Iran J. Insur. Res., 7(1): 13-26, Winter 2018, Serial #23

Iranian Journal of Insurance Research
(IJIR)
Homepage: https://ijir.irc.ac.ir/?lang=en

Insurance Research Center (IRC)

ORIGINAL RESEARCH PAPER

Investigation and analysis of fraudulent factors in the third-party civil liability car Insurance (Third-party insurance-physical damage)

Z. Seydi Agilabadi[1,*], S. Sehat[2], R. Salehi[3]

[1] Department of Business Management, Department of Insurance Management, Islamic Azad University, Science and Research Unit, Tehran, Iran
[2] Department of Business Management, School of Management, Allameh Tabatabai University, Tehran, Iran
[3] Department of Automobile Insurance, Sina Insurance Company, Tehran, Iran

ARTICLE INFO

Article History
Received: 13 January 2017
Revised: 13 February 2017
Accepted: 03 January 2018

Keywords
Fraudulent claims; Data Mining; Classification; Decision Tree; Support Vector Machine; Neural Network; Fraud.

ABSTRACT

The insurance industry, by its very nature, is susceptible to fraud. In car insurance, the insurer covers all damages caused to third parties by the car or car load. In recent years, due to the growth of this type of insurance, it has become necessary to identify the influencing factors on the decisions that deal with the falsity of a damage claim. One of the ways to detect and deal with this type of fraud is to check the information in the files that have claimed damages through a third party insurance policy. Data mining is a suitable method to interact with such databases and leads to the discovery of valuable knowledge from them; In this research, by examining 142 third party cases and 6 variables, it has been tried to discover fraud patterns in third party insurance. The research results show that the decision tree algorithm and neural networks have performed better than the support vector machine algorithm in identifying fraudulent, non-fraudulent, and suspicious cases.

***Corresponding Author:**
Email: *z.seidi85@gmail.com*
DOI: 10.22056/ijir.2018.01.02

نشریه علمی پژوهشنامه بیمه

سایت نشریه: https://ijir.irc.ac.ir/?lang=fa

مقاله علمی

بررسی و تحلیل عوامل کلاهبردارانه در بیمهٔ مسئولیت مدنی دارندگان وسایل نقلیهٔ زمینی در قبال اشخاص ثالث (بیمهٔ شخص ثالث -خسارات بدنی)

زهرا صیدی عقیل آبادی[۱،*]، سعید صحت[۲]، رسول صالحی[۳]

[۱]گروه مدیریت بازرگانی، گرایش مدیریت بیمه، دانشگاه آزاد اسلامی، واحد علوم و تحقیقات، تهران، ایران

[۲]گروه آموزشی مدیریت بازرگانی، دانشکدهٔ مدیریت، دانشگاه علامهٔ طباطبائی، تهران، ایران

[۳]گروه بیمه‌های اتومبیل، شرکت بیمهٔ سینا، تهران، ایران

اطلاعات مقاله

تاریخ دریافت: ۲۴ دی ۱۳۹۵

تاریخ داوری: ۲۵ بهمن ۱۳۹۵

تاریخ پذیرش: ۱۳ دی ۱۳۹۶

کلمات کلیدی

ادعای کلاهبردارانه

داده‌کاوی

رده‌بندی

درخت تصمیم

ماشین بردار پشتیبان

شبکهٔعصبی، تقلب

چکیده:

صنعت بیمه با توجه به ماهیت خود، مستعد ابتلا به کلاهبرداری و تقلب است. در بیمهٔ اتومبیل، بیمه‌گر کلیهٔ خسارت‌هایی را که به‌واسطهٔ خودرو یا بار خودرو به اشخاص ثالث وارد می‌شود، تحت پوشش قرار می دهد. در طی سال‌های اخیر با توجه به رشد این نوع بیمه، تشخیص عوامل تأثیرگذار بر روی تصمیم‌هایی که به جعلی‌بودن یک ادعای خسارت می‌پردازد به امری ضروری مبدل شده است. یکی از راه‌های کشف و مقابلهٔ با این نوع تقلب‌ها بررسی اطلاعات موجود در پرونده‌هایی است که از طریق بیمه‌نامهٔ شخص ثالث ادعای خسارت کرده‌اند. داده‌کاوی روش مناسبی برای تعامل با چنین بانک‌های اطلاعاتی است و منجر به کشف دانشی ارزشمند از آن‌ها می‌شود؛ در این تحقیق با بررسی ۱۴۲ پروندهٔ شخص ثالث و ۶ متغیر، سعی شده است تا الگوهای تقلب در بیمهٔ شخص ثالث کشف شود. نتایج تحقیق نشان می‌دهد الگوریتم درخت تصمیم و شبکه‌های عصبی در شناسایی پرونده‌های تقلبی، غیرتقلبی، و مشکوک نسبت به الگوریتم ماشین بردار پشتیبان عملکرد بهتری داشته‌اند.

***نویسنده مسئول:**

ایمیل: *z.seidi85@gmail.com*

DOI: 10.22056/ijir.2018.01.02

مقدمه

کلاهبرداری جرمی است که در آن از اعتماد افراد سوءاستفاده می‌شود. کلاهبرداری بیمه‌ای، عملی است که با هدف کلاهبرداری از بیمه‌گر برای کسب منافع مالی انجام می‌گیرد. در صنعت بیمه نیز کلاهبرداریهای متعددی دیده می‌شود، به این صورت که موضوعات بیمه‌شده، خود تبدیل به ابزاری برای کلاهبرداران شده است و این امر هرساله خسارتهای بسیاری بر شرکتهای بیمه تحمیل می‌کند. کلاهبرداری و تقلب در صنعت بیمه ممکن است در مراحل مختلف و توسط اشخاص مختلفی رخ دهد: بیمه‌گذاران جدید، بیمه‌گذاران فعلی، اشخاص ثالث زیان‌دیده و یا متخصصانی که به بیمه‌گذاران خدمات ارائه می‌دهند (Terisa, 2010 به نقل از فیروزی و همکاران، ۱۳۹۰).

برای حفاظت از منافع مشتریان و حفظ کیفیت خدمات در صنعت بیمه لازم است حتی پزشکان و پزشکان قانونی و همچنین دستگاه‌های قانونی و اجرایی نیز در این زمینه مشارکت کنند تا بتوانند نحوهٔ کلاهبرداری مشتریان (بیمه‌گذاران) از بیمه را در شرایط خاص پیشگویی کنند. به طور خلاصه، بروز تقلب از سوی مشتری منجر به ایجاد اثرات منفی بر حق‌بیمه و ادعای خدمات با توجه به ملاحظهٔ فرایندهای تنظیم‌کنندهٔ بیمه در شرکتهای بیمه و در میان عموم مردم جامعه می‌شود.

پژوهش حاضر به دنبال یافتن پاسخی برای سؤالهای پیش رو به شرح زیر است:

چگونه می‌توان تقلبهای شناسایی‌شده در خسارات بدنی بیمه‌های اتومبیل را با کمک الگوریتمهای داده‌کاوی رده‌بندی کرد؟

مهم‌ترین متغیرهای متقلبانه در ادعاهای خسارت بدنی، کدام خصیصه است؟

بهترین الگوریتم داده‌کاوی در شناسایی عوامل کلاهبردارانه کدام است؟

مبانی نظری پژوهش

تقلب (کلاهبرداری)

تقلب بیمه‌ای زمانی رخ می‌دهد که مردم، نمایندگی یا شرکت بیمه را فریب می‌دهند تا پولی را به دست آورند که درواقع حقشان نیست. این عمل، عملی مجرمانه و غیراخلاقی است که برای به‌دست‌آوردن منفعت باید عمداً مطالب نادرستی را ارائه داد.

