پژوهشنامهٔ بیمه/ سال سی و یکم / شمارهٔ ۳/ پاییز ۱۳۹۵/ شمارهٔ مسلسل ۱۲۳/ صفحات ۸۹ تا ۱۱۰
شناسه دیجیتال (DOI): 10.22056/jir.2017.41377

# ارزیابی ریسک تقلب در مزایای بیمهٔ بیکاری سازمان تأمین اجتماعی

**تاریخ دریافت مقاله:** ۱۳۹٤/۱۰/۲۲
**تاریخ پذیرش مقاله:** ۱۳۹۵/۰٦/۱۵

**آتوسا گودرزی**[۱]
**سید جواد طباطبایی‌منش**[۲]

## چکیده

*تقلب در حقوق و مزایای بیمهٔ بیکاری همواره یکی از موضوعات حساس و مورد توجه در حوزهٔ بیمه‌های اجتماعی است که طبق قوانین، جزو جرائم کیفری بوده و قابل پیگیری است. در حال حاضر بهترین روش به منظور ارزیابی تقلب، کنترل آن در همان مراحل اولیهٔ شکل‌گیری و به کمک اطلاعات تقلبهای کشف‌شدهٔ گذشته است. در این مقاله ابتدا مراحل استاندارد کنترل تقلب در تقاضاهای بیمه‌ای مورد بررسی قرار گرفته و سپس با توجه به وجود پایگاه دادهٔ مناسب در خصوص مقرری‌بگیران بیمهٔ بیکاری سازمان تأمین اجتماعی، از دو روش داده‌کاوی شبکه‌های عصبی و درخت تصمیم به منظور یافتن الگوهایی مناسب استفاده شده است که می‌تواند به‌عنوان ابزاری سودمند همراه با کاهش قابل توجه در وقت و هزینه به ارزیابی به موقع تقلب در تقاضاهای بیمهٔ بیکاری کمک کند. در فرایند مطالعهٔ تجربی، این روشها بر روی داده‌های واقعی شامل اطلاعات ۱۵۹۸۳ تقاضای جدید و جاری مقرری بیمهٔ بیکاری آزمایش و کارایی هر روش سنجیده شده است. روش شبکه‌های عصبی با دقت ۸۸ درصد در ارزیابی صحیح متقلبانه یا عادی‌بودن تقاضاها، بهترین کارایی را در مقایسه با روش درخت تصمیم با دقت ۸٤ درصد در برداشته است. بر این اساس، مهم‌ترین متغیرهای مؤثر بر تقاضاهای متقلبانه در روش شبکه‌های عصبی به‌ترتیب، متغیرهای شغل قبلی بیمه‌شده، سابقهٔ پرداخت حق بیمه، سن و در روش درخت تصمیم، متغیرهای محل جغرافیایی شعبه، جنسیت، و تعداد افراد تحت تکفل متقاضی هستند.*

**واژگان کلیدی:** *ریسک تقلب، بیمهٔ بیکاری، داده‌کاوی، شبکه‌های عصبی، درخت تصمیم، تأمین اجتماعی.*

---

۱. استادیار مؤسسهٔ آموزش عالی بیمهٔ اکو، دانشگاه علامه طباطبائی، atousagoodarzi@yahoo.com
۲. کارشناس ارشد آکچوئری سازمان تأمین اجتماعی (نویسندهٔ مسئول)، jtabatabaey@gmail.com

**۱. مقدمه**

بیمهٔ بیکاری، کارگرانی را که به طور 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 به موقع از نیروی کار مستعد و خانوادهٔ ایشان تا زمان یافتن شغلی مناسب است. افراد باید از شرایط قانونی لازم به منظور استفاده از این مزایا برخوردار باشند، کارگر و کارفرما هر دو در خصوص لزوم رعایت قانون در استفاده از مزایای مذکور مسئول هستند. مزایای بیمهٔ بیکاری علاوه بر کمکهای مالی شامل کمکهای درمانی به خانوادهٔ افراد بیکار و همچنین ادامهٔ سوابق بیمه‌ای فرد نیز می‌شود، لذا همواره انگیزهٔ استفادهٔ متقلبانه از مزایای چنین سیستم بیمهٔ بیکاری وجود دارد. تقلب خطری غیرآشکار است و افراد متقلب همواره به دنبال مخفی‌کردن آن هستند (Lee, 1993).

ارزیابی تقلب در تقاضاهای بیمهٔ بیکاری باید با جدیت مورد بررسی و توجه قرار گیرد؛ چون این عمل مجرمانه باعث ایجاد هزینه‌های هنگفت و درنتیجه اتلاف منابع صندوق بیمهٔ بیکاری می‌شود. دریگ[1] (۲۰۰۲)، چهار اصل را برای تعریف تقلب بیان کرده‌اند: تقلب باید عمدی باشد؛ علیه قانون باشد؛ در آن منفعت مالی وجود داشته باشد و اطلاع‌رسانی نادرست نیز اتفاق افتاده باشد. کلاهبرداریهای بیمه‌ای در قوانین بسیاری از کشورهای جهان، جرم کیفری تلقی و در صورت اثبات، مرتکبین آن علاوه بر بازگرداندن وجوه ناشی از کلاهبرداری، محکوم به جریمهٔ مالی و حتی حبس می‌شوند.

تقلبهای مالی به شکل چشمگیری توجه عموم جوامع، رسانه‌ها، سرمایه‌گذاران، جامعهٔ مالی و قانونگذاران را به خود جلب کرده است و دلیل این جلب توجه، به سبب تقلبات کلانی 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

1. Derrig

افزوده است. تقلب، یک دلیل مهم و مؤثر در شکست بسیاری از شرکتها قلمداد می‌شود، زیرا سرمایه‌گذاران، اعتباردهندگان و تحلیل‌گران مالی برای سیاست‌گذاری، به صورتهای مالی متکی هستند و به آنها اعتماد می‌کنند (Ata and Seyrek, 2009).

براساس آخرین تقسیم‌بندی مرکز مطالعات و تحقیقات جرائم مالی ایالات متحده در سال ۲۰۰۷ دسته‌بندی کلی انواع تقلبات مالی مطابق با جدول ۱ ارائه شده است.

**جدول۱. دسته‌بندی کلی انواع تقلبهای مالی**

| تقلب مالی | فعالیتهای متقلبانه |
|---|---|
| تقلب بیمه‌ای | بیمهٔ بیکاری، بیمهٔ سلامت، بیمهٔ عمر و حوادث و ... |
| تقلب بانکی | رهن، پولشویی، اختلاس، خلق پول و... |
| تقلب کالا و اوراق بهادار | اوراق قرضه، قاچاق کالا و ... |
| سایر | تقلبهای شرکتی، تقلب با استفاده از رسانه‌ها و ... |

**منبع: مرکز مطالعات و تحقیقات جرائم مالی ایالات متحده در سال ۲۰۰۷**

در کشورهایی که بیمه دارای ریشه‌های عمیق است، همواره نظارت و کنترل دقیق بر ریسک تقلب حاکم بر تعهدات مالی بیمه‌ها اعم از کوتاه‌مدت و بلندمدت به نحو بارزی دیده می‌شود؛ اما در کشورهایی که بیمه مانند سایر علوم، یک پدیدهٔ وارداتی است، فرایند مذکور از اهمیت قابل ملاحظه‌ای برخوردار است.

