دوفصل‌نامه علمی منادی امنیت فضای تولید و تبادل اطلاعات (افتا)
سال ۱۴۰۱، جلد ۲۰، شماره ۲ (۱۱-۲۱)
https://monadi.isc.org.ir

ارائه‌شده در نوزدهمین کنفرانس بین‌المللی انجمن رمز ایران در دانشگاه گیلان، ۹ و ۱۰ شهریور ۱۴۰۱

تشخیص پولشویی الکترونیکی در تراکنش‌های تامین‌کنندگان خدمات پرداخت★

علی نظری* و بابک صادقیان

دانشکده مهندسی کامپیوتر، دانشگاه صنعتی امیر کبیر، تهران، ایران

اطلاعات مقاله

کلمات کلیدی:
پولشویی الکترونیکی
پول غیر قانونی
تراکنش نامتعارف
ماشین بولتزمن محدود
استدلال مبتنی بر مورد
پی‌جویی جرم

dor: 20.1001.1.24763047.1401.11.2.2.1

نوع مقاله: پژوهشی

چکیده

مجرمان پولشویی در پوشش کسب و کارهای قانونی با سوء استفاده از خدمات شرکت‌های ارائه کننده خدمات پرداخت (PSP) اقدام به پولشویی الکترونیکی می‌نمایند. به منظور فورنسیک پولشویی در تراکنش‌های مالی شرکت‌های PSP، روشی توسط حجتی و همکاران ارائه شده است که از طریق تشخیص تراکنش‌های خارج از الگوی صنف فروشندگان و با روش تحلیل درون گروهی انجام می‌گردد. بررسی‌های ما نشان می‌دهد که استفاده از روش مطرح شده در تشخیص تراکنش‌های خارج از الگوی صنف، نرخ اعلام‌های مثبت نادرست حدوداً ۱۳٪ را در تشخیص پولشویی نتیجه می‌دهد. ما در این مقاله با اصلاح راه‌حل مطرح شده، نرخ اعلام‌های مثبت نادرست را ۱۲٪ درصد کاهش داده و به کمتر از ۱٪ رساندیم. برای این منظور در تحلیل درون گروهی، حجم تراکنش‌های مالی فروشندگان را در کنار حجم بازدیدکنندگان وب‌سایت‌های آنها مورد تحلیل قرار دادیم و بر اساس تعداد بازدیدکنندگان وب‌سایت‌های صنف مربوطه، حجم تراکنش‌های هر فروشنده را تخمین زدیم و حجم فروش بیش از تخمین را نامتعارف در نظر گرفتیم. با استفاده از ماشین بولتزمن محدود دقت تشخیص تراکنش‌های خارج از الگوی صنف را ارتقاء دادیم و با کمک استدلال مبتنی بر مورد، نرخ اعلام‌های منفی نادرست را کاهش دادیم. سیستم پیشنهادی ما، از یک پنجره لغزان چهارهفته‌ای برای تشخیص برخط پولشویی استفاده می‌نماید. نتایج ارزیابی‌ها نشان داد که راه‌حل پیشنهادی ما دارای دقت تشخیص ۹۹٪ می‌باشد.

۱ مقدمه

توسعه فنآوری یک شمشیر دو طرفه است. از یک سو باعث بهبود سطح زندگی انسان می‌شود، از طرف دیگر امکان تخلف را برای مجرمان فراهم می‌نماید. ظهور پول الکترونیکی [1] و پرداخت اینترنتی [2]، فضای بیشتری برای مجرمان پولشویی فراهم کرده است. سیستم پرداخت الکترونیکی قادر است پول زیادی را به سرعت در مسافت‌های طولانی جابجا نماید. اتخاذ تکنیک‌های رمزگذاری و امکانات انتقال از راه دور به طور خارق‌العاده‌ای ویژگی بی‌نامی [3] پول را افزایش داده است که مطلوب پولشویان است [۷]. از آنجا که پولشویان پس از تطهیر پول کثیف [4]، با درآمدهای حاصل از آن، چرخه اقتصاد کشور را تحت تأثیر قرار می‌دهند، عملاً مانعی در راه تحقق فضای رقابتی کشور محسوب گردیده و منبع مالی عمده برای انجام فعالیت‌های «سیاه و پنهان» بشمار می‌روند. بر اساس ارزیابی‌های منتشر شده توسط مؤسسه بازل که به صورت سالانه گزارش میزان خطر پولشویی کشورهای جهان را منتشر می‌کند ایران هرساله در صدر کشورهای دارای ریسک پولشویی بالا قرار گرفته است [۶]. از سال ۲۰۱۸ در پی قطع

★از کمیته علمی نوزدهمین کنفرانس بین‌المللی انجمن رمز ایران برای داوری این مقاله تشکر می‌شود.

*نویسنده مسئول
آدرس‌های رایانامه: nazariali@aut.ac.ir (علی نظری)، basadegh@aut.ac.ir (بابک صادقیان)

[1] E-money [2] electronic payment

[3] anonymity [4] dirty money

همکاری طرفین، آمار مربوط به ایران از گزارش‌های این مؤسسه حذف گردیده است.

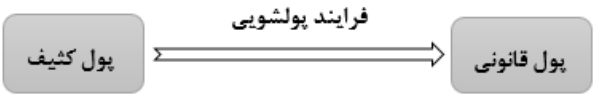
شرکت‌های تامین کننده خدمات پرداخت (PSP) متولی انجام تراکنش‌های بین بانکی ارجاع شده از وب‌سایت‌های تجارت الکترونیکی[۱] هستند. مجرمان با راه‌اندازی وب‌سایت‌های پوششی[۲] اقدام به پولشویی الکترونیکی از طریق خدمات پرداخت این شرکت‌ها می‌نمایند. پژوهشگران حوزه امنیت پرداخت الکترونکی با بهره‌گیری از روش‌های داده‌کاوی و یادگیری ماشین در پی شناسایی تراکنش‌های غیرقانونی هستند. تحلیل درون گروهی (PGA[۳]) روشی برای تشخیص انحراف رفتار[۴] یک فروشنده در مقایسه با رفتار فروشندگان همتا است. رفتار فروشندگان صنف مبنای رفتار قانونی است و انحراف از الگوی صنف ممکن است بیانگر رفتار غیر قانونی باشد. در پژوهش [۱] روشی برای تشخیص تراکنش‌های خارج از الگوی صنف ارائه شده است که با توجه به فاصله اقلیدسی مبالغ تراکنش‌های هر فروشنده از فروشندگان همتا، انحراف از رفتار صنف تشخیص داده می‌شود. در راه‌حل پیشنهادی، به همسایگان نزدیک فروشنده وزن بیشتری داده می‌شود تا انحراف فروشنده از الگوی صنف با توجه به همسایگان نزدیک معین گردد. از این طریق اثر منفی همسایگان دوری که احتمالا دارای برچسب شغلی اشتباه هستند کاهش داده می‌شود.

