فصلنامه مطالعات کمی در مدیریت

دوره ۱۳، شماره دو، تابستان ۱۴۰۱، صص ۱۶۹-۲۲۳

ارزیابی و کشف تقلب در فرایند زنجیره تامین مبتنی بر رویکرد تلفیقی ANN-Big data

الناز علیخانی زنجانی[۱]، فرید عسکری[۲]، امیر نجفی[۳] و بابک حاجی‌کریمی[۴]

تاریخ دریافت: ۱۴۰۱/۴/۹ و تاریخ پذیرش: ۱۴۰۱/۶/۵

چکیده

در دهه‌های اخیر، رقابت جهت ارائه ارزش برتر به مشتریان، از رقابت میان بنگاه‌ها به سمت رقابت میان زنجیره‌های تامین سوق پیدا کرده است. طراحی مناسب زنجیره تامین با توجه به ابعاد اقتصادی، اجتماعی و زیست‌محیطی در سطوح استراتژیک، تاکتیکی و عملیاتی ضامن بقا و توسعه پایدار بنگاه‌های فعال در هر بخش از زنجیره تامین است. یکی از مسائل مهم در زنجیره تامین، وجود تقلب و مخاطرات مرتبط با آن‌ها در سرتاسر زنجیره است. همچنین با توجه به حجم پژوهش‌های علمی و تجربی صورت گرفته در سال‌های اخیر، مساله کشف و ارزیابی تقلب با استفاده از روش‌های محاسباتی به تنهایی یک موضوع با اهمیت برای پژوهش می‌باشد. از مهم‌ترین چالش‌های کشف و ارزیابی تقلب می‌توان به در دسترس نبودن مجموعه‌های داده‌ای واقعی، وجود مجموعه‌های داده‌ای نامتقارن، عظیم بودن اندازه مجموعه‌های داده‌ای، رفتار پویای متقلب و پراکندگی رخدادهای تقلب‌آمیز اشاره نمود که نتیجه این دو عدم اطمینان و ابهام در تصمیم‌سازی است. لذا هدف این پژوهش، ارائه مدلی جهت کشف و ارزیابی تقلب در فرایند زنجیره تامین مبتنی بر رویکرد تلفیقی ANN-Big data می‌باشد. این پژوهش از نظر هدف، یک پژوهش توسعه‌ای و کاربردی است. نخست به دلیل ایجاد یک مدل جامع، با در نظر گرفتن منابع مختلف اخذ تصمیم در خصوص قانونی یا تقلب‌آمیز بودن تراکنش‌های کارت الکترونیک و الزامات تحلیل بزرگ داده، سعی به توسعه دانشی نظری در این حوزه دارد، این پژوهش را می‌توان از لحاظ ماهیت یک پژوهش تحلیلی ریاضی در نظر گرفت، بر همین اساس این پژوهش از نظر رویکرد، یک پژوهش کمی است. فرآیند استاندارد میان صنعتی داده‌کاوی یا به اختصار CRISP DM (چپمن و همکاران ۲۰۰۰)، به‌عنوان روش‌شناسی تحلیل مورد استفاده قرار کرفته است. از مدل‌های خوشه‌بندی و الگوریتم‌های K-means، شبکه عصبی Kohone ، کشف ناهنجاری مبتنی بر خوشه‌بندی، مدل مخفی مارکوف، روش پردازش موازی‌نگاشت کاهش و مدل هم‌جوشی دمپستر شیفر فازی استفاده می‌شود. تجزیه و تحلیل با استفاده از نرم‌افزارهای MATLAB و R انجام شده است. مدل پیشنهادی توانسته است از نظر خروجی و زمان اجرا، نسبت به مدل‌های دیگر عملکرد برتری را به نمایش گذارد.

کلمات کلیدی: کشف و ارزیابی تقلب، زنجیره تامین، ANN, Big data

[۱] دانشجوی دکتری مدیریت صنعتی_مالی، واحد ابهر، دانشگاه آزاد اسلامی، ابهر، ایران.

[۲] نویسنده مسئول، استادیار گروه اقتصاد مالی، واحد ابهر، دانشگاه آزاد اسلامی، ابهر، ایران.

آدرس پست الکترونیکی: **fi.asgarii@gmail.com**

[۳] دانشیار گروه مدیریت، واحد زنجان، دانشگاه آزاد اسلامی، زنجان، ایران.

[۴] استادیار گروه مدیریت، واحد زنجان، دانشگاه آزاد اسلامی، زنجان، ایران.

مقدمه

آثار زیان‌باری که رسوایی‌های مالی در سال‌های اخیر به وجود آورده، توجه به بحث تقلب را اجتناب ناپذیر کرده است. رسوایی‌های مالی شرکت‌هایی همچون انرون و ورلدکام نمونه هایی است که بازار سرمایه آمریکا را دچار نوسان کرده است. این رسوایی‌ها نه تنها هزینه‌ها و زیآن‌های اقتصادی مانند متضرر کردن اعتباردهندگان، سرمایه‌گذاران و سهامداران داشته، بلکه هزینه‌های سیاسی، قضایی و ... را نیز به بار آورده است. کاهش سرمایه‌گذاری مولد در بلندمدت، ایجاد شکاف طبقاتی در سازمان، اتلاف منابع، ایجاد نارضایتی عمومی، نارضایتی مشتریان، ضعف اعتماد عموم مردم و کاهش منابع، کاهش ارزش سهام و ورشکستگی از پیامدهای تقلب است (صدیقی کمال، ۱۳۹۲). ایران نیز از آسیب و رشد این پدیده اقتصادی مستثنی نبوده، به گونه‌ای که در دو دهه اخیر، تقلب بحث غالب بازارهای مالی و نهادهای اقتصادی بوده است. اولین چیزی که در ذهن تمام آحاد جامعه پس از این تقلب‌ها متبلور شد، این سؤال بود که پس نقش دستگاه‌های نظارتی در این سازمان‌ها چه می‌شود؟ آیا سازمان‌ها روش‌های مؤثر جهت جلوگیری از تقلب را به‌کار نمی‌گرفتند؟ با وجود این سؤالات اما توجه کمتری نسبت به شناسایی عوامل اثرگذار بر تقلب و اجرای روش‌های پیشگیری آن شده و نیز پژوهشی گسترده در مورد چرایی عدم اجرای مؤثر این روش‌های پیشگیری صورت نپذیرفته است. این در حالی است که اگر تقلب از طریق روش‌هایی مانند نظارت و کنترل و یا تطبیق حساب کشف و کنترل شود، هزینه‌های آن خیلی کمتر از زمانی است که از طریق روش‌های منفعل تشخیص داده شود. هم چنین در یک محیط تجاری مبتنی بر فن‌آوری که با تغییرات سریع همراه است، نیاز مبرمی به روش‌های مؤثر برای پیشگیری و کنترل تقلب وجود دارد (خواجوی و ابراهیمی، ۱۳۹٦). عدم درستی سیستم‌های حسابداری و گزارش‌گری مالی واحد مورد رسیدگی و تعریف نشدن کنترل‌های مناسب شامل کنترل‌های مالی، کنترل‌های مربوط به رعایت قوانین و کنترل‌های هشداردهنده خطر، باعث ایجاد و ثبت هرگونه تقلب و سوء استفاده از منابع مالی و غیرمالی شرکت‌ها می‌شود (بلکو و محمودی، ۱۳۹۵). در نتیجه عوامل مؤثر بر تقلب و اجرای روش‌های پیشگیری آن به منظور کارایی و اثربخشی این روش‌ها حائز اهمیت می‌شوند. یکی از مهم‌ترین بخش‌های یک سازمان، مدیریت زنجیره تامین می‌باشد که تقلب می‌تواند در عملکرد آن تاثیر به‌سزائی بگذارد، لذا سازمان‌ها را بر آن داشته تا برای غلبه بر شرایط نامطمئن به تحقیق جهت جلوگیری تقلب در مدیریت زنجیره تأمین بپردازند، چرا که تمامی فعالیت‌های سازمان را به منظور تولید محصولات و ارائه خدمات مورد نیاز مشتریان تحت

تأثیر قرار می‌دهد به گونه‌ای که تأمین‌کنندگان بایستی قطعات و مواد را با بهترین کیفیت و کمترین هزینه تولید نمایند و تولیدکنندگان و توزیع‌کنندگان محصولات نیز باید رابطه نزدیکی با سیاست‌های توسعه بازار داشته باشند(کرباسیان،۱۳۹۳). تقلب و تصمیمات مالی یکی از چالش‌های بسیار مهم زنجیره تأمین است که اگر وقوع آن در زنجیره پیش‌بینی نشده باشد و برای مقابله با آن راه‌کار مناسبی اتخاذ نشود، اهداف این زنجیره را متأثر می‌کند، آن را مختل می‌کند و درنتیجه موجب تهدید بقای سازمان می‌شود(دهنوی و همکاران، ۱۳۹۰).

تامین مالی زنجیره تامین(SCF)[1] جریان تازه‌ای از تحقیقات است که با هدف بهینه‌سازی جریان‌های مالی در سطح بین‌سازمانی (هافمن، ۲۰۰۵) از طریق راه‌حل‌های انجام شده توسط موسسات مالی (کامرینلی، ۲۰۰۹) یا ارائه‌دهندگان فن‌آوری انجام می‌شود. هدف نهایی هم‌ترازی جریان‌های مالی با جریان‌های اطلاعاتی و محصول در زنجیره تامین و بهبود مدیریت جریان نقدی از منظر زنجیره تامین است (ووتک وهمکاران، ۲۰۱۳). بررسی تحقیقات جدید، وجود دیدگاه‌های مختلف در تامین مالی زنجیره تامین را نشان می‌دهد. **مورد اول** بر جنبه‌های مالی متمرکز است و رویکردهای تامین مالی زنجیره تامین را به عنوان مجموعه‌ای از راه‌حل‌های مالی مورد توجه قرار می‌دهد. **مورد دوم** نقش همکاری میان اعضای زنجیره تامین را تأیید می‌کند و مرزهای تامین مالی زنجیره تامین را فراتر از راه‌حل‌های مالی گسترش می‌دهد. به همین ترتیب تامین مالی زنجیره تامین همچنین راه‌حل‌های زنجیره تامین را برای کاهش ریسک‌های مالی ارائه می‌دهد. در انجام این کار، تامین مالی زنجیره تامین بر روی زنجیره تامین و مشارکت مالی متمرکز است تا بتواند پس‌انداز، ایجاد سود و کارایی مدیریت دارایی را برای همه اعضای شبکه تامین کند (هوف و روگرز،۲۰۱۵). **علاوه بر مسائل فوق،** تنگناهای مالی پیش روی بنگاههای اقتصادی در ایران، اهمیت ملاحظه ابعاد مالی در تصمیمات مختلف سازمانی را بیشتر نموده است. نرخ بالای معوقات بانکی صنایع کوچک و متوسط نشان‌دهنده چالش‌های عمده در تصمیمات مالی این بنگاه‌هاست. لذا اخذ تصمیمات اثربخش مالی و متعاقب آن حداکثرسازی کارایی در منابع مالی از مهم‌ترین اهداف مدیران در صنایع مختلف است. در این بین ایجاد یکپارچگی میان تصمیمات مالی سازمان و تصمیمات فیزیکی زنجیره تامین این امکان را به سرمایه‌گذاران می‌دهد تا با رویکردی جامع‌تر به انتخاب استراتژیک در خصوص سرمایه‌گذاری در

[1] Supply Chain Finance

زنجیره تامین و گزینه‌های دیگر بپردازند. در حال حاضر در بسیاری از سازمان‌ها عدم یکپارچگی میان تصمیمات مالی سازمانی و تصمیمات فیزیکی زنجیره تامین منجر به تصمیم‌گیری بخشی شده است که این سازمان‌ها را از نگرشی جامع در تصمیم‌گیری محروم نموده‌است. این در حالیست که تصمیم‌گیری در مورد سرمایه‌گذاری در زنجیره تامین بخشی از مجموعه تصمیمات سازمان است که ماهیت مالی دارند.

مبانی نظری و پیشینه پژوهش

جدول ۱ نتیجه تجمیع کار گای و همکارانش با سایر بررسی های صورت پذیرفته در زمینه کشف و ارزیابی تقلب بانکی در طی پژوهش حاضر است .

جدول ۱ – مرور ادبیات در خصوص استفاده از داده‌کاوی در کشف تقلب

طبقه‌های تقلب مالی	فعالیت‌های تقلب‌آمیز	مدل‌ها	روش‌ها	منبع
تقلب بانکی	تقلب کارت اعتباری	طبقه‌بندی	الگوریتم آدا بوست، درخت‌های تصمیم، کارت، ریپر، شبکه‌های بیزی	(مایز و همکاران ۱۹۹۳)، (وانگ و همکاران ۲۰۰۳)، (فان ۲۰۰٤)، (ویان و همکاران ۲۰۰٤)، (گادی و همکاران ۲۰۰۸)، (کندو و همکاران ۲۰۰۹)، (متان و همکاران ۲۰۱۰)، راج و پرتیا ۲۰۱۱). ۲۰۱۱)، (سینگ و همکاران
			منطق ترکیبی، ابر طبقه‌بندی	استوفیلو و همکاران ۱۹۹۷)، (چان و همکاران ۱۹۹۹)، (فوا و همکاران ۲۰۰٤)، (کندو و همکاران ۲۰۰۹)، (راج و پرتیا ۲۰۱۱)، (زریپور و همکاران ۲۰۱۲)، (سینگ و همکاران ۲۰۱۱). (زریپور و شمس المعالی ۲۰۱۵).
			شبکه‌های عصبی	(مایز و همکاران ۱۹۹۳)، (گوش و رایلی ۱۹۹٤)، (ازاوا و نورتن ۱۹۹٦) (السکروف وا همکاران ۱۹۹۷)، (دور نسرو و همکاران ۱۹۹۷)، سیدا و همکاران ۲۰۰۲)، (کیم و کیم ۲۰۰۲)، چنا و گادی و)۲۰۰۷ هوانگ ۲۰۰۳)، (شن و همکاران همکاران ۲۰۰۸)، (راج و پرتیا ۲۰۱۱)، (پتیدار و شارما ۲۰۱۱)، (زریپور و همکاران ۲۰۱۲)، (شرلی ۲۰۱۲) ، (وثوق، تقوی فرد و البرزی ۱۳۹۳)

ماخذ: یافته‌های تحقیق

ادامه جدول ۱ – مرور ادبیات در خصوص استفاده از داده‌کاوی در کشف تقلب

طبقه‌های تقلب مالی	فعالیت‌های تقلب‌آمیز	مدل‌ها	روش‌ها	منبع
تقلب بانکی	تقلب کارت اعتباری	طبقه‌بندی	KNN مدل‌های لجستیکی، بیز ساده	(یه و لین ۲۰۰۸)، (گادی و همکارا ۲۰۰۸)
			الگوریتم ژنتیک	بیدگلی و همکاران ۲۰۰۳) (دومان و ازلیک ۲۰۱۱)، (راماکالیانی و اومادوی ۲۰۱۲).
			سیستم‌های ایمنی مصنوعی	هانت و همکاران ۱۹۹۸)، (رنگ و همکاران ۲۰۱۲)حلوایی و اکبری ۲۰۱٤)
			فازی داروینی	بنتلی و همکاران ۲۰۰۰)
			رگرسیون	(فاستر و استین ۲۰۰٤)
			ماشین برداری پشتیبان	چن و همکاران ۲۰۰٦)
		خوشه‌بندی	مدل مخفی مارکوف نقشه‌های خود سازمان‌دهنده	سریواستاوا و همکاران ۲۰۰۸)، (بوساری و پاتیل ۲۰۱۱)، (کومار و راج ۲۰۱۲)، (میشرا و همکاران (زاسلاوسکی و استریژاک ۲۰۰٦)، (کوا و ۲۰۱۳ سریگانش ۲۰۰۸)، (اولرووسکی ۲۰۱٤)
		کشف ناهنجاری	با Tآماره فاصله هلینگر	بولتون و هند ۲۰۰۱
	پول‌شویی	طبقه‌بندی	تحلیل شبکه	گائو و یه ۲۰۰۷

ماخذ: یافته‌های تحقیق

برای کشف و ارزیابی تقلب در کارت‌های پرداخت از روش‌ها و الگوریتم‌های متعددی استفاده می‌شود که در زیر به بعضی از آنها اشاره شده است: الگوریتم‌های هم‌جوشی اطلاعات مانند تئوری گواه دمپستر و شیفر و یادگیری بیزی (کندو و همکاران ۲۰۰۹؛ راج و پرتیا ۲۰۱۱؛ سینگ و همکاران ۲۰۱۱)، ایجاد منطق ترکیبی (زریپور و همکاران ۲۰۱۲)، استفاده از مدل مخفی مارکوف (سریواستاوا و همکاران ۲۰۰۸؛ بوساری و پاتیل ۲۰۱۱)، شبکه‌های عصبی (راج و پرتیا ۲۰۱۱؛ زریپور و همکاران ۲۰۱۲)، الگوریتم ژنتیک (بیدگلی و همکاران ۲۰۰۳ء دومان و ازلیک ۲۰۱۱؛ راماکالیانی و اومادوی

۲۰۱۲)، و سیستم‌های ایمنی مصنوعی (هانت و همکاران ۱۹۹۸؛ ونگ و همکاران ۲۰۱۲). همان‌گونه که در بخش مبانی نظری بررسی شد، الگوریتم‌های یاد شده در قالب مدل‌های با سرپرستی، نیمه سرپرستی، بدون سرپرستی و ترکیبی قرار می‌گیرند.

اساساً بیشترین تکنیک‌های داده‌کاوی که در حوزه کشف تقلب مورد استفاده قرار گرفته‌اند به ترتیب عبارتند از: مدل‌های لگاریتمی، شبکه‌های عصبی، شبکه‌های بیزی، و درخت‌های تصمیم که همگی در حوزه طبقه‌بندی جای می‌گیرند. مدل‌های لگاریتمی، مدل‌های تعمیم یافته خطی هستند که برای رگرسیون دو قطبی مورد استفاده قرار می‌گیرند، در این مدل‌ها متغیر پیش‌بینی‌کننده می‌تواند هم عددی و هم طبقه‌ای باشد (یه و لین ۲۰۰۸) بیشترین استفاده از این مدل کشف تقلب‌های سازمانی و بیمه‌ای است. شبکه‌های عصبی در کارت‌های اعتباری، بیمه اتومبیل، و تقلب‌های سازمانی مورد استفاده قرار گرفته‌اند. شبکه‌های بیزی نیز مانند شبکه‌های عصبی در کارت‌های اعتباری، بیمه اتومبیل، و سازمانی استفاده شده‌اند. درخت‌های تصمیم: در کارت‌های اعتباری، بیمه های اتومبیل، سازمانی و امنیتی (گای و همکاران ۲۰۱۱).