گیل[۱] و همکاران (۱۹۹۴)، کلاهبرداری بیمه‌ای را این‌گونه تعریف کرده‌اند: اعلام عمدی خسارتهای جعلی، اعلام خسارت بیش از مقدار واقعی آن، یا هر روش دیگر برای به‌دست‌آوردن مبلغی بیش از آنچه که بیمه‌گذار قانوناً مستحق دریافت آن باشد (فیروزی و همکاران، ۱۳۹۰). اغلب اوقات ارائهٔ دادخواهی از سوی شرکت بیمه تنها روشی است که پس از آن می‌توان ارزیابی مطمئنی از یک کلاهبرداری بیمه‌ای انجام داد. البته در عمل، دادخواستهای مشکوک به‌ندرت توسط دادگاه تأیید و درنهایت به توافق منجر می‌شوند و در عوض از طریق مذاکرات خصوصی بین شرکت بیمه و بیمه‌شده حل‌وفصل می‌شوند (Brockett et al., 2002). همچنین به این دلیل، هیچ‌گونه آمار دقیقی در زمینهٔ اینکه چه مقدار کلاهبرداری در بازار کنونی بیمه اتفاق می‌افتد، وجود ندارد. بسیاری از آمارهای موجود ناشی از حدسهای منطقی هستند و کمتر مرتبط با رویه‌های ارزیابی رسمی و دقیق هستند (Ai et al., 2013).

از لحاظ بین‌المللی، متخصصان و کارشناسان حوزهٔ بیمه به مشکلاتی در خصوص ارزیابی میزان و هزینهٔ کلاهبرداریهای بیمه‌ای اشاره می‌کنند (NICB, 2000؛ اتحاد علیه کلاهبرداریهای بیمه‌ای (CAIF[۲])، ۲۰۰۳). حداقل سه مانع در این زمینه وجود دارد. اول، ماهیت پنهان کلاهبرداری (Dionne, 2000)، دوم حساسیت نسبت به تغییر (CAIF, 2003) و سرانجام فقدان توافقی در مورد اینکه چه موضوعی به طور دقیق کلاهبرداری بیمه‌ای محسوب می‌شود و بایستی بر کدام نوع از کلاهبرداری تمرکز کرد (Derrig, 2002).

فعالیتهای کلاهبردارانه به طور معمول در سه گروه رده‌بندی می‌شوند: ۱. اغراق در یک ادعای قانونی، ۲. ایجاد ساختگی یک شکایت یا ادعا، و ۳. افشای کلاهبردارانه یا عدم ارائهٔ اطلاعات صحیح در مورد مدارک. این کلاهبرداریها تنها در صورتی قابل شناسایی هستند که شرکتهای بیمه دارای ابزار بررسی مشخصی، مانند وجود گروه تحقیقاتی مجرب، برای تشخیص ماهیت و میزان کلاهبرداری باشند (Cross and Blackshaw, 2015).

---

۱. Gill

۲. Coalition Against Insurance Fraud

فرصت‌طلبی و برنامه‌ریزی و دعاوی خسارتی کلاهبردارانه

فرصت‌طلبی و برنامه‌ریزی در تقلب دو نوع از ادعاهای کلاهبردارانه است (Dionneand Gagne, 2002). فرصت‌طلبی در تقلب شامل تلاش مشتری در ادعاهای مندرج در بیمه‌نامه برای بروز یک رخداد است. در حالی که تقلب برنامه‌ریزی‌شده به معنای تلاش در کسب جبران خسارت با تحریف حادثه است.

کلاهبردارِ فرصت‌طلب معمولاً فردی است که متوجه فرصتی برای ارتکاب کلاهبرداری می‌شود. برای مثال، این فرد ممکن است تصور کند که بیمه‌گران وجوه نامحدودی دارند و ممکن است برای اینکه هزینه‌های حق‌بیمه‌ای پرداخت‌شده در سالهای پیشین را زمانی که هیچ خسارتی نداشته است بازیافت کند، جبران خسارات از طریق کلاهبرداری را قابل قبول و منطقی بیابد. در این بین کلاهبردار حرفه‌ای، درآمد خویش را از طریق ارتکاب به کلاهبرداری به دست می‌آورد یا کلاهبرداری، مکملی برای درآمد وی است. وی ممکن است تا زمانی که شناسایی شود به ارتکاب کلاهبرداری ادامه دهد و ممکن است تعدادی از بیمه‌گران را مورد هدف قرار دهد. حالت بسط‌یافتهٔ کلاهبرداری حرفه‌ای، تبهکاری سازمان‌یافته است که گروهی از افراد را در ارتکاب کلاهبرداریهای پیچیده و گسترده درگیر می‌کند. وجوهی که از طریق کلاهبرداری به دست می‌آید، ممکن است برای تأمین مالی سایر اقدامات مجرمانه نیز استفاده شوند.

عدم پوشش خسارت و ادعای کلاهبردارانه مشتری

مشتریان تنها زمانی که به طور واقعی با ازدست‌دادن اموال خود مواجه می‌شوند می‌توانند تحت پوشش بیمه، هزینهٔ خسارت خودشان را از بیمه بازپرداخت کنند. برخی از مشتریان ممکن است سعی در به‌دست‌آوردن پول از بیمه حتی در زمان بروز تلفاتِ غیر تحت پوشش بیمهٔ شخص ثالث (به‌عنوان مثال: سقوط از بلندی و جایگزینی آن در قالب تصادفات، عدم پوشش افراد مستقر در قسمت بار وسایل نقلیهٔ بارکش و ادعای دروغین در خصوص محل استقرار افراد در زمان حادثه و غیره) داشته باشند.

مروری بر پیشینه پژوهش

تانگ[۱] (۲۰۱۳)، سیستم نظارتی مبارزه با تقلب در بیمهٔ اتومبیل را با استفاده از روشهای داده‌کاوی طراحی کرد. بینگ[۲] (۲۰۱۳) به تحلیل و مقایسهٔ چندین مدل داده‌کاوی تقلب بیمه‌ای پرداخت و برای تفکیک و پردازش داده‌ها از مدل خطی تعمیم‌یافتهٔ پیشنهادی بهره برد و درنهایت به مدل داده‌کاوی مناسب‌تری دست یافت.

ایشیدا[۳] و همکاران (۲۰۱۶)، شدت اخلاق‌گرایی، آگاهی اخلاقی، و زمینه‌های اخلاقی را در بروز تقلبهای بیمه‌ای مورد بحث و بررسی قراردادند. آنها با علم به اینکه تمامی تقلبهای بیمه‌ای به شکلی یکسان ایجاد نمی‌شوند، در مدلی (۲*۲) به بررسی تغییرپذیری برداشتها مبنی بر شدت اخلاق‌گرایی در شرایط مختلف تقلب بیمه‌ای پرداختند و نتایج بین دو نمونهٔ مستقل (افراد جوان و مسن) را مورد مقایسه قرار دادند. نتایج مبین این است که جوانان قضاوتهای سهل‌انگارانه‌تری از تخطیهای اخلاقی مشارکتی دارند.

فیروزی و همکاران (۱۳۹۰)، تقلب در بیمه‌های اتومبیل را با استفاده از روش داده‌کاوی شناسایی کردند. در این بررسی ۶ متغیر مستقل شامل سابقهٔ بیمه‌ای هر یک از بیمه‌گذاران در شرکتهای بیمه، تعداد ادعای خسارت بیمه‌گذاران در طول دورهٔ سابقه بیمه، فاصلهٔ زمانی وقوع حادثه تا اعلام خسارت، وضعیت کروکی خسارت، جانی و مالی‌بودن خسارت، و بالابودن مبلغ خسارت را در شناسایی تقلب در بیمهٔ شخص ثالث و بدنه مورد ارزیابی قرار دادند. آنها از سه روش درخت تصمیم، رگرسیون لجستیک و بیز ساده در شناسایی خسارت تقلبی استفاده کردند که درنهایت، روش بیز ساده برتری بیشتری نسبت به بقیهٔ روشها داشته است.