ما راه‌حل مطرح شده را از دو جهت ارتقاء دادیم. اولاً همسایگان نزدیک هر فروشنده را با اولویت نزدیکی تعداد بازدیدکنندگان وب‌سایت‌ها و سپس نزدیکی حجم تراکنش‌های مالی مد نظر قرار دادیم که باعث شد همسایه‌های نزدیک را دقیق‌تر معین نمائیم و از این طریق نرخ اعلام‌های مثبت نادرست را کاهش دهیم. ثانیاً روش آماری محاسبه انحراف فروشنده از الگوی صنف را که صرفا با توجه به مبالغ فروش و عدم لحاظ تعداد تراکنش‌ها انجام می‌شد را به روش یادگیری عمیق رفتار تراکنشی صنف فروشندگان توسط ماشین بولتزمن محدود تغییر دادیم که علاوه بر حجم مبالغ فروش از تعداد تراکنش‌ها نیز استفاده می نماید. اگرچه تعداد تراکنش‌ها به تنهایی معیار مشابهت نیست و به میزان شهرت فروشندگان و فعالیت‌های بازاریابی آنها بستگی دارد اما در ترکیب با جمع مبالغ فروش معیاری برای تشخیص رفتار خارج از الگوی صنف محسوب می‌شود.

برای ارتقای دامنه تشخیص پولشویی، از روش استدلال مبتنی بر مورد (CBR) استفاده نمودیم که با رویکرد فرامحلی، ضمن مقایسه بردار رفتار فروشنده هدف با بردار رفتار همه فروشندگان دارای سابقه پولشویی، موارد جدید مشکوک به پولشویی را تشخیص می‌دهد. سیستم پیشنهادی ما دارای قابلیت تشخیص برخط پولشویی الکترونیکی است و از پنجره لغزان چهار هفته‌ای برای تحلیل حجم نامتعارف تراکنش‌ها بهره برده است. این پنجره در هر هفته، با تراکنش‌های هفته جدید بروز گردیده و تراکنش‌های هفته قدیمی‌تر حذف می‌شود.

به منظور توصیف روش پیشنهادی، در بخش ۲ مقاله مفاهیم اولیه مورد

[1]E-commerce website [2]shell website [3]Peer Group Analysis [4]behavior deviation



شکل ۱. مفهوم پولشویی

نیاز مانند مفهوم پولشویی، ماشین بولتزمن محدود و روش استدلال مبتنی بر مورد را توضیح می‌دهیم. سپس، در بخش ۳ مقاله برای درک جایگاه و ارزش دستاوردهای این مقاله، به سابقه پژوهش در این حوزه پرداخته و معرفی مختصری از سابقه روش‌های بکار رفته برای تشخیص رفتار غیرقانونی و نیز توصیف روش حجتی و همکاران می‌پردازیم. در بخش ۴، به توضیح روش پیشنهادی خود و در بخش ۵ به ارزیابی این روش با استفاده از دادگان واقعی می‌پردازیم. در انتها و در بخش ۶ جمع‌بندی مطالب ارائه شده، تقدیم می‌گردد.

۲ مفاهیم اولیه

در این قسمت مفاهیم اولیه لازم برای فهم بهتر مطالب مقاله بیان شده است.

۱.۲ پولشویی

پولشویی فرایند تبدیل پول غیر قانونی یا کثیفِ حاصل از فعالیت مجرمانه به پول قانونی و بهره‌مندی از مزایای آن توسط مجرمان پولشویی تعریف شده است [۶]. به شکل ۱ توجه نمائید.

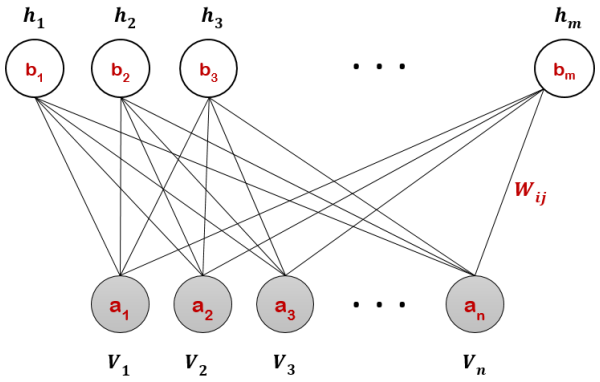
باید توجه داشت که پولشویی فرایندی پیچیده است. چرخه انجام آن طولانی بوده و شامل سه مرحله اصلی است [۱۰]:

- جایگذاری[۵]: واریز پول نقد کثیف در حساب‌های بانکی
- لایه‌گذاری[۶]: ایجاد لایه‌های پیچیده‌ای از تراکنش‌های مالی برای پنهان سازی رابطه بین پول‌های غیر قانونی و منبع جنایی آن
- یکپارچه‌سازی[۷]: پول تجمیع شده در تجارت قانونی سرمایه گذاری شده و سود آن برای فعالیت‌های مجرمانه به کار گرفته می‌شود تا چرخه جدیدی آغاز گردد.

پولشویی به روش‌های متنوعی انجام می‌گردد. مهمترین روش‌های انجام پولشویی به شرح زیر است:

- پولشویی از طریق قاچاق وجه نقد
- پولشویی از طریق مؤسسات بانکی
- پولشویی از طریق مؤسسات بیمه
- پولشویی از طریق معاملات ملکی و حراج اشیای قیمتی
- پولشویی از طریق تجارت بین‌المللی
- پولشویی از طریق شرکت‌های پوششی
- پولشویی از طریق مؤسسات خیریه

[5]placement [6]layering [7]integration



شکل ۲. گراف ماشین بولتزمن محدود

- پولشویی الکترونیکی

پولشویی الکترونیکی با سوء استفاده از خدمات پرداخت اینترنتی انجام می‌پذیرد که مهمترین روش‌های انجام آن به شرح زیر است:

- از طریق خدمات پرداخت برخط
- از طریق کارت‌های پیش پرداخت
- از طریق قمار اینترنتی
- از طریق حراج‌های اینترنتی
- از طریق پرداخت‌های تلفن همراه
- از طریق پول مجازی

۲.۲ ماشین بولتزمن محدود

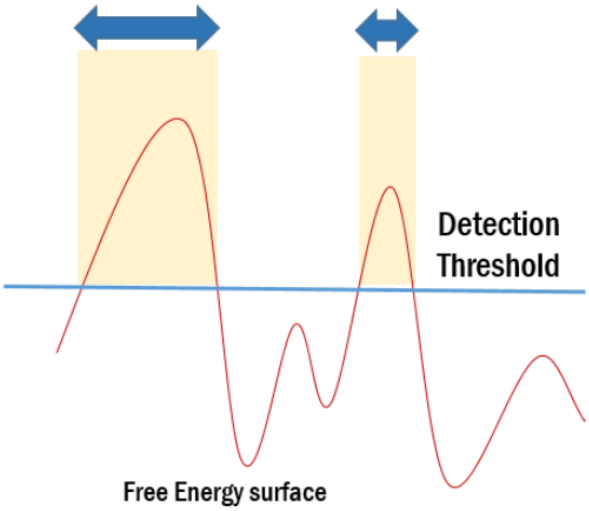
ماشین بولتزمن یک مدل گراف احتمالاتی[1] غیرجهت‌دار و در دسته شبکه‌های عصبی تصادفی قرار دارد. قدرت محاسباتی بالاتر و فرایند یادگیری سریع‌تر آن باعث شده است که برای حل مسایل یادگیری ماشین مورد توجه پژوهشگران قرار گیرد. نوع خاصی از این ماشین با عنوان ماشین بولتزمن محدود (RBM[2]) یک گراف دو بخشی است که یال‌های آن بین دو لایه آشکار[3] و پنهان قرار دارد و هر گره لایه آشکار به همه‌ی گره‌های لایه پنهان[4] متصل می‌شود و هیچ دو گره از یک لایه یکسان با یکدیگر مرتبط نیستند [۸]. گراف شکل ۲ ماشین بولتزمن محدودی را نمایش می‌دهد که دارای n گره آشکار $\{v_1,\dots,v_n\}$ و m گره پنهان $\{h_1,\dots,h_m\}$ بوده و با توجه به مقادیر i و j در محدوده $i \in \{1,\dots,n\}$ و $j \in \{1,\dots,m\}$، وزن یال‌ها با W_{ij} و بایاس لایه آشکار با a_i و بایاس لایه پنهان با b_j نشان داده شده است.