بیشترین روش‌هایی که در حوزه کارت‌های اعتباری و هوشمند مورد استفاده قرار گرفته‌اند، عبارتند از تکنیک‌های استخراج قاعده، درخت تصمیم، شبکه‌های عصبی مصنوعی، SVM، رگرسیون لگاریتمی، و روش‌های فراابتکاری. البته بیشتر این روش‌ها از الگوریتم‌های با سرپرستی استفاده می‌کنند (ساهین، بولکان و دومان ۲۰۱۳). مایز و همکارانش برای کشف تقلب در تراکنش‌های کارت‌های پرداخت از شبکه‌های عصبی و شبکه‌های بیزی استفاده کرده و این دو را با هم مقایسه کرده‌اند (مایز و همکاران ۱۹۹۳). به عقیده آنها شبکه‌های عصبی دقت بالاتری را در طبقه‌بندی از خود نشان داده‌اند. ازاوا و نورتون یک شبکه بیزی در چهار مرحله با دو پارامتر را برای کشف تقلب استفاده می‌کنند. آنها بیان کردند که رگرسیون، نزدیکترین همسایه و شبکه‌های عصبی بسیار کند هستند و درخت‌های تصمیم با برخی از متغیرهای گسسته دچار دشواری‌هایی می‌شوند (ازاوا و نورتن ۱۹۹۶). گوش و رایلی با استفاده از شبکه عصبی جلوسوی شعاعی، روشی برای کشف تقلب در کارت‌های الکترونیکی ارائه دادند. در روش پیشنهاد شده توسط آنها شبکه با استفاده از یک نمونه داده‌ای بزرگ آموزش داده می‌شود. همان‌گونه که مشخص است در این روش کلیه تراکنش‌های کارت دارای برچسب هستند. در نمونه مورد استفاده تراکنش‌ها دارای برچسب های گم شدن فیزیکی کارت، تقلب در استفاده از برنامه کاربردی مرتبط با کارت، کارت‌های جعلی و ... هستند (گوش و رایلی

۱۹۹٤). این مدل یکی از اولین کارهای استفاده‌کننده از الگوریتم پس انتشار خطایی است که گزارش شده است. البته در این مدل صرف تقلب‌هایی که شبکه برای آنها آموزش دیده است، کشف می‌شوند. السکروف با تیم همکارش و یک سیستم کاوش پایگاه داده برای یافتن تقلب در کارت‌های الکترونیکی با استفاده از شبکه‌های عصبی به‌عنوان پاینده کارت ایجاد کردند (السکروف و همکاران ۱۹۹۷). سیدا و همکاران ۲۰۰۲ یک شبکه عصبی فازی بر روی ماشین‌های موازی برای بالا بردن سرعت تولید قواعد کشف تقلب کارت‌های الکترونیکی ارائه داده‌اند (سیدا، ژانگ و پن ۲۰۰۲). به گزارش آنها موازی‌سازی توانسته است سرعت قابل قبولی را فراهم کند؛ البته تعریف وظایف یکسان دشوار گزارش شده است.

بارس و همکارانش نیز از یک شبکه عصبی جلوسو با استفاده از حافظه رهگیری نمایی برای مدیریت وابستگی‌های زمانی در تراکنش‌های مربوط به تقلب در درخواست های نوار های ویدیو استفاده کرده‌اند (بارس، وارنستروم و جانسون ۲۰۰۳). ویژگی مهم تر پژوهش بارس، ارائه یک روش‌شناسی مورد پذیرش برای شبیه‌سازی داده‌های مصنوعی برای مدل‌سازی با استفاده از تعداد کمی داده واقعی تولید شده در محیط آزمایشگاهی است.

مدل‌سازی آماری مانند رگرسیون برای کشف تقلب به شدت مورد استفاده قرار گرفته است. در همین راستا، فاستر و استین از رگرسیون کمترین مجذور خطا و انتخاب گام به گام استفاده کرده و نشان داده‌اند که روش‌های آماری استاندارد نیز همچنان رقابت‌پذیر هستند. این نسخه رگرسیون کاملا خودکار گام به گام سه مرحله دارد؛ نخست سازماندهی محاسبات برای تطبیق با تعاملات، دوم استخراج معیارهای نظری پیشرفته برای انتخاب روش پیش‌بینی، و سوم تخمین محتاطانه ارزش احتمال برای مدیریت داده‌های پراکنده و پاسخ دوقطبی قبل از پیش‌بینی (فاستر و استین ۲۰۰٤).

بلهاجی و همکارانش بهترین شاخص و ویژگی‌های تقلب را با استفاده از روشی سه مرحله‌ای که مرحله نخست آن پرسش از خبرگان، دوم استخراج احتمالات شرطی تقلب برای هر شاخص و سوم رگرسیون پروبیت برای تعیین مهم‌ترین شاخص است، استخراج کرده‌اند. مولفین همچنین از رگرسیون پروبیت برای پیش‌بینی تقلب و تنظیم آستانه برای تطابق رویه‌های تقلب شرکت‌های بیمه‌ای در مورد بیمه اتومبیل، سود برده‌اند (بلهاجی، دیون و تار خانی ۲۰۰۰). بعضی از روش‌های دیگر نیز وجود دارند که از کنار هم چیدن سیستم‌های خبره، قواعد انجمنی، الگوریتم ژنتیک و ... استفاده می‌کنند.

به‌عنوان مثال پژیک بچ نمایی کلی از یک سیستم نمایه‌سازی برای کشف و ممانعت از تقلب ارائه داده است (پژیک بچ ۲۰۱۰).

ارمرد و همکاران یک تولیدکننده قاعده را برای پالایش وزن‌های شبکه‌های بیزی در یافتن تقلب‌های بیمه‌ای پیشنهاد می‌کنند (ارمرد و همکاران ۲۰۰۳). هی و همکاران یک الگوریتم ژنتیک برای تخمین وزن‌های بهینه ویژگی‌های تعداد K نزدیکترین همسایه برای طبقه‌بندی داده‌های تقلب پزشکی ارائه داده‌اند؛ آنها روش خود را با استدلال مورد محور، مقایسه و نتایج بهتری را گزارش کرده‌اند (هی و همکاران ۱۹۹۹) استو فیلو و همکاران یک سیستم کشف تقلب را برای استفاده از روش‌های ابر یادگیری مدل‌های تراکنش‌های کارت‌های الکترونیکی تقلب‌آمیز ارائه داده‌اند (استو فیلو و همکاران ۲۱۹۹۷). همان گروه بر روی مدلی مبتنی بر هزینه کار کردند. آنها با استفاده از یک سیستم داده‌کاوی توزیع شده برای یافتن تقلب‌های کارت‌های الکترونیکی یک مدل ساختند (استوفیلو و همکاران ۲۰۰۰). زریپور و شمس المعالی با استفاده از داده‌های یک مسابقه کشف تقلب که در سال ۲۰۰۹ در سان دیگر انجام شد، مدلی با استفاده از آموزش همزمان چند طبقه‌بند و تلفیق نتایج با استفاده از تجمیع خود را اندازه رویکردی برای آموزش همزمان دو طبقه‌بند برای طبقه‌بندی مشاهدات) و درخت تصمیم ارائه کردند (زریپور و شمس المعالی ۲۰۱۵).

کندو و همکارانش اظهار می‌کنند که در درجه اول هر دارنده کارت رفتار خرید ویژه‌ای دارد که می‌توان بر اساس آن یک پروفایل ایجاد کرد (کندو و همکاران ۲۰۰۹). از سوی دیگر استفاده از این منبع در زمان تغییر رفتار دارنده کارت ناکارآمد خواهد بود. بنابراین برای غلبه بر نقص ناکافی بودن منابع داده‌ای، آنها از دو سیستم طبقه‌بندی بر اساس شبکه‌های بیزی، یکی برمبنای شباهت به رفتار قانونی مالک کارت و دیگری شباهت به رفتار تقلب‌آمیز ایجاد کردند. نتیجه دو رویکرد طبقه‌بندی، با تئوری گواه دمپستر و شیفر ترکیب شده که حاصل گزارش میزان اعلان اشتباه پایین بوده است (کندو و همکاران ۲۰۰۹). یک رویکرد برای مدیریت مسائل پیچیده، ترکیب مدل‌های طبقه‌بندی و ساخت یک ابر طبقه‌بند برای استخراج رفتارهای موضعی متفاوت، و دست‌یابی به یک مدل طبقه‌بندی کلی است (هداوندی، شهرابی و هایاشی ۲۰۱۵). در همین راستا هداوندی و همکاران (۲۰۱۵) رویکردی با سرپرستی را براساس ترکیب خبرگان، یادگیری همبستگی منفی)، و روش طرح زیرفضاها جهت ترکیبی تقویت‌کننده ارائه داده‌اند. موتور هوش مصنوعی یا به عبارتی طبقه‌بند اصلی استفاده شده در این روش، شبکه عصبی توام است (هداوندی، شهرابی و هایاشی ۲۰۱۵). روش ترکیب خبرگان، این

امکان را به هر طبقه‌بند می‌دهد که بر روی زیر مجموعه‌ای افراز شده از فضای آموزش تمرکز کرده و آن را به خوبی اکتشاف کند؛ به‌علاوه یک شبکه پویا نیز بر اساس کارایی هر خبره (طبقه‌بند) وزن‌ها را برای ترکیب نهایی تخمین می‌زند هداوندی، شهرابی و هایاشی ۲۰۱۵). یادگیری همبستگی منفی نیز یک قاعده برای آموزش یا به عبارتی افزایش تعمیم‌پذیری مدل، ارائه می‌دهد. نهایتا روش طرح زیر فضاها، رویکردی برای افراز فضای آموزش است که می‌تواند گوناگونی را در نظر بگیرد. این روش توانسته است نقایص جداگانه هر یک از روش‌ها را برای ترکیب طبقه‌بندها جبران کند (هداوندی، شهرابی و هایاشی ۲۰۱۵). رویکرد ارائه شده، برای مسائل طبقه‌بندی رگرسیون چند هدفه‌ای که اولا دارای تعداد زیاد ویژگی هستند، ثانیا دارای پیچیدگی هستند، و نهایتا مشاهدات کمی در اختیار دارند، ارائه شده است (هداوندی، شهرابی و هایاشی ۲۰۱۵) .

سریواستاوا هم با استفاده از فرآیندهای تصادفی و مدل مخفی مارکوف، مدلی را برای کشف تقلب ارائه می‌دهد که نیاز به داده‌های برچسب‌دار نداشته و با استفاده از عادت استفاده دارنده کارت، اقدام به کشف تقلب در کارت‌های الکترونیکی می‌کند. در این روش توالی تراکنش‌های کارت بانکی به وسیله فرآیندهای تصادفی مارکوف مخفی استخراج می‌شود. از آنجا که ریزاقلام خریداری شده توسط صاحب کارت در یک تراکنش معمولا برای بانک صادرکننده کارت نمایان نیست، با استفاده از یک روش خوشه‌بندی، تراکنش‌ها را صرفا از لحاظ میزان مالی در سه خوشه کم، متوسط و زیاد جای دادند. تغییر در هر خوشه به عنوان توالی مشاهده در نظر گرفته شد. ضمنا خوشه‌ای که دارای بیشترین فراوانی عضویت تراکنش بود، به‌عنوان پروفایل در نر گرفته شده و نقش اساسی در فرآیند مقداردهی اولیه برای آموزش مدل مارکوف بازی کرد. در خود تراکنش فرآیندهای تصادفی بر اساس مقدار پول خرج شده ساخته می‌شوند. آنها نرخ اعلان اشتباه پایینی را گزارش داده‌اند (سریواستاوا و همکاران ۲۰۰۸).

اساسا در پژوهش‌های اخیر مشاهده می‌شود که استفاده‌کنندگان از مدل مخفی مارکوف برای کشف و ارزیابی تقلب به شدت در حال کاهش نرخ اعلان اشتباه در تراکنش‌های کشف شده هستند (میشرا و همکاران ۲۰۱۳). اپیلارد و بوگیلا با استفاده از یک مدل مخفی مارکوف در داده‌های حاصل از پردازش تصویر دوربین‌های شهری به دنبال کشف ناهنجاری بوده‌اند (اپیلارد و بوگیلا ۲۰۱۶). در مدل ارائه شده توسط آنها برای محاسبه احتمالات ماتریس انتشار از تعمیم دیریکله و توزیع لیوویل بتا استفاده کرده و عملکرد بهتری را در یافتن نقاط ناهنجار نسبت به مدل مارکوف سنتی گزارش

داده‌اند. در پژوهشی دیگر، یو و وانگ از روشی بر مبنای سنجش فاصله برای یافتن مقدار حدی در میان داده‌ها به‌عنوان معیاری برای تقلب استفاده کرده‌اند. در این روش با بهره بردن از میزان فاصله با نزدیک‌ترین همسایه، مقادیر مرکزی و مقادیر حدی تراکنش‌ها مشخص می‌شود. این پژوهش صحت را در بهترین حالت با ارزشی برابر با ٪۸۹ نشان داده و اشاره ای به نرخ مثبت حقیقی و اعلان اشتباه ندارد (یو و وانگ ۲۰۰۹).

نظریه گواه قابلیت حل مسئله با چندین تصمیم‌گیر و انواع داده‌های ناهمگن را دارد. این نظریه به این علت که توانایی حل مسائل با بی خبری و اطلاعات از بین رفته، ناکامل و نامطمئن که معمولا در مسئله ارزیابی زیاد اتفاق می‌افتد را داراست، در این حوزه به شکل فراوانی به‌شدت مورد استفاده قرار گرفته است (آواستی و ساتیاریر ۲۰۱۱). هر چند نظریه گواه، توانایی حل مسائل با داده نامطمئن را دارد؛ ولی در حل مسائلی که داده‌هایی با چند نوع ابهام در آنها وجود دارد، ناتوان است. از نمونه‌این پژوهش‌ها می‌توان به کارهای زیر اشاره کرد. یانگ و همکارانش از نظریه گواه برای تجمیع اطلاعات ناپایدار، ناکامل و نامطمئن ارزیابی ریسک تیغه‌های توربین هواپیما که از چندین خبره به‌دست می‌آمده استفاده کردند. آنها نظریه گواه را برای حل این مسئله به‌کار گرفته؛ به طوری که برای هر خبره و فاکتور تاثیرگذار بر روی ریسک در مسئله، وزنی در نظر گرفتند (یانگ و همکاران ۲۰۱۱). پوپسکو از نظریه گواه برای داده ورودی مبهم و ناکامل استفاده کرده و نشان داد که چگونه می‌توان با آن الگوهای منفی مثبت ورودی را مشاهده کرد (پوپسکو و همکاران ۲۰۱۰). بائه و همکارانش نظریه گواه را به شکل سودمندی با ارائه الگوریتم کم هزینه محاسباتی برای ساختارهای طراحی مهندسی بر روی تحلیل کمی‌سازی نااطمینانی به کار گرفتند. الگوریتم آنها روش‌های بهینه‌سازی و تقریبی برای ارزیابی کارای توابع باور و مقبولیت بدون از دست دادن دقت به کار گرفته است (بائه و همکاران ۲۰۰٤) هی و همکارانش از نظریه گواه و روش بیزی مونت کارلو، برای تخمین سلامت و زمان مفید باقی‌مانده برای باتری‌های لیتیوم استفاده کردند. آنها پارامترهای مختلفی را برای ارزیابی استخراج، و بر مبنای نظریه گواه آنها را آموزش داده و سپس از طریق روش مونت کارلو آنها را به‌روزرسانی کردند. نهایتا مدت زمان باقی‌مانده عمر باتری را پیش‌بینی کردند (هی و همکاران ۲۰۱۱). دیموا از ترکیب نظریه گواه و فازی برای تصمیم‌گیری در مورد خرید و فروش سهام در بورس ورشو استفاده، و ادعا کردند که سیستم آنها در عمل سودآوری خوبی به دست آورده است (دی موا، سواستیانف و بارت ۲۰۱۱). دنگ و همکارانش نیز از ترکیب نظریه‌های فازی و گواه برای حل مسائل تصمیم‌گیری

چندشرطی استفاده کردند؛ در پژوهش آنها، نظرات خبرگان با استفاده از اعداد فازی نمایش داده و سپس از طریق رتبه‌بندی برای تولید باور پایه به کار رفته است. در نهایت آنها داده‌های هر شرط را به وسیله قاعده ترکیب دمپستر هم جوش کرده و از طریق تبدیل احتمال ضمانتی، رتبه‌بندی نهایی آنها را مشخص کردند (دنگ و همکاران ۲۰۱۱). طبسیان و همکارانش از ترکیب یادگیری ماشینی و نظریه گواه برای طبقه‌بندی در حالتی که برچسب داده‌ها ناقص و دارای ابهام است، استفاده کردند. آنها از شبکه عصبی و با استفاده از داده آموزشی تخصیص باور پایه نظریه گواه را انجام دادند. نتایج عددی نشان داد که روش ترکیبی آنها کارایی خوبی نسبت به حالت استفاده مجزا در شناسایی برچسب و طبقه‌بندی داده دارد (طبسیان، قادری و ابراهیم پور ۲۰۱۲). خطیبی و منتظر از ترکیب این دو نظریه برای ارزیابی بیماری‌های قلبی استفاده کرده‌اند. روش پیشنهادی آنها در عمل نتایج خوبی نسبت به روش‌های قبلی به‌دست آورده است (خطیبی و منتظر ۲۰۱۰). منصوری و همکارانش با استفاده از ترکیب گواه و فازی نتایج حاصل از دو سیستم استنتاج فازی، برای تشخیص کیفیت خدمت سرویس‌های صدا بر روی اینترنت استفاده کردند. آنها نخست دو مفهوم کیفیت خدمت و کیفیت تجربه را اندازه‌گیری کرده و سپس هر دو را با ترکیب گواه، فازی هم‌جوش کردند (منصوری و همکاران ۲۰۱۵). لذا می‌توان بیان داشت که کشف و ارزیابی تقلب مبتنی بر رویکرد تلفیقی ANN-Big Data تاکنون صورت نگرفته و تلفیقشان می‌تواند نتایج بهتری را تولید کند.

روش‌شناسی پژوهش

این پژوهش **از نظر هدف**، یک پژوهش توسعه‌ای و کاربردی است. نخست به دلیل ایجاد یک مدل جامع، با در نظر گرفتن منابع مختلف اخذ تصمیم در خصوص قانونی یا تقلب‌آمیز بودن تراکنش‌های کارت الکترونیک، و الزامات تحلیل بزرگ داده، سعی به توسعه دانشی نظری در این حوزه دارد که تاکنون هیچ مطالعه‌ای نسبت به آن اقدام نکرده است؛ بنابراین می‌توان این پژوهش را توسعه‌ای در نظر گرفت. به‌علاوه، پیاده‌سازی این مدل، می‌تواند برای سازمان‌های فعال در این زمینه و حتی سازمان‌های نظارتی، مفید بوده و عملاً راه‌حلی برای مساله تقلب در آنها باشد؛ بنابراین از جمله پژوهش‌های کاربردی محسوب می‌شود. با توجه به اینکه در پژوهش حاضر داده‌های مورد استفاده درون یک پایگاه داده بوده، و یا با استفاده از ابزارهای شبیه‌سازی تولید می‌شوند، و اینکه اساس مدل بر مدل‌های خوشه‌بندی و طبقه‌بندی داده‌کاوی و الگوریتم‌های هوش مصنوعی از یک سو و مدل‌های

محاسباتی توزیع شده و موازی از سوی دیگر بوده که همگی اصالتاً مبنای ریاضی دارند، این پژوهش را می‌توان **از لحاظ ماهیت** یک پژوهش تحلیلی ریاضی در نظر گرفت. برهمین اساس این پژوهش **از نظر رویکرد**، یک پژوهش کمی است. با عنایت به اهمیت زیاد گردآوری داده‌ها و ابزارهای آن مانند: مشاهده، پرسشنامه، مطالعه استاد و مدارک، و مصاحبه، در این پژوهش در گام نخست با استفاده از بررسی اسناد و مدارک، مطالعه تجربیات صورت پذیرفته از طریق پایگاه‌های مقالات اینترنتی مدل مفهومی ساخته شده، سپس برای پاسخ به سوالات زیر، گروهی کانونی متشکل از خبرگان در حوزه کارت‌های الکترونیک استفاده شده است. در این پژوهش ٦ خبره که دارای ویژگی‌های زیر هستند، از هر دو سازمان انتخاب شدند.