تقوی‌فرد و جعفری (۱۳۹۴)، با بهره‌مندی از یک سیستم خبرهٔ فازی به کشف تقلب در بیمهٔ بدنهٔ خودرو پرداختند. به نظر آنها به دلیل ماهیت پنهان پدیدهٔ تقلب، دانش و قضاوت خبرگان مبتنی بر شواهد و اطلاعات کیفی است که اغلب واژه‌های کلامی را برای توصیف رفتار متقلبانه به کار می‌برند. در مدل ارائه‌شده، از میان ۶۱ معیار کمی و کیفی شناسایی‌شده کشف تقلب در بیمهٔ بدنهٔ خودرو، بر اساس نظر خبرگان ۱۷ معیار که از اولویت بالایی برخوردار بودند، در قالب ۸ عامل دسته‌بندی شدند. درنهایت، پس از طراحی و پیاده‌سازی سیستم در

---

۱. Tang

۲. Bing

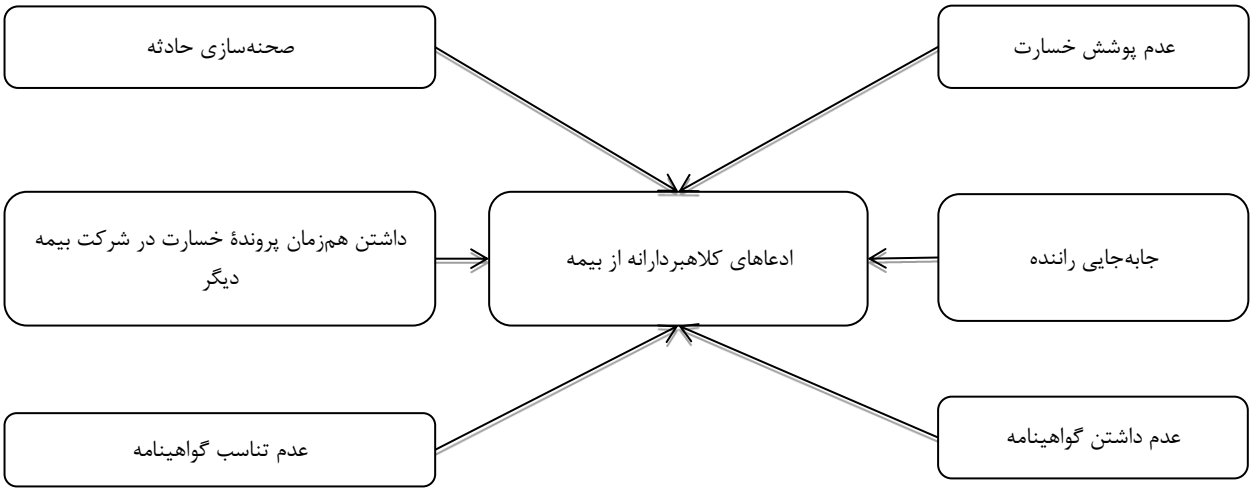
۳. Ishida

یکی از شرکتهای بیمهٔ خصوصی ایران، نتیجهٔ آن مؤید این نکته است که اعتبار کلی سیستم پیشنهادی به میزان شایان توجهی از قابلیت شناسایی تقلب برخوردار است.

مطالعات مربوط به تقلبهای بیمه‌ای، مفاهیم گسترده‌تر این پدیده را مورد بررسی قرار داده‌اند و علی‌رغم اینکه مشارکت کلاهبرداران بیمه‌ای در این خصوص امری به اثبات رسیده است اما چگونگی ورود این افراد به تقلبهای بیمه‌ای هنوز نامشخص است زیرا تحقیقها آن‌گونه که باید به این موضوع نپرداخته‌اند. آنچه گفته شد الزاماً به‌منزلهٔ نقدی بر مطالعات صورت گرفته نیست، بلکه مبین دشواریهای موجود در مفهوم‌سازی تقلب‌های بیمه‌ای است.

مدل مفهومی تحقیق

با توجه به مباحث نظری ارائه‌شده، می‌توان مدل مفهومی در شکل ۱ را برای ارتباط متغیرهای مستقل با متغیر وابسته (کلاهبرداری بیمه‌ای) ترسیم کرد.



شکل ۱: مدل مفهومی تحقیق بر اساس چارچوب نظری پژوهش

روش‌شناسی پژوهش

روش پژوهش از نظر هدف، کاربردی است و از دید ماهیت و روش، در دستهٔ پژوهشهای توصیفی قرار دارد. به منظور تحلیل، مطالعهٔ موردی این پژوهش به بررسی ۱۴۲ پروندهٔ یک شرکت بیمهٔ خصوصی در کل کشور اختصاص دارد. شایان ذکر است در این پرونده‌ها تقلب در بیمهٔ شخص ثالث در بازهٔ زمانی ۲ سالهٔ ۱۳۹۳-۱۳۹۵ بررسی شده است. بر این اساس و با توجه به معیارهای ذکرشده و همچنین مدل مفهومی ارائه‌شده و بررسی اطلاعات ۱۴۲ پروندهٔ مشکوک به کلاهبرداری، در این پژوهش ۶ متغیر مستقل به شرح ذیل شناسایی شد:

فقدان گواهینامه؛

عدم تناسب گواهینامه با وسیلهٔ نقلیهٔ مورد بیمه؛

وجود خسارت مشابه در سایر شرکتهای بیمه؛

جابه‌جایی راننده؛

عدم پوشش قانون شخص ثالث؛

صحنه‌سازی حادثه به‌قصد سوءاستفاده از شرکتهای بیمه‌ای.

همچنین متغیر وابسته در پژوهش حاضر ادعای کلاهبردارانه تلقی شد. برای رده‌بندی متغیرها از پرسشنامه‌ای به شرح جدول ۱ استفاده و برای هر متغیر مقادیری تخصیص داده شد. متغیرهای زبانی و مقادیرآنها مطابق جدول ۱ است.

جدول ۱: معرفی متغیرها و مقادیر داده‌شده

متغیرهای زبانی	مقادیر زبانی				نوع متغیر
مقصر حادثه فاقد گواهینامه است.	موافقم ۱			مخالفم ۰	مستقل
گواهینامه مقصر حادثه با وسیلهٔ نقلیه تناسب ندارد.	موافقم ۱			مخالفم ۰	مستقل
خسارت مشابه در سایر شرکت‌های بیمه‌ای داشته است.	موافقم ۱	مشخص نشد ۲		مخالفم ۰	مستقل
در پروندهٔ موردبررسی جابه‌جایی راننده اتفاق افتاده است.	موافقم ۱			مخالفم ۰	مستقل
خسارت اعلامی خارج از شمول قانون شخص ثالث بوده است.	موافقم ۱			مخالفم ۰	مستقل
خسارت اعلامی صحنه‌سازی و به قصد سوءاستفاده از شرکت بیمه بوده است.	موافقم ۱			مخالفم ۰	مستقل
آیا خسارت تقلبی بوده است یا خیر و یا مشکوک است؟	جعلی ۳	مشکوک ۲		غیرجعلی ۱	وابسته
کلاهبرداری به‌صورت فردی صورت گرفته است یا گروهی؟	غیرقابل‌تشخیص ۱	انفرادی ۲		گروهی ۳	مستقل

در این مقاله، سه الگوریتم داده‌کاوی یعنی درخت تصمیم‌گیری، ماشین بردار پشتیبان (SVMs)[۱]، و شبکه‌های عصبی مصنوعی برای پیشگویی ادعای خسارت متقلبانه تحلیل شد و بهترین راهکار در پیشگویی این امر شناسایی شد. در این پژوهش از نرم‌افزار کلمنتاین نسخهٔ ۱۲[۲] برای داده‌کاوی داده‌های موجود و متغیرها استفاده شد.

