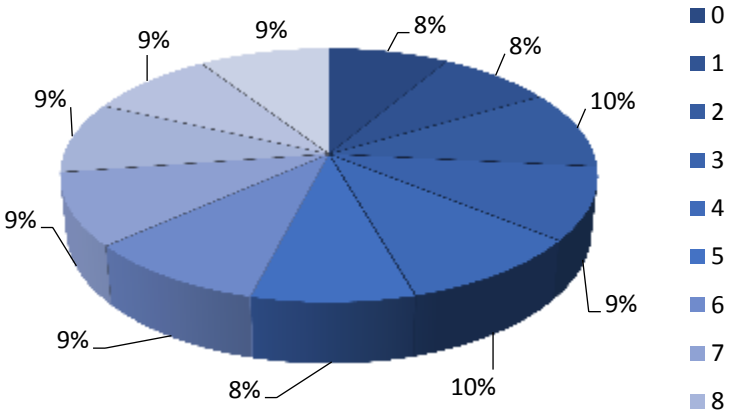
شکل (۴-۲): مشتریان بیمه به تفکیک جنسیت

اگر بخواهیم ادعای خسارت بیمه‌شدگان را برحسب وجود یا عدم وجود کروکی صحنه تصادف به دودسته تقسیم کنیم طبق شکل (۴-۳) مشاهده می‌کنیم که ۵۱٪ مشتری دارای کروکی بوده‌اند و ۴۹٪ کروکی‌ای از صحنه تصادف خود نداشته‌اند.



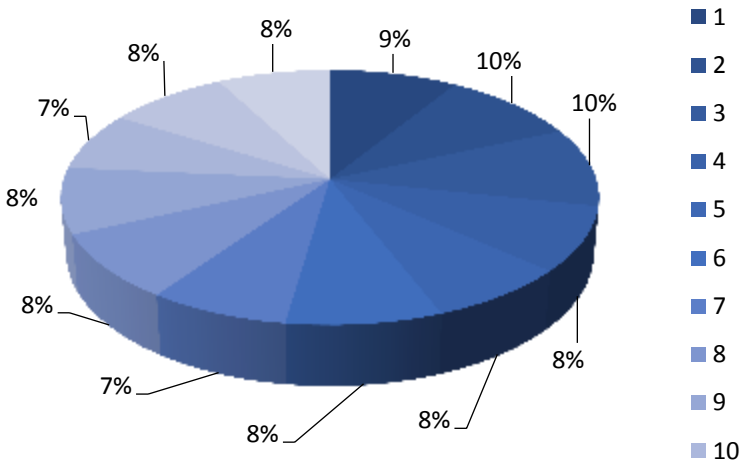
شکل (۴-۳): مشتریان بیمه به تفکیک وجود و عدم وجود کروکی

در تقسیم‌بندی دیگری سعی شده است که مشتریان بیمه را از جهت سابقه سالیانه بیمه‌نامه، تفکیک کرد و طبق شکل (۴-۴) نتایج زیر حاصل شد.



شکل (۴-۴): مشتریان بیمه به تفکیک سابقه سالیانه بیمه

در تفکیک مشتریان بر اساس اعتبار ماهیانه بیمه‌نامه آن‌ها تا پایان سال ۱۳۹۴، طبق شکل (۴-۵) نتایج زیر حاصل شد.

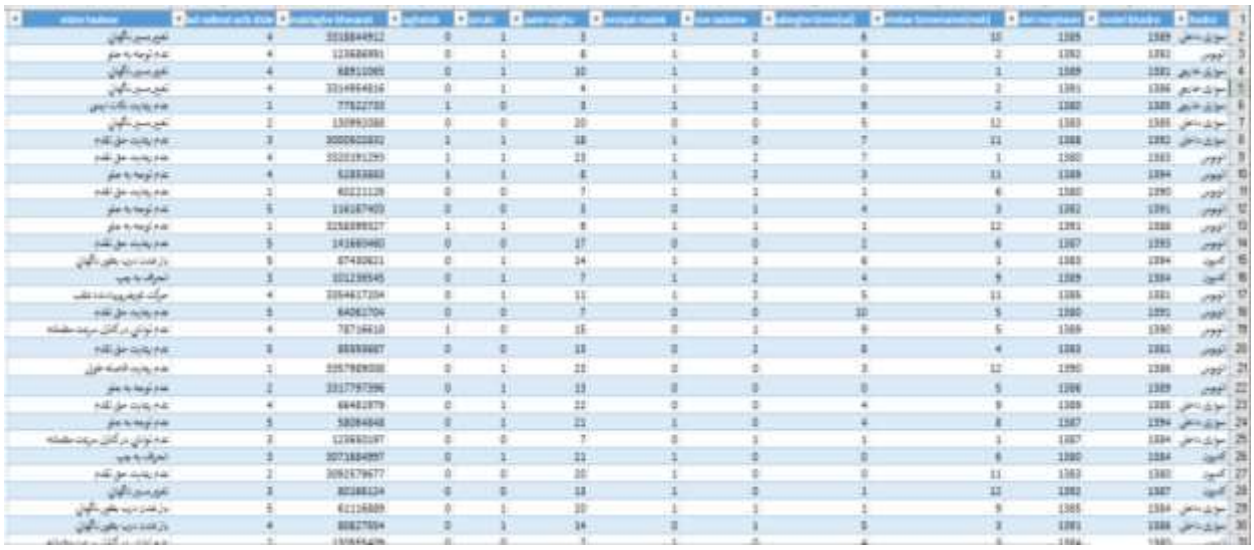


شکل (۴-۵): مشتریان بیمه به تفکیک اعتبار ماهیانه بیمه‌نامه

۴-۳-آمار استنباطی

۴-۳-۱- آماده سازی داده ها

با توجه به این که داده های این تحقیق متغیری مبنی بر تقلبی بودن یا نبودن ادعای خسارت را شامل میشوند این متغیر را بعنوان متغیر هدف درنظر گرفته و با استفاده از روش دسته بندی به پیش بینی تقلبات صورت گرفته خواهیم پرداخت ودر این راستا از ۶۰ در صد داده های موجود بعنوان داده های آموزشی واز ۴۰ درصد باقیمانده برای تست مدل استفاده خواهیم کرد. قسمتی از داده هارا در شکل (۴-۶) مشاهده میکنید.



شکل (۴-۶): داده‌های اولیه در اکسل

فرآیند نرمال‌سازی داده‌ها به‌طور خودکار در نرم‌افزار کلمنتان مورداستفاده در این تحقیق انجام می‌شود. همچنین بانک اطلاعاتی مورداستفاده در این تحقیق تعدادی اطلاعات ازدست‌رفته در رکوردهای مختلف دارد. به این دلیل که در بعضی از پرونده‌ها اطلاعاتی که موردنیاز این تحقیق است موجود نبوده و یا در هنگام ثبت اطلاعات سه‌ها وارد نشده‌اند؛ که نرم‌افزار کلمنتاین مشکل این داده‌های ازدست‌رفته را با تشکیل ابر گره به‌طور خودکار برطرف می‌کند.