عملکرد ماشین بولتزمن محدود بر اساس محاسبه انرژی است و مقدار انرژی آن با تابع ذکر شده در فرمول (۱) محاسبه می‌گردد.

$$E(v,h) = -\sum_{i\in visible} a_i v_i - \sum_{j\in hidden} b_j h_j - \sum_{i,j} v_i h_j w_{(i,j)} \qquad (۱)$$

ماشین RBM به هر زوج ممکن از بردار ورودی v (آشکار) و بردار پنهان h مطابق با فرمول (۲) یک مقدار احتمال منتسب می‌نماید:

[1]probabilistic graphical model [2]Restricted Boltzmann Machine [3]visible layer [4]hidden layer



شکل ۳. حد آستانه سطح انرژی آزاد

$$p(v,h) = \frac{1}{Z} e^{-E(v,h)} \qquad (۲)$$

Z ضریب نرمال‌سازی بوده و با استفاده از فرمول (۳) محاسبه می‌گردد:

$$Z = \sum_{v,h} e^{-E(v,h)} \qquad (۳)$$

مقدار احتمال هر بردار ورودی v از حاصل‌جمع احتمال‌های زوج بردار v با همه بردارهای پنهان h طبق فرمول (۴) محاسبه می‌شود:

$$p(v) = \frac{1}{Z} \sum_{h} e^{-E(v,h)} \qquad (۴)$$

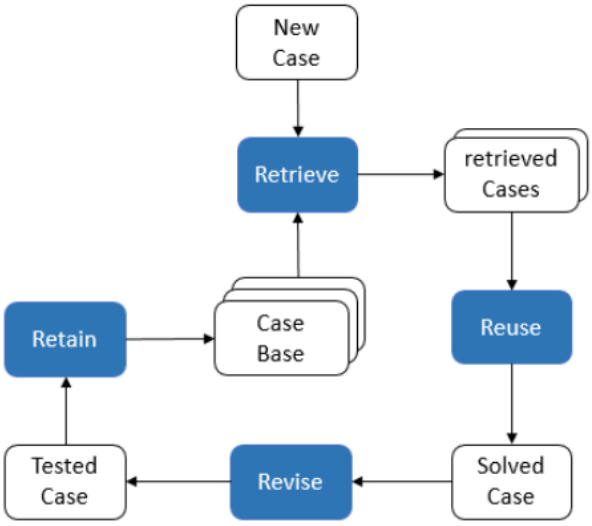
با کنترل میزان تغییرات وزن یال‌ها و مقادیر بایاس لایه‌های آشکار و پنهان می‌توان مقدار احتمال داده‌های ورودی را در فرایند آموزش مدل افزایش داد، در این صورت مقدار انرژی آزاد برای داده‌های اکثریت کاهش یافته و برای داده‌های اقلیت افزایش می‌یابد از این طریق می‌توان با محاسبه میزان انرژی آزاد ، موارد نامتعارف را با توجه به سطح انرژی آزاد تشخیص داد. سطح انرژی آزاد بردار ورودی v از فرمول (۵) محاسبه می‌گردد:

$$F(v) = -\ln\left(\sum_{h} e^{-E(v,h)}\right) \qquad (۵)$$

با توجه به شکل ۳ اگر سطح انرژی آزاد بردار ورودی از حد آستانه انرژی آزاد بیشتر گردد نامتعارف شناخته می‌شود.

تشخیص موارد نامتعارف توسط ماشین‌های بولتزمن محدود دارای امتیازات متعددی است که به شرح زیر است [۴].

- نیازی به برچسب دادگان آموزش ندارد.
- برای پردازش دادگان حجیم مناسب است.
- زمان پاسخ‌دهی آن خطی است.
- دقت و سرعت تشخیص بالاتری نسبت به سایر مدل‌های یادگیری عمیق غیر نظارت شده نظیر Auto-Encoder دارد.



شکل ۴. مراحل اجرای روش استدلال مبتنی بر مورد

- مناسب داده‌های چند بعدی و بردارهای چند متغیره است.

۳.۲ استدلال مبتنی بر مورد

روش استدلال مبتنی بر مورد (CBR [۱]) مطابق با رفتار انسان‌ها است که در مواجه با مسایل جدید از تجارب پیشین خود برای حل مسئله جدید بهره می‌برند. واژه «استدلال» به معنای نتیجه‌گیری از موارد ثبت شده پیشین است. واژه «مورد» به پرونده یا تجربه‌ای اطلاق می‌گردد که برای استخراج راه‌حل مسئله جدید به کار گرفته می‌شود [۹]. متدولوژی CBR دارای چهار مرحله بازیابی [۲]، استفاده مجدد [۳]، اصلاح [۴] و ذخیره پرونده [۵] جدید است که توالی اجرای آنها در شکل ۴ نمایش داده شده است.

در مرحله بازیابی، نزدیکترین پرونده به پرونده جدید با توجه به میزان مشابهت ویژگی‌های آنها انتخاب می‌گردد. حل مسئله فعلی مستلزم برخی اصلاحات و انطباق‌های بیشتر بر روی حل مسئله مشابه پیشین است تا برای حل مسئله جدید مناسب گردد. در مرحله استفاده مجدد، تغییراتی در پرونده منتخب اعمال می‌گردد تا بر پرونده جدید منطبق گردد. در مرحله اصلاح، راه‌حل پرونده مشابه با اصلاحاتی برای پرونده جدید بکار گرفته می‌شود و درستی آن مورد ارزیابی قرار می‌گیرد. در مرحله ثبت، پرونده جدید در پایگاه سوابق [۶] ذخیره می‌شود تا برای حل مسائل جدیدتر مورد استفاده قرار گیرد.