تحلیل داده‌ها

در این بخش، داده‌ها را مدل‌سازی کرده و تحلیلهای لازم در خصوص اعتبارسنجی مدل و کارایی آن ارائه می‌شود. نتایج نشان داد که ۴۵/۵۸ درصد از پرونده‌های کلاهبرداری به صورت گروهی و ۲۷/۴۶ درصد به صورت انفرادی صورت گرفته است. در ۲۵/۳۵ درصد از پرونده‌های مورد بررسی مقصرین حادثه فاقد گواهینامه بوده‌اند و در ۱۱/۹۷ درصد از پرونده‌ها، گواهینامهٔ مقصر با وسیلهٔ نقلیهٔ مورد بیمه تناسب نداشته است. ۱۲/۶۸ درصد از پرونده‌های بررسی‌شده در سایر شرکتهای بیمه‌ای نیز دارای پروندهٔ خسارتی مشابه بوده‌اند. در ۳۳/۱ درصد، جابه‌جایی راننده صورت گرفته است و ۲۶/۰۶ درصد تحت پوشش بیمهٔ شخص ثالث نبوده‌اند، ۶۶/۲ درصد از پرونده‌های مورد بررسی صحنه‌سازی به قصد سوءاستفاده از شرکت بیمه محرز بوده است و درنهایت از بین ۱۴۲ پروندهٔ مشکوک به کلاهبرداری، جعلی‌بودن ۵۱/۴۱ درصد پرونده اثبات شده است.

الگوریتم درخت تصمیم C5.0

درخت تصمیم روشی معروف برای رده‌بندی است که نتایج آن در یک فلوچارت شبیه ساختار درخت ارائه شده است، که هر گره[۳] نشانگر یک آزمون بر روی ارزش مشخصه و هر شاخه، خروجی هر آزمون را نمایش می‌دهد؛ برگهای درخت نیز نمایانگر رده‌ها هستند (Tan and Steinbach, 2006). در ساخت درختهای تصمیم، به طور معمول داده‌ها را به دو دسته تقسیم می‌کنند: ۱) داده‌های آموزشی[۴] که برای ساخت مدل مورد استفاده قرار می‌گیرند، و ۲) داده‌های آزمون[۵] که برای آزمون و ارزیابی مدل ساخته‌شده کاربرد دارند. این الگوریتم برای ساخت درخت تصمیم یا مجموعهٔ قوانین استفاده می‌شود و این فرایند تا زمانی که نتواند زیرنمونه را شاخه بزند ادامه می‌یابد. در پایان، پایین‌ترین سطح شاخه‌ها آزموده می‌شوند و شاخه‌هایی که نقش مهمی در مدل ندارند حذف یا هرس می‌شوند. میزان دقت الگوریتم، با توجه به میزان تکرارهای آزمایش (دقت اعمال الگوریتم بر روی داده‌ها) به صورت جدول ۲ است.

---

[1]. Support Vector Machines

[۲]. SPSS Clementine 12

[۳]. Node

[۴]. Train Data

[۵].Test Data

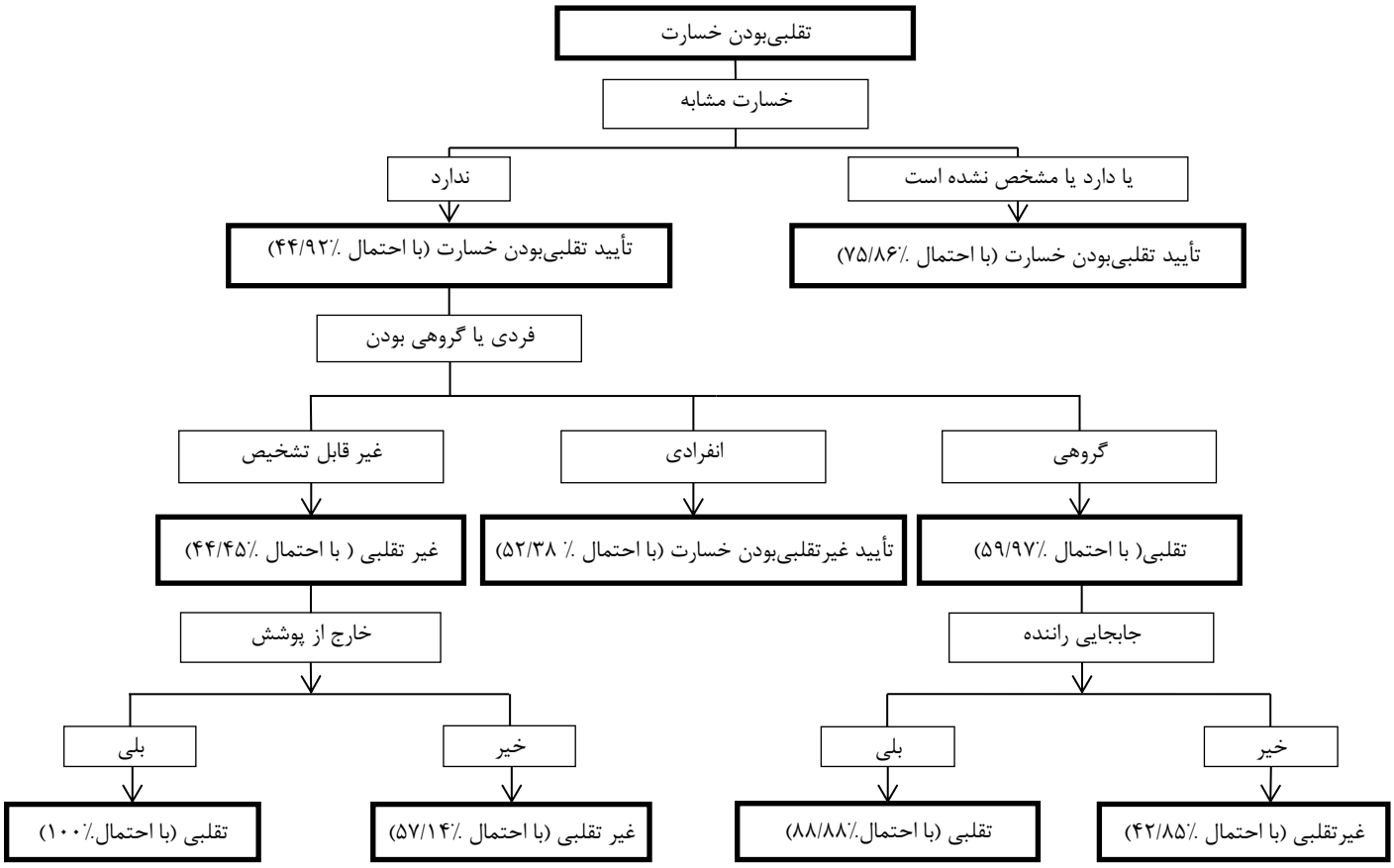
جدول ۲: دقت الگوریتم درخت تصمیم C5.0

دسته‌بندی	آموزشی	آزمایشی
نمونه‌های رده‌بندی‌شدهٔ صحیح (دقت مدل)	٪۶۵/۳۱	٪۶۳/۶۴
نمونه‌های رده‌بندی‌شدهٔ نادرست (خطای مدل)	٪۳۴/۶۹	٪۳۶/۳۶

برای ساخت درخت از ۷ مشخصه که به‌عنوان مشخصهٔ ورودی تعریف شده بود، استفاده شد. همچنین، پارامترهای لازم در الگوریتم درخت تصمیم با توجه به مطالبی که در جدول ۱ بیان شد، تنظیم شد تا بتوان درخت موردنظر را ایجاد کرد. درخت ابتدا با استفاده از ۷۰ درصد داده‌ها ایجاد شد و سپس مدل ایجادشده با ۳۰ درصد باقی‌ماندهٔ داده‌ها مورد ارزیابی قرار گرفت.