کلاهبرداری و تقلب در استفاده از مزایای بیمهٔ بیکاری به روشهای مختلف صورت می‌گیرد، براساس آخرین گزارش تحقیقی دفتر ارزیابی مزایای بیمهٔ بیکاری وزارت کار ایالات متحده در سال ۲۰۱۰ که بر روی جزئیات انواع مختلف تقلب در بیمهٔ بیکاری طی سالهای ۲۰۰۵- ۲۰۰۹ صورت پذیرفته است، انواع تقلب در بیمهٔ بیکاری به شش گروه اصلی طبقه‌بندی می‌شود که در جدول ۲ ارائه شده است.

**جدول۲. جزئیات انواع تقلب در بیمهٔ بیکاری**

| نوع تقلب | درصد ارتکاب نوع تقلب | | | | |
|---|---|---|---|---|---|
| | ۲۰۰۵ | ۲۰۰۶ | ۲۰۰۷ | ۲۰۰۸ | ۲۰۰۹ |
| مخفی‌کردن سایر درآمدهایی که فرد به واسطهٔ اشتغال پنهان دریافت می‌کند، زمانی که فرد مقرری‌بگیر بیمهٔ بیکاری است. | ٦٢/٦٤ | ٥٤/٤٠ | ٦٠/٠٦ | ٦٧/٣٢ | ٦٥/٨٩ |
| عدم جستجو وپیگیری‌هی لازم به منظور یافتن شغل مناسب توسط فرد، در زمانی که فرد مقرری بگیر بیمهٔ بیکاری است. | ٤/٥٥ | ٤/١٥ | ٤/٩٥ | ٣/٠٢ | ٢/٧٥ |
| عدم پذیرش و رد پیشنهادات شغلی مناسب توسط فرد، زمانی که فرد مقرری بگیر بیمهٔ بیکاری است. | ٠/٦٣ | ٠/٣٦ | ٠/٨٠ | ٠/٣٦ | ٠/٧٧ |
| ترک شغل و خروج از کار به طور کاملاً ارادی و با برنامهٔ قبلی و با استفاده از فرمهای قرارداد کار ساختگی در تاریخ و زمانهای مشخص. | ١٢/٧٨ | ١٦/٤١ | ٧/٠٦ | ٥/٠٤ | ٥/١٤ |
| بیکاری به واسطهٔ اخراج از محل کار، به طوری که اخراج ناموجه بوده و به دلایل ساختگی و تبانی با کارفرما برنامه‌ریزی شده باشد. | ٤/٢٧ | ٤/٦٠ | ١٣/٢٩ | ١٢//٦٩ | ٩/٦١ |
| تظاهر به عدم توانایی انجام کار توسط فرد بیکار در خصوص مشاغل پیشنهادی از طرف مراکز معتبر کاریابی. | ٤/٩٤ | ٦/٩٥ | ٤/١٧ | ٤/٦٠ | ٧/٣٨ |
| سایر | ١٠/٢٠ | ١٣/١٤ | ٩/٦٧ | ٦/٩٧ | ٨/٤٦ |
| کل | ١٠٠ | ١٠٠ | ١٠٠ | ١٠٠ | ١٠٠ |

**منبع: گزارش دفتر ارزیابی مزایای بیمهٔ بیکاری، وزارت کار ایالات‌متحده (۲۰۱۰)**

به طور کلی، متقاضیان بیمهٔ بیکاری در دو موقعیت مرتکب تقلب می‌شوند: مورد اول شرایطی است که در آن فرد کاملاً آگاهانه از همان ابتدا به طور برنامه‌ریزی‌شده اقدام به ایجاد سابقه و پرداخت حق بیمه می‌کند و سپس به محض حصول شرایط مناسب موردنظر، نسبت به خروج از حالت اشتغال و تقاضای مقرری بیمهٔ بیکاری اقدام می‌کند. مورد دوم، شرایطی است که در آن، فرد صرفاً به دلیل وجود پشتوانهٔ بیمهٔ بیکاری و محق‌دانستن خود در استفاده از مزایای بیمهٔ بیکاری ناآگاهانه دست به اعمالی می‌زند که موجبات اخراج موجه خود را از شغل موردنظر فراهم می‌کند. وجود بیمه و تعهد به جبران غرامت باعث می‌شود

بیمه‌شدگان ریسکهایی را پذیرش کنند که در صورت عدم وجود بیمه، از این ریسکها دوری می‌جستند (Wilson, 2003). معمولاً مسئولیت بازرسی و تشخیص ادعاها و تقاضاهای متقلبانهٔ بیمه‌ای بر عهدهٔ افرادی است که در اغلب موارد کم-تجربه بوده و آموزشهای ضروری را در خصوص کشف تقلب سپری نکرده‌اند (Doig et al., 1999).

از آنجایی که روند ارزیابی تعهدات مالی ارائه‌شده توسط سازمانهای بیمه‌گر اجتماعی اغلب به صورت ممیزی (بازرسی) بعد از پرداخت[1] صورت می‌پذیرد، لذا بسیاری از تقاضاهای متقلبانه یا شناسایی نمی‌شوند، یا بسیار دیرهنگام و غالباً با اتمام کامل دورهٔ استحقاقی مقرری و پس از پرداخت مبالغ قابل توجهی در قالب تعهدات بیمه‌ای و درمانی به بیمه‌شده، شناسایی می‌شوند.

در این مقاله ابتدا مراحل کنترل تقلب، در مدل رایج در روند ارزیابی تقاضاهای بیمه‌ای مرور شده است؛ در ادامه دو روش شبکه‌های عصبی و درخت تصمیم که از ابزارهای مورد استفاده در حوزهٔ داده‌کاوی[2] هستند، به‌عنوان تکنیکهای معتبر برای ارزیابی شدت متقلبانه‌بودن تقاضاهای بیمهٔ بیکاری معرفی شده‌اند. طبق مستندات تجربی و داده‌های واقعی، مدلهای خودیادگیرنده و پویا براساس این دو روش ارائه خواهد شد، که به کشف متقلبانه‌بودن یک تقاضای جدید یا جاری بیمهٔ بیکاری کمک می‌کنند. برای این منظور اطلاعات گذشته و مستندات ثبت‌شده ۱۵۹۸۳ مقرری‌بگیر بیمهٔ بیکاری، که در ۳۱۱۸ مورد از آنها تقلب صورت‌گرفته، مورد استفاده قرار گرفته است. در بخش پایانی مقاله، نتایج به‌کارگیری این دو روش با یکدیگر مقایسه تطبیقی شده و روش بهینه و دلایل ارجحیت آن مورد بحث و بررسی قرار خواهد گرفت.