۴ -۳-۲ - مدل‌سازی

همان‌طور که پیش‌تر بیان شد در این پژوهش از دو الگوریتم ماشین بردار پشتیبان و جنگل تصادفی برای دسته‌بندی مشتریان بیمه و تائید صحت ادعای خسارت آن‌ها و پیش‌بینی ادعاهای آتی استفاده می‌شود که در ادامه به بیان چگونگی مدل‌سازی آن می‌پردازیم.

۴-۳-۲-۱-مدل الگوریتم ماشین بردار پشتیبان (SVM)

SVM یا ماشین بردار پشتیبان، یک دسته‌بند یا مرزی است که با معیار قرار دادن بردارهای پشتیبان، بهترین دسته‌بندی و تفکیک داده‌ها را برای ما مشخص می‌کند. در SVM فقط داده‌های قرارگرفته در بردارهای پشتیبان مبنای یادگیری ماشین و ساخت مدل قرار می‌گیرند و این الگوریتم به سایر نقاط داده حساس نیست و هدف آن‌هم یافتن بهترین مرز در بین داده‌هاست به‌گونه‌ای که بیشترین فاصله ممکن را از تمام دسته‌ها (بردارهای پشتیبان آن‌ها) داشته باشد.

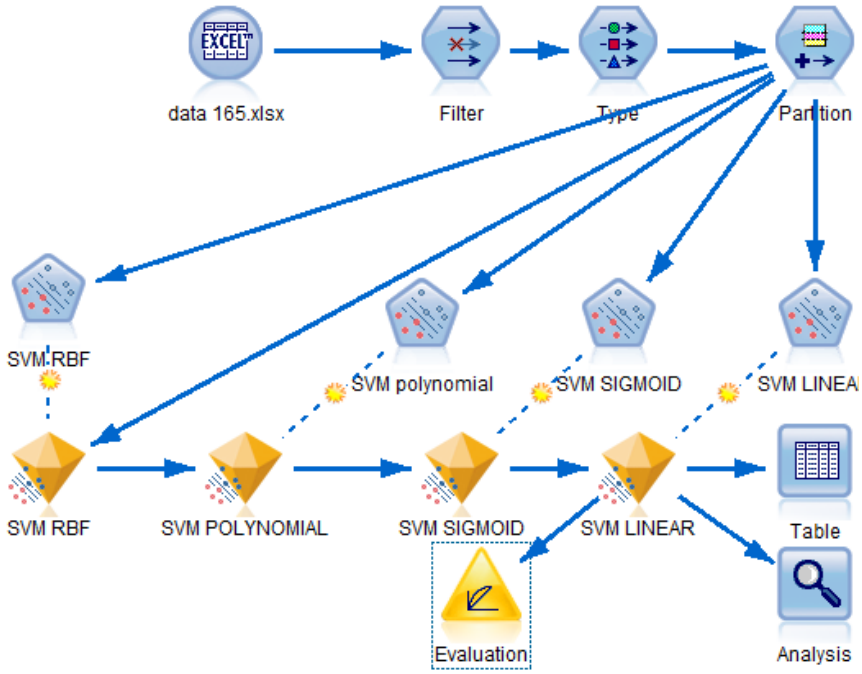
این الگوریتم از طریق روش ازدحام ذرات بهینه‌سازی شده است. الگوریتم ماشین بردار پشتیبان، الگوریتم بسیار قدرتمندی در دسته‌بندی و تفکیک داده‌ها هستند بخصوص زمانی که با سایر روش‌های یادگیری ماشین تلفیق شود به‌شرط اینکه توابع نگاشت و کرنل‌ها را به‌درستی انتخاب کنیم، بسیار خوب عمل می‌کند.

ماشین‌های بردار پشتیبان برای حل مسائل غیرخطی، ابعاد مسئله را از طریق توابع کرنل تغییر می‌دهند. انتخاب کرنل برای SVM به حجم داده‌های آموزشی و ابعاد بردار ویژگی بستگی دارد. به‌عبارت‌دیگر، باید با توجه به این پارامترها تابع کرنلی را انتخاب نمود که توانایی آموزش برای ورودی‌های مسئله را داشته باشد. در عمل چهار نوع کرنل خطی[1] کرنل چندجمله‌ای[2]، کرنل

[1] .Linear kernel
[2] . Polynomial kernel

سیگموییدی[1] کرنل گوسی[2] (RBF)بکار گرفته می‌شوند؛ که هرکدام دارای دقت متفاوتی هستند و ما در این پژوهش تمامی این کرنل هارا بررسی نموده‌ایم تا تابعی با بهترین دقت در تست را بیابیم.

نمایشی از گره‌ها در ساخت مدل را در تصویر (۴-۷) مشاهده می‌کنید.



شکل (۴-۷): مدل‌های SVMدر کلمنتاین

در جداول زیر به تفکیک کرنل‌های مختلف ماشین بردار پشتیبان دقت هر کرنل در آموزش و تست داده‌ها بررسی‌شده است.

جدول (۴-۱): الگوریتم SVM کرنل RBF

'Partition'	1_Training		2_Testing	
Correct	580	98.14%	299	73.28%
Wrong	11	1.86%	109	26.72%
Total	591		408	

[1] . Sigmoid kernel
[2] . Radial Base Function kerne

طبق آنچه جدول بالا مشخص است ۵۹۱ نفر به‌عنوان داده‌های آموزشی در نظر گرفته‌شده‌اند که از این تعداد ۵۸۰ نفر به‌درستی پیش‌بینی‌شده‌اند و ۴۰۸ نفر به‌عنوان داده‌های آزمایشی لحاظ شده‌اند که از این تعداد ۲۹۹ نفر به‌درستی پیش‌بینی شدند. ازاین‌رو دقت تست مدل ما با استفاده از این کرنل ۷۳.۲۸٪ می‌باشد

جدول (۴-۲): الگوریتم SVM کرنل Polynomial

'Partition'	1_Training		2_Testing	
Correct	586	99.15%	323	79.17%
Wrong	5	0.85%	85	20.83%
Total	591		408	

در این مدل در جدول (۴-۲)، ۳۲۳ نفر از ۴۰۸ نفر مورد تست قرارداد شده به‌درستی پیش‌بینی‌شده‌اند و دقت این مدل در تست ۷۹.۱۷٪ محاسبه‌شده است.