با توجه به درخت ایجادشده در شکل ۲، متغیر خسارت مشابه در سایر شرکتهای بیمه به‌عنوان اولین و مهم‌ترین عامل در بررسی پرونده‌های مشکوک به کلاهبرداری در نظر گرفته می‌شود. این متغیر در ریشهٔ درخت به دو شاخهٔ وجود یا عدم وجود تقسیم می‌شود. اگر پروندهٔ مشکوک به خسارت تقلبی، در سایر شرکتهای بیمه نیز خسارت مشابه داشته باشد، با اطمینان ۷۵/۸۶ درصد می‌تواند خسارت اعلامی جعلی باشد. اگر پروندهٔ مشکوک به خسارت تقلبی، در سایر شرکتهای بیمه خسارت مشابه نداشته باشد و به صورت انفرادی صورت گرفته باشد، با اطمینان ۵۲/۳۸ درصد پرونده غیرجعلی خواهد بود.

اگر پروندهٔ مشکوک به خسارت تقلبی، در سایر شرکتهای بیمه خسارت مشابه نداشته باشد و امکان تشخیص گروهی و یا انفرادی‌بودن آن فراهم نباشد و در عین حال اصل حادثه از شمول بیمهٔ شخص ثالث خارج باشد، با اطمینان ۱۰۰ درصد می‌توان پرونده را جعلی فرض کرد. سایر قوانین هم به همین ترتیب قابل استخراج است.



شکل ۲: خروجی گرافیکی درخت تصمیم بر اساس الگوریتم C5

در ادامه با اعمال این مدل بر روی داده‌های اولیه، نتایج مندرج در جدول ۳ برای بررسی دقت مدل به‌دست‌آمده است.

جدول ۳: دقت مدل در شناسایی وضعیت پرونده‌های خسارتی با استفاده از درخت تصمیم‌گیری

وضعیت پرونده	میزان حساسیت (آموزش)	میزان حساسیت (آزمایش)
۱(غیر جعلی)	۰/۸۰ ٪	۰/۸۰٪
۲(مشکوک)	۰٪	۰٪
۳(تقلبی)	۷۵/۴۵٪	۸۰٪
میزان دقت	۶۵/۳۱٪	۶۳/۶۴٪
خطای رده‌بندی	۳۴/۶۹٪	۳۶/۳۶٪

لازم به توضیح است که در الگوریتم تصمیم‌گیری C5، مهم‌ترین عوامل کلاهبردارانه به شرح ۱. عدم پوشش بیمه شخص ثالث، ۲. وجود خسارت مشابه در سایر شرکتهای بیمه، ۳. جابه‌جایی راننده، و ۴. صحنه‌سازی به قصد سوءاستفاده از شرکتهای بیمه، رده‌بندی شد که این موضوع، آگاهی سازمانی ارزشمندی را در اختیار شرکتهای بیمه قرار می‌دهد تا شاخصهای با اهمیت بالا را مبنای تصمیم‌گیری قرار دهند.

الگوریتم ماشین بردار پشتیبان

این روش از جملهٔ روش‌های نسبتاً جدیدی است که در سالهای اخیر کارایی خوبی نسبت به روشهای قدیمی‌تر برای رده‌بندی نشان داده است که هم برای رده‌بندی داده‌های خطی و هم غیرخطی به کار می‌رود. بر اساس جدول ۴ می‌توان گفت الگوریتم استخراج‌شده قابلیت رده‌بندی رفتارهای متقلبانه در صنعت بیمه و در این پژوهش را داراست.

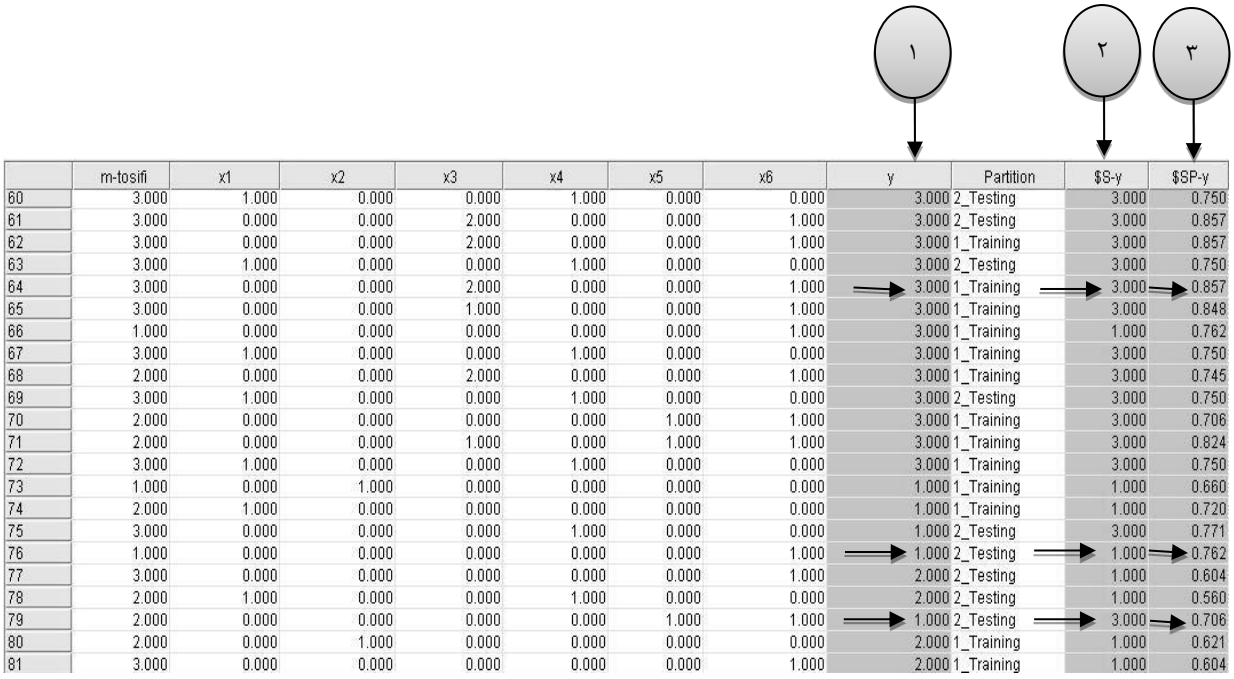
جدول ۴: دقت مدل در الگوریتم ماشین بردار پشتیبان

دسته‌بندی	آموزشی		آزمایشی	
نمونه‌های رده‌بندی‌شدهٔ صحیح (دقت مدل)	۷۰	۷۱/۴۳٪	۲۷	۶۱/۳۶٪
نمونه‌های رده‌بندی‌شدهٔ نادرست (خطای مدل)	۲۸	۲۸/۵۷٪	۱۷	۳۸/۶۴٪
کل	۹۸		۴۴	

همان‌گونه که در شکل ۳ نیز قابل‌مشاهده است، متغیرهای پراهمیت در الگوریتم ماشین بردار پشتیبان، عدم پوشش شخص ثالث، خسارت مشابه، صحنه‌سازی به قصد گرفتن خسارت از شرکت بیمه، عدم تناسب گواهینامه با وسیلهٔ نقلیه، فاقد گواهینامه‌بودن راننده و درنهایت جابه‌جایی راننده است.

