



نخستین کنفرانس ملی توسعه مدیریت پولی و بانکی

تحول در مدیریت نظام بانکی، گامی به سوی حمایه اقتصادی



۸ و ۹ بهمن ماه ۱۳۹۲ تهران - مرکز همایش‌های بین‌المللی صدا و سیما

تقلب در بانکداری الکترونیکی در ایران و ارائه راهکارها

شبنم بختیاری*، دکتر محمد تقی تقوی فرد، دکتر مریم اخوان خرازیان

کارشناسی ارشد، مدیریت فناوری اطلاعات، واحد الکترونیکی، دانشگاه آزاد اسلامی، تهران، ایران

استادیار، دانشکده مدیریت و حسابداری، دانشگاه علامه طباطبائی، تهران، ایران

استادیار، دانشکده مدیریت و حسابداری، دانشگاه شهید بهشتی، تهران، ایران

چکیده

امروزه استفاده از خدمات بانکداری الکترونیکی در اکثر جوامع افزایش چشمگیری داشته است و مردم از طریق کانالهای مختلف آن، امور مالی و بانکی خود را انجام می‌دهند. یکی از این روشها، کارتهای الکترونیکی می‌باشد که با افزایش میزان استفاده از آن، متأسفانه موارد تقلب و سوء استفاده نیز بیشتر شده است و این امر علاوه بر ایجاد ضرر به دارنده کارت، قابلیت اطمینان و حسن شهرت بانک را تحت تاثیر قرار می‌دهد. بنابراین لزوم استفاده از روشهایی برای شناسایی تقلب امری بدیهی است. در این تحقیق، روشی برای شناسایی تقلب مشتریان کارتهای بانکی مورد بررسی قرار گرفته می‌شود. در این روش تراکنش‌های بانک به دو دسته عادی و مشکوک تقسیم می‌شوند. برای اجرای روش دسته بندی در داده کاوی دو تکنیک درخت تصمیم و نزدیکترین همسایگی به کار گرفته شد و نتایج بدست آمده مبین این موضوع است که درخت تصمیم با بالاترین دقت به نحو مناسبتری تقلب را شناسایی می‌کند.

کلمات کلیدی: تقلب، کارت بانکی، درخت تصمیم، نزدیکترین همسایگی

dataacademy.ir

* شبنم بختیاری، نشانی: تهران خیابان رسالت خیابان سلمان طرقي پلاک ۱۴ واحد ۸ - ۷۷۸۰۳۱۸۱

Shabnam_1631@yahoo.com



۱- مقدمه

امروزه، دورنمای رقابتی در صنعت بانکداری به طور قابل توجهی تغییر یافته است. این امر به علت نیروهای 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 بانک ها، باعث ایجاد مباحث جدیدی در حوزه بانکی همچون تقلب شده است. توسعه فناوری های جدید راه های زیادی را برای متقلبان و مجرمان باز کرده است که بتوانند مرتکب تقلب شوند. ایجاد یک سیستم اطلاعاتی جدید، تکنیک های شناسایی تقلب، علاوه بر آنکه تقلب ها و کلاهبرداری های صورت گرفته در یک سازمان را شناسایی کرده و مورد تجزیه و تحلیل قرار می دهد، به نوعی با شناخت رفتار کاربران یا مشتریان سعی در پیش بینی رفتار آتی آنها داشته و ریسک انجام تقلب ها را کاهش می دهد. (حاتمی راد، شهریاری، ۱۳۹۰)

به دلیل هزینه های بسیار مستقیم و غیر مستقیم تقلب، بانکها و موسسات مالی و پولی به شدت به دنبال تسریع و سرعت عمل در شناخت فعالیتهای کلاهبرداران و متقلبان می باشند. این امر به دلیل اثر مستقیم آن روی خدمت رسانی به مشتریان بانکها و موسسات، کاهش هزینه های عملیاتی و باقی ماندن به عنوان یک ارائه دهنده خدمات مالی معتبر و قابل اطمینان است. لذا بکارگیری تکنیک های شناسایی تقلب به منظور جلوگیری از اقدامات متقلبان در سیستم بانکداری، اجتناب ناپذیر است.

امروزه پیچیدگی سازمان ها و تراکنش ها باعث افزایش فرصت برای استفاده های سودجویانه و تقلب شده است. این تقلب ها 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 ببیند و سطح اعتماد مشتریان نسبت به خدمات جدید کاهش یابد. در بعد درون سازمانی نیز ممکن است فرآیند مدیریتی سازمان با اختلال مواجه شود و اخلاق و وجدان کارکنان تحت اثر منفی آن قرار گیرد. خدمات نوین ارائه شده توسط بانک ها نیز حوزه های جدیدی از تقلب را گشوده است و اثرات منفی آن ضرورت اقدامات شناسایی تقلب را توجیه پذیر ساخته است. فنون شناسایی تقلب شامل شگردهای جست و جوی پیچیده ای است که از طریق بررسی تراکنش های حساب مشتریان و رفتار مصرفی مشتری الگوهای تقلب را کشف و به موقع اعلام می کند، از طرفی آگاهی از انواع مختلف تقلبات بانکی نیز جهت اقدامات پیشگیرانه می تواند مفید باشد و بانکها و موسسات مالی در صورت آشنایی با انواع مختلف تقلبات بانکی می توانند فرآیندهای نظارتی مناسبی را طراحی کنند.

در صنعت بانکداری نیز به تبع توسعه و رشد بانکداری نوین، پیشرفت های فناوری اطلاعات و در دسترس بودن امکانات کامپیوتری پیشرفته جهت ذخیره داده ها، حجم عظیمی از داده ها در دسترس تصمیم گیرندگان قرار دارد که با توجه به وضعیت رقابتی موجود، تصمیم گیری سریع، امکان استفاده از فرصت ها و تهدید ها به فرصت های طلایی، صنعت بانکداری را به سمت استفاده از تکنیک های داه کاوی ترغیب می کند. داده کاوی به عنوان تکنیکی خاص در جهت دستیابی به اطلاعات مفید و مناسب از اهمیت ویژه ای در سالیان اخیر برخوردار شده است که به صورت گسترده ای در صنایع مختلف مانند بانکداری، هوا و فضا، بهداشت و شناسایی الگوهای مفید و ارتباطات بین داده های ثبت شده و غیره مورد استفاده قرار می گیرد. هدف داده کاوی کشف و شناسایی الگوی جدید در داده هاست. در این رابطه نوعی احساس خلاء در شناسایی تقلبات در صنعت بانکداری با رویکرد داده کاوی به چشم می خورد. بنابر اعلام دانشگاه MIT دانش نوین داده کاوی یکی از ده دانش در حال توسعه ای است که دهه آینده را با انقلاب تکنولوژی مواجه می سازد. این تکنولوژی امروزه دارای کاربرد بسیار وسیعی در حوزه های مختلف است به گونه ای که امروزه حد و مرزی برای کاربرد این دانش در نظر نگرفته اند و زمینه های کاری این دانش را از ذرات کف اقیانوسها تا اعماق فضا می دانند.

نتایج تحقیقات ارنست و یانگ عوامل افزایش تقلب در سازمان ها را افزایش پیچیدگی سازمان ها، تغییر در فرآیندهای کسب و کار، کنترل های داخلی غیر موثر معرفی می کند که سبب افزایش فرصت برای استفاده های سود جویانه شده است. در این راستا بانکها و



نخستین کنفرانس ملی توسعه مدیریت پولی و بانکی

تحول در مدیریت نظام بانکی، گامی به سوی حساسه اقتصادی



۸ و ۹ بهمن ماه ۱۳۹۲ تهران - مرکز همایش‌های بین‌المللی صدا و سیما

موسسات مالی، با تجهیز به سیستم های کشف تقلب می توانند به پیشگیری بپردازند. این تحقیق سعی دارد با تکنولوژی داده کاوی روشی نوین در کشف تراکنش های مشکوک را به نمایش در آورد.

در رابطه با این موضوع کوثری لنگری و همکاران تحقیقی را تحت عنوان بکارگیری الگوریتم های درخت تصمیم گیری جهت کشف رفتارهای مشکوک در بانکداری الکترونیک انجام داده اند. در این تحقیق از دانش خیرگان و دسته بندی الگوی رفتاری کاربران توسط الگوریتم درخت تصمیم استفاده شده است. و چهار الگوریتم chaid، chaid_ex، c4.5، c5 مورد مقایسه قرار گرفته شده است و الگوریتم C5 با دقت ۹۱ درصد الگوریتم بهینه محسوب شده است. همچنین فیروزی و همکاران در تحقیقی تحت عنوان شناسایی تقلب در بیمه اتومبیل با استفاده از داده کاوی از سه الگوریتم رگرسیون لجستیک، بیز ساده، 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.

۲- ادبیات تحقیق

بانک ها جزو سازمان هایی می باشند که مستقیماً با مشتریان در تعامل هستند. بنابراین تحلیل رفتاری مشتریان برای افزایش وفاداری آنها از اهمیت بالایی برخوردار است. در سال های اخیر با افزایش دسترسی به داده های مشتریان و بهبود قابلیت های تحلیل داده ها بوسیله روش های هوشمند، فعالیت های مختلفی به منظور تحلیل رفتار مشتریان انجام شده است. یکی از این فعالیت ها، استفاده از سیستم های هوشمند در کشف تقلبات بانکی است. تقلبات بانکی در حال حاضر طیف گسترده ای یافته و باعث ضررهای مادی و غیر مادی بسیاری به بانک ها و مشتریان بانک شده اند. لذا آشنایی با این گونه تقلبات علاوه بر اینکه در پیشگیری از وقوع آنها مورد استفاده قرار می گیرد، در

حوزه تحلیل و طراحی سیستم های اطلاعاتی مورد نیاز نیز می تواند بسیار مفید باشد. (معین - داده، ۱۳۹۰)

dataacademy.ir

انواع تقلبات بانکی عبارتند از:

۱. تقلب کارتهای پرداخت
۲. مسئولین فاسد
۳. وام های تقلبی
۴. تقلبات سیمی
۵. اسناد دستکاری شده و تقلبی
۶. سپرده های مشکوک
۷. تقلب برات دیداری
۸. چک های دستکاری شده و تقلبی
۹. چکهای مفقود شده
۱۰. تقلب در صورتحساب های بانکی
۱۱. چک بازی
۱۲. چکهای بی محل
۱۳. کارتهای پرداخت مفقود شده
۱۴. دزدی اطلاعات کارتهای بانکی
۱۵. جعل هویت
۱۶. درخواست وام های تقلبی



نخستین کنفرانس ملی توسعه مدیریت پولی و بانکی

تحول در مدیریت نظام بانکی، گامی به سوی حمله اقتصادی



۸ و ۹ بهمن ماه ۱۳۹۲ تهران - مرکز همایش‌های بین‌المللی صدا و سیما

۱۷. بازرسان بانکی جعلی

۱۸. تقلبات اینترنتی و فیشینگ

۱۹. پولشویی

۱-۲- انواع کارتهای بانکی

شاید بتوان از آمریکا به عنوان اولین کشور ارائه دهنده کارت های اعتباری نام برد. پدید آمدن ایده کارت اعتباری به سالها پیش بر می گردد که اولین بار در سال ۱۸۶۰ توسط ادوارد پلای دانشمند بزرگ آمریکایی ارائه گردید. (یو و همکاران، ۲۰۰۲)

در طول این سالها کارتهای اعتباری، کارتهای بدهی، کارتهای پیش پرداخت و کارتهای هوشمند عرضه گردیده است. تمرکز تحقیق حاضر بر روی کارتهای بدهی می باشد. 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 شیوه ای برای پرداخت در آینده هستند کارت بدهی ابزار پرداخت در لحظه خرید و در بسیاری از سوپر مارکت ها، فروشگاهها و رستوران های مورد پذیرش است. از این رو کارت بدهی شیوه مناسبی برای کنترل و مدیریت هزینه است زیرا فقط می توانید به اندازه ای که پول در حساب دارید خرج کنید و بر خلاف کارت های اعتباری نمی توانید پرداخت ها را به آخر ماه یا ماههای بعد موکول کنید. البته در بسیاری از کشورها کارت بدهی یک کارت چند منظوره است و به طور مثال، به عنوان کارت خودپرداز برای دریافت پول از دستگاههای خودپرداز نیز مورد استفاده قرار می گیرد. در هر حال انواع کارتهای بدهی از ابزار های رایج پرداخت اینترنتی هستند که استفاده از آنها در سراسر دنیا از جمله ایران رو به گسترش است (باتالا و همکاران، ۲۰۰۳)

dataacademy.ir

۲-۲- راههای تقلب در کارتهای بانکی

در ساده ترین این روشها کارمند یک سازمان می تواند یک کپی از رسید مشتری را نزد خود نگه دارد و از آن برای مقاصد بعدی خود استفاده کند. کپی کردن از اطلاعات کارت و سپس سرقت از حساب مشتریان سابقه زیادی در حوزه کلاهبرداری از طریق خودپردازها دارد. به طوری که، این مساله در سال ۲۰۱۰ روی بیش از ۲۵ درصد عابر بانک هایی که کمتر از شش ماه از نصب آنها گذشته بود، اتفاق افتاده است.

اصولا در حوزه خودپردازها تقلب امری رایج محسوب می شود. در کارتهای مغناطیسی امکان سرقت اطلاعات از طریق دوربین، نصب نرم افزاری درون خودپرداز یا نصب پایانه فروشگاهی کاذب روی کارتخوان اصلی وجود دارد. نوع دیگری از کلاهبرداری زمانی روی می دهد که افراد به سیستم نرم افزاری بانک نفوذ می کنند و با استفاده از اطلاعات سرقت شده مشتریان، از راههای مختلفی از جمله صدور کارتهای تقلبی به برداشت از حساب مشتریان می پردازد. (شهرابی، ۱۳۹۰)

برخی از تحقیق هایی که به بررسی عملکرد مدل های شناسایی تقلب پرداخته اند، تفاوت بین انواع مختلف سرقت ها را در نظر نگرفته اند. وارد کردن چنین مسئله ای به مدل ها اگر چه ممکن است منجر به پیچیده تر شدن مدل گردد اما جواب های قابل قبول تری به سازمان ارائه می دهد. به طور معمول پس از آن که فردی از گم شدن کارت خود اطلاع پیدا می کند در اولین فرصت موضوع را به بانک صادرکننده ی کارت اطلاع می دهد تا از سوء استفاده های احتمالی جلوگیری کند. آسان ترین راه سرقت در حالی اتفاق می افتد که فرد یابنده قبل از اعلام گم شدن کارت به بانک و باطل کردن آن، با امتحان کردن تعدادی رمز از حساب فرد پول برداشت کند. (شهرابی، ۱۳۹۰)

در روشی دیگر، هنگامی که اطلاعات یک کارت سرقت می شود فرد سارق می تواند ماه ها از کارت فردی دیگر استفاده کند بدون اینکه صاحب اصلی اطلاعی از این موضوع داشته باشد. این نوع از کلاهبرداری ها هنگامی اتفاق می افتد که شماره ی کارت، تاریخ انقضا، نام



نخستین کنفرانس ملی توسعه مدیریت پولی و بانکی

تحول در مدیریت نظام بانکی، گامی به سوی حساسه اقتصادی



۸ و ۹ بهمن ماه ۱۳۹۲ تهران - مرکز همایش‌های بین‌المللی صدا و سیما

صاحب کارت و یا کد اعتبارسنجی کارتی، ندانسته در اختیار فردی دیگر قرار می‌گیرد. این مسئله بخصوص هنگام خرید از فروشگاهها می‌تواند رخ دهد. (شهرابی، ۱۳۹۰)

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 غیر حضوری با یکدیگر معامله می‌کنند، فروشنده راهی برای تشخیص این که آیا فردی که اطلاعات را به او داده است صاحب اصلی کارت است یا نه، ندارد.

در نوع دیگری از سرقت‌ها، اطلاعات کارتی که به تازگی صادر و به فرد متقاضی فرستاده شده است، قبل از رسیدن به فرد اصلی سرقت می‌شود. معمولاً بانک‌ها برای فعال کردن اینترنتی کارتها از متقاضیان خود اطلاعاتی مانند تاریخ تولد را سوال می‌کنند 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، به خصوص در کشورهایی که ارائه‌ی خدمات بانکی و بیمه‌ای بیشتر به اعتبار افراد بستگی دارد، سال‌ها زمان لازم است تا فرد قربانی اعتبار قبلی خود را نزد بانک به دست آورد.

اساساً در تمام دنیا کلاهبرداری از طریق خودپردازها و دیگر ابزارهای الکترونیک وجود دارد. معمولاً فردی که اطلاعات او به سرقت رفته است تا زمان اطلاع یافتن از خالی شدن حساب خود از این مساله بی‌خبر می‌ماند. بانک‌ها نیز احتمالاً با برداشت‌های غیرعادی و مشکوکی روبرو نمی‌شوند. به این معنی که به ظاهر همه‌ی تراکنش‌های مالی فرد، معمولی و در سقف مجاز خود به نظر می‌آید. امروزه با پیشرفت تکنولوژی که منجر به پیچیده‌تر شدن روش‌های انجام سرقت 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 طرفی دیگر، افزایش تعداد مشتریان، ریسک انجام سرقت و نفوذ به پایگاه‌های داده‌ی سازمان‌ها را نیز افزایش داده است. تا زمانی که داده‌های یک بانک به صورت 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 مسئله رفتارسنجی مشتریان به وسیله تکنیک‌های داده‌کاوی به یکی از مسائل مهم مدیران و کارشناسان بانکی تبدیل شده است. (شهرابی، ۱۳۹۰)

۲-۳- داده کاوی در بانکداری

در طول دهه‌های گذشته حجم زیادی از داده‌ها در پایگاه داده‌ها انباشته و ذخیره شده‌اند. در واقع سازمان‌ها در اطلاعات غرق شده‌اند در حالیکه تشنه دانش هستند. این امر نشانگر آن است که سازمانها توانسته‌اند از دانش درون داده‌ها به نحو مناسب استفاده نمایند. با توجه به تنوع زیاد مخاطبین، 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 اثر بخش از میان انبوهی از داده‌ها برای سازمان‌ها امری ضروری و حیاتی بوده و یک تخصص و هنر محسوب می‌شود. در واقع داده‌کاوی پاسخی به این نیاز سازمانها و موسسات است. هرچه حجم داده‌ها بیشتر و روابط میان آنها پیچیده‌تر باشد، دسترسی به اطلاعات نهفته در داده‌ها مشکل‌تر شده لذا نقش داده‌کاوی به عنوان یکی از روشهای کشف دانش، روشن‌تر می‌شود. امروزه، استفاده از روشهای سنتی جمع‌آوری 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 بسیار حیاتی 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 تصمیم‌گیری جهت 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 جهت تصمیم‌گیری‌های سازمانی مطرح می‌باشند.

برای اولین بار مفهوم داده‌کاوی در کارگاه IJCAI¹ در زمینه کشف دانش از پایگاه داده توسط Shapir مطرح گردید. به دنبال آن در سالهای ۱۹۹۱ تا ۱۹۹۴، کارگاههای کشف دانش از پایگاه داده مفاهیم جدیدی را در این شاخه از علم ارائه کردند بطوری که بسیاری از علوم و مفاهیم با آن مرتبط گردیدند. به مرور زمان، استخراج و کشف سریع و دقیق اطلاعات با ارزش و پنهان از پایگاه داده‌ها، به عنوان داده‌کاوی مورد توجه قرار گرفت. به این شکل بود که فرایند داده‌کاوی به عنوان فرایند آماری، تجزیه و تحلیل در فرایند کشف دانش در پایگاه داده‌ها (KDD) پررنگ شد، به حدی که گاه داده‌کاوی^۲ (DM) به عنوان مترادف کشف دانش در پایگاه داده‌ها مورد استفاده قرار می‌گرفت. (آهوپای، ۱۳۹۰)

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 جست و جو 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 سایر متغیرها 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 تحت شرایط درست می‌تواند منجر به بینش درست شود. بسیاری از مسائل محیط اطراف خود را می‌توان در یک قالب گنجانده، به عبارت دیگر برای تبدیل یک مسئله به یک مسئله داده‌کاوی باید آن را به یکی از فعالیت‌های داده‌کاوی تبدیل نمود، یکی از این فعالیتها دسته‌بندی می‌باشد. هدف دسته‌بندی داده‌ها، سازماندهی و تخصیص داده‌ها به کلاس‌های مجزا است. در این پروسه بر اساس مجموعه داده‌های آموزشی، مدل اولیه‌ای ایجاد می‌شود، سپس این مدل برای دسته‌بندی داده‌های جدید مورد استفاده قرار می‌گیرد، به این ترتیب با بکارگیری مدل بدست آمده تعلق داده‌های جدید به دسته معین قابل پیش‌بینی است. به عبارت دیگر دسته‌بندی شامل بررسی ویژگی‌های یک شیء جدید و تخصیص آن به یکی از مجموعه‌های از قبل تعیین شده است.

۳- روش شناسی تحقیق

داده‌کاوی عمدتاً با ساختن مدل‌ها مرتبط است و یک مدل اساساً به الگوریتمی از قوانین گفته می‌شود که مجموعه‌ای از ورودی‌ها را با هدف یا مقصد خاصی مرتبط می‌نماید. بسیاری از مسائل محیط اطراف خود را می‌توان در قالب یک مدل گنجانده، به عبارت دیگر برای تبدیل یک مسئله به یک مسئله داده‌کاوی باید آن را به یکی از فعالیت‌های داده‌کاوی تبدیل نمود. یکی از فعالیت‌های متداول در داده‌کاوی دسته‌بندی می‌باشد.

هدف دسته‌بندی داده‌ها، سازماندهی و تخصیص داده‌ها به کلاس‌های مجزا است. در این پروسه بر اساس مجموعه داده‌های آموزشی، مدل اولیه‌ای ایجاد می‌شود، سپس این مدل برای دسته‌بندی داده‌های جدید مورد استفاده قرار می‌گیرد، به این ترتیب با بکارگیری مدل بدست آمده تعلق داده‌های جدید به دسته معین قابل پیش‌بینی است. به عبارت دیگر دسته‌بندی شامل بررسی ویژگی‌های یک شیء جدید و تخصیص آن به یکی از مجموعه‌های از قبل تعیین شده است. از جمله مدلها و رویکردهای دسته‌بندی می‌توان به درخت تصمیم و k نزدیکترین همسایگی اشاره کرد. در تحقیق حاضر این دو مدل با هم مقایسه شده‌اند.

1. Knowledge Discovery From Database (KDD)

2. Data Mining

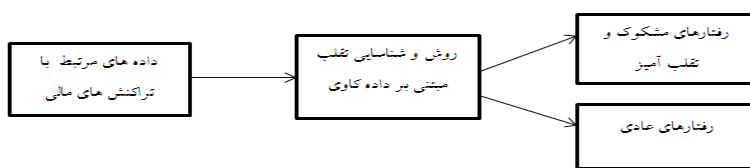


درخت تصمیم یکی از ابزارهای قوی و متداول برای پیش بینی و دسته بندی است. درختان تصمیم، یک ساختار درختی شبیه فلوچارت دارد. بالاترین گره در درخت، گره ریشه است و گره های برگ، دسته ها یا توزیع دسته ها را نشان می دهند. نمونه ها را با مرتب کردن آن ها در درخت از گره ریشه به سمت گره های برگ دسته بندی می کنند. هر گره داخلی در درخت، صفتی از نمونه را آزمایش می کند و هر شاخه ای که از آن گره خارج می شود متناظر یک مقدار ممکن برای آن صفت می باشد. همچنین به هر گره برگ، یک دسته بندی منتسب می شود. 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، دسته بندی می شود. این فرآیند برای هر زیر درختی که گره جدید ریشه آن می باشد تکرار می شود. (تیمورپور، ۱۳۹۰)

در مورد خصوصیات درخت تصمیم به موارد زیر می توان اشاره کرد:

- روش درخت تصمیم در تقسیم بندی داده ها به گروه های مختلف، به گونه ای است که هیچ داده ای حذف نمی شود.
- استفاده از درخت تصمیم آسان است.
- درک مدل ایجاد شده توسط درخت تصمیم آسان است. به عبارت دیگر، با وجود این که فهمیدن روش کار الگوریتم های سازنده درخت، چندان ساده نیست ولی فهمیدن نتایج به دست آمده از آن ها آسان است.
- دسته بندی هایی که توسط درخت تصمیم ایجاد می شوند، از روی شباهت داده های ذخیره شده در پارامترهای پیش بینی کننده، قابل انجام است.

الگوریتم k نزدیکترین همسایگی، یک الگوریتم تعلیم با سرپرستی است. در حالت کلی از این الگوریتم به دو منظور استفاده می شود: برای تخمین تابع چگالی توزیع داده های تعلیم و برای طبقه بندی داده های تست بر اساس الگوهای تعلیم. در این تحقیق سعی شده است به بررسی تقلب در بانکداری الکترونیکی در شاخه کارتهای الکترونیکی و ارائه راه کارهایی در این زمینه پرداخته شود. فرآیند مورد بررسی در این تحقیق شامل سه مرحله آماده سازی داده ها، دسته بندی تراکنشهای مالی و شناسایی تقلب است. مدل مفهومی تحقیق در شکل یک نشان داده شده است.



شکل ۱-مدل مفهومی

۳-۱- پیش پردازش داده ها

قبل از کشف دانش از هر منبع داده ای، باید اول داده ها به منظور سازگاری با الگوریتم های یادگیری و بیان صحیح واقعیت، در جهت توانایی استخراج دانش مفید از آنها آماده شوند. معمولاً گفته می شود که آماده سازی داده ۷۰ تا ۸۰ درصد از روند داده کاوی است. داده های خام معمولاً دچار مشکلاتی هستند و استفاده از آن ها به همین صورت موجب تضعیف مدل خواهد شد. پیش پردازش داده ها شامل تبدیلات پیچیده ای که برای کاهش ابعاد داده ها مورد استفاده قرار می گیرد. به طور خلاصه می توان گفت پیش پردازش داده ها شامل همه تبدیلاتی است که بر روی داده های خام صورت می گیرد و آن ها را به صورتی در می آورد که برای پردازش های بعدی نظیر استفاده در دسته بندی مورد استفاده قرار می گیرد. ابزارها و روش های مختلفی برای پیش پردازش وجود دارد مانند بهنجار کردن، که داده



نخستین کنفرانس ملی توسعه مدیریت پولی و بانکی

تحول در مدیریت نظام بانکی، گامی به سوی حساسه اقتصادی



۸ و ۹ بهمن ماه ۱۳۹۲ تهران - مرکز همایش‌های بین‌المللی صدا و سیما

ها را به داده‌هایی جدید با بازه تغییرات و یا توزیع مناسب تبدیل می‌کند، کاهش ابعاد، که برای حذف داده‌های تکراری، اضافه و یا نامربوط برای دسته‌بندی استفاده می‌شود.

مجموعه آموزشی، به مجموعه اطلاق می‌شود که در آن داده‌هایی که به طور پیش فرض در دسته‌های مختلفی قرار دارند، همراه با ساختار دسته‌بندی خود وارد سیستم می‌شود و سیستم بر اساس آنها به خود آموزش می‌دهد. از مجموعه آموزشی است که الگوریتم‌ها یادگیری می‌کنند و این مجموعه هم دارای داده‌های معتبر و هم دارای داده‌های نامعتبر می‌باشد.

مجموعه آزمایشی یا آزمون، به مجموعه اطلاق می‌شود که شامل داده‌هایی می‌باشد که برای تست الگوریتم استفاده شده است و به عبارتی میزان عملکرد الگوریتم مشخص شده است.

در این تحقیق داده‌ها به دو دسته مجموعه آموزشی که شامل ۸۰٪ داده‌هاست و مجموعه آزمایشی یا آزمون که شامل ۲۰٪ داده‌هاست تقسیم شده است.

۳-۲- داده‌های تحقیق

از آنجایی که ما از الگوریتم‌های دسته‌بندی برای تحلیل داده‌ها استفاده می‌کنیم، داده‌ها را در دو دسته به قرار زیر تعریف می‌کنیم (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 غیر عادی انجام شده است. تراکنش‌های مشکوک این تحقیق که برچسب تقلب دارند عبارتند از:

- کارت‌هایی که در بازه زمانی کوتاه شهری که تراکنش در آن صورت گرفته تغییر کند.
- تراکنش‌های بالای سه میلیون تومان
- کارتهایی که روند برداشت آنها در یک بازه زمانی کوتاه تغییر کرده است. به طور مثال کارت‌هایی در بیشتر مواقع برداشتهای با مبالغ پایین داشته است و به یکباره در یک بازه زمانی کوتاه برداشتهایی با مبالغ بالا دارد.

مولفه‌ها	شرح
نوع تحقیق	اکتشافی
جامعه آماری	مشتریان دارنده کاتهای بانکی
روش نمونه‌گیری	تصادفی
آماده‌سازی داده	قابلیت نرم افزار Weka روش Resample
روش تجزیه و تحلیل آماری	مدلهای داده کاوی گروه الگوریتم‌های دسته بندی

جدول ۱: روش اجرای تحقیق

۳-۳- نرم افزار و محیط پیاده سازی

عمده فعالیت‌های عملی این تحقیق با استفاده از نرم افزار Weka به انجام رسیده است. این نرم افزار به عنوان یک نرم افزار تجاری در زمینه فعالیت‌های داده کاوی شناخته شده است و تحلیل‌هایی که در خصوص خوشه بندی و دسته بندی و کارهای مرتبط با آن مطرح شده به کمک این نرم افزار پیاده سازی شده است. این نرم افزار با فراهم نمودن محیط بصری مناسب از داده‌ها، روشها و ابزارهای داده



کاوی به صورت شبکه ای از گره ها و جریان ها، به سادگی امکان مدلسازی و اجرای تحلیل های مختلف داده کاوی را فراهم می آورد. الگوریتم های موجود داده ها را به صورت یک جدول رابطه ای با فرمت Arff دریافت می کنند.

۳-۴- ارزیابی الگوریتم ها

متغیرهایی که برای ارزیابی الگوریتم ها در نظر گرفته شده است به قرار زیر می باشد:

۱. مثبت واقعی (TP) اگر الگوریتم تراکنش عادی را به درستی در گروه تراکنش عادی قرار دهد.
۲. منفی واقعی (FP) اگر الگوریتم تراکنش مشکوک را به درستی در گروه تراکنش مشکوک قرار دهد.
۳. مثبت کاذب (TN) اگر تراکنش عادی در دسته تراکنش مشکوک قرار گیرد.
۴. منفی کاذب (FN) اگر تراکنش مشکوک در دسته تراکنش عادی قرار گیرد.

الگوریتمی که کمترین خطا را در دسته بندی ما داشته باشد یعنی کمترین کاذب مثبت و کاذب منفی را دارا باشد الگوریتم بهینه ما خواهد بود.

برای تحلیل رفتار تراکنشی مشتری که عوامل تأثیر گذار در شناسایی رفتار مشکوک و تقلب برانگیز مشتری است، الگوریتم های دسته بندی به منظور بیان دانش و استخراج قوانین در ارتباط با مشتریان به کار می روند. الگوریتم ها را می توان از چند منظر مختلف بررسی نمود مانند سرعت پیش بینی، دقت، درستی و غیره. در این تحقیق قصد داریم تا دو الگوریتم مطرح شده که شامل نزدیک ترین همسایگی و درخت تصمیم هستند را با هم از نظر میزان درستی و دقت بررسی کنیم. هدف از مقایسه این الگوریتم ها با هم از منظر درستی و دقت بدست آوردن الگوریتم بهینه در شناسایی رفتار مشکوک و تقلب آمیز مشتریان بانک است. الگوریتمی که کمترین خطا را در دسته بندی داشته باشد یعنی کمترین کاذب مثبت و کاذب منفی دارا باشد الگوریتم بهینه ما خواهد بود.

برای محاسبه میزان درستی الگوریتم ها از فرمول های زیر استفاده می شود.

$$\text{میزان درستی} = TP + TN$$

$$\text{میزان نادرستی} = FP + FN$$

برای محاسبه میزان دقت الگوریتم ها پس از وارد کردن داده ها در نرم افزار و پیاده سازی الگوریتم بر روی داده ها با توجه به خروجی بدست آمده که به ماتریس در هم ریختگی معروف است دقت الگوریتم ها محاسبه می گردد.

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



نخستین کنفرانس ملی توسعه مدیریت پولی و بانکی

تحول در مدیریت نظام بانکی، گامی به سوی حماسه اقتصادی



۸ و ۹ بهمن ماه ۱۳۹۲ تهران - مرکز همایش‌های بین‌المللی صدا و سیما

Confusion Matrix	true negative	True positive	Class precision
Predicted negative	TN	FN	TN/(TN+FN)
Predicted positive	FP	TP	TP/(TP+FP)
Class recall	TN/(TN+FP)	TP/(TP+FN)	
Accuracy	(TN+TP)/(TN+TP+FN+FP)		

جدول ۲- ماتریس درهم ریختگی

برای شناسایی تقلب مشتریان و بدست آوردن الگوریتم بهینه، داده‌های آزمون مورد تحلیل قرار داده شده است، شاخص دقت بدست آمده از الگوریتم‌ها به قرار زیر می‌باشد. مقادیر هر یک از این معیارها عددی بین صفر و یک است که هرچه به یک نزدیکتر باشد، نشان دهنده این است که مدل عملکرد و دقت بهتری دارد و درخت تصمیم در بین این دو الگوریتم بالاترین دقت را دارد.

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area
	0.977	0.047	0.949	0.977	0.963	0.973
	0.953	0.023	0.979	0.953	0.966	0.973
Weighted Avg.	0.964	0.034	0.965	0.964	0.964	0.973

=== Confusion Matrix ===

```

a  b  <-- classified as
129 3 | a = YES
7 141 | b = NO
    
```

شکل ۳- خروجی الگوریتم درخت تصمیم

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area
	0.932	0.095	0.898	0.932	0.914	0.919
	0.905	0.068	0.937	0.905	0.921	0.919
Weighted Avg.	0.918	0.081	0.919	0.918	0.918	0.919

=== Confusion Matrix ===

```

a  b  <-- classified as
123 9 | a = YES
14 134 | b = NO
    
```

شکل ۴- خروجی الگوریتم نزدیکترین همسایگی

۴- نتایج

همان طوری که اشاره شد با توجه به رشد روز افزون خدمات مالی بانکها به صورت الکترونیکی در سطح کشور و افزایش استفاده از خدمات بانکداری الکترونیک، رویکرد کلاهبرداران و متقلبان به سمت بانکداری الکترونیک رو به افزایش است و در صورت عدم استفاده از سازوکارهای تشخیص و جلوگیری از تقلب باید شاهد افزایش آمار تقلب‌ها در بانکداری الکترونیک باشیم. در این مطالعه الگوریتم‌های



نخستین کنفرانس ملی توسعه مدیریت پولی و بانکی

تحول در مدیریت نظام بانکی، گامی به سوی حمله اقتصادی



۸ و ۹ بهمن ماه ۱۳۹۲ تهران - مرکز همایش‌های بین‌المللی صدا و سیما

داده کاوی نظیر الگوریتم درخت تصمیم و نزدیکترین همسایگی بررسی شده است. با توجه به نتایج بدست آمده می توان از الگوریتم درخت تصمیم در شناسایی تراکنشهای مشکوک مشتریان کارت استفاده کرد و با تولید نرم افزارهایی، تمهیدات لازم را برای پیشگیری از تقلبهایی که ممکن است در آینده رخ دهد، جلوگیری کرد.

مولفه‌هایی که برای شناسایی رفتار مشکوک و ضریب تاثیر پذیری در نظر گرفته شده است. با توجه به نتایج بدست آمده از نرم افزار به قرار زیر می باشد.

۱. شهری که تراکنش رخداد داده است

۲. زمان تراکنش

۳. تاریخ وقوع

در واقع ما در این تحقیق به شناسایی چند نوع از تراکنش های مشکوک و پیش بینی و شناسایی آنها پرداخته ایم که این بررسی پیش مقدمه‌ای برای جلوگیری از تقلب است. در واقع راهکار ارایه شده در این تحقیق مبین این مسئله است که بانک باید ابتدا با سیستم هوشمند به بررسی تراکنش های مشکوک بپردازد برای مثال برداشت بیش از حد و غیر عادی در فواصل زمانی کوتاه و غیره. از طرفی با توجه به یادگیری ماشین و داده کاوی می توان هشدار در سیستم ایجاد کرد تا در صورت روبرو شدن با این تراکنش های مشکوک سوال امنیتی را از صاحب کارت بپرسد و امنیت بالاتری را برای کاربران ایجاد نماید.

dataacademy.ir

۵-منابع

۱. آهوپای، ر. (۱۳۹۰). "روشهای کشف تقلب در استفاده از کارتهای اعتباری". ایران دیتا.
۲. تیمورپور، ب.، علیزاده، س. (۱۳۹۰). "داده کاوی و کشف، دانش". انتشارات دانشگاه علم و صنعت ایران.
۳. حاتمی راد، ع و شهریاری، ح. (۱۳۸۹). "روشها و راهکارهای شناسایی تقلب در بانکداری الکترونیک". تازه های اقتصاد
۴. شهرابی، ج. (۱۳۹۰). "داده کاوی در صنعت بانکداری". انتشارات جهاد دانشگاهی واحد صنعتی امیرکبیر.
۵. فیروزی، م. شکوری، م. کاظمی، لیلا. (۱۳۹۰). "شناسایی تقلب در بیمه اتومبیل بر استفاده از روش داده کاوی". نشریه پژوهشکده بیمه.
۶. کوثری لنگری، ر. زودآیند. (۱۳۹۰). "بکارگیری الگوریتمهای درخت تصمیم گیری جهت کشف رفتارهای مشکوک در بانکداری اینترنتی". پژوهشنامه پردازش و مدیریت اطلاعات.
۷. معین زاد، ح. (۱۳۹۰). بانکداری الکترونیک و کشف تخلفات. مجله بانک ملی ایران.

8. Bhatl Tej P. Prabhu V & Dua A.(2003). "Understanding Credit Card Frauds, Card Business Review.United Nations. Economic and Social Council. Economic Comission for Africa.

9. Yu H.C,His K.H and Kuo.P.J ,(2002), Electronic Payment System: Analysis and Comparasion of Type Technology in Society