جدول (۴-۳): الگوریتم SVM کرنل Sigmoid

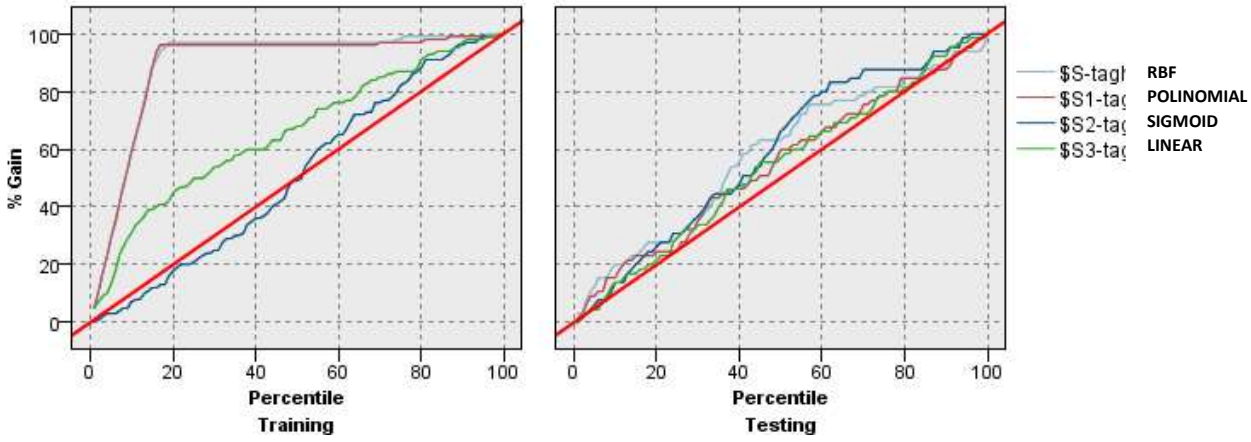
'Partition'	1_Training		2_Testing	
Correct	491	83.08%	343	84.07%
Wrong	100	16.92%	65	15.93%
Total	591		408	

طبق جدول (۴-۳) در کرنل سیگموییدی نیز ۳۴۳ نفر از ۴۰۸ نفر تست شده بدرستی پیش بینی شده اند که معادل ۸۴.۰۷٪ دقت مدل در تست است.

جدول (۴-۴): الگوریتم SVM کرنل Linear

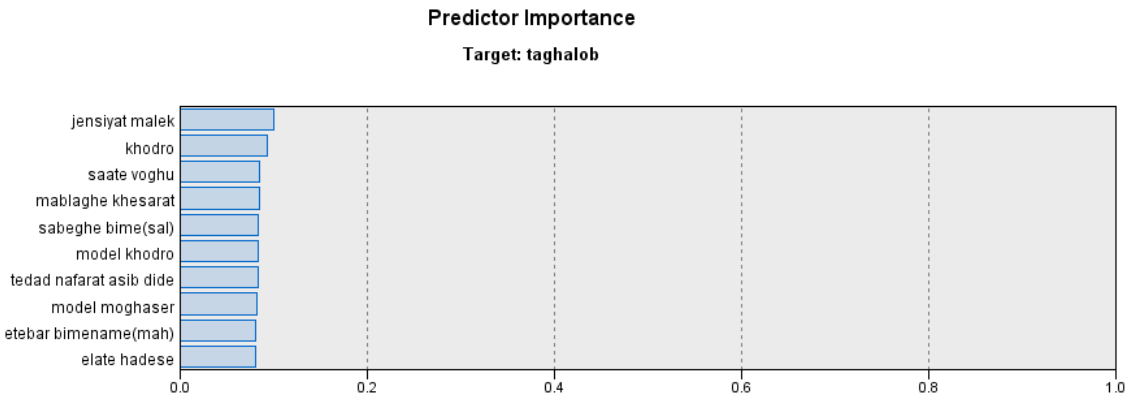
'Partition'	1_Training		2_Testing	
Correct	487	82.4%	326	79.9%
Wrong	104	17.6%	82	20.1%
Total	591		408	

طبق جدول (۴-۴)، در این کرنل صحت ادعای ۳۲۶ نفر از خسارت دیدگان به‌درستی تعیین‌شده که معادل ۷۹.۹٪ دقت در تست این مدل است.



شکل (۴-۸): مقایسه نمودارهای کرنل‌های مختلف svm

طبق شکل (۴-۸)، با استفاده از نمودار Gain به بررسی و مقایسه گرافیکی چهار کرنل استفاده‌شده پرداختیم. طبق آنچه مشاهده می‌شود هر چهار نمودار مربوطه بالای خط قرمز نیمسازی قرارگرفته‌اند که نشان‌دهنده مطلوب و موردقبول بودن هر چهار کرنل است؛ و در بین این چهار نمودار مشاهده می‌شود که نمودار کرنل Sigmoidاز سایرین بالاتر و مطلوب‌تر و دقیق‌تر است.



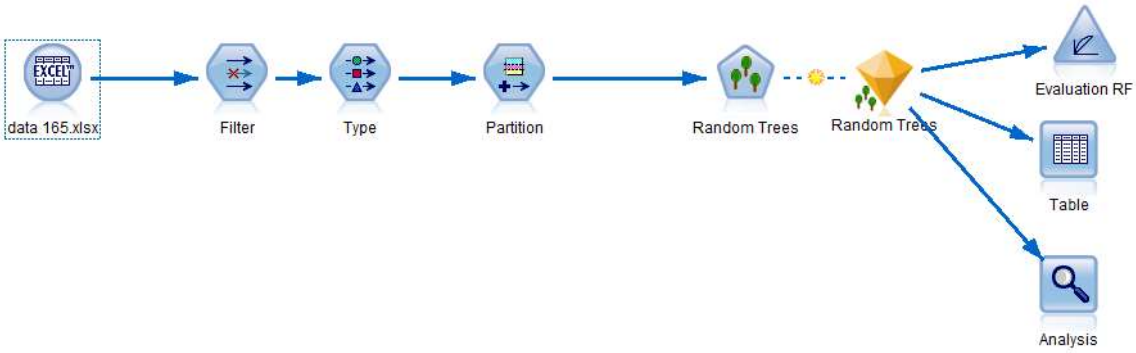
شکل (۴-۹): فیلدهای مهم در مدل SVM SIGMOID

در نمودار بالا در شکل (۴-۹) موارد مهم تاثیرگزار برساخت مدل کرنل سیگموییدی ماشین بردار پشتیبان به ترتیب اهمیت آن‌ها آمده است. طبق نمودار مهم‌ترین عامل تاثیرگذاز در ساخت این مدل جنسیت مالک و کم‌اهمیت‌ترین عامل علت حادثه قلمداد شده است.

۴-۳-۲-۲-مدل جنگل تصادفیRandom forest

الگوریتم جنگل تصادفی از کل مجموعه داده‌ها با ویژگی‌های گوناگون، زیرمجموعه‌هایی را در نظر گرفته و درخت‌هایی را ایجاد می‌کند و سپس بر اساس رای گیری، بهترین تصمیم را از درخت‌ها اتخاذ می‌کند که نهایتاً شخص موردنظر متعلق به کدام‌یک از کلاس‌ها می‌باشد که آیا شخص متقلب است یا خیر؟

نمایشی از گره‌ها جهت ساخت مدل در شکل (۴-۱۰) نمایان است.