شکل ۳: اهمیت شاخص‌ها در الگوریتم ماشین بردار پشتیبان

صحت کلی پیشگویی موارد تقلب در الگوریتم ماشین بردار پشتیبان در شکل ۴ برای ۳۰ رکورد به‌عنوان نمونه به نمایش درآمده است.



	m-tosifi	x1	x2	x3	x4	x5	x6	y	Partition	$S-y	$SP-y
60	3.000	1.000	0.000	0.000	1.000	0.000	0.000	3.000	2_Testing	3.000	0.750
61	3.000	0.000	0.000	2.000	0.000	0.000	1.000	3.000	2_Testing	3.000	0.857
62	3.000	0.000	0.000	2.000	0.000	0.000	1.000	3.000	1_Training	3.000	0.857
63	3.000	1.000	0.000	0.000	1.000	0.000	0.000	3.000	2_Testing	3.000	0.750
64	3.000	0.000	0.000	2.000	0.000	0.000	1.000	3.000	1_Training	3.000	0.857
65	3.000	0.000	0.000	1.000	0.000	0.000	1.000	3.000	1_Training	3.000	0.848
66	1.000	0.000	0.000	0.000	0.000	0.000	1.000	3.000	1_Training	1.000	0.762
67	3.000	1.000	0.000	0.000	1.000	0.000	0.000	3.000	1_Training	3.000	0.750
68	2.000	0.000	0.000	2.000	0.000	0.000	1.000	3.000	1_Training	3.000	0.745
69	3.000	1.000	0.000	0.000	1.000	0.000	0.000	3.000	2_Testing	3.000	0.750
70	2.000	0.000	0.000	0.000	0.000	1.000	1.000	3.000	1_Training	3.000	0.706
71	2.000	0.000	0.000	1.000	0.000	1.000	1.000	3.000	1_Training	3.000	0.824
72	3.000	1.000	0.000	0.000	1.000	0.000	0.000	3.000	1_Training	3.000	0.750
73	1.000	0.000	1.000	0.000	0.000	0.000	0.000	1.000	1_Training	1.000	0.660
74	2.000	1.000	0.000	0.000	0.000	0.000	0.000	1.000	1_Training	1.000	0.720
75	3.000	0.000	0.000	0.000	1.000	0.000	0.000	1.000	2_Testing	3.000	0.771
76	1.000	0.000	0.000	0.000	0.000	0.000	1.000	1.000	2_Testing	1.000	0.762
77	3.000	0.000	0.000	0.000	0.000	0.000	1.000	2.000	2_Testing	1.000	0.604
78	2.000	1.000	0.000	0.000	1.000	0.000	0.000	2.000	2_Testing	1.000	0.560
79	2.000	0.000	0.000	0.000	0.000	1.000	1.000	1.000	2_Testing	3.000	0.706
80	2.000	0.000	1.000	0.000	0.000	0.000	0.000	2.000	1_Training	1.000	0.621
81	3.000	0.000	0.000	0.000	0.000	0.000	1.000	2.000	1_Training	1.000	0.604

شکل ۴: صحت کلی پیشگویی ۳۰ رکورد در الگوریتم ماشین بردار پشتیبان

در شکل ۴، شمارهٔ ۱ در بالای جدول، مقادیر واقعی رده‌های تقلب (جعلی، غیرجعلی، و مشکوک‌بودن) را نشان می‌دهد، شمارهٔ ۲ مقادیر پیشگویی‌شده برای رده توسط نرم‌افزار را نشان می‌دهد و شمارهٔ ۳ نمرهٔ تمایل به پیشگویی و احتمال این را که پیشگویی برای این رکورد خاص درست باشد، نشان می‌دهد. به‌عنوان مثال در رکورد شمارهٔ ۶۴، مقدار واقعی رده برابر با ۳ است (خسارت متقلبانه است) و نرم‌افزار هم مقدار ۳ را با اطمینان ۸۵/۷ درصد تشخیص داده است؛ بنابراین ادعای تقلبی‌بودن تأیید شده است.

در ادامه با اعمال این مدل بر روی داده‌های اولیه، نتایج در جدول ۵ برای بررسی دقت مدل به‌دست‌آمده است. همان‌طور که در جدول ۵ قابل مشاهده است میزان حساسیت مدل در شناسایی پرونده‌های خسارت جعلی در بخش آموزش ۸۳/۳۳ درصد بوده است. ضمن اینکه دقت مدل در شناسایی صحیح جعلی یا غیرجعلی‌بودن هر پرونده در این بخش برابر ۷۱/۴۳ درصد و خطای رده‌بندی ۲۸/۵۷ درصد بوده است.

جدول ۵: دقت مدل در شناسایی وضعیت پرونده‌های خسارتی با استفاده از الگوریتم ماشین بردار پشتیبان

وضعیت پرونده	میزان حساسیت (آموزشی)	میزان حساسیت (آزمایشی)
۱(غیر جعلی)	۸۳/۳۳٪	۶۶/۶۷٪
۲(مشکوک)	۰٪	۰٪
۳(تقلبی)	۸۴/۹۱٪	۸۵٪
میزان دقت	۷۱/۴۳٪	۶۱/۳۶٪
خطای رده‌بندی	۲۸/۵۷٪	۳۸/۶۴٪

الگوریتم شبکه‌های عصبی

به طور معمول سه قسمت در یک شبکهٔ عصبی وجود دارد: یک لایهٔ ورودی، یک یا چند لایهٔ پنهان و یک لایهٔ خروجی. این شبکه‌ها قادر به مدل‌سازی توابع غیرخطی هستند.

برای ایجاد مدل شناسایی تقلب در خسارت بدنی شخص ثالث بیمهٔ اتومبیل، یک متغیر افراز[۱] ایجاد شد تا بتوان داده‌ها را به دو بخش آموزشی و اعتبارسنجی تقسیم‌بندی کرد. در این پژوهش ۷۰ درصد (۹۸ مورد) از داده‌ها در بخش آموزشی و ۳۰ درصد (۴۴ مورد) در بخش آزمایش برای اعتبارسنجی مدل به صورت تصادفی تخصیص داده شد. برای تخصیص تصادفی داده‌ها به مجموعه‌های یادشده، از توزیع برنولی با احتمال ۶۰ درصد برای متغیر افراز استفاده شده است.

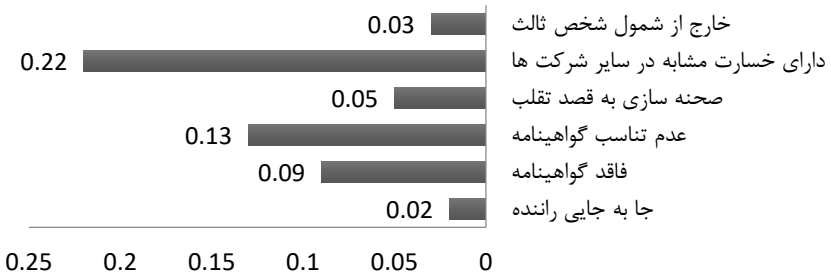
در این پژوهش با توجه به اینکه تعداد داده‌های متقلبانه (متغیر وابسته با مقدار ۳) محدود است، تلاش بر این بوده است که هیچ‌یک از داده‌های متقلبانه از فرایند آموزش حذف نشود. بر اساس جدول ۶ دقت مدل در بخش آزمایشی ٪۶۳/۶۴ و در بخش آموزشی ٪۶۴/۲۹ و همچنین خطای مدل به‌ترتیب ٪۳۶/۳۶ و ٪۳۵/۷۱ بوده است.

جدول ۶: دقت مدل در الگوریتم شبکه‌های عصبی

افراز	آموزشی		آزمایشی	
نمونه‌های رده‌بندی‌شدهٔ صحیح (دقت مدل)	۶۳	٪۶۴/۲۹	۲۸	٪۶۳/۶۴
نمونه‌های رده‌بندی‌شدهٔ نادرست (خطای مدل)	۳۵	٪۳۵/۷۱	۱۶	٪۳۶/۳۶
کل	۹۸		۴۴	

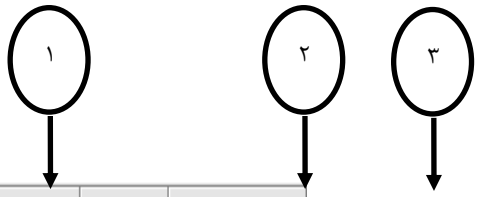
در شکل ۵ مشاهده می‌شود که متغیرهای پراهمیت در الگوریتم شبکه‌های عصبی، خسارت مشابه در سایر شرکتهای بیمه، عدم تناسب گواهینامه با وسیلهٔ نقلیه، فاقد گواهینامه‌بودن راننده، صحنه‌سازی به قصد گرفتن خسارت از شرکت بیمه، عدم پوشش شخص ثالث و درنهایت جابه‌جایی راننده هستند.