۱–حداقل دارای مدرک کارشناسی ارشد باشند.

۲– حداقل ۵ سال در زمینه معماری کارت‌های هوشمند، استانداردها و ساختارهای پرداخت، ترمینال‌های کارت‌خوان، معماری داده‌ای، و فرآیند مدیریت کارت سابقه داشته باشند.

۳– با مفاهیم تقلب مالی و رویکردهای مختلف آن آشنا بوده و در این زمینه دارای تجربه باشند.

٤– حتی‌المقدور دارای ترکیبی از دانش فنی از منظرهای مختلف، کارت، شبکه، پایگاه داده و فرآیندهای کارت باشند. در جلسات گروه کانونی مد نظر، سوالات زیر مطرح شده و فرآیند مباحثه میان خبرگان تا دست‌یابی به اجماع ادامه یافته است. محقق در جلسات صرفا سوالات را مطرح و سعی کرده است ساختار فرآیند را مطابق با روش گروه‌های کانونی دنبال کند. چه ویژگی‌هایی از تراکنش که در بخش پارامترهای مستقل کشف الگو، در فصل دو به آنها پرداخته شد، می‌تواند توصیف کننده الگوی رفتاری باشد؟ میزان تغییر یاد شده چه مقدار است؟ وزن خوشه‌های ایستا و پویا در استنتاج نهایی چه مقدار است؟

با توجه به جامعه اشاره شده، در این پژوهش داده‌های جامعه انتخاب شده به‌صورت نمونه‌ای از کل سیستم مورد بررسی قرار گرفته است. با تمرکز بر حوزه‌ی مدیریت زنجیره تأمین (SCM) بر آن شدیم تا این پرسشنامه را از طریق یک نظرسنجی که ۸٤ مدیر را شامل می‌شود، به منظور بررسی عملکردی کلان داده‌ها در تصمیم‌گیری مدیریت زنجیره تامین اعتبار سنجی کنیم.

با توجه به هدف، با استفاده از یادگیری ماشین برای کشف تقلب، به همراه ابزارهای تحلیل بزرگ داده؛ مانند پارادایم نگاشت کاهش، تحلیل داده‌ها برای ارائه مدل نهایی صورت خواهد پذیرفت. در همین راستا، از مدل‌های خوشه‌بندی و الگوریتم‌های K -means، شبکه عصبی Kohonen کشف ناهنجاری مبتنی بر خوشه‌بندی، مدل مخفی مارکوف، روش پردازش موازی نگاشت کاهش، و مدل هم‌جوشی دمپستر شیفر فازی استفاده می‌شود. روش‌شناسی فرآیند استاندارد میان صنعتی داده‌کاوی یا به اختصار CRISP DM (چپمن و همکاران ۲۰۰۰)، به‌عنوان روش‌شناسی تحلیل مورد استفاده قرار می‌گیرد. با توجه به توضیح کاملی که در این خصوص در فصل دو داده شده است، در ادامه به بررسی مساله پژوهش بر اساس گام‌های CRISPپرداخته می‌شود. تنها نکته‌ای که باقی‌مانده است، این است که به اعتقاد چپمن، این روش‌شناسی، یک روش‌شناسی رفت و برگشتی است، یعنی فرآیند میان گام‌ها به قدری تکرار می‌شود تا بتوان به اهداف پژوهش دست یافت (چپمن و همکاران ۲۰۰۰).

بنابر اهداف پژوهش، تحلیل اطلاعات در **سه فاز** مورد بررسی قرار می‌گیرد:

- ✓ روش تجزیه و تحلیل در **فاز اول** پژوهش حاضر مساله کشف تقلب در تراکنش‌های کارت‌های الکترونیک به صورت یک مساله کشف ناهنجاری بدون‌سرپرستی با ورود الزامات تحلیل بزرگ‌داده در نظر گرفته می‌شود. مدل ارائه شده جهت حل این مساله نیز، از هم‌جوشی استنتاجات ناشی از سه منبع سابقه تراکنشی موجودیت مالک تراکنش، سابقه تراکنشی خوشه‌ایستا و سابقه تراکنشی خوشه پویا سود می‌برد. برای این منظور از مدل مخفی مارکوف به همراه خوشه‌های چندبعدی برای ایجاد سابقه تراکنشی موجودیت مالک؛ و از ترکیب k-means با شبکه عصبی کوهونن به‌صورت متوالی، برای ساخت سوابق تراکنشی خوشه‌های ایستا و پویا استفاده خواهد شد. ضمناً، الزامات بزرگ‌داده با تغییر شکل مدل‌ها به صورت فرآیندهای مستقل، بهره‌برداری از چارچوب محاسباتی نگاشت‌کاهش، در نظر گرفته شده است. نهایتاً برای هم‌جوشی منابع نیز تئوری گواه دمپستر–شیفر پیشنهاد می‌شود.
- ✓ در **فاز دوم**، رقابت جهت ارائه ارزش بهتر به مشتریان، از رقابت میان بنگاه‌ها به سمت رقابت میان زنجیره‌های تامین سوق پیدا کرده‌است. به منظور بررسی فاز دوم پژوهش برای مدل‌سازی از برنامه‌ریزی خطی عدد صحیح بهره گرفته خواهد شد. پس از طراحی شبکه زنجیره تامین در سطح استراتژیک، برنامه اصلی زنجیره‌تامین در سطح تاکتیکی

طراحی خواهد شد. در این سطح جریان فیزیکی و جریان مالی زنجیره‌تامین در چهارچوب توسعه پایدار بهینه خواهد شد. برای بهینه‌سازی از رویکرد برنامه‌ریزی آرمانی استفاده می‌شود که آرمان‌ها شامل آرمان‌های اقتصادی با رویکر بهینه‌سازی درآمد و شاخص‌های مالی، آرمان اجتماعی و آرمان زیست‌محیطی می‌باشند.

- ✓ در **فاز سوم** به منظور بررسی عملکردی اولویت‌بندیِ کلان داده‌ها در تصمیم‌گیری مدیریت زنجیره تامین با یک نظرسنجی از مدیران اعتبارسنجی آن پرداخته شده است.

روش تجزیه و تحلیل با استفاده از نرم‌افزارهای تحلیلی مورد نیاز از جمله MATLAB و R انجام گردیده است.

نتایج و بحث

منابع داده

همان‌گونه که اشاره شد، برای این پژوهش از دو مجموعه داده‌ای استفاده شده است. مجموعه داده‌ای نخست مربوط به داده‌های خودروهای نفت گازسوز و سوخت‌گیری آنها در سامانه هوشمند سوخت و مجموعه دوم مربوط به سیستم بانکی دریافت شده از یک شرکت ارائه دهنده خدمات پرداخت است. در زیر به نحوه گردآوری و تحلیل هر کدام از این مجموعه‌ها پرداخته می‌شود.

– **سامانه هوشمند سوخت:** از این سامانه تعداد ۱۵۷۹۷ کارت با توزیعی نزدیک به یکنواخت از طبقات ۱۱گانه خودروهای نفت گاز سوز (دیزلی) به همراه مشخصاتشان بدون دست‌یابی به اطلاعات هویتی و شماره کارت، و با لحاظ کردن کلیه موارد محرمانگی انتخاب شدند. در انتخاب این کارت‌ها تنها نکته‌ای که لحاظ شد، این بود که، از مجموع تراکنش‌های آنها حداقل ۲ درصد تراکنش برچسب دار تقلب‌آمیز وجود داشته باشد (زریپور و شمس المعالی ۲۰۱۵؛ باچمایر ۲۰۰۸: بولتن و همکاران ۲۰۰۲؛ دال – پوزولو و همکاران ۲۰۱٤؛ ون ولاسلر و همکاران ۲۰۱۵؛ ویدروو و همکاران ۲۰۰۸؛ گادی و همکاران ۲۰۰۸). توزیع این تراکنش‌ها و نیز انتخاب مابقی کارت‌ها به‌صورت کاملا تصادفی در نظر گرفته شد. مقطع تاریخی استخراج تراکنش‌ها از اول مهرماه ۱۳۹٤ الی پایان دی ماه همان سال بود. حاصل این انتخاب جدول زیر است.

جدول ۲ - مجموعه اولیه استخراج شده از سامانه هوشمند سوخت

تعداد کارت	کل تراکنش‌ها	تراکنش‌های قانونی	تراکنش‌های تقلب‌آمیز
۱۵۷۹۷	۷۲۲۹۸۸	۷۰۳۱۵۱	۱۹۸۳۷

ماخذ: یافته‌های تحقیق

ویژگی‌های کلیدی کارت‌ها به‌شرح جدول زیر می‌باشد.

جدول ۳ - ویژگی‌های تراکنشی استفاده شده در پژوهش‌های پیشین

ویژگی	منبع
زمان تراکنش	بنتلی و همکاران ۲۰۰۰؛ گادی و همکاران ۲۰۰۸؛ کندو و همکاران ۲۰۰۹؛ ؛ حلوایی و اکبری ۲۰۱٤؛ اولزوسکی ۲۰۱٤؛ زریپور و شمس المعالی ۲۰۱۵؛ کرا بنسن و همکاران ۲۰۱٦)
مبلغ تراکنش	(السکروف و همکاران ۱۹۹۷؛ بنتلی و همکاران ۲۰۰۰؛ زاسلاوسکی و استریژاک ۲۰۰٦؛ سریواستاوا و همکاران ۲۰۰۸؛ گادی و همکاران ۲۰۰۸؛ کوآ و سریگانش ۲۰۰۸؛ کندو و همکاران ۲۰۰۹؛ کومار و راج ۲۰۱۲؛ میشرا و همکاران ۲۰۱۳؛ حلوایی و اکبری ۲۰۱٤؛ اولزوسکی ۲۰۱٤ زریپور و شمس المعالی ۲۰۱۵؛ کرا بنسن و همکاران ۲۰۱٦)
مبلغ قبلی	باچمایر ۲۰۰۸؛ گادی و مکاران ۲۰۰۸؛ حلوایی و اکبری ۲۰۱٤
آدرس تراکنش	گادی و همکاران ۲۰۰۸؛ کندو و همکاران ۲۰۰۹؛ حلوایی و اکبری ۲۰۱٤؛ اولروسکی ۲۰۱٤؛ زریپور و شمس المعالی ۲۰۱۵
آدرس قبلی تراکنش	گادی و همکاران ۲۰۰۸؛ حلوایی و اکبری ۲۰۱٤)
تعداد تراکنش در طول زمان	(زاسلاوسکی و استریژاک ۲۰۰٦؛ ویترو و همکاران ۲۰۰۸؛ تاسولیس و آدامز ۲۰۰۸ء باچاریا ۲۰۱۱)
جمع ارزش پولی در طول D ساعت	(زاسلاوسکی و استریژاک ۲۰۰٦؛ ویترو و همکاران ۲۰۰۸؛ تاسولیس و آدامز ۲۰۰۸؛ باچاریا ۲۰۱۱
متوسط ارزش پولی تراکنش	سریواستاوا و همکاران ۲۰۰۸؛ تاسولیس و آدامز ۲۰۰۸؛ باچاریا ۲۰۱۱؛ کومار و راج ۲۰۱۲؛ میشرا و همکاران ۲۰۱۳)
متوسط تعداد تراکنش	ویترو و همکاران ۲۰۰۸؛ تاسولیس و آدامز ۸ - ۲۰: باچاریا ۲۰۱۱)
فاصله زمانی با تراکنش قبلی	السکروف و همکاران ۱۹۹۷؛ بنتلی و همکاران ۲۰۰۰؛ گادی و همکاران ۲۰۰۸؛ ویترو و همکاران ۲۰۰۸ باچمایر ۲۰۰۸؛ کندو و همکاران ۲۰۰۹؛ حلوایی و اکبری ۲۰۱٤)
متوسط ساعات انجام تراکنش	(ویترو و همکاران ۲۰۰۸؛ کرابنسن و همکاران ۲۰۱٦)

شهر تراکنش	(زاسلاوسکی و استریژاک ۲۰۰٦؛ گادی و همکاران ۲۰۰۸؛ حلوایی و اکبری ۲۰۱٤)
کشور تراکنش	(زاسلاوسکی و استریاک ۲۰۰٦؛ مادی و همکاران ۲۰۰۸ حلوایی و اکبری ۲۰۱٤)
نوع تراکنش	(زاسلاوسکی و استریژاک ۲۰۰٦؛ گادی و همکاران ۲۰۰۸؛ میشرا و همکاران ۲۰۱۳؛ حلوایی و اکبری ۲۰۱٤)
شماره ترمینال	(زاسلاوسکی و استریژاک ۲۰۰٦؛ گادی و همکاران ۲۰۰۸؛ حلوایی و اکبری ۲۰۱٤)

ماخذ: یافته‌های تحقیق

جدول ۳ – ویژگی‌های تراکنشی استفاده شده در پژوهش‌های پیشین

ویژگی	منبع
زمان تراکنش	بنتلی و همکاران ۲۰۰۰؛ گادی و همکاران ۲۰۰۸؛ کندو و همکاران ۲۰۰۹؛ ؛ حلوایی و اکبری ۲۰۱٤؛ اولزوسکی ۲۰۱٤؛ زریپور و شمس المعالی ۲۰۱۵؛ کرا بنسن و همکاران ۲۰۱٦)
مبلغ تراکنش	(السکروف و همکاران ۱۹۹۷؛ بنتلی و همکاران ۲۰۰۰؛ زاسلاوسکی و استریژاک ۲۰۰٦؛ سریواستاوا و همکاران ۲۰۰۸؛ گادی و همکاران ۲۰۰۸؛ کوآ و سریگانش ۲۰۰۸؛ کندو و همکاران ۲۰۰۹؛ کومار و راج ۲۰۱۲؛ میشرا و همکاران ۲۰۱۳؛ حلوایی و اکبری ۲۰۱٤؛ اولزوسکی ۲۰۱٤ زریپور و شمس المعالی ۲۰۱۵؛ کرا بنسن و همکاران ۲۰۱٦)
مبلغ قبلی	باچمایر ۲۰۰۸؛ گادی و مکاران ۲۰۰۸؛ حلوایی و اکبری ۲۰۱٤
آدرس تراکنش	گادی و همکاران ۲۰۰۸؛ کندو و همکاران ۲۰۰۹؛ حلوایی و اکبری ۲۰۱٤؛ اولروسکی ۲۰۱٤؛ زریپور و شمس المعالی ۲۰۱۵
آدرس قبلی تراکنش	گادی و همکاران ۲۰۰۸؛ حلوایی و اکبری ۲۰۱٤)
تعداد تراکنش در طول زمان	(زاسلاوسکی و استریژاک ۲۰۰٦؛ ویترو و همکاران ۲۰۰۸؛ تاسولیس و آدامز ۲۰۰۸ء باچاریا ۲۰۱۱)
جمع ارزش پولی در طول D ساعت	(زاسلاوسکی و استریژاک ۲۰۰٦؛ ویترو و همکاران ۲۰۰۸؛ تاسولیس و آدامز ۲۰۰۸؛ باچاریا ۲۰۱۱
متوسط ارزش پولی تراکنش	سریواستاوا و همکاران ۲۰۰۸؛ تاسولیس و آدامز ۲۰۰۸؛ باچاریا ۲۰۱۱؛ کومار و راج ۲۰۱۲؛ میشرا و همکاران ۲۰۱۳)
متوسط تعداد تراکنش	ویترو و همکاران ۲۰۰۸؛ تاسولیس و آدامز ۸ – ۲۰: باچاریا ۲۰۱۱)

ماخذ: یافته‌های تحقیق

ادامه جدول ۳ – ویژگی‌های تراکنشی استفاده شده در پژوهش‌های پیشین

ویژگی	منبع
فاصله زمانی با تراکنش قبلی	السکروف و همکاران ۱۹۹۷؛ بنتلی و همکاران ۲۰۰۰؛ گادی و همکاران ۲۰۰۸؛ ویترو و همکاران ۲۰۰۸ باچمایر ۲۰۰۸؛ کندو و همکاران ۲۰۰۹؛ حلوایی و اکبری ۲۰۱٤)
متوسط ساعات انجام تراکنش	(ویترو و همکاران ۲۰۰۸؛ کرابنسن و همکاران ۲۰۱٦)
شهر تراکنش	(زاسلاوسکی و استریژاک ۲۰۰٦؛ گادی و همکاران ۲۰۰۸؛ حلوایی و اکبری ۲۰۱٤)
کشور تراکنش	(زاسلاوسکی و استریاک ۲۰۰٦؛ مادی و همکاران ۲۰۰۸ حلوایی و اکبری ۲۰۱٤)
نوع تراکنش	(زاسلاوسکی و استریژاک ۲۰۰٦؛ گادی و همکاران ۲۰۰۸؛ میشرا و همکاران ۲۰۱۳؛ حلوایی و اکبری ۲۰۱٤)
شماره ترمینال	(زاسلاوسکی و استریژاک ۲۰۰٦؛ گادی و همکاران ۲۰۰۸؛ حلوایی و اکبری ۲۰۱٤)

ماخذ: یافته‌های تحقیق

نکته حائز اهمیت این است که با توجه به متحرک بودن وسایل نقلیه، خبرگان دو ویژگی تعداد ترمینال‌ها در طول بازه زمانی ۱٦۸ ساعت و تعداد شهرها در طول همین بازه زمانی را نیز به داده‌ها اضافه کردند. علاوه بر اینکه تعداد دفعات رسیدن به آستانه نیز مد نظر قرار گرفت. جدول زیرتلفیقی از ویژگی‌های خام و تجمیعی تراکنش‌های سامانه هوشمند سوخت است.

جدول ٤ – ویژگی‌های تجمیعی انتخاب شده برای توصیف رفتار تراکنش

ویژگی	متوسط زمان تراکنش	مقدار تراکنش	جمع مقداری	جمع تعداد تراکنش	فاصله زمانی از تراکنش قبلی	ترمینال	شهر	تعداد ترمینال	تعداد شهرها	تعداد دفعات رسیدن به آستانه
دوره زمانی	٦۰ ساعت	–	٦۰ ساعت	٦۰ ساعت	–	–	–	۱٦۸ ساعت	۱٦۸ ساعت	٦۰ ساعت

ماخذ: یافته‌های تحقیق

با فرض وجود برچسب برای داده‌ها، ۳ ویژگی اضافه شده همراه با ۱۱ ویژگی دیگر با استفاده از روش رتبه‌بندی ویژگی مبتنی بر تئوری اطلاعات و اطلاعات متقابل شرح داده شده در (شهرابی و ذوالقدر شجاعی ۱۳۸۸) است.