شکل (۴-۱۰): مدل جنگل تصادفی در کلمنتاین

- Analysis
  - Analysis of data 165.xlsx (Oct 9, 2016 7:32:39 PM)
  - Analysis of data 165.xlsx (Oct 9, 2016 7:35:05 PM)
    - Number of records: 591
    - Analysis Accuracy: 98.985%
- Fields
  - Target
    - taghalob
  - Inputs
- Build Settings
  - Use partitioned data: true
  - Partition: Partition
  - Target*: taghalob
  - Predictors*: [khodro, model khodro, model moghaser, etebar bimename(mah), sabeghe bime(sal), noe sadame, jensiyat malek, saate voghu, koruk
  - Number of models to build: 100
  - Sample size: 1.0
  - Handle imbalanced data: false
  - Use weighted sampling for variable selection: false
  - Maximum number of nodes: 10,000
  - Maximum tree depth: 10
  - Minimum child node size: 5
  - Specify number of predictors to use for splitting: false
  - Total number of predictors: 12
  - Number of predictors for splitting: 1
  - Stop building when accuracy can no longer be improved: true
  - Use misclassification costs: false
  - Default cost increase for ordinal target: Linear
  - Maximum percentage of missing values: 70

شکل (۴-۱۱): جزئیات ساخت مدل جنگل تصادفی

در شکل (۴-۱۱) جزئیات ساخت مدل جنگل تصادفی آمده است. دقت آنالیز مدل در قسمت آموزش ۹۸.۹۸٪ تعیین‌شده. همچنین در جدول (۴-۵) می‌بینیم که دقت مدل در قسمت تست و آزمایش ۸۰.۳۹٪ تعیین‌شده.

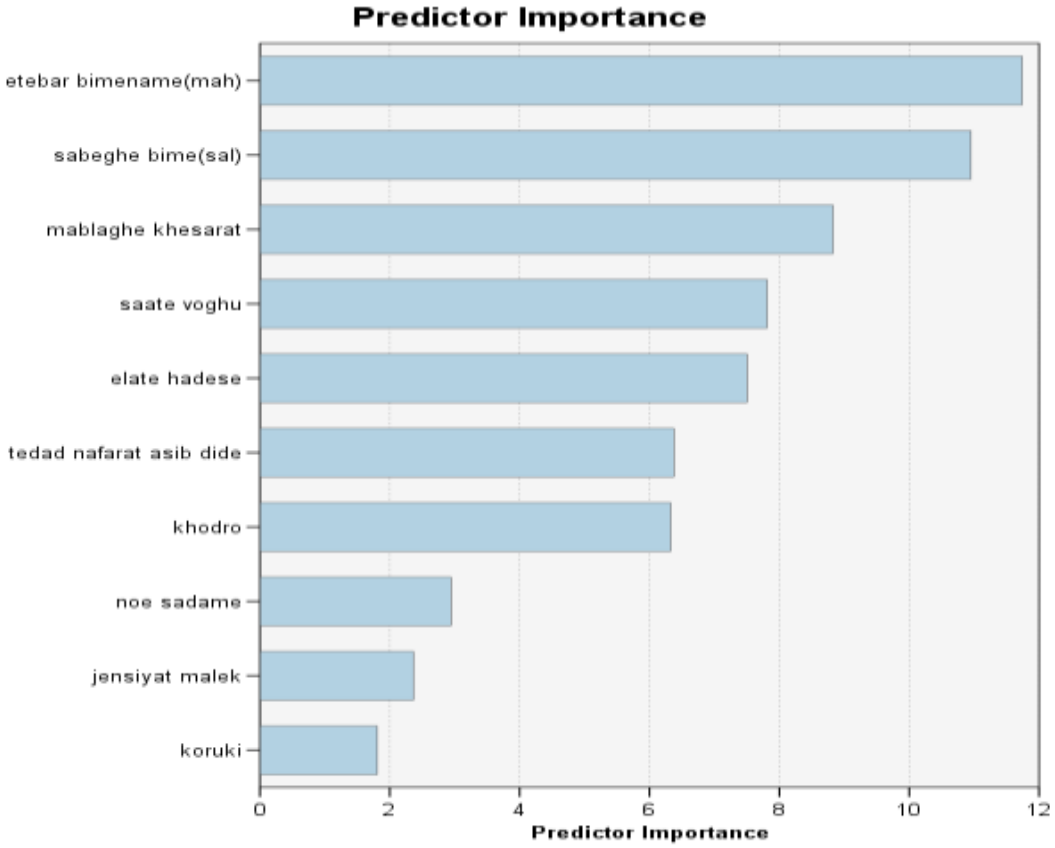
جدول (۴-۵): الگوریتم جنگل تصادفی

'Partition'	1_Training		2_Testing	
Correct	585	98.98%	328	80.39%
Wrong	6	1.02%	80	19.61%
Total	591		408	

جدول (۴-۶): اطلاعات مدل RF

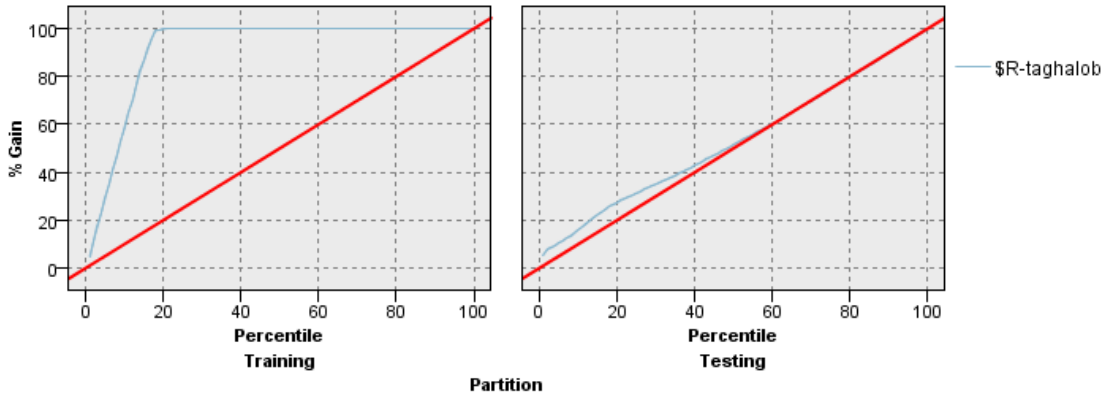
Model Information	
Target Field	taghalob
Model Building Method	Random Trees Classification
Number of Predictors Input	10
Model Accuracy	0.749
Misclassification Rate	0.251

همان‌طور که در جدول(۴-۶) مشخص است دقت مدل ۷۵٪ محاسبه‌شده و نرخ خطای این مدل ۲۵٪ می‌باشد.