۱. Partition Variable



شکل ۵: اهمیت شاخص‌ها در الگوریتم شبکه‌های عصبی

صحت کلی پیشگویی موارد تقلب در الگوریتم شبکهٔ عصبی در شکل ۶ برای ۳۰ رکورد به‌عنوان نمونه به نمایش درآمده است. بر این اساس، شمارهٔ ۱ مقادیر واقعی رده‌های تقلب را نشان می‌دهد. شمارهٔ ۲ مقادیر پیشگویی‌شده برای رده‌ها را نشان می‌دهد و شمارهٔ ۳، نمرهٔ تمایل به پیشگویی و احتمال این را که پیشگویی برای این رکورد خاص درست باشد، نشان داده که عددی بین ۰ تا ۱ است. به‌عنوان مثال، در رکورد ۲ مقدار واقعی کلاس ۳ است و نرم‌افزار هم مقدار ۳ را با اطمینان ۹۵ درصد برای این رکورد پیشگویی کرده است.



	m-tosifi	x1	x2	x3	x4	x5	x6	y	Partition	$N-y	$NC-y
1	3.000	0.000	0.000	1.000	0.000	1.000	1.000	3.000	1_Training	3.000	0.959
2	3.000	0.000	0.000	1.000	0.000	1.000	1.000	3.000	1_Training	3.000	0.959
3	2.000	0.000	0.000	1.000	0.000	1.000	1.000	3.000	1_Training	3.000	0.433
4	2.000	0.000	0.000	1.000	1.000	1.000	1.000	3.000	2_Testing	3.000	0.505
5	2.000	0.000	0.000	0.000	0.000	0.000	1.000	3.000	1_Training	1.000	0.498
6	2.000	0.000	0.000	2.000	0.000	1.000	1.000	3.000	1_Training	3.000	0.249
7	2.000	0.000	0.000	2.000	0.000	1.000	1.000	3.000	1_Training	3.000	0.249
8	2.000	0.000	1.000	2.000	1.000	0.000	0.000	3.000	1_Training	1.000	0.406
9	2.000	0.000	1.000	0.000	1.000	0.000	0.000	3.000	1_Training	1.000	0.512
10	3.000	0.000	0.000	0.000	0.000	0.000	1.000	3.000	1_Training	1.000	0.148
11	1.000	0.000	0.000	0.000	0.000	0.000	1.000	3.000	1_Training	1.000	0.503
12	3.000	1.000	0.000	0.000	1.000	0.000	0.000	3.000	1_Training	3.000	0.740
13	3.000	0.000	0.000	2.000	0.000	1.000	1.000	3.000	2_Testing	3.000	0.942
14	3.000	0.000	0.000	2.000	0.000	1.000	1.000	3.000	1_Training	3.000	0.942
15	3.000	1.000	0.000	0.000	1.000	0.000	0.000	3.000	2_Testing	3.000	0.740
16	3.000	0.000	0.000	1.000	0.000	0.000	1.000	3.000	2_Testing	3.000	0.946

شکل ۶: صحت کلی پیشگویی ۳۰ رکورد در الگوریتم شبکه‌های عصبی

دقت کلی مدل از لحاظ حساسیت در دو بخش آموزشی و آزمایشی در جدول ۷ نشان داده شده است.

جدول ۷ :دقت مدل در شناسایی وضعیت پرونده‌های خسارتی با استفاده از الگوریتم شبکه‌های عصبی

وضعیت پرونده	میزان حساسیت (آموزشی)	میزان حساسیت (آزمایشی)
۱(غیر جعلی)	۸۶/۶۷٪	۷۳/۳۳٪
۲(مشکوک)	۰٪	۰٪
۳(تقلبی)	۶۹/۸۱٪	۸۵٪
میزان دقت	۶۴/۲۹٪	۶۳/۶۴٪
خطای رده‌بندی	۳۵/۷۱٪	۳۶/۳۶٪

بر اساس جدول ۷، میزان حساسیت مدل در شناسایی پرونده‌های خسارت جعلی در بخش آموزشی ٪۶۹/۸۱ بوده است. ضمن اینکه دقت مدل در شناسایی صحیح جعلی یا غیرجعلی‌بودن هر پرونده در این بخش برابر ٪۶۴/۲۹ و خطای رده‌بندی ٪۳۵/۷۱ بوده است.

مقایسهٔ مدل‌ها و مدل پیشنهادی

در این مطالعه، سه روش داده‌کاوی مبتنی بر دقت الگوریتمها مقایسه شدند و هدف نهایی دستیابی به مدلی با بالاترین میزان دقت بود. از آنجا که این پـژوهش بـه دنبـال مـدل پیشنهاددهنده‌ای بـا دقـت زیـاد برای شناسایی عوامل کلاهبردانه در پرونده‌های خسارت بدنی بیمهٔ اتومبیل با رویکرد دسته‌بندی است، می‌توان مدل نهایی را بر اساس دقت روشهای دسته‌بندی مختلف بـه دست آورد. در دو جدول ۸ و ۹ خلاصه‌ای از مقایسهٔ بین این سه مدل ازنظر دقت و حساسیت آورده شده است.

جدول ۸: مقایسهٔ دقت الگوریتم‌ها در کشف خسارات تقلبی بیمهٔ شخص ثالث

نام الگوریتم	دقت مدل در بخش آموزشی	دقت مدل در بخش آزمایشی
درخت تصمیم C5.0	۶۵/۳۱٪	۶۳/۶۴٪
ماشین بردار پشتیبان	۷۱/۴۳٪	۶۱/۳۶٪
شبکه‌های عصبی	۶۴/۲۹٪	۶۳/۶۴٪

همان‌گونه که از نتایج برمی‌آید، الگوریتم درخت تصمیم و شبکه‌های عصبی در شناسایی پرونده‌های تقلبی، غیرتقلبی، و مشکوک نسبت به الگوریتم ماشین بردار پشتیبان با دقت اطمینان ٪۶۳/۶۴ عملکرد بهتری داشته‌اند.

جدول ۹: دقت کلاس ۳ (تقلبی) با استفاده از معیار حساسیت

نام الگوریتم	دقت مدل در بخش آموزشی	دقت مدل در بخش آزمایشی
درخت تصمیم C5.0	۷۵/۴۷٪	۸۰٪
ماشین بردار پشتیبان	۸۴/۹۱٪	۸۵٪
شبکه‌های عصبی	۶۹/۸۱٪	۸۵٪

همان‌گونه که از جدول ۹ مشخص است در بخش حساسیت، الگوریتم شبکه‌های عصبی و ماشین بردار پشتیبان با دقت ٪۸۵ برای پرونده‌های خسارت تقلبی (کلاهبرداری) بهترین عملکرد را داشته‌اند.