1. Post-Payment Audits
2. Data Mining

## ۲. مروری بر ادبیات نظری

از دههٔ ۱۹۷۰ میلادی تاکنون تحقیقات متعددی در زمینهٔ کشف و ارزیابی تقلب در استفاده از مزایای بیمه‌های اجتماعی صورت پذیرفته است. از این میان می‌توان به تحقیق شاول و ویس[۱] (۱۹۷۹) در رابطه با طراحی سیستمهای بهینهٔ پرداخت مقرری در مدت زمان استحقاق بیمهٔ بیکاری، اشاره کرد. بیلی[۲] (۱۹۷۸) تحقیقات مشابهی را با استفاده از روش کالیبراسیون و کنترل بهینه در کشور آمریکا انجام داد. همچنین هوپن‌هاین و نیکولینی[۳] (۱۹۹۷) تحقیقاتی را در این زمینه با استفاده از اطلاعات تقلبهای ثبت‌شدهٔ گذشتهٔ مربوط به بیکاران تحت‌پوشش وزارت کار ایالات‌متحده، انجام داده‌اند. دریگ و استاسزیوسکی[٤] (۱۹۹۵)، ویزبرگ و دریگ[٥] (۱۹۹۳) و براکت[٦] و همکاران (۱۹۹۸) در مقالات خود تکنیکهایی برای شناسایی ادعاهای متقلبانه و دسته‌بندی کلاهبرداریهای بیمه‌ای ارائه داده‌اند. سول و ونکات[۷] (۱۹۹۵)، 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، به طوری که با ترکیب دسته‌گرهای[۸] مختلف و همچنین روشهای بهینه‌سازی مدل از جمله اعتبارسنجی مقطعی و کلی[۹] به یک مدل ایده‌آل در کشف و ارزیابی تقلب دسترسی پیدا کردند.

باست و وینبرگ[۱۰] (۱۹۹۸) شبکه‌های عصبی بازگشتی[۱۱] را برای تمیز صورتهای مالی عادی از متقلبانه به کار بردند. آنها به کمک قانون بنفرد[۱] در تحلیل داده‌ها،

1. Shavell and weiss
2. Baily
3. Hopenhayn and Nicolini
4. Derrig and Ostazewski
5. Weisberg and Derrig
6. Brocket
7. Sohl and Vankat
8. Classifiers
9. Bagging and Cross Validation
10. Bust and Weinberg
11. Back Propagation Neural Networks

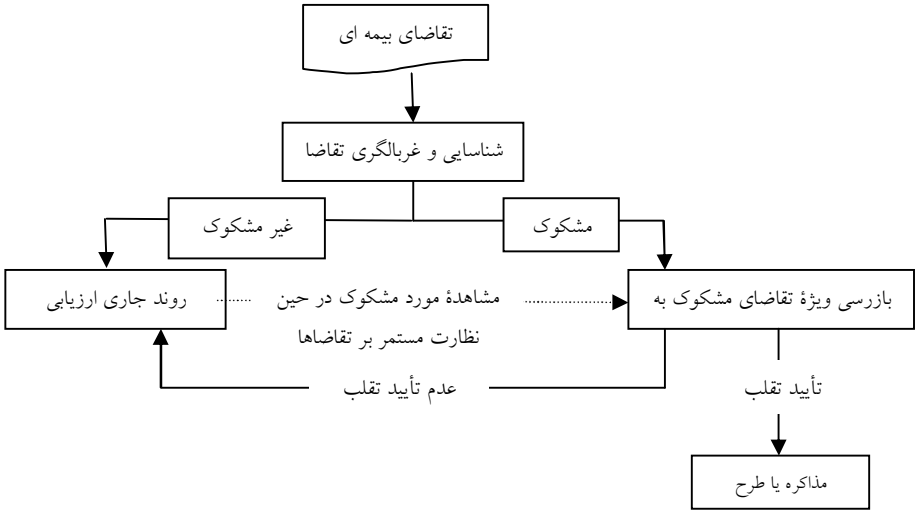
توزیع عددی مبالغ مالی را مورد ارزیابی قرار دادند. با مشخص‌شدن توزیع داده‌ها، با استفاده از قابلیت خاص روش شبکه‌های عصبی در برآوردهای آماری که با پردازش فوق‌العادهٔ به‌هم‌پیوسته و لایه لایه بر روی توزیع داده‌ها صورت می‌پذیرد؛ با تلفیق ساختار الگوریتم نورونهای پردازش لایهٔ درونی شش مدل شبکهٔ عصبی و همچنین افراز کلاس مقادیر ورودی متغیرها به بازهٔ بین صفر و یک، توانستند مدل کارا و بهینه‌ای را در پیش‌بینی و ارزیابی ریسک تقلب ارائه کنند که امکان آموزش و یادگیری مداوم برای مدل فراهم می‌شود و سیستم ارزیابی حاصل از آن مرتباً دانش قبلی‌اش را به روز رسانی می‌کند. نتایج ارزنده و کاربردی این مدل، در ارزیابی و کشف به موقع فرایندهای متقلبانه جریانات مالی چندین هلدینگ نفتی و دارویی در ایالات‌متحده استفاده شده است.

ویلیامز و هووانگ[۲] (۱۹۹۷)، از روشهای داده‌کاوی شبکهٔ بیزی و شبکهٔ عصبی بازگشتی و تکنیک شناسایی داده‌های پرت، مدلی را در ارزیابی ریسک تقلب ارائه کرده‌اند که مبتنی بر پاسخ معین قبلی نبوده و براساس توزیع تجمعی داده‌ها و شناسایی خوشه‌های داده‌های پرت به‌عنوان خوشه‌ای که می‌تواند شامل موارد متقلبانه باشد، اقدام به مدل‌سازی ریسک تقلب کرده‌اند؛ این روش و مدل حاصل از آن در مواردی که اطلاعات کامل و صحیحی از وقوع تقلب در داده‌های مستند قبلی نباشد، بسیار کاراست. از نتایج این مدل در ارزیابی خسارات ناشی از تقلب در تعهدات بیمه‌های درمانی کشور تایوان استفاده شده است. این افراد از روش ماتریس اغتشاش[۳] دقت هر مدل را در پیش‌بینی مقادیر مشاهده‌شدهٔ واقعی، اندازه‌گیری کردند. مدل رایج (شکل ۱) برای کنترل تقلبها، شامل مراحل شناسایی و غربالگری، تحقیق و بررسی، مذاکره با بیمه‌شده یا طرح دعوی است که در روند شناسایی تقاضای جدید و یا جاری مزایای بیمه‌ای اجرا می‌شود.

