به‌کارگیری الگوریتم‌های درخت تصمیم‌گیری جهت کشف رفتارهای مشکوک در بانکداری اینترنتی

روح‌الله کوثری لنگری[1] | کارشناس ارشد، مدیریت فناوری اطلاعات، دانشگاه پیام نور

نصرالله مقدم چرکری[2] | عضو هیئت علمی، دانشگاه تربیت مدرس، گروه مهندسی برق و کامپیوتر

داود وحدت*| عضو هیئت علمی، دانشگاه پیام نور، گروه مهندسی فناوری اطلاعات

دریافت: ۱۳۹۰/۰۹/۲٤ | پذیرش: ۱۳۹۱/۰۲/۱۳

فصلنامهٔ علمی پژوهشی
پژوهشگاه علوم و فنّاوری اطلاعات ایران
شاپا(چاپی) ۸۲۲۳-۲۲۵۱
شاپا(الکترونیکی) ۸۲۳۱-۲۲۵۱
نمایه در LISA، SCOPUS و ISC
http://jipm.irandoc.ac.ir
دوره ۲۸ | شمارهٔ ۳ | ص ص ۶۸۱- ۷۰۰
بهار ۱۳۹۲

نوع مقاله: پژوهشی

چکیده: دگرگونی جهان به واسطه گسترش فناوری اینترنت، رقابتی دانش‌محور را در عرصه تجارت الکترونیکی به وجود آورده است. به‌طوری‌که با افزایش نرخ مبادلات تجاری، تضمین امنیت، منوط به تحقق نظام پویای بانکداری الکترونیکی است. بانکداری اینترنتی به عنوان یک فرصت بالقوه و رکن اساسی، در فضای سایبر با تهدیدهای گوناگونی مواجه است که یکی از این چالش‌ها، عدم قطعیت در تضمین امنیت تراکنش‌ها و وجود رفتارهای غیرمتعارف از سوی شیادان الکترونیکی است. از آنجایی که رفتار کاربران در سامانه اینترنتی با عدم قطعیت همراه بوده و سوابق تراکنش‌ها، تنها راه کنترل این حرکات است، دانش استنتاجی خبرگان و دسته‌بندی الگوی رفتاری کاربران توسط الگوریتم درخت تصمیم، جهت کشف تقلب و رفتارهای مشکوک راهگشاست. در این مقاله، ابتدا متغیرهای مؤثر در تولید قوانین رفتاری تعیین شده است و در نهایت، روند چهار الگوریتم Chaid، ex_Chaid، C4.5 و C5.0 مورد مقایسه قرار گرفته است. نتایج پژوهش نشان می‌دهد که الگوریتم C5.0 با دقت ۹۱ درصد می‌تواند به عنوان روش ماشینی مطمئن جهت کشف الگوهای مشکوک موجود روی تراکنش‌های بانک محسوب شود.

کلیدواژه‌ها: بانکداری اینترنتی[3]، دسته‌بندی[4]، درخت تصمیم[5]، رفتار مشکوک، تقلب[6]

1. Kosari58@yahoo.com
2. Charkari@modares.ac.ir
*Vahdat@pnu.ac.ir
3. internet banking
4. classification
5. decision tree
6. fraud

۱. مقدمه

گرایش مردم به خرید برخط روز به روز در حال افزایش است. مطابق مطالعه ای‌سی‌نیلسن[1]، یک دهم مردم جهان به صورت برخط خرید می‌کنند (Kelly 2003). بنا به گزارش آنکتاد[2]، تعداد کاربران اینترنتی در سال ۲۰۰۹ نزدیک به ۱.۷۳ میلیارد نفر از جمعیت دنیا را تشکیل می‌دهد که این رقم رشد ۱۸درصدی را در مقایسه با آمار پایان سال قبل نشان می‌دهد (UNCTAD 2009). گسترش این روند به عنوان یک فرصت، عامل تهدیدی بزرگ خواهد بود چرا که فرصت جدیدی را جهت ارتکاب تقلب و سوءاستفاده‌های مالی شیادان فراهم خواهد ساخت. با توجه به پژوهش‌های گروه KPMG در سال ۲۰۰۳، نتایج نشان از افزایش خطر و نرخ تقلب، نسبت به گذشته دارد. در این پژوهش‌ها ۷۵٪ سازمان‌های تحت مطالعه، نمونه‌هایی از تقلب را تجربه کرده بودند. این نسبت، ۱۳٪ افزایش نسبت به سال ۱۹۹۸ داشته است (KPMG 2003). به تازگی، موسّسه SAS[3]، تخمین خود را از هزینه فریب در اقتصاد انگلستان از ۱۴ میلیارد پوند به ۱۸ میلیارد پوند در سال افزایش داده است (Dorrington 2003).

نتایج پژوهش‌های ارنست و یانگ[4] عوامل افزایش تقلب را در سازمان‌ها افزایش پیچیدگی سازمان‌ها، تغییر در فرآیندهای کسب و کار، کنترل‌های داخلی غیرمؤثر معرفی می‌کند که سبب افزایش فرصت برای استفاده‌های سودجویانه شده است (حسنی و دیگران ۱۳۸۷). اثرات سازمانی این تقلب‌ها علاوه بر ضررهای اقتصادی، بر روی شهرت سازمان و سطح رضایت و اعتماد مشتریان نیز وارد است (Ernest and Young 2003). آمارها نشان می‌دهد که بانکداری الکترونیک سریع‌ترین رشد استفاده‌کنندگان از اینترنت را دارد و طبق این آمارها ۴۴٪ از مردم آمریکا از این سرویس استفاده می‌کنند (Quah and Striganesh 2007). مجموع تقلّب برخط سال ۲۰۰۶ در ایالات متّحده بیش از ۳ میلیارد برآورد شده است و در خارج از ایالات متّحده به ترتیب ۱.۶ تا ۱.۷ میلیارد دلار است (Waite and Harrison 2004).

با افزایش خدمات بانک‌ها دراینترنت و رشد تراکنش‌های برخط توسط مشتریان، میزان بروز جرائم در صنعت بانکداری اینترنتی نیز به سرعت در حال رشد است، به‌طوریکه آهنگ رشد جرائم برخط بین ۸ تا ۹ درصد در سال تخمین زده می‌شود (Quah and Sriganesg 2007). آمارها نشان می‌دهد که میزان ضرر مالی بانک‌ها از این جرائم در انگلستان در سال ۲۰۰۷ میلادی بالغ بر ۲۲.۶ میلیون پوند بوده است (Young 2008). در این راستا بانک‌ها و مؤسسات

1 . A.C. Nielsen 2. UNCTAD 3. Statements on Auditing Standards 4. Ernst andYoung

مالی، با تجهیز به سامانه‌های کشف تقلب می‌توانند به پیشگیری و حذف فرآیندهای حامی این گونه اعمال بپردازند. این پژوهش سعی دارد با فناوری داده‌کاوی و رفتارشناسی اینترنتی، روشی نوین در کشف ماشینی رفتارهای مشکوک در تراکنش‌های مالی را به نمایش در آورد. همچنین، دامنه کاربرد روش‌های ماشینی با توجه به سوابق پژوهشی دیگر بانک‌های دنیا، حاکی از آن است که در وجود عدم تفاهم در قوانین مالی، سوددهی و پارامترهای پردازش اینترنتی بانک‌های ربوی نسبت به بانکداری اسلامی که قواعد بومی را رعایت می‌کنند، نشان از تطابق کاربرد و تشابه نتایج در به‌کارگیری این روش‌ها جهت کشف رفتارهای مشکوک دارد.

۲. ادبیات پژوهش

بانکداری اینترنتی علاوه بر مزیت بهبود کارایی و سرعت خدمات، در راستای شکل‌گیری شبکه اینترنت با مباحث جدیدی پیرامون محرمانگی و امنیت اطلاعات مواجه است. نظام بانکداری الکترونیکی باید الزام تصدیق اصالت، محرمانگی، یکپارچگی و انکارناپذیری و دیگر عوامل امنیتی را در نظر داشته باشد و تضمین کند که فقط افراد مجاز به اطلاعات محرمانه و حساب‌های مشتریان دسترسی داشته و نیز معاملات صورت‌گرفته غیرقابل ردیابی و رسیدگی باشند (فرد ۱۳۸۴). تکمیل پروتکل‌های امنیتی در لایه‌های مختلف به‌خصوص در سطح خدمات اینترنتی و تضمین فرایند پرداخت الکترونیکی مستلزم پاسخ به سه مشکل اصلی در فضای ایمن مالی است:

الف) کلاهبرداری[1]

چگونه می‌توانیم به مشتری این اطمینان را بدهیم که با ورود به سایت و انجام معامله در آن، شماره رمز کارت اعتباری وی مورد سرقت و جعل قرار نخواهد گرفت؟

ب) شنود[2]

چگونه می‌توانیم مطمئن شویم که اطلاعات شماره حساب مشتری هنگامی که برای یک معامله امن

در وب اقدام می‌کند، قابل دستیابی برای متخلفان نیست؟

ج) دگرگون کردن اطلاعات[3]

چگونه می‌توانیم مطمئن باشیم که اطلاعات شخصی مشتری توسط متخلفان قابل تغییر نیست؟

۲-۱. روش‌های معمول تقلب در بانکداری الکترونیکی

1. Spoofing 2. Eavesdropping 3. data alteration

برای مقابله با عوامل نفوذ در بانکداری الکترونیکی راه‌های مختلفی وجود دارد که به‌کارگیری آنها مستلزم احاطه کامل به مفاهیم حمله[1]، فریب[2]، جرم[3]، جرم شناسی و رفتارهای مشکوک و جلوگیری از آنهاست. فریب که یکی از مصادیق جرم عمومی به‌شمار می‌آید با هدف به‌دست آوردن سود ناروا و ناصحیح صورت می‌گیرد (Phua 2003). در کل، جرائم الکترونیکی به جرائمی اطلاق می‌شود که به نوعی با تجارت الکترونیکی مرتبط باشد و ارتباط بین جرم و قربانی تحت یک شبکه رایانه‌ای انجام شود (Shah et al. 2005).

۲-۱-۱. جعل هویت[4]

جعل عنوان از جمله موارد سوءاستفاده به‌شمار می‌آید که از مشخصه‌های فردی شخصی دیگر مانند: نام، شماره ملی، شماره کارت اعتباری، به منظور انجام امور مجرمانه، کلاهبرداری یا سرقت سوءاستفاده شود. مطابق با گزارش‌های ارائه‌شده میزان خسارت ناشی از جعل عنوان در سال ۲۰۰۵ در انگلستان معادل ۱/۷۲ میلیارد پوند برآورد شده است (حسنی، سلطانی، و ضرابیه ۱۳۸۷).

۲-۱-۲. به‌دست آوردن حساب[5]

این نوع کلاهبرداری یکی از انواع رایج جعل عنوان است به طوری که فرد شیاد پس از به‌دست آوردن اطلاعات مشخص، شماره حساب و تغییر نشانی ایمیل رسمی طعمه خود و با ارسال ایمیلی به بانک مبنی بر گم‌شدن یا دزدیده شدن کارت، تقاضای کارت جدیدی می‌نماید.

۲-۱-۳. بیراهه کشانی[6]

حمله نفوذگر[7] به منظور تغییر ترافیک وب‌سایت به یک وب‌سایت جعلی دیگر است. در این بخش از شیادی، با دستکاری سرویس‌دهنده دامین[8] توسط فرد شیاد که در اصطلاح فنیّ «سمیّ شدن[9]» سرویس‌دهنده دامین معروف است، منجر می‌شود. کاربر به تصور اینکه وارد سایت اصلی بانک می‌شود، وارد سایت جعلی فرد شیاد می‌گردد و اطلاعات محرمانه بانکی اعم از شماره حساب، شماره کارت و کلمه عبور را وارد می‌کند (حسنی، سلطانی، و ضرابیه ۱۳۸۷).

۲-۱-٤. سرقت هویت[10]

فرایندی است که متخلف با جلب اعتماد کاربر، اطلاعات شخصی، کلمه عبور و اطلاعات مالی محرمانه را در اختیار شیاد قرار دهد. در این فرایند، اطلاعات در قالب فرم‌ها با عنوان بانک، مؤسسه‌های وابسته به بانک یا دولت برای طعمه ارسال می‌شود و طعمه بدون اطلاع از

---

1. Attack
2. Fraud
3. Crime
4. identify theft
5. account takeover
6. Pharming
7. Hacker
8. DNS (Domain Name Service)
9. Poisoned
10. phishing

اینکه فرم دریافتی جعلی است، ناآگاهانه اطلاعات مورد نظر را برای شیاد ارسال می‌نماید (حسنی، سلطانی، و ضرابیه ۱۳۸۷).

۲-۱-۵. تقلب سیمی[1]

تقلب سیمی، جرمی است که با استفاده از وسائل ارتباط الکترونیکی از راه دور همانند پست الکترونیکی جهت فریب قربانیان صورت می‌گیرد. نمونه بارز شبکه‌های سیمی، انتقال اطلاعات مالی SWIFT[2] است، چرا که ارتباط مالی در این شبکه توسط شبکه‌ای بین‌المللی از ارتباطات واسط صورت می‌پذیرد. تقلب در این نظام بدین صورت است که کاربر سوئیفت پیام تقاضای جعلی را به طرف بانک سپرده‌گذار ارسال کند. این نوع کلاهبرداری به طور معمول در بانک‌های آفشور[3] (کشورهایی که بدلیل ضعیف بودن اقتصادشان اقدام به جذب شرکت‌های خارجی می‌کنند و از طریق شرکت‌های واسط به فعالیت‌های مالی می‌پردازند.) رخ می‌دهد (Phua 2003).

۲-۱-٦. تقلب کارت‌های پرداخت٤

تقلب‌های کارت‌های اعتباری شامل طیف وسیعی مربوط به تقلب در سمت مشتری، فروشنده و بانک است. در سمت مشتریان، با افزایش مقطعی حجم تراکنش‌ها، اعتبار بیشتری از سوی بانک کسب می‌کند و در ادامه، خریدهای بدون اعتبار بیشتری انجام می‌دهد. در سمت فروشنده، فروشنده می‌تواند با نرخ های سود مختلف بهره بیشتری از این تراکنش‌ها داشته باشد. در سمت بانک، کارمند بانک با افزایش زودهنگام اعتبار مشتری و عدم اطلاع وی از این منابع مالی استفاده می‌نماید (Fravdlab n.d).

۲-۱-۷. دزدی اطلاعات کارت‌های بانکی[5]

در این حالت، فرد متقلب اقدام به نصب پایانه فروش[6] و یا کارت‌خوان می‌کند و زمانی که دارنده کارت از این دستگاه استفاده می‌کند، اطلاعات موجود در نوار مغناطیسی کارت کپی و از شماره‌های روی کارت به‌وسیله دوربین کوچک تعبیه‌شده در دستگاه عکس گرفته می‌شود. یک نمونه از این نوع تقلب که به‌وسیله خودپرداز[7] صورت گرفته بود به این شکل

1. wire fraud
2. Society for Worldwide Interbank Financial Telecommunication
3. Offshore
4. payment card fraud
5. duplication or skimming of card information
6. POS
7. ATM

صورت گرفت: زمانی که کارت در داخل دستگاه قرار می‌گیرد، اطلاعات موجود در نوار مغناطیسی کپی و سپس، این اطلاعات از روی کارت پاک و پس از آن، عکسی از شماره‌های روی کارت گرفته و از سوی دیگر، رمز کارت ذخیره می‌شود. فرد متقلب سپس اقدام به تولید یک کارت جدید با همان اطلاعات می‌کند و این در حالی است که کارت اصلی به دلیل پاک شدن اطلاعاتش غیر فعال شده است (Fravdlab n.d).

۲-۲. مروری بر فنّ درخت تصمیم‌گیری

بانک‌های فراهم‌کننده خدمات اینترنتی، روش‌های مختلفی را برای شناسایی مشتریان خود جهت جلوگیری از تقلب به‌کار می‌برند. یکی از این روش‌ها، استفاده از فنّ داده‌کاوی است که بر تحلیل‌های آماری و استفاده از الگوهای تراکنش جهت کشف رفتار مشتریان و شناسایی جرم تمرکز دارد (Quah and Striganesh 2007). این روش‌ها مبتنی بر یادگیری هستند و قادرند شاخص‌های رفتار فریب‌آمیز را از پایگاه تراکنش کاربران کشف کنند. این شاخص‌ها برای ایجاد سامانه‌های ناظر[1] استفاده می‌شوند تا رفتارهای غیرمعمول مشتریان را ثبت و آنها را شناسایی کنند.

روش مبتنی بر درخت تصمیم، یکی از ابزارهای قوی برای دسته‌بندی و کشف محسوب می‌شود. درخت از نظریه اطلاع[2] و مقدار آنتروپی[3] جهت انتخاب بهترین متغیر و شروع پیمایش استفاده می‌کند. هر رأس درخت به عنوان یک کلاس یا قاعده، نمایشگر یک آزمایش یا تصمیم یکتا است. یال‌های هر رأس، متناظر با احتمال‌های حاصل از آزمایش روی رأس است. پیشامدها داده‌ها را به چندین زیرمجموعه افراز می‌کند که توسط برگ‌های این درخت شناخته می‌شود. درخت برخلاف سایر فنون به تولید قانون می‌پردازد و پیش‌بینی خود را در قالب قوانین توضیح می‌دهد، در حالی‌که در سایر فنون پیش‌بینی نهایی بدون چگونگی اجرا بیان می‌شود. همچنین، نسبت به سایر مدل‌های دسته‌بندی زودتر محاسبه می‌شود و از دقت بالاتری برخوردار است (Mitchell 1997). درخت قابل درک است و به شکل مجموعه‌ایی از قوانین "اگر – آنگاه[4]" نمایش داده می‌شود. در این پژوهش، جهت صحت رفتارشناسی کاربران اینترنتی، ابتدا مطابق با مدل مفهومی پژوهش، متغیرهای مؤثر که در بانک‌های ایرانی نقش به‌سزایی را در مشتری‌شناسی دارند استخراج و نقش و ضریب اثر هرکدام در پایگاه داده تراکنش‌های اینترنتی

1. monitoring 2. information gain 3. entropy 4. if- then- else

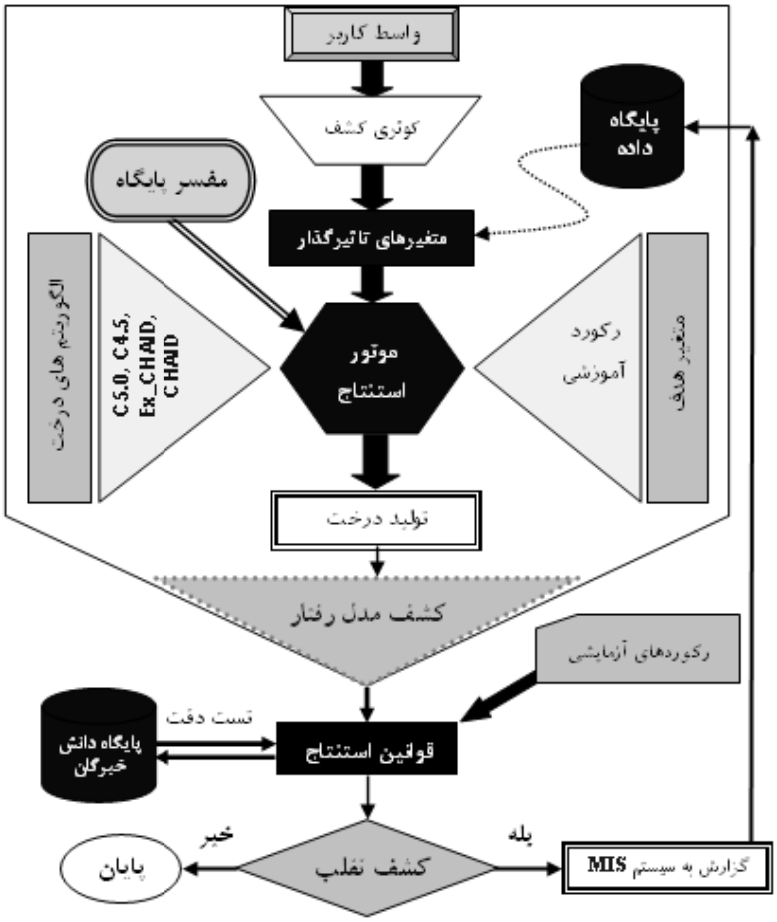
تعیین شد. در ادامه، با توجه به نظر مهندسان و خبرگان اینترنتی بانک دسته‌بندی تهدید رفتارها و ضریب دقت هرکدام صورت گرفت و در راستای اعمال داده‌های آموزشی و آزمایشی، دقت و صحت چهار الگوریتم معروف درخت تصمیم، یعنی Chaid، ex_Chaid، C4.5 و C5.0 جهت کشف رفتارهای مشکوک مورد مقایسه قرار گرفت.

۳. مدل مفهومی پژوهش

با توجه به اینکه تشخیص عوامل مؤثر در زمینه شناسایی رفتار و دسته‌بندی الگوهای رفتاری کاربران در سامانه اینترنتی، شامل ملاحظاتی پیرامون ویژگی‌های روان‌شناختی و شخصیتی کاربران است و با عدم قطعیت همراه است، بنابراین در نظر گرفتن تمامی عوامل و پارامترهای مؤثر غیرممکن است و سوابق به‌جای مانده از تراکنش‌ها فقط راه شناسایی و کنترل این حرکات خواهد بود. یوفنگ و همکارانش در سال ۲۰۰۲ به بررسی فنون شناسایی تقلّب در سه حوزه، کارت‌های اعتباری، شناسایی تجاوز کامپیوتری و سامانه اینترنتی پرداختند و رویکرد شبکه‌های عصبی را ابزار بسیار معمول برای شناسایی تقلّب معرفی کردند (Yufeng 2002). همچنین، جهت شناسایی و تشخیص جرم در بانکداری اینترنتی از شبکه‌های بیزین[1]، الگوریتم‌های ژنتیک و شبکه‌های عصبی مصنوعی[2] استفاده شد که قادر به استخراج الگو از پایگاه حاوی تراکنش‌های گذشته مشتریان بود که با کمک فرایند یادگیری با نظارت و استفاده از مجموعه‌های یادگیرنده توانست با قابلیت ۹۰ درصد تطابق، برای ساختن مدلی از تراکنش‌های فریب‌آمیز استفاده گردد. این شبکه‌ها آموزش‌پذیر و قابلیت انطباق با شکل‌های جدید جرم را نیز دارا هستند (Quah and Striganesh 2007). به عنوان نمونه در پژوهشی دیگر، یک شبکه عصبی با کمک مجموعه بزرگی از تراکنش‌ها که می‌توانند طی فرایندی الگوی فعالیت‌های غیرقانونی و مجرمانه را شناسایی کنند، طراحی شده است (Bignell 2006). در پژوهشی که در سال ۲۰۰۵ روی سایت e-Bay صورت گرفت، از دو فنّ C4.5 و C5 که مربوط به الگوریتم‌های درخت تصمیم‌ساز است در جهت تشخیص کلاهبرداری در حراج الکترونیکی سایت e-Bay استفاده شد که تا ۸۳٪ در ارائه الگوی‌های متناظر با رفتار مشتریان در جهت کشف کلاهبرداری مؤثر واقع گردید (Chau and Faloutsos 2005).

با توجه به مطالعات مختلف انجام‌شده در این زمینه و الگو گرفتن از مقالات مرتبط مدل مفهومی بومی‌شده جهت کشف رفتارهای مشکوک به صورت شکل ۱ ارائه شده است.

1. Bayesian belief networks
2. artificial neural network



شکل ۱. مدل مفهومی کشف رفتار مشکوک

همچنین، با توجه به سوابق مشتریان، بیش از ۲۰ متغیر مربوط به الگوی رفتاری توسط خبرگان اینترنتی انتخاب شد که در نهایت، با توجه به دسته‌بندی‌های انجام‌شده ۷ عامل به عنوان عوامل مؤثر جهت شناسایی رفتارهای مشکوک در بانکداری اینترنتی مطابق با نظر خبرگان اینترنتی چند بانک خصوصی در کشور در راستای مدل ارائه‌شده به اثبات رسیده است. متغیرهای ورودی و خروجی مدل مفهومی پژوهش مطابق جدول۱ معرفی شده است. همچنین، میزان تأثیر عوامل پژوهش و نسبت هر یک در بروز رفتار به شکل دانش استنتاجی در ادامه آمده است.

جدول ۱. توضیح متغیرهای ورودی و خروجی مدل مفهومی پژوهش

ردیف	متغیر ورودی	برچسب	توضیح
۱	تعداد خطا	Err_Cnt	تعداد تلاش ناموفق کاربر جهت ورود به سامانه
۲	ورود موفق	Login_Cnt	تعداد دفعات ورود موفق کاربر به سامانه
۳	آی‌اس‌پی	ISP_Cnt	تعداد مراکز سرویس اینترنتی مورد استفاده کاربر
۴	مرورگر	Browser_Cnt	تعداد مرورگرهای مختلفی که کاربر برای ورود به سامانه استفاده کرده است.
۵	آی‌پی	IP_Cnt	تنوع تعداد IPهای مختلفی است که در زمان ورود کاربر به سامانه ثبت شده است.
۶	تعداد حواله	FT_Cnt	بیشترین تعداد دفعاتی است که کاربر حواله با سقف پایین انجام داده است.
۷	مبلغ حواله	FT_Amnt_Cnt	میانگین مبلغی است که کاربر، حواله با سقف پایین انجام داده است.
۸	رفتار	Result	متغیر خروجی که نشانگر نوع و شدت رفتار کاربر بنا به غیرمعمول بودن رفتارها از عادی تا خطرناک دسته‌بندی می‌شود.

٤. روش‌شناسی پژوهش

روش این پژوهش، روش مطالعه موردی است، زیرا به طور ویژه بر روی یک بانک خصوصی مورد پژوهش تمرکز دارد.

الف) پژوهش از نظر هدف، از نوع کاربردی است، زیرا نتایج آن برای آگاهی مدیران بانک از عوامل مؤثر بر میزان پذیرش یکی از خدمات بانکداری الکترونیکی توسط مشتریان کاربرد دارد.

ب) پژوهش از نظر مکانی از نوع میدانی است، زیرا داده‌های آن با حضور در جامعه و یا نمونه آماری و با استفاده از ابزارهای مصاحبه و پرسشنامه گردآوری می‌شود.

ج) پژوهش از نظر روش، در زمره پژوهش‌های توصیفی- پیمایشی است، زیرا به بررسی توزیع ویژگی‌های رفتاری مشتریان بانک مورد پژوهش می‌پردازد و عناصر و متغیرهای پژوهش و نحوه ارتباط میان آنان را در چارچوب مشخص توصیف می‌کند.

٤-١. جامعه آماری

جامعه اول پژوهش شامل تمام تراکنش‌های اینترنتی مشتریان مرتبط با بانک خصوصی مورد پژوهش است که از سامانه اینترنتی استفاده می‌کنند و جامعه دوم مورد بررسی در حقیقت، مدیران و خبرگان اینترنتی بانک هستند که باید دانش استنتاجی آنان در جهت تعیین متغیرهای هدف گردآوری شود.

٤-٢. حجم نمونه

با توجه اینکه روزانه به طور متوسط ۳,۰۰۰,۰۰۰ رکورد درسامانه اینترنتی مورد پژوهش ثبت می‌شود و تشابه رفتاری نزدیکی در بیشتر رکوردها قابل مشاهده است، جهت دقت و اعتماد به نتیجه پژوهش به تعداد ۱۰۰،۰۰۰ رکورد به عنوان نمونه آموزشی و از ۷۰ رکورد دیگر به عنوان نمونه آزمایشی استفاده شده است. همچنین، حجم نمونه خبرگان نیز ۳۰ نمونه است.

٤-٣. ابزار گردآوری داده‌ها

اطلاعات لازم در پژوهش حاضر با استفاده از روش‌های مطالعات کتابخانه‌ای شامل کتب، مقالات، مجلات، گزارش‌های پژوهشی، اسناد موجود و اینترنت گردآوری شده است و همچنین بخشی از آن مربوط به پرسشنامه و مصاحبه با خبرگان و کارشناسان سامانه اینترنتی است. داده‌های آموزشی پژوهش به تعداد ۱۰۰،۰۰۰ رکورد اولیه به عنوان نمونه واقعی از بانک خصوصی دریافت شد و پیش‌پردازش اولیه آن نیز توسط نرم‌افزار SQL Server 2005 صورت گرفت.

٤-٤. روایی آزمون

برای اطمینان از روایی پرسشنامه پژوهش، از نظرات اساتید و خبرگان اینترنتی بانک که بیشتر مدیران باسابقه بخش بانکداری الکترونیکی، مهندسان نرم‌افزار و شبکه آگاه به فنّ داده‌کاوی و کارشناسان علوم بانکی که مسئول نگهداری پایگاه داده اینترنتی بانک هستند، مورد استفاده قرار گرفت و پس از اصلاحات لازم، پرسشنامه نهایی تهیه گردید.

٤-٥. پایایی آزمون

در این پژوهش، از روش آلفای کرونباخ برای پایایی استفاده شده است. برای محاسبه ضریب آلفای کرونباخ، ابتدا یک نمونه ۳۰ نفری که مشتمل بر نظرات خبرگان اینترنتی بانک خصوصی است، گردآوری شد و با استفاده از تجزیه و تحلیل داده‌های نمونه مقدماتی از طریق نرم‌افزار آماری SPSS، ضریب آلفای پرسشنامه ۰.۹ محاسبه گردید.

٤-٦. سؤالات اصلی پژوهش

۱. الگوهای رفتاری غیرمتعارف یا تهدیدآمیز در بانکداری اینترنتی چیست؟
۲. ویژگی‌های مورد نیاز از تراکنش‌های اینترنتی در بانکداری الکترونیکی به چه صورت استخراج می‌شود؟
۳. چگونگی نگاشت الگوهای مرتبط بر روی درخت تصمیم‌ساز به چه صورت است؟
۴. اعتبار الگوریتم درخت تصمیم‌گیری در نزد خبرگان اینترنتی بانک تا چه حدی است؟

۵. تجزیه و تحلیل داده‌ها

در این بخش از مقاله، فرایند گردآوری قوانین استنتاجی و تعیین میزان تأثیر متغیرهای ورودی جهت کشف رفتار معرفی می‌گردد.

۵-۱. دسته‌بندی رفتار

هدف بخشی از این پژوهش، شناسایی طیف وسیعی از رفتارهای عادی تا مشکوک کاربران در سامانه اینترنتی است. بنابراین برای رسیدن به تعریف واضحی از درجه مشکوک بودن رفتار و اتخاذ راهبردهای صحیح جهت جلوگیری و کشف رفتارهای تقلب‌آمیز، باید تعریف و دسته‌بندی دقیقی از انواع رفتار ارائه داد که این تفکیک بر اساس تجربه‌ای که از خبرگان اینترنتی بانک به‌دست آمده است، مطابق جدول ۲ انجام گرفته است:

۱. رفتار عادی: گویای تلاش بدون خطا و اشتباه کاربر در زمان ورود به سامانه است.
۲. رفتار کمی مشکوک: رفتاری که کاربر با خطا وارد سامانه شود و در حین انجام تراکنش نیز مشخصات مکانی کاربر نامشخص باشد (مشخصه IP سامانه کاربر که گویای موقعیت جغرافیایی درخواست است به‌طور جعلی فیلتر شود و قابل شناسایی نباشد).
۳. رفتار مشکوک: رفتاری که کاربر هنگام ورود، خطای پی‌درپی داشته و به تلاش غیرمجاز مظنون است و یا اینکه ورودی‌های متعددی بدون انجام تراکنش مالی داشته است.
۴. رفتار بسیار مشکوک: رفتاری که به طور کامل گویای تلاش مکرر غیرمجاز کاربر به سامانه و عملیاتی که غیرمعمول است، تکرار گردد.
۵. رفتار خطرناک: رفتاری که گویای تحرک غیرمجاز کاربر است و تلاش به انجام تراکنشی غیرعادی را تکرار نماید.

جدول ۲. معرفی پارامترهای متغیر هدف (نوع رفتار)

ردیف	نام پارامتر	برچسب	درصد تقلب
۱	عادی	NS	0
۲	کمی مشکوک	LS	% 25
۳	مشکوک	S	% 50
۴	خیلی مشکوک	VS	% 75
۵	تقلب	F	% 100

۵-۲. تعیین متغیرهای ورودی رفتار

اولویت‌بندی و وزن‌دهی سؤالات پرسشنامه، شامل تعیین میزان میانگین تأثیر متغیرهای ورودی جهت اعتبار بروز رفتار اینترنتی است که مطابق با نظر خبرگان در جدول ۳ الی ۹ آمده است.

جدول۳. معرفی پارامترهای متغیر ورودی خطای کاربر

ردیف	نام پارامتر	برچسب	شناسه	وزن میانگین	شاخص ورودی
۱	بدون اشتباه	Z	۰	0.7	X ≤ 1
۲	کم اشتباه	L	۱	2.9	1 <X≤ 3
۳	چند اشتباه	N	۲	5.65	3 <X≤ 6
۴	پراشتباه	H	۳	8.3	6 <X≤ 8
۵	بسیار پراشتباه	VH	۴	12	X>8

جدول ٤. معرفی پارامترهای متغیر ورودی تعداد ورود موفق

ردیف	نام پارامتر	برچسب	شناسه	وزن میانگین	شاخص ورودی
۱	کم	L	۱	4.55	X ≤ 5
۲	متوسط	M	۲	9.7	5 <X≤ 10
۳	زیاد	H	۳	13.8	X > 10

جدول ۵. معرفی پارامترهای متغیر ورودی تعداد ISP

ردیف	نام پارامتر	برچسب	شناسه	وزن میانگین	شاخص ورودی
۱	کم	L	۱	1.6	X ≤ 2
۲	متوسط	M	۲	3.65	2 <X≤ 4
۳	زیاد	H	۳	6.05	X > 4

جدول ٦. معرفی پارامترهای متغیر ورودی تعداد مرورگر

ردیف	نام پارامتر	برچسب	شناسه	وزن میانگین	شاخص ورودی
۱	کم	L	۱	1.5	X ≤ 2
۲	متوسط	M	۲	3.51	2 <X≤ 4
۳	زیاد	H	۳	4.4	X > 4

جدول ۷. معرفی پارامترهای متغیر ورودی تعداد IP

ردیف	نام پارامتر	برچسب	شناسه	وزن میانگین	شاخص ورودی
۱	کم	L	۱	2.35	X ≤ 2
۲	متوسط	M	۲	3.7	2 <X≤ 4
۳	زیاد	H	۳	7.1	X > 4

جدول ۸. معرفی پارامترهای متغیر ورودی تعداد حواله اینترنتی

ردیف	نام پارامتر	برچسب	شناسه	وزن میانگین	شاخص ورودی
۱	کم	L	۱	3.2	X ≤ 3
۲	متوسط	M	۲	5.55	3 <X≤ 6
۳	زیاد	H	۳	9.5	6 <X≤ 10
۵	خیلی زیاد	VH	۴	15.8	X > 10

جدول ۹. معرفی پارامترهای متغیر ورودی مبلغ حواله اینترنتی

ردیف	نام پارامتر	برچسب	شناسه	وزن میانگین	شاخص ورودی(ریال)
۱	کم	L	۱	4650000	X ≤ 5000000
۲	متوسط	M	۲	7500000	5000000 <X≤ 10000000
۳	زیاد	H	۳	18000000	10000000 <X≤ 20000000
۵	خیلی زیاد	VH	۴	36000000	X > 20000000

۵-۳. تولید پایگاه قواعد رفتار

در مدل پژوهش، پایگاه دانشی[1] که از ترکیب دانش خبرگان امر به شکل اجتماع قواعد متغیرهای زبانی گردآوری شده است، از مهم‌ترین بخش‌های سیستم به‌شمار می‌آید. هدف

1. knowledge base

اصلی، به‌دست آوردن قواعد مفید، از تمام قواعدی است که هر یک توانایی تولید خروجی رفتار مورد نظر را دارند. در پایگاه دانش سامانه ارائه‌شده، با به‌کارگیری و ترکیب هفت متغیر ورودی به صورت (اگر- آنگاه) درکل، تعداد ۱۲۸ قاعده تولید گردید تا به عنوان مهم‌ترین ابزار مدل‌سازی جهت شناسایی نوع رفتار کاربر مورد استفاده قرار گیرد. خلاصه‌ای از این قواعد در جدول ۱۰ و نتایج آن روی نمونه آموزشی که ۱۰۰،۲۷۰ رکورد اعمال شده، در جدول ۱۱ ارائه شده است.

جدول ۱۰. خلاصه‌ای از پایگاه دانش استنتاجی حاصل از نظر خبرگان

NO	Err_Cnt	Login_Cnt	ISP_Cnt	Browser_Cnt	IP_Cnt	Ft_Cnt	Ft_Amnt	Result
۱	Z	L	L	L	L	L	L	NS
۲	Z	M	L	L	L	M	L	NS
۳	Z	L	L	L	L	H	L	NS
۴	Z	L	L	L	L	L	H	NS
۵	Z	M	L	M	L	L	M	LS
۶	Z	H	L	L	L	L	M	LS
۷	Z	H	L	L	M	L	M	LS
۸	Z	H	L	L	L	VH	L	LS
۹	Z	L	L	L	L	VH	L	S
۱۰	Z	H	M	L	H	L	M	S
۱۱	L	M	L	L	L	VH	L	S
۱۲	L	L	L	L	L	VH	L	S
۱۳	L	M	H	L	H	M	L	VS
۱۴	N	M	L	L	L	L	L	VS
۱۵	VH	L	H	H	L	H	H	VS
۱۶	VH	L	M	H	L	H	H	VS
۱۷	VH	L	H	M	H	VH	VH	F
۱۸	H	H	H	M	H	H	H	F
۱۹	VH	H	H	H	H	VH	VH	F
۲۰	VH	H	H	H	L	H	H	F

جدول ۱۱. نتایج حاصل از قواعد استنتاجی در SQL Server 2005

نوع رفتار	نتیجه جستجو	درصد سهمیه
عادی (NS)	99872	٪ 99.87
کمی مشکوک (LS)	118	٪ 0.11
مشکوک (S)	262	٪ 0.26
خیلی مشکوک (VS)	16	٪ 0.01
تقلب (F)	3	٪ 0.003

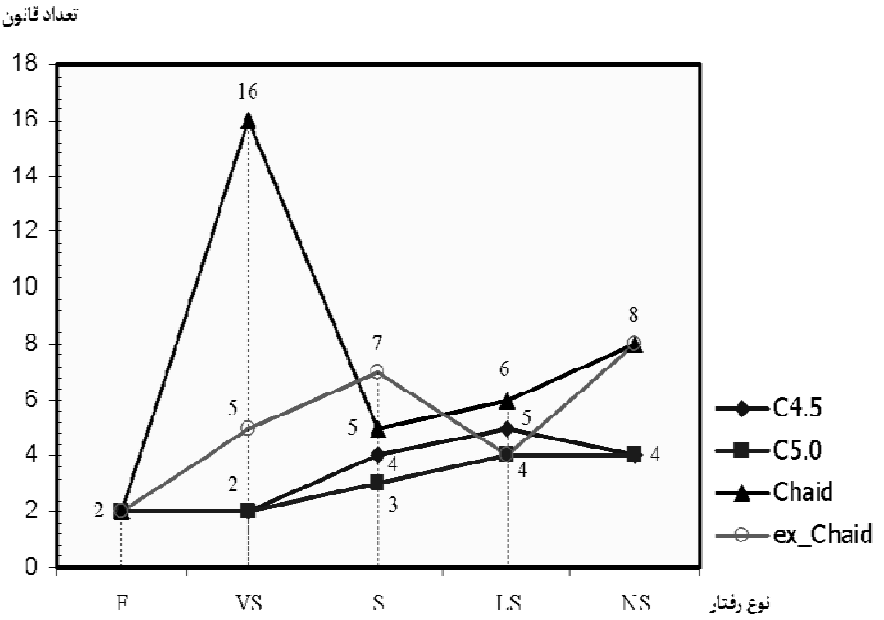
٦. انتخاب ابزار داده‌کاوی

جهت پیشبرد این پژوهش، نرم‌افزارهای منبع باز[1] در اولویت بوده است و ابزار تحلیلی داده‌کاوی SPSS_Clementine Client 11.1 را که از بیشتر الگوریتم‌های فنّ درخت تصمیم پشتیبانی می‌کند، جهت تحلیل درخت و تولید قوانین رفتاری استفاده شد (LBM n.d).

۷. ارزیابی سامانه

با توجه به اینکه هر کدام از الگوریتم‌های Chaid، ex_Chaid، C4.5 و C5.0 از هرس درخت جهت حذف شاخه‌های تکراری و تولید بهینه‌ترین قوانین استفاده می‌کنند، با این حال از طریق داده‌های آموزشی تا ۱۰ مرحله کیفیت طبقه‌بندی را مورد سنجش قرار می‌دهد تا زیردرخت منطقی تولید شود. اما در روند هر الگوریتم، تعداد قوانین لازم جهت کشف هر رفتار متفاوت بوده است و گاهی با خلاصه نمودن برخی قوانین استنتاجی، در کوتاه‌ترین مسیر و کمترین قانون، تمام رفتارهای مجموعه یک کلاس شناسایی شدند که این مقایسه در نمودار شکل ۲ مشهود است. همچنین، در افزایش اطمینان به نتایج سامانه، دسته‌بندی میزان تشخیص غلط در محدوده قبل و بعد و تحلیل خطر در مرز تعیین نوع رفتار مهم است. بنابراین، جهت برآورد دقت نهایی سامانه، علاوه بر اینکه ماتریس محاسبه هزینه غلط، از ۷۰ رکورد آزمایشی نیز به عنوان منبع مقایسه و مورد قبول خبرگان بانکداری اینترنتی استفاده گردید و خروجی الگوریتم، به همراه سناریوی کامل اجرای تحلیل درخت جهت بررسی دقت نهایی سامانه، به تأیید خبرگان اینترنتی درآمد.

1. open source



شکل ۲. مقایسه تعداد قوانین تولیدشده توسط الگوریتم‌ها جهت کشف رفتار

۸. نتایج پژوهش

با توجه به نوع پژوهش و برآورد تحلیل خطر، هریک از فنون داده‌کاوی و حتی الگوریتم‌های درخت تصمیم، با در نظر گرفتن نوع متغیرهای دخیل در خروجی سامانه، از اطمینان و خروجی متفاوتی برخوردار است. در این پژوهش نیز با توجه به ماهیت دسته‌بندی و کشف رفتار، خروجی چهار الگوریتم مطرح، به همراه سناریوی کامل اجرای آن در نرم‌افزار کلمنتاین[1] مورد مطالعه و مقایسه قرار گرفت که در نهایت، خروجی سامانه جهت بررسی و تعیین میزان دقت آن، در اختیار خبرگان اینترنتی بانک قرار گرفت. جدول ۱۲ حاکی از نتیجه مقایسه و میزان دقت و قدرت چهار الگوریتم مطرح در کشف و شناسایی رفتار مشکوک است.

1. Clementine

جدول ۱۲. مقایسه نتایج دقت الگوریتم‌های درخت تصمیم با داده‌های آموزشی و آزمایشی

	مقیاس الگـــــوریتم	C4.5	C5.0	Chaid	Ex_Chaid
شرایط عمومی	نوع داده ورودی	گسسته و پیوسته	گسسته و پیوسته	گسسته و پیوسته	گسسته و پیوسته
	تعداد شاخه‌ها در گره	۲ یا بیشتر	۲ یا بیشتر	۲ یا بیشتر	۲ یا بیشتر
	روش هرس درخت	نرخ خطای تخمین	نرخ خطای تخمین	بدون هرس	بدون هرس
	معیار جداسازی	نرخ منفعت	نرخ منفعت	تست مربع کای	تست مربع کای
	قابلیت خروجی قواعد	می‌دهد	می‌دهد	نمی‌دهد	نمی‌دهد
	تعیین منفعت (Gain)	نمی‌دهد	نمی‌دهد	می‌دهد	می‌دهد
	نمایش گرافیکی درخت	دارد	دارد	دارد	دارد
داده‌های آموزشی	تعریف تابع هزینه	بله	بله	بله	بله
	میزان دقت اتوماتیک	۹۹.۷۰	۹۹.۹۲	۹۹.۹۹	۹۹.۹۹
	تعداد سطوح درخت	۸	۸	۷	۷
	تعداد قواعد اصلی	۱۵	۱۴	۲۵	۲۶
	متغیر شروع در پیمایش درخت	Err_cnt	Err_cnt	Err_cnt	Err_cnt
	متغیر هزارد	Browser+IP	Browser+IP	Browser_cnt	Browser_cnt
	متغیرهای مؤثر در کشف تقلب	Err+Login+ft	Err+Login+ft	Err+Log+Amnt	Err+Log+Amnt
داده‌های آزمایشی	تعداد رکوردهای آزمایشی	۷۰	۷۰	۷۰	۷۰
	تعداد مغایرت در کشف خبرگان	۸	۷	۸	۷
	تعداد قوانین جهت کشف F	۲	۲	۲	۲
	تعداد قوانین جهت کشف VS	۲	۲	۱۶	۵
	تعداد قوانین جهت کشف S	۴	۳	۵	۷
	تعداد قوانین جهت کشف LS	۵	۴	۶	۴
	تعداد قوانین جهت کشف NS	۴	۴	۸	۸
	تعداد کل قوانین جهت کشف	۱۷	۱۵	۲۶	۲۶
	سرعت عملیات پیمایش کشف	۷۴ t	۶۸ t	۸۹ t	۷۱ t
	درصد دقت نهایی کشف	٪۸۹	٪۹۱	٪۸۹	٪۹۰

جمع‌بندی تحلیل‌های اشاره‌شده حاکی از آن است که در این پژوهش، از بین مجموع درخت‌های به‌دست آمده از چهار الگوریتم مطرح، فقط الگوریتم C5.0 به عنوان یک روش کلاسیک توانسته با ۹۱ درصد، بیشترین دقت را از طریق تعریف ماتریس محاسبه غلط و هرس زیردرخت‌های بی‌اهمیت کسب کند و نسبت به روش‌های مشابه با تعداد قواعد کمتری تمام رفتارهای مشکوک ممکن را تحت پوشش شناسایی قرار دهد.

۹. پیشنهادها

۹-۱. با توجه به اینکه عدم قطعیت و تعداد پارامترهای کیفی در این نوع پژوهش بالاست، استفاده از استدلال‌های ریاضی و شبکه بیزین جهت دسته‌بندی رفتار دقت بالاتری دارد.

۹-۲. اگر چنانچه رفتار هر مشتری به‌طور جداگانه در یک بازه زمانی طولانی مورد تجزیه و تحلیل قرار گیرد، رفتارهای ثبت‌شده به عنوان اهلیت هر مشتری قابل مدیریت است.

۹-۳. با گسترش سامانه مالی اینترنتی و ردگیری تعداد حواله انجام‌شده در تراکنش‌ها می‌توان سامانه‌ای مدیریتی جهت پیگیری انتقال پول در حساب‌ها جهت پول‌شویی طراحی نمود.

۹-۴. یکی از موارد کاربردی سامانه‌های کشف تقلب، شناسایی برخط تراکنش و جلوگیری از ارتکاب تقلب است. بنابراین با فراهم نمودن بستر پردازشی سامانه و مرتبط نمودن سامانه به یک سامانه خبره می‌توان از ارتکاب تقلب در محیط برخط جلوگیری نمود.

۹-۵. اگر چنانچه از دیگر پارامترهای شخصیتی مشتری استفاده شود، می‌توان با استفاده از فنّ درخت تصمیم و تلفیق آن با الگوریتم‌های خوشه‌بندی، تحلیل خطر کرد و تمام مشتریان را به دو گروه خوش‌حساب و بدحساب تقسیم کرد.

۹-۶. با توجه به اینکه بیشتر مشتریان از کاناهای حساب‌های متمرکز همچون خودپرداز (ATM) و کارت‌خوان (POS) نیز استفاده می‌کنند، بنابراین تلفیق رفتارهای حاصل از ثبت تراکنش از این سه کانال نیز پایگاه دانش مفیدی خواهد بود که به دقت بالاتری در رفتار مشتری و کشف رفتارهای مشتری می‌انجامد.

۱۰. نتیجه‌گیری

اثرات منفی تقلب، ضرورت طراحی سامانه‌های کشف تقلب را توجیه‌پذیرتر ساخته است. پیشرفت‌های فناورانه اخیر باعث تسهیل در توسعه سامانه‌های اطلاعاتی کشف تقلب شده است. فنون کشف تقلب شامل فنون جستجوی پیچیده‌ای است که از طریق بررسی تراکنش‌های

حساب مشتریان و رفتار مصرفی مشتری، الگوهای تقلب را کشف و به‌موقع اعلام می‌نماید. بانک‌ها با شناسایی فرایندهایی که احتمال انجام تقلب در آنها وجود دارد و ارتباط آن با سامانه‌های ناظر اطلاعات مدیریت، می‌توانند به نحوه مناسب‌تری سامانه‌های هشداردهنده و کشف تقلب را در این فرایندها به‌کار گیرند.

در این پژوهش نیز با توجه به آمار بالای استفاده کاربران از بانکداری اینترنتی، ابتدا با پیش‌پردازش اولیه پایگاه داده بانک، متغیرهای تأثیرگذار که هرکدام معرف نوعی از رفتار مشتری اینترنتی است، استخراج گردید. خروجی این پژوهش با وجود عدم شباهت پارامترهای انتخابی، متغیرهای مؤثر و قوانین موجود در نظام بانکی ایران با دیگر بانک‌های غیراسلامی، نشان از مطابقت و دقت کشف این الگوریتم‌ها روی تراکنش‌های بانکی دارد. بنابراین، سوابق دیگر پژوهش‌ها و این پژوهش روی یکی از بانک‌های بومی نشان می‌دهد که می‌توان از روش‌های هوشمند ماشینی جهت تسهیل و افزایش دقت کشف دانش استفاده نمود. مهم‌ترین مزیت این سیستم نسبت به روش‌های به‌کاررفته در سایر مقالات، امکان مدل‌سازی رفتار کاربران در پنج دسته مختلف است که با دقت بیشتری نوع رفتار کاربر را پیش‌بینی می‌کند. همچنین، در جهت عملیاتی نمودن این پژوهش می‌توان با طراحی و تلفیق روش دسته‌بندی درخت و نظام خبره فازی، یک سامانه خدمت‌گرا طراحی کرد تا شناسایی رفتارهای مشکوک در سامانه‌های تحت وب به‌صورت برخط نیز امکان‌پذیر باشد.

این مقاله با استفاده از اعتبارات و حمایت دانشگاه پیام نور انجام شده است.

۱۱. منابع

ف، حسنی. س، سلطانی. و ف ضرابیه. ۱۳۸۷. *مدیریت بانکداری الکترونیک*. تهران: انتشارات سبزان.

فرد.م .۱۳۸۴. *خدمات بانکداری الکترونیک و نیازهای اجرایی آن در مقایسه تطبیقی هزینه‌های عملیاتی خدمات مختلف بانکی*. تهران: پژهشکده پولی و بانکی.

Bignell, K. B. 2006. *Authentication in an internet banking environment; towards developing a strategy for fraud detection*. Internet Surveillance and Protection, 2006. ICISP "6 International Conference on Cote dazur. IEEEE explore.

Chau,D.H and Faloutsos,C. 2005. Fraud detection in electronic auction. Human Computer Interaction Institute, School of Computer Science.

Dorrington, P.2003. Detecting and eliminating the risk of Fraud. SAS white paper.

Ernst and Young .2003., Fraud: the unmanaged risk --- 8th global survey,http://www.ey.com/global/content.nsf/International/Home. http://www.spss.com/clementine/

IBM. SPSS Software. Access at http: //www.SPSS.com/clementine (accessed 10 March 2012).

Kelly, C. 2003. *Electronic government strategics* , Meta Group, advisory Service.

KPMG.2003. Fraud Survey of 2003. http://www.kpmg.com

Mitchell, Tom M. 1997. Machine Learning 21 Jan. 2013 McGraw-Hill.

Online banking. 2005. A pew internet project data memo. http://www.pewinternet.org/PPF/r/149/report_display.asp.

Phua, C. W. C. 2003. Investigative data mining in fraud detection. Schol of Business Systems. Moash University.

Quah.Jon T.S, Sriganesh. M .2007. Real-time credit card fraud detection using computational intelligence, Expert Systems with Applications. doi:10.1016/j.eswa.2007.08.093.

Shah, M. H., Branganza, A., Khan, S. & Xu, M. 2005. A survey of critical success factors in e-banking. European and mediterrance conference on information systems.

Statistics for general and online fraud, ePaynews.com. http://www.epaynews.com/statistics/fraud.html#2, Jan 2007.

UNCTAD. 2009. E-commerce and Development Report, Geneva and New York: United nations Publication. www.unctad.org/e-commerce

Yufeng .Kou, Lu. Chang-Tien, Sirirat Sirwongwattana. 2002. Survey of Fraud detection techniques, Virginia Polytechnic Institute and State University.

Young.T. 2008. Lords call for e-crime shakeup

Waite,K.Harrison,T. 2004. Online banking information : What we want and what we get. *Qualitative Market Research. An International Journal* 7 (1): 67-79

Introducing a Model for Suspicious Behaviors Detection in Electronic Banking by Using Decision Tree Algorithms

Rohulla Kosari Langari[1]
MS in Management of Information Technology

Nasrolla Moghaddam[2]
Computer Engineering PhD

Davood Vahdat*
IT Engineering PhD Student

Iranian Journal of Information Processing & Management

Iranian Research Institute
For Science and Technology
ISSN 2251-8223
eISSN 2251-8231
Indexed in LISA, SCOPUS & ISC
Vol.28 | No.3 | pp: 681-700
Spring 2013

Abstract: Change the world through information technology and Internet development, has created competitive knowledge in the field of electronic commerce, lead to increasing in competitive potential among organizations. In this condition The increasing rate of commercial deals developing guaranteed with speed and light quality is due to provide dynamic system of electronic banking until by using modern technology to facilitate electronic business process. Internet banking is enumerate as a potential opportunity the fundamental pillars and determinates of e-banking that in cyber space has been faced with various obstacles and threats. One of this challenge is complete uncertainty in security guarantee of financial transactions also exist of suspicious and unusual behavior with mail fraud for financial abuse. Now various systems because of intelligence mechanical methods and data mining technique has been designed for fraud detection in users' behaviors and applied in various industrial such as insurance, medicine and banking. Main of article has been recognizing of unusual users behaviors in e-banking system. Therefore, detection behavior user and categories of emerged patterns to paper the conditions for predicting unauthorized penetration and detection of suspicious behavior. Since detection behavior user in internet system has been uncertainty and records of transactions can be useful to understand these movement and therefore among machine method, decision tree technique is considered common tool for classification and prediction, therefore in this research at first has determinate banking effective variable and weight of everything in internet behaviors production and in continuation combining of various behaviors manner draw out such as the model of inductive rules to provide ability recognizing of different behaviors. At least trend of four algorithm Chaid, ex_Chaid, C4.5, C5.0 has compared and evaluated for classification and detection of exist models on the real transactions of private bank.

Keywords: electronic banking, internet banking, security, suspicious behaviors, decision tree

1. kosari58@yahoo.com 2. charkari@modares.ac.ir
*Corresponding author: vahdat@pnu.ac.ir