نتایج و بحث

جمع بندی و پیشنهادها

داده‌کاوی قادر به کشف و استخراج دانش جدید از داده‌های گذشته‌نگر است. در این مقاله سه روش داده‌کاوی الگوریتم درخت تصمیم C5.0، ماشین‌بردار پشتیبان SVM و الگوریتم شبکه‌های عصبی برای ساخت مدلهایی برای شناسایی ادعاهای خسارت تقلبی در خسارت بدنی بیمه‌های اتومبیل معرفی شدند. در ادامه این روشها بر روی ۱۴۲ دادهٔ واقعی جمع‌آوری‌شده از خسارات بدنی مشکوک و تعیین تکلیف شدهٔ یک

شرکت بیمهٔ خصوصی، آزمایش و کارایی هر روش سنجیده شد. درخت تصمیم‌گیری، ماشین بردار پشتیبان و شبکهٔ عصبی با وجود سادگی نتایج، دقت قابل قبولی در داده‌کاوی داده‌های جمع‌آوری‌شده در رابطهٔ با تقلب در بیمه ارائه دادند؛ دقت الگوریتم درخت تصمیم و شبکه‌های عصبی در شناسایی پرونده‌های جعلی، غیرجعلی و مشکوک نسبت به الگوریتم ماشین بردار پشتیبان عملکرد بهتری داشته‌اند و در بخش حساسیت، الگوریتم شبکه‌های عصبی و ماشین بردار پشتیبان برای پرونده‌های خسارت تقلبی (کلاهبرداری) بهترین عملکرد را داشته‌اند. با توجه به نتایج می‌توان الگوریتم شبکه‌های عصبی را به‌عنوان الگوریتم برتر انتخاب کرد. نتیجهٔ به‌دست‌آمده در این پژوهش از این جهت با نتیجهٔ به‌دست‌آمده از تحقیق فیروزی و همکاران (۱۳۹۰) همسوست که در تحقیق آنها الگوریتم درخت تصمیم نسبت به روش بیز ساده و رگرسیون لجستیک در شناسایی خسارات تقلبی دقت پایین‌تری داشته است. آنها به این نتیجه رسیدند که هرچه مبلغ خسارت بیشتر باشد، و نوع خسارت مالی باشد، مورد خسارت جعلی خواهد بود و اگر خسارت جانی باشد و در عین حال سابقهٔ بیمه‌ای فرد، کوچکتر و یا مساوی ۲/۵ سال باشد مورد خسارت جعلی خواهد بود.

با توجه به نتایج به‌دست‌آمده از این پژوهش و مقالهٔ فیروزی و همکاران (۱۳۹۰)، می‌توان متغیرهای دیگری نیز برای شناسایی و رده‌بندی تقلبهای خسارت بدنی اضافه کرد به‌گونه‌ای که نتایج به‌دست‌آمده جامع‌تر شود. بر این اساس پیشنهاد می‌شود متغیرهایی نظیر سن و جنسیت بیمه‌گذار یا طرفین دخیل در حادثه، زمان و مکان حادثه، سابقهٔ گواهینامهٔ مقصر حادثه، سوابق کیفری طرفین حادثه، سوابق بیمه‌ای بیمه‌گذار و غیره همراه با متغیرهای پژوهش حاضر در تحقیقات دیگر مورد بررسی قرار گیرد و نتایج آن با نتایج به‌دست‌آمده دربارهٔ تقلبهای بیمه‌ای مقایسه و در صورت امکان راهکارهایی برای پیشگیری از این دست تخلفها ارائه شود.

از آنجایی که نفس بیمه تقلب‌پذیر است و با توجه به تعدد رشته‌های بیمه‌ای و امکان تقلب در آنها و همچنین متفاوت‌بودن معیارهای شناسایی تقلب در هر یک از رشته‌های بیمه‌ای (مانند فاکتورسازی در بیمهٔ درمان و ...) پیشنهاد می‌شود شرکتهای بیمه با همکاری یکدیگر و همچنین نظارت نهاد محترم ناظر اقدام به طراحی سیستم یکپارچه به منظور شناسایی عوامل مؤثر در تخلفهای بیمه‌ای پرداخته تا حتی‌الامکان پرونده‌هایی از این دست شناسایی شوند و همچنین با همکاری واحدهای محترم قضایی در جهت حفظ منافع بیمه‌گران قدم برداشت.

پیشنهاد می‌شود برای افزایش آگاهی عموم مردم در خصوص استفادهٔ افراد سودجو از بیمه‌نامهٔ شخص ثالث آنها به‌منظور دریافت خسارات متقلبانه، از طریق رسانه‌های عمومی اطلاع‌رسانی لازم توسط مراجع زیربط انجام گیرد. همچنین به نظر می‌رسد ایجاد سازمانی به‌عنوان مثال تحت عنوان «انجمن خبرگان ضد تقلب در صنعت بیمه» متشکل از افسران خبرهٔ تحقیق و پزشکان پزشکی قانونی در این مسیر می‌تواند یاری‌ده شرکتهای بیمه باشد که با برون‌سپاری پرونده‌های مشکوک به تقلب برای بررسی بیشتر و حتی‌الامکان در صورت تأیید به طور مستقیم با ارجاع به مراجع قضایی اعمال قانون دربارهٔ متقلبان صورت گیرد.

و در پایان و با توجه به موارد مورد بحث، به نظرمی رسد یکی از مهم‌ترین عواملی که منجر به موفقیت دریافت خسارات کلاهبردارانه از شرکتهای بیمه‌ای توسط افراد سودجو می‌شود این است که آنها به سهولت دریافت خسارت از جانب شرکتهای بیمه به دلیل وجود قوانین اجبارانه از سوی نهاد ناظر و دستگاه‌های قضایی در برابر شرکتهای بیمه پی برده و این امر نیز می‌تواند عواقب مالی سنگینی را بر شرکتهای بیمه تحمیل و سرمایه‌های ارزشمند مالی را از شرکت خارج کند؛ بنابراین ضروری است همکاری متقابل مؤثر بین شرکتهای بیمه و دستگاه‌های قضایی به منظور شناسایی این افراد و جلوگیری از اعمال کلاهبردارانه و سودجویی آنها بیش از گذشته صورت گیرد.

منابع و ماخذ

تقوی‌فرد، س.م. جعفری، ز.، (۱۳۹۴). کشف تقلب در بیمهٔ بدنهٔ خودرو با بهره‌مندی از سیستم خبرهٔ فازی. مدیریت فناوری اطلاعات، دورهٔ ۷، شمارهٔ ۲، صص ۲۳۹-۲۵۸.

فیروزی، م. شکوری، م.کاظمی، ل. زاهدی، س.، (۱۳۹۰). شناسایی تقلب در بیمهٔ اتومبیل با استفاده از روش‌های داده‌کاوی. پژوهشنامه بیمه، سال ۲۶، شمارهٔ ۳. صص ۱۰۳-۱۲۸.

Ai, J.; Brockett, P.L.; Golden, L.L.; Guillén, M., (2013). A robust unsupervised method for fraud rate estimation. Journal of Risk and Insurance, 80(1), pp.121-143.

Bing X., (2013). Study of Fraud Identification Model for Vehicle Insurance [D]: Jilin: Jilin University.

Brockett, P.L.; Derrig, R.A.; Golden, L.L.; Levine, A.; Alpert, M., (2002). Fraud classification using principal component analysis of RIDITs. Journal of Risk and Insurance, 69(3), pp.341-371.

Cross, C.; Blackshaw, D., (2015). Improving the police response to online fraud. Policin: A Journal of Policy and Practice, 9, pp. 119-128.

Derrig, R.A., (2002). Insurance fraud. Journal of Risk and Insurance, 69(3), pp. 271-287.

Dionne, G., (2000). The empirical measure of information problems with emphasis on insurance fraud. In Handbook of Insurance (pp. 395-419). Springer, Dordrecht.

Dionne, G.; Gagne, R., (2002). Replacement cost endorsement and opportunistic fraud in automobile insurance. Journal of Risk and Uncertainty, 24(3), pp. 213-230.

Gill, K.M.; Woolley, A.; Gill, M., (2005). Insurance fraud: the business as a victim?. In Crime at work (pp. 73-82). Palgrave Macmillan, London.

Ishida, C.; Chang, W.; Taylor, S., (2016). Moral intensity, moral awareness and ethical predispositions: The case of insurance fraud. Journal of Financial Services Marketing, 21(1), pp.4-18.

Tang, J., (2013). Construction of auto insurance anti-fraud system based on data mining technology. Shanghai Insurance, pp.39-42.

Tan, P.N.; Steinbach, M., (2006). Kumar. Introduction to Data Mining.

Terisa, R., (2010). Improving the defense lines: the future of fraud detection in the insurance industry (with fraud risk models, text mining, and social networks). In SAS Global forum, Washington.