جدول ۵ - کدگذاری ویژگی‌های استخراج شده از سامانه هوشمند سوخت

ویژگی	کد	ویژگی	کد	ویژگی	کد
زمان تراکنش	X۱	تعداد تراکنش در طول زمان	X٦	شماره ترمینال	X۱۱
مبلغ تراکنش	X۲	مقدار تراکنش در طول زمان	X۷	تعداد ترمینال ها	X۱۲
مبلغ قبلی	X۳	متوسط مقدار تراکنش	X۸	تعداد شهرها	X۱۳
فاصله زمانی با تراکنش قبلی	X٤	متوسط تعداد تراکنش	X۹	تعداد دفعات رسیدن به آستانه	X۱٤
متوسط ساعات انجام تراکنش	X٥	شهر تراکنش	X۱۰		

ماخذ: یافته‌های تحقیق

جدول ٦ - نتیجه انتخاب ویژگی توسط تئوری اطلاعات در داده‌های سامانه هوشمند سوخت

رتبه	ویژگی
۱	شماره ترمینال
۲	شهر تراکنش
۳	مبلغ تراکنش
٤	مقدار تراکنش در طول زمان
٥	مبلغ قبلی
٦	تعداد تراکنش در طول زمان
۷	فاصله زمانی با تراکنش قبلی
۸	تعداد ترمینال ها
۹	تعداد دفعات رسیدن به آستانه
۱۰	متوسط ساعات انجام تراکنش

ماخذ: یافته‌های تحقیق

جدول فوق ، ۱۰ ویژگی را که روش تئوری اطلاعات به‌عنوان موثرترین برای طبقه‌بندی تقلب‌آمیز، قانونی بودن تراکنش می‌شناسد، به تصویر می‌کشد. با مقایسه اطلاعات این جدول با نتایج حاصل از

خبرگان مشخص می‌شود؛ که تنها ویژگی مبلغ قبلی تراکنش در این روش به دست آمده است. با توجه به بدون سرپرستی بودن مدل پیشنهادی نهایتا در این پژوهش ویژگی‌های انتخاب شده توسط خبرگان مبنا قرار می‌گیرند. در خصوص ابعاد ایستای کارت هوشمند سوخت که نشان‌دهنده ویژگی‌های یک کارت فارغ از تراکنش‌هایش هستند؛ با توجه به رکوردهای موجود در پایگاه داده اجماع نظرات خبرگان، بر روی ویژگی‌های جدول بالا، حاصل شد.از سوی دیگر نشان دهنده ویژگی‌های پویای توصیف کننده یک کارت است، که از بررسی کلیه تراکنش‌هایش به‌صورت تجمیعی حاصل شده است، این ویژگی‌ها نیز حاصل توافق خبرگان است.

جدول ۷ - ویژگی‌های پویای توصیف‌کننده کارت

ویژگی	جمع سهمیه کارکردی	متوسط تجمیعی زمان تراکنش	متوسط مجموع مقداری	متوسط مجموع تعدادی	متوسط فاصله زمانی از تراکنش قبلی	متوسط تعداد ترمینال‌ها	متوسط تعداد شهرها	متوسط تعداد دفعات رسیدن به آستانه
دوره زمانی	-	٦٠	٦٠	٦٠	-	١٦٨	١٦٨	٦٠

ماخذ: یافته‌های تحقیق

- **سیستم بانکی:** برای سنجش بیشتر مدل‌های توسعه یافته در این پژوهش، از داده‌های سیستم بانکی نیز استفاده شده است. البته در اینجا هدف یافتن تقلب در تراکنش‌های دریافتی از پایانه‌های فروش و بر اساس الگوی رفتاری تراکنشی این پایانه‌ها است. بنابراین تعداد ٦٠ پایانه فروش از ٦ کسب و کار در استان‌های مختلف کشور، به صورت تصادفی انتخاب شدند. کلیه ستون‌های اطلاعاتی که در خصوص پایانه‌های یاد شده و تراکنش‌های آنها در سمت پذیرنده نگهداری می‌شوند، به استثنای موارد محرمانه و با رعایت سیاست‌های امنیتی مد نظر و با دادن تعهد عدم افشای اطلاعات از پایگاه داده مربوطه استخراج شدند. تراکنش‌های پایانه‌های فوق از تاریخ اول فروردین ماه ۱۳۹۵ تا پایان اردیبهشت ماه همان سال جمع‌آوری شدند. جدول ۸ خلاصه‌ای از داده‌های اولیه را نشان می‌دهد .

جدول ۸ - اطلاعات استخراج شده از سیستم بانکی

تعداد پایانه	کل تراکنش‌ها
٦٠	١٠١١٨٣٢

ماخذ: یافته‌های تحقیق

لازم به ذکر است، پایانه‌های مد نظر در ٦ دسته کسب و کار به شرح زیر، قرار دارند.

جدول ۹ - توزیع کسب و کاری پایانه‌های فروش

کسب و کار	تعداد	کسب و کار	تعداد
رستوران	۱۰	سوپرمارکت	۱۰
قنادی	۱۰	فروشگاه کالای خاص	۱۰
عمده فروشی	۱۰	سایر فروشگاه ها	۱۰

ماخذ: یافته‌های تحقیق

در اینجا نیز هدف دست‌یابی به ویژگی‌هایی ترکیبی از تراکنش است. لذا این امر هم در گروه کانونی خبرگان مطرح شد، و نهایتا پس از برگزاری یک جلسه کامل، بر روی ویژگی‌های مندرج در جدول زیر در مورد رفتار تراکنشی یک پایانه فروش توافق حاصل شد. نکته قابل توجه این است که کدهای پردازشی تراکنش را می‌توان به سه دسته موفقیت انجام تراکنش، عدم موفقیت و دلایل امنیتی تقسیم‌بندی کرد. از آنجا که دلایل امنیتی توسط سویچ‌های پرداخت شناسایی شده و اقدامات لازم بر روی آنها صورت می‌پذیرند، دیگر در مساله کشف تقلب موضوعیتی ندارند، و از این پژوهش نیز حذف می‌شوند. به‌علاوه در مساله کشف تقلب پذیرنده‌ها مقدار شهر، کشور، شماره ترمینال، آدرس و آدرس قبلی ثابت است و از مدل حذف می‌شوند. علاوه بر اینکه اطلاعاتی در زمینه مبلغ قبلی تراکنش صورت گرفته توسط کارت در اختیار پذیرنده نیست، بنابراین این ویژگی نیز حذف می‌شود. در خصوص ویژگی‌های باقی‌مانده با خبرگان سیستم بانکی جلسه گروه کانونی برقرار شد، که نتیجه آن در جدول زیر، گزارش شده است. برهمین اساس دو ویژگی تعداد تراکنش موفق و تعداد تراکنش ناموفق دوره‌ای نیز به مدل افزوده شد.

جدول ۱۰ - ویژگی‌های تجمیعی توافق شده توسط خبرگان در مورد رفتار تراکنشی پایانه فروش

ویژگی	متوسط زمان تراکنش	مقدار تراکنش	جمع مقداری تراکنش‌ها	جمع تعدادی تراکنش‌ها	فاصله زمانی از تراکنش قبلی	تعداد تراکنش موفق دوره‌ای	تعداد تراکنش ناموفق‌دوره‌ای	زمان تراکنش
دوره زمانی	٦٠	-	٦٠	٦٠	-	٦٠	٦٠	-

ماخذ: یافته‌های تحقیق

با توجه به اینکه در خصوص داده‌های سیستم بانکی اولا مبنای برچسب‌گذاری نظر خبرگان است. و ثانیهٔ ویژگی‌های انتخاب شده تقریبا همه ویژگی‌های باقی‌مانده موجود در ادبیات را در بر گرفته‌اند، لذا از روش کمی برای انتخاب ویژگی در این مورد استفاده نشده است. در خصوص ابعاد ایستای پایانه فروش که نشان‌دهنده ویژگی‌های یک پایانه فارغ از تراکنش‌هایش هستند؛ با توجه به رکوردهای موجود در پایگاه داده حاصل شد.

جدول ۱۱ - ویژگی‌های پویای توصیف کننده رفتار پایانه فروش

ویژگی	متوسط تجمیعی زمان تراکنش	متوسط تجمیعی مقدار تراکنش	متوسط جمع مقداری تراکنش‌ها	متوسط جمع تعدادی تراکنش‌ها	متوسط فاصله زمانی از تراکنش قبلی	متوسط تعداد تراکنش موفق دوره‌ای	متوسط تعداد تراکنش ناموفق دوره‌ای
دوره زمانی	٦٠	-	٦٠	٦٠	-	٦٠	٦٠

ماخذ: یافته‌های تحقیق

پیش‌پردازش داده

در این مرحله هر دو مجموعه داده‌ای برای مدل‌سازی آماده می‌شوند.

-پیش‌پردازش داده‌های سامانه هوشمند سوخت

اولین نکته‌ای که در مورد داده‌های سامانه هوشمند سوخت وجود دارد این است که در این تراکنش‌ها مقدار تهی وجود ندارد، لذا مساله با مقادیر تهی روبرو نیست. به‌علاوه برای دست‌یابی به اثربخشی مدل‌سازی بر اساس نظر زاسلاوسکی و استریژاک (۲۰۰٦)، کارت‌هایی که زیر ۵۰ تراکنش در بازه مدنظر داشتند، حذف شدند. نکته دیگری که می‌تواند بر عملکرد الگوریتم‌ها تاثیر گذارد، وجود

داده‌های پرت در مجموعه آموزش خصوصا در مورد تراکنش‌های قانونی است. روشی که در این پژوهش استفاده شده است، قاعده چبیشف مندرج در معادله ۱ است (لاروس ۲۰۰۵).

$$Outlier \in \{\forall x | x > \mu + ۳\delta \cup x < \mu - ۳\delta\}$$ معادله ۱

یعنی همه داده‌هایی که از نظر ویژگی‌های مد نظر، بیشتر یا کمتر از سه انحراف معیار با میانگین فاصله دارند حذف می‌شوند. بررسی داده‌های سامانه هوشمند سوخت نشان داد که دو ویژگی مقدار تراکنش و اخیر بودن دارای ضریب چولگی زیادی هستند.

جدول ۱۲- مقادیر ویژگی‌های دارای پراکندگی در داده‌های سامانه هوشمند سوخت

ویژگی	میانگین	انحراف معیار	ضریب چولگی
مقدار (لیتر)	۱٤٦	۱۰۲.٦۸	٤.۷۳
فاصله زمانی با تراکنش قبلی	۱.۸۰۸	۲.٥۹۸	۸.۲۲۰

ماخذ: یافته‌های تحقیق

تراکنش‌های قانونی که در فاصله بیشتر از ۳ انحراف معیار از میانگین دو ویژگی فوق بودند، حذف شدند. حاصل در جدول زیر ذکر شده است.

جدول ۱۳ - داده‌های سامانه هوشمند سوخت، پس از عملیات تمیز کردن

تعداد کارت	کل تراکنش‌ها	تراکنش‌های قانونی	تراکنش‌های تقلب‌آمیز
۸۹۰۰	٥۸۰٥۰۸	٥٦۸۸۳۸	۱۱٦۷۰

ماخذ: یافته‌های تحقیق

اکنون داده‌های تراکنشی کارت‌ها آماده مدل‌سازی هستند. از مجموع تراکنش‌های جدول فوق، با توجه به رویکردهای غالب در الگوریتم‌های یادگیری ماشین و نیز نظر خبرگان تقریبا ۹۰ درصد برای ساخت مدل و ۱۰ درصد برای اعتبارسنجی و آزمون طبقه‌بندی مورد استفاده قرار خواهند گرفت. جدول زیر، توزیع قانونی و تقلب‌آمیز بودن را در خصوص هر دو مجموعه داده‌ای مدل‌سازی و طبقه‌بندی نشان می‌دهد.

جدول ۱٤ - توزیع انواع تراکنش در مجموعه‌های مدل‌سازی و طبقه‌بندی

نوع مجموعه داده‌ای	تعداد کل	قانونی	تقلب‌آمیز
مدل‌سازی (آموزش)	٥۲۷۲۳٤	٥۱٦٦۸٦	۱۰٥٤۸
طبقه‌بندی (آزمون)	٥۳۲۷٤	٥۲۱٥۲	۱۱۲۲

ماخذ: یافته‌های تحقیق

-پیش‌پردازش داده‌های سیستم بانکی

در خصوص تراکنش‌های استخراج شده از سیستم بانکی، نخستین نکته آن است که بر طبق نظر خبرگان، مهم‌ترین کد تراکنشی که احتمال تقلب در آن وجود دارد و باید تحت بررسی قرار گیرد، کد تراکنشی مربوط به خرید است. بنابراین تنها تراکنش‌های این کد در نظر گرفته شده و بقیه از تحلیل حذف شدند. علاوه بر اینکه در این داده‌ها نیز ارزش تهی وجود ندارد. در خصوص مقادیر پرت در این مجموعه داده‌ای، همه ویژگی‌ها به جز مبلغ دارای ضریب چولگی نزدیک به نرمال بودند.

جدول ۱۵- مقادیر ویژگی‌های دارای پراکندگی در سیستم بانکی

ویژگی	میانگین	انحراف معیار	ضریب چولگی
مبلغ	۲۱٦۱۷۷	٤۱۰۰۸۱۹	۱۸۰.٤۹

ماخذ: یافته‌های تحقیق

جدول زیر تراکنش‌های باقی‌مانده این سیستم را نشان می‌دهد.

جدول ۱٦- تعداد تراکنش‌های بانکی پس از عملیات تمیز کردن

تعداد پایانه	کل تراکنش‌ها
٦۰	٥۳۱۹۲۷

ماخذ: یافته‌های تحقیق

اما نکته‌ای که قبلا نیز به آن اشاره شد این بود که در داده‌های سیستم بانکی تراکنش بر چسب دار وجود نداشت، لذا با مطرح کردن این موضوع در گروه کانونی خبرگان، چند قاعده به شرح جدول زیر، برای تولید تعدادی تراکنش تقلب‌آمیز برچسب دار با نرخ حداکثر ۲ درصد مورد استفاده قرار گرفتند.

جدول ۱۷ - قواعد نشان‌دهنده ناهنجاری برای برچسب‌گذاری تراکنش‌های سیستم بانکی بر اساس نظر خبرگان

قاعده	شرح
۱	تغییرات قابل توجه در مبلغ تراکنش در فاصله های زمانی کم (۲.۵ انجراف معیار از میانگین)
۲	داشتن تراکنش در ساعات خارج از زمان کسب و کار
۳	افزایش تعداد تراکنش‌های ناموفق (عدم کفایت موجودی، رمز اشتباه)
٤	کاهش فاصله زمانی تراکنش

ماخذ: یافته‌های تحقیق

براساس بررسی‌های خبرگان قواعد جدول بالا به صورت زیر عددی شدند .

جدول ۱۸ - قواعد عددی برچسب گذاری تراکنش‌های سیستم بانکی

قاعده	شرح
۱	بیش از سه برابر شدن یا به سقف رسیدن تراکنش در زمان کمتر از ۱۰ ثانیه
۲	برای هر کسب و کار ساعات پر ریسک مشخص شدند
۳	بیشتر از ۱۰ تراکنش متوالی
٤	کمتر از ٤ ثانیه

ماخذ: یافته‌های تحقیق

با اعمال چهار قاعده فوق بر روی داده‌ها، تعداد ۱۰۷۱۲ تراکنش دارای برچسب تقلب‌آمیز شدند. نهایت، با همان رویکرد تقسیم داده‌های سامانه هوشمند سوخت با نرخ ۹۰ درصد برای مدل‌سازی و ۱۰ درصد برای آزمون طبقه‌بندی، جدول زیرحاصل شد.

جدول ۱۹ - توزیع انواع تراکنش در مجموعه‌های مدل‌سازی و طبقه‌بندی

نوع مجموعه داده‌ای	تعداد	قانونی	تقلب‌آمیز
مدل‌سازی (آموزش)	٤٧٨٧٣٤	٤٦٩٢٣٢	۹۵۰۲
طبقه‌بندی (آزمون)	۵۳۱۹۲	۵۱۹۸۲	۱۲۱۰

ماخذ: یافته‌های تحقیق

مدل‌سازی ارزیابی و کشف تقلب در زنجیره تامین

توجه به مدل‌سازی‌های انجام شده در تحقیقات قبلی، به نظر می‌رسد، استفاده صرف از یک الگوریتم حتی اگر به صورت ترکیبی نیز باشد، با اینکه می‌تواند الزامات بزرگ داده را ارضاء کند، اما برای شاخص‌های طبقه‌بندی در مساله مد نظر پژوهش حاضر کافی نیست. زیرا حتی نرخ‌های حاصله از اعلان اشتباه و عدم اعلان اشتباه، در حجم تراکنش‌های کارت می‌تواند سبب کاهش قابل توجه کارایی سیستم کشف تقلب شود. لذا در این مدل‌سازی، استنتاجاتی مستقل براساس منابع مختلف صورت می‌پذیرد. نتیجه هر یک از استنتاج‌ها نیز با استفاده از رویکرد هم‌جوشی ترکیب می‌شود. با این مقدمه، پیشنهاد پژوهش حاضر از یک سو، استفاده از منطق فازی برای یافتن راه‌حلی در جهت کاهش ابهام است. از سوی دیگر سه منبع جداگانه برای طبقه‌بندی تراکنش جدید تعریف می‌شوند. منبع نخست، همان پروفایل ایجاد شده توسط تراکنش‌های تاریخی موجودیت مالک تراکنش تا زمان t_n است. برای

ساخت این منبع از الگوریتم ساخت پروفایل براساس مدل مخفی مارکوف و خوشه‌بندی چندبعدی استفاده شده و ارزش $\theta_{۱}(v) = P(O|\lambda)$ مطابق الگوریتم یاد شده محاسبه و به همراه مدل آموزش‌دیده مارکوف مخفی و ماتریس خوشه‌های تراکنش نگهداری می‌شود.

منبع دوم، استفاده از رفتار موجودیت‌هایی است که بر اساس ویژگی‌های ایستا، به موجودیت مد نظر شباهت دارند. بنابراین در اینجا نخست یک خوشه‌بندی میان همه موجودیت‌ها فارغ از رفتار تراکنشی آنها، با ابعادی که در طول زمان یا تغییر نمی‌کنند یا میزان تغییر کمی دارند، صورت می‌پذیرد. در گام بعدی رفتار تراکنشی هر یک از خوشه‌ها که تجمیعی از رفتار تراکنشی موجودیت‌های تشکیل‌دهنده آن است، به‌عنوان پروفایل هر خوشه ایجاد می‌شود. این منبع را می‌توان یک منبع سطح بالاتر در نظر گرفت. در هنگام مقایسه نخست خوشه موجودیت مالک (به‌عنوان مثال دارنده کارت) تعیین، سپس تراکنش ورودی با پروفایل این خوشه سنجیده می‌شود. خوشه‌بندی موجودیت‌ها با استفاده از الگوریتم k-means انجام می‌شود. پروفایل سازی رفتار خوشه‌ها نیز به وسیله الگوریتم کوهونن نگاشت کاهشی و با استفاده از همه تراکنش‌های موجودیت‌های تشکیل‌دهنده یک خوشه صورت می‌پذیرد. سنجش فاصله تراکنش ورودی با پروفایل خوشه نیز با الگوریتم پیشنهادی انجام می‌شود.

منبع سوم، مشابه منبع دوم استفاده از رفتار موجودیت‌هایی است که با موجودیت مالک تراکنش، شباهت رفتاری دارند. در اینجا از خوشه‌بندی پویا، یعنی خوشه‌بندی بر اساس ویژگی‌هایی که در طول زمان تغییر می‌کنند، استفاده می‌شود. برای این منبع نیز مانند منبع قبلی عمل می‌شود؛ یعنی یک گام تشکیل خوشه‌هایی متشکل از مالکین تراکنش، و در گام دوم تشکیل پروفایل‌هایی با استفاده از تراکنش‌های موجودیت‌های تشکیل‌دهنده یک خوشه برای خوشه مد نظر. الگوریتم‌ها نیز به همان ترتیب منبع دوم هستند.

۱- برای همه مالکین تراکنش‌های کارت، خوشه‌بندی ایستا و پویا در قالب دو فرآیند مستقل و موازی انجام می‌شود .