---

1. Benfords' Law
2. Wiliams and Huang
3 .Confusion Matrix

**شکل ۱. فرایند کنترل کلاهبرداری و تقلب در تقاضاهای بیمه‌ای**




**منبع:** (Viaene and Dedene , 2004)

## ۳. روش‌شناسی کشف تقلب

گام نخست در این خصوص، تعریف واضح مسئله است. این گام بسیار زمانبر است و اساساً شامل تصفیه و دستکاری داده‌ها می‌شود. سپس، یک مدل داده‌کاوی که برای ارزیابی تقلب، مناسب‌تر از بقیه به نظر می‌رسد، انتخاب می‌شود و مورد استفاده قرار می‌گیرد. در این مرحله، عموماً داده‌ها به دو مجموعهٔ آموزش و آزمون (تأیید) تقسیم می‌شوند. داده‌های آموزشی برای استخراج قواعد و روابط و نهایتاً برازش مدل، مورد استفاده قرار می‌گیرند، در حالی که داده‌های تأیید (آزمون) به این منظور استفاده می‌شوند که مشخص شود قواعد استخراج‌شده در یک مجموعهٔ متفاوت از داده‌ها، به چه کیفیتی عمل می‌کنند. بر مبنای نتایج اجرای مدل بر روی همین داده‌ها، عملکرد مدل ارزیابی می‌شود؛ درنتیجه مدل حاصل یا اصلاح می‌شود یا برای حل مسئله به کار گرفته می‌شود.

در ادامه دو روش شبکه‌های عصبی و درخت تصمیم که از ابزارهای رایج و پرکاربرد در کلاس داده‌کاوی تفحصی هستند، به منظور ارزیابی و مدل‌سازی ریسک تقلب انتخاب شده‌اند. در روش شبکه عصبی و درخت تصمیم فرایند ارزیابی و مدل‌سازی براساس پردازش ترکیبی و تلفیقی به‌هم‌پیوسته از سایر

روشهای مرسوم مدل‌سازی آماری از جمله روش مدلهای ساختاری، تحلیل مؤلفه‌ای و انواع مدلهای رگرسیونی صورت می‌پذیرد که به مراتب کاراتر از کاربرد جداگانه هر یک از روشهای آماری مذکور است. اگر چه شباهتهایی بین روش معادلات ساختاری با برخی از انواع روشهای داده‌کاوی به‌خصوص شبکه‌های عصبی دارد، ولی باید توجه داشت که در روش معادلات ساختاری عمدتاً به رابطهٔ بین متغیرهای مستقل و توضیح‌دهندگی آنها در وزنهای ثابت پرداخته می‌شود در صورتی‌که در مدلهای شبکهٔ عصبی و درخت تصمیم تمرکز بر ارتقا توان پیش‌بینی مدل است. با توجه به ماهیت کاربردی تحقیق حاضر، دو روش شبکه عصبی و درخت تصمیم به‌خوبی می‌توانند به کمک ساختار الگوریتمهای چندمنظورهٔ خود، بر پیچیدگیهای درونی و پنهانی داده‌های واقعی فائق آیند، لذا این دو روش انتخاب شدند تا به کمک آن، مدلهایی با هدف شناسایی و دسته‌بندی ادعاهای متقلبانه در بیمهٔ بیکاری، برازش داده شوند.

### ۳-۱. مرور پیشینهٔ تجربی

به منظور طراحی و ساخت یک مدل ریاضی، وجود داده‌هایی از هر دو دستهٔ تقاضاهای متقلبانه و عادی الزامی است. در این بخش با استفاده از اطلاعات ۱۵۹۸۳ مقرری‌بگیر بیمهٔ بیکاری (شامل ادعاهای جدید و جاری) که در سال ۱۳۹۳ در پایگاه دادهٔ مکانیزهٔ سازمان تأمین اجتماعی استان اصفهان ثبت شده‌اند (جدول ۳)، مدلهای دقیقی به منظور شناسایی تقلب در تقاضاهای بیمهٔ بیکاری ارائه شده است.

**جدول۳. توصیف آماری داده‌ها**

| عنوان | شرح | توصیف عددی |
|---|---|---|
| جنسیت | مرد | ۱۲۸۳۸ |
| | زن | ۳۱٤٥ |
| متأهل | مرد | ۱۱٦۳۰ |
| | زن | ۲٥۰۲ |
| مجرد | مرد | ۱۲۰۸ |
| | زن | ٦٤۳ |
| سن | – | بین ۲۵ تا ۵۰ سال |

| عنوان | شرح | توصیف عددی |
|---|---|---|
| **شعبه** | کلیه مناطق تحت پوشش شعب بیمه‌ای استان | ۳۷ |
| **افراد تحت تکفل** | – | بین ۰ تا ٤ نفر |
| **شغل** | عنوان فعالیت کارگاه‌های تحت بررسی | ۹۰۰ عنوان شغلی |
| **سوابق پرداخت حق بیمه** | سنوات کامل پرداخت حق یمه، مشمول قانون | بین ۱ تا ۳۰ سال |
| **پروندهٔ تقلب** | براساس موارد قطعی | ۳۱۱۸ |
| **پروندهٔ عادی** | براساس موارد قطعی | ۱۲۸٦٥ |
| **تعداد کل** | شامل پرونده‌های متقلبانه و عادی | ۱٥۹۸۳ |

**منبع:** آمارهای ثبتی تعهدات کوناه‌مدت، سیستم اتوماسیون بیمه‌ای سازمان تأمین اجتماعی استان اصفهان (سبا)، سال ۱۳۹۳.

مطابق با استاندارد کدینگ جدولهای پایهٔ سیستم اتوماسیون اداری سازمان تأمین اجتماعی، ۱۸ کد مختلف به‌عنوان علل قطع مقرری بیکاری تعریف شده است. طبق نظر کارشناسان و مسئولین بیمهٔ بیکاری تعداد ۸ مورد (٤٥ درصد) از کدها به‌عنوان مصادیق بارز تقلب محسوب می‌شوند و زمانی مورد استفاده قرار می‌گیرند که علت قطع مقرری بیکاری به طور یقین، به واسطهٔ ارتکاب تقلب توسط فرد متقاضی بوده است. بر همین اساس تعداد ۳۱۱۸ پرونده (اطلاعات ثبت‌شدهٔ مقرری‌بگیران)، پرونده‌هایی هستند که وقوع تقلب در آنها توسط کارشناسان ذیربط تشخیص داده شده و علت قطع آن با یکی از ۸ کد مصداق تقلب ثبت شده است. بدیهی است تعداد ۱۲۸٦٥ پروندهٔ باقی‌مانده به‌عنوان تقاضاهای عادی مورد استفاده قرار گرفته‌اند (لازم به ذکر است این امکان وجود دارد که در بین پرونده‌های تقاضاهای عادی نیز موارد متقلبانه‌ای وجود داشته باشد که به هر دلیلی شناسایی نشده‌اند).