۳ سابقه پژوهش

رویکرد مقالاتی که در این بخش مورد توجه قرارگرفته‌اند به روش تشخیص درون گروهی رفتارهای غیرقانونی است. پیش فرض این روش، مشخص بودن گروه همتای فروشنده هدف است. در مقاله [۲] روشی برای دسته‌بندی صنف فروشندگان و تعیین اعضای گروه همتا با استفاده از کلمات کلیدی استخراج شده از وب‌سایت‌های آنها ارائه شده است. در پژوهش‌های متعددی نظیر [۱، ۱۱–۱۳] از روش تحلیل درون گروهی (PGA) برای تشخیص فریبکاری مالی استفاده شده است. در پژوهش [۱] محمد حسین حجتی و همکاران روشی برای تشخیص تراکنش‌های مالی خارج از الگوی صنف فروشندگان ارائه نموده‌اند. دقت این روش تحت الشعاع تراکنش‌های پرت فروشندگانی قرار می‌گیرد که دارای برچسب شغلی اشتباه هستند. برای کاهش اثر منفی برچسب‌های اشتباه از محاسبه میانگین وزنی میزان انحراف یک هدف نسبت به همسایگانش استفاده شده است. فاصله اقلیدسی نرمال شده همسایگان مبنای تشخیص همسایگان نزدیک است به همسایگانی که نزدیک به هدف قرار دارند وزن بیشتری نسبت به همسایگان دورتر داده شده است تا تأثیر مقادیر پرت کاسته شود. میانگین وزنی همسایگان از فرمول (۶) حساب می‌شود.

$$P_{ij} = \frac{\sum_{p \in PG_i(t_j)} W_{pj} * S_{pj}}{\sum_{p \in PG_i(t_j)} W_{pj}}, \quad j \geq ۱,\ p \neq i \qquad (۶)$$

$PG_i(t_j)$ گروه نظیر فروشنده iام در زمان jام است و S_{pj} میزان فروش همسایه pام در زمان j است. W_{pj} نیز وزن انحراف pامین همسایه نزدیک نسبت به هدف i در زمان j است که از فرمول (۷) محاسبه می‌گردد:

$$W_{pj} = ۱ - \frac{S_{pj} - \min(\text{dist}_{PG_i})}{\max(\text{dist}_{PG_i}) - \min(\text{dist}_{PG_i})}, \qquad (۷)$$
$$j \geq ۱,\ p \neq i$$

dist_{PG_i} آرایه‌ای از تفاضل قیمت‌های ارائه شده توسط همسایگان فروشنده iام است. در نهایت با داشتن میانگین وزنی P_{ij} و انحراف معیار V_{ij} که از فرمول (۸) محاسبه شده است؛ میزان انحراف نسبی قیمت عضو iام از قیمت‌های اعضای گروه همتا توسط فرمول (۹) محاسبه می‌گردد.

$$V_{ij} = \frac{\sum_{p \in PG_i(t_۱)} W_{pj}(S_{pj} - P_{ij})^۲}{\left(\sum_{p \in PG_i(t_j)} W_{pj}\right) - ۱}, \qquad (۸)$$
$$j \geq ۱,\ p \neq i$$

$$T_{ij} = \frac{S_{ij} - P_{ij}}{\sqrt{V_{ij}}} \qquad (۹)$$

S_{ij} خلاصه آماری میزان فروش فروشنده iام در زمان j می‌باشد.

در پژوهش‌های دیگری نظیر [۳، ۴] از ماشین بولتزمن محدود به عنوان یک روش یادگیری عمیق برای شناسایی فریبکاری مالی استفاده گردیده است. در مقاله [۳] ابتدا بردار رفتار مشتریان با توجه به ویژگی‌های تراکنش‌های آنها نظیر بازه زمانی تراکنش، نوع تراکنش، مبلغ تراکنش، شناسه واریزکننده، میزان موجودی حساب مبدأ پیش از انجام تراکنش، میزان موجودی حساب مبدأ پس از انجام تراکنش، شناسه دریافت کننده مبلغ، میزان موجودی حساب مقصد پیش از انجام تراکنش و میزان موجودی حساب مقصد پس از انجام تراکنش ساخته گردیده و پس از نرمال‌سازی به مدل‌های یادگیری عمیق Auto-Encoder و RBM اعمال

[1]Case-Based Reasoning [2]retrieve [3]reuse [4]revise [5]retain [6]case base

می‌گردد تا فریبکاری مالی مشتریان موبایل بانک شناسایی گردد. نتایج ارزیابی‌ها، دقت بالاتری برای مدل RBM نشان داده است.

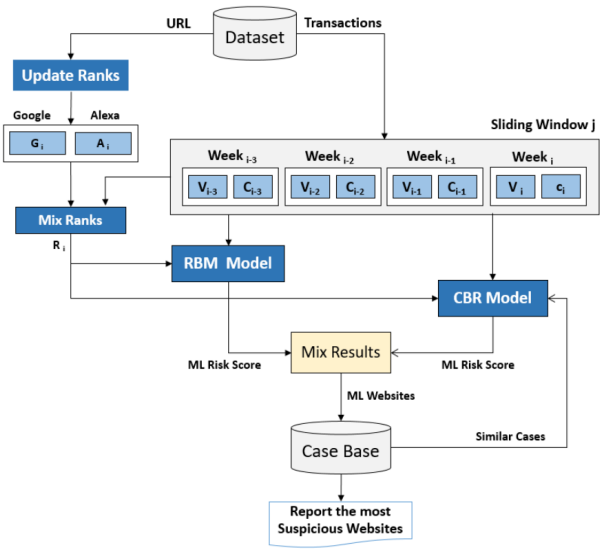
در مقاله [۴] نیز با استفاده از ماشین بولتزمن محدود و شبکه باور عمیق [۱] که از تعدادی ماشین بولتزمن محدود تشکیل می‌شود؛ روش جدیدی برای تشخیص موارد نامتعارف در بردار رفتار مشتریان با ویژگی‌های دارای انواع داده نامتجانس [۲] ارائه داده است. انواع داده نامتجانس، شامل مقادیر پیوسته مانند بازه زمانی تراکنش‌ها، مقادیر دودیی مانند درست و غلط، مقادیر گسسته مانند مبالغ تراکنش‌های مالی و مقادیر اسمی مانند جنسیت مشتریان بوده است. در مقاله [۵] از بردار تراکنش‌های مالی مشتریان در کنار شبکه گردش مالی بین حساب‌های مشتریان و حساب‌های فروشندگان برای تشخیص الگوهای پولشویی استفاده شده است. الگوهایی نظیر حجم مبالغ بیش از ۱۰۰ میلیون دلار در طول دوره ۱۰ روزه، وجود تعداد بالای تراکنش‌های با مبالغ کم، حساب‌های بانکی نیمه فعال و جرم خیز بودن ناحیه صدور تراکنش‌ها مبنای تشخیص موارد مشکوک به پولشویی قرار گرفته است.

۴ راه‌حل پیشنهادی

می‌دانیم که انجام پولشویی نیازمند زمان است و مجرمان برای اجتناب از حساس شدن سامانه‌های ضد پولشویی، پول‌های کثیف خود را در حجم کم و زمان طولانی‌تر به گردش در می‌آورند. بررسی‌های ما بیانگر آن است که کوتاه‌ترین دوره زمانی مفید برای تشخیص پولشویی نمی‌تواند کمتر از چهار هفته باشد و دقت لازم را نخواهد داشت. بنابراین ما از یک پنجره لغزان چهار هفته‌ای استفاده کرده‌ایم که با تجمیع تراکنش‌های هفته جدید، این پنجره یک هفته به جلوتر جابجا می‌گردد. بردار رفتار فروشندگان را نیز دو متغیره در نظر گرفته‌ایم که شامل تعداد و جمع مبالغ تراکنش‌های ارجاع شده به سوئیچ پرداخت در طی چهار هفته پنجره لغزان جاری است.