معادله ۲
$$k-means^{parallel}(D_{static}) \in \{v_j \mid j = ۱, \dots, k_s\}$$

در معادله ۲، Dstatic ماتریس ابعاد ایستای همه موجودیت‌های مالک تراکن و ks تعداد خوشه‌های ایستای مد نظر و از پیش تعریف شده است.

معادله ۳
$$k-means^{parallel}(D_{dynamic}) \in \{v_j \mid j = ۱, \dots, k_d\}$$

در معادله ۳،Daynamic ماتریس ابعاد ایستای همه موجودیت‌های مالک تراکنش و kd تعداد خوشه‌های ایستای مد نظر و از پیش تعریف شده است.

۲- برای همه خوشه‌های ایستا (k_s) و پویا (k_d)، به صورت موازی و با ماتریس های تراکنشی تجمیعی و نرمال شده ماتریس تراکنش‌های مالکین متعلق به خوشه k اجرای موازی معادلات:

معادله ٤ $$SOM^{parallel}\ (P^j) = |W^j|\ ; j = ۱, \dots, k_s$$

معادله ٥ $$SOM^{parallel}\ (P^j) = |W^j|\ ; j = ۱, \dots, k_d$$

۳- برای دارنده تراکنش کارت i تا زمان tn ساخت پروفایل براساس مدل مخفی مارکوف و خوشه‌بندی چندبعدی به صورت موازی اجرا، و مقدار $\theta_۱\ (v) = P(O|\lambda)$ محاسبه می‌شود .

پایان

چارچوب تشخیص تخمین قانونی یا تقلب‌آمیز بودن تراکنش ورودی است. معادله زیر این چارچوب را نشان می‌دهد. در این معادله؛ θ_f احتمال تقلب‌آمیز بودن و $\theta_۱$ احتمال قانونی بودن تراکنش جدید است .

معادله ٦ $$\theta = \{\theta_f\ , \theta_l\}$$

حال برای اینکه هر یک از طبقه‌بندهای توضیح داده شده در فوق تبدیل به یک تابع انتساب باور پایه شوند، مجموعه معادله ٦ باید تبدیل به مجموعه توانی مطابق با معادله ۷ شود:

معادله ۷ $$\{\emptyset\ , \theta_۱\ , \theta_f, \{\theta_۱\ , \theta_f\}\ \}$$

حال هر یک از طبقه‌بندها که بتوانند نگاشتی از مجموعه معادله ٤-٥۷ به فاصله $[۰,۱]$ ایجاد کنند، به شرطی که :

الف - برای هر یک از گزاره ها، باور مقداری بزرگتر یا مساوی صفر داشته :

معادله ۸ $$\forall\ A\ \in\ ۲^{\theta}\ \rightarrow m(A)\ \geq\ ۰$$

ب- هیچ درجه ای از باور برای گزاره تهی در نظر گرفته نشود :

معادله ۹ $$m(\emptyset) = ۰$$

ج- مجموع کل باور برابر با یک باشد :

معادله ۱۰ $$\sum_{A \in ۲\theta} m(A) = ۱$$

یک تابع انتساب باور پایه خواهد بود. از سوی دیگر روش استفاده شده در این پژوهش دمپستر شیفر فازی است. بنابراین خروجی هریک از طبقه‌بندها که میزانی احتمالی است، باید به فرم فازی تبدیل

شود. تا در گام بعد ترکیب نتایج، اندازه‌گیری فاصله های باور و مقبولیت، وافازی سازی و نهایتا استنتاج نهایی صورت پذیرد. نخستین طبقه‌بندی که با اندکی تغییر این ویژگی را خواهد داشت الگوریتم است.

الگوریتم پیشنهادی برای طبقه‌بندی تراکنش ورودی براساس مدل مخفی مارکوف است. در این الگوریتم یک نرخ احتمال اولیه $\theta_1\ (v) = P(O|\lambda)$ برای توالی پروفایل محاسبه می‌شود، حال اگر نرخ احتمال توالی جدید، $\theta_2\ (v) = P(O^{new}|\lambda)$ کمتر از مقدار احتمال قبلی به اندازه یک آستانه بود، تراکنش تقلب‌آمیز و در غیر این‌صورت، قانونی طبقه‌بندی می‌شود. از سوی دیگر، در فاصله احتمال θ_1 تا میزان آستانه δ مقداری ناباوری وجود دارد که مربوط به احتمال قانونی یا تقلب‌آمیز بودن، توام است. با کمی تطبیقی تر کردن و فازی سازی استنتاج فوق می‌توان آن را به یک تابع انتساب باور پایه تبدیل کرد. برای این منظور، مقدار احتمال θ_1 به‌عنوان احتمال اولیه، θ_2 به‌عنوان میزان احتمال توالی مورد آزمون، و θ_{min} به‌عنوان کمترین احتمال پذیرفته شده، در نظر گرفته می‌شوند. اعداد فازی معادلات تشکیل یک تابع عضویت فازی را می دهند که خروجی مدل مارکوف مخفی، الگوریتم پیشنهادی برای طبقه‌بندی تراکنش ورودی براساس مدل مخفی مارکوف را تبدیل به یک تابع باور پایه می‌کند.

معادله ۱۱
$$m_1(\theta_f) = \mu_{\theta_f}(\theta_2) = \begin{cases} 1, & \theta_2 < \theta_{min} \\ \frac{\theta_1 - \theta_2}{\theta_1 - \theta_{min}}, & \theta_{min} \le \theta_2 < \theta_1 \\ 0, & \theta_2 \ge \theta_1 \end{cases}$$

معادله ۱۲
$$m_1(\theta_1) = \mu_{\theta_1}(\theta_2) = \begin{cases} 0, & \theta_2 \le \theta_{min} \\ \frac{\theta_2 - \theta_{min}}{\theta_1 - \theta_{min}}, & \theta_{min} \le \theta_2 < \theta_1 \\ 1, & \theta_2 \ge \theta_1 \end{cases}$$

و نهایتا معادله ۱۳
$$m_1(\{\theta_1, \theta_f\}) = 1 - \left[\mu_{\theta_1}(\theta_2) + \mu_{\theta_f}(\theta_2)\right]$$

در خصوص مدل شبکه عصبی کوهونن، الگوریتم پیشنهادی برای سنجش فاصله تراکنش ورودی با پروفایل دارنده کارت، که برای هر دو استنتاج نهایی ایستا و پویا مورد استفاده قرار می‌گیرد، نیز تبدیل آن به یک تابع باور پایه در ادامه آمده است. همان‌گونه که قبلا عنوان شد، در روش پیشنهادی، پارامتر یا به عبارتی فاصله نامتشابه‌ترین نرون با همسایه‌هایش با مرکز توری شبکه کوهونن حاصل می‌شود، نیز میانگین فاصله نامتشابه‌ترین نرون با همسایه‌هایش است محاسبه می‌شود. و نهایتا فاصله بردار

تراکنش ورودی P_{n+1} با نزدیک‌ترین نرون به آن است که از طریق فاصله اقلیدسی سنجیده شده و با dist نشان داده می‌شود. با داشتن این پارامترها و اعداد فازی معادلات، خروجی الگوریتم کوهونن مد نظر به یک تابع انتساب باور پایه تبدیل می‌شود.

$$m_2(\theta_f) = \mu_{\theta_f}(dist) = \begin{cases} 0, & dist < \tau - \rho \\ \frac{dist - \tau + \rho}{2\rho}, & \tau - \rho \le dist < \tau + \rho \\ 1, & dist \ge \tau + \rho \end{cases}$$ معادله ۱۴

$$m_2(\theta_1) = \mu_{\theta_1}(dist) = \begin{cases} 1, & dist < \tau - \rho \\ \frac{\tau + \rho - dist}{2\rho}, & \tau - \rho \le dist < \tau + \rho \\ 0, & dist \ge \tau + \rho \end{cases}$$ معادله ۱۵

$$m_2(\{\theta_1, \theta_f\}) = 1 - \left[\mu_{\theta_1}(dist) + \mu_{\theta_f}(dist)\right]$$ و نهایتا معادله ۱۶

لازم به ذکر است که تابع باور پایه رm نیز دقیقا مشابه m با پارامترهای متفاوت است. با داشتن ارزش های توابع باور پایه فوق، در درجه اول برای دو منبع خوشه‌ای ایستا و پویا، با استفاده از قاعده میانگین‌گیری معادله ۱۷ ترکیب نتایج خوشه‌ها با وزن‌های از پیش تعریف شده توسط خبرگان انجام می‌شود.

$$m_{1\dots n} = \frac{1}{n} \sum_{i=1}^{n} w_i m_i$$ معادله ۱۷

سپس نتیجه حاصله با نتیجه حاصل از سنجش سابقه تراکنشی موجودیت مالک تراکنش از طریق قاعده ترکیب دمپستر شیفر موجود در معادله ۱۸ ترکیب می‌شود.

$$m(A) = \frac{\sum_{c_i \cap c_j = A} m_1(C_i)\, m_2(C_J)}{1 - \sum_{c_i \cap c_j = \emptyset} m_1(C_i)\, m_2(C_J)}, A \neq \emptyset$$ معادله ۱۸

مقدار ترکیبی احتمال هریک از زیرمجموعه‌های مجموعه۰ به جز مقدار ت‌هی، محاسبه می‌شود. سپس با استفاده از معادله ۱۹ برای اندازه‌گیری میزان باور هر گزاره و معادله ۲۰ برای اندازه‌گیری میزان مقبولیت هر گزاره استفاده می‌شود .

$$Bel(A) = \sum_{c_i \in 2^\theta, C_i \subset A} m(C_i)$$ معادله ۱۹

$$Pls(A) = \sum_{C_i \cap A \neq \emptyset} m(C_i)$$ معادله ۲۰

نهایتا مقادیر وافازی شده و حال آن گزاره ای که بازه باور تا مقبولیت بیشتری داشت به‌عنوان طبقه تراکنش ورودی در نظر گرفته می‌شود.

د- الگوریتم پیشنهادی برای طبقه‌بندی تراکنش براساس رویکرد موازی گواه فازی

آغاز

۱ -برای بردار ورودی P_{n+1}^{i} (i شماره موجودیت مالک تراکنش کارت در زمان ۱(tn+)، موارد زیر به‌صورت سه فرآیند نگاشت موازی سطح بالا اجرا می‌شوند .

محاسبه احتمالات همه حالات تصمیم بر اساس الگوریتم مارکوفی موازی.

تابع انتساب باور توضیح داد شده برای تراکنش‌های تاریخی موجودیت مالک .

محاسبه احتمالات همه حالات تصمیم بر اساس شبکه عصبی کوهونن موازی و تابع انتساب باور توضیح داد شده بر روی پروفایل خوشه استاتیک موجودیت مالک

محاسبه احتمالات همه حالات تصمیم، براساس شبکه عصبی کوهونن موازی و تابع انتساب باور توضیح داد شده بر روی پروفایل خوشه دینامیک موجودیت مالک

۲- ترکیب گواه و محاسبه مقادیر باور و مقبولیت همه حالات تصمیم در قالب یک فرآیند کاهش

۳- استنتاج بر اساس مقادیر باور و مقبولیت حالات

- اگر بیشترین باور مربوط به قانونی یا تقلب‌آمیز باشد: برچسب بر اساس آن انتخاب می‌شود.
- در غیر این‌صورت: برچسب بیشترین مقبولیت در نظر گرفته می‌شود

در این بخش برای اعتبارسنجی مدل‌های گام قبل، برای هر دو مجموعه داده‌ای، ابتدا پروفایل موجودیت‌ها بر اساس الگوریتم جدول ٤-٦٨ مدل موازی پیشنهادی برای ساخت پروفایل‌های چندگانه برای مالکین تراکنش‌های کارت و با تعیین تعداد خوشه‌های ایستا و پویا، و همان طول توالی، تعداد حالات مخفی و تعداد خوشه‌های تراکنشی مدل مارکوفی با خوشه‌بندی چندگانه معرفی شده در مرحله قبل مدل‌سازی ساخته می‌شود. سپس با استفاده از الگوریتم پیشنهادی برای طبقه‌بندی تراکنش براساس رویکرد موازی گواه فازی، تراکنش‌های جدید، طبقه‌بندی می‌شوند. با توجه به اینکه در مدل پیشنهادی دو خوشه‌ایستا و پویا وجود دارد، تعیین تعداد این خوشه‌ها دو پارامتر است که باید از قبل به مدل معرفی شوند. در خصوص سامانه هوشمند سوخت، برای تعیین عدد خوشه‌ها برای ۸۹۰۰ کارت نخست با استفاده از ویژگی‌های ایستا کارت‌ها به وسیله الگوریتم x-means و با تعداد خوشه‌های ۲ تا ۲۰ خوشه‌بندی شدند. نتیجه ۱۰ خوشه‌ایستا شد.

جدول ۲۰ – نتیجه خوشه‌بندی ایستای داده‌های سامانه هوشمند سوخت

شماره خوشه	تعداد مشاهده
۱	۱۱۰۱
۲	۷٦۹
۳	۸۹۳
٤	۸٦٥
٥	٦۹۱
٦	۱۰۷۲
۷	۷۳۹
۸	۹۰٦
۹	۹٥۱
۱۰	۹۱۳

ماخذ: یافته‌های تحقیق

جدول ۲۱ – نتیجه خوشه‌بندی پویای داده‌های سامانه هوشمند سوخت

شماره خوشه	تعداد مشاهده
۱	٦۱۳
۲	٦٦۹
۳	٥۹۳
٤	۷۷۰
٥	٦۹۸
٦	٦۱۱
۷	۷٦۲
۸	۹۰٦

ماخذ: یافته‌های تحقیق

ادامه جدول ۲۱ – نتیجه خوشه‌بندی پویای داده‌های سامانه هوشمند سوخت

شماره خوشه	تعداد مشاهده
۹	۹۵۱
۱۰	۹۱۳
۱۱	۲۰۸
۱۲	٥١٤
۱۳	٦٩٢

ماخذ: یافته‌های تحقیق

بنابراین با مشخص شدن این دو پارامتر و اینکه بقیه پارامترها مطابق روال قبل تخمین زده می‌شوند. وزن‌های ترکیب خوشه‌های ایستا و پویا نیز به ترتیب وزن ۰.۸ و ۰.۲ از نظر خبرگان انتخاب شد. پارامترهای لازم برای ایجاد پروفایل‌های سه‌گانه برای داده‌های سامانه هوشمند سوخت به شرح جدول زیر است.

جدول ۲۲ – پارامترهای مدل ترکیبی پیشنهادی برای داده‌های سامانه هوشمند سوخت

پارامتر	شرح	ارزش
k_s	تعداد خوشه‌های ایستا	۱۰
k_d	تعداد خوشه‌های پویا	۱۳
M	تعداد ارزش های انتشار	٦
N	تعداد حالات مخفی	۱۰
R	طول توالی مشاهدات	۱٥
L	طول توالی آموزش	٥۰
Q	تعداد نرون های کوهنن	٤٩
W_s	وزن خوشه‌ایستا	۰.۸
Wd	وزن خوشه پویا	۰.۲

ماخذ: یافته‌های تحقیق

در این مرحله با در نظر گرفتن پارامترهای مندرج در جدول ۲۲، نخست ٤ مثال تجزیه و تحلیل شده و سپس عملکرد کلی مدل مورد بحث قرار می‌گیرد.

مطالعه موردی ۱- این مثال مربوط به تراکنش ورودی کارتی است که با پروفایل‌های ساخته شده بر مبنای هر سه رفتار تراکنشی کارت مزبور، رفتار خوشه‌ایستایی که کارت به آن تعلق دارد، و رفتار خوشه پویای کارت‌هایی که مانند کارت مورد نظر رفتار کرده بودند، فاصله داشته و از نظر هرسه تابع انتساب باور پایه تقلب‌آمیز ارزیابی شده است. جدول ۲۳ مقادیر انتساب باور پایه حاصل از اجرای مدل بر روی مثال ۱ را به تصویر می‌کشد.

جدول ۲۳ - نتیجه انتساب باور پایه سه‌گانه

منبع	تقلب‌آمیز	قانونی	ناباوری
رفتار تراکنشی موجودیت	۰.۸۹	۰.۱	۰.۰۱
رفتار خوشه‌ایستا	۰.۸۳	۰.۱۵	۰.۰۲
رفتار خوشه پویا	۰.۹۵	۰.۰٤	۰.۰۱

ماخذ: یافته‌های تحقیق

با استفاده از روش میانگین‌گیری وزنی نتایج رفتار دو خوشه‌ایستا و پیا به شرح جدول زیر است.

جدول ۲٤ - ترکیب وزنی باورهای خوشه‌ای

منبع	درجه باور پایه			ضریب وزنی
	تقلب‌آمیز	قانونی	ناباوری	
رفتار خوشه‌ایستا	۰.۸۳	۰.۱۵	۰.۰۲	۰.۸
رفتار خوشه پویا	۰.۹۵	۰.۰٤	۰.۰۱	۰.۲
نتیجه هم‌جوشی	۰.۸٥٤	۰.۱۲۸	۰.۰۱۸	-

ماخذ: یافته‌های تحقیق

نهایتا با استفاده از ترکیب گواه و محاسبه باور (بدبینانه) و مقبولیت خوشبینانه) استنتاج نهایی به‌صورت تقلب‌آمیز در نظر گرفته می‌شود. لازم به ذکر است بر چسب این مطالعه موردی تقلب‌آمیز است.

جدول ۲۵- استنتاج نهایی بر اساس ترکیب گواه

تراکنشی / ایستا – پویا		تقلب‌آمیز	قانونی	ناباوری
		۰.۸۵٤	۰.۱۲۸	۰.۰۱۸
تقلب‌آمیز	۰.۸۹	۰.۷٦	۰.۱۱٤	۰.۰۱٦
قانونی	۰.۱	۰.۰۸۵	۰.۰۱۳	۰.۰۰۲
ناباوری	۰.۰۱	۰.۰۰۸۵	۰.۰۰۱	۰
باور	–	۰.۷۸۵	۰.۰۱٦	۰.۲
مقبولیت	–	۰.۹۸٤	۰.۲۱۵	–
نتیجه گیری		تقلب‌آمیز		

ماخذ: یافته‌های تحقیق

مطالعه موردی ۲- این مطالعه موردی مربوط به تراکنش ورودی کارتی است که با پروفایل ساخته شده بر مبنای رفتار تراکنشی فاصله نسبتا کمی دارد. اما با پروفایل‌های ناشی از دو منبع خوشه‌ای این فاصله زیاد دارد. جدول ۲٦ مقادیر انتساب اور پایه حاصل از اجرای مدل بر روی مطالعه موردی ۲ را نشان می‌دهد.

جدول ۲٦ - نتیجه انتساب باور پایه سه‌گانه

منبع	تقلب‌آمیز	قانونی	ناباوری
رفتار تراکنشی موجودیت	۰.٤٦	۰.۲۷	۰.۲۷
رفتار خوشه‌ایستا	۰.۹۱	۰.۰۹	۰.۰
رفتار خوشه پویا	۰.۸۳	۰.۱۵	۰.۰۲

ماخذ: یافته‌های تحقیق

با استفاده از روش میانگین‌گیری وزنی نتایج رفتار دو خوشه‌ایستا و پویای مطالعه موردی ۲ به شرح جدول زیر است.