نکتهٔ حائز اهمیت این است که در این مطالعه به منظور ارتقاء سطح کیفی فرایند تحلیل داده در برازش بهترین مدل شناسایی تقلب، در طول دورهٔ موردنظر (سال ۱۳۹۳)، تعداد دفعات ورود و خروج افراد به صندوق بیمهٔ بیکاری همراه با تغییرات ناشی از آن مورد توجه قرار گرفته است. سوابق پرداخت حق بیمهٔ افراد انتخاب شده بین ۱ تا ۲۷ سال متغیر بوده است و برخی از این افراد در طول سوابق پرداخت حق بیمه‌شان ۱ تا ٤ بار از بیمهٔ بیکاری استفاده کرده‌اند. سن افراد بین ۲٥ تا ٥۰ سال قرار دارد که برحسب زن، مرد، مجرد و متأهل طبقه‌بندی شده‌اند. تعداد

افراد تحت تکفل افراد متأهل نیز به طور استاندارد بین ۰ تا ٤ نفر در نظر گرفته شده است.

**۳-۲. متغیرهای مدل**

در هر یک از دو مدلی که در این بخش برای ارزیابی تقلب معرفی خواهند شد، متقلبانه یا عادی‌بودن یک تقاضای بیمهٔ بیکاری، به‌عنوان متغیر وابسته (پاسخ) در نظر گرفته می‌شود. مقدار ۱ برای متغیر وابسته به معنای متقلبانه‌بودن تقاضا و مقدار۰ به معنای عادی‌بودن آن تقاضاست. در این مطالعه، فرایند شناسایی تقلب با استفاده از ۹ متغیر مستقل صورت گرفته است. برای انتخاب متغیرها از روش فاصله‌ای هاسدورف[1] استفاده شده است. روش فاصله‌ای هاسدرف ابتدا مجموعه‌هایی بسته و غیرتهی از کلاس متغیرهای مستقل و کلاس متغیر پاسخ در یک فضای متریک ایجاد می‌کند، سپس با درنظرگرفتن مکان هر متغیر در فضای متریک، شدت رابطهٔ هر متغیر از یک مجموعه را به‌عنوان نقطه‌ای از یک فضای متریک با سایر متغیرها در مجموعه‌های دیگر به‌عنوان سایر نقاط فضای متریک اندازه‌گیری می‌کند و با تکرار این فرایند به دفعات زیاد مطابق با الگوریتم هاسدرف که کدنویسی آن نیز در نرم‌افزار آماری R انجام شده است، مناسب‌ترین متغیرها را به منظور استفاده در عملیات داده‌کاوی ارائه می‌کند. الگوریتم هاسدورف بر این اساس که در توپولوژی، فضایی است که در آن بتوان نقاط را با همسایگی جدا کرد، شکل گرفته است؛ 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، اگر بتوان هر دو نقطه در X را توسط همسایگیهایی از هم جدا کرد، فضای X هاسدورف است. در الگوریتم هاسدورف به کار رفته در این مطالعه، تعداد متغیرهای مرتبط با مجموع داده‌ها با k نشان داده شده است. در گام (۱) فضای

1. Hausdorff Distance Feature Selection

متریک هاسدورف برای $v_i$ها در دو همسایگی $x_1$ و $x_2$ با متر $H_i$ برای تمام $i$ها و k متغیر محاسبه می‌شود؛ در گام (۲) $H_i$های محاسبه‌شده براساس مجموعه‌های تهی $(s_1,\dots,s_k)$ نظیربه‌نظیر مرتب می‌شوند (تعداد kها و $H_i$های محاسبه‌شده با هم برابرند)، در این گام شرط اساسی تهی‌بودن همسایگیها با مقایسه تناظر $H_i(.,.)$ و $S_i$ها کنترل می‌شود تا مشخص شود کدام $v_i$ها در فضای هاسدورف قرار می‌گیرد؛ در گام (۳) و ۳(a) هر کدام از $v_i$هایی که طی فرایند قبلی، انتخاب می‌شود، جایگزین مقادیر تهی $S_i$ها می‌شود؛ در مراحل ۳(b) و ۳(c) سطح اطمینان براساس مدل انتخابی که قرار است متغیرها در آن به کار برده شوند، تعیین می‌شود؛ مراحل ۳(d) و ۳(e) به منظور تکرار فرایند گام ۳ تا تکمیل آخرین اندیس و متغیر، به کار می‌روند و نهایتاً گام (٤) متغیرهای نهایی و قابل استفاده را ارائه می‌دهد.

**جدول ٤. متغیرهای استفاده‌شده در مدل‌سازی ریسک تقلب**

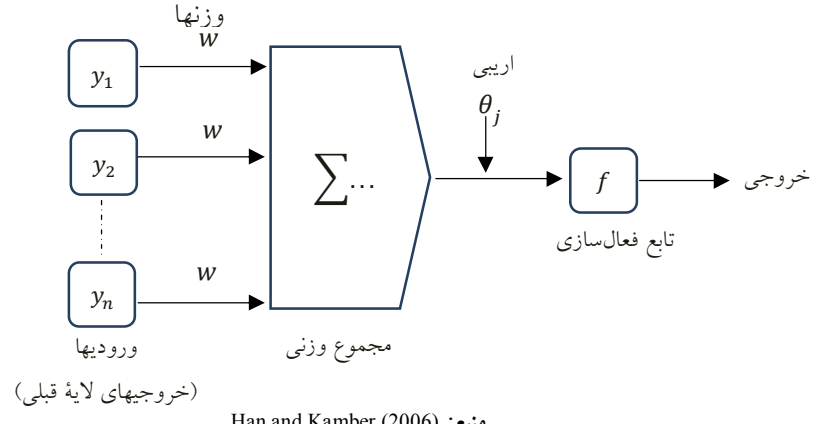
| متغیر | نام | نوع | سطوح اندازه‌گیری | نقش در مدل |
|---|---|---|---|---|
| X1 | جنس | گسسته | دوحالتی | توضیحی- (مستقل) |
| X2 | وضعیت تأهل | گسسته | دوحالتی | توضیحی- (مستقل) |
| X3 | سن | پیوسته | فاصله‌ای | توضیحی- (مستقل) |
| X4 | تعداد تبعی | گسسته | فاصله‌ای | توضیحی- (مستقل) |
| X5 | سوابق پرداخت حق بیمه | پیوسته | فاصله‌ای | توضیحی- (مستقل) |
| X6 | شعبه | گسسته | رده‌ای- اسمی | توضیحی- (مستقل) |
| X7 | علت بیکاری | گسسته | رده ای- اسمی | توضیحی- (مستقل) |
| X8 | تعداد تکرار مقرری | گسسته | فاصله ای | توضیحی- (مستقل) |
| X9 | شغل | گسسته | رده ای- اسمی | توضیحی- (مستقل) |
| **Y** | شناسایی | گسسته | دوحالتی | پاسخ- (وابسته) |

جدول ٤ انواع متغیرهای استفاده‌شده در مدل‌سازی ریسک تقلب را همراه با توضیحات مختصری در خصوص ساختار هر کدام ارائه می‌کند.