بررسی‌های ما نشان داد که حجم فروش وب‌سایت‌های تجارت الکترونیکی با تعداد بازدیدکنندگان آنها دارای رابطه همبستگی مستقیم غیر خطی است، در این راستا یکی از شواهد وقوع پولشویی [۳]، تشخیص نامتناسب بودن حجم فروش وب‌سایت با تعداد بازدیدکنندگان آن است. برای تعیین تعداد بازدیدکنندگان وب‌سایت‌ها از ترکیب رتبه‌های جستجوی گوگل [۴] و الکسای [۵] آنها استفاده نمودیم. با ترکیب وزن‌دار این دو رتبه تخمین دقیق‌تری از میزان بازدید وب‌سایت‌ها به دست آمد. معماری سیستم پیشنهادی ما در شکل ۵ نمایش داده شده است. ابتدا تراکنش‌های پنجره لغزان جاری را که پنجره jام می‌نامیم از مجموعه دادگان در دسترس استخراج نموده و با توجه به آدرس دامنه وب‌سایت‌ها، رتبه‌های گوگل و الکسای آنها را بازیابی نمودیم، سپس با اجرای موازی و مستقل مدل‌های RBM و CBR بر روی تراکنش‌های پنجره لغزان جاری، موارد مشکوک به پولشویی را تشخیص داده و نتایج دو مدل را ترکیب

[1]DBN [2]heterogeneous data type [3]money laundering evidence [4]Google page rank [5]Alexa rank



شکل ۵. معماری سیستم پیشنهادی تشخیص پولشویی

نمودیم و نمرات خطر پولشویی [۶] وب‌سایت‌های مشکوک پنجره لغزان جاری را محاسبه کردیم و نتایج را در پایگاه سوابق پولشویی ذخیره نمودیم.

برای استخراج و گزارش وب‌سایت‌های با بالاترین خطر پولشویی، با مراجعه مجدد به پایگاه سوابق پولشویی، مجموع وزن‌دار نمرات خطر پولشویی همه وب‌سایت‌ها را محاسبه نمودیم. به این ترتیب که وزن نمرات خطر محاسبه شده در پنجره‌های لغزان جدیدتر را بالاتر در نظر گرفتیم تا سوابق پولشویی جدیدتر وزن بیشتری در محاسبه مجموع داشته باشند. در نهایت با مرتب‌سازی نزولی وب‌سایت‌ها بر اساس جمع نمرات خطر، وب‌سایت‌های با بالاترین خطر پولشویی را تشخیص دادیم.

دلیل استفاده از مدل تشخیص CBR آن است که این مدل مواردی مشکوکی را تشخیص می‌دهد که مدل RBM به دلیل رویکرد محلی قادر به تشخیص آنها نبوده است. در سیستم پیشنهادی، ماشین بولتزمن محدود دارای رویکرد پنجره محور است و موارد نامتعارف تراکنش‌ها را با تحلیل درون گروهی تراکنش‌های پنجره لغزان جاری تشخیص می‌دهد. بنابراین در مواردی که تراکنش‌های یک وب‌سایت در مقایسه با تراکنش‌های پنجره لغزان جاری متعارف محسوب شده اما در مقایسه با تراکنش‌های پنجره‌های قبلی نامتعارف باشد، قادر به تشخیص درست نیست. از آنجا که مدل CBR تراکنش‌های پنجره لغزان جاری را با تراکنش‌های پولشویی پنجره‌های لغزان قبلی مقایسه می‌نماید، دارای رویکرد فرامحلی است.

۱.۴ تخمین رتبه بازدید وب‌سایت‌ها

شرکت‌های گوگل و الکسا از دیدگاه‌های متفاوتی رتبه محبوبیت وب‌سایت‌ها را محاسبه می‌نمایند. رتبه شرکت گوگل برای وب‌سایت‌ها عددی در بازه صفر تا ده می‌باشد که مقدار کمتر نشان‌دهنده رتبه پایین‌تر است. ایده مدیران گوگل این بوده است که صفحات با لینک ورودی زیاد صفحات مهمتر و با اهمیت‌تر هستند. برای پیشگیری از فریبکاری

[6]money laundering risk score

افزایش دروغین رتبه وب‌سایت‌ها از الگوریتم رتبه‌بندی شخصی شده[1] استفاده گردیده است که اعتبار هر وب‌سایت را با توجه به پیوندهای دریافت شده از وب‌سایت‌های معتبر تشخیص می‌دهد. شرکت الکسا برای سال‌های متمادی وب‌سایت‌های فعال در شبکه اینترنت را با توجه به ترافیک ارائه شده توسط «پانل داده جهانی» رتبه‌بندی نموده است که از بازدیدهای مرورگرهایی نظیر IE، Firefox، chrome جمع‌آوری می‌گردد. مبنای الکسا برای محاسبه رتبه وب‌سایت‌ها، ترکیبی از میانگین تخمینی بازدیدکنندگان روزانه و تعداد تخمینی بازدید صفحات آن در طی سه ماه اخیر است.

برای ترکیب دو رتبه گوگل و الکسا، ابتدا مقادیر آنها را بین ۰ و ۱ نرمال‌سازی نموده و سپس میانگین وزنی آنها را به عنوان رتبه ترکیبی در نظر گرفتیم. رتبه موتور جستجوی گوگل به صورت افزایشی و رتبه الکسا به صورت کاهشی است بنابراین برای نرمال‌سازی رتبه گوگل از فرمول (۱۰) و برای نرمال‌سازی رتبه الکسا از فرمول (۱۱) استفاده نموده‌ایم:

$$G_i = \frac{x - \min}{\max - \min} \qquad (۱۰)$$

$$A_i = ۱ - \frac{x - \min}{\max - \min} \qquad (۱۱)$$

برای ترکیب مقادیر نرمال شده و محاسبه رتبه نهایی از فرمول (۱۲) استفاده نمودیم:

$$R = \alpha * G_i + (۱ - \alpha) * A_i \qquad (۱۲)$$

G_i و A_i مقادیر نرمال شده رتبه‌های گوگل و الکسا و α ضریب تأثیر است که مقدار آن با توجه به ضریب همبستگی حجم تراکنش‌های پنجره لغزان جاری با مقادیر رتبه‌های گوگل و الکسای وب‌سایت‌های پنجره لغزان طبق فرمول‌های (۱۳)، (۱۴) و (۱۵) تعیین می‌گردد.

$$\rho_۱ = \frac{\sigma_{xy}}{\sigma_x \times \sigma_y} \quad x : \text{Google ranks},\ y : \text{Sale volume} \qquad (۱۳)$$

$$\rho_۲ = \frac{\sigma_{xy}}{\sigma_x \times \sigma_y} \quad x : \text{Alexa ranks},\ y : \text{Sale volume} \qquad (۱۴)$$

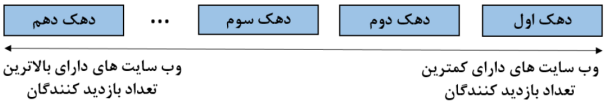
$$\alpha = (\rho_۱ - \rho_۲) + ۰/۵ \qquad (۱۵)$$

$\rho_۱$ ضریب همبستگی حجم تراکنش‌های پنجره لغزان جاری با رتبه‌های گوگل و $\rho_۲$ ضریب همبستگی حجم تراکنش‌های پنجره لغزان جاری با رتبه‌های الکسا می‌باشد. در صورتی که مقادیر $\rho_۱$ و $\rho_۲$ برابر باشد مقدار α برابر ۰/۵ است در غیر این صورت طبق فرمول (۱۵) به اندازه فاصله دو ضریب همبستگی $\rho_۱$ و $\rho_۲$ از مقدار ۰/۵ منحرف می‌گردد.

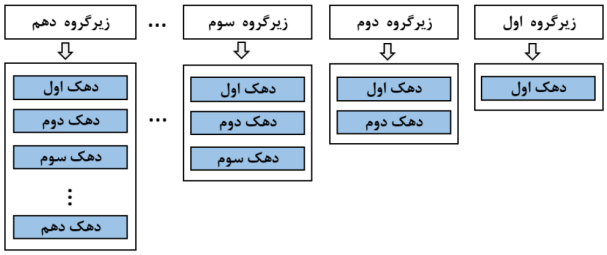
۲.۴ مدل تشخیص مبتنی بر RBM

نوآوری ما در این بخش ایجاد زیرگروه‌هایی از وب‌سایت‌های موجود در پنجره لغزان جاری بر اساس رتبه‌های بازدید آنها و تشخیص موارد مشکوک به پولشویی با تحلیل درون گروهی در این زیرگروه‌ها با استفاده از ماشین RBM بوده است. برای این منظور، ابتدا همسایگان نزدیک هر وب‌سایت

[1] personalized page rank



شکل ۶. دهک‌بندی وب‌سایت‌ها بر اساس تعداد بازدیدکنندگان



شکل ۷. ایجاد زیرگروه‌های دهگانه از وب‌سایت‌های یک صنف

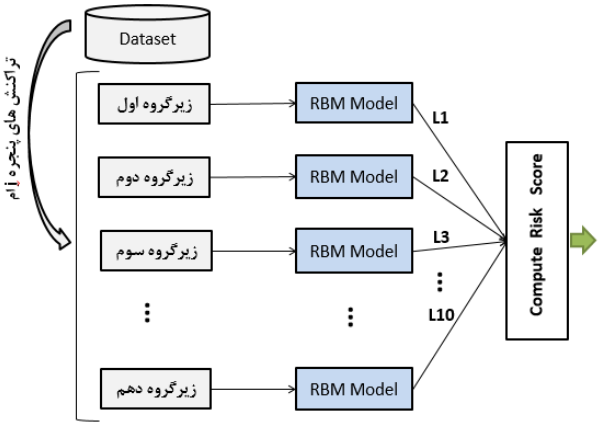
را با رویکرد دهک‌بندی مشخص نمودیم. اولین دهک شامل وب‌سایت‌های با پایین‌ترین رتبه بازدید و آخرین دهک شامل وب‌سایت‌های با بالاترین رتبه بازدید می‌باشد. به شکل ۶ در این مورد توجه نمائید.

دلیل استفاده از ایده دهک‌بندی این بوده است که با بررسی انجام شده مشخص گردید که تعداد زیاد بازه‌ها باعث افت تعداد وب‌سایت‌های موجود در بازه‌ها شده و تحلیل درون گروهی را دچار مشکل می‌نماید و تعداد کمتر از ده بازه نیز دقت محاسبه نمرات خطر پولشویی را کاهش می‌دهد.

در مرحله بعد وب‌سایت‌های این دهک‌ها را طبق شکل ۷ به صورت تجمعی ترکیب نموده و ده زیرگروه ایجاد کردیم. اولین زیرگروه شامل وب‌سایت‌های دهک اول بوده و دومین زیرگروه شامل وب‌سایت‌های دهک‌های اول و دوم است و نهایتا آخرین زیرگروه شامل وب‌سایت‌های همه دهک‌ها خواهد بود. بنابراین در زیرگروه اول وب‌سایت‌هایی حضور خواهند داشت که رتبه بازدید و تراکنش‌های مالی آنها بسیار پائین است اما در زیرگروه دوم، وب‌سایت‌هایی با رتبه بازدید و حجم تراکنش‌های بالاتر افزوده می‌شوند و این روند به صورت تجمعی تا آخرین زیرگروه ادامه می‌یابد تا تمام وب‌سایت‌ها تجمیع گردند.

سپس طبق شکل ۸ در هر زیر گروه با استفاده از ماشین RBM موارد مشکوک به پولشویی را شناسایی نموده و در لیست‌های $L_۱$ تا $L_{۱۰}$ قرار دادیم. مدل RBM از حد آستانه سطح انرژی آزاد برای تشخیص موارد نامتعارف تراکنش‌ها استفاده می‌نماید. تعیین غیردقیق حد آستانه، نرخ اعلام‌های نادرست را افزایش می‌دهد. برای تعیین حد آستانه سطح انرژی آزاد ابتدا بالاترین سطوح انرژی آزاد را که دارای فاصله معناداری از بقیه سطوح هستند، به عنوان مقادیر پرت کنار گذاشتیم. سپس میانگین مقادیر باقیمانده را به عنوان حد آستانه سطح انرژی آزاد در نظر گرفتیم. با مشخص شدن حد آستانه انرژی آزاد، وب‌سایت‌هایی که سطح انرژی آزاد آنها بیش از حد آستانه تعیین شده باشد، به عنوان موارد مشکوک به پولشویی در نظر گرفته شد.

در مرحله بعد بررسی نمودیم که هر یک از وب‌سایت‌ها در نتایج چه



شکل ۸. تشخیص موارد مشکوک در زیرگروه‌های دهگانه

تعداد از زیرگروه‌ها به عنوان مورد مشکوک به پولشویی حضور داشته‌اند. هرگاه وب‌سایتی صرفا در نتایج زیرگروه اول نامتعارف تشخیص داده شود نمره خطر پولشویی آن ۱ است و اگر در نتایج خروجی زیرگروه‌های اول و دوم حضور داشته باشد نمره خطر پولشویی آن ۲ است، و نهایتا اگر در نتایج همه زیرگروه‌ها حضور داشته‌باشد نمره خطر پولشویی آن ۱۰ می‌باشد. فرمول‌های (۱۶) تا (۱۹) روش محاسبه نمرات خط پولشویی توسط مدل RBM را نشان می‌دهد. $RBM_{S_{ij}}$ نمره خطر پولشویی وب‌سایت iام در پنجره jام است که توسط مدل RBM بدست آمده است.

$$RBM_{S_{ij}} = ۱ \quad \text{if } web_i \in L_۱ \tag{۱۶}$$
$$RBM_{S_{ij}} = ۲ \quad \text{if } web_i \in L_۱ \& L_۲ \tag{۱۷}$$
$$RBM_{S_{ij}} = ۳ \quad \text{if } web_i \in L_۱ \& L_۲ \& L_۳ \tag{۱۸}$$
$$\vdots$$
$$RBM_{S_{ij}} = ۱۰ \quad \text{if } web_i \in L_۱ \& L_۲ \& \cdots \& L_{۱۰} \tag{۱۹}$$

۳.۴ مدل تشخیص مبتنی بر CBR

همانطور که پیش از این ذکر گردید، مدل استدلال مبتنی بر مورد را به این علت مورد استفاده قرار دادیم که بتوانیم دامنه تشخیص وب‌سایت‌های مشکوک به پولشویی در پنجره لغزان جاری را ارتقاء داده و نرخ اعلام‌های منفی نادرست را کاهش دهیم. هرگاه وب‌سایت iام از پنجره لغزان jام را با عنوان «وب‌سایت هدف» و وب‌سایت‌های بازیابی شده از پایگاه سوابق را با عنوان «وب‌سایت‌های مشابه» نام‌گذاری نمائیم، مراحل مدل CBR پیشنهادی به شرح زیر است.