جدول ۲۷ – ترکیب وزنی باورهای خوشه‌ای

منبع	تقلب‌آمیز	قانونی	ناباوری	ضریب وزنی
رفتار خوشه‌ایستا	۰.۹۱	۰.۰۹	۰.۰	۰.۸
رفتار خوشه پویا	۰.۸۳	۰.۱۵	۰.۰۲	۰.۲
نتیجه هم‌جوشی	۰.۸۹٤	۰.۱۰۲	۰.۰۰٤	–

ماخذ: یافته‌های تحقیق

نهایتا با استفاده از ترکیب گواه و محاسبه باور (بد بینانه) و مقبولیت خوشبینانه) تراکنش مطالعه موردی ۲ به‌عنوان تقلب‌آمیز طبقه‌بندی می‌شود. لازم به ذکر است بر چسب این مطالعه موردی نیز تقلب‌آمیز است.

مطالعه موردی ۳ – مطالعه موردی ۳ مربوط به تراکنش ورودی کارتی است که از نظر رفتار تاریخی تراکنش در نقطه میانی قرار دارد، یعنی عملا در این شرایط امکان استنتاج وجود ندارد. اما فاصله با دو پروفایل دیگر میل بیشتر به سمت طبقه قانونی را نشان می‌دهد. جدول ۲۸ مقادیر انتساب باور پایه حاصل از اجرای مدل بر روی این مطالعه موردی را به تصویر می‌کشد.

جدول ۲۸ – نتیجه انتساب باور پایه سه‌گانه

منبع	تقلب‌آمیز	قانونی	ناباوری
رفتار تراکنشی موجودیت	۰.۵۱	۰.٤۷	۰.۰۲
رفتار خوشه‌ایستا	۰.۰۹	۰.۵٦	۰.۳۵
رفتار خوشه پویا	۰.۱٦	۰.٦٦	۰.۱۸

ماخذ: یافته‌های تحقیق

مطابق قبل، با استفاده از میانگین‌گیری وزنی نتایج رفتار دو خوشه‌ایستا و پویا به شرح جدول ۲۹ است .

جدول ۲۹ – ترکیب وزنی باورهای خوشه‌ای

منبع	تقلب‌آمیز	قانونی	ناباوری	ضریب وزنی
رفتار خوشه‌ایستا	۰.۰۹	۰.۵٦	۰.۳۵	۰.۸
رفتار خوشه پویا	۰.۱٦	۰.٦٦	۰.۱۸	۰.۲
نتیجه هم‌جوشی	۰.۱۰٤	۰.۵۸	۰.۳۱٦	–

ماخذ: یافته‌های تحقیق

مطالعه موردی ٤ – مطالعه موردی ٤ مربوط به تراکنش ورودی کارتی است که از نظر رفتار تاریخی تراکنش با عضویت کمی در منطقه قانونی قرار دارد. اما فاصله با دو پروفایل دیگر میل بیشتر به سمت طبقه قانونی را نشان می‌دهد. جدول ۳۰ مقادیر انتساب باور پایه حاصل از اجرای مدل بر روی این مطالعه موردی را به تصویر می‌کشد.

جدول ۳۰ – نتیجه انتساب باور پایه سه‌گانه

منبع	تقلب‌آمیز	قانونی	ناباوری
رفتار تراکنشی موجودیت	۰.۱۲	۰.٤۹	۰.۳۹
رفتار خوشه‌ایستا	۰.۲۳	۰.۵۸	۰.۱۹
رفتار خوشه پویا	۰.۲۸	۰.٦۱	۰.۱۱

ماخذ: یافته‌های تحقیق

با استفاده از روش میانگین‌گیری وزنی نتایج رفتار خوشه‌ای مطالعه موردی ٤ به شرح جدول ۳۱ است.

جدول ۳۱ – ترکیب وزنی باورهای خوشه‌ای

منبع	تقلب‌آمیز	قانونی	ناباوری	ضریب وزنی
رفتار خوشه‌ایستا	۰.۲۳	۰.۵۸	۰.۱۹	۰.۸
رفتار خوشه پویا	۰.۲۸	۰.٦۱	۰.۱۱	۰.۲
نتیجه هم‌جوشی	۰.۲٤	۰.۵۸٦	۰.۱۷٤	–

ماخذ: یافته‌های تحقیق

نهایتا با استفاده از ترکیب گواه، محاسبه باور و مقبولیت، تراکنش ورودی مطالعه موردی ٤ به‌عنوان قانونی طبقه‌بندی می‌شود. لازم به ذکر است بر چسب این مطالعه موردی قانونی است.

همان‌گونه که در مطالعه موردی‌های ٤گانه فوق نمایان است؛ ورود دو منبع توانسته است، استنتاج کامل‌تری را در خصوص تعیین طبقه تراکنش ورودی انجام دارد. این استنتاج، خصوصا در شرایط قرار داشتن در مرز طبقات می‌تواند تصمیم با ابهام کمتر و کامل‌تری را انجام دهد.. همان‌گونه که در این جدول مشاهده می‌شود، میزان فراخوانی با بهبودی قابل ملاحظه به ۹۱ درصد، افزایش و نرخ اعلان نادرست به ٤ درصد کاهش یافته است. مقدار F به ٤۷ درصد و شاخص g-mean , نیز به میزان ۹۳ درصد ارتقا یافته است .

جدول ۳۲ - نتیجه آزمون مدل پیشنهادی ترکیبی در داده‌های سامانه هوشمند سوخت

روش	فراخوانی	اعلان نادرست	F	g-mean۲
پیشنهادی	۹۱	٤	٤٧	۹۳

ماخذ: یافته‌های تحقیق

جدول ۳۳، نیز ماتریس درهم ریختگی این اجرا وجود دارد. همانطور که مشاهده می‌شود؛ تعداد عدم اعلان نادرست و اعلان نادرست به شدت کاهش یافته است.

جدول ۳۳ - ماتریس در هم ریختگی اجرای مدل پیشنهادی بر روی داده‌های سامانه هوشمند سوخت

	شرایط درست	شرایط نادرست
پیش‌بینی مثبت	۱۰۲۱	۲۲۳٤
پیش‌بینی منفی	۹۰	٤۹۹۱۸

ماخذ: یافته‌های تحقیق

عملیات فوق برای ٦۰ ترمینال سیستم بانکی تکرار می‌شود. یعنی یک بار تعداد خوشه‌های ایستا و یک بار تعداد خوشه‌های پویا تعیین شدند؛ که نتایج به شرح جداول زیر است.

جدول ۳٤ - نتیجه خوشه‌بندی ایستای داده‌های سیستم بانکی

شماره خوشه	تعداد مشاهده
۱	۹
۲	۱۳
۳	۱۰
٤	۱۲
٥	۱٦

ماخذ: یافته‌های تحقیق

جدول ۳٥ - با نتیجه خوشه‌بندی پویای داده‌های سیستم بانکی

شماره خوشه	تعداد مشاهده
۱	۱۲
۲	۱۰
۳	۸
٤	۱٦
٥	۱٤

ماخذ: یافته‌های تحقیق

بنابراین با مشخص شدن این دو پارامتر و اینکه بقیه پارامترها مطابق روال قبل تخمین زده می‌شوند؛ وزن‌های ترکیب خوشه‌های ایستا و پویا نیز به ترتیب وزن ۰.۸ و ۰.۲ از نظر خبرگان انتخاب شد. پارامترهای لازم برای ایجاد پروفایل‌های سه‌گانه برای داده‌های سیستم بانکی به شرح جدول ۳۶ است.

جدول ۳۶ - پارامترهای مدل ترکیبی پیشنهادی برای داده‌های سیستم بانکی

پارامتر	شرح	ارزش
k_s	تعداد خوشه‌های ایستا	۵
k_d	تعداد خوشه‌های پویا	۵
M	تعداد ارزش های انتشار	۱۲
N	تعداد حالات مخفی	۱۰
R	طول توالی مشاهدات	۲۵
L	طول توالی آموزش	۱۰۰
Q	تعداد نرون های کوهنن	۱۵۰
W_s	وزن خوشه‌ایستا	۰.۸
Wd	وزن خوشه پویا	۰.۲

ماخذ: یافته‌های تحقیق

همان‌گونه که در این جدول مشاهده می‌شود، میزان فراخوانی با بهبودی قابل ملاحظه به ۸۳ درصد، افزایش و نرخ اعلان نادرست به ٦ درصد کاهش یافته است. مقدار F به ۳۹ درصد و شاخ g-mean , نیز به میزان ۸۹ درصد ارتقا یافته است.

جدول ۳۷ - نتیجه آزمون مدل پیشنهادی ترکیبی در داده‌های سیستم بانکی

روش	فراخوانی	اعلان نادرست	F	g-mean۲
پیشنهادی	۸۳	٦	۳۹	۸۹

ماخذ: یافته‌های تحقیق

جدول زیر، نیز ماتریس در هم ریختگی این اجرا وجود دارد. همان‌طور که مشاهده می‌شود؛ تعداد عدم اعلان نادرست و اعلان نادرست به شدت کاهش یافته است.

جدول ۳۸ - ماتریس در هم ریختگی اجرای مدل پیشنهادی بر روی داده‌های سیستم بانکی

	شرایط درست	شرایط نادرست
پیش‌بینی مثبت	۱۰۰۵	۲۹۵٦
پیش‌بینی منفی	۲۰۵	٤۹۰۲٦

ماخذ: یافته‌های تحقیق

به طور خلاصه نتیجه آزمون مدل ارائه شده نهایی بر روی هر دو مجموعه داده‌ای به صورت زیر خلاصه می‌شود:

- ✓ همه سنجه‌های عملکرد طبقه‌بندی نسبت به روش‌های قبل، بهبود قابل ملاحظه‌ای یافتند. این بهبود خصوصا در میزان اعلان نادرست به خوبی نمایان است.
- ✓ رویکرد هم‌جوشی استفاده شده به خوبی استنتاجات مستقل را به یکدیگر پیوند زده است.
- ✓ با توجه به اجرای موازی، مدل از نظر سرعت اجرا عملکرد قابل قبولی داشته است..

اعتبارسنجی مدل

در این بخش به سنجش اعتبار مدل ارائه شده در مقایسه با مدل‌های پیشین با استفاده از داده‌های استاندارد پرداخته می‌شود. یکی از مهم‌ترین مشکلاتی که در خصوص ارائه مدل‌های کشف تقلب در کارت‌های الکترونیک عنوان شده است، فقدان یا به عبارتی کمبود داده‌های استاندارد و به گزینی برای سنجش کارایی مدل‌های توسعه یافته است (کیبی و جونهوا ۲۰۱۱ گای و همکاران ۲۰۱۱؛ سانجیو، مونتسرات، و وستلند ۲۰۱۲؛ زریپور و شمس المعالی ۲۰۱۵ و کرابنسن و همکاران ۲۰۱٦). بنابراین برای پژوهش حاضر نیز مجموعه داده‌ای که دقیقا بتواند ویژگی‌های موردنیاز مدل را پوشش دهد یافت نشد. همان‌گونه که در بخش پیش شرح داده شد، مدل نیاز به اطلاعات تراکنشی یک موجودیت به همراه اطلاعاتی در خصوص ویژگی‌های آن موجودیت دارد. به طور دقیق‌تر هم داده‌های تراکنشی کارت و هم اطلاعاتی در خصوص دارنده کارت موردنیاز مدل است. نخستین مجموعه داده‌ای که بررسی شد، مجموعه داده‌ای واقعی ارائه شده توسط (گادی و همکاران ۲۰۰۸) است که در کارهای پژوهشی فراوانی به‌عنوان مجموعه استاندارد مورد استفاده قرار گرفته است، رکوردهای جمع‌آوری شده توسط آنها از یکی از بانک‌های برزیلی از تاریخ ۱٤ جولای ۲۰۰٤- تا ۱۲ سپتامبر ۲۰۰٤- استخراج شده است، که تقلب‌هایی مانند دزدی، استفاده از کارت‌های گم‌شده، تقلب‌های درخواست

و ... را در خود دارد. پژوهشگران یک نمونه شامل ٤١٦٤٧ تراکنش که ۳.۷٤ درصد از آنها تقلب‌آمیز بودند را پالایش و انتخاب کردند (گادی و همکاران ۲۰۰۸). در این مجموعه داده‌ای انتخاب ویژگی انجام شده، و همه ستون ها به‌صورت عددی در آمده‌اند. اما نکته‌ای که این مجموعه را برای پژوهش حاضر بلااستفاده می سازد، فقدان وجود ستونی برای یکتاسازی شماره دارنده کارت است. یعنی هر تراکنش به صورت مستقل مورد بررسی قرار می‌گیرد. بنابراین هیچ یک از مولفه‌های مدل قابل پیاده‌سازی در این مجموعه داده‌ای نیستند، زیرا به دلیل عدم یکتاسازی دارنده کارت، امکان ایجاد پروفایل رفتاری تراکنشی برای او و نیز امکان ایجاد خوشه‌های ایستا و پویا وجود ندارد. بنابراین با خلاصه‌سازی مدل نیز این داده‌ها برای پژوهش حاضر کاربردی نیستند. دومین مجموعه داده‌ای بررسی شده، از کار (دال پازولو و همکاران ۲۰۱٤) استخراج شد. در این مجموعه نیز تلفیقی از تراکنش‌های قانونی و تقلب‌آمیز وجود داشته و مجموعه داده‌ای دارای برچسب است. کل تراکنش‌ها ۲۸٤۸۰۷ عدد، حاوی ۳۱ ویژگی عددی شامل، زمان، مقدار و مکان تراکنش است. داده‌ها به شدت نامتوازن هستند، یعنی ۰.۱۷۲ درصد از آنها (٤۹۲ تراکنش تقلب‌آمیز) دارای برچسب تقلب هستند. در اینجا نیز شماره دارنده کارت وجود نداشته و هر تراکنش مستقل از دیگری قابل تحلیل است. بنابراین از این مجموعه داده‌ای نیز به دلایل یاد شده در مجموعه قبلی، نمی‌توان در مدل ارائه شده در پژوهش حاضر استفاده کرد. با توجه به اینکه حتی با تماس با محققین قبلی مجموعه داده‌ای ناشی از تراکنش‌های کارت‌های اعتباری و مناسب برای آزمون مدل یافت نشد، مجموعه‌های استاندارد برای کشف ناهنجاری مورد توجه قرار گرفتند اولین مجموعه داده‌ای که به‌عنوان یک مجموعه کشف ناهنجاری مورد توجه قرار گرفت. مجموعه داده‌های مسابقات کشف دانش ۹۹ است. این مجموعه در پژوهش‌های بسیاری برای سنجش کارایی مدل‌های کشف ناهنجاری مورد استفاده قرار گرفته است، یکی از آخرین پژوهش‌هایی که از این مجموعه داده‌ای استفاده کرده بود، پژوهش ادیبی و شهرابی در خصوص یافتن برخط ناهنجاری با ترکیبی از شبکه عصبی کوهونن و خوشه‌بندی بردار پشتیبان است (ادیبی و شهرابی ۲۰۱۵). مجموعه داده‌ای ۹۹ KDD Cup حاصل ارزیابی داده‌های سیستم کشف تقب دارپا است که توسط آزمایشگاه لینکلن MIT تحلیل شده است. این مجموعه شامل ٤۹٤۰۰۰۰ رکورد سابقه ارتباط و دارای ٤۱ ویژگی است. علاوه بر اینکه در هر رکورد، برچسب قانونی یا تقلب مشخص شده است. این تقلب‌ها ناشی از حملات ۲٤گانه شبکه هستند. در این مجموعه نیز امکان یکتاسازی یک موجودیت و متعاقبا امکان استخراج رفتار تراکنشی آن موجودیت

و خوشه‌بندی وجود ندارد. بنابر این امکان از این مجموعه داده‌ای برای آزمون مدل ارائه شده در این پژوهش نمی‌توان استفاده کرده مجموعه داده‌ای دوم مورد استفاده (ادیبی و شهرابی ۲۰۱۵) برای کشف ناهنجاری، مجموعه جمع‌آوری شده تحت عنوان KeyStroke Dynamics dataset است. این مجموعه شامل ۲۰٤۰۰ تراکنش مربوط به زمانبندی ورود کلمه عبور ۵۱ تایپیست است، این رکوردها ۳٤ ستون دارند. در این مجموعه برچسب وجود ندارد و با جابجا کردن رکوردهای تایپیست‌ها ناهنجاری ایجاد و متعاقبا کشف می‌شود. نکته مهم در استفاده از این مجموعه، وجود کلید یکتای موجودیت مالک تراکنش است، بنابراین تنها مجموعه یافته شده‌ای که با خلاصه‌سازی مدل می‌توان از آن استفاده کرد، همین مجموعه داده‌ای است که در ادامه به تشریح آن و روش استفاده در پژوهش حاضر پرداخته خواهد شد. در KeyStroke Dynamics dataset هر یک از تایپیست‌ها با کد S۰۰۲ تا S۰۵۷ مشخص شده و برای هر یک ٤۰۰ رکورد تراکنش وجود دارد. برای آماده کردن مجموعه داده‌ای جهت آزمون مدل تراکنش‌های مربوط به ۲۰ تایپیست اول یعنی ۸۰۰۰ تراکنش به‌عنوان تراکنش قانونی در نظر گرفته شدند. از سوی دیگر ٤۰۰ تراکنش تایپیست S۰٤۹ به‌عنوان تراکنش تقلب‌آمیز انتخاب شد. با نرخ تصادفی ۲ برای تقلب، و ۹۸٪ برای داده قانونی، ۸۰۰۰ رکورد به صورت بدون جای‌گذاری از مجموعه ۱۲۰۰۰ رکورد انتخاب شد) .٪۹۸ از رکوردها مربوط به ۲۰ تایپیست اول و ۲ مربوط به تایپیست (S۰٤۹ با در نظر گرفتن ۷۰٪ داده‌ها برای مدل‌سازی و ۳۰٪ داده‌ها برای آزمون، نتیجه به شرح جدول ۳۹ است.

جدول ۳۹ – توزیع داده‌های نمونه مورد استفاده

نوع مجموعه داده‌ای	تعداد کل	قانونی	تقلب‌آمیز
مدل‌سازی (آموزش)	٥٦۰۰	٥٤۸۳	۱۱۷
طبقه‌بندی (آزمون)	۲٤۰۰	۲۳٥٤	٤٦

ماخذ: یافته‌های تحقیق

در مدل‌های استفاده شده در این مقاله، از همه ویژگی‌ها استفاده شده است، تنها پیش‌پردازش انجام شده، ترمال‌سازی داده‌ها در فاصله ۰ تا ۱ است، برای خوشه‌بندی تجمیعی نیز متوسط ارزش هر یک از ویژگی‌ها در طول زمان به‌عنوان ویژگی تجمیعی مورد استفاده قرار گرفته است. نخستین مدلی که مورد بررسی قرار می‌گیرد، مدل شبکه عصبی کوهونن برای کشف تقلب در تراکنش‌های کارت‌های اعتباری ارائه شده توسط زاسلاوسکی و استریژاک (۲۰۰٦) است. همان‌گونه که در همین فصل و در

بخش مدل‌سازی اول توضیح داده شد، این روش یک پروفایل با استفاده از ترون های شبکه عصبی کوهونن می‌سازد. فاصله تراکنش ورودی با شبیه‌ترین نرون اندازه‌گیری شده، و در صورتی که بیشتر از یک آستانه از پیش تعیین شده باشد، به‌عنوان تقلب‌آمیز، در غیر این‌صورت قانونی ارزیابی می‌شود. جدول ٤٠ پارامترهای اولیه برای آزمون مدل را به تصویر می‌کشد.