### ۳-۳. شبکه‌های عصبی مصنوعی[1]

شبکهٔ عصبی روشی است که با استفاده از گره‌های به هم مرتبط، از فرایند کارکرد مغز انسان تقلید می‌کند. این سیستم از شمار زیادی عناصر پردازش فوق‌العاده به‌هم پیوسته با نام نورون تشکیل شده است که برای حل یک مسئلهٔ هماهنگ با یکدیگر عمل می‌کنند. شکل ۲ یک واحد پردازش نورون که مجموع وزنهای سیگنال ورودی را محاسبه می‌کند، نشان می‌دهد.

**شکل۲. یک واحد پردازش نورون**




**منبع:** Han and Kamber (2006)

شبکه‌های عصبی مصنوعی با توجه به داده‌های موجود در تقاضاهای بیمهٔ بیکاری، امکان پیش‌بینی ارتکاب تقلب را فراهم می‌آورند. شبکه‌های عصبی مصنوعی نیازی به روزآمدی ثابت ندارد اما با تحلیل اطلاعات خاصی، امکان یادگیری مداوم را فراهم می‌کند. در این روش هر تقاضای ایجادشده مورد تحلیل قرار گرفته و یک برآورد آماری مبنی بر اینکه آیا داده‌ها با برون‌دادهای مطلوب مطابقت داشته‌اند یا خیر، ارائه می‌شود. آموزش سیستم با استفاده از شواهد تجربی و مستندات مربوط به تقلبهایی که از قبل در سیستم وجود داشته (به منظور کشف کلاهبرداری) به انجام می‌رسد. به محض تحقق این امر، سیستم از دانش قبلی برای کشف و تشخیص متقلبانه‌بودن تقاضاها، استفاده می‌کند. از این سیستم می‌توان هم

1. Artificial Neural Network

در کشف تقلبهای قبل از پرداخت و هم بعد از پرداخت مقرری بیمهٔ بیکاری استفاده کرد. راهکارهایی که این رویکرد در کشف تقلب از آن استفاده می‌کند، شامل نمایه‌سازی داده‌ها[1]، مدل تحلیل پیشرفته[2] و رده‌بندی امتیازی[3] است.

### ۳-٤. درخت تصمیم[4]

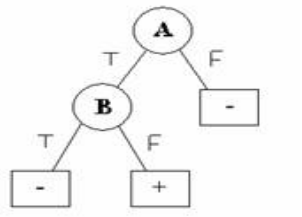
درخت تصمیم یکی از ابزارهای متداول داده‌کاوی است که در رده‌بندی داده‌های کیفی، مورد استفاده قرار می‌گیرد. یک درخت معمولاً از ریشه، شاخه‌ها، گره‌ها[5] (جایی که شاخه‌ها منشعب می‌شوند و در سه نوع کلی شامل گرهٔ تصمیم یا تعیین هدف، گرهٔ تصادفی یا عملیاتی و گرهٔ پایانی یا نتیجهٔ نهایی)، و برگها تشکیل شده است. درختهای تصمیم، به صورت مشابه از گره‌ها که با دایره و شاخه‌ها که با پاره-خطهای اتصال بین گره‌ها نشان داده می‌شوند، تشکیل شده‌اند. درخت تصمیم را به منظور سادگی در رسم، معمولاً از چپ به راست یا از بالا به پایین رسم می‌کنند به طوری که ریشه در بالا قرار بگیرد. گرهٔ اول را ریشه می‌گویند. انتهای یک زنجیره "ریشه – شاخه – گره –... گره" را یک "برگ" می‌نامند.

از هر یک از گره‌های داخلی (یعنی هر گره‌ای که برگ نباشد)، دو یا چند شاخهٔ دیگر می‌توانند منشعب شوند. هر گره مربوط به یک خصوصیت معین است و شاخه‌ها به معنای بازه‌ای از مقادیر هستند. این بازه‌های مقادیر باید بخشهای مختلف مجموعهٔ مقادیر معلوم برای خصوصیتها را به دست دهند. توسط درختهای تصمیم می‌توان خصوصیات کمی و خصوصیات کیفی را به طور همزمان پردازش کرد. بنابراین درخت تصمیم نشان‌دهندهٔ یک مدل منطقی از پدیدهٔ موردتحقیق است. هنگامی که دقیقاً دو شاخه از یک گره داخلی منشعب شوند (چنین درختی را درخت دوحالته[6] می‌گویند)، همان‌طور که در شکل ۳ نشان داده شده است، هر یک

1. Data Profiling
2. Advanced Analytical Model
3. Rank Scoring
4. Decision Tree
5. Node
6. Binary

از این دو شاخه می‌تواند نمایندهٔ یک عبارت درست یا غلط برحسب خصوصیات معلوم باشد.

**شکل۳. درخت تصمیم دوحالته**




با ایجاد درخت تصمیم براساس داده‌های پیشین که ردهٔ آنها معلوم شده است، می‌توان داده‌های جدید را دسته‌بندی و پیش‌بینی کرد. هدف از ایجاد درخت تصمیم در این تحقیق، طبقه‌بندی اطلاعات با استفاده از مستندات گذشتهٔ مقرری‌بگیران به منظور تصمیم‌گیری در مورد متقلبانه یا عادی‌بودن ادعاهای جدید و یا جاری بیمهٔ بیکاری است. معیارهای مختلفی برای تعیین صفتی که خردکردن داده‌ها باید براساس آن انجام شود، وجود دارد که از آن جمله می‌توان به معیارهای بهره اطلاعاتی[1]، نسبت بهره[2] و شاخص جینی[3] اشاره کرد.

## ٤. مدل‌سازی و تحلیل داده‌ها

به منظور مدل‌سازی، همهٔ اطلاعات گذشته ادعاهای بیمهٔ بیکاری شامل متقلبانه و عادی در یک فایل تلفیق می‌شوند، به طوری که برای برنامه قابل فهم باشد. در این مدل، داده‌ها با اطلاعات منابع مختلف مقایسه می‌شوند تا سرانجام الگویی برای کلاهبرداری یا سوء استفاده از مزایای بیمهٔ کاری تعیین شود.

تمامی ابزارهای داده کاوی استفاده‌شده در این تحقیق به منظور مدل‌سازی و تصمیم‌گیری به کمک نرم‌افزار IBM Spss Modeler v.14 صورت پذیرفته است.

1 .Information Gain
2 .Gain Ratio
3 .Gini Index

### ٤-۱. مدل شبکه‌های عصبی

در مواردی که هنوز رابطهٔ بین متغیرها مشخص نشده باشد و امکان تعیین رابطهٔ بین متغیرها از طریق حدس امکان‌پذیر نباشد، رویکرد نمایه‌سازی داده‌ها در مدل‌سازی شبکه‌های عصبی مصنوعی به‌خوبی پاسخگو بوده و می‌تواند روابط بین ورودیها و خروجیهای مختلف را شناسایی کند. رویکرد نمایه‌سازی تکنیکی است که داده‌هایی را که به منظور اهداف خاص محاسباتی از روی پایگاه داده‌های بزرگ استخراج می‌شوند را از لحاظ میزان دقت، کامل‌بودن و یکپارچگی مورد ارزیابی قرار داده و با ارائهٔ اطلاعات و آمارهای دقیق، قلمرو داده‌هایی را که موجب بروز خطا و اشتباه در سازماندهی داده‌ها خواهند شد را مشخص می کند. کاربرد و توسعهٔ این تکنیک، به طور قابل توجهی کیفیت داده‌ها را افزایش می‌دهد. این تکنیک به طور گسترده در خصوص آن دسته از آمارهای ثبتی که از پایگاه داده‌های بزرگ استخراج می‌شوند، مورد استفاده قرار می‌گیرد. در این شرایط، می‌توان موارد جدید را وارد و رفتار مدل را پیش‌بینی کرد. هنگامی که مناسب‌ترین مدل براساس این روش برازش داده شد، با ورود اطلاعات جدید در مدل، مقدار متغیر خروجی که همان شدت متقلبانه‌بودن یک تقاضاست به صورت درصدی بین ۰ تا ۱۰۰ ارزیابی می‌شود.

با اعمال مدل شبکه عصبی برازش‌داده‌شده بر روی آن بخش از اطلاعات و مستندات تجربی مقرری‌بگیران که در مراحل قبل به منظور آزمون دقت مدل انتخاب شده بودند، جدول ٦ در این خصوص تنظیم شده است. همان‌طور که دراین جدول مشاهده می‌شود، دقت مدل در شناسایی تقاضاهای متقلبانه ۸۹ درصد و در شناسایی تقاضاهای عادی ۸۷ درصد بوده است. ضمن اینکه دقت مدل در شناسایی صحیح متقلبانه یا عادی‌بودن هر تقاضا برابر ۸۸ درصد است.

**جدول ٦. دقت مدل شبکه‌های عصبی در ارزیابی تقاضاهای بیمهٔ بیکاری**

| ماتریس ارزیابی | | وضعیت برآوردشدهٔ تقاضاها | | درصد صحیح |
|---|---|---|---|---|
| | | عادی | متقلبانه | |
| وضعیت مشاهده‌شدهٔ تقاضاها | عادی | ۱۹٤۷ | ۲۹۰ | ۸۷ |
| | متقلبانه | ۱۰٦ | ۸٥٤ | ۸۹ |
| کل | | | | ۸۸ |

### ٤-۲. مدل درخت تصمیم

در این تحقیق با کمک نرم‌افزار IBM Spss Modeler v.14 و با استفاده از معیار بهرهٔ اطلاعاتی، درخت تصمیم‌گیری رسم شده است، معیار بهرهٔ اطلاعاتی مشخص می‌کند که هر ویژگی یا متغیر، علاوه بر مقدار مشخصهٔ خود، تا چه میزان اطلاعات منظم و ساختاریافته‌ای را در ارتباط با سایر ویژگیها و متغیرها در خود جای داده است؛ به عبارتی، یک ویژگی تا چه اندازه می‌تواند از اطلاعات نهفته در سایر ویژگیها بهره‌مند باشد. هر چه میزان بهرهٔ اطلاعاتی یک ویژگی بالاتر باشد میزان آنتروپی آن کمتر است. الگوریتم ساختاری درخت تصمیم براساس میزان بهرهٔ اطلاعاتی، داده‌ها را خرد و گره‌سازی می‌کند.

با توجه به درخت تدوین‌شده، متغیر اطلاعات جغرافیایی شعب سازمان تأمین اجتماعی به‌عنوان اولین و مهم‌ترین عامل در بررسی متقلبانه یا عادی‌بودن تقاضاهای بیمهٔ بیکاری در نظر گرفته شده است. این متغیر در ریشهٔ درخت به دو شاخهٔ کلی شامل کد شعب تقسیم می‌شود. در ابتدا شاید این انتخاب ابهام‌آمیز به نظر برسد، ولی باید به این نکته توجه داشت که ساختار فرهنگی، اقتصادی و اجتماعی جمعیت تحت پوشش هر شعبه، نقش بسزایی در افزایش یا کاهش ارتکاب اعمال مجرمانه از قبیل تقلب در مزایای بیمه‌های اجتماعی به دنبال خواهد داشت. آنچه که مشخص است در برگ هر شاخه، شدت (درصد) متقلبانه‌بودن تقاضا مشخص شده است. در سطح دوم درخت، پارامتر سوابق پرداخت حق بیمهٔ متقاضی قرار دارد، که هر یک براساس مقادیر کمتر و بیشتر از ۹ سال در شاخهٔ سمت چپ و مقادیر بیشتر و کمتر از ۱۳ سال در شاخهٔ سمت راست تقسیم می‌شوند و به همین ترتیب درخت تصمیم تا سه سطح ادامه می‌یابد. براساس درخت حاصل می‌توان قوانینی به صورت اگر– آنگاه به شرح زیر استخراج کرد:

– اگر تقاضای مقرری در شعبه ۱۹۲۰ باشد و سابقهٔ پرداخت حق بیمهٔ فرد متقاضی بیشتر از ۹ سال باشد و سن فرد کمتر از ۳۲ سال باشد، آنگاه احتمال متقلبانه‌بودن تقاضای مربوطه ۹۰ درصد است.

– سایر قوانین نیز به همین ترتیب قابل استخراج است.

در ادامه با اعمال این مدل بر روی آن دسته از اطلاعات و مستندات مقرری-بگیران که به‌عنوان داده‌های آزمون در نظر گرفته شده‌اند، جدول ۷ به منظور بررسی دقت مدل استخراج شد.

**جدول ۷. دقت مدل درخت تصمیم در ارزیابی تقاضاهای بیمهٔ بیکاری**

| ماتریس ارزیابی | | وضعیت برآوردشدهٔ تقاضاها | | درصد صحیح |
|---|---|---|---|---|
| | | عادی | متقلبانه | |
| وضعیت مشاهده‌شدهٔ تقاضاها | عادی | ۱۸۵۷ | ۳۸۰ | ۸۳ |
| | متقلبانه | ۱۵۹ | ۹۰۱ | ۸۵ |
| کل | | | | ۸٤ |

همان‌طور که در این جدول مشاهده می‌شود، دقت مدل در شناسایی تقاضاهای متقلبانه ۸۵ درصد و در شناسایی تقاضاهای عادی ۸۳ درصد بوده است. ضمن اینکه دقت مدل در شناسایی صحیح متقلبانه یا عادی‌بودن هر تقاضا برابر ۸٤ درصد است.

## ۵. نتیجه‌گیری و پیشنهادها

در این مقاله دو روش داده‌کاوی شبکه‌های عصبی و درخت تصمیم برای ساخت مدلهایی برای شناسایی و ارزیابی تقاضاهای متقلبانه بیمهٔ بیکاری معرفی شدند. این روشها بر روی داده‌های واقعی که به طور کاملاً تصادفی انتخاب شده بودند، آزمایش و کارایی هر روش مورد سنجش قرار گرفت. مدل شبکه‌های عصبی با دقت ۹٦ درصد در ارزیابی صحیح متقلبانه یا عادی‌بودن تقاضاهای بیمهٔ بیکاری جدید و جاری بهترین کارایی را در مقایسه با روش درخت تصمیم با دقت کلی ۸۸ درصد، ارائه کرده است. البته باید به این نکته توجه داشت که در روش شبکه‌های عصبی تمامی متغیرها به طور مستقیم در مدل حضور دارند. این در حالی است که در مدل درخت تصمیم صرفاً از چهار متغیر اصلی (و همچنین سایر متغیرهایی که بیشترین وابستگی را با متغیر وابسته دارند) استفاده شده است. مدل درخت تصمیم با توجه به روش شهودی آن در زمانهایی که نیاز به سرعت در تصمیم‌گیری وجود داشته باشد، قابلیت بهتری را نشان

می‌دهند. علاوه بر متغیرهای استفاده‌شده در این تحقیق، متغیرهای دیگری نیز وجود دارند که به شرط ثبت اطلاعات آنها در تقاضاهای پیشین، می‌توانند با ورود به مدل به تشخیص بهتر تقلب در تقاضاهای مقرری بیمهٔ بیکاری کمک کنند. برای مثال، می‌توان به سطح تغییرات دستمزد، میزان سواد، تعداد دفعات استفاده از سایر مزایای بیمه‌ای و تعداد افراد معرفی‌شده به‌عنوان بیکار توسط کارگاه محل اشتغال فرد متقاضی در طول دوره‌های گذشته، اشاره کرد.

با توجه به هزینه‌های هنگفتی که سالانه سازمانهای بیمه‌گر اجتماعی بابت تقاضاهای متقلبانه در خصوص استفاده از تعهدات بلندمدت و کوتاه‌مدت از جمله مزایای بیمهٔ بیکاری متحمل می‌شوند و همچنین مطالعات اندک انجام‌گرفته در زمینهٔ راه‌های کشف و کاهش تقلب، این تحقیق می‌تواند مبنایی علمی برای کشف تقلب، پیش روی مدیران صنعت بیمه و پژوهشگران علاقه‌مند به این رشته از مطالعات بگذارد.

باتوجه به دقت مناسبی که این روشها در شناسایی صحیح متقلبانه یا عادی‌بودن تقاضاهای بیمهٔ بیکاری دارند، موارد ذیل در این خصوص پیشنهاد می‌شود:

– در صفحهٔ اصلی ثبت اطلاعات تقاضای مقرری‌بگیران بیمهٔ بیکاری در سیستم اتوماسیون شعب سازمان تأمین اجتماعی، یک سطر یا منو یا یک کلید تابعی با عنوان ارزیابی ریسک تقلب تقاضا تعبیه شود. در این صورت همزمان با ثبت اطلاعات فرد متقاضی توسط کارشناس بیمهٔ بیکاری، ریسک تقلب آن تقاضا توسط مدلهای ارزیابی، به صورت عددی بین ۰ تا ۱۰۰ ارزیابی می‌شود.

– با توجه به اینکه به کمک مدلهای ارزیابی، شدت ریسک تقلب در تقاضاهای بیمهٔ بیکاری در همان مراحل اولیه ثبت تقاضا به صورت برخط شناسایی می‌شود، لذا می‌توان با منطقی و هدفمندساختن فرایند انجام بازرسیهای فنی بیمهٔ بیکاری واحد امور بیمه‌شدگان سازمان تأمین اجتماعی و همچنین از طرف دیگر با بررسیهای مجدد و دقیق‌تر مدارک و مستندات ارائه‌شدهٔ فرد متقاضی و احیاناً ارجاع آن به واحدهای مرتبط، در خصوص آن دسته از تقاضاهایی که مشکوک به تقلب

هستند، اقدامات به‌موقع و پیشگیرانهٔ لازم را به منظور کنترل و مقابله جدی با ارتکاب تقلب انجام داد و از هدررفتن هنگفت منابع مالی و تعهدات ایجادشده ناشی از آن در سطحی گسترده جلوگیری به عمل آورد.

– روشها و تکنیکهایی که در این تحقیق به منظور کشف و ارزیابی تقلب مورد استفاده قرار گرفتند را می‌توان در سایر شاخه‌های مزایای بیمه‌های اجتماعی از جمله تعهدات بلندمدت مستمری‌بگیران، تعهدات کوتاه‌مدت ایام بارداری، حوزهٔ بیمهٔ سلامت و درمان و حتی متقاضیان استفاده از طرحهای بیمه‌ای خاص و تخفیف‌دار نیز مورد بهره برداری قرار داد.

– با توجه به دخالت عامل انسانی در پدیدهٔ تقلب در مزایای بیمه‌ها، بررسیهای بیشتر و استفاده از نظر خبرگان و متخصصان بیمه، به کاربردی‌ترشدن نتایج این تحقیق خواهد انجامید.

## منابع

1. Ata, A.H., Seyrek, I.H., 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*, 14(2), pp. 157-170.

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

3. Baily, M., 1978. Some aspect of optimal unemployment insurance. *Journal of Public Economics*, 10, pp.379 -402.

4. 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*, pp.245-274.

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

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

7. Doig, A., Jones, B. and Wait, B., 1999. The insurance industry response to fraud. *Security Journal*, 12, pp.19 -30.

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

9. Hopenhayn, H., and Nicolini J.P., 1997: Optimal unemployment Insurance. *Journal of Political Economy*, 105(2), pp.412 -438.

10. Sohl, J.E. and Venkatachalam, A.R., 1995. A neural network approach to forecasting model selection. *Information & Management*, 29(6), pp.297-303.

11. Han, J. and Kamber, M., 2006. *Data mining: concepts and techniques*, Second ed., Morgan Kaufmann Publishers, pp.285–464.

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

13. Shavell, S. and Weiss, I., 1979. The optimal payment of unemployment insurance benefits over time. *Journal of Political Economy*, 87, pp.1347 -1362.

14. U.S. Department of Labor, 2010. Benefit Accuracy Measurement program (BAM), http://www.ows.doleta.gov/unemploy/bqc.asp

15. Viaene, S. and Dedene, G., 2004. Insurance fraud: issues and challenges. *Geneva Papers on Risk and Insurance and Practice*, 29, pp.313-33.

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

17. Williams, G. and Huang, Z., 1997. Mining the knowledge mine: the hot spots methodology for mining large real world databases. *Lecture Notes in Artificial Intelligence*, Springer-Verlag.

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