شکل (۴-۱۲): فیلدهای مهم در مدل Random forest

طبق نموداربالادر شکل (۴-۱۲) که نشان‌دهنده‌ی ترتیب اهمیت متغیرهای مختلف در ساخت مدل جنگل تصادفی برای تشخیص صحت ادعای خسارت متقاضیان بیمه است، مهم‌ترین عامل تاثیرگزار اعتبار بیمه‌نامه متقاضی است و کم اهمیت‌ترین عامل داشتن یا نداشتن کروکی است.

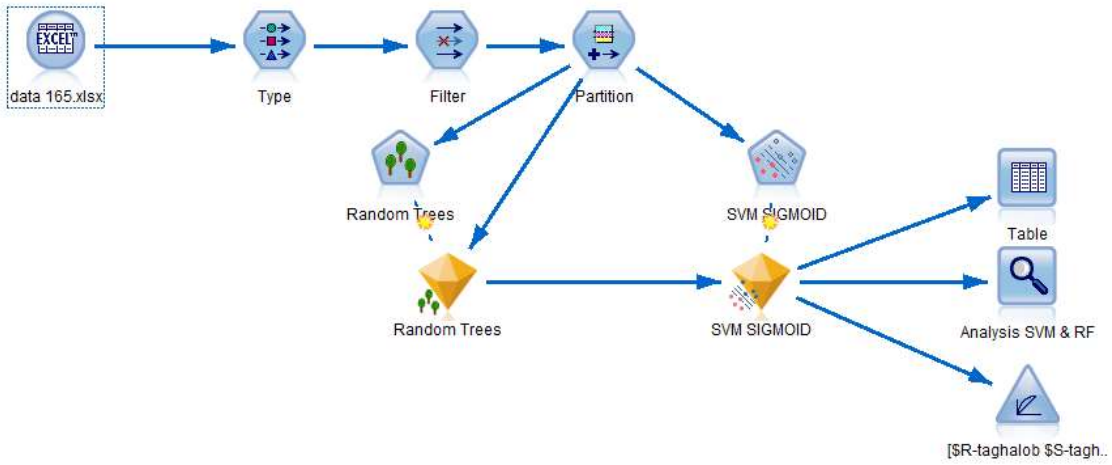


شکل (۴-۱۳): نمودار آموزش و تست مدل جنگل تصادفی

نمودار Gainدر شکل (۴-۱۳) مربوط به مدل جنگل تصادفی است و طبق آنچه مشاهده می‌شود به علت این‌که هردو نمودار آموزش و تست بالای خط نیمساز قرارگرفته‌اند پس این مدل برای دسته‌بندی و پیش‌بینی صحت ادعاها قابل‌قبول و مناسب است.

۴-۳-۲-۳-مدل ترکیب الگوریتم ماشین بردار پشتیبان SVM و الگوریتم جنگل تصادفی RF

در این مدل دو الگوریتم SVM با کرنل Sigmoidکه دقت بالاتری در تست نسبتاً به سایر کرنل‌ها داشت را با الگوریتم جنگل تصادفی که آن‌هم دقت مطلوبی داشت در نرم‌افزار کلمنتاین با یکدیگر ادغام می‌کنیم. سپس الگوریتم ترکیبی را اجرا می‌کنیم، که نمایش گرافیکی ساخت این مدل را در شکل (۴-۱۴) مشاهده می‌کنید.

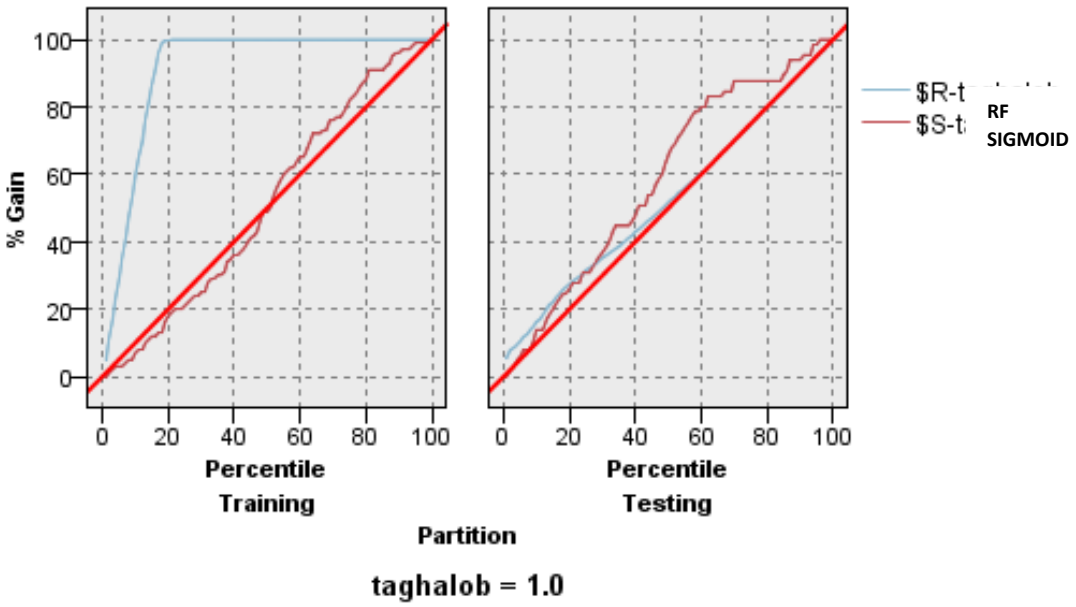


شکل (۴-۱۴): ترکیب دو مدل SVM SIGMOID & RF

با توجه به جدول (۴-۷) مشخص می‌شود که از ۳۷۵ نفر بیمه‌شده که مورد تست قرارگرفته‌اند ادعای خسارت ۳۱۹ نفر به‌درستی تشخیص داده‌شده است و دقت این مدل ترکیبی ۸۵.۰۷٪ تعیین‌شده که از دقت مدل سیگموییدی و جنگل تصادفی به‌طور جداگانه از یکدیگر بالاتر و مطلوب‌تر است.

جدول (۴-۷): الگوریتم SVM کرنل Sigmoid و RF

'Partition'	1_Training		2_Testing	
Correct	486	99.79%	319	85.07%
Wrong	1	0.21%	56	14.93%
Total	487		375	



شکل (۴-۱۵): نمودار آموزش و تست مدل ترکیبی Svm SIGMOID & RF

با توجه به نمودار بالا در شکل (۴-۱۵) مشاهده می‌شود که هرچند که طبق انتظارمان نمودار مدل سیگموییدی بالاتر از نمودار جنگل تصادفی قرارگرفته است که حاکی از دقت بالاتر آن از جنگل تصادفی است. اما به علت اینکه هر دو نمودار بالای خط قرمز نیمساز قرارگرفته‌اند پس دقت مدل ترکیبی در تست قابل‌قبول است.