جدول ٤٠ - پارامترهای اولیه برای آزمون مدل زاسلاوسکی و استریژا (۲۰۰٦)

پارامتر	مقدار
شعاع همسایگی	۲
ضریب آموزش اولیه	۰.۲
ضریب همسایگی اولیه	۰.۵
اندازه شبکه	۱۰ در ۱۰ (۱۰۰نرون)
تعداد تکرار	۲۰۰

ماخذ: یافته‌های تحقیق

این مدل در مدل‌سازی، مدل موازی پیشنهادی برای ساخت پروفایل‌های چندگانه برای مالکین تراکنش‌های کارت و الگوریتم پیشنهادی برای طبقه‌بندی تراکنش براساس رویکرد موازی گواه فازی بطور کامل توضیح داده شد. اما برای استفاده از این مدل در داده‌های اعتبارسنجی جمع‌آوری شده، باید ساده‌سازی صورت پذیرد. همان‌گونه که در تشریح داده‌های KeyStroke Dynamics dataset توضیح داده شد. این مجموعه داده‌ای، ابر داده در خصوص تایپیست‌ها ندارد، بنابراین ویژگی ایستایی در خصوص آنها وجود ندارد تا بتوان بر روی آنها خوشه‌بندی انجام داد. بنابراین در مدل استفاده شده در این بخش، منبع خوشه ایستا حذف می‌شود. علاوه بر اینکه ویژگی‌ها نیز به همان صورت ویژگی‌های ۳۱ گانه مورداستفاده قرار گرفته و برای خوشه‌بندی پویا عدد متوسط آنها به‌عنوان مقدار ویژگی مورد استفاده قرار می‌گیرد. جدول زیر پارامترهای اولیه برای مدل نهایی را نشان می‌دهد .

جدول ٤١ - پارامترهای اولیه برای مدل نهایی پژوهش

پارامتر	شرح	ارزش
K_d	تعداد خوشه‌های پویا	۷
M	تعداد ارزش های انتشار	۵
N	تعداد حالات مخفی	۱۰
R	طول توالی مشاهدات	۲۰
L	طول توالی آموزش	۱۰۰
Q	تعداد نرون های کوهنن	۱۰۰

ماخذ: یافته‌های تحقیق

جدول ٤٢ - نتیجه آزمون مدل‌ها با استفاده از داده‌های KeyStroke Dynamics dataset

روش	فراخوانی	اعلان نادرست	F	g-means ۲	متوسط زمان طبقه‌بندی (میلی ثانیه)
کوهنن (زاسلاوسکی و استریژاک ۲۰۰٦)	۳۸.۷۱	۱۹	۱۲.۵۲	۵۵.٤۸	۱۲۱
کوهنن با ماتریس یو	٦۳.۱	۹.۳	۳۲.۱۵	۷۵.٦۵	۱۲٦
مارکوف پیشنهادی با خوشه‌بندی چند بعدی	۵۸.۳۲	۸.۷	۳۱.۲۷	۷۲.۹۷	٤۹
مدل نهایی با نگاشت کاهش و هم‌جوشی	٦۸.۹	٤.۸۱	٤۷.۹۲	۸۰.۹۹	۲۷

ماخذ: یافته‌های تحقیق

با بررسی نتایج موجود در جدول فوق مشاهده می‌شود، مدل نهایی خلاصه شده بدون در نظر گرفتن رفتار خوشه‌ایستا، همچنان در کلیه پارامترها بهتر از مدل‌های دیگر عمل کرده است. با توجه به معماری نگاشت کاهشی، این مدل سرعت اجرای بالاتری را نسبت به مدل‌های دیگر نیز داشته است.

نتیجه‌گیری و پیشنهادها

در این پژوهش نخست مساله پژوهش به طور کامل تبیین شده و مدل‌های مختلف برای کشف تقلب تراکنش‌های کارت‌های الکترونیکی مورد بررسی قرار گرفت. در خلال بررسی ها مشخص شد که رویکردهای با سرپرستی دارای نقائصی هستند، در حالیکه مدل‌های بدون سرپرستی با وجود دقت کمتر، نقص های اشاره شده را ندارند. از سوی دیگر پیچیدگی های عنوان شده در مساله کشف تقلب سبب ایجاد ابهام و عدم اطمینان در تصمیم شده است. نکته دیگری که در خصوص این مساله مشهود است، وجود شرایط بزرگ داده در آن است. با این اوصاف سه شکاف علمی مد نظر قرار گرفتند که عبارت بودند از :

- مدل‌های استفاده شده در ایران برای کشف تقلب در تراکنش‌های کارت‌های الکترونیکی، به‌صورت با سرپرستی بوده‌اند.
- هیچ یک از مدل‌های بررسی شده، منبع رفتاری خوشه‌های مالک تراکنش را برای افزایش دقت در نظر نگرفته‌اند.
- پژوهش‌های اندکی الزامات بزرگ داده را در مساله کشف تقلب در کارت‌های الکترونیکی در نظر گرفته‌اند.

در این راستا همانطور که در فصل سه عنوان شد، رویکرد پژوهش کمی، روش پژوهش تحلیلی ریاضی و مقطع زمانی نیز مقطعی در نظر گرفته شد. برای جمع‌آوری داده‌های مورد نیاز مدل‌سازی و اعتبارسنجی آن، نخست ویژگی‌های مشخص کننده رفتار دارنده کارت استخراج شده از ادبیات موضوع، در گروه کانونی مطرح شده و مهم‌ترین آنها انتخاب شدند. با در نظر گرفتن شرایط وجود برچسب، این انتخاب ویژگی با روش تئوری اطلاعات نیز انجام شده و مورد مقایسه قرار گرفت. سپس برای مدل‌سازی، داده‌های برچسب دار، زیرسامانه نفت گاز سامانه هوشمند سوخت؛ مربوط به ٤ ماه از اول مهر ۱۳۹٤ الی پایان دی ماه همان سال جمع‌آوری شد. به موازات این داده‌ها، تراکنش‌های مربوط به پایانه‌های پرداخت سیستم بانکی نیز از یک شرکت ارائه دهنده خدمات پرداخت از تاریخ

اول فروردین ۱۳۹۵ تا پایان اردیبهشت همان سال دریافت شد. داده‌های این شرکت دارای برچسب نبود؛ که با در نظر گرفتن نظر خبرگان تعدادی از تراکنش‌های آن دارای برچسب شدند.

با توجه به بررسی ادبیات موضوع و شرایط مدل‌های با سرپرستی، مدل‌های بدون سرپرستی نزدیکی بیشتری به واقعیت مسئله کشف تقلب در تراکنش‌های کارت‌های الکترونیکی دارند. به‌علاوه مقایسه مدل ارائه شده در این پژوهش، با مدل‌های بدون سرپرستی موجود در ادبیات موضوع، آشکار ساخت؛ که در نظر گرفتن منبع رفتار تاریخی موجودیت مالک تراکنش برای استنتاج نهایی، نمی‌تواند دقت بالایی در طبقه‌بندی را ارائه دهد. این عدم دقت، ناشی از ذات داده‌های تراکنش‌های کارت‌های الکترونیک است، که سبب ابهام و عدم اطمینان در تصمیم نهایی می‌شود. بنابراین مدل باید از چند منبع مستقل و رویکردی برای هم‌جوشی نتایج استفاده کند. در این تحقیق دو منبع رفتار تاریخی خوشه‌های ایستا و پویای در بر گیرنده موجودیت مالک تراکنش، به منبع معمول استفاده شده در رویکردهای پیشین، افزوده شد. جهت هم‌جوشی نتایج، برای کاهش عدم اطمینان از ترکیب دمپستر–شیفر استفاده و جهت کاهش ابهام، از منطق فازی بهره برداری شد. نتیجه مقایسه مدل با مدل‌های دیگر، در دو مجموعه داده‌ای حاصل از سیستم بانکی و سامانه هوشمند سوخت، نشان داد که مدل پیشنهادی توانسته است دقت بالاتری را نسبت به مدل‌های دیگر، ارائه دهد. این دقت خصوصاً در کاهش نرخ اعلان اشتباه کاملا مشهود است. به‌علاوه سنجش اعتبار مدل ساده شده (بدون در نظر گرفتن خوشه‌ایستا) برتری آن را با شدت کمتر، در داده‌های استاندارد نیز به نمایش گذاشت. لازم به ذکر است، کاهش دقت نسبی، ناشی از عدم در نظر گرفتن منبع استنتاج خوشه‌های ایستا و نیز تفاوت ماهوی داده‌های جمع‌آوری شده برای اعتبارسنجی مدل است .تصمیم دیگری که در افزایش دقت طبقه‌بندی نقش ایفا می‌کند، استفاده از آستانه برش تطبیقی در مقابل استفاده از آستانه برش ثابت است. در پژوهش حاضر، معیاری فازی تطبیقی در محاسبه مقدار آستانه ها استفاده شد. نتایج آزمون‌ها و اعتبارسنجی در هر سه مجموعه داده‌ای، دقت بالاتر روش اول را نشان دادند. این دقت هم ناشی از کاهش ابهام به دلیل استفاده از منطق فازی و هم به دلیل ایجاد یک معیار مطابق با رفتار هر موجودیت به صورت جداگانه است.

نهایتا پس از دست‌یابی به دقت بالاتر نسبت به مدل‌های دیگر، نوبت به در نظر گرفتن الزامات بزرگ داده، برای افزایش سرعت اجرای مدل، می‌رسد. مدل ارائه شده با تغییر شکل الگوریتم‌ها، برای ایجاد زیر مجموعه‌ای از وظایف مستقل، و استفاده از چارچوب محاسباتی نگاشت کاهش، به دنبال رسیدن

به دو هدف اجرای موازی و مقیاس‌پذیری است. آزمون مدل در شرایط مختلف در هر سه مجموعه داده‌ای نشان داده که از یک سو، با افزایش ابعاد مدل، زمان اجرا تغییر معنا داری نمی‌کند. از سوی دیگر، در هر سه مجموعه داده‌ای هم زمان ساخت مدل، و هم زمان طبقه‌بندی تراکنش جدید، کاهش چشم‌گیری داشته است. بنابراین مدل پیشنهادی توانسته است از نظر زمان اجرا نیز، نسبت به مدل‌های دیگر عملکرد برتری را به نمایش گذارد. به طور خلاصه، مهم‌ترین دلایل عملکرد قابل قبول مدل ارائه شده را می‌توان به شرح زیر دانست:

- استفاده از چند منبع برای استنتاج که توانسته است، نقص عدم اطمینان طبقه‌بندی تراکنش به تقلب‌آمیز و قانونی بودن را کاهش دهد .
- استفاده از منطق فازی که توانسته است، نقص ابهام در طبقه‌بندی را کاهش دهد.
- استفاده از معیار طبقه‌بندی تطبیقی، که دقت طبقه‌بندی را با شخصی سازی آن، افزایش داده است.
- استفاده از نگاشت کاهش که توانسته است مدل را به صورت موازی و مقیاس پذیر اجرا کرده و زمان اجرا را کاهش دهد.

مهم‌ترین مولفه‌هایی که می‌توانند در مدل شرکت کنند، نخست الگوی تراکنش تاریخی موجودیت مالک تراکنش‌هاست. استفاده از این منبع در همه ادبیات مطالعه شده، رایج است. پیچیدگی رفتار متقلب، و تغییرات زیاد رفتاری تراکنش، استفاده صرف از این منبع را با محدودیت مواجه می سازد. لذا در نظر گرفتن دو منبع رفتار خوشه‌ای برای افزایش دقت، مناسب خواهند بود. مولفه مهم دیگر استفاده از آستانه های تطبیقی و فازی برای افزایش دقت است. و نهایتا قابلیت مقیاس پذیری و اجرای موازی مدل برای دست‌یابی به الزامات بزرگ داده است .

در مساله کشف تقلب در تراکنش‌های کارت الکترونیکی، مهم‌ترین معیار استفاده از ترکیبی از ویژگی‌های تجمیعی و خام برای استخراج الگوی رفتاری موجودیت مالک تراکنش در طول یک بازه زمانی است. همانطور که در ادبیات بررسی شد، استفاده از ویژگی‌های خام تراکنش نمی‌تواند دانش کافی را از نظر اکتشاف الگوی رفتار موجودیت مالک تراکنش‌ها به تصویر بکشد. به‌علاوه، به نظر مدل‌های با سرپرستی با توجه به حجم زیاد اطلاعات موجود در بزرگداده و سرعت به روزرسانی آن، به سرعت نامعتبر می‌شوند. بنابراین رویکردهای بدون سرپرستی برای این دسته از داده‌ها مناسب تر خواهند بود. فقدان وجود برچسب در همه شرایط، دلیل دیگری است که استفاده تنها از مدل‌های

با سرپرستی را ناکارآمد می سازد. ترکیب دمپستر–شیفر فازی، توانسته است، نقائص ابهام و عدم اطمینان را که در ذات تراکنش‌های کارت‌های الکترونیک وجود دارد، تا حد قابل قبولی رفع کند. این نقائص سبب بالا رفتن مقدار اعلان اشتباه می‌شوند. نهایت نگاشت کاهش با داشتن قابلیت اجرای موازی و توزیع شده وظایف، توانسته است عملکرد مناسبی را در پژوهش حاضر به نمایش بگذارد. این مدل با ایجاد ویژگی‌های مقیاس پذیری و اجرای موازی، به الزامات بزرگ داده دست یافته است .

بر اساس یافته های این پژوهش، پیشنهاد می‌شود، سازمان‌هایی که دارای حجم انبوهی از تراکنش‌های مالی ناشی از استفاده از کارت‌های الکترونیکی بوده، و به نوعی با مسئله تقلب روبرو هستند، در مرحله اول از مدل پیشنهادی به صورت کشف کننده استفاده کرده و عملکرد آن را در شرایط عملیاتی مورد آزمون قرار دهند. این امر می‌تواند مخاطره اعلان اشتباه را کاهش دهد، علاوه براینکه سبب شناسایی متقلبین در کمترین زمان ممکن شود. سازمان می‌تواند با ایجاد ساز و کار پیگیری و بررسی مناسب، صحت عملکرد سیستم را مورد آزمون قرار دهد. به‌علاوه سرعت بالای مدل، در صورت رضایت بخش بودن نتایج آزمون ها، این امکان را فراهم می سازد که با پیکره بندی مناسب، بتوان از آن به‌صورت سلبی در نشست ایجاد تراکنش استفاده کرد. این قابلیت مانع از انعقاد تراکنش مشکوک به تقلب شده که خود می‌تواند بخش عمده ای از زیان ناشی از تقلب را کاهش داده، و برای مشتریان سبب اطمینان بیشتر شود. به‌علاوه، از آنجا که مدل‌های با سرپرستی می‌توانند نسبتا دقت بالاتری داشته باشند، در صورت دست‌یابی به مصادیق کافی تقلب، در سازمان می‌توان مدل با سرپرستی مناسبی با در نظر گرفتن الزامات بزرگ داده‌ایجاد، و به‌عنوان منبع چهارم تصمیم در استنتاج نهایی وارد ساخت. سازمان می‌تواند نتایج تحلیل های ناشی از پیاده‌سازی بر خط مدل را در یک ابزار تحلیلی هوش کسب و کار، به نمایش گذارد؛ تا مدیران و تحلیل گران در سطوح مختلف، در کمترین زمان ممکن به آنها دسترسی داشته باشند. این امر می‌تواند به شکل چشم گیری در مدیریت مخاطرات و جلوگیری از آسیب پذیری ها به سازمان یاری رساند. متعاقبا، از زیآن‌های ناشی از این آسیب پذیری ها جلوگیری شده و سطح اعتماد مشتری نیز افزایش می یابد. همان‌گونه که در ادبیات عنوان شد، رفتار متقلبین در طول زمان در حال تغییر است. بنابراین در صورت استفاده از ساز و کارهایی مانند مدل پیشنهادی، به روز نگاه داشتن آن نیز، همواره باید مد نظر قرار گیرد. بنابراین قابلیت ساخت سریع مدل، در اینجا می‌تواند برای سازمان‌ها بسیار مفید باشد. زیرا هر بروز رسانی در کمترین زمان

صورت می‌پذیرد، که خود دغدغه کارشناسان فعال در بخش فن‌آوری اطلاعات را برای استفاده از منابع سخت افزاری بسیار پرقدرت به طور چشم‌گیری کاهش می‌دهد.

برای سامانه هوشمند سوخت، به طور خاص پیشنهاد می‌شود، از تحلیل‌های ناشی از مدل، برای مبارزه با قاچاق فرآورده نفت گاز استفاده کند. زیرا اساس مدل توسعه‌یافته بر مبنای کشف ناهنجاری است، که به عقیده خبرگان امر، قاچاق نیز نوعی انحراف از مصرف، و ناهنجاری محسوب می‌شود. قاعدتا این تحلیل‌ها می‌تواند نه تنها برای سازمان، بلکه برای کشور نیز سودمند باشد. به هر دو کسب و کار پیشنهاد می‌شود، ابر داده‌های غنی‌تری در مورد کسب و کار پذیرندگان خود جمع‌آوری کنند. این امر می‌تواند تحلیل خوشه‌های ایستا را بسیار موثرتر سازد. با تقویت این منبع استنتاج دقت مدل نیز افزایش خواهد یافت. از سوی دیگر عنصر داده‌ای فاصله میان پذیرنده‌ها، می‌تواند در مورد فاصله زمانی میان تراکنش‌ها اطلاعات مفیدتری را عرضه کند؛ که متعاقبا صحت تصمیم بالاتر خواهد رفت. نهایتا به سازمان‌های نظارتی پیشنهاد می‌شود از مدل ارائه شده برای بررسی عملکرد سیستم‌های تحت نظارت خود استفاده کنند.

در طول پژوهش حاضر مانند هر پژوهش دیگری، محدودیت هایی وجود داشتند که رفع آنها می‌توانست بر کیفیت کار اثر قابل ملاحظه‌ای داشته باشد. نوین بودن حوزه بزرگ داده و گستردگی آن این محدودیتها را پر رنگ تر ساخته است. نخستین محدودیت کمبود داده‌های در دسترس برای پژوهش در زمینه کشف تقلب است. دلایل امنیتی و کسب و کاری، به همراه رقابتی بودن راه‌حل‌های کشف تقلب، به شدت بر این محدودیت اثر گذاشته اند. این امر در ادبیات پژوهش به وضوح مشاهده شد، در حالیکه در طی انجام پژوهش حاضر نیز به چشم خورد. در کنار کمبود داده، فقدان داده‌های به گزینی برای مقایسه کارایی الگوریتم‌ها و مدل‌ها را نیز می‌توان به عنوان یک محدودیت و چالش اساسی برشمرد. نبود داده‌های استاندارد، اعتبارسنجی کامل تحقیقات در این زمینه را بسیار محدود می سازد. در بخشی از تحقیق که با داده‌های بدون برچسب سر و کار داشتند، گردآوری نشدن مصادیق تقلب، و فقدان برچسب، برچسب گذاری را به یک تصمیم نظری تبدیل ساخت که مستقیما بر ارزیابی مدل تاثیر می گذارد. در خصوص سیستم بانکی نیز مشاهده شد، که ابرداده کافی در خصوص کسب و کارهای پذیرنده وجود ندارد. این امر تحلیل های ناشی از خوشه‌های ایستا را که به‌صورت بالقوه می‌توانند، یک منبع اساسی برای تصمیم باشند، محدود می سازد. ضمنا، در خصوص هر دو کسب و کار سامانه هوشمند سوخت و سیستم بانکی، موقعیت جغرافیایی و مسافت میان

جایگاه‌های سوخت یا پذیرنده‌ها محاسبه نشده است. در صورتیکه وجود این عناصر داده‌ای بردقت تصمیم، اثرگذار است.

در راستای ارتقاء عملکرد مدل، می‌توان تغییر متغیرهای تصمیم بر روی مصادیق دیگر تقلب و ناهنجاری پیشنهادی را به‌عنوان یک حوزه تحقیقاتی آتی در نظر گرفت.

منابع

- Alshawabkeh, M., Jang, B & Kaeli, D. (۲۰۱۰). Accelerating the Local Outlier Factor Algorithm on a GPU for Intrusion Detection Systems .ACM, ۱۰٤-۱۱۰.

-Awasthi, A & ,.Satyaveer , S. C. (۲۰۱۱). Using AHP and Dempster–Shafer theory for evaluating sustainable transport solutions. Environmental Modelling & Software. ۷۸۷-۷۹٦.

-Bai, M., Wang, X., Xin, J & Wang, G. (۲۰۱٦). An efficient algorithm for distributed density-based outlier detection on big data. Neurocomputing, ۱۹-۲۸.

-Bhattacharyya, S., Jha, S., Tharakunnel, K & Westland, J. C. (۲۰۱۱). Data mining for credit card fraud: A comparative study. Decision Support Systems, ٦۰۲-٦۱۳.

-Bhusari, V & ,.Patil, S. (۲۰۱۱). Study of Hidden Markov Model in Credit Card Fraudulent Detection. International Journal of Computer Applications.

-Bitterer, A. (۲۰۱۱). Hype Cycle for Business Intelligence. Gartner Inc.

-Bolton, R & ,.Hand, D. (۲۰۱۱). Unsupervised Profiling Methods for Fraud Detection.Credit Scoring and Credit Control.

-Boyd, D & ,.Crawford, K. (۲۰۱۱). Six provocations for Big Data. Symposium on the Dynamics of the Internet and Society, (۱-۱۷). London: Oxford Internet Institute.

-Brown, B., Chui, M & Manyika, J. (۲۰۱۱). Are you ready for the era of big data? McKinsey quarterly.

-Chaudhary , K., Yadav, J & Mallick, B. (۲۰۱۲). A review of Fraud Detection Techniques: Credit Card. International Journal of Computer Applications.

-Chen, H., Chiang, R. H & Storey, V. C. (۲۰۱۲). Business Intelligence and Analytics: From Big Data to Big Impact. MIS Quarterly, ۱۱٦٥-۱۱۸۸.

-Chen, R., Chen, T & Lin, C. (۲۰۰٦). A new binary support vector system for increasing detection rate of credit card fraud. International Journal of Pattern Recognition and Artificial Intelligence, ۲۲۷-۲۳۹.

-Correa Bahnsen, A., Aouada, D., Stojanovic, A & Ottersten, B. (۲۰۱٦). Feature engineering strategies for credit card fraud detection. Expert Systems with Applications, ۱۳٤-۱٤۲.

-Correa Bahnsen, A., Stojanovic, A., Aouada, D & Ottersten, B. (۲۰۱۳). Cost sensitive credit card fraud detection using bayes minimum risk. In Proceedings of the ۲۰۱۳ ۱۲th international conference on machine learning and applications, ۳۳۳-۳۳۸.

-Correa Bahnsen, A., Stojanovic, A., Aouada, D & Ottersten, B. (۲۰۱٤). Improving credit card fraud detection with calibrated probabilities. In Proceedings of the fourteenth siam international conference on data mining, ٦٧٧-٦٨٥.
-Dal Pozzolo, A., Caelen, O., Le Borgne, Y. A., Waterschoot, S & Bontempi, G. (۲۰۱٤). Learned lessons in credit card fraud detection from a practitioner perspective. Expert Systems with Applications, ٤٩١٥-٤٩٢٨.
-Deng, Y., Chan, T. S., Wu, Y & Wang, D. (۲۰۱۱). A new linguistic MCDM method based on multiple-criterion data fusion. Expert Systems with Applications, ٦٩٨٥-٦٩٩٣.
-Doan, A., Ramakrishnan, R & Halevy, A. Y. (۲۰۱۱). Crowd Crowdsourcing. Communication of the ACM, ٨٦-٩٦.
-Duman, E & ,.Ozcelik, M. H. (۲۰۱۱). Detecting credit card fraud by genetic algorithm and scatter search. Expert Systems with Applications, ١٣٠٥٧-١٣٠٦٣.
-Duru Aral, K., Güvenir, H. A., Sabuncuoglu, I & Akar, A. R. (۲۰۱۲). A prescription fraud detection model. Computer methods and programs in biomedicine, ٣٧-٤٦.
-Dymova, L., Sevastianov, P & Bart, P. (۲۰۱۱). A new approach to the rule-base evidential reasoning: Stock trading expert system application. Expert systems with Applications, ٥٥٦٤-٥٥٧٦.
-Epaillard, E & ,.Bouguila, N. (۲۰۱٦). Proportional data modeling with hidden Markov models based on generalized Dirichlet and Beta-Liouville mixtures applied to anomaly detection in public areas. Pattern Recognition, ١٢٥-١٣٦.
-European payment cards fraud report Payments Card. (۲۰۱۰). European payment cards fraud report Payments, Cards and Mobiles. LLP & Author.
-González, P. C & Velásquez, J. D. (۲۰۱۳). Characterization and detection of tax payers with false invoices using data mining techniques., Expert systemps with applications, ١٤٢٧-١٤٣٦.
-Graaff, A. J & Engelbrecht, A. P. (۲۰۱۱). The Artificial Immune System for Fraud Detection in the Telecommunications Environment.
-Gravina, R., Alinia, P., Ghasemzadeh, H & Fortino, G. (۲۰۱۷). Multi-sensor fusion in body sensor networks: State-of-the-art and research challenges. Information Fusion ,٦٨-٨٠.
-Hadavandi, E., Shahrabi, J & Hayashi, Y. (۲۰۱٥). SPMoE: a novel subspace-projected mixture of experts' model for multi-target regression problems. Soft Computing, ٢٠٤٧-٢٠٦٥.

-Hafezi, R., Shahrabi, J & Hadavandi, E. (۲۰۱۵). A bat-neural network multi-agent system (BNNMAS) for stock price prediction: Case study of DAX stock price. Applied Soft Computing, ۱۹۶-۲۱۰.

-Halvaiee, N. S & Akbari, M. K. (۲۰۱٤). A novel model for credit card fraud detection using Artificial Immune Systems. Applied soft computing, ٤۰-٤۹.

-He, W., Williard, N., Osterman, M & Pecht, M. (۲۰۱۱). Prognostics of lithium-ion batteries based on Dempster–Shafer theory and the Bayesian Monte Carlo method. Journal of Power Sources, ۱۰۳۱٤-۱۰۳۲۱.

-Herodotou, H., Lim, H., Luo, G., Borisov, N., Dong, L., Cetin, F. B. (۲۰۱۱). Starfish: A Selftuning System for Big Data Analysis. Biennial Conference on Innovative Data Systems Research, ۲٦۱-۲۷۲.

-Ho, G. T., Ip, W. H., Lee, C. K & Mou, W. L. (۲۰۱۲). Customer grouping for better resource allocation using GA based clustering technique. Expert Systems with Applications, ۱۹۷۹-۱۹۸۷.

-Huang, J., Zhu, Q., Yang, L & Feng, J. (۲۰۱٦). A non-parameter outlier detection algorithm based on Natural Neighbor. Knowledge-Based Systems, ۷۱-۷۷.

-Huang, S. Y., Tsaih, R. H & Yu, F. (۲۰۱٤). Topological pattern discovery and feature extraction for fraudulent financial reporting. Expert systems with applications, ٤۳٦۰-٤۳۷۲.

-Jha, S., Guillen, M & Westland, J.C. (۲۰۱۲). Employing transaction aggregation strategy to detect credit card fraud. Expert Systems with Applications, ۱۲٦٥۰-۱۲٦٥۷.

-Kesselheim, A. S., Studdert, D. M & Mello, M. M. (۲۰۱۰). Whistle-blowers 'experiences in fraud litigation against pharmaceutical companies. New England Journal of Medicine, ۱۸۳۲-۱۸۳۹.

-Khaleghi, B., Khamis, A., Okar, F & Razavi, S. N. (۲۰۱۳). Multisensor data fusion: A review of the state-of-the-art. Information Fusion, ۲۸-٤٤.

-Khatibi, V & ,.Montazer, G. A. (۲۰۱۰). A fuzzy-evidential hybrid inference engine for coronary heart disease risk assessment. Expert Systems with Applications, ۸٥۳٦-۸٥٤۲.

-Khreich, W., Granger, E., Miri, A & Sabourin, R. (۲۰۱۰). On the memory complexity of the forward–backward algorithm. Pattern Recognition Letters, ۹۱-۹۹.

-Kotu, V & ,.Deshpande, B. (۲۰۱٥). Predictive analytics and data mining; Concepts and practice with RapidMiner. San Francisco: Morgan Kaufmann.

-Kumar, R & ,.Raj, S. (۲۰۱۲). Design and analisys of credit card fraud detection based on HMM. International Journal of Engineering and Innovative Technology, ۳۳۲-۳۳٥.

-Kundu, A., Panigrahi, S., Sural, S & Majumdar, A. K. (۲۰۰۹). Credit card fraud detection: A fusion approach using Dempster-Shafer theory and Bayesian learning. Special Issue on Information Fusion in Computer Security, ۳۵٤-۳٦۳.

-Loshin, D. (۲۰۱۳). Big data analytucs; from strategic planning to enterprise integration with tools, technigues, NoSQL, and graph. Morgan Kaufmann.

-Lusch, R. F., Liu, F & Chen, Y. (۲۰۱۰). The Phase Transition of Markets and Organizations: The New Intelligence and Entrepreneurial Frontier. IEEE Intelligent Systems, ۷۱-۷۵.

-Manovich, L. (۲۰۱۱). Trending: The Promises and the Challenges of Big Social Data. Debates in the Digital Humanities. Minneapolis: University of Minnesota Press.

-Mansouri, T., Nabavi, A., Ravasan, A. Z & Ahangarbahan, H. (۲۰۱۵). A practical model for ensemble estimation of QoS and QoE in VoIP services via fuzzy inference systems and fuzzy evidence theory. Telecommunication Systems, ۱-۱۳.

-Mansouri, T., Zare Ravasan, A & Gholamian, M. R. (۲۰۱٤). A Novel Hybrid Algorithm Based on K-Means and Evolutionary Computations for Real Time Clustering. International Journal of Data Warehousing and Mining, ۱-۱٤.

-Metan, G., Sabuncuoglu, I & Pierreval, H. (۲۰۱۰). Real time selection of scheduling rules and knowledge extraction via dynamically controlled data mining. International Journal of Production Research, ٦۹۰۹-٦۹۳۸.

-Minegishi, T & ,.Niimi, A. (۲۰۱۱). Proposal of Credit Card Fraudulent Use Detection by Online-type Decision Tree Construction and Verification of Generality. International Journal for Information Security Research, ۲۲۹-۲۳۵.

-Mishra, J. S., Panda, S & Mishra, A. K. (۲۰۱۳). A Novel Approach for Credit Card Fraud Detection Targeting the Indian Market. International Journal of Computer Science Issues, ۱۷۲-۱۷۹.

-Nasiri, N & ,.Minayi, B. (۲۰۱۱). Data mining methods for credit card fraud detection first International conference on E-Citizen & Cellphone, Tehran, ۲۸-۲۹.

-Ngai, E. W., Hu, Y., Wong, Y. H., Chen, Y & Sun, X. (۲۰۱۱). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. Decision Support Systems, ٥٥۹-٥٦۹.

-Olszewski, D. (۲۰۱٤). Fraud detection using self-organizing map visualizing the user profiles. Knowledge-Based Systems, ۳۲٤-۳۳٤.

-Oracle. (۲۰۱۲). Oracle: Big Data for the Enterprise. Oracle white paper. Oracle. (۲۰۱۳). Big Data Analytics; Advanced Analytics in Oracle Database.www.oracle.com.

-Patil, D. D., Wadhai, V. M & Gokhale, J. A. (۲۰۱۰). Evaluation of Decision Tree Pruning Algorithms for Complexity and Classification Accuracy. International Journal of Computer Applications, ۲۳-۳۰.
-Pejic-Bach, M. (۲۰۱۰). Profiling intelligent systems applications in fraud detection and prevention: survey of research articles. Proceedings of International Conference on Intelligent Systems: Modelling and Simulation, ۸۰-۸۵.
-Popescu, D. E., Lonea, M & Zmaranda, D. (۲۰۱۰). Some aspects about vagueness &imprecision in computer network fault-tree analysis. International Journal of Computer Commun Control, ۵۵۸-۵٦٦.
-Powers, D. M. (۲۰۱۱). Evaluation: From Precision, Recall and F-Measure to ROC Informedness, Markedness & Correlation. Journal of Machine Learning Technologies, ۳۷-٦۳.
-Qibei, L & ,.Chunhua, J. (۲۰۱۱). Research on Credit Card Fraud Detection Model Based on Class Weighted Support Vector Machine. Journal of Convergence Information Technology, ٦۲-٦۸.
-Raj, B. E & Portia, A. (۲۰۱۱). Analysis on Credit Card Fraud Detection Methods. International Conference on Computer, Communication and Electrical Technology.
-RamaKalyani, K & ,.UmaDevi, D. (۲۰۱۲). Fraud Detection of Credit Card Payment System by Genetic Algorithm. International Journal of Scientific & Engineering Research.
-Sahin, Y., Bulkan, S & Duman, E. (۲۰۱۳). A cost-sensitive decision tree approach for fraud detection. Expert Systems with Applications, ۵۹۱٦-۵۹۲۳.
-Sanjeev, J., Montserrat, G & Westland, J. C. (۲۰۱۲). Employing transaction aggreagation strategy to detect credit card fraud. Expert system with application, ۱۲٦۵۰-۱۲٦۵۷.
-Sawant, N & ,.Shah, H. (۲۰۱۳). Big data application, architecture Q&A. Apress.
-Seyedhossein, L & ,.hashemi, M. R. (۲۰۱۰). Mining Information from Credit Card Time Series for Timelier Fraud Detection. IEEE-۵th International Symposium on Telecommunications.
-Sherly, K. K. (۲۰۱۲). A comparative assessment of supervised data mining techniques for fraud prevention. International Journal of Science and Technology, ۱-٦.
-Singh, S. P., Shukla, S. S., Rakesh, N & Tyagi, V. (۲۰۱۱). Problem Reduction in Online Payment System Using Hybrid Model. International Journal of Managing Information Technology.
-Tabassian, M., Ghaderi, R & Ebrahimpour, R. (۲۰۱۲). Combination of multiple diverse classifiers using belief functions for handling data with imperfect labels. Expert Systems with Applications, ۱٦۹۸-۱۷۰۷.

-Tankard, C. (۲۰۱۲). Big data security. Network Security, ۵-۸.
-Tripathi, K. K & Pavaskar, M. A. (۲۰۱۲). Survey on Credit Card Fraud Detection Methods. International Journal of Emerging Technology and Advanced Engineering, ۷۲۱-۷۲٦.
-Tripathi, K. K & Ragha, L. (۲۰۱۳). Hybrid Approach for Credit Card Fraud Detection. International Journal of Soft Computing and Engineering, ۸-۱۱.
-Weng, X & ,.Shen, J. (۲۰۰۸). Detecting outlier samples in multivariate time series dataset. Knowledge-Based Systems, ۸۰۷-۸۱۲.
-Weston, D. J., Hand, D. J., Adams, N. M., Whitrow, C & Juszczak, P. (۲۰۰۸). Plastic card fraud detection using peer group analysis. Advances in Data Analysis and Classification, ٤٥-٦۲.
-Witten, I. H., Frank, E & Hall, M. A. (۲۰۱۱). Data mining; Practical machine learning; Tools and techniques. San Francisco: Morgan Kaufmann.
-Wong, N., Ray, P., Stephens, G & Lewis, L. (۲۰۱۲). Artificial immune systems for the detection of credit card fraud. Info Systems.
-Wu, C. H., Tzeng, G. H., Goo, Y. J & Fang, W. C. (۲۰۰۷). A real-valued genetic algorithm to optimize the parameters of support vector machine for predicting bankruptcy. Expert Systems with Applications, ۳۹۷-٤۰۸.
-Yang, J., Huang, H. Z., He, L. P., Zhu, S. P & Wen, D. (۲۰۱۱). Risk evaluation in failure mode and effects analysis of aircraft turbine rotor blades using using Dempster–
-Shafer evidence theory under uncertainty .Engineering Failure Analysis, ۲۰۸٤-۲۰۹۲.
-Yu, W. F & Wang, N. (۲۰۰۹). Research on Credit Card Fraud Detection Model Based on Distance Sum. International Joint Conference on Artificial Intelligence.
-Zareapoor, M & ,.Shamsolmoali, P. (۲۰۱٥). Application of credit card fraud detection: based on bagging ensemble classifier. Procedia computer science, ٦۷۹-٦۸٥.
-Zareapoor, M., Seeja, K. R & Alam, M. A. (۲۰۱۲). Analysis of Credit Card Fraud Detection Techniques: based on Certain Design Criteria. International Journal of Computer Applications, ۳٥-٤۲.

Evaluation and detection of fraud in the supply chain process based on the integrated approach of ANN-Big data

Elnaz Alikhani Zanjani[۱], Farid Askari[۲], Amir Najafi[۳] and Babak Hajikarimi[۴]

Abstract

In recent decades, competition to provide superior value to customers has shifted from competition between companies to competition between supply chains. Appropriate design of the supply chain with regard to the economic, social and environmental dimensions at the strategic, tactical and operational levels guarantees the survival and sustainable development of enterprises active in each part of the supply chain. One of the important issues in the supply chain is the existence of fraud and the risks associated with it throughout the chain. Also, according to the amount of scientific and experimental research conducted in recent years, the problem of detecting and evaluating fraud using computational methods alone It is an important topic for research. The most important challenges of fraud detection and evaluation are the unavailability of real data sets, the existence of asymmetric data sets, the large size of data sets, the dynamic behavior of fraudsters and the dispersion of fraudulent events, which are the result of these two Uncertainty and ambiguity in decision making. Therefore, the aim of this research is to provide a model to detect and evaluate fraud in the supply chain process based on the ANN-Big data integrated approach. In terms of its purpose, this research is a developmental and applied research. First, due to the creation of a comprehensive model, taking into account the various sources of decision-making regarding the legitimacy or fraud of electronic card transactions, and the requirements of big data analysis, it tries to develop theoretical knowledge in this field, this research can be considered Considering the nature of a mathematical analytical research, on the basis of this research, in terms of approach, it is a quantitative research. The inter-industry standard process of data mining or CRISP DM (Chapman et al. ۲۰۰۰) has been used as the analysis methodology. Clustering models and K-means algorithms, Kohonen neural network, anomaly detection based on clustering, hidden Markov model, parallel processing method of reduction mapping and fuzzy Dempster cipher fusion model are used. The analysis was done using MATLAB and R software. The proposed model has been able to show superior performance compared to other models in terms of output and execution time.

Keywords: ANN, Big Data, Supply Chain and Evaluation and detection of fraud.

[۱]Ph.D. Student of Industrial Management-Finance, Abhar Branch, Islamic Azad University, Zanjan, Iran.
[۲]Corresponding Author, Assistant Professor, Department of Financial Economics, Abhar Branch, Islamic Azad University, Zanjan, Iran. Email: **fi.asgarii@gmail.com**.
[۳]Associate Professor, Department of Management, Zanjan Branch, Islamic Azad University, Zanjan, Iran.
[۴]Assistant Professor, Department of Management, Abhar Branch, Islamic Azad University, Zanjan, Iran.