بازیابی: ابتدا با جستجو در پایگاه سوابق پولشویی مجموعه وب‌سایت‌هایی که دارای رتبه بازدید یکسان با وب‌سایت هدف هستند را بازیابی نمودیم

انتخاب نزدیکترین موارد مشابه: با مقایسه تعداد و جمع مبالغ تراکنش‌های وب‌سایت هدف با تعداد و جمع مبالغ تراکنش‌های وب‌سایت‌های مشابه، نزدیکترین وب‌سایت‌ها را از لحاظ تعداد و حجم تراکنش‌ها با استفاده از فاصله منهتن معین نمودیم.

محاسبه نمره خطر پولشویی: نمره خطر پولشویی وب‌سایت هدف با توجه به میزان بالاتر بودن حجم تراکنش‌های وب‌سایت هدف از وب‌سایت مشابه محاسبه می‌شود. طبق فرمول (۲۰) به نمره خطر پولشویی وب‌سایت مشابه kام به اندازه ضریب W_{ik} افزوده‌ایم تا نمره خطر پولشویی وب‌سایت هدف محاسبه گردد.

$$S_{ik} = S_k + W_{ik} * S_k \tag{۲۰}$$

W_{ik} وزن فاصله حجم تراکنش‌های وب‌سایت هدف از وب‌سایت مشابه kام است که با فرمول (۲۱) محاسبه شده است. هر چه مقادیر V_{ij} و C_{ij} که حجم فروش و تعداد تراکنش وب‌سایت هدف هستند از مقادیر حجم فروش و تعداد تراکنش‌های وب‌سایت مشابه Kام یعنی V_k و C_k بیشتر باشد؛ نمره خطر پولشویی وب‌سایت هدف را بالاتر لحاظ نموده‌ایم.

$$W_{ik} = \frac{|V_{ij} - V_k| + |C_{ij} - C_k|}{|V_k + C_k|} \tag{۲۱}$$

در آخرین مرحله از مدل CBR، میانگین n نمره خطر پولشویی محاسبه شده با فرمول (۲۰) را با استفاده از فرمول میانگین (۲۲) محاسبه نموده و به عنوان نمره خطر پولشویی وب‌سایت هدف لحاظ نمودیم.

$$CBR_{S_{ij}} = MEAN_{k=۱}^{n}(S_{ik}) \tag{۲۲}$$

$CBR_{S_{ij}}$ نمره خطر پولشویی وب‌سایت iام در پنجره jام است که توسط مدل CBR محاسبه گردیده است.

۴.۴ ترکیب نتایج دو مدل

در ترکیب نتایج خروجی مدل‌های RBM و CBR، هرگاه وب‌سایتی در هر دو لیست حضور داشته باشد، طبق فرمول (۲۳) مقدار بیشینه نمرات خطر پولشویی انتخاب می‌گردد تا سطح خطر بالاتر مد نظر قرار بگیرد.

$$S_{ij} = \mathrm{MAX}(RBM_{S_{ij}}, CBR_{S_{ij}}) \tag{۲۳}$$

در صورتی که وب‌سایتی صرفا در یکی از دو لیست خروجی باشد. نمره خطر پولشویی آن تغییر نمی‌کند. پس از ترکیب نتایج خروجی مدل‌ها، لیست وب‌سایت‌های مشکوکِ پنجره لغزان جاری در پایگاه سوابق پولشویی ثبت می‌گردد.

۵.۴ گزارش وب‌سایت‌های با خطر پولشویی بالا

برای گزارش نتیجه، ابتدا با استفاده از فرمول (۲۴) جمع نمرات خطر پولشویی همه‌ی وب‌سایت‌های دارای سابقه پولشویی را محاسبه نمودیم. با استفاده از وزن $\frac{k}{j}$ هرچه پنجره لغزان kام به پنجره لغزان جاری j نزدیک‌تر باشد، وزن بیشتری در محاسبه مجموع نمرات خواهد داشت.

$$Final_{S_{ij}} = \sum_{k=۱}^{j} \left(\frac{k}{j} \times S_{ik} \right) \tag{۲۴}$$

جدول ۱. مشخصه‌های مجموعه دادگان

ردیف	ویژگی	نوع	توضیحات
۱	Vendor-id	عدد	شناسه فروشنده
۲	Vendor-website	رشته	آدرس اینترنتی فروشنده
۳	Job-class	عدد	کد صنف فروشنده
۴	amount	عدد	مبلغ تراکنش مالی

جدول ۲. مشخصات اصناف مورد ارزیابی

تعداد نامتعارف	تعداد وب‌سایت‌ها	عنوان صنف
۱۸	۳۳۱	آژانس فروش بلیط هواپیما، اتوبوس و قطار
۱۷	۱۶۵	دلال اوراق بهادار
۱۵	۳۰۹	خدمات شبکه و اینترنت
۵۰	۸۰۵	

S_{ik} نمره خطر پولشویی وب‌سایت iام در پنجره Kام است. در آخرین مرحله با مرتب‌سازی نزولی مقادیر $Final_{S_{ij}}$ وب‌سایت‌های با خطر پولشویی بالا در ابتدای لیست قرار خواهند گرفت.

۵ ارزیابی

مجموعه دادگان واقعی مورد استفاده در پژوهش‌های [۱، ۲] در این پژوهش نیز مورد استفاده واقع شده است. این مجموعه دادگان شامل ۲۲ میلیون تراکنش ارجاع شده به یکی از شرکت‌های ارائه کننده خدمات پرداخت داخل کشور می‌باشد که دارای ۹۶۶۵ شناسه فروشنده و ۱۰۲۵۲ آدرس وب‌سایت است و بیش از ۸۰ درصد فروشندگان صرفا دارای یک دامنه اینترنتی هستند. تعداد اصناف ۲۰۵ مورد بوده که ۱۸۲ صنف دارای تراکنش مالی بوده‌اند. مجموعه دادگان دارای ۱۴ ویژگی اصلی است که ما صرفا از ویژگی‌های جدول ۱ استفاده نموده‌ایم.

برای ارزیابی سیستم پیشنهادی، سه صنف با بیشترین تعداد وب‌سایت‌ها را انتخاب نمودیم. در جدول ۲، تعداد کل وب‌سایت‌ها و تعداد وب‌سایت‌های با برچسب پولشویی اصناف منتخب آورده شده است.

در نتایج ارزیابی جهت تشخیص سیستم پیشنهادی مقاله حاضر از سیستم پیشنهادی [۱]، از عناوین RBM-CBR و PGA استفاده شده است. عنوان RBM-CBR برای سیستم پیشنهادی ما و عنوان PGA برای سیستم پیشنهادی پژوهش [۱] است. ماتریس درهم‌ریختگی نتایج هر دو مدل در جدول ۳ نمایش داده شده است.

با توجه به ماتریس درهم‌ریختگی، نرخ اعلام‌های مثبت نادرست مدل PGA طبق فرمول (۲۵) برای اصناف مورد بررسی برابر ۱۱ درصد بوده است:

جدول ۳. ماتریس درهم‌ریختگی نتایج ارزیابی

اصناف	TP		FP		TN		FN	
	RBM - CBR	PGA	RBM - CBR	PGA	RBM - CBR	PGA	RBM - CBR	PGA
آژانس فروش بلیط هواپیما، اتوبوس و قطار	۱۸	۱۸	۲	۵۲	۳۱۱	۲۶۱	۰	۰
دلال اوراق بهادار	۱۷	۹	۲	۱	۱۴۶	۱۴۷	۰	۸
خدمات شبکه و اینترنت	۱۵	۱۵	۳	۳۱	۲۹۱	۲۶۳	۰	۰
	۵۰	۴۲	۷	۸۴	۷۴۸	۶۷۱	۰	۸

$$FPR = \frac{FP}{FP+TN} = \frac{۸۴}{۸۴+۶۷۱} = ۰/۱۱ \qquad (۲۵)$$

همانطور که از فرمول (۲۶) مشهود است؛ نرخ اعلام‌های مثبت نادرست مدل RBM-CBR با حدود ۱۱ درصد بهبود برابر ۰/۰۱ شده است:

$$FPR = \frac{FP}{FP+TN} = \frac{۷}{۷+۷۴۸} = ۰/۰۱ \qquad (۲۶)$$

دقت تشخیص با استفاده از فرمول (۲۷) محاسبه می‌گردد. میزان دقت مدل PGA برای اصناف مورد بررسی برابر ۰/۸۹ بوده است که در مدل RBM-CBR به میزان ۱۰ درصد بهبود یافته و ۰/۹۹ شده است.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \qquad (۲۷)$$

نکته حائز اهمیت دیگر کاهش نرخ اعلام‌های منفی نادرست در مدل RBM-CBR است. دلیل این کاهش، تأثیر مؤلفه CBR در افزایش دامنه تشخیص موارد پولشویی است. میزان اعلام‌های منفی نادرست مدل PGA برابر ۱۶ مورد بوده است که به صفر کاهش پیدا کرده است.

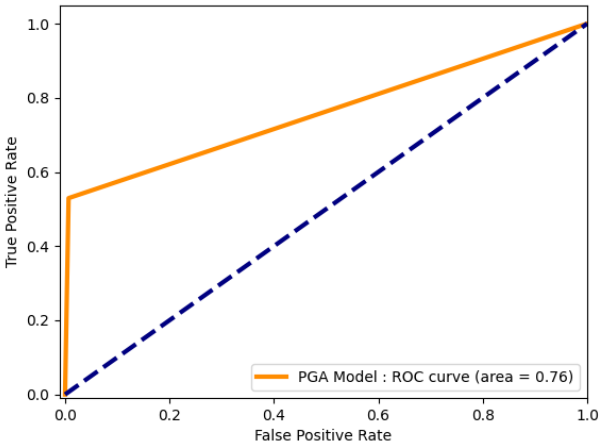
در جدول ۴ نمرات معیارهای F1، Recall، precision هر دو مدل آورده شده است. همانطور که از نتایج ارزیابی مشهود است؛ مدل RBM_CBR نمرات بالاتری کسب نموده است. امتیاز F1 مدل PGA برابر ۰/۵۲ بوده است که با بهبودهای انجام شده در مدل RBM-CBR نرخ FPR به اندازه زیادی کاهش یافته و در نتیجه آن مقدار Precision مدل افزایش یافته است. با بهبود Precision، امتیاز F1 نیز طبق فرمول (۲۸) افزایش یافته و برابر ۰/۹۳ گردیده است که بهبود زیادی را نشان می‌دهد:

$$F۱\_score = ۲ \times \frac{Precision \times Recall}{Precision + Recall} \qquad (۲۸)$$

می‌دانیم که در نمودار ROC، دو شاخص نرخ اعلام‌های مثبت درست (TPR) و نرخ اعلام‌های مثبت نادرست (FPR) مدل ترکیب شده و توازن آنها در یک نمودار نشان داده می‌شود. توجه داشته باشید که هرچه نرخ TPR مدل نزدیک به ۱ و نرخ FPR آن نزدیک به صفر باشد گوشه بالا و سمت چپ منحنی تیزتر بوده و سطح زیر منحنی بیشتری را پوشش خواهد داد.

جدول ۴. نتایج ارزیابی

اصناف	Precision		Recall		F1 measure	
	RBM - CBR	PGA	RBM - CBR	PGA	RBM - CBR	PGA
آژانس فروش بلیط هواپیما، اتوبوس و قطار	۰/۹۰	۰/۲۶	۰/۹۹	۰/۹۲	۰/۹۵	۰/۴۰
دلال اوراق بهادار	۰/۸۹	۰/۵۲	۰/۹۹	۰/۷۶	۰/۹۴	۰/۶۷
خدمات شبکه و اینترنت	۰/۸۳	۰/۳۳	۰/۹۹	۰/۹۵	۰/۹۰	۰/۴۹
	۰/۵۲	۰/۹۳	۰/۸۸	۰/۹۹	۰/۹۳	۰/۵۲



شکل ۹. نمودار ROC-AUC مدل PGA

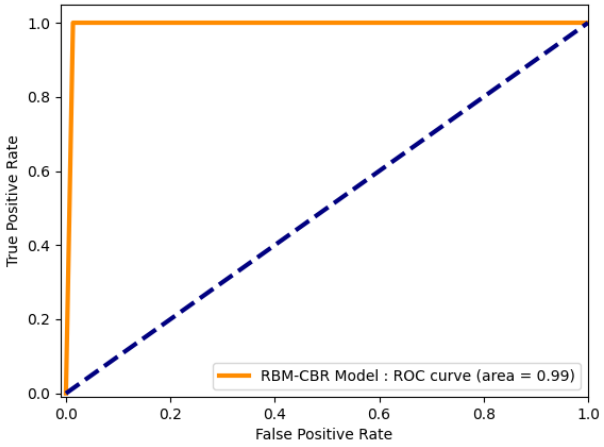
شکل ۹ شاخص عملکرد سیستم پیشنهادی PGA برای تشخیص موارد پولشویی در صنف دلال اوراق بهادار را نشان می‌دهد که سطح زیر منحنی آن ۰/۷۶ است.

شکل ۱۰ نیز شاخص عملکرد سیستم پیشنهادی ما برای تشخیص موارد پولشویی در صنف دلال اوراق بهادار را نشان داده است. همانطور که مشهود است، سطح زیر منحنی ۰/۹۹ بوده و افزایش قابل توجهی نسبت به نمودار ۹ نشان داده شده است. این بهبود ناشی از کاهش نرخ اعلام‌های مثبت نادرست (FPR) در سیستم پیشنهادی ما بوده است.

پیچیدگی محاسباتی راه‌حل پیشنهادی: با توجه به این که مرتبه زمانی الگوریتم‌های جستجو و مرتب سازی که در قسمت‌هایی از راه‌حل پیشنهادی استفاده شده‌اند حداکثر از مرتبه $n \times \log n$ می‌باشند، و مرتبه زمانی ماشین بولتزمن محدود در حالتی که تعداد متغیرهای لایه‌های آشکار و پنهان آن کم باشد خطی است [۴]، و با توجه به اینکه حداکثر نