فصلنامه روند، سال بیست و سوم، شماره ۷۶، زمستان ۱۳۹۵، صفحات ۲۰۴-۱۷۳

کشف رفتارهای مشکوک در بانکداری الکترونیکی

دکتر مریم اخوان خرازیان* و شبنم بختیاری **

چکیده

امروزه یکی از مهم‌ترین موانع استفاده از خدمات بانکداری الکترونیکی، نبود امنیت و برخی سوء استفاده‌ها در مسیر انجام مبادلات مالی است. به همین دلیل، استفاده از روش‌هایی برای شناسایی رفتارهای مشکوک از مسائل مهم در مؤسسات مالی و بانک‌هاست. در این پژوهش، سعی می‌شود با استفاده از تکنولوژی داده‌کاوی، روشی نوین در کشف تقلب در کارت‌های بانکی ارائه شود. هر چند آمار دقیقی از تقلب در کارت‌های بانکی معتبر کشور وجود ندارد، ولی به نظر می‌رسد همزمان با توسعه بانکداری الکترونیک تقلب در کارت‌های بانکی رو به رشد بوده و در آینده‌ای نه چندان دور به یکی از معضلات سیستم بانکی کشور تبدیل خواهدشد. بدین‌روی، در این پژوهش پس از مصاحبه با خبرگان در زمینه کارت‌های بانکی، شناسایی انواع تقلب‌های رایج در زمینه کارت‌های بانکی از طریق اعمال تغییرات معنادار روی داده‌ها برای تهیه تراکنش‌های متقلبانه، با بهره‌گیری از روش دسته‌بندی در داده‌کاوی، سه تکنیک شبکه‌های عصبی، درخت تصمیم و نزدیکترین همسایگی، مدلی برای طبقه‌بندی تراکنش‌ها به تراکنش‌های سالم و متقلبانه ارائه می‌شود. این مدل، علاوه بر این که مبتنی بر داده‌های سیستم بانکی کشور است، توانسته است با دقت ۹۸ درصد، عملکرد نسبتاً خوبی در طبقه‌بندی یادشده داشته باشد.

واژگان کلیدی: بانکداری الکترونیکی، رفتارهای مشکوک، کارت بانکی، داده کاوی.

طبقه‌بندی JEL : M15, C38 , K42 , G21 .

* استادیار دانشکده مدیریت و حسابداری، دانشگاه شهید بهشتی m.akhavan.kharazian@gmail.com

** کارشناسی ارشد مدیریت فناوری اطلاعات sh.bakhtiari@karafarinbank.ir

۱. مقدمه

امروزه، دورنمای رقابت در صنعت بانکداری به‌طور قابل توجهی تغییر یافته است. این امر به علت نیروهایی همچون قوانین جدید، جهانی شدن، رشد فناوری، تبدیل شدن خدمات بانکی به محصول و افزایش قابل توجه تقاضای مشتریان است. تحول در فعالیت‌های بانکی و افزایش پیچیدگی بانک‌ها، باعث ایجاد مباحث جدیدی در حوزه بانکی همچون تقلب شده است. توسعه فناوری‌های جدید راه‌های زیادی را برای متقلبان و مجرمان باز کرده که بتوانند مرتکب تقلب شوند. ایجاد یک سیستم اطلاعاتی جدید و تکنیک‌های شناسایی تقلب، علاوه بر آنکه تقلب‌ها و کلاهبرداری‌های صورت گرفته در یک سازمان را شناسایی کرده و مورد تجزیه و تحلیل قرار می‌دهد، به نوعی با شناخت رفتار کاربران یا مشتریان سعی در پیش‌بینی رفتار آتی آنها داشته و ریسک انجام تقلب‌ها را کاهش می‌دهد.[۱]

به دلیل وجود هزینه‌های فراوان مستقیم و غیرمستقیم تقلب، بانک‌ها و مؤسسات مالی و پولی به‌شدت به دنبال تسریع و سرعت عمل در شناخت فعالیت‌های کلاهبرداران و متقلبان هستند. این امر به دلیل اثر مستقیم آن روی خدمت‌رسانی به مشتریان بانک‌ها و مؤسسات، کاهش هزینه‌های عملیاتی و باقی‌ماندن به عنوان یک ارائه‌دهنده خدمات مالی معتبر و قابل اطمینان است. بنابراین، به‌کارگیری تکنیک‌های شناسایی تقلب به منظور جلوگیری از اقدامات متقلبانه در سیستم بانکداری، اجتناب‌ناپذیر است.

امروزه پیچیدگی سازمان‌ها و تراکنش‌ها باعث افزایش فرصت برای استفاده‌های سودجویانه و تقلب شده است. این تقلب‌ها علاوه بر ضررهای اقتصادی، اثرات روانی گوناگونی بر روی خود بانک و مشتریان آن خواهد داشت. ممکن است شهرت بانک و سطح رضایت مشتریان آسیب دیده و سطح اعتماد مشتریان نسبت به خدمات جدید کاهش یابد. در بُعد درون سازمانی نیز ممکن است، فرآیند مدیریتی سازمان با اختلال مواجه شده و اخلاق و وجدان کاری کارکنان تحت اثر منفی آن قرار گیرد. خدمات نوین ارائه‌شده توسط بانک‌ها نیز حوزه‌های جدیدی از تقلب را گشوده و اثرات منفی آن

۱. حاتمی راد، شهریاری .(۱۳۹۰).

ضرورت اقدامات شناسایی تقلب را توجیه‌پذیر ساخته است. فنون شناسایی تقلب، شامل شگردهای جست‌وجوی پیچیده‌ای است که از طریق بررسی تراکنش‌ها و رفتار مصرفی مشتری، الگوهای تقلب را کشف و به موقع اعلام می‌کند. از سوی دیگر، آگاهی از انواع مختلف تقلبات بانکی نیز برای اقدامات پیشگیرانه می‌تواند مفید باشد و بانک‌ها و مؤسسات مالی در صورت آشنایی با انواع مختلف تقلبات بانکی می‌توانند فرآیندهای نظارتی مناسبی را طراحی کنند.

در صنعت بانکداری نیز به دنبال رشد و توسعه بانکداری نوین، پیشرفت‌های فناوری اطلاعات و در دسترس بودن امکانات کامپیوتری پیشرفته به‌منظور ذخیره داده‌ها، حجم عظیمی از داده‌ها در دسترس تصمیم‌گیرندگان قرار دارد که با توجه به وضعیت رقابتی موجود، تصمیم‌گیری سریع، امکان تبدیل فرصت‌ها و تهدیدها به فرصت‌های طلایی، صنعت بانکداری را به سمت استفاده از تکنیک‌های داده‌کاوی ترغیب می‌کند. داده‌کاوی به عنوان تکنیکی خاص در جهت دستیابی به اطلاعات مفید و مناسب از اهمیت ویژه‌ای در سال‌های اخیر برخوردار شده است که به صورت گسترده‌ای در صنایع مختلف مانند بانکداری، هوا و فضا، بهداشت و شناسایی الگوهای مفید و ارتباطات بین داده‌های ثبت شده و غیره مورد استفاده قرار می‌گیرد. هدف داده‌کاوی، کشف و شناسایی الگوی جدید در داده‌هاست. در این رابطه، نوعی احساس خلأ در شناسایی تقلب‌ها در صنعت بانکداری با رویکرد داده‌کاوی به چشم می‌خورد. بنا بر اعلام دانشگاه MIT دانش نوین داده‌کاوی یکی از ده دانش در حال توسعه‌ای است که دهه آینده را با انقلاب تکنولوژی مواجه می‌سازد. این تکنولوژی امروزه کاربرد بسیار وسیعی در حوزه‌های مختلف دارد؛ به‌گونه‌ای که امروزه حد و مرزی برای کاربرد این دانش در نظر نگرفته‌اند و زمینه‌های کاری این دانش را از ذرات کف اقیانوس‌ها تا اعماق فضا می‌دانند.

تاکنون در سیستم بانکی کشور سازوکار و برنامه جامعی برای شناسایی و جلوگیری از تقلب‌های مربوط به تراکنش‌های مبتنی بر کارت وجود نداشته است،[1] در صورتی که یکی از وظایف مهم بانک‌ها نظارت بر صحت تراکنش‌های بانکی، حفظ مشتریان و کاهش ریسک بانکی است. از این رو ایجاد و پیاده‌سازی سیستمی توسط بانک‌ها به منظور شناسایی تقلب کارت‌های بانکی ضروری است.

---

۱. نوبرزاده .(۱۳۹۱)؛ حاتمی راد و شهریاری .(۱۳۹۰).

در کشورهای دیگر هم بـه دلیـل گسـتردگی اسـتفاده از کـارت‌هـای اعتبـاری، پـژوهش‌هـای انجام‌شده بیشتر بر این کارت‌ها تمرکز داشته است، در حالی که استفاده از این کارت‌هـا در کشـور مـا مرسوم نبوده و همه تراکنش‌ها براساس کارت‌های پرداخت نقدی صورت می‌گیرد. بنابراین، اسـتفاده از مدل‌های طراحی شده در پژوهش‌های کشورهای دیگر چندان مقدور نیست.

با توجه به حجم گسترده تراکنش‌های بانکی روزانه، تغییر در فرآیندهای بانکی و سیسـتم‌هـای جدید عملیات کارت، نیاز به تشخیص به موقع تقلب‌ها و جلوگیری از وقوع آنها، به یکی از فرآینـدهای مهم در عملیات بانکی تبدیل شده و چون در عمل، شناسایی دستی امکان‌پذیر نیست و مستلزم صـرف زمان و نیروی کاری کارآمد خواهد بود، در این راستا، بانک‌ها و مؤسسات مالی، با تجهیز به سیستم‌های کشف تقلب می‌توانند به پیشگیری بپردازند. در این پژوهش، سعی می‌شـود بـا اسـتفاده از تکنولـوژی داده‌کاوی، روشی نوین در کشف تقلب در کارت‌های بانکی ارائه شود.

هدف اصلی پژوهش، شناسایی تقلب در سیستم بـانکی بـه منظـور حفـظ و نگهـداری بانـک و مشتریان است. مزیتی که برای بانک دارد، حفظ مشتریان و ایجاد اطمینـان خـاطر بیشـتر بـرای آنهـا به‌منظور استفاده از خدمات بانک و مزیتی که برای مشتری دارد، کاهش ضرر و زیان وارده به مشتریان است.

هدف دیگر پژوهش، استفاده از داده‌کاوی در مبحـث شناسـایی تقلـب در کـارت‌هـای بـانکی و شناسایی بهتر تراکنش‌های متقلبانه از میان تراکنش‌ها و کارایی در سیستم بانکداری است.

۲. پیشینه پژوهش

بانک‌ها جزء سازمان‌هایی هستند که به‌طور مسـتقیم بـا مشـتریان در تعامـل هسـتند. بنـابراین، تحلیـل رفتاری مشتریان برای افزایش وفاداری آنها از اهمیت بالایی برخوردار است. در سال‌های اخیر با افـزایش دسترسی به داده‌هـای مشـتریان و بهبـود قابلیـت‌هـای تحلیـل داده‌هـا بـه‌وسـیله روش‌هـای هوشـمند، فعالیت‌های مختلفی به منظور تحلیل رفتار مشتریان انجام شده است. یکی از این فعالیت‌هـا، اسـتفاده از سیستم‌های هوشمند در کشف تقلب‌های بانکی است. تقلب‌های بانکی در حال حاضر طیـف گسـترده‌ای یافته و باعث ضررهای مادی و غیرمادی بسیاری به بانک‌ها و مشتریان بانک شده‌اند. بنابراین، آشنایی بـا

این گونه تقلب‌ها علاوه بر اینکه در پیشگیری از وقوع آنها استفاده می‌شود، در حوزه تحلیل و طراحی سیستم‌های اطلاعاتی مورد نیاز نیز می‌تواند بسیار مفید باشد.[۱]

در مقالات و منابع علمی، تقلب در کارت‌های بانکی به روش‌های گوناگونی تعریف شده است که چکیده این تعاریف را می‌توان این گونه جمع‌بندی کرد: تقلب در کارت‌های بانکی به کلاهبرداری یا تقلب به وسیله کارت بانکی یا هرگونه ساز وکار پرداخت مشابه گفته می‌شود که از منبع متقلبانه در تراکنش انجام می‌شود.[۲]

به دلیل کمبودهای امنیتی سیستم پردازش کارت‌های بانکی مرسوم، تقلب در آنها روند افزایشی دارد. تقلب در کارت‌های بانکی به یکی از منابع جذاب کسب درآمد برای مجرمان تبدیل شده است. به همین دلیل مسأله تقلب برای بانک‌ها و مؤسسات مالی اهمیت ویژه‌ای دارد.[۳] پیشگیری و شناسایی تقلب بخش مهمی از مدیریت ریسک بانک‌هاست. هدف از شناسایی سریع تقلب، متوقف کردن آن در کوتاه‌ترین زمان ممکن پس از وقوع آن است. انواع تقلب‌های بانکی که تاکنون شناسایی شده است، تقلب کارت‌های پرداخت، مسئولان فاسد، وام‌های تقلبی، تقلب‌های سیمی، اسناد دستکاری شده و تقلبی، سپرده‌های مشکوک، تقلب برات دیداری، چک‌های دستکاری شده و تقلبی، چک‌های مفقودشده، تقلب در صورتحساب‌های بانکی، چک بازی، چک‌های بی‌محل، کارت‌های پرداخت مفقودشده، دزدی اطلاعات کارت‌های بانکی، جعل هویت، درخواست وام‌های تقلبی، بازرسان بانکی جعلی، تقلب‌های اینترنتی و فیشینگ و پولشویی است. البته، تقلب به موارد یادشده خلاصه نشده و ممکن است روش‌های دیگری هم از تقلب وجود داشته باشد که هنوز کشف نشده است. هر چه بانک‌ها برای پایین آوردن ریسک و تأمین امنیت روش‌هایی را ارائه می‌دهند، از سوی دیگر، جاعلان در پی یافتن راه‌های جدید تقلب هستند.

در طول دهه‌های گذشته حجم زیادی از داده‌ها در پایگاه داده‌ها انباشته و ذخیره شده‌است. در واقع، سازمان‌ها در اطلاعات غرق شده‌اند، در حالی که تشنه دانش هستند. این موضوع نشان می‌دهد که

---

۱. معین زاده .(۱۳۹۰).

2. Delamaire, et al. (2009)., Patidar and Sharma. (2011)., Phua, etal. (2010).

۳. نصیری و مینایی .(۱۳۸۹).

سازمان‌ها نتوانسته‌اند از دانش درون داده‌ها به نحو مناسب استفاده کنند. با توجه به تنوع زیاد مخاطبان، مشتریان، بازارها، تنوع و پیچیدگی خدمات و محیط‌های کسب و کار و لزوم دسترسی به اطلاعات مناسب برای تصمیم‌گیری صحیح و به موقع، استفاده از راهکارهای مناسب برای طبقه‌بندی و یافتن اطلاعات کاربردی و اثربخش از میان انبوهی از داده‌ها برای سازمان‌ها ضروری و حیاتی بوده و یک تخصص و هنر محسوب می‌شود. در واقع، داده‌کاوی پاسخی به این نیاز سازمان‌ها و مؤسسات است. هرچه حجم داده‌ها بیشتر و روابط میان آنها پیچیده‌تر باشد، دسترسی به اطلاعات نهفته در داده‌ها مشکل‌تر شده، بنابراین، نقش داده‌کاوی به عنوان یکی از روش‌های کشف دانش، روشن‌تر می‌شود.[۱]

روش‌های داده‌کاوی به عنوان یکی از اصلی‌ترین ابزارهای شناسایی تقلب در کارت‌های بانکی استفاده می‌شود. داده‌کاوی، فرآیند کشف روابط ناشناخته و الگوی درون داده‌هاست، در واقع، فعالیتی است که به‌طور اساسی با آمار و تحلیل دقیق داده‌ها انطباق دارد.[۲]

راهبردهای کلان مسائل شناسایی تقلب در حوزه کارت‌های بانکی را نیز می‌توان منطبق با راهبردهای داده‌کاوی دانست. دو راهبرد کلان برای فرآیند داده‌کاوی وجود دارد: ۱. یادگیری نظارت شده[۳] و ۲. یادگیری نظارت نشده.[۴] روش‌های نظارت شده، از یک پایگاه داده شامل موارد متقلبانه و غیرمتقلبانه ساختاریافته استفاده می‌کنند و در موارد جدید مشکوک به تقلب به کار می‌روند. یادگیری نظارت شده از داده‌های گذشته یاد می‌گیرد و دانش آموخته‌شده را در موارد بعدی به کار می‌برد. این فرآیند تلاش می‌کند الگوهای از پیش تعریف‌شده معین از فعالیت تراکنش‌هایی را شناسایی کند که برای مطابقت دادن با فعالیت‌های متقلبانه به کار می‌روند.

---

۱. شهرابی .(۱۳۸۶).

۲. آذر، احمدی و سبط .(۱۳۸۹).

3 . Supervised Learning

4. Unsupervised Learning

در روش نظارت نشده، سیستم بدون در اختیار داشتن داده‌های خروجی و بدون کمک خارجی، درستی یا نادرستی سیگنال‌های خروجی خود را مشخص می‌کند.[۱] تکنیک‌ها و مدل‌های مورد استفاده درخصوص مسأله کشف تقلب در سال ۲۰۰۲ توسط بولتون و هند[۲] بررسی شد. آنها مدل‌های کشف تقلب در حوزه کارت‌های اعتباری را با دو رویکرد کلی نظارتی و غیرنظارتی طبقه‌بندی کردند. به کمک این روش، حساب‌هایی که در این بازه زمانی مشخص الگوی رفتاری متفاوتی از خود نشان می‌دهند، شناسایی می‌شوند.

درخت تصمیم، یکی از روش‌های داده‌کاوی با قابلیت فهم زیاد و سرعت مناسب در یادگیری الگو است.[۳] شن و همکارانش علاوه بر چارچوب‌های دیگر ارائه شده، درخت تصمیم را نیز آزموده و با مطالعات دیگر مقایسه کردند.[۴] درخت تصمیم یکی از روش‌های طبقه‌بندی است. هر تراکنش دارای مجموعه مشخصاتی است که بر اساس مقادیر آنها، تراکنش به یک طبقه تعلق می‌گیرد، بنابراین، هدف از طبقه‌بندی، ساختن تابعی است که هر تراکنش را بر اساس مقادیر مشخصاتش به یکی از چندین گروه از پیش تعیین شده، نگاشت کند. در مورد این موضوع، کوثری لنگری و همکاران، پژوهشی را با عنوان "به‌کارگیری الگوریتم‌های درخت تصمیم‌گیری به‌منظور کشف رفتارهای مشکوک در بانکداری الکترونیک"، انجام داده‌اند. در این پژوهش از دانش خبرگان و دسته‌بندی الگوی رفتاری کاربران توسط الگوریتم درخت تصمیم استفاده شده و چهار الگوریتم chaid، chaid_ex، c4.5، c5 مقایسه‌شده و الگوریتم C5 با دقت ۹۱ درصد الگوریتم بهینه محسوب شده است. لئونارد از سیستم خبره مبتنی بر قوانین، برای شناسایی تقلب کارت اعتباری استفاده کرده است.[۵] پژوهشی دیگر، دو نظریه داده‌کاوی پیشرفته ماشین بردار پشتیبان[۶] و جنگل‌های تصادفی[۷] را با استفاده از رگرسیون لجستیک[۸] ارزیابی

1. Paasch. (2008).
2. Bolton and Hand. (2002).
۳. البرزی، محمد پورزندی و خان بابایی .(۱۳۸۹).
4. Shen, Tong and Deng. (2007).
5. Leonard. (1995).
6. Support Vector Machine
7. Random Forests
8. Logistic Regression

کرده است. این پژوهش بر اساس داده‌های واقعی تراکنش‌های بین‌المللی کارت اعتباری انجام شده است.[۱]

در سال ۲۰۱۱ مدلی مبتنی بر قوانین برای شناسایی و مقابله با تراکنش‌های متقلبانه (برای تقلب‌های بدون استفاده از کارت) در سیستم‌های پرداخت الکترونیکی ارائه شده است. در این روش، با تعریف الگوریتم یادگیری مبتنی بر قوانین، به طبقه‌بندی تراکنش‌ها به تراکنش‌های "سالم" و "متقلبانه" پرداخته شده است.[۲] استفاده از شبکه‌های عصبی مصنوعی برای طبقه‌بندی در بسیاری از زمینه‌ها، کاربرد فراوانی دارد که یکی از ویژگی‌های آنها، خاصیت یادگیری نظارت نشده است.[۳] شبکه‌های عصبی مصنوعی نیز یکی از روش‌هایی است که برای شناسایی تقلب در کارت‌های بانکی استفاده می‌شود. برتری شبکه‌های عصبی نسبت به روش‌های دیگر این است که می‌تواند از تراکنش‌های گذشته بیاموزد و با گذشت زمان نتایج را بهبود دهد. همچنین، می‌تواند قوانین را استخراج کرده و رفتار آینده را بر اساس وضعیت فعلی پیش‌بینی کند.[۴] آگوئلکا برنامه شبکه عصبی مصنوعی کاربردی ای برای خوشه‌بندی طراحی کرد، این برنامه می‌تواند از حجم بزرگی از داده‌های تراکنش‌ها استفاده کند. در این پژوهش از چهار خوشه با ریسک زیاد، متوسط، پایین و کم ریسک استفاده شده است؛ به این شکل که تراکنش‌های پردازش شده در یکی از این خوشه‌ها قرار خواهد گرفت.[۵]

همچنین، چان و همکارانش الگوریتمی را به منظور پیش‌بینی رفتار مشکوک ایجاد کردند. در حالی که مطالعات دیگر از ارزیابی مبتنی بر درصد پیش‌بینی، درصد مثبت درست و درصد منفی نادرست استفاده می‌کنند، اساس این پژوهش ارزیابی به کمک مدل هزینه است.[۶] در پژوهش فیروزی و همکاران با عنوان "شناسایی تقلب در بیمه اتومبیل با استفاده از داده‌کاوی از سه الگوریتم رگرسیون

1. Battacharyya, et al. (2011)..
2. Al-khatib. (2011).
۳. قاسمی و اصغری زاده .(۱۳۹۳).
۴. نصیری و مینایی .(۱۳۸۹).
5. Ogwueleka. (2011).
6. Chan, et al.(1999)..

لجستیک"، از بیز ساده و درخت تصمیم برای پیداکردن الگو و شناسایی تقلب استفاده شده است. الگوریتم بیز با ۹۰,۲۸ درصد در شناسایی تقلب بهترین کارایی را دارد.

محقر و همکارانش، روش‌های کشف تقلب در بانکداری را به دو دسته اصلیِ "روش‌های آماری" و "روش‌های هوش مصنوعی" تقسیم کرده و به بررسی امکان استفاده از روش مبتنی بر هوش کسب وکار پرداختند.[۱]

کوندو و همکارانش از روشی مبتنی بر مدل مخفی مارکوف[۲] برای تشخیص تقلب استفاده کرده‌اند. این روش نیز پارامترهای مربوط به تراکنش‌ها را به کار برده و با استفاده از رفتارهای دریافتی مشتریان الگوهایی را می‌سازد و سپس با استفاده از امتیازدهی سریع به هر تراکنش وارده بر اساس پارامترهای از پیش مشخص‌شده، تصمیم می‌گیرد که تراکنش پذیرفته و یا رد شود. این روش یک روش آماری است که از مقادیر پارامترها در یک محدوده خاص استفاده کرده و با محاسبه امتیاز تراکنش بر اساس مقادیر منصوب‌شده به پارامترها، احتمال متقلبانه بودن آن را تعیین می‌کند. این مقدار بین صفر و یک است. هر چه عدد به‌دست آمده نزدیک به یک باشد، مشکوک‌تر بوده و می‌تواند مؤلفه هشدار سیستم تشخیص تقلب را راه‌اندازی نماید.[۳]

دومان و همکارش از روشی مبتنی بر تطبیق دنباله‌ها برای تشخیص تقلب در کارت‌های اعتباری استفاده کرده‌اند. تطبیق دنباله‌ها در بایوانفورماتیک برای شباهت بین رشته‌های ژنی استفاده می‌شود و ابزاری برای اندازه‌گیری و ارزیابی شباهت بین دو یا چند دنباله است. در پردازش تراکنش‌های کارت‌های اعتباری، دنبالهٔ استفاده (رفتارهای پیشین مشتری) شامل اطلاعاتی راجع به مبلغ تراکنش، فاصله زمانی از آخرین خرید، روز، هفته و جزاینها، برای صادرکننده کارت در دسترس است. هر انحراف دنبالهٔ استفاده از دنبالهٔ موجود (تراکنش مورد بررسی) می‌تواند از طریق تطبیق دنباله‌ها محاسبه شود. در این پژوهش، به علت عمل تطبیق و مقایسه تک‌تک دنباله‌ها با یکدیگر، عمل

---

۱. محقر، لوکس، حسینی و منشی .(۱۳۸۷).

2. Markov

3. Kundo, et al. (2009)..

پردازش وقت‌گیر است، ضمن اینکه در این پژوهش به نکته تقلید رفتار فرد متقلب از رفتار صاحب اصلی کارت به هنگام عمل تقلب، که عموماً یکی از ترفندهای متقلبان است، توجهی نشده است.[1]

در دسته‌بندی این روش‌ها، مرزبندی چندان دقیقی وجود ندارد، چرا که هریک از این روش‌ها فقط شکلی از یک روش علمی است و برخی از آنها می‌توانند به یکدیگر تبدیل شوند. یادآوری می‌شود، هیچ یک از این روش‌ها به تنهایی نمی‌توانند تقلب را حذف کنند، در واقع، هر روش توانایی یک سیستم را در شناسایی تقلب افزایش می‌دهد.

1. Duman and Hamdi. (2011)..

جدول ۱. خلاصه‌ای از پژوهش‌های خارجی

پژوهشگر	سال انتشار	عنوان	نتیجه
Emanuel Mineda Carneiro; Luiz Alberto Vieira Dias	۲۰۱۵	Cluster Analysis and Artificial Neural Networks: A Case Study in Credit Card Fraud Detection	از مدل شبکه عصبی و تجزیه و تحلیل خوشه‌ای برای کشف تقلب کارت‌های اعتباری استفاده شده است. نتایج به‌دست آمده در استفاده از این دو مدل بدین صورت است که ورودی های عصبی را می‌توان با ویژگی‌های خوشه‌ای کاهش داده و قدم‌های مثبتی در کشف تقلب برداشت.
Dr. H. Ali ATA . Brahim .Dr H. SEYREK	۲۰۱۲	The Use Of Data Mining Techniques Detecting Fraudulent Financial Statements	طبقه‌بندی: درخت‌های تصمیم‌گیری و شبکه‌های عصبی استفاده کرده است. برای تجزیه و تحلیل داده‌ها تقریباً نیمی از موارد به‌طور تصادفی انتخاب شده (۴۷ از ۱۰۰) برای آموزش مدل و موارد باقی مانده (۵۳ از ۱۰۰) استفاده شد. برای اعتبار مدل در مرحله آموزش، طبقه‌بندی درخت تصمیم ، ۹۵،۷۴ درصد از موارد درست و ۴،۲۶ درصد اشتباه بود. شبکه داده‌های عصبی، مدلی بهتر از مدل درخت تصمیم است.
Rekha Bhowmik	۲۰۱۰	Detecting Auto Insurance Fraud by Data Mining Techniques	برای مدل پیش‌بینی از الگوریتم‌های بیز ساده و بی‌تکلف و درخت تصمیم استفاده کرده و این پیش‌بینی برای تشخیص تقلب و تعیین درستی از ماتریس سردرگمی استفاده می‌کند.
S.Bhattacharyya, S.Jha, K.Tharakunnel, and J.Westland	۲۰۰۹	Data Mining for Credit Card Fraud: A Comparative Study	روش‌های داده‌کاوی به نام های ماشین‌های بردار پشتیبان (SVM)، جنگل تصادفی (RF) و رگرسیون منطقی (LR) برای کشف تقلب تشریح و سپس مقایسه شده است. مجموعه داده استفاده‌شده در این پژوهش نیز شامل ۷۹ میلیون تراکنش‌های بانکی مربوط به سال ۲۰۰۸ از یک کشور خاص بوده است. پژوهشگران برای اینکه ارزیابی دقیق‌تری از الگوریتم‌ها داشته باشند، داده‌های آموزشی را به ۴ برش تقسیم کردند تا الگوریتم‌ها را با نرخ تقلب متفاوت مقایسه کنند. همچنین، آنها از معیارهای متفاوتی برای مقایسه الگوریتم‌ها بهره برده‌اند.
N.C.Hsieh	۲۰۰۸	Hybrid Mining Approach in Design of Credit Scoring Models	به دنبال ایجاد مدل ترکیبی برای امتیازدهی به اعتبار کارت‌های اعتباری بوده است. وی هم با ترکیب الگوریتم K-means از نمونه الگوریتم‌های خوشه‌بندی و الگوریتم شبکه عصبی به دنبال تأثیر منفی مجموعه داده پرت بر روی مدل نهایی بوده است. مدل ارائه‌شده برای برآورد اعتبار کارت‌ها بر روی دو پایگاه داده آزمون شده است. در واقع، با به کارگیری الگوریتم‌های K-means و فازی در کنار یکدیگر میزان دقت مدل به‌دست آمده تا ۹۹ درصد بهبود یافته است.

مأخذ: یافته‌های پژوهش.

جدول ۲. خلاصه‌ای از پژوهش‌های داخلی

پژوهشگر	سال انتشار	عنوان	نتیجه
کوثری لنگری، مقدم چرکری	۱۳۸۹	به‌کارگیری الگوریتم‌های درخت تصمیم‌گیری جهت کشف رفتارهای مشکوک در بانکداری الکترونیک	از دانش خبرگان و دسته‌بندی الگوی رفتاری کاربران توسط الگوریتم درخت تصمیم استفاده شده و چهار الگوریتم chaid، chaid_ex، c4.5، c5 مقایسه شده و الگوریتم C5 با دقت ۹۱ درصد الگوریتم بهینه محسوب شده است.
فیروزی، شکوری، کاظمی	۱۳۹۰	شناسایی تقلب در بیمه اتومبیل با استفاده از داده‌کاوی	از سه الگوریتم رگرسیون لجستیک، بیز ساده و درخت تصمیم برای پیدا کردن الگو و شناسایی تقلب استفاده شده است. الگوریتم بیز با ۹۰،۲۸ درصد در شناسایی تقلب بهترین کارایی را دارد.
سلطانی، اکبری، سرگلزایی	۱۳۹۱	ارائه مدل پیاده سازی سیستم کشف تقلب کارت‌های اعتباری در محاسبات ابری با استفاده از سیستم ایمنی مصنوعی	مدل پیشنهادی با دید کاربردی به الگوریتم سیستم ایمنی مصنوعی نگریسته و درصدد بهبود دقت کشف تقلب است. در مدل معرفی شده با بهره‌گیری از مفهوم انتراب منفی، سلول‌های بهتر در سیستم ابقا می‌شوند. همچنین، به روزرسانی مدل در فاز کشف موجب یادگیری روش‌های تقلب جدید می‌شود.
ابراهیمی، منجمی، سرورنژاد	۱۳۹۲	تشخیص تقلب در تراکنش‌های کارت‌های اعتباری با الگوریتم سلول‌های دندریتیک	الگوریتم سلول‌های دندریتیک، الگوریتمی مناسب با دقت مناسب ۹۰ درصد به‌منظور تشخیص تقلب در تراکنش‌های کارت‌های اعتباری است. این الگوریتم به دلیل اعمال علامت امن که باعث کاهش اثر علامت PAMP و خطر شده و تعداد شناسایی اشتباه تراکنش‌های نرمال به عنوان تقلب را کاهش می‌دهد، میزان هشدار مثبت نادرست پایینی دارد و به علت ماهیت نامتوازن بودن پایگاه داده‌های تقلب میزان هشدار منفی نادرست نسبتاً بالایی را نیز تولید می‌کند. به دلیل نداشتن فاز آموزش برای پایگاه داده‌های آنلاین (پرکاربرد در تشخیص تقلب آنلاین) نیز بسیار مناسب است.
سیادت، رباب اسمعیلی، عزیزی	۱۳۹۴	بررسی و مقایسه روش‌های داده‌کاوی برای کشف تقلب در تراکنش‌های کارت‌های بانکی	یکی از دلایل محدودکردن تعداد مقالات در زمینه کشف تقلب را سختی به‌دست آوردن داده‌های پژوهش می‌دانند. از سوی دیگر، برای شروع کار باید داده‌های آموزش دیده مدل را به دو دسته متقلب و غیرمتقلب تقسیم کرد که این خود چالشی مهم است. با وجود تکنیک‌های داده‌کاوی موجود که به‌کار گرفته می‌شوند، این تکنیک‌ها هم معایب و محدودیت‌های خود را دارند؛ چرا که هر یک از این روش‌ها دامنه کاربرد محدودیت‌های ویژه‌ای را شامل می‌شوند.

مأخذ: یافته‌های پژوهش.

۳. انواع کارت‌های بانکی

شاید بتوان از آمریکا به عنوان نخستین کشور ارائه‌دهنده کارت‌های اعتباری نام برد. پدیدآمدن ایده کارت اعتباری به سال‌ها پیش بر می‌گردد که نخستین بار در سال ۱۸۶۰ توسط ادوارد پلای دانشمند بزرگ آمریکایی ارائه شد.[۱]

در طول این سال‌ها کارت‌های اعتباری، کارت‌های بدهی، کارت‌های پیش پرداخت و کارت‌های هوشمند عرضه شده است. تمرکز این پژوهش بر روی کارت‌های پرداخت نقد است. این نوع کارت، امکانی برای پرداخت وجه به جای ارائه پول نقد در بسیاری از خریدهای رایج است. بنابراین، در مقایسه با کارت‌های اعتباری که شیوه‌ای برای پرداخت در آینده هستند، کارت پرداخت نقد ابزار پرداخت در لحظه خرید و در بسیاری از سوپر مارکت‌ها، فروشگاه‌ها و رستوران‌های مورد پذیرش است. از این رو کارت پرداخت نقد شیوه مناسبی برای کنترل و مدیریت هزینه است؛ زیرا فقط می‌توانید به اندازه‌ای که پول در حساب دارید، خرج کنید و بر خلاف کارت‌های اعتباری نمی‌توانید پرداخت‌ها را به پایان ماه یا ماه‌های بعد موکول کنید. البته، در بسیاری از کشورها کارت پرداخت نقد یک کارت چند منظوره است، به‌طور مثال، به عنوان کارت خودپرداز برای دریافت پول از دستگاه‌های خودپرداز نیز مورد استفاده قرار می‌گیرد. در هر حال، انواع کارت‌های پرداخت نقد از ابزارهای رایج پرداخت اینترنتی هستند که استفاده از آنها در سراسر دنیا از جمله ایران رو به گسترش است.[۲]

۴. راه‌های تقلب در کارت‌های بانکی

در ساده‌ترین این روش‌ها کارمند یک سازمان می‌تواند تصویری از رسید مشتری را نزد خود نگه داشته و از آن برای مقاصد بعدی خود استفاده کند. کپی‌کردن از اطلاعات کارت و سپس، سرقت از حساب مشتریان سابقه زیادی در حوزه کلاهبرداری از طریق خودپردازها دارد؛ به‌طوری که، این مسأله در سال ۲۰۱۰ روی بیش از ۲۵ درصد عابر بانک‌هایی که کمتر از شش ماه از نصب آنها گذشته بود، اتفاق افتاده است.

1. Yu, et al. (2002)..
2. Bhatl, et al. (2003).

اصولاً در حوزه خودپردازها تقلب امری رایج محسوب می‌شود. در کارت‌های مغناطیسی امکان سرقت اطلاعات از طریق دوربین، نصب نرم‌افزاری درون خودپرداز یا نصب پایانه فروشگاهی کاذب روی کارتخوان اصلی وجود دارد. نوع دیگری از کلاهبرداری زمانی روی می‌دهد که افراد به سیستم نرم‌افزاری بانک نفوذ کرده و با استفاده از اطلاعات سرقت شده مشتریان، از راه‌های مختلفی از جمله صدور کارت‌های تقلبی به برداشت از حساب مشتریان می پردازند.[۱]

برخی از پژوهش‌هایی که به بررسی عملکرد مدل‌های شناسایی تقلب پرداخته‌اند، تفاوت بین انواع مختلف سرقت‌ها را در نظر نگرفته‌اند. واردکردن چنین مسأله‌ای به مدل‌ها اگر چه ممکن است به پیچیده‌تر شدن مدل منجر شود، اما پاسخ‌های قابل قبول‌تری به سازمان ارائه می‌دهد. به‌طور معمول پس از آن که فردی از گم‌شدن کارت خود اطلاع پیدا می‌کند، در نخستین فرصت موضوع را به بانک صادرکننده کارت اطلاع می‌دهد تا از سوء استفاده‌های احتمالی جلوگیری کند. آسان‌ترین راه سرقت در حالی اتفاق می‌افتد که فرد یابنده پیش از اعلام گم‌شدن کارت به بانک و باطل‌کردن آن، با امتحان‌کردن تعدادی رمز از حساب فرد پول برداشت کند.[۲]

در روشی دیگر، هنگامی که اطلاعات یک کارت سرقت می‌شود، فرد سارق می‌تواند ماه‌ها از کارت فردی دیگر استفاده کند، بدون اینکه صاحب اصلی اطلاعی از این موضوع داشته باشد. این نوع از کلاهبرداری‌ها هنگامی اتفاق می‌افتد که شماره کارت، تاریخ انقضا، نام صاحب کارت و یا کد اعتبارسنجی کارتی، ندانسته در اختیار فردی دیگر قرار می‌گیرد. این مسأله به‌ویژه هنگام خرید از فروشگاه‌ها می‌تواند رخ دهد.[۳]

گاهی ممکن است خریدها با اطلاعات کارتی که وجود ندارد، انجام شود. برای مثال در خریدهای اینترنتی و تلفنی که خریدار و فروشنده به صورت غیرحضوری با یکدیگر معامله می‌کنند،

۱. شهرابی .(۱۳۹۰).
۲. شهرابی .(۱۳۹۰).
۳. همان مأخذ.

فروشنده راهی برای تشخیص این که آیا فردی که اطلاعات را به او داده است، صاحب اصلی کارت است یا خیر، ندارد.

در نوع دیگری از سرقت‌ها، اطلاعات کارتی که به تازگی صادر و به فرد متقاضی فرستاده شده است، پیش از رسیدن به فرد اصلی سرقت می‌شود. معمولاً بانک‌ها برای فعال‌کردن اینترنتی کارت‌ها از متقاضیان خود اطلاعاتی مانند تاریخ تولد را می‌پرسند؛ بنابراین، در این روش فرد سارق نیاز به دانستن برخی از اطلاعات شخصی صاحب کارت دارد. این نوع کلاهبرداری، ریسک زیادی را برای بانک و صاحب کارت به دنبال دارد. سارق می‌تواند تمام موجودی حساب را یکباره برداشت کند. پس از آن، به‌ویژه در کشورهایی که ارائه خدمات بانکی و بیمه‌ای بیشتر به اعتبار افراد بستگی دارد، سال‌ها زمان لازم است تا فرد قربانی اعتبار قبلی خود را نزد بانک به‌دست آورد.

اساساً در تمام دنیا کلاهبرداری از طریق خودپردازها و ابزارهای دیگر الکترونیک وجود دارد. معمولاً فردی که اطلاعات او به سرقت رفته است، تا زمان اطلاع یافتن از خالی‌شدن حساب خود از این مسأله بی‌خبر می‌ماند. بانک‌ها نیز احتمالاً با برداشت‌های غیرعادی و مشکوکی روبه‌رو نمی‌شوند. به این معنا که به ظاهر همه تراکنش‌های مالی فرد، معمولی و در سقف مجاز خود به نظر می‌آید. امروزه با پیشرفت تکنولوژی که به پیچیده‌تر شدن روش‌های انجام سرقت منجر شده است، شناسایی چنین فعالیت‌هایی نیاز به روش‌های نوین در این زمینه دارد. از این رو، تنها راه جلوگیری از بروز چنین سرقت‌هایی، آگاهی بانک‌ها از ویژگی‌های رفتاری مشتریان خود است. از سوی دیگر، افزایش تعداد مشتریان، ریسک انجام سرقت و نفوذ به پایگاه‌های داده سازمان‌ها را نیز افزایش داده است. تا زمانی که داده‌های یک بانک به‌صورت رکوردهای ماهانه و یا تراکنش‌های روزانه وجود دارد، تحلیل چنین داده‌هایی و کشف موارد مشکوک بدون استفاده از ابزارهای مناسب، بسیار مشکل و چه بسا نشدنی خواهد بود. از این رو امروزه مسأله رفتارسنجی مشتریان به وسیله تکنیک‌های داده‌کاوی به یکی از مسائل مهم مدیران و کارشناسان بانکی تبدیل شده است.[۱]

۱. همان مأخذ.

۵. داده‌کاوی در بانکداری

در طول دهه‌های گذشته حجم زیادی از داده‌ها در پایگاه داده‌ها انباشته و ذخیره شده‌است. در واقع، سازمان‌ها در اطلاعات غرق شده‌اند، در حالی که تشنه دانش هستند. این امر نشان می‌دهد که سازمان‌ها نتوانسته‌اند از دانش درون داده‌ها به نحو مناسب استفاده کنند. با توجه به تنوع زیاد مخاطبان، مشتریان، بازارها، تنوع و پیچیدگی خدمات و محیط‌های کسب و کار و لزوم دسترسی به اطلاعات مناسب برای تصمیم‌گیری درست و به موقع، استفاده از راهکارهای مناسب برای طبقه‌بندی و یافتن اطلاعات کاربردی و اثربخش از میان انبوهی از داده‌ها برای سازمان‌ها ضروری بوده و یک تخصص و هنر محسوب می‌شود. در واقع، داده‌کاوی پاسخی به این نیاز سازمان‌ها و مؤسسات است. هرچه حجم داده‌ها بیشتر و روابط میان آنها پیچیده‌تر باشد، دسترسی به اطلاعات نهفته در داده‌ها مشکل‌تر شده، بنابراین، نقش داده‌کاوی به عنوان یکی از روش‌های کشف دانش، روشن‌تر می‌شود. امروزه، استفاده از روش‌های سنتی گردآوری و تحلیل داده به دلیل اتلاف زمان و ایجاد هزینه‌های بسیار زیاد، مناسب نبوده و از این رو استفاده از روش‌های جدید آنالیز داده مانند داده‌کاوی ضروری به نظر می‌رسد.

رقابت‌های جهانی، بازارهای پویا و چرخه‌های نوآوری و فناوری که به سرعت در حال کوتاه شدن هستند، همگی چالش‌های مهمی را برای صنعت مالی و بانکداری ایجاد کرده‌اند و نیاز به استفاده از سیستم‌های پشتیبان از تصمیم به‌منظور بهبود فرآیندهای تصمیم‌گیری در این سازمان‌ها بیش از پیش اهمیت پیدا کرده است. در این میان، داده‌هایی که در پایگاه‌های اطلاعاتی این سازمان‌ها نگهداری می‌شوند، به عنوان منابع ارزشمند اطلاعات و دانش مورد نیاز برای تصمیم‌گیری‌های سازمانی مطرح هستند.

برای نخستین بار مفهوم داده‌کاوی در کارگاه IJCAI در زمینه کشف دانش از پایگاه داده توسط Shapir مطرح شد. به دنبال آن در سال‌های ۱۹۹۱ تا ۱۹۹۴، کارگاه‌های کشف دانش از پایگاه داده مفاهیم جدیدی را در این شاخه از علم ارائه کردند؛ به‌طوری که بسیاری از علوم و مفاهیم با آن مرتبط شدند. به مرور زمان، استخراج و کشف سریع و دقیق اطلاعات با ارزش و پنهان از پایگاه داده‌ها، به عنوان داده‌کاوی مورد توجه قرار گرفت. در ادامه فرایند داده‌کاوی به عنوان فرایند آماری، تجزیه و

تحلیل در فرایند کشف دانش در پایگاه داده‌ها (KDD)[۱] پررنگ شد، به حدی که گاه داده‌کاوی[۲] (DM) به عنوان مترادف کشف دانش در پایگاه داده‌ها مورد استفاده قرار می گرفت.[۳]

هدف نهایی داده‌کاوی، ایجاد سیستم‌های پشتیبانی تصمیم‌گیری سازمانی است. داده‌کاوی، به استخراج اطلاعات مفید و دانش از حجم زیاد داده‌ها می‌پردازد. داده‌کاوی، الگوهای حاوی اطلاعات را در داده‌های موجود جست‌وجو می‌کند و این الگوها و الگوریتم‌ها، می‌توانند توصیفی باشند، یعنی داده‌ها را توصیف کنند و یا جنبه پیش‌بینی داشته باشند، یعنی متغیرها برای پیش‌بینی ارزش‌های ناشناخته متغیرها ی دیگر به کار روند.

داده‌کاوی عمدتاً با ساختن مدل‌ها مرتبط است. یک مدل اساساً به الگوریتم یا مجموعه‌ای از قوانینی گفته می‌شود که مجموعه‌ای از ورودی‌ها را با هدف یا مقصد خاصی مرتبط می‌نماید. یک مدل در شرایط درست می‌تواند به بینش درست منجر شود. بسیاری از مسائل محیط اطراف خود را می‌توان در یک قالب گنجاند، به بیان دیگر، برای تبدیل یک مسأله به یک مسأله داده‌کاوی باید آن را به یکی از فعالیت‌های داده‌کاوی تبدیل کرد، یکی از این فعالیت‌ها دسته‌بندی است. هدف دسته‌بندی داده‌ها، سازماندهی و تخصیص داده‌ها به کلاس‌های جداگانه است. در این فرآیند، بر اساس مجموعه داده‌های آموزشی، مدل اولیه‌ای ایجاد می‌شود، سپس، این مدل برای دسته‌بندی داده‌های جدید استفاده می‌شود، به این ترتیب با به‌کارگیری مدل به‌دست آمده، تعلق داده‌های جدید به دسته معین قابل پیش‌بینی است. به بیان دیگر، دسته‌بندی شامل بررسی ویژگی‌های یک شیء جدید و تخصیص آن به یکی از مجموعه‌های از پیش تعیین شده است.

۶. مدل مفهومی

مدل مفهومی، توصیف غیر نرم‌افزاری خاصی از مدل است که اهداف، ورودی، خروجی و محتوای مدل را تشریح می‌کند.[۴] به‌طور خلاصه، پس از تعیین هدف اصلی پژوهش با عنوان ایجاد مدل با قابلیت

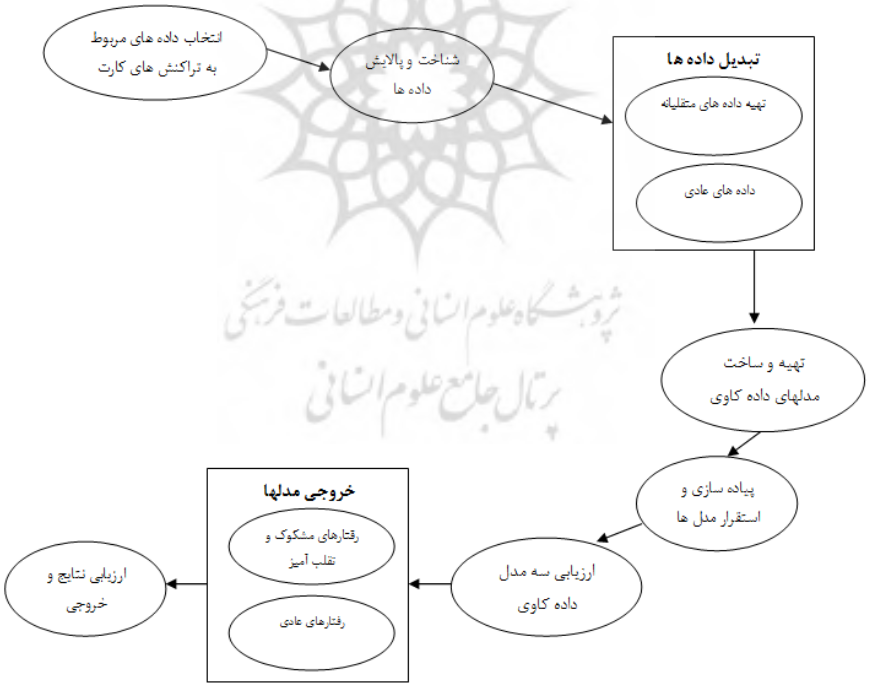
1. Knowledge Discovery From Database (KDD)
2 .Data Mining.

۳. آهوپای .(۱۳۹۰).

4. Robinson. (2004)..

اطمینان مناسب به منظور کشف تقلب در کارت‌های بانکی، نحوه گردآوری، پردازش و آماده‌سازی داده‌ها برای ایجاد مدل، تشریح می‌شود و داده‌های تقلب (تراکنش‌های متقلبانه یا مشکوک به تقلب) که به کمک دانش خبرگان در زمینه کارت‌های بانکی و ادبیات موضوع تهیه شده است، برای مدل‌سازی آماده خواهد شد. سپس، با استفاده از قابلیت‌های نرم‌افزار یک‌سری از داده‌های پژوهش که به صورت تصادفی انتخاب شده‌اند، آموزش داده می‌شوند و مدل اصلی پژوهش برای طبقه‌بندی تراکنش‌ها به عادی و مشکوک به‌دست خواهد آمد.

شکل۱. مدل مفهومی



۷. روش‌شناسی پژوهش

این پژوهش که از نوع اکتشافی است، به دنبال دستیابی به اطلاعاتی است که با کمک آنها می‌توان موضوع پژوهش را به خوبی شناسایی کرد. برای این منظور از مشتریان دارنده کارت‌های بانک به صورت تصادفی نمونه‌گیری صورت گرفته و با استفاده از روش Resample در نرم‌افزار Weka که یک واسط همگون برای بسیاری از الگوریتم‌های یادگیری متفاوت، فراهم می‌کند، استفاده شده که از طریق آن روش‌های پیش‌پردازش، پس از پردازش و ارزیابی نتایج طرح‌های یادگیری روی همه مجموعه‌های داده موجود قابل‌اعمال بوده و روش‌های تجزیه و تحلیل در مدل‌های داده‌کاوی و الگوریتم‌های دسته‌بندی نتایج به‌دست آمده است.

از قابلیت‌های نرم‌افزار «وکا» اعمال یک روش یادگیری به یک مجموعه داده و تحلیل خروجی آن برای شناخت چیزهای بیشتری راجع به آن اطلاعات است. راه دیگر، استفاده از مدل یادگیری شده برای تولید پیش‌بینی‌هایی در مورد نمونه‌های جدید است. سومین راه، اعمال یادگیرنده‌های مختلف و مقایسه کارآیی آنها به منظور انتخاب یکی از آنها برای برآورد است.

۸. داده‌های پژوهش

داده‌های اصلی پژوهش از تراکنش‌های ثبت شده کارت‌های بانکی مشتریان در پایگاه داده یکی از بانک‌های خصوصی به دست آمده و از آن برای طراحی چارچوب شناسایی تقلب در کارت‌های بانکی استفاده شده است. این تراکنش‌ها شامل برداشت، خرید و انتقال وجه کارت‌های شتاب مشتریان است. چون امکان دسترسی به داده‌ها و اطلاعات پایگاه داده اصلی بانک وجود نداشت، حدود ۱۴۰ هزار تراکنش به صورت تصادفی از پایگاه داده در نظر گرفته شد.

کارت‌هایی که در این پژوهش برچسب تقلب دارند، یا به عبارتی در معرض ریسک بیشتری هستند، عبارتند از:

- کارت‌هایی که از نظر تراکنش، پرتراکنش‌ترین کارت‌ها هستند،
- کارتی که در بازه زمانی کوتاه شهری که تراکنش در آن صورت گرفته، تغییر کند،

- کارت‌هایی که روند برداشت آنها در یک بازه زمانی کوتاه تغییر کرده است. به‌طور مثال، کارتی در بیشتر مواقع برداشت‌های با مبالغ پایین داشته است و به یک‌باره در یک بازه زمانی کوتاه برداشت‌هایی با مبالغ بالا دارد.

از آنجا که در این پژوهش داده‌های مشکوک وجود ندارد، با استفاده از ادبیات موضوع و بهره‌گیری از مصاحبه با خبرگان در زمینه کارت‌های بانکی و خدمات بانکداری الکترونیکی، داده‌های متقلبانه برای استفاده در مدل آماده شد. سازوکار تهیه تراکنش مشکوک به تقلب در این پژوهش، از طریق اعمال تغییرات معنادار روی داده‌های گردآوری شده از طریق گزارش‌های موجود در خصوص موارد مشکوک، مصاحبه با کارشناسان واحدهای مربوط و همچنین تقلب‌هایی که ممکن است بر روی کارت‌های بانکی رخ دهد، تهیه شده است.

داده‌ها را به قرار زیر تعریف می‌شود (تراکنش‌ها به دو دسته تقسیم می‌شوند):

۱. تراکنش عادی، شامل عملیاتی است که به صورت عادی و بدون اشتباه و کامل انجام شده است.

۲. تراکنش مشکوک شامل عملیاتی است که به صورت غیرعادی انجام شده است.

در جدول زیر، متغیرهای برگرفته از داده‌های مربوط به تراکنش‌های بانکی به‌طور کامل توضیح داده شده است.

جدول۳. متغیرهای پژوهش

ردیف	نام متغیرها	توضیحات
۱	LOCAL_TIME	زمان تراکنش
۲	FINANCIAL_DATE	تاریخ تراکنش
۳	CH_CITY	شهری که تراکنش رخ داده است
۴	CARD_NO	شماره کارت
۵	AMT	مبلغ تراکنش
۶	TRAN_TYPE_DESC	نوع تراکنش
۷	BANK_NAME	نام بانک

۹. گردآوری و آماده‌سازی داده‌ها

داده‌کاوی عمدتاً با ساختن مدل‌ها مرتبط بوده و یک مدل اساساً به الگوریتمی از قوانین گفته می‌شود که مجموعه‌ای از ورودی‌ها را با هدف یا مقصد خاصی مرتبط می‌کند. بسیاری از مسائل محیط اطراف خود را می‌توان در قالب یک مدل گنجاند، به بیان دیگر، برای تبدیل یک مسأله به یک مسأله داده‌کاوی باید آن را به یکی از فعالیت‌های داده‌کاوی تبدیل کرد. یکی از فعالیت‌های متداول در داده‌کاوی، دسته‌بندی است.

هدف دسته‌بندی داده‌ها، سازماندهی و تخصیص داده‌ها به کلاس‌های جداگانه است. در این فرآیند، بر اساس مجموعه داده‌های آموزشی، مدل اولیه‌ای ایجاد می‌شود. سپس، این مدل برای دسته‌بندی داده‌های جدید مورد استفاده قرار می‌گیرد، به این ترتیب با به‌کارگیری مدل به‌دست آمده تعلق داده‌های جدید به دسته معین قابل پیش‌بینی است. به بیان دیگر، دسته‌بندی شامل بررسی ویژگی‌های یک شیء جدید و تخصیص آن به یکی از مجموعه‌های از پیش تعیین شده است. از جمله مدل‌ها و رویکردهای دسته‌بندی می‌توان به درخت تصمیم، k نزدیکترین همسایگی و شبکه عصبی اشاره کرد. در این پژوهش، این مدل‌ها با هم مقایسه شده‌اند.

درخت تصمیم یکی از ابزارهای قوی و متداول برای پیش‌بینی و دسته‌بندی است. درختان تصمیم، یک ساختار درختی شبیه روندنما دارد. بالاترین گره در درخت، گره ریشه است و گره‌های برگ، دسته‌ها یا توزیع دسته‌ها را نشان می‌دهند. نمونه‌ها را با مرتب‌کردن آنها در درخت از گره ریشه به سمت گره‌های برگ دسته‌بندی می‌کنند. هر گره داخلی در درخت، صفتی از نمونه را آزمایش می‌کند و هر شاخه‌ای که از آن گره خارج می‌شود، متناظر یک مقدار ممکن برای آن صفت است. همچنین، به هر گره برگ، یک دسته بندی منتسب می‌شود. هر نمونه، با شروع از گره ریشه درخت و آزمایش صفت مشخص شده توسط این گره و حرکت در شاخه متناظر با مقدار صفت داده شده در نمونه، دسته‌بندی می‌شود. این فرآیند برای هر زیردرختی که گره جدید ریشه آن است، تکرار می‌شود.[۱]

---

۱. تیمورپور .(۱۳۹۰).

در مورد ویژگی‌های درخت تصمیم می‌توان به موارد زیر اشاره کرد:

- روش درخت تصمیم در تقسیم‌بندی داده‌ها به گروه‌های مختلف، به‌گونه‌ای است که هیچ داده‌ای حذف نمی‌شود،
- استفاده از درخت تصمیم آسان است،
- درک مدل ایجادشده توسط درخت تصمیم آسان است. به بیان دیگر، با وجود این که فهمیدن روش کار الگوریتم‌های سازنده درخت، چندان ساده نیست، ولی فهمیدن نتایج به دست آمده از آنها آسان است،
- دسته‌بندی‌هایی که توسط درخت تصمیم ایجاد می‌شوند، از روی شباهت داده‌های ذخیره شده در پارامترهای پیش‌بینی کننده، قابل انجام است.

الگوریتم k نزدیکترین همسایگی، یک الگوریتم تعلیم با سرپرستی است. در حالت کلی از این الگوریتم به دو منظور استفاده می‌شود: برای برآورد تابع چگالی توزیع داده‌های تعلیم و برای طبقه‌بندی داده‌های آزمون بر اساس الگوهای تعلیم. در این پژوهش تلاش شده است به بررسی تقلب در بانکداری الکترونیکی در شاخه کارت‌های الکترونیکی و ارائه راهکارهایی در این زمینه پرداخته شود. فرآیند مورد بررسی در این پژوهش شامل سه مرحله آماده‌سازی داده‌ها، دسته‌بندی تراکنش‌های مالی و شناسایی تقلب است.

شبکه‌های عصبی با الهام‌گیری از الگوی مغز انسان ضمن فرآیند آموزش، اطلاعات مربوطه را درون شبکه ذخیره می‌کنند. این شبکه امکان یادگیری داشته و همانند شبکه‌های زیستی می‌تواند با توجه به اطلاعات اولیه چیزی را بیاموزد و یا بر اساس آموخته‌های خود تصمیم‌گیری کند. شبکه‌های عصبی به‌طور خاصی استفاده شده، چرا که آنها ابزار مؤثری برای مدلسازی مسائل بزرگ و پیچیده که ممکن است در آنها صدها متغیر پیش‌بینی کننده که فعل و انفعالات زیادی دارند وجود داشته باشد (شبکه‌های عصبی زیستی به‌طور غیرقابل مقایسه‌ای پیچیده‌تر هستند). شبکه‌های عصبی می‌توانند در مسائل طبقه‌بندی یا حدس‌های بازگشتی (که در آنها متغیر خروجی پیوسته است) استفاده شوند. یک شبکه عصبی با یک لایه داخلی شروع می‌شود که در آن هر گره به یک متغیر پیشگو منسوب می‌شود.

این گره‌های ورودی به یک تعداد از گره‌ها در لایه پنهان متصل می‌شوند. گره‌ها در لایه پنهان می‌توانند به گره‌هایی در یک لایه پنهان دیگر یا به یک لایه خروجی متصل شوند. لایه خروجی خود شامل یک یا چند متغیر پاسخ است.

پیش از کشف دانش از هر منبع داده‌ای، باید نخست داده‌ها به منظور سازگاری با الگوریتم‌های یادگیری و بیان درست واقعیت، در جهت توانایی استخراج دانش مفید از آنها آماده شوند. معمولاً گفته می‌شود که آماده‌سازی داده ۷۰ تا ۸۰ درصد از روند داده‌کاوی است.

داده‌های خام معمولاً دچار مشکلاتی هستند و استفاده از آنها به همین صورت موجب تضعیف مدل خواهد شد. پیش‌پردازش داده‌ها شامل تبدیلات پیچیده‌ای است که برای کاهش ابعاد داده‌ها مورد استفاده قرار می‌گیرد. به‌طور خلاصه می‌توان گفت پیش‌پردازش داده‌ها شامل تمام تبدیلاتی است که بر روی داده‌های خام صورت می‌گیرد و آنها را به صورتی در می‌آورد که برای پردازش‌های بعدی نظیر استفاده در دسته‌بندی به‌کار برده می‌شود. ابزارها و روش‌های مختلفی برای پیش‌پردازش وجود دارد؛ مانند به‌هنجارکردن، که داده‌ها را به داده‌هایی جدید با بازه تغییرات و یا توزیع مناسب تبدیل می‌کند و کاهش ابعاد، که برای حذف داده‌های تکراری، اضافه و یا نامربوط برای دسته‌بندی استفاده می‌شود.

"مجموعه آموزشی"، به مجموعه‌ای گفته می‌شود که در آن داده‌هایی که به‌طور پیش‌فرض در دسته‌های مختلفی قرار دارند، همراه با ساختار دسته‌بندی خود وارد سیستم شده و سیستم بر اساس آنها به خود آموزش می‌دهد. از مجموعه آموزشی است که الگوریتم‌ها یادگیری کرده و این مجموعه هم دارای داده‌های معتبر و هم دارای داده‌های نامعتبر است. "مجموعه آزمایشی" یا آزمونه، به مجموعه‌ای گفته می‌شود که شامل داده‌هایی بوده که برای آزمون الگوریتم استفاده شده و به عبارتی میزان عملکرد الگوریتم مشخص شده است.

در این پژوهش داده‌ها به دو دسته مجموعه آموزشی که شامل ۸۰ درصد داده بوده و مجموعه آزمایشی یا آزمونه که شامل ۲۰ درصد داده است، تقسیم شده است.

۱۰. یافته‌های پژوهش

برای تحلیل رفتار تراکنشی مشتری که عوامل تأثیرگذار در شناسایی رفتار مشکوک و تقلب برانگیز مشتری است، از الگوریتم‌های طبقه‌بندی که به بیان دانش و استخراج قوانین در ارتباط با مجموعه‌ای از اطلاعات می‌پردازد، استفاده شده است. از آنجا که الگوریتم‌های ارائه شده در این پژوهش به منظور طبقه‌بندی انتخاب شده‌اند، باید با معیارهای طبقه‌بندی ارزیابی شوند. برای ارزیابی عملکرد الگوریتم‌ها باید طبقه‌بندی واقعی تراکنش‌های کارت‌های بانکی با طبقه‌بندی انجام شده توسط نرم‌افزار مقایسه شده و توانایی الگوریتم‌ها را در شناسایی تراکنش‌های متقلبانه آزمود. معیارهای ارائه‌شده در جدول ۴ برای ارزیابی عملکرد سیستم طبقه‌بندی استفاده شده است.[1]

جدول ۴. معیارهای ارزیابی الگوریتم‌ها

نام معیار	توضیح	فرمول محاسبه
نسبت مثبت درست[2] (حساسیت یا فراخوانی)	نسبت موارد مثبتی که به درستی طبقه‌بندی شده‌اند.	$TPR = \frac{TP}{TP + FN}$
نسبت منفی درست[3] (ویژگی یا فراخوانی)	نسبت موارد منفی است که به درستی طبقه‌بندی شده‌اند.	$TNR = \frac{TN}{TN + FP}$
نسبت مثبت نادرست[4]	نسبت موارد منفی که به نادرست، مثبت طبقه‌بندی شده‌اند.	$FPR = \frac{FP}{FP + TN}$
نسبت منفی نادرست	نسبت موارد مثبتی که به نادرست، منفی طبقه‌بندی شده‌اند.	$FNR = \frac{FN}{FN + TP}$
صحت[5]	نسبت تعداد مثبت‌های درست به کل نتایج مثبت (هم مثبت درست و هم مثبت نادرست)	$P = \frac{TP}{TP + FP}$
صحت	نسبت تعداد مثبت‌های نادرست به کل نتایج منفی(هم مثبت نادرست و هم منفی نادرست)	$P = \frac{TN}{TN + FN}$
دقت[6]	نسبت نتایج درست (هم مثبت درست و هم منفی درست) به کل جامعه	$AC = \frac{TP + TN}{TP + TN + FP + FN}$
شاخص F	F	$F = 2 * \frac{P * TPR}{P + TPR}$
میانگین هندسی[7]	G-mean	$\sqrt{TPR * P}$

مأخذ: یافته‌های پژوهش.

1. Brodersen, et al. (2010)..
2. True - positive Ratio
3. True Negative Ratio
4. False Positive Ratio
5. Precision
6. Accuracy
7. Geometric Mean

متغیرهایی که برای ارزیابی الگوریتم‌ها در نظر گرفته شده است به صورت زیر است:

۱. مثبت واقعی (TP)[1] اگر تراکنش عادی به‌درستی در گروه تراکنش عادی قرار گیرد.
۲. منفی واقعی (FP)[2] اگر تراکنش مشکوک در دسته تراکنش عادی قرار گیرد.
۳. مثبت کاذب (TN)[3] تراکنش مشکوک به‌درستی در گروه تراکنش مشکوک قرار گیرد.
۴. منفی کاذب (FN)[4] اگر تراکنش عادی در دسته تراکنش مشکوک قرار گیرد.

الگوریتم‌ها را می‌توان از چند منظر مختلف بررسی کرد؛ مانند سرعت پیش‌بینی، دقت در تشخیص تراکنش‌ها و درستی. زمانی که تعداد منفی‌ها از تعداد مثبت‌ها بیشتر باشد، ممکن است که معیار دقت طبقه‌بندی معیار مناسبی برای ارزیابی عملکرد نباشد. بنابراین، معیارهای دیگری مانند میانگین هندسی و F را برای ارزیابی عملکرد طبقه‌بندی می‌توان در نظر گرفت. در جدول ۵، نتایج پیش‌بینی الگوریتم‌ها ارائه شده است.

جدول ۵. میزان پیش‌بینی مدل‌ها

الگو	درخت تصمیم	نزدیکترین همسایگی	شبکه عصبی
پیش‌بینی	٪۹۶	٪۹۱	٪۹۵

مأخذ: یافته‌های پژوهش.

در این پژوهش تلاش می‌شود تا سه الگوریتم مطرح شده را که شامل نزدیک‌ترین همسایگی، درخت تصمیم و شبکه عصبی هستند، با هم بررسی شود. هدف از مقایسه این الگوریتم‌ها، شناسایی تقلب در کارت‌های بانکی و شناسایی بهتر تراکنش‌های متقلبانه از میان تراکنش‌ها و به‌دست آوردن الگوریتم بهینه در شناسایی رفتار مشکوک و تقلب‌آمیز مشتریان بانک است.

از ۱۴۰ هزار تراکنش که به عنوان داده‌های تحقیق در نظر گرفته شده است ۸۰ درصد به عنوان داده‌های آموزش انتخاب شده‌اند. برای شناسایی تقلب مشتریان و به‌دست آوردن الگوریتم

1. True-positive
2. False-positive
3. True-negative
4. False-negative

بهینه، داده‌های آزمونه در نرم‌افزار پیاده‌سازی و تحلیل شد که نتایج به‌دست آمده با توجه بـه مـاتریس درهم ریختگی الگوریتم‌ها به صورت زیر است:

جدول ۶. ماتریس درهم ریختگی الگوریتم‌ها

	نوع الگوریتم	مشاهدات	سالم	تقلب
نمونه اعتبارسنجی	درخت تصمیم	سالم	۱۲,۹۰۰	۳۰۰
		تقلب	۷۰۰	۱۴,۱۰۰
	نزدیکترین همسایگی	سالم	۱۲,۳۰۰	۹۰۰
		تقلب	۱,۴۰۰	۱۳,۴۰۰
	شبکه عصبی	سالم	۱۲,۹۰۰	۳۰۰
		تقلب	۱,۰۰۰	۱۳,۸۰۰

مأخذ: یافته‌های پژوهش.

با توجه به خروجی به‌دست آمـده از نـرم‌افـزار معیارهـای میـزان دقـت، صـحت، TPR، FNR، میانگین هندسی و شاخص F تحلیل شده است. مقادیر هر یک از این معیارها عددی بین صـفر و یـک است که هر چه به یک نزدیکتر باشد، نشان می‌دهد که معیار خروجی بهتری داشته است.

با توجه به معیارهای عملکرد مدل‌های داده‌کاوی که در جدول ۴ ارائه شـده اسـت، محاسـبات مربوط به سه مدل انجام شد و با توجـه بـه نتـایج بـه‌دسـت آمـده، الگـوریتم درخـت تصـمیم از بـین الگوریتم‌های مورد ارزیابی توانسته است در بیشتر موارد تراکنش‌ها را به‌درسـتی بـه دو دسـته سـالم و متقلبانه طبقه‌بندی کند. در ادامه، محاسبات مربوط به درخت تصمیم به‌طور کامل آورده شده است.

$$TP=12900, \quad FP=700, \quad FN=300, \quad TN=14100$$

$$TPR = \frac{TP}{TP + FN} = \frac{12900}{12900 + 300} = 0.977$$

$$FPR = \frac{FP}{FP + TN} = \frac{700}{700 + 12900} = 0.051$$

$$FNR = \frac{FN}{FN + TP} = \frac{300}{300 + 12900} = 0.021$$

$$TNR = \frac{TN}{TN + FP} = \frac{14100}{14100 + 700} = 0.952$$
$$AC = \frac{TP + TN}{TP + TN + FP + FN} = \frac{12900 + 14100}{12900 + 14100 + 700 + 300} = 0.964$$
$$P = \frac{TP}{TP + FP} = \frac{12900}{12900 + 700} = 0.948$$
$$F = 2 * \frac{P * TPR}{P + TPR} = 2 * \frac{0.948 * 0.977}{0.948 + 0.977} = 0.962$$
$$G - mean = \sqrt{TPR * P} = \sqrt{0.977 * 0.948} = 0.962$$

جدول ۷. معیارهای عملکرد دو مدل شبکه عصبی و نزدیکترین همسایگی

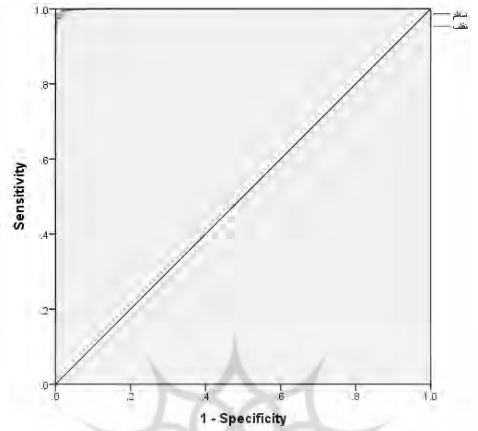
	TPR	FPR	FNR	TNR	AC	P	F	G-mean
نزدیکترین همسایگی	٪۹۳	٪۹	٪۶	٪۹۰	٪۹۱	٪۹۳	٪۹۳	٪۹۳
شبکه عصبی	٪۹۶	٪۶	٪۳	٪۹۴	٪۹۴	٪۹۶	٪۹۶	٪۹۶

مأخذ: یافته‌های پژوهش.

نمودار شکل ۲، منحنی ROC[1] مربوط به مدل درخت تصمیم را نمایش می‌دهد. منحنی ROC، نموداری گرافیکی است که عملکرد یک سیستم طبقه‌بندی را با آستانه افتراق متفاوت نمایش می‌دهد. این نمودار با ترسیم نسبت مثبت‌های درست در مقابل نسبت مثبت‌های نادرست با مقادیر برش مختلف، به دست می‌آید. در این نمودار نقطه (۱و۰) مختص به بهترین طبقه‌بندی کننده بوده که می‌تواند تمام موارد مثبت و منفی را به‌درستی طبقه‌بندی کند. سطح زیر نمودار ROC معیار اصلی ارزیابی عملکرد طبقه‌بندی است. سطح زیر نمودار برای این پژوهش برای سه مدل به صورت زیر است:

1. Receiver Operating Characteristic

شکل ۲. منحنی ROC مدل درخت تصمیم



جدول ۸. نمودار ROC

	درخت تصمیم	نزدیکترین همسایگی	شبکه عصبی
نمودار ROC	٪۹۸	٪۹۱	٪۹۶

مأخذ: یافته‌های پژوهش.

۱۱. نتیجه‌گیری و ارائه پیشنهاد سیاستی

همان طور که اشاره شد، امروزه تقلب رشد روزافزونی در سیستم بانکی داشته است و همزمان با آن روش‌های متفاوتی برای مقابله با تقلب معرفی می‌شود. در این پژوهش، مدلی برای شناسایی تقلب مشتریان کارت بر اساس الگوریتم‌های داده‌کاوی نظیر الگوریتم درخت تصمیم، شبکه عصبی و نزدیکترین همسایگی ارائه شده است.

در دنیای امروز یکی از تصمیمات ضروری برای بانک‌ها، نحوه عملیاتی‌کردن مدل‌های شناسایی تقلب است. پیشنهاد سیاستی این است که بانک‌ها باید تصمیم بگیرند که مدل شناسایی تقلب را برخط استفاده کنند یا خیر. در صورت استفاده برخط از مدل، از انجام تراکنش‌های مشکوک به تقلب جلوگیری شده و حتی ممکن است کارت مشتری باطل شود. بنابراین، این ریسک برای بانک وجود دارد که نارضایتی مشتریان را در مواردی که مدل به نادرست تراکنش آنها را مشکوک شناسایی کرده

است، شاهد باشد. از سوی دیگر، چنانچه شناسایی تراکنش‌های مشکوک به صورت برخط نباشد، این ریسک برای بانک وجود دارد که تراکنش‌های متقلبانه صورت گرفته، قابل برگشت نباشد. در طبقه‌بندی تراکنش‌های مشکوک و سالم، سیاست بانک نقش تعیین‌کننده‌ای دارد که مشخص می‌شود از چه روشی با چه میزان دقت استفاده شود.

شاخص‌های معرفی شده در این پژوهش شامل شهری که تراکنش در آن رخ داده، زمان تراکنش و تاریخ وقوع است. نتایج به‌دست آمده از الگوریتم درخت تصمیم تأیید می‌کند که مدل با دقت نسبتاً بالایی در بازشناسایی احتمال تقلب موفق است. به نظر می‌رسد یکی از کاستی‌های نظام بانکداری الکترونیک و نظام‌های پرداخت شبکه بانکی، تجهیز نبودن زیرساخت‌های نرم‌افزاری بانک‌ها به این چنین نرم‌افزاهایی است؛ بنابراین، یافته‌های این پژوهش می‌تواند در پیاده‌سازی این رهیافت در شبکه بانکی مؤثر باشد.

این قابلیت اطمینان در شناسایی تراکنش‌های مشکوک مشتریان کارت، به مدیران فناوری اطلاعات بانک‌ها این امکان را می‌دهد با تولید نرم‌افزارهایی از مدل یادشده و اتصال آن به شبکه بانکداری الکترونیک اقدامات مشکوک را شناسایی کنند؛ همچنین، تمهیدات لازم را برای پیشگیری از تقلب‌هایی که ممکن است در آینده رخ دهد، برنامه‌ریزی کنند.

در پژوهش‌های آتی با توجه به یادگیری ماشین و داده‌کاوی می‌توان هشداری در سیستم ایجاد کرد تا در صورت روبه‌روشدن با تراکنش‌های مشکوک پرسش امنیتی از صاحب کارت پرسیده شود و در صورت پاسخ درست از سوی صاحب کارت، مراحل بعدی تراکنش انجام شده و امنیت بالاتری برای کاربران ایجاد شود. از سوی دیگر، می‌توان متغیری وابسته با طبقات "تراکنش سالم"، "تراکنش با ریسک کم"، "تراکنش با ریسک بالا" و "تراکنش متقلبانه" تعریف کرد که در صورت رخ دادن هر یک از موارد با توجه به سیاست بانک و اهمیت آن به‌صورت مستقیم کارت را مسدود کرد یا به صورت دستی تحت بررسی کارشناسان قرار داد.

منابع

- البرزی، محمود؛ پورزندی، محمد؛ ابراهیم، محمد و خان بابایی، محمد. (۱۳۸۹). به‌کارگیری الگوریتم ژنتیک در بهینه‌سازی درختان تصمیم‌گیری برای اعتبارسنجی مشتریان بانک. فصلنامه مدیریت فناوری اطلاعات دانشگاه تهران.
- آذر، عادل؛ احمدی، پرویز و سبط، محمدوحید. (۱۳۸۹). طراحی مدل انتخاب نیروی انسانی با رویکرد داده کاوی. فصلنامه مدیریت فناوری اطلاعات دانشگاه تهران.
- آهوپای، رها. (۱۳۹۰). روش‌های کشف تقلب در استفاده از کارت‌های اعتباری. ایران دیتا.
- تیمورپور، بابک و علیزاده، سمیه. (۱۳۹۰). داده‌کاوی و کشف دانش. انتشارات دانشگاه علم و صنعت ایران.
- حاتمی‌راد، علی و شهریاری، حمیدرضا. (۱۳۸۹). روش‌ها و راهکارهای شناسایی تقلب در بانکداری الکترونیک. تازه های اقتصاد.
- شهرابی، جمال. (۱۳۸۶). داده‌کاوی ۱ و ۲. انتشارات جهاد دانشگاهی واحد صنعتی امیرکبیر.
- شهرابی، جمال. (۱۳۹۰). داده‌کاوی در صنعت بانکداری. انتشارات جهاد دانشگاهی واحد صنعتی امیرکبیر.
- فیروزی، مهدی؛ شکوری، مرتضی و کاظمی، لیلا. (۱۳۹۰). شناسایی تقلب در بیمه اتومبیل با استفاده از روش داده‌کاوی. نشریه پژوهشکده بیمه.
- قاسمی، احمدرضا و اصغری زاده، عزت اله. (۱۳۹۳). به‌کارگیری روش شبکه‌های عصبی مصنوعی خودسامانده اصلاح شده در تعیین سطح سرآمدی شرکت‌های پتروشیمی کشور. فصلنامه مدیریت فناوری اطلاعات دانشگاه تهران.
- کوثری لنگری، روح اله، مقدم چرکری ،نصراله ، وحدت، داود .(۱۳۸۹). به‌کارگیری الگوریتم‌های درخت تصمیم‌گیری جهت کشف رفتارهای مشکوک در بانکداری اینترنتی. پژوهشنامه پردازش و مدیریت اطلاعات.

- محقر، علی؛ لوکس، کارو؛ حسینی، فرید و منشی، آصف علی. (۱۳۸۷). کاربرد هوش تجاری به عنوان یک تکنولوژی اطلاعات استراتژیک در بانکداری: بازرسی و کشف تقلب. فصلنامه مدیریت فناوری اطلاعات دانشگاه تهران.
- معین‌زاد، حسین. (۱۳۹۰). بانکداری الکترونیک و کشف تخلفات. مجله بانک ملی ایران.
- نصیری، نوید و مینایی، بهروز. (۱۳۸۹). روش‌های داده‌کاوی در شناسایی تقلب در کارت‌های اعتباری. کنفرانس بین‌المللی شهروند الکترونیک و تلفن همراه تهران.
- نوبرزاده، محمد. (۱۳۹۱). کشف تقلب در کارت‌های اعتباری بر اساس مدل الگوریتم ژنتیک و جستجوی پراکنده. پایان‌نامه کارشناسی ارشد مؤسسه علوم بانکی.
- Al-Khatib, A..(2011). Detect CNP Fraudulent Transactions. World of Computer Science and Information Technology Journal.
- Bhatl Tej Paul, Prabhu Vikram and Dua Amit, (2003). Understanding Credit Card Frauds, Card Business Review. United Nations, Economic and Social Council, Economic Comission for Africa.
- Bhattacharyya, S., Jha, S., Tharakunnel, K., and Westland, J., C. (2011). Data Mining for Credit Card Fraud: A Comparative Study. Decision Support Systems. 50(3): pp. 602-613.
- Bolton, R. and Hand, D. (2002). Statistical Fraud Detection: A Review (with Discussion). Statistical Science, 17(3): pp. 235-255.
- Brodersen, K. H., Ong, C. S., Stephan, K. E., and Buhmann, J. M. (2010). The balanced accuracy and its posterior distribution. Proceedings of 20th International Conference on Pattern Recognition, pp. 3121-3124.
- Chan P., Fan, W., Prodromidis, A., and Stolfo, S. (1999). Distributed Datamining in Credit Card Fraud Detection. IEEE Intelligent Systems, 14: pp. 67-74.
- Delamaire, L., Abdou H., and Pointon J. (2009). Credit Card Fraud and Detection Techniques: A Review. Banks and Bank Systems, 4(2):pp. 57-68.

- Duman, E., and Hamdi, M. (2011). Detecting Credit Card Fraud by Genetic Algorithm and Scatter Search. Journal of Expert Systems with Applications,38, pp. 13057-13063.
- Kundu, A., Panigrahi, S., Sural, S., and Majumdar, A. K. (2009). BLASTSSAHA Hybridization for Credit Card Fraud Detection, IEEE Transactions on Dependable and Secure Computing, 6(4), pp. 309-315.
- Leonard, K. J. (1995). The Development of A Rule Based Expert System Model for Fraud Alert in Consumer Credit. European Journal of Operational Research, 80(2): pp. 350-356.
- Ogwueleka, F. N. (2011). Datamining Application in Credit Card Fraud Detection System. Journal of Engineering Science and Technology, 6(3): pp. 311-322.
- Paasch, C. A. W. (2008). Credit Card Fraud Detection Using Artificial Neural Network Tuned by Genetic Algorithms (Doctoral dissertation). Retrieved from the HKUST Institutional Repository (Thesis ISMT (2008) Paasch).
- Patidar, R. and Sharma L. (2011). Credit Card Fraud Detection Using Neural network. International Journal of Soft Computing and Engineering, 1 (NCAI2011): pp. 2231-2307.
- Phua, C., Lee, V., Smith, K., and Gayler, R. (2010). A Comprehensive Survey of Data Mining-based Fraud Detection Research. Computing Research Repository, abs/1009. 6119.
- Robinson S. (2004). Simulation: The Practice of Model Development and Use. England: John Wiley and Sons Publisher.
- Shen, A., Tong, R. and Deng, Y. (2007). Application of Classification Model on Credit Card Fraud Detection. Proceedings of International Conference on Services Systems and Services Management (ICSSSM). 9-11 June 2007, Chengdu.
- Yu H.C,His K.H and Kuo.P.J .(2002). Electronic Payment System: Analysis and Comparasion of Type ,Technology in Society.