۴-۴ - جمع‌بندی

در این فصل به تجزیه‌وتحلیل اطلاعات پرونده‌های بیمه‌ای ۱۰۰۰ نفر از بیمه‌شدگان شخص ثالث بیمه دی تهران در سال ۹۴ پرداختیم و پس از آماده‌سازی و پالایش داده‌ها به مدل‌سازی پرداختیم و در مرحله مدل‌سازی از مدل‌های ماشین بردار پشتیبان با چهار کرنل مختلف و مدل جنگل تصادفی و مدل ترکیبی با دقت نرین کرنل ماشین بردار پشتیبان و جنگل تصادفی پرداختیم و سپس به آنالیز و تجزیه‌وتحلیل نتایج پرداختیم. در فصل بعدی به نتایج حاصل از این تجزیه‌وتحلیل می‌پردازیم.

فصل پنجم

نتیجه‌گیری

فصل ۵: نتیجه‌گیری

۵-۱-مقدمه

پژوهش حاضر به دنبال ارائه مدل برتری است که به کمک آن بتوان ادعاهای خسارت بیمه‌ای را که مشکوک به تقلب هستند را با دقت بالا مشخص کرده و سپس با استفاده از تحقیقات گسترده‌تر خبرگان و کارشناسان بیمه‌ای صحت این ادعاها را به‌طور قطعی مشخص کنیم. درواقع چنین مدلی باعث صرفه‌جویی در وقت و هزینه تحقیقات بیمه‌ای می‌شود و همچنین می‌تواند از تقلبات بیمه‌ای که ضررهای هنگفتی را متحمل شرکت‌های بیمه‌ای می‌کنند جلوگیری به عمل آورد. در فصل چهارم مدل‌های استفاده‌شده را مشاهده کردیم. در این فصل قصد داریم که از نتایج مدل‌ها نتیجه‌گیری کلی کنیم و مدل برتر را معرفی نماییم و در پایان پیشنهادها سازنده‌ای را برای آینده این‌گونه تحقیقات مطرح کنیم.

۵-۲ -مقایسه نتایج

همان‌طور که در جدول زیر مشاهده می‌کنید همه‌ی مدل‌های بکار گرفته‌شده دقت قابل قبولی در پیش‌بینی صحت ادعای خسارت بیمه‌شدگان از خود نشان داده‌اند؛ اما در بین کرنل‌های مختلف ماشین بردار پشتیبان، کرنل سیگموییدی دقت بالاتری را از سایر کرنل‌ها از خود نشان داد؛ که دقت آن حتی از دقت مدل جنگل تصادفی که آن‌هم دقت بالا و قابل‌توجهی از خود نشان داده بالاتر است؛ اما مشاهده می‌شود که وقتی از ترکیب دو مدل سیگموییدی و جنگل تصادفی باهم برای این پژوهش بهره می‌بریم دقت به ۸۵.۰۴٪ صعود می‌کند که دقت بسیار بالا و مطلوب و قابل قبولی است و می‌توان این مدل را به‌عنوان مدل برتر معرفی کرد و از آن بهره جست.

جدول (۵-۱): مقایسه دقت مدل‌های مختلف

مدل‌ها	دقت مدل در آموزش	دقت مدل در تست
SVM RBF	۹۸.۱۴٪	۷۳.۲۸٪
SVM POLYNOMIAL	۹۹.۱۵٪	۷۹.۱۸٪
SVM SIGMOID	۸۳.۰۸٪	۸۴.۰۷٪
SVM LINEAR	۸۲.۴٪	۷۹.۹٪
RF	۹۸.۹۸٪	۸۰.۳۹٪
SIGMOID & RF	۹۹.۷۹٪	۸۵.۰۴٪

۵-۳ -نتیجه‌گیری

در این تحقیق مقوله داده‌کاوی و تکنیک های آن موردبررسی قرار گرفت. با توجه به حساسیت صنعت بیمه به مقوله‌ی تقلب، علم داده‌کاوی به کمک ما آمد تا بتوانیم با دقت مطلوب ممیزی تقلب را در بیمه شخص ثالث یکی از شرکت‌های بیمه تهران به نام بیمه دی به کار ببندیم. در این پژوهش درصدد یافتن مدلی بودیم که بتواند با دقت بالا تقلبی بودن یا نبودن ادعاهای خسارت مشتریان بیمه شخص ثالث را تشخیص دهد و ازاین‌رو برای پیش‌بینی تقلبات آتی به یاری این شرکت بیمه بشتابد تا در وقت و هزینه و انرژی کارشناسان بیمه صرفه‌جویی شود و این کارشناسان بتوانند پس‌ازاین نوع غربالگری بر روی مشخصات بیمه‌شدگان با اطمینان بیشتری فرآیند قانونی تشخیص تقلب را طی نمایند. ازاین‌رو طبق آنچه نرم‌افزار کلمنتاین با ساختن مدل‌های ماشین بردار پشتیبان با کرنل‌های مختلف و مدل جنگل تصادفی و مدل ترکیبی سیگمویدی ماشین بردار پشتیبان و جنگل تصادفی به ما ارائه دادند، می‌توان این نتیجه را

برداشت کرد که وقتی از دو الگوریتم قدرتمند ماشین بردار پشتیبان و جنگل تصادفی به‌صورت ترکیبی استفاده می‌کنیم با دقت بالاتری می‌توانیم پیش‌بینی کنیم چراکه در این پژوهش پیش‌بینی تقلب با مدل ترکیبی ۸۵.۴٪ محاسبه شد که نسب به سایر مدل‌ها بالاتر است و خطای این مدل هم ۱۴.۹۳٪ بوده که از خطای سایر مدل‌ها کمتر است.

۵-۴ -پیشنهادات

در این تحقیق با توجه به بعضی از محدودیت‌ها تنها از یک شرکت بیمه و تنها در شهر تهران برای جمع‌آوری اطلاعات استفاده شد که می‌توان برای تحقیقات آینده از شرکت‌های بیمه دیگر و یا از شعبات دیگر بیمه دی استفاده‌شده در این پژوهش هم بهره برد. از طرفی می‌توان مشابه این پژوهش را بر روی بیمه‌های مختلفی مثل بیمه بدنه یا بیمه آتش‌سوزی و ... انجام داد.

در این تحقیق از متغیرهایی شبیه نوع خودرو و اعتبار بیمه‌نامه و سلیقه بیمه و ... به‌عنوان متغیر پیشگو بهره بردیم که می‌توان در تحقیقات آتی از متغیرهای متنوع دیگری هم استفاده کنیم مانند مشخصه‌های جمعیت شناختی، زیرا این مشخصات می‌توانند توصیف‌کننده میزان خطرپذیری مشتریان باشند و همچنین جمع‌آوری این اطلاعات از مشتریان ساده است و منع قانونی هم ندارند. ازجمله این مشخصات جمعیت شناختی می‌توان به موارد زیر اشاره کرد:

داده‌های مربوط به جرائم رانندگی، نوع جرائم رانندگی مشتری، تعداد سانحه‌های رانندگی مشتری در سال، محل زندگی مشتری، محل کار مشتری، میزان فاصله محل کار تا محل زندگی مشتری، میزان تراکم جمعیت محل کار و محل زندگی مشتری، تعداد افراد استفاده‌کننده از وسیله نقلیه، داشتن یا نداشتن

پارکینگ اختصاصی برای مشتری و ... درنتیجه می‌توان مشتریان را بهتر شناسایی کرد و برای آن‌ها تعرفه بیمه‌نامه‌ای با توجه به ویژگی‌های آن‌ها و سطح خطرپذیری آن‌ها تعیین کرد.

همچنین می‌توان مشابه این پژوهش را در سازمان‌های دیگری مانند تأمین اجتماعی، پلیس راهنمایی و رانندگی، بانک و غیره که در معرض خطر متقلبان قرار دارند پیاده‌سازی کرد؛ و ازآنجایی‌که در پژوهش حاضر شاهد افزایش دقت در مدل ترکیبی بودیم می‌توان پیشنهاد داد که از ترکیب مدل‌های معرفی‌شده در این پژوهش با مدل‌ها و الگوریتم‌های دیگرِ استفاده کنیم تا بهترین مدل را بتوانیم در عملیات داده‌کاوی در حوزه کشف تقلب بکار بندیم.

منابع

[1].B.Manjula, S.S.V.N.Sarma, Dr.A.Govardhan and R.LakshmanNaik **(1998)**"DFFS: Detecting Fraud in Finance Sector"International **J. of Advanced Engineering Sciences and Technologies, 9, 2, 178-182.**

{۲}. راه چمنی. ابوالقاسم.(۱۳۸۵)"تقلب و کلاه‌برداری تهدید همیشگی صنعت بیمه"**فصلنامه آسیا،۲**، ۱۶، ص ۱۹

[3].S.Viaene, M.Ayuso, M.Guillen,D.Van Gheel and G.Dedene**.(2007)**"Strategies for detecting fraudulent claims in the automobile insurance industry"**European J. of Operational Research,176,1,pp656-583**

[4]. Han, J. and Kamber, M.(**2006**) " Data Mining : Concepts and Techniques"Second Edition, Morgan Kaufman Publisher,**100,2**.

[5].Patil B. M. R. C. J. and Durga T.(**2010**)" Association rule for classification of type- 2 diabec pa ents. In Machine Learning and Compung (ICMLC)"

[6].Ngaie.W.T.and Yong Hu,Y.H.and Wong,Yijun Chen.and Xin Sun.(**2010**)"The Application of Data Mining Techniques in Financial Fraud Detection" Framework and an Academic Review of literature;Decision Support Systems,**50,3,pp559-569**

[7].Dionne.G, Gagne.R.(**2000**)"Replacement Cost Endorsement And Opportunistic Fraud In Automobile Insurance"Working **pp 00-01**

{۸}. محمد. بیگی، علی. اعظم (۱۳۸۴)"بحثی مقدماتی درباره تقلب بیمه ای: مورد بیمه شخص ثالث" **تازه های جهان بیمه**،۸۹، ص ۲۵-۳۶

[9]. Morley, B.(**2006**)" How the detection of insurance fraud. Psychology, Crime &amp" Law**12:163-180**

[10].Kassem.R and Higson.A.(**2012**)"The New Fraud Triangle Model"Emerging Ttends In Economics And Management Sciences,**15,8,pp191-195**

[11].Hebenton, B. (**2007**)" Insurance fraud in Taiwan: Reflections on regulatory effort and criminological complexity" International Journal of the Sociology of Law, **35,3, pp127-142**

{۱۲}. قانون اصلاح قانون بیمه اجباری مسئولیت مدنی دارندگان وسایل نقلیه موتوری زمینی در مقابل شخص ثالث،مصوب ۱۳۸۷.

{۱۳}. بیمه بدنه و شخص ثالث، رکورددار تقلب، سرویس اقتصادی جهان نیوز، ۱۳۸۹

{۱۴}. کاوه. حمیدرضا (۲۰۱۳)" ارزش های اخلاقی زیربنای اعتماد در صنعت بیمه" **بیمه ایران**،۱۲

[15].Derring.R .and A,Johnson .and D.J,Sprinkel.A.E.(**2006**)" AUTO INSURANCE FRAUD:MEASUREMENTS AND EFFORTS TO COMBAT IT" Risk Management And Insurance,Working ,**9,2,pp109-130**

[16].Dionne.G.and Gagne.R.(**2000**)"Replacement Cost Endorsement And Opportunistic Fraud In Automobe Insurance".Working **pp 00-01**.

[17].Viaene.S and Ayuso.M and Guillen.M and Vangeel.D.(**2007**)"Strategies For Detecting Fraudulent Claims In The Automobile Insurance Industry" **European J. Of Operation Research,pp565-83**

[18].Stephan.k,and Wilson . V,.(**2011**)" Online Banking Fraud Detection Based and Local on Global Behavior"www.thinkmind.org.

[19].T. Fawcett and F. Provost,(**1997**)"**Adaptive fraud detection**", Data Mining and Knowledge Discovery Journal, Kluwer Academic Publishers, **Vol. 1, No. 3, , pp. 291-316.**

[20].Lunt. T.F and et al. (**1990**)"A Real-Time intrusion Detection Expert System (IDES)"Final Technical Report ,Technical Report. SRl Computer Science Laboratory .SRl International, from http://www.wenke.gtisc.gatech .edu

[21].Anderson. D, and Frivold,T., and Tamaru. A. and Valdes,A(**1994**)" Next generation intrusion detection expert system (NDES)"software user's manual,beta-update release)

Technical Report SRIXSL-9547. Computer Science /Laboratory, SRI International, from www.thc.org/root .docs/intrusion-detect

[22].Burge,P and et al(**1999**)" Fraud Detection and Management inMobile Tele communications Networks".London:Royal Holloway University

[23].Ghosh, S and et al.(**1994**)"Credit card fraud detection with a neural-network" 27th Annual Hawaii International Conference on System Science. Los Alami, CA: IEEE Computer Society

[24].David .H.and Heikki M .and Padhraic S(**2001**)" Principles of Data Mining" The MIT Press

[25]. Iigun,K..and Kemmerer, R. A. and Porras, P. A.(**1995**)." State transition analysis: A rule-based intrusion detection approach" Software Engineering, **21,3,pp 181-199**.

[26].Iigun, K.(**1993**)" USTAT A Real-time intrusion detection system for UNIX. IEEE Symposium on Research in Security and Privacy",Oakland, CA: IEEE Symposium on Research in Securiry and Privacy.**pp 16-28**.

[27].Danna, A.and Gandy, O ,(2002)" All That Glitters is Not Gold: Digging Beneath the Surface of Data Mining," Journal of Business Ethics, 40,pp 373–386.

[28].Jeffery W. (2004)" Analyst in information science and Technology Policy, ' Data Mining : An Overview '14.2,pp152-160.

[29].Frawley. W, Piatetsky .G.(**1992**)" Knowledge Discovery I DataBases.ISSN **0738-4602**

[30].Hand. D.J (**1998**) "Review of Data mining", The American statistician**, 52, 112-118**

[31].Michael.S and Vipin K (**2006**)."Cluster Analysis: Basic Concepts and Algorithms. In Introduction to Data Mining by Pang-Ning Tan, Michael Steinbach and Vipin Kumar". Minnesota: Addison-Wesley Companion Book Site, p. **487-568**

[32]. Steinbach .M, Kumar V **(2006)**" Introduction to data mining". Pearson Addison-Wesley

[33]. Alizadeh S, Ghazanfari M,and Teimorpour B (**2011**)"Data Mining and Knowledge Discovery" Publication of Iran University of Science and Technology .۲nd ed.

{۳۴}.شهرابی،جمال،(۱۳۹۰)**"داده کاوی در مهندسی کیفیت و پایایی"**چاپ اول ،انتشارات جهاددانشگاهی واحد صنعتی امیرکبیر.ص۱۲۰.

[35]. Eapen, A.G.(**2004**)" **Application of Data mining in Medical Applications**". University of Waterloo,Ontario,Canada .

[36]. Beryy J.and Michael J.and Linoff Gordan S.,(**2004**)" Data Mining Techniques for Marketing Sales and Customer Relationship Management.John".Wiley and Sons Publishing Inc

[37]. Guo, L.(**2002**)"Applying data mining techniques in property" Casualty Insurance, **13 ,2, pp230-472.**

[38]..Hand D, . and Heikki M,and Padhraic S.(**2001**)" Mining. The MIT Press; Principles of Data"

[39].Sternberg .M.and Reynolds .R.G.(**1997**)" Using cultural algorithms to support reengineering of rule-based expert systems, in dynamic performance environments a case study in fraud detection" IEEE Transactions on Evolutionary ComputationI **,4,pp225-243.**

[40].Weisberg.H and Derrig .R(**1998**)"Quantitative methods for detecting fraudulent automobile bodily injury claims" Risques ,**35,pp75-101.**

[41].Belhadji E.B. and Dionne G. and Tarkhani, F. (**2000**) "A model for the detection of insurance Fraud" **The Geneva Papers On Risk and Insurance 25,4,pp 517-538.**

[42]. Vilene. S, and Derig. R and Baesens.B,and Dedene.G (**2002**)" A comparison of state of the art classification techniques for expert automobile insurance claim fraud detection" The **j. of risk and insurance, vol. 69,No.3,pp373-421**

[43]. Artis M and Ayuso M, and Guillén M.,(**2002**) " Detection of automobile insurance fraud with discrete choice models and misclassified claims". **The J. of Risk and Insurance, 55,pp325-340**

[44].Pathak. J.and Vidyarthi N.and , Summers L.(2005)"A fuzzy-based algorithm for auditors to

detect elements of fraud in settled ince claims" **Managerial Auditing Journal**

2,6,pp۷ 32-644.

[45].Vilene ,S.,and Dedene G,and Derig R.(**2005**)" Auto claim fraud detection using Bayesian

learning neural networks" Expert Systems with Applications **29,3,pp 653-666**

[46].Vilene, S and Ayes, M and Guillen, M(**2007**)" Wan Ghee, G. Deaden, Strategies for detecting fraudulent claims in the automobile insurance industry" European **J. of Operational Research 176 ,1,pp 565-583**

[47]. Bermudez .L and . Pérez J.M and . Ayes. M,and Gómez. E,. Vázquez ,F.J, (**2008**)"A. Bayesian Dichotomous, Model with asymmetric link for fraud in insurance**" Insurance:**

Mathematics and Economics 42,2, pp779-786.

[48].Rekha B, (**2011**)"Detecting Auto Insurance Fraud by Data Mining Techniques" **J. of Emerging Trends in Computing and Information Sciences Vol 2 No.4.pp156-162**

[49].Izadparast.S.and Farahi. A and Fath Nejad F .(**2012**) "Using Data Mining Techniques to Predict the Detriment Level of Car Insurance Customer"**J, of Information Processing and Management ,27,3,pp 699-722**

{۵۰}.فیروری، م، شکوری ،م، کاظمی، ل، زاهدی ،س، (۱۳۹۰)، "شناسایی تقامب در بیمه اتومبیل با استفاده از روش های داده کاوی" **پژوهشنامه صنعت بیمه** ،۲۶، ۱۰۳-۱۲۸

[51].J, Solomori,(**2001**), "Sepport Wector Machines For Phone Ile Classification," Master Of Science, Schools Of Artificial Intellengce Division Of Informatics University Of Edinburg, pp. **34,.**

[52].Schicollkopf, B.and Burges, C.and SIIl cola.A.,(**1998**) "Introduction. la support vector learning." In B. Schiolkopf C. Burges, and A. Smola, editors Advances in Kernel Methods- Support Wector Learning" MIT Press.

[53].P. Watanachaturaporn, M. Arora .K., Varshney. P. K.(**2004**) "Evaluation Of Factors Affecting Support Vector Machines For Hyperspectral Classification." in Proc. American Soc. Photogrammetry & Remote Sensing (ASPRS) Annual Conf.

[54].Lee M. C. To Chang,(**2010**) "Comparison Of Support Vector Machine And Back Propagation Neural Network In EvaluatingThe Enterprise Financial Distress." International **J. of Artificial Intelligence & Applications (IJAIA)**, Vol.1, No.3.

[55].Breiman,L.(**1996**)"**Bagging predictors.Machine Learning**"Vol.**24,2,pp 40-123**

[56].Breiman, L**.(2001**)" Random Forests. Machine Learning," Vol**. 45,1, pp. 5-32,**

[57].Peters. J. and Baets, B.D.and Verhoest. N. E. C.and Samson, R. and Degroeve. S. and Becker. P. D.(**2007**)" Random Forests as a tool for ecohydrological distribution modelling"**J. of Ecol Model, Vol. 207(2–4), pp. 304-18.**

Summary

Insurance companies offer different products, each of which can cover part of the risks, but unfortunately sometimes forgotten philosophy insurance and people with different tricks to "trying to cheat" and misappropriation of insurance "third party" are. "Fraud" of important issues losses to insurance companies and the insured is in all insurance fields. One way of identifying fraud "losses declared" the use of information fraud is discovered in the past. The data mining methods are widely used in discovering patterns in data. Using these methods can detect "damages fraud" in the "insurance industry" is useful. . In an experimental study these methods on real data, including information on the 1000 "case of damage" insurance third-party testing and performance of each method was used. The results showed that the accuracy of "algorithm" in education, 99.79% and 85.07%, which is in the testing of two basic algorithm is higher.

Key words: damage insurance, guided learning, data mining, fraud

Faculty of Industrial Engineering and Management

M.Sc. Thesis in Industrial Management

Assessment and recognition the trueness of the assurance claims using data mining techniques based on the supervised learning

By: Negar Sadeghian

Supervisor:

Dr. Mohammad Hassan Hosseini

Advisor:

Dr.Aliakbar Hasani

November 2016