



به نام خدا



گزارش کارشناسی
ارزیابی ریسک قلب
در مزایای بیمه بیکاری



وزارت بهداشت، درمان و آموزش پزشکی

معاونت بیمه بیکاری





موسسه عالی پژوهش تأمین اجتماعی

ارزیابی ریسک تقلب در مزایای بیمهٔ بیکاری

واژه‌های کلیدی:

مجری: سید جواد طباطبائی منش
ناظر علمی: دکتر امین حسن‌زاده، هاله فرزانه

ارزیابی تقلب
بیمهٔ بیکاری
بیمه‌های اجتماعی
ریسک

شماره مسلسل: ۹۵(۱۶)AC۰۴ DF

تاریخ انتشار گزارش: ۹۵/۱۰/۵



فهرست مطالب	
۴	پیشگفتار مؤسسه
۶	چکیده
۸	مقدمه
۱۱	پیشینه ارزیابی ریسک تقلب در مزایای بیمهٔ بیکاری
۱۵	سابقهٔ تقلب در مزایا در بیمهٔ بیکاری ایران
۱۸	متدولوژی و داده‌ها
۲۳	نتایج
۳۱	منابع
۳۳	پیوست ۱ نتایج مقایسه‌ای عملکرد ارزیابی‌ها
۳۴	پیوست ۲ الگوریتم
۳۵	پیوست ۳ الگوریتم
۳۸	پیوست ۴ مدل

پیشگفتار مؤسسه

یکی از دغدغه‌های اصلی شاغلین، تأمین معیشت در دوره بیکاری بوده و ایران جزء اولین کشورهای خاورمیانه است که بیمه بیکاری را برای افراد تحت پوشش قانون کار و قانون تأمین اجتماعی خود پیش‌بینی کرده است. به موجب قانون بیمه بیکاری کشور، آن دسته از بیمه‌شدگانی که بدون میل و اراده، بیکار شده و آماده به کار باشند از مستمری بیمه بیکاری برخوردار می‌شوند. در حال حاضر غیرارادی بودن بیکاری، ملاک اصلی برقراری حمایت‌های قانونی برای بیمه‌شدگان بیکار است و تنها افرادی می‌توانند از مقرری بیمه بیکاری استفاده نمایند که به صورت غیرارادی از کار بیکار شده و خروج از اشتغال به اراده و خواست آنها نبوده باشد.

در سال‌های اخیر تقلب‌های بیمه‌ای از مسائل مهم و خسارت‌زا در تمام رشته‌های بیمه‌ای بوده، به‌ویژه در بیمه‌های اجتماعی که ورود و خروج افراد به این بیمه‌ها باز است و منابع مالی این بیمه‌ها به راحتی در اختیار افراد جامعه قرار می‌گیرد و خود می‌تواند منجر به ایجاد فضای بی‌اعتمادی در اجرای تعهدات آتی سازمان بیمه‌گر گردد. تقلب در حقوق و مزایای بیمه بیکاری همواره یکی از موضوعات حساس و مورد توجه در حوزه





بیمه‌های اجتماعی است که طبق قوانین، جزء جرائم کیفری بوده و قابل پیگیری است. هرچند قوانین تأمین اجتماعی به مقررات کیفری و ضمانت اجرای مربوط به تقلب و سوءاستفاده از اموال صندوق تأمین اجتماعی پرداخته است؛ ولی نوع و شدت ضمانت اجرای مقرر به حدی پایین است که عملاً ناکارآمد است. بنابراین یکی از بهترین روش‌ها ارزیابی تقلب، کنترل آن در همان مراحل اولیه شکل‌گیری است.

باتوجه به هزینه‌های هنگفتی که سالانه سازمان تأمین اجتماعی بابت تقاضاهای متقلبانه در استفاده از مزایای بیمه بیکاری متحمل می‌شود، مؤسسه عالی پژوهش تأمین اجتماعی بر خود دانسته تا به ماهیت ریسک تقلب و انواع روش‌های متقلبانه در استفاده من غیر حق از مزایای بیمه بیکاری نگاه دقیق‌تری بیندازد که امید است گزارش حاضر مورد توجه و استفاده تصمیم‌گیران، تصمیم‌سازان و متخصصین این حوزه قرار گیرد.

دکتر شهرام غفاری

رئیس مؤسسه

زمستان ۱۳۹۵



تقلب خطری است پنهان و غیر آشکار که افراد، همواره به دنبال مخفی کردن آن هستند. آنچه در حال حاضر توجه عموم جوامع، رسانه‌ها، سرمایه‌گذاران و قانون‌گذاران را به خود جلب کرده است، گستردگی و وسعت دامنه تبعات ارتکاب اعمال متقلبانه در موضوعات مالی است. بیمه بیکاری در زمره یکی از مهم‌ترین تعهدات سازمان تأمین اجتماعی است که با پرداخت به موقع کمک‌های مالی و حمایت‌های درمانی تا زمان یافتن شغلی مناسب، از نیروی کار مستعد و خانواده ایشان حمایت می‌کند. این تعهدات به دلیل وسعت تأثیرات اقتصادی و اجتماعی در جامعه و کشور، از اهمیت بسزایی برخوردار است. تقلب در حقوق و مزایای بیمه بیکاری همواره یکی از موضوعات حساس و مورد توجه در حوزه بیمه‌های اجتماعی است که طبق قوانین، جزء جرائم کیفری بوده و قابل پیگیری است. در حال حاضر بهترین روش به منظور ارزیابی تقلب، کنترل آن در همان مراحل اولیه شکل‌گیری و با توجه به اطلاعات تقلب‌های کشف شده گذشته است.

در این گزارش، ماهیت ریسک تقلب و انواع روش‌های متقلبانه در استفاده من غیر حق از مزایای بیمه بیکاری بیان و اثرات نامطلوب و بلندمدت این ریسک به عنوان چالشی مهم در مسیر مدیریت صحیح و تخصیص بهینه منابع و مصارف صندوق بیمه بیکاری مورد توجه و ارزیابی قرار گرفته است؛ همچنین با توجه به وجود پایگاه داده مناسب در خصوص مقرری‌بگیران بیمه بیکاری سازمان تأمین اجتماعی از سه

روش داده‌کاوی تفحصی شامل شبکه عصبی، شبکه بیزی و درخت تصمیم به‌منظور مدل‌سازی ریسک تقلب و یافتن الگوهای مناسب جهت کشف، کنترل و پیش‌بینی به‌موقع ادعاهای متقلبانه استفاده شده‌است؛ هر کدام از این مدل‌ها متناسب با الگوریتم محاسباتی خود پس از دریافت اطلاعات موردنیاز در قالب متغیرهای ورودی، شدت ریسک متقلبانه بودن یک ادعای بیمه بیکاری را با عددی در بازه احتمالی (۱ و ۰) مشخص می‌کند، این تشخیص می‌تواند به‌عنوان ابزاری سودمند همراه با کاهش قابل‌توجه در وقت و هزینه به کشف و ارزیابی به‌موقع تقلب در ادعاهای جدید و جاری بیمه بیکاری کمک نماید.

در فرایند مطالعه مستندات تجربی، این روش‌ها بر روی داده‌های واقعی شامل اطلاعات ۱۵,۹۸۳ ادعای بیمه بیکاری جدید و جاری، آزمایش و کارایی هر روش سنجیده شده است. روش‌های شبکه عصبی و شبکه بیزی به ترتیب با دقت ۸۸٪ و ۸۷٪ در ارزیابی صحیح متقلبانه یا عادی بودن ادعاها، بهترین کارایی را در مقایسه با روش درخت تصمیم با دقت ۸۴٪ در برداشته‌اند. همچنین بر اساس نتایج به‌دست‌آمده از تحلیل ساختاری دو مدل برتر شبکه عصبی و شبکه بیزی در می‌یابیم که متغیرهای سن، سابقه پرداخت حق بیمه، موقعیت جغرافیایی کارگاه، علت بیکاری و نوع شغل، صرف‌نظر از اولویت، بیشترین تأثیر را بر احتمال تقلب در بیمه بیکاری دارند.

در بخش پایانی گزارش و با استفاده از مدل‌های ارزیابی، ریسک تقلب یک نمونه ۲۳ تایی کاملاً تصادفی و جدید، اندازه‌گیری و سپس با نتایج حاصل از تحقیقات میدانی بازرسان بیمه بیکاری، نظیر به‌نظیر مقایسه شده‌است.



در ارتباط با واژه تقلب، معانی و تعاریف متعددی وجود دارد که هر کدام متناسب با شرایط وقوع تقلب از یکدیگر متفاوت‌اند. آنچه بیش از همه توجه عموم افراد جامعه، رسانه‌ها، سرمایه‌گذاران، جامعه مالی و قانون‌گذاران را به خود جلب کرده است، گستردگی و وسعت دامنه تبعات ارتکاب اعمال متقلبانه در موضوعات مالی است. لذا واژه تقلب به‌عنوان یک ریسک یا خطر همواره مورد توجه قرار داشته است.

در حال حاضر یکی از پرکاربردترین و عمومی‌ترین تعاریفی که در رابطه با واژه تقلب عنوان می‌گردد به‌صورت زیر است:

تقلب خطری است پنهان و غیر آشکار که افراد، همواره به دنبال مخفی کردن آن هستند. از این تعریف مشخص می‌شود که اولاً تقلب یک ریسک یا خطر بوده و ثانیاً به دلیل ماهیت پنهانی و مخفی کاری، به‌مراتب با ساختاری پیچیده‌تر و گسترده‌تر از سایر ریسک‌ها مورد ارزیابی قرار می‌گیرد.^۱

بدیهی است که ریسک تقلب نیز مانند سایر ریسک‌ها به‌طور مطلق حذف‌شدنی نیست ولی یکی از خصوصیات قابل توجه که در رابطه با این نوع خاص ریسک وجود دارد، خاصیت «فراگیر مثبت»^۲ بودن این ریسک است.^۳ به‌طوری که در حقیقت با شروع فرایند ارزیابی و شناسایی ریسک تقلب چه در ابتدای کار و چه در حین انجام کار، این ریسک خودبه‌خود و به‌طور ناگهانی، به‌شدت و به‌صورت گروهی و جمعی کاهش می‌یابد. این خاصیت ریسک تقلب است

1. Lee, 1993

2. Positive Extension

3. Abagnale, 2001



که از درون واکاوی شده و به صورت سلسله مراتبی^۱ به دیگر موارد تسری می‌یابد. اکثر مطالعات جامعه‌شناسی و روان‌شناختی نیز که در دنیا که با موضوع ریسک تقلب و ارتکاب اعمال متقلبانه صورت پذیرفته، بر این خاصیت تأکید داشته و آن را تأیید نموده است.^۲

زمانی که افراد متقلب از وجود یک سیستم ساختاریافته ارزیابی و کشف تقلب آگاهی یافته به نحوی که خود را هر لحظه در معرض این ارزیابی و شناسایی می‌یابند و از سوی دیگر به دلیل اینکه طبق قانون، تقلب یک جرم کیفری بوده و در صورت اثبات، مرتکبین آن علاوه بر بازگرداندن وجوه ناشی از کلاهبرداری، محکوم به جبران خسارت و جریمه مالی و حتی حبس می‌شوند، افراد خودبه‌خود و ناخواسته از ارتکاب اعمال متقلبانه و حتی فکر کردن به آن دوری می‌جویند.^۳

بیمهٔ بیکاری، کارگرانی را که به‌طور غیرارادی شغل خود را از دست داده‌اند مورد حمایت قرار می‌دهد؛ این حمایت با پرداخت کمک‌های مالی موقت به‌عنوان مقرری بیمهٔ بیکاری صورت می‌پذیرد؛ هدف از پرداخت این مقرری، حمایت به‌موقع از نیروی کار مستعد و خانواده‌ایشان تا زمان یافتن شغلی مناسب هست؛ افراد باید از شرایط قانونی لازم به‌منظور بهره‌مندی از این مزایا برخوردار باشند؛ کارگر و کارفرما هر دو در خصوص لزوم رعایت قانون در استفاده از مزایای بیمهٔ بیکاری مسئول هستند.^۴

مزایای بیمهٔ بیکاری علاوه بر کمک‌های مالی شامل حمایت‌های درمانی به افراد تحت تکفل فرد بیکار و ادامه سوابق بیمه پردازی وی نیز می‌گردد؛ بدیهی است به دلیل وجود مزایای بالا در چنین سیستم بیمهٔ بیکاری و از طرف دیگر دسترسی آسان عموم افراد به این مزایا، همواره انگیزه زیادی را در استفاده متقلبانه از این مزایا به وجود می‌آورد.

1. Hierarchical
 2. Hopenhayn & Nicolini, 1997
 3. Viaene, 2004
 4. Hopenhayn & Nicolini, 1997



در حال حاضر مزایای بیمه بیکاری در ایران به‌عنوان بخشی از خدمات و تعهدات سازمان تأمین اجتماعی ارائه می‌گردد که به دلیل وسعت اثرات اقتصادی و اجتماعی، از اهمیت بسزایی برخوردار است؛ اما نکته تأمل برانگیز و جالب توجه اینجاست که در سیستم تأمین اجتماعی ایران علاوه بر وجود همه خصوصیات مشترک بین سازمان‌های بیمه‌گر اجتماعی از جمله توزیع عادلانه ثروت و درآمد بین اقشار جامعه، عدم کسب سود و ... خصوصیت منحصر و خاص دیگری نیز وجود دارد که صرفاً مختص این سیستم و تعهدات آن است و آن دسترسی بسیار ساده و اقتضایی عموم افراد جامعه به منابع ناشی از تعهدات این سازمان است؛ به بیانی دیگر روش‌های دسترسی به منابع سازمان به راحتی در اختیار عموم افراد جامعه قرار دارد که این امر عمدتاً از قوانین و مقررات غیر کارشناسی و سهل‌الوصولی نشأت می‌گیرد که به‌مرور زمان و تحت فشارهای سیاسی طاقت‌فرسا به سازمان تحمیل گردیده است.

به‌عنوان مثال چنانچه یک فرد تصمیم بگیرد تا حقوق و مزایای بیمه بیکاری دریافت نماید به راحتی می‌تواند تا حداقل یک ماه و حداکثر دو ماه آتی از حقوق و مزایای مرتبط با آن (چه‌بسا تا آخرین روز استحقاق) بهره‌مند شود. به طوری که اگر فرد چند ماه قبل‌تر تصمیم به چنین کاری بگیرد حتی می‌تواند میزان مبلغ دریافتی خود را نیز با ارائه ریز دستمزدها و فهرست‌های خلاف ارسالی تا قبل از محاسبات مزایای بیمه بیکاری به گونه‌ای تنظیم کند که اولاً کمترین حق بیمه را به سازمان پرداخت نماید و ثانیاً در زمان بیکاری به راحتی مبالغ دلخواه خود را دریافت نماید. استمرار این فرایند متقلبانه، عامل اساسی و خمیرمایه اصلی فرایند شکست در مدیریت صحیح و تخصیص بهینه منابع و مصارف صندوق بیمه بیکاری و به تبع آن ایجاد چالش در مسیر ارائه تعهدات سازمان تأمین اجتماعی به جامعه مخاطبین خواهد بود که اثرات مخرب و جبران‌ناپذیری را نیز بر وضعیت اقتصاد جامعه، کارگران و مشاغل مختلف در پی خواهد داشت.

تقلب‌های مالی به شکل چشم‌گیری توجه عموم جوامع، رسانه‌ها، سرمایه‌گذاران و قانون‌گذاران را به خود جلب نموده است، دلیل این جلب توجه، به سبب تقلب‌هایی است که در شرکت‌های بزرگی از قبیل انرون و لوسنت رخ داده است؛ از طرف دیگر، تقلب‌های اخیر بیمه و بانکی در ایران، نیز بر اهمیت موضوع افزوده است. تقلب، یک دلیل مهم و مؤثر در شکست بسیاری از سازمان‌ها و شرکت‌ها قلمداد می‌شود، زیرا سرمایه‌گذاران، اعتباردهندگان و تحلیل‌گران مالی جهت سیاست‌گذاری، به صورت‌های مالی متکی هستند و به آن‌ها اعتماد می‌کنند^۱.

بر اساس آخرین تقسیم‌بندی مرکز مطالعات و تحقیقات جرائم مالی ایالات متحده در سال ۲۰۰۷ دسته‌بندی کلی انواع تقلب‌های مالی مطابق با جدول شماره (۱) ارائه گردیده است.

شرح	نوع تقلب
بیمه بیکاری، بیمه سلامت، بیمه عمر و حوادث و ...	تقلب بیمه‌ای
رهن، پول شویی، اختلاس، خلق پول و...	تقلب بانکی
اوراق قرضه، قاچاق کالا و ...	تقلب کالا و اوراق بهادار
تقلب‌های شرکتی، تقلب با استفاده از رسانه‌ها و ...	سایر

△ جدول ۱. دسته‌بندی کلی انواع تقلب‌های مالی.
منبع: مرکز مطالعات و تحقیقات جرائم مالی ایالات متحده در سال ۲۰۰۷

بر این اساس، تقلب بیمه‌ای به‌عنوان یکی از اصلی‌ترین تقلب مالی محسوب می‌شود؛ که در این میان ارزیابی ریسک تقلب در بیمه بیکاری به‌عنوان بخشی از مزایای بیمه‌های اجتماعی با توجه به سازوکار حاکم بر آن از اهمیت بسزایی برخوردار است، زیرا بیمه‌های اجتماعی، مسئولیت توزیع عادلانه ثروت و درآمد را نیز به عهده‌دارند. در کشورهایی که بیمه ریشه‌های عمیقی دارد، همواره نظارت و کنترل دقیق بر ریسک تقلب حاکم بر تعهدات مالی بیمه‌ها اعم از کوتاه‌مدت و بلندمدت به نحو بارزی دیده

1. Enron & Lucent
2. Ata & Seyrek, 2009

می‌شود؛ اما در کشورهایی که بیمه مانند سایر علوم، یک پدیده وارداتی است، فرایند مذکور از اهمیت قابل ملاحظه‌ای برخوردار نیست.^۱

از ۱۹۷۰ میلادی تاکنون تحقیقات متعددی در زمینه کشف و ارزیابی تقلب در استفاده از مزایای بیمه‌های اجتماعی صورت پذیرفته است. از این میان می‌توان به تحقیق شاول و ویس^۲ در رابطه با طراحی سیستم‌های بهینه پرداخت مقرری در مدت زمان استحقاق بیمه بیکاری اشاره نمود. بیلی^۳ تحقیقات مشابهی را با استفاده از روش کالیبراسیون و کنترل بهینه^۴ در کشور آمریکا انجام داد. همچنین هوپن هاین و نیکولینی^۵ تحقیقاتی را در این زمینه با استفاده از اطلاعات تقلب‌های ثبت شده گذشته مربوط به بیکاران تحت پوشش وزارت کار ایالات متحده، انجام دادند. دریگ و استاسزیوسکی^۶، ویزبرگ و دریگ^۷ و براکت و همکارانش^۸ در مقالات خود تکنیک‌هایی برای شناسایی ادعاهای متقلبانه و دست‌بندی کلاهبرداری‌های بیمه‌ای ارائه نمودند. سول و ونکات^۹، روش جدیدی را به منظور افزایش دقت و کارایی توضیح‌دهندگی مدل کشف تقلب در استفاده از روش درخت تصمیم و شبکه‌های عصبی به کار بردند، بطوریکه با ترکیب دسته‌گرهای^{۱۰} مختلف و همچنین روش‌های بهینه‌سازی مدل از جمله اعتبار سنجی مقطعی و کلی^{۱۱} به یک مدل ایده‌آل در کشف و ارزیابی تقلب دسترسی پیدا کردند.

با توجه به تنوع حجم و نوع داده‌ها، روش‌های آماری بسیاری برای کشف تقلب‌ها وجود دارند. این روش‌ها می‌توانند با ناظر یا بی‌ناظر باشند. در روش‌های با ناظر، نمونه‌هایی از موارد تقلبی و غیر تقلبی موجود است و مدلی ساخته می‌شود که بر اساس آن تقلبی یا غیر تقلبی بودن نمونه‌های جدید مشخص می‌شود. این روش جهت تشخیص تقلب‌هایی مناسب است که از قبل وجود داشته‌اند. روش‌های بی‌ناظر، به

1. Viaene, 2004
2. Shavell & Weiss, 1979
3. Baily, 1978
4. Calibration and Optimal Control
5. Hopenhayn & Nicolini, 1997
6. Derrig & Ostazewski, 1995
7. Weisberg & Derrig, 1993
8. Brocket et al, 1998
9. Sohl & Vankat, 1995
10. Classifiers
11. Bagging and Cross Validation

دنبال کشف نمونه‌هایی هستند که کمترین شباهت را با نمونه‌های نرمال دارند.^۱

بوست و وین برگ^۲ شبکه‌های عصبی بازگشتی^۳ را برای تفکیک صورت‌های مالی عادی از متقلبانه به کار بردند. آن‌ها به کمک قانون بن فرد^۴ در تحلیل داده‌ها، توزیع عددی مبالغ مالی را مورد ارزیابی قرار دادند. با مشخص شدن توزیع داده‌ها، با استفاده از قابلیت خاص روش شبکه‌های عصبی در تخمین‌های آماری که با پردازش فوق‌العاده روی توزیع داده‌ها به هم پیوسته شده‌اند و با تلفیق ساختار الگوریتم نورون‌های پردازش لایه درونی^۵ شش مدل شبکه عصبی و همچنین افزاز کلاس مقادیر ورودی متغیرها به بازه بین صفر و یک، توانستند مدل کارا و بهینه‌ای را در پیش‌بینی و ارزیابی ریسک تقلب ارائه نمایند که امکان آموزش و یادگیری مداوم برای مدل فراهم شده و سیستم ارزیابی حاصل از آن مرتباً دانش قبلی‌اش را به روزرسانی کند؛ نتایج ارزنده و کاربردی این مدل، در ارزیابی و کشف به موقع فرایندهای متقلبانه جریانات مالی چند هلدینگ نفتی و دارویی در ایالات متحده استفاده شده است.

ویلیامز و هووانگ^۶ با استفاده از روش‌های داده‌کاوی شبکه بیزی و شبکه عصبی بازگشتی و تکنیک شناسایی داده‌های پرت، مدلی را در ارزیابی ریسک تقلب ارائه نموده که مبتنی بر پاسخ معین قبلی نباشد بلکه و بر اساس توزیع تجمعی داده‌ها و شناسایی خوشه‌ای داده‌های پرت به عنوان خوشه‌ای که می‌تواند شامل موارد متقلبانه باشد، اقدام به مدل‌سازی ریسک تقلب نماید. این روش و مدل حاصل از آن در مواردی که اطلاعات کامل و صحیحی از وقوع تقلب در داده‌های مستند قبلی در دسترس نباشد، بسیار کاراست. از نتایج این مدل در ارزیابی خسارات ناشی از تقلب در تعهدات بیمه‌های درمانی کشور تایوان استفاده شده است. این افراد با استفاده از روش ماتریس اغتشاش^۷ دقت هر مدل را در پیش‌بینی مقادیر مشاهده شده واقعی اندازه‌گیری کردند.

از آنجائی که روند ارزیابی تعهدات مالی ارائه شده توسط سازمان‌های بیمه‌گر

1. Bolton & Hand, 2002
2. Bust, & Weinberg, 1998
3. Back Propagation Neural Networks
4. Benfords' Law
5. Hidden Layer
6. Williams, & Huang, 1997
7. Confusion Matrix

اجتماعی عمدتاً به صورت ممیزی (بازرسی)^۱ پس از پرداخت صورت می‌پذیرد لذا بسیاری از تقاضاهای متقالبانه شناسایی نشده یا بسیار دیر و غالباً با اتمام کامل دوره استحقاقی مقرر و پس از پرداخت مبالغ قابل توجهی در قالب تعهدات بیمه‌ای و درمانی به بیمه‌شده، شناسایی می‌گردند. این روش ارزیابی صرفاً فیزیکی بوده و در آن از سیستم‌های کامپیوتری به ندرت استفاده می‌شود. همچنین با توجه به اینکه همواره شیوه‌های جدیدی در کلاهبرداری به کار گرفته می‌شود، روش‌های مورد استفاده در کشف تقلب باید قابلیت کافی برای شناسایی تقلب را داشته باشند. هر چند کشف کاملاً خودکار کلاهبرداری‌ها در عمل ممکن نیست، اما استفاده از اطلاعات تقلب‌های کشف شده در گذشته و بهره‌گیری از تکنیک‌های آماری می‌تواند به کارشناسان در کشف و ارزیابی به موقع تقاضای متقالبانه که منجر به ایجاد تعهدات مالی می‌گردد کمک نماید.

در ایران، از گذشته دور تاکنون غالباً هیچ فرایند ارزیابی و کشف تقلب مؤثر و دارای زیرساخت مجدانه، مشخص و مدونی نه تنها در مزایای مرتبط با بیمه بیکاری بلکه در هیچ‌یک از تعهدات بلندمدت بین نسلی و هنگفت سازمان تأمین اجتماعی وجود نداشته است، این بدان معنا نیست که مطلقاً هیچ‌گونه تلاشی در این خصوص صورت نپذیرفته، بلکه بیانگر آن است که همواره به اقدامات و تلاش‌های مقطعی، اقتضایی، محدود و گذرا اکتفا شده و متأسفانه عمدتاً به دلیل اتکای بیش از حد به انجام بازرسی‌های وابسته به نیروی انسانی و مشکلات متعدد غالب بر آن و عدم استفاده از زیرساخت‌های علمی پایدار و قابل توسعه در حوزه تخصصی، این تلاش‌ها و اقدامات همواره به صورت چالشی و موقت در سازمان اجرا و پس از مدتی بسیار کوتاه کارایی خود را از دست داده و در کاربرد و عمل، اجرای آن به پایان رسیده است.

از طرفی در حال حاضر بزرگ‌ترین ابزار کنترلی سازمان همان بازرسی‌های فنی در اختیار سازمان است که با حضور فیزیکی افراد و به صورت میدانی صورت می‌پذیرد. در این میان معمولاً مسئولیت بازرسی و تشخیص ادعاها و تقاضاهای متقالبانه بیمه‌ای بر عهده افرادی است که بعضاً کم‌تجربه بوده و اغلب آموزش‌های لازم را در خصوص کشف تقلب، ندیده‌اند.

به‌طور کلی، متقاضیان بیمهٔ بیکاری در دو حالت با انگیزه‌های قوی مرتکب تقلب می‌شوند.^۱

حالت اول:

شرایطی که در آن فرد کاملاً آگاهانه و از همان ابتدا به‌طور کاملاً برنامه‌ریزی شده اقدام به ایجاد سابقه و پرداخت حق بیمه نموده و سپس به‌محض حصول حداقل شرایط ممکن، نسبت به خروج از حالت اشتغال و ادعای بیکاری و نهایتاً دریافت مقرری بیمهٔ بیکاری اقدام می‌نماید. متأسفانه این در حالی است که این افراد از ابتدای فرایند مذکور یعنی ارائه تقاضا تا به نتیجه رسانیدن آن و دریافت مقرری غالباً با کوچک‌ترین مانع و یا مشکلی که ناشی از ارزیابی عملکرد آن‌ها باشد، مواجه نمی‌شوند و با خیالی آسوده فرایند متقلبانه خود را تا پایان ادامه می‌دهند.

حالت دوم:

زمانی که در آن فرد صرفاً به دلیل وجود مزایای وسوسه‌انگیز و قابل توجه بیمهٔ بیکاری و همچنین محق دانستن خود در استفاده از این مزایا و بهره‌مندی از مزایای داشتن شغلی دیگر، آگاهانه دست به اعمالی می‌زند که موجبات اخراج موجه خود را از شغل فعلی‌اش فراهم می‌سازد. اگرچه ممکن است شخص به‌طور مستقیم انگیزه استفاده متقلبانه از مزایای بیمهٔ بیکاری را نداشته باشد ولی با این حال اقدام به انجام مجموعه فعالیت‌هایی نموده است که در صورت نداشتن بیمهٔ بیکاری این فعالیت‌ها را انجام نمی‌داد که این خود مصداق بارز ارتکاب تقلب و اعمال مجرمانه مرتبط با آن است. چرا که همان‌طور که می‌دانیم وجود بیمه و تعهد به جبران غرامت، موجب باعث می‌شود بیمه‌شدگان آگاهانه ریسک‌هایی را متحمل شوند که در صورت عدم وجود بیمه از این ریسک‌ها دوری می‌جستند.

کلاهبرداری و تقلب در استفاده از مزایای بیمهٔ بیکاری به‌روش‌های مختلفی صورت می‌پذیرد. مطابق با استاندارد کدگذاری جداول پایه سیستم اتوماسیون بیمه‌ای شعب سازمان تأمین اجتماعی (جدول ۲)، ۱۸ کد مختلف به‌عنوان علل قطع مقرری بیکاری به شرح زیر تعریف شده‌اند.

1. Wilson, 2003

کد	شرح	کد	شرح
۱	پایان استحقاق بیمه‌شده	۱۰	قطع مقرری به دلیل عدم احراز شرایط
۲	اشتغال مجدد	۱۱	اعزام به خدمت سربازی
۳	درخواست وزارت کار	۱۲	عدم شرکت در دوره‌های آموزش فنی و حرفه‌ای
۴	احراز شرایط بازنشتگی مقرری‌بگیر	۱۳	ارادی بودن بیکاری
۵	احراز شرایط از کارافتادگی مقرری‌بگیر	۱۴	عدم شرکت در کلاس‌های سوادآموزی
۶	فوت مقرری‌بگیر	۱۵	استفاده از حداکثر مدت استحقاقی
۷	خودداری از پذیرش شغل مشابه	۱۶	انتقال پرونده بیمه‌شده به سایر شعب
۸	اعاده به کار مقرری‌بگیر با دریافت مزد	۱۷	اشتغال با پرداخت تسهیلات بانک رفاه
۹	عدم مراجعه بیمه‌شده	۱۸	اشتغال با تسهیلات صندوق مهر امام رضا (ع)

Δ جدول ۲. کدینگ علل قطع بیمه بیکاری.
منبع: استاندارد کدگذاری جداول پایه، سیستم اتوماسیون بیمه‌ای سازمان تأمین اجتماعی (سبا).

طبق آیین‌نامه اجرایی قانون بیمه بیکاری (مصوب ۱۳۶۹/۱۰/۱۲ هیئت وزیران)، تعداد ۸ مورد (۴۵٪) از این کدها (جدول ۳) به‌عنوان مصادیق بارز تقلب در نظر گرفته‌شده و زمانی مورد استفاده قرار می‌گیرند که علت قطع مقرری بیکاری به‌واسطه ارتکاب تقلب توسط فرد متقاضی به اثبات رسیده باشد.

نتایج بررسی‌های انجام‌شده بر روی بانک اطلاعات اصلاح‌شده مقرری‌بگیران بیمه بیکاری^۱ استان اصفهان در سال ۱۳۹۳ نشان می‌دهد، از تعداد کل ۱۵,۹۸۳ پرونده مقرری‌بگیر بیمه بیکاری شامل پرونده‌های جدید و جاری که به‌طور قطعی در این پایگاه به‌عنوان مقرری‌بگیر فعال ثبت‌شده‌اند و در طول سال حداقل یک ماه و بیشتر از مزایای بیمه بیکاری استفاده نموده‌اند، تعداد ۳,۱۱۸ پرونده (۱۹/۵٪ از کل پرونده‌ها) که وقوع تقلب در آن‌ها توسط بازرسان فنی سازمان به‌طور مستند و قطعی محرز گردیده و با یکی از کدهای استاندارد ۸ گانه مصداق تقلب (جدول ۳)، در سیستم اتوماسیون بیمه‌ای سازمان تأمین اجتماعی (سبا) ثبت‌شده‌اند، به‌عنوان تعداد قطعی پرونده‌های تقلب در مزایای بیمه بیکاری شناسایی و مدنظر قرار گرفته‌اند. بر همین

۱. سیستم مدیریت اطلاعات (MIS)، آمارهای ثبتی تعهدات کوتاه‌مدت، سیستم اتوماسیون بیمه‌ای سازمان تأمین اجتماعی استان اصفهان (سبا)، سال ۱۳۹۳.

اساس می‌توان گفت تقریباً به میزان ۲۰٪ از کل پرونده‌های فعال جاری لغایت پایان سال ۱۳۹۳ متقلبانه بوده و این در حالی است که یقیناً در بین پرونده‌های بیمه بیکاری که به عنوان پرونده‌های عادی شمارش شده‌اند، پرونده‌های متقلبانه فراوانی وجود دارد که به هر دلیل شناسایی نشده است.

کد	شرح	درصد فراوانی
۱	اشتغال مجدد	۶۵٪
۲	درخواست وزارت کار	۱۵٪
۳	ارادی بودن بیکاری	۱۰٪
۴	قطع مقرری به دلیل عدم احراز شرایط	۰٪
۵	خودداری از پذیرش شغل مشابه	۵٪
۶	عدم شرکت در دوره‌های آموزش فنی و حرفه‌ای	۵٪
۷	عدم مراجعه	۰٪
۸	عدم شرکت در کلاس‌های سوادآموزی	۰٪

Δ جدول ۳. مصادیق تقلب هشت گانه و درصد فراوانی مرتبط در سال ۱۳۹۳.
منبع: اطلاعات آمارهای ثبتی کمک‌های کوتاه‌مدت، سیستم اتوماسیون بیمه‌ای سازمان تأمین اجتماعی استان اصفهان (سبا)، سال ۱۳۹۳.

با استفاده از شاخص‌ها و متغیرهای به کار رفته در فرمول محاسباتی حقوق بیمه بیکاری شامل حداقل و حداکثر دستمزد سال ۱۳۹۳، وضعیت تأهل، سوابق پرداخت حق بیمه، دستمزد مبنای کسر حق بیمه و مدت زمان باقی‌مانده دریافت مقرری تا پایان دوره استحقاق (در صورت عدم شناسایی)، نسبت به محاسبه بر آورد مالی تقلب در سال ۱۳۹۳ اقدام شد که متأسفانه بر این اساس اعلام گردید، که تنها در استان اصفهان و صرفاً بر اساس تقلب‌های شناسایی شده مبلغ ۴۵,۰۰۰,۰۰۰,۰۰۰ ریال سالیانه از منابع سازمان تأمین اجتماعی صرف جبران تعهداتی می‌گردد که ماهیتی کاملاً متقلبانه دارند. همچنین هزینه‌های سنگین ناشی از ارائه تعهدات درمانی به فرد بیکار و خانواده وی و پرداخت حق بیمه نیابتی سازمان به فرد بیکار در طول مدت زمان بیکاری محاسبه نگردیده که می‌باید پس از محاسبه، به مبلغ فوق اضافه گردد. بدیهی است بار مالی ناشی از تقلب در مقیاس‌های بزرگ‌تر و کشوری به مراتب مبالغ، بالاتر و قابل تأمل‌تر است.

به‌عنوان مثالی ساده و کاربردی می‌توان به انواع مزایای طرح‌های بیمه‌ای قالی‌بافان و کارگران ساختمانی اشاره نمود که همواره به دلیل مزایا و تعهدات بالای این نوع بیمه‌ها در مقابل حق بیمه بسیار ناچیز آن، به شدت مورد طمع‌ورزی و توجه متقلبین و افراد سودجو قرار داشته به طوری که متأسفانه بر اساس آخرین مستندات بازرسی کشف‌شده تعداد قابل‌توجهی پزشک، وکیل پایه‌یک دادگستری، تاجر، پلافروش، عمده‌فروش آهن‌آلات و... در حال استفاده من‌غیر حق و متقلبانه از مزایای این طرح‌ها بوده و هستند.

متأسفانه در ابتدای اجرای طرح مذکور، عمدتاً به دلیل عدم وجود سازوکار بازرسی‌های فنی مناسب و یا کم‌رنگ بودن آن، حجم گسترده‌ای از این افراد توانستند از طرح‌ها و مزایای مرتبط با آن استفاده نمایند اما زمانی که پس از گذشت سال‌ها، چند دستورالعمل مبنی بر افزایش کمی و کیفی بازرسی‌های فنی ابلاغ گردید، خود باعث تغییرات کاهشی و ناگهانی قابل‌توجهی در روند جاری ادعاها و تقاضاهای متقلبانه شد که با کمترین تبعات و چالش‌های سیاسی، اقتصادی و اجتماعی پیرامون سازمان نیز همراه بود و مهم‌تر از همه لازم به ذکر است که این نتایج در شرایطی به‌دست آمده است که به دلیل وجود مشکلات متعدد در روش‌های سنتی جاری بازرسی حضوری بازرسان فنی سازمان، بسیاری از این ارزیابی‌ها را در شعب به‌صورت ناقص و سلیقه‌ای اجرا می‌کنند و همان‌طور که ذکر شد اگر یک ارزیابی ساده نیز انجام شود می‌تواند بسیار تأثیرگذار و منجر به کاهش چشمگیری در شدت وقوع تقلب و اعمال متقلبانه در ادعاهای بیمه‌ای داشته‌باشد.

متدولوژی و داده‌ها

جامعه آماری تحت بررسی شامل تمام افرادی است که بر اساس قانون بیمه بیکاری (مصوب ۱۳۶۹/۰۶/۲۶ مجلس شورای اسلامی)، در بانک اطلاعاتی مقررری‌بگیران سازمان تأمین اجتماعی استان اصفهان در سال ۱۳۹۳ ثبت گردیده‌اند. این جامعه آماری شامل اطلاعات ۱۵,۹۸۳ نفر مقررری‌بگیر بیمه بیکاری است، که یا دارای پرونده فعال هستند و مقررری دریافت می‌کنند یا متقاضی جدید می‌باشند و یا صرفاً به علت ارتکاب اعمال متقلبانه مقررری آن‌ها قطع گردیده است. لازم به ذکر است که این اطلاعات پس از انجام فرایند آماده‌سازی داده‌ها شامل برچسب‌گذاری، به‌روزرسانی، حذف و پالایش برخی موارد مخدوش و ناقص، در فاز محاسباتی مورد استفاده قرار گرفته‌اند.

نکته حائز اهمیت این است که در این مطالعه به منظور ارتقاء سطح کیفی فرایند تحلیل داده در برآزش بهترین مدل شناسایی تقلب در طول دوره مورد نظر (سال ۱۳۹۳)، تعداد دفعات ورود و خروج افراد به صندوق بیمه بیکاری به همراه تغییرات ناشی از آن مورد ارزیابی قرار گرفته است. سوابق پرداخت حق بیمه افراد انتخاب شده بین ۱ تا ۲۷ سال متغیر بوده که این افراد در برخی موارد در طول سوابق پرداخت حق بیمه شان به دفعات مختلف بین ۱ تا ۴ بار از بیمه بیکاری استفاده نموده اند. سن افراد بین ۲۵ تا ۵۰ سال قرار دارد که شامل زن، مرد، مجرد و متأهل است. تعداد تبعی افراد متأهل نیز به طور استاندارد بین صفر تا چهار نفر در نظر گرفته شده است.

به منظور ساخت مدلی ریاضی، نیاز به داده‌هایی از هر دودسته تقاضاهای متقلبانه و عادی داریم. بر همین اساس متقلبانه یا عادی بودن یک تقاضای بیمه بیکاری، به عنوان متغیر وابسته (پاسخ) در نظر گرفته می‌شود. مقدار یک برای متغیر وابسته به معنای متقلبانه بودن تقاضا و مقدار صفر به معنای عادی بودن آن تقاضا است. در این مطالعه، فرایند شناسایی تقلب با استفاده از نه متغیر مستقل صورت گرفته است. به منظور انتخاب متغیرها از روش فاصله‌ای هاسدرف^۱ استفاده شده است. روش فاصله‌ای هاسدرف ابتدا مجموعه‌هایی بسته و غیر تهی از کلاس متغیرهای مستقل و کلاس متغیر پاسخ در یک فضای متریک ایجاد می‌کند سپس با در نظر گرفتن مکان هر متغیر در فضای متریک، شدت رابطه هر متغیر از یک مجموعه را به عنوان نقطه‌ای از یک فضای متریک با سایر متغیرها در مجموعه‌های دیگر به عنوان سایر نقاط فضای متریک اندازه‌گیری می‌نماید و با تکرار این فرایند به دفعات زیاد مطابق با الگوریتم هاسدرف (پیوست ۲) که کدنویسی آن نیز در نرم‌افزار آماری R انجام شده است، مناسب‌ترین متغیرها را به منظور استفاده در عملیات داده کاوی ارائه می‌نماید. الگوریتم هاسدرف بر این اساس شکل گرفته که در توپولوژی، فضایی است که در آن بتوان نقاط را با همسایگی جدا کرد؛ فرض کنید X یک فضای توپولوژیکی باشد، می‌گوییم X و Y توسط همسایگی‌هایی از هم جدا شده‌اند اگر همسایگی مانند U از X و همسایگی مانند V از Y موجود باشند به طوری که اشتراک این دو همسایگی تهی باشد، در صورتی که هر دو نقطه در X را توسط همسایگی‌هایی از هم جدا نمود، فضای X هاسدرف است. در الگوریتم هاسدرف به کاررفته در این مطالعه،

1. Hausdorff Distance Method

تعداد متغیرهای مرتبط با مجموع داده‌ها با k نشان داده شده است. در گام (۱)، فضای متریک هاسدرف برای v_i ها در دو همسایگی x_1 و x_2 با متر H_i برای تمام آنها و k متغیر محاسبه می‌شود؛ در گام (۲)، H_i های محاسبه شده بر اساس مجموعه‌های تهی (s_1, \dots, s_k) نظیر به نظیر مرتب می‌شوند (تعداد k ها و H_i های محاسبه شده با هم برابرند). در این گام شرط اساسی تهی بودن همسایگی‌ها با مقایسه تناظر $(\dots, H_i, S_i$ ها کنترل می‌شود تا مشخص گردد کدام v_i ها در فضای هاسدرف قرار می‌گیرد؛ در گام (۳)، و (a) هر کدام از v_i هایی که در فرایند قبل، انتخاب می‌شوند، جایگزین مقادیر تهی S_i ها می‌گردند؛ در مراحل (b) و (c) سطح اطمینان بر اساس مدل انتخابی که متغیرها در آن به کار برده می‌شوند، تعیین می‌گردد؛ مراحل (d) و (e) به منظور تکرار فرایند گام (۳) تا تکمیل آخرین اندیس و متغیر، به کار می‌روند و در نهایت گام (۴) متغیرهای نهایی و قابل استفاده را ارائه می‌دهد. علاوه بر این در این خصوص به منظور بهبود نتایج ارزیابی‌ها از نظرات کارشناسان مرتبط با بیمه بیکاری و امور فنی بیمه شدگان نیز بهره گرفته شده است. جدول (۴) انواع متغیرهای استفاده شده در مدل سازی ریسک تقلب را همراه با توضیحات مختصری در خصوص ساختار هر یک ارائه می‌نماید.

متغیر	نام	نوع	سطوح اندازه گیری	نقش در مدل
X_1	جنس	گسسته	باینری	توضیحی-مستقل)
X_2	وضعیت تأهل	گسسته	باینری	توضیحی-مستقل)
X_3	سن	پیوسته	فاصله‌ای	توضیحی-مستقل)
X_4	تعداد تبعی	گسسته	فاصله‌ای	توضیحی-مستقل)
X_5	سوابق پرداخت حق بیمه	پیوسته	فاصله‌ای	توضیحی-مستقل)
X_6	شعبه	گسسته	رده‌ای - اسمی	توضیحی-مستقل)
X_7	علت بیکاری	گسسته	رده‌ای - اسمی	توضیحی-مستقل)
X_8	تعداد تکرار مقرر	گسسته	فاصله‌ای	توضیحی-مستقل)
X_9	شغل	گسسته	رده‌ای - اسمی	توضیحی-مستقل)
Y	شناسایی	گسسته	باینری	پاسخ-وابسته)

Δ جدول ۴. متغیرهای استفاده شده در مدل سازی ریسک تقلب.

در فرایند آماده‌سازی داده‌ها به‌منظور انجام عملیات داده‌کاوی از بین ۱۵,۹۸۳ پرونده مقرر بگیر بیکاری مورد مطالعه، به روش نمونه‌گیری تصادفی ساده بدون جایگزاری، ۸۰٪ پرونده‌ها (۱۲,۷۸۶ مورد) به‌منظور مدل‌سازی و ۲۰٪ باقی‌مانده (۳,۱۹۷ مورد) به‌منظور آزمون دقت ارزیابی مدل‌های برازش داده‌شده مورد استفاده قرار گرفته‌اند.

تمامی ابزارهای داده‌کاوی استفاده‌شده در این تحقیق به‌منظور مدل‌سازی و تصمیم‌گیری به کمک نرم‌افزار IBM Spss Modeler (ver. 14.2) صورت پذیرفته است.

پس از گردآوری و آماده‌سازی داده‌ها، به‌منظور مدل‌سازی ارزیابی ریسک تقلب در فاز محاسباتی از سه روش شبکه عصبی^۱، شبکه استنباط بیزی^۲ و درخت تصمیم^۳ استفاده شده است. هر کدام از این روش‌ها متناسب با منطق الگوریتم محاسباتی خود (پیوست ۳)، به مدل‌سازی و پیش‌بینی ریسک تقلب می‌پردازد.

باتوجه به اینکه هریک از این روش‌ها دارای دقت و توان متفاوتی در ارزیابی ریسک تقلب هستند، با استفاده از روش ماتریس اغتشاش^۴ به‌عنوان یک روش رایج در ارزیابی دقت مدل‌ها، کارآمدترین و دقیق‌ترین مدل، انتخاب و به‌منظور استفاده کاربردی در ارزیابی ریسک تقلب در مزایای بیمه بیکاری مورد استفاده قرار می‌گیرد.

هریک از روش‌های مدل‌سازی باتوجه به داده‌های موجود در ادعاهای بیمه بیکاری، امکان پیش‌بینی ارتکاب تقلب را فراهم می‌آورند، به‌طوری‌که نیازی به روزآمدی ثابت نداشته اما با تحلیل اطلاعات خاصی، امکان یادگیری مداوم را فراهم می‌آورند. در این روش‌ها هر تقاضای ایجادشده مورد تحلیل قرار گرفته و یک تخمین آماری مبتنی بر اینکه آیا داده‌ها با پرونده‌های مطلوب مطابقت داشته‌اند یا خیر، ارائه می‌گردد. آموزش سیستم با استفاده از شواهد تجربی و مستندات مربوط به تقلب‌هایی که از قبل در سیستم وجود داشته، (به‌منظور کشف تقلب) به انجام می‌رسد. به‌محض تحقق این امر، سیستم از دانش قبلی برای کشف و تشخیص متقلبانه بودن تقاضاها، استفاده می‌نماید. از این سیستم می‌توان هم در کشف تقلب‌ها قبل از پرداخت و هم بعد از پرداخت مقرر بیمه بیکاری استفاده نمود. راهکارهای مورد استفاده رویکرد در کشف تقلب،

-
1. Neural Network
 2. Bayesian Network
 3. Decision Tree
 4. Confusion Matrix

شامل نمایه‌سازی داده‌ها^۱، مدل تحلیل پیشرفته^۲ و رده‌بندی امتیازی^۳ است.

در محاسبات از روش نمایه‌سازی داده‌ها استفاده شده است. در این روش تمام اطلاعات گذشته تقاضاهای بیمه بیکاری شامل متقالبانه و عادی در یک فایل تلفیق می‌شوند به طوری که برای برنامه قابل فهم باشد. به منظور شناخت الگوی تقلب در این روش از مدل‌های تحلیل پیشرفته استفاده شده است. در این مدل‌ها، داده‌ها با اطلاعات منابع مختلف مقایسه می‌شوند تا سرانجام الگویی یافت شود که تقلب یا سوءاستفاده از مزایای بیمه بیکاری را پیشنهاد نمایند. به عبارت دیگر برای مواردی که هنوز رابطه بین متغیرها مشخص نشده باشد و افراد نتوانسته‌اند رابطه بین متغیرها را حدس بزنند، مدل‌های حاصل از روش‌های سه‌گانه، به خوبی پاسخگو بوده و می‌تواند روابط بین ورودی‌ها و خروجی‌های مختلف را شناسایی نمایند. در این زمان است که می‌توان موارد جدید را وارد و رفتار آن‌ها را پیش‌بینی نمود. در هر یک از این روش‌ها، قابلیت یادگیری بر اساس تجربه فراهم می‌شود که این عمل موجب روزآمدی و یادگیری خودکار سیستم نیز می‌شود. در اینجا ما برعکس سیستم‌های خبره، برنامه‌نویسی نداشته (به عبارت دیگر لازم نیست قواعدی را در پایگاه دانش ذخیره نماییم) بلکه تقاضاهایی (مثلاً تقاضای شماره ۱، ۲ و...) تعریف نموده و ارتباط میان اجزای مختلف این تقاضا به منظور ردیابی تقلب مشخص می‌شود؛ در این مرحله، عموماً داده‌ها به دو مجموعه آموزش و آزمون (تأیید) تقسیم می‌شوند. داده‌های آموزشی برای استخراج قواعد و روابط و نهایتاً برآزش مدل، مورد استفاده قرار می‌گیرند، در حالی که داده‌های آزمون (تأیید) به منظور مشخص شدن کیفیت قواعد استخراج شده در یک مجموعه متفاوت از داده‌ها، به چه کیفیتی عمل می‌کنند. بر مبنای نتایج اجرای مدل بر روی همین داده‌ها، عملکرد مدل ارزیابی شده؛ در نتیجه مدل حاصل یا اصلاح می‌گردد یا به منظور حل مسئله به کار برده می‌شود.

زمانی که مناسب‌ترین مدل بر اساس این روش برآزش داده شد، با ورود اطلاعات جدید در مدل، مقدار متغیر خروجی که همان شدت متقالبانه بودن یک تقاضا است به صورت درصدی بین صفر تا ۱۰۰ ارزیابی می‌گردد.

1. Data Profiling
2. Advanced Analytical Model
3. Rank scoring

پس از برازش مدل‌ها (پیوست ۴)، دقت هریک مورد ارزیابی قرار گرفته که بر این اساس مدل شبکه عصبی با دقت ۸۸٪ در تشخیص ادعاهای متقلبانه جدید در رتبه اول و به‌عنوان بهترین مدل ارزیابی ریسک تقلب انتخاب می‌گردد. همچنین مدل شبکه بیزی با دقت ۸۷٪ در تشخیص ادعاهای متقلبانه جدید، رتبه دوم را به خود اختصاص می‌دهد و درنهایت درخت تصمیم با دقت تقریبی ۸۴٪ جایگاه آخر و رتبه سوم را در این تشخیص به خود اختصاص می‌دهد.

یکی از مهم‌ترین موضوعاتی که پس از مدل‌سازی ریسک تقلب مورد توجه قرار می‌گیرد، تعیین و اولویت‌بندی عوامل تأثیرگذار در احتمال وقوع این ریسک است. هر کدام از روش‌های به‌کاررفته در مدل‌سازی ریسک تقلب، باتوجه‌به ساختار مدل و بر اساس شدت رابطه بین متغیرهای به‌کاررفته در مدل، قابلیت آن را دارند که به‌طور جداگانه مهم‌ترین عوامل و فاکتورهای تأثیرگذار بر احتمال وقوع تقلب را تعیین نمایند، بر همین اساس در مدل شبکه عصبی که به‌عنوان دقیق‌ترین مدل ارزیابی شده‌است، متغیرهای نوع شغل، سابقه پرداخت حق بیمه، سن، موقعیت جغرافیایی کارگاه و علت بیکاری به ترتیب بیشترین تأثیر را بر احتمال تقلب در بیمه بیکاری دارند. در مدل شبکه بیزی که به لحاظ دقت ارزیابی در جایگاه دوم قرار دارد، متغیرهای سن، سابقه پرداخت حق بیمه، موقعیت جغرافیایی کارگاه، علت بیکاری و جنسیت به ترتیب بیشترین تأثیر را بر احتمال تقلب در بیمه بیکاری دارند؛ در مدل درخت تصمیم که در جایگاه سوم دقت تشخیص قرار می‌گیرد، متغیرهای موقعیت جغرافیایی کارگاه، جنسیت، تعداد افراد تحت تکفل، علت بیکاری و وضعیت تأهل به ترتیب بیشترین تأثیر را بر احتمال تقلب در بیمه بیکاری دارند.

اگر به‌دقت مهم‌ترین عوامل تأثیرگذار بر احتمال تقلب را به‌صورت مستقل بررسی کنیم، درمی‌یابیم که بر اساس نتایج حاصل از دو مدل برتر شبکه عصبی و شبکه بیزی، متغیرهای سن، سابقه پرداخت حق بیمه، موقعیت جغرافیایی کارگاه، علت بیکاری و نوع شغل، صرف‌نظر از اولویت، بیشترین تأثیر (بالا‌تر از ۲۰٪) را بر احتمال تقلب در بیمه بیکاری دارند. همان‌طور که در جدول (۵) مشاهده می‌شود هر کدام از این

متغیرها در هریک از مدل‌ها با تفاوتی اندک در اولویتی متفاوت قرار گرفته‌اند. لازم به ذکر است، سایر متغیرها و عواملی که در مدل‌سازی به کار گرفته شده و در جدول (۵) نیز عنوان گردیده‌اند، دارای تأثیری کمتر از ۲۰٪ بر احتمال تقلب در بیمهٔ بیکاری می‌باشند.

ردیف	متغیر (عامل تحت بررسی)	مدل شبکه عصبی	مدل شبکه بیزی	مدل درخت تصمیم
۱	سن	۳۱٪	۳۴٪	۱۳٪
۲	سابقه پرداخت حق بیمه	۳۴٪	۲۶٪	۱۵٪
۳	موقعیت جغرافیایی کارگاه	۲۰٪	۲۶٪	۲۷٪
۴	علت بیکاری	۲۰٪	۲۵٪	۲۲٪
۵	نوع شغل	۳۴٪	۲۰٪	۱۰٪

△ جدول ۵، میزان و عوامل تأثیرگذار بر احتمال تقلب در بیمهٔ بیکاری.

بر اساس محاسبات صورت گرفته بر توزیع فراوانی کدهای علت قطع، ثبت شده در بانک اطلاعات مقررری‌بگیران استان اصفهان صورت پذیرفته است، انواع تقلب در مزایای بیمهٔ بیکاری را می‌توان به‌طور عمده به چهار اولویت اصلی زیر طبقه‌بندی نمود:

▲ اولویت اول با درصد فراوانی نسبی ۶۵٪:

مخفی کردن سایر درآمدهایی که فرد به‌واسطهٔ اشتغال پنهان دریافت می‌نماید.

▲ اولویت دوم با درصد فراوانی نسبی ۲۰٪:

ترک شغل و خروج از کار به‌طور کاملاً ارادی و با برنامه قبلی و با استفاده از برگه‌های قرارداد کار ساختگی در تاریخ و زمان‌های مشخص.

▲ اولویت سوم با درصد فراوانی نسبی ۱۰٪:

اخراج از محل کار، به‌طوری‌که اخراج ناموجه بوده و به دلایل ساختگی و تباری با کارفرما بر نامه‌ریزی شده‌باشد.

▲ اولویت چهارم با درصد فراوانی نسبی ۵٪:

عدم جستجو و پیگیری‌های لازم به منظور یافتن شغل مناسب توسط فرد، با سرباززدن از شرکت در کلاس‌های آموزش فنی و حرفه‌ای.

با توجه به دقت بالا این مدل‌ها در شناسایی صحیح متقلبانه یا عادی بودن ادعاهای بیمه‌بیکاری، موارد زیر در قالب اهداف کلی این گزارش ارائه می‌گردد.

در صفحه‌ی اصلی ثبت اطلاعات تقاضای مقررری‌بگیران بیمه‌بیکاری در سیستم اتوماسیون شعب سازمان تأمین اجتماعی، یک سطر یا منو یا یک کلید تابعی با عنوان ارزیابی ریسک ثقلب تقاضا تعبیه گردد. در این صورت همزمان با ثبت اطلاعات فرد متقاضی توسط کارشناس بیمه‌بیکاری، ریسک ثقلب آن تقاضا توسط مدل‌های ارزیابی، به صورت عددی بین صفر تا ۱۰۰ ارزیابی می‌گردد.

با توجه به اینکه به کمک مدل‌های ارزیابی، شدت ریسک ثقلب در ادعاهای بیمه‌بیکاری در همان مراحل اولیه ثبت ادعا به صورت برخط شناسایی می‌گردد، لذا می‌توان با منطقی هدفمند ساخت فرایند انجام بازرسی‌های فنی بیمه‌بیکاری واحد امور بیمه‌شدگان سازمان تأمین اجتماعی و همچنین از طرف دیگر با بررسی‌های مجدد و دقیق‌تر مدارک و مستندات ارائه‌شده فرد متقاضی و احیاناً ارجاع آن به واحدهای مرتبط، در خصوص آن دسته از ادعاهایی که مشکوک به ثقلب می‌باشند، اقدامات به موقع و پیشگیرانه لازم را به منظور کنترل و مقابله جدی با ارتکاب ثقلب انجام داد و از هدر رفت هنگفت منابع مالی ناشی از تعهدات ایجادشده از آن در سطحی گسترده جلوگیری نمود.

روش‌ها و تکنیک‌هایی که در این گزارش به منظور کشف و ارزیابی ثقلب به کاررفته‌اند را می‌توان در سایر شاخه‌های مزایای بیمه‌های اجتماعی از جمله، تعهدات بلندمدت مستمری‌بگیران، تعهدات کوتاه‌مدت ایام بارداری، حوزه بیمه سلامت و درمان و حتی متقاضیان استفاده از طرح‌های بیمه‌ای خاص و تخفیف‌دار نیز مورد استفاده قرار داد.

با توجه به دخالت عامل انسانی در پدیده ثقلب در مزایای بیمه‌ها، بررسی‌های بیشتر و استفاده از نظرات خبرگان و متخصصان بیمه، به کاربردی‌تر شدن نتایج این گزارش کمک می‌کند.

مدل و الگوهای ارائه‌شده در گزارش حاضر بر اساس شبکه احتمالی استنباط بیزی، الگوریتم شبکه‌های عصبی بازگشتی و همچنین ساختار منطقی درختان تصمیم، به‌گونه‌ای ساخته‌شده‌اند که با حداکثر دقت ممکن بتوانند احتمال یا میزان شدت متقلبانه بودن یک ادعای بیمه‌بیکاری را پیش‌بینی نمایند. همان‌طور که در بخش‌های قبل نیز توضیح داده شد، این مدل‌ها بر اساس تلفیقی از محاسبات پیشرفته آماری و ریاضیات، به‌دست آمده‌اند که پایه محاسبات عددی آن‌ها شامل حجم وسیعی از اطلاعات و داده‌های مستندشده و تجربه‌شده سازمان تأمین اجتماعی در ادوار گذشته است. این مدل‌ها در حقیقت الگو و نقشه راه‌گویی هستند، که منجر به کوتاه شدن و یا حتی حذف بسیاری از فرایندهای کاری طاقت‌فرسا و پرهزینه‌ای بوده که نشأت گرفته از روش‌های کنونی و سنتی ارزیابی ریسک تقلب در سازمان است. این مدل‌ها علاوه بر اینکه خود یادگیرنده، هوشمند و خود به‌روزرسان هستند، دارای روش‌های استفاده کاربردی و عملیاتی بسیار ساده و روان نیز می‌باشند، به‌طوری‌که اگر بخواهیم این مدل‌ها را در فاز اجرایی در شعب سازمان به کار ببریم، مطلقاً نیازی به هیچ‌گونه زیرساخت نرم‌افزاری یا سخت‌افزاری خاص و یا حتی کوچک‌ترین تغییری در روال جاری دستورالعمل‌ها، بخشنامه‌ها و دستورات اداری در حوزه بیمه‌بیکاری و پیرامون آن نخواهیم داشت.

اگرچه در طول فرایند ساخت این مدل‌ها از روش‌های نسبتاً پیچیده‌ای استفاده شده است ولی روش استفاده و بهره‌برداری از این مدل‌ها به‌قدری ساده است که یک کاربر عادی و با سطح دانش عمومی به‌راحتی می‌تواند به کمک یک فایل اکسل معمولی کم‌حجم و مختصر و یا حتی یک برگ کاغذ A4 (در صورت استفاده از مدل‌های درختی) از قابلیت ارزشمند این مدل‌ها در راستای کشف و ارزیابی شدت متقلبانه بودن یک ادعای بیمه‌بیکاری استفاده نماید.

روش انجام کار به این صورت است که این مدل‌ها در حقیقت اطلاعات موردنیاز خود را در قالب ۹ تا ۱۱ مشخصه ورودی بر اساس اسناد و مدارک ارائه‌شده توسط فرد متقاضی، از کاربر دریافت نموده و پس از پردازش در لایه‌های درونی مدل و به‌دور از چشم کاربر، شاخص ریسک تقلب را به‌صورت درصدی بین صفر تا ۱۰۰

ارائه می‌نمایند. این عدد در واقع خروجی مدل و همان شدت یا احتمال متقلبانه بودن یک ادعای بیمه بیکاری است که به صورت انفرادی برای تک‌تک افراد با مشخصات مختلف ارائه می‌گردد.

در صورتی که از مدل‌های درختی استفاده شود، کاربر می‌تواند مدل درختی مربوطه را که بیشتر همانند نقشه‌ای دقیق از مبدأ تا مقصد بوده و جزئیاتی در آن ترسیم شده است را چاپ گرفته و در کنار دست خود قرار دهد و هر بار مشخصات افراد را از ساختار منطق تصمیم‌ساز آن که مشابه حرکت از ریشه درخت به برگ‌های مختلف آن، به عنوان اهداف اصلی است، را عبور داده تا به عدد شاخص ریسک تقلب یا همان درصد ارزیابی صفر تا ۱۰۰ در خصوص متقلبانه بودن ادعای مربوطه برسد.

شاخص ریسک تقلب عددی است بین صفر تا ۱۰۰ درصد که از نظر احتمالی همان عدد بین صفر و یک است و در اصل همانند چراغ راهنمایی رانندگی، ملاک عملی هوشمندانه برای کاربر بیمه بیکاری و شعب اجرایی سازمان خواهد بود چراکه بدین وسیله می‌توان در خصوص پذیرش یک ادعا، رد و یا قطع یک ادعا و یا انجام بررسی‌های دقیق‌تر ادعاهای مطرح شده در خصوص افراد متقاضی بیمه بیکاری تصمیم‌گیری نمود.

در هر بار ارزیابی این روش، نتایج نظری مدل‌ها توسط بازرسان فنی بیمه بیکاری با بررسی‌های دقیق از محل کار و سکونت افراد مدعی بیکاری، مجدد و به صورت حضوری مورد ارزیابی و مقایسه قرار می‌گیرند. به عنوان مثال اگر مدل، شدت ریسک تقلب ادعای فردی را ۷۰٪ ارزیابی نماید و پس از انجام بازرسی‌های حضوری و فنی مشخص شود این پیش‌بینی درست بوده است یا خیر، در هر حالت نتیجه به مدل بازگردانده می‌شود چراکه در حقیقت در این روش، مدل خودبه‌خود تحت آموزش قرار گرفته، به‌روزرسانی شده و پیوسته ارتقاء می‌یابد. به عبارت دیگر با این کار هر بار دانش و هوش مدل نسبت به وضعیت موجود توسعه یافته و بر این اساس همواره مدل پویاتر و کاراتر از قبل و با اعتماد و اتکا بیشتر به‌طور مستقل قابل استفاده است.

بر همین اساس و در جهت اجرای یافته‌های گزارش حاضر، تصمیم گرفته شد به منظور تعیین میزان اعتماد و باورمندی مدل‌های به‌دست آمده، نتایج ارزیابی این مدل‌ها را در دنیای واقعی و با اطلاعات و داده‌های کاملاً جدید مورد آزمایش و ارزیابی قرار

دهیم، لذا ضمن انجام هماهنگی‌های لازم، واحد بیمهٔ بیکاری شعبه سه تأمین اجتماعی استان اصفهان، به دلیل وجود تنوع جامعه آماری افراد بیکار به‌عنوان مأخذ آزمون و پایلوت مدنظر قرار گرفته شد. به‌این‌ترتیب یک نمونه ۲۳ تایی از ادعاهای جدید و جاری انتخاب و سپس با استفاده از فایل اکسلی که به تابع مدل مجهز شده بود نسبت به ارزیابی ریسک تقلب افراد انتخاب‌شده بر اساس نمونه اقدام گردید. به‌طوری‌که نتیجه هر ارزیابی برای هر نفر در زمانی کمتر از پنج تا شش ثانیه مشخص و بلافاصله در مستطیلی جداگانه درج گردیده است. ریز نتایج عملکرد این آزمون در پیوست (۱) ارائه شده‌است. در این ارزیابی از دو مدل شبکه‌های عصبی و درخت تصمیم استفاده‌شده به‌طوری‌که برای هر کدام از ۲۳ نمونه انتخابی، مدل‌ها ارزیابی‌های مختلفی را بین صفر تا ۱۰۰ ارائه نموده‌اند.

از طرف دیگر و بدون هیچ‌گونه اطلاع قبلی، شماره بیمه‌ها در اختیار بازرسان فنی بیمهٔ بیکاری شعبه قرار گرفت تا همان‌طور که بیان شد بدون اطلاع قبلی از نتایج نظری به‌دست‌آمده، نسبت به بررسی استحقاق بیمهٔ بیکاری این افراد از طریق همان روش‌های بازرسی موجود و در اختیار سازمان، نسبت به بررسی‌های لازم اقدام نمایند. به‌این‌ترتیب و پس از گذشت یک فرایند کاری دوهفته‌ای نتایج حاصل از این فرایند با نتایج قبلی به‌دست‌آمده از مدل‌ها مقایسه گردید که نتایج بسیار چشمگیر و درخور توجهی را در برداشت. بیش از ۷۳٪ از پیش‌بینی‌های مدل در خصوص متقلبان بودن ادعاها با نتایج بازرسی حضوری بازرسان سازمان کاملاً منطبق بوده و تنها در ۲۷٪ موارد، نتایج با آنچه در واقعیت موجود است تفاوت دارد که قطعاً با بررسی‌های دقیق‌تر بازرسان، این درصد نیز به مقادیر کمتر از آن خواهد رسید. این نتایج در ستون درستی‌انطباق نتایج ارزیابی مدل با بازرسی فنی به‌وضوح نمایش داده‌شده‌اند.

نکته جالب‌تر در خصوص این بررسی و آزمون مقایسه، نتایج ارزنده دیگری است که در اثر خاصیت توسعه‌ای یا همان خاصیت فراگیر مثبت بودن ارزیابی ریسک تقلب به‌دست‌آمده است. زمانی که بر اساس این نتایج لایه‌های عمیق‌تری را در خصوص مستندات و مدارک ارائه‌شده افراد بررسی می‌کنیم به مصادیق و اعمال مجرمانه‌تر و متقلبانانه‌تر بسیار مدون، ساختاریافته و سازمان‌یافته‌ای دست می‌یابیم که بسیار حیرت‌انگیزند. به‌عنوان مثال در بررسی بیشتر بر روی کارگاه‌هایی که نسبت

به معرفی افراد بیکار اقدام نموده بودند و این شماره بیمه‌ها در آن‌ها ثبت گردیده مشخص شد این شرکت‌ها با ماهیت حقوقی و با کدهای فعالیتی مانند کد ۴۱۰ (شرکت‌های ساختمان‌سازی) و یا کد ۶۱۰ (فعالیت‌های بازرگانی) کاملاً قانونی به ثبت رسمی رسیده‌اند ولی متأسفانه به جای آنکه در ارتباط با موضوع اساس‌نامه خود فعالیت داشته باشند، آزادانه و به شکلی کاملاً سازمان‌یافته بازاریاب‌های متعددی را با حداقل حقوق استخدام و آن‌ها را در سطح شهر و مناطق مختلف پراکنده نموده و افراد را مجدانه و پیگیرانه ترغیب می‌نمایند تا به کمک این شرکت‌ها و با استفاده از انواع اعمال و اقدامات مجرمانه و متقلبانه‌ای که این شرکت‌ها از طریق ایجاد سوابق جعلی و سوری، ادعاهای کذب کارفرمایی و سابقه‌سازی و ... در مقام کارفرمایی ارائه می‌دهند، شرایطی را فراهم می‌آورند که از آن طریق عموم افراد به راحتی می‌توانند نسبت به دریافت مقرری بیکاری، دریافت انواع غرامت از جمله غرامت ایام بارداری، تهیه ادعاهای کذب اشتغال در زمان‌های گذشته و استفاده از ماده ۱۴۸ قانون کار و ایجاد سوابق جعلی و نهایتاً برقراری انواع مستمری و ... اقدام نمایند. از سوی دیگر، این شرکت‌ها بازار بسیار پرسود و پررونقی را برای خود تصاحب نموده‌اند که از طریق اخذ پورسانت‌های کلان در قالب دریافت چند حقوق اول بیمه بیکاری، چند حقوق مستمری و یا حتی دریافت مابه‌التفاوت چند ده‌میلیونی واریز مستمری معوق به حساب خود به‌عنوان حق‌الزحمه، فرایند متقلبانه خود را تکمیل و منابع سازمان را به ناحق در اختیار عموم قرار می‌دهند. به طوری که در همین نمونه ۲۳ تایی کاملاً تصادفی، تعداد نه مورد از آن توسط دو مورد از این شرکت‌ها معرفی شده‌اند.

بدیهی است در صورت جستجو بیشتر در لایه‌های عمیق‌تر مستندات و مدارک ارائه شده آن دسته افرادی که از ریسک بالای تقلب برخوردارند، به‌مواردی به‌مراتب تأسف بارتر از این خواهیم رسید. لذا همواره باید به این موضوع مهم توجه داشت که ارزیابی انواع ریسک‌های مالی از جمله ریسک مالی تقلب در سازمان امری حیاتی بوده که می‌تواند و عاملی اساسی در کنترل و حفظ منابع موجود سایر شاخه‌ها و گروه‌های مزایای بیمه‌ای باشد.

در نهایت همان‌طور که بیان شد روش‌های انجام کار و پیاده‌سازی این گزارش بسیار ساده، روان، کاربردی و از همه مهم‌تر بدون هزینه هستند به طوری که متدولوژی و روش‌های موجود با کمی تغییرات می‌تواند به راحتی در شعب مختلف اجرایی سازمان

مورد استفاده قرار گیرد. بدیهی است هرچقدر این روش‌ها در سطحی گسترده‌تر مورد استفاده قرار گیرند هم‌توان اجرایی و قابلیت این روش‌ها بالاتر می‌روند و هم مقابله هوشمندانه و ریشه‌ای با تقلب و افراد سودجو که در لایه‌های مختلف جامعه مخاطب سازمان پراکنده شده‌اند، قاطعانه‌تر و در حجمی گسترده صورت می‌پذیرد.

بدیهی است علاوه بر متغیرهای استفاده شده در این گزارش، متغیرهای دیگری نیز وجود دارند که به شرط ثبت اطلاعات آن‌ها در تقاضاهای پیشین، می‌توانند با ورود به مدل به تشخیص بهتر تقلب در ادعاهای بیمه بیکاری جدید و جاری کمک نمایند، متغیرهایی مانند سطح تغییرات دستمزد، میزان سواد، تعداد دفعات استفاده از سایر مزایای بیمه‌ای و تعداد افراد معرفی شده به‌عنوان بیکار توسط کارگاه محل اشتغال فرد متقاضی در طول دوره‌های گذشته.

باتوجه به هزینه‌های هنگفتی که سالانه سازمان‌های بیمه‌گر اجتماعی بابت تقاضاهای متقلبانه در خصوص استفاده از تعهدات بلندمدت و کوتاه‌مدت از جمله مزایای بیمه بیکاری متحمل می‌شوند و همچنین مطالعات اندک انجام گرفته در زمینه راه‌های کشف و کاهش تقلب، این گزارش می‌تواند مبنایی علمی جهت کشف تقلب، پیش روی مدیران صنعت بیمه و پژوهشگران علاقه‌مند به مطالعه قرار دهد.

بی‌تردید، نظرات سازنده و حمایت‌های مدیران، متخصصین و خبرگان حوزه صنعت بیمه به‌خصوص همکاران حوزه بیمه‌های اجتماعی می‌تواند در پیشرفت و بهبود هر چه بیشتر این گزارش و موارد مشابه آن مفید و راه‌گشا باشد.



Abagnale, F. (2001). *The Art of the Steal: How to Protect Yourself and Your Business from Fraud*. Transworld Publishers, NSW, Australia.



Ata, A., Ibrahim, H., Seyrek (2009). *The Use of data mining Techniques in Detecting Fraudulent Financial Statements: An Application on Manufacturing Firms, Suleyman Demirel University*. The Journal of Faculty of Economics and Administrative sciences, vol. 14(2), pp. 157-170.

Bust, B., Weinberg, R. (1998). *Using Benfords law and neural networks as a review procedure*. Managerial Auditing Journal 13 (6), pp. 356-366.

Baily, M. (1978). *Some Aspect of Optimal Unemployment Insurance*. Journal of Public Economics, 10, 379-402.

Bolton, R.J., Hand, D.J.(2002), *Statistical fraud detection: a review*. Statistical Science, vol. 17, no. pp. 235-255.

Brockett, P., Xia, X. and Derrig, R. (1998). *Using Kohonen's Self Organising Feature Map to Uncover Bodily Injury Claims Fraud*. Journal of Risk and Insurance, USA.

Derrig, R.A., Ostaszewski, K.M. (1995). *Fuzzy techniques of pattern recognition in risk and claim classification*, The Journal of Risk and Insurance, pp. 447-82.

FBI, Federal Bureau of Investigation, Financial Crimes Report to the Public Fiscal Year, Department of Justice, United States, (2007), http://www.fbi.gov/publications/financial/fcs_report2007/financial_crime_2007.htm.

Hopenhayn, H. and J.P. Nicolini 'Optimal Unemployment Insurance', Journal of Political Economy, 105(2), 412-438.



Sohl, J.E., Venkatachalam, A.R. (1995). *A neural network approach to Forecasting Model Selection*. Information and management 29 (6) 297-303.



Han, J., Kamber, M. (2006). *Data Mining: Concepts and Techniques, Second Edition*. MorganKaufmann Publishers, 2006, pp. 285–464.



Khoavi, R., Frasva, B. (1994). *Useful Feature Subsets and Rough Sets Reducts*. Third international workshop on rough sets and soft computing (RSSC 94).

Lee, R. M. (1993). *Doing research on sensitive topics*. Sage publications, London.

Modrzejewski, M. (1993). *Feature Selection Using Rough Sets Theory*. European Conference on Machin Learning, pp. 213-220.

Shavell, S. and I. Weiss (1979): 'The optimal Payment of Unemployment Insurance Benefits over Time,' *Journal of Political Economy*, 87, pp. 1347 -1362.

Viaene, S., Dedene, G. (2004). *Insurance fraud: issues and challenges*. Geneva Papers on Risk and Insurance Issues and Practice 29 (2), pp. 313–333.

Weisberg, H.I., Derrig, R.A. (1993). *Quantative methods for detecting fraudulent bodily insurance claims* AIB cost Containment/Fraud Filing, pp. 49-82.

Williams, G., Huang, Z. (1997). *Mining the Knowledge Mine: The Hot Spots Methodology for Mining Large Real World Databases*. Lecture Notes in Artificial Intelligence, Springer-Verlag.

Wilson, H.J. (2003). *An analytical approach to detecting insurance fraud*. *Journal of Finance and accountancy*, pp. 1-15.





ردیف	شماره بیمه	نوع تقاضا	مدل ارزیابی	ریسک تقلب شناسایی شده توسط مدل	نتیجه بازرسی بیمه‌بیکاری واحد فنی	درست‌نمایی انطباق نتایج ارزیابی مدل با بازرسی فنی
۱	*****۷۴	جاری	شبکه عصبی	٪۸۵	متقلبانه	صحیح
۲	*****۲۶	جاری	درخت تصمیم	٪۹۰	متقلبانه	صحیح
۳	*****۷۲	جاری	درخت تصمیم	٪۵۶	مشکوک	نیاز به بررسی بیشتر
۴	*****۷۲	جدید	درخت تصمیم	٪۸۵	متقلبانه	صحیح
۵	*****۷۲	جدید	شبکه عصبی	٪۹۶	متقلبانه	صحیح
۶	*****۷۲	جدید	درخت تصمیم	٪۸۲	متقلبانه	صحیح
۷	*****۷۲	جدید	شبکه عصبی	٪۹۳	متقلبانه	صحیح
۸	*****۷۳	جاری	درخت تصمیم	٪۲۵	عادی	صحیح
۹	*****۷۳	جدید	شبکه عصبی	٪۹۷	متقلبانه	صحیح
۱۰	*****۷۳	جاری	درخت تصمیم	٪۱۰	عادی	صحیح
۱۱	*****۷۴	جاری	شبکه عصبی	٪۴۳	مشکوک	نیاز به بررسی بیشتر
۱۲	*****۷۴	جاری	شبکه عصبی	٪۹۱	متقلبانه	صحیح
۱۳	*****۷۴	جاری	شبکه عصبی	٪۸۹	متقلبانه	صحیح
۱۴	*****۷۴	جاری	درخت تصمیم	٪۷۵	متقلبانه	صحیح
۱۵	*****۷۴	جاری	شبکه عصبی	٪۱۸	عادی	صحیح
۱۶	*****۷۴	جاری	درخت تصمیم	٪۳۲	عادی	نیاز به بررسی بیشتر
۱۷	*****۷۴	جاری	شبکه عصبی	٪۷۶	متقلبانه	صحیح
۱۸	*****۷۴	جاری	شبکه عصبی	٪۵۱	مشکوک	نیاز به بررسی بیشتر

Δ جدول ۶. نتایج مقایسه‌ای عملکرد ارزیابی‌ها

ردیف	شماره بیمه	نوع تقاضا	مدل ارزیابی	ریسک تقلب شناسایی شده توسط مدل	نتیجه بازرسی بیمه‌بیکاری واحد فنی	درست‌نمایی انطباق نتایج ارزیابی مدل با بازرسی فنی
۱۹	*****۷۴	جاری	درخت تصمیم	٪۷۳	مشکوک	نیاز به بررسی بیشتر
۲۰	*****۷۴	جاری	شبکه عصبی	٪۸۸	متقلبانه	صحیح
۲۱	*****۷۵	جاری	درخت تصمیم	٪۶۴	مشکوک	نیاز به بررسی بیشتر
۲۲	*****۷۶	جاری	شبکه عصبی	٪۱۵	عادی	صحیح

△ ادامه جدول ۶.

پیوست ۲



الگوریتم هاسدرف

Algorithm FSH (feature selection with Hausdroff distance)

Variables in data;

S = set of all input variables =

- (1) Set $i = 1$; while $i < k + 1$, do
 - (a) Calculate for .
 - (b) $i = i + 1$.
 - (c) Go to 1(a).
- (2) Sort in ascending order, with the corresponding features ().
- (3) Set $j = k$; until stopping criterion is met, do
 - (a) $S = S +$
 - (b) Induce DM model with input S.
 - (c) Evaluate quality DM model.
 - (d) $j = j - 1$
 - (e) Go to 3(a).
- (4) Return final set of features.



الگوریتم شبکه استنباط بیزی در مدل سازی ریسک تقلب

Step	Description
1	<p>Assume that there are two output classes, fraud and legal. Given an instance $X = (x_1, x_2, \dots, x_n)$ with A_1, A_2, \dots, A_n attributes. Maximise the probability of the two classes based on Bayes theorem using.</p> $P(\text{fraud}/X) = \frac{P(X/\text{fraud})P(\text{fraud})}{P(X)} \text{ and}$ $P(\text{legal}/X) = \frac{P(X/\text{legal})P(\text{legal})}{P(X)}$
2	$P(\text{fraud}) = \frac{s_f}{s}$ <p>where s_f is the number of training examples of class fraud and s is the total number of training examples.</p>
3	<p>In order to reduce the computation in evaluating $P(X/\text{fraud})$, the naive assumption of no dependent relationships among the attributes is made. Thus,</p> $P(X/\text{fraud}) = \prod_{i=1}^n P(x_i/\text{fraud})$ <p>The probabilities $P(x_1/\text{fraud}), P(x_2/\text{fraud}), \dots, P(x_n/\text{fraud})$ can be estimated from training examples using.</p> $P(x_i/\text{fraud}) = \frac{s_{fi}}{s_f}$ <p>where s_{fi} is the number of training examples of class fraud having the value x_i for A_i, and s_f is the number of training examples belonging to class fraud.</p>
4	For $P(X/\text{legal})$, the computation is the same as Step 3.
5	Only $P(X/\text{fraud})P(\text{fraud})$ and $P(X/\text{legal})P(\text{legal})$ need to be maximised as $P(X)$ is constant for the two classes. Therefore, the classifier will predict that X belongs to a class which has the highest posterior probability, conditioned on X . That is, $P(\text{fraud}/X)P(\text{fraud}) > P(\text{legal}/X)P(\text{legal})$ or vice versa.

منبع: Δ (1999, p82-89) Witten and Frank, (2006, p296-299) Han and Kamber.

Step	Description
1	Assume that there are two output classes, fraud and legal. The tree starts as a single node N representing the training examples. If the examples' class is all fraud, then the node becomes a leaf and is labelled as fraud. The same is true if class is all legal.
2	If the examples are not of the same class, information gain measure points out the attribute with the highest gain A to separate the examples into individual classes. The expected information needed to classify a given example X is, $I(\text{fraud_examples, legal_examples}) = - \left(\frac{\text{Number_of_fraud_examples}}{\text{Number_of_examples}} \right) \log_2 \left(\frac{\text{Number_of_fraud_examples}}{\text{Number_of_examples}} \right) - \left(\frac{\text{Number_of_legal_examples}}{\text{Number_of_examples}} \right) \log_2 \left(\frac{\text{Number_of_legal_examples}}{\text{Number_of_examples}} \right)$
3	The entropy, or expected information based on the partitioning into subsets by test attribute E , $E(A) = \sum \left\{ \left[\left(\frac{\text{Number_of_test_attribute_fraud_values}}{\text{Number_of_examples}} \right) + \left(\frac{\text{Number_of_test_attribute_legal_values}}{\text{Number_of_examples}} \right) \right] \times \left[I(\text{test_attribute_fraud_values, test_attribute_legal_values}) \right] \right\}$ The smaller the entropy value, the greater the purity of the subset partitions.
4	The expected reduction in entropy caused by knowing the test attribute A is, $\text{Gain}(A) = I - E(A)$ The attribute with the highest gain is the test attribute.
5	A branch is created for each known value of the test attribute, and the examples are partitioned accordingly. The algorithm uses the same process iteratively to form a decision tree for the examples at each partition. Once an attribute has occurred at a node, it need not be considered in any of the node's descendants.
6	The iterative partitioning stops only when one of the following conditions is true: (a) All examples for a given node belong to the same class, or (b) There are no remaining attributes on which examples can be further partitioned. If this is the case, a leaf is created with the majority class in examples. (c) There are no examples for the branch the known value in the test attribute. If this is the case, a leaf is created with the majority class in examples.

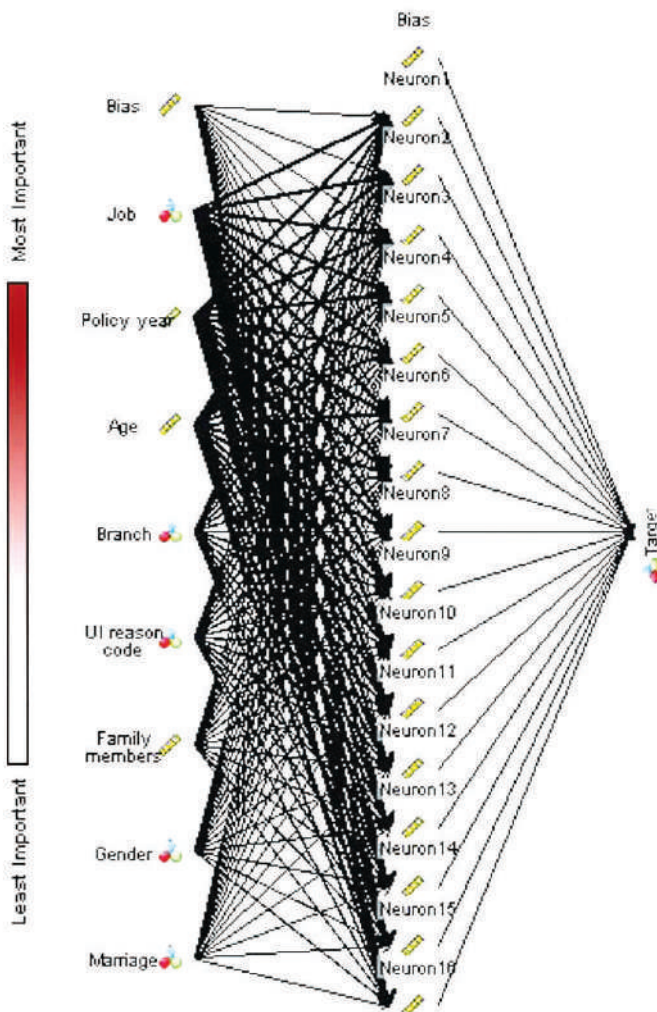
△ منبع: (Han and Kamber (2006, p248-293):

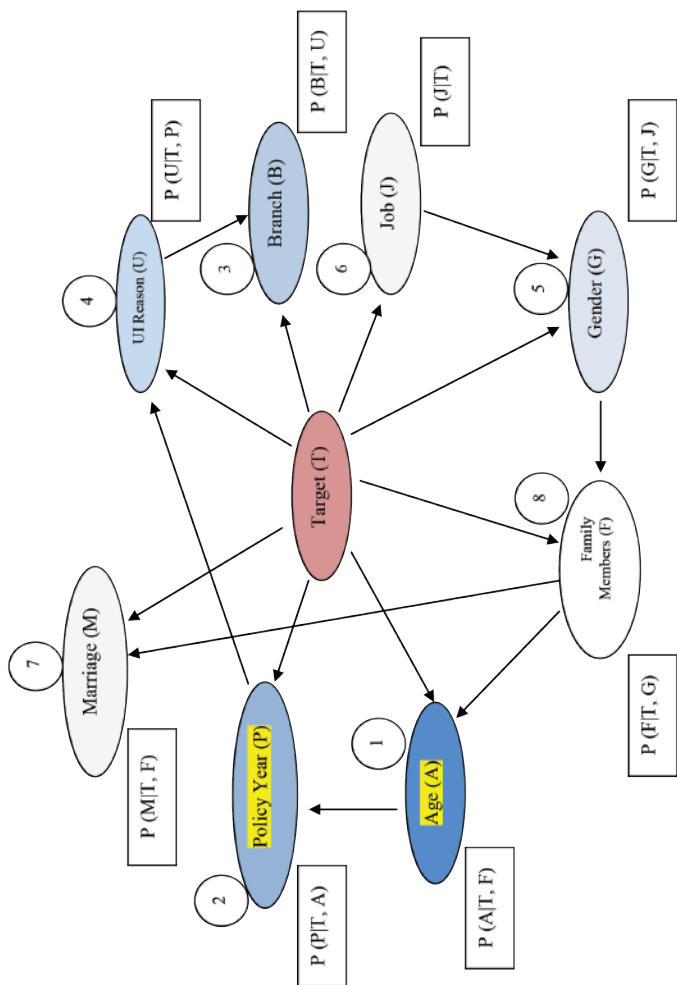
Step	Description
1	Assume that there are two output classes, fraud and legal. Initialise weights in the network to small random numbers between 0 and 1. Present an N dimensional example X to the input layer, including an extra input of -1 for the threshold.
2	Each hidden neuron's net inputs is determined by, $net_j^h = \sum_{i=1}^{N-1} w_{ij} x_i$ where w_{ij} is the weight of the connection from unit i in the input layer to unit j in the hidden layer, x_i is the output value of unit i from input layer. Each hidden neuron's outputs is determined by applying logistic function on net input, $y_j = \frac{1}{1 + e^{-net_j^h}}$ where it maps a large input domain onto a smaller range of 0 to 1.
3	The output neuron's (one output neuron for two classes) net input is determined by, $net_o = \sum_{j=1}^{J-1} v_j y_j$ where v_j is the weight of the connection from unit j in the hidden layer to the output neuron, y_j is the output value of unit j from hidden layer. The output neuron's (one output neuron for two classes) output is determined by applying logistic function on net input, $o_o = \frac{1}{1 + e^{-net_o}}$
4	The error, or learning signal, for the output neuron is, $r_o^o = \lambda(d_o - o_o) \alpha_o(1 - o_o)$ where o_o is the actual output and d_o is the desired output. $\alpha_o(1 - o_o)$ is the derivative of the logistic function The error, or learning signal, for the hidden neurons are, $r_j^h = \lambda(r_o^o v_j) \beta_j(1 - y_j)$ where v_j is the weight of the connection from the output neuron to unit j in the hidden layer. r_o^o is the error of the output neuron.
5	The weights in the output layer are updated with, $v_j(t+1) = v_j(t) + c \lambda(d_o - o_o) \alpha_o(1 - o_o) \beta_j(t)$ The weights in the hidden layer are updated with, $w_{ij}(t+1) = w_{ij}(t) + c \lambda(r_j^h x_i) \beta_j(t)$
6	The training terminates when either: (a) a prespecified number of epochs has expired, or (b) $E \leftarrow E + \sum (r_o^o)^2$ in the previous epoch is smaller than the prespecified threshold. E is the error of each epoch.

منبع: Δ Han and Kamber (2006, p305-310), Smith (1999, p45-51)



مدل شبکه‌های عصبی در ارزیابی ریسک ثقل در بیمه بیکاری





مدل درخت تصمیم در بیمه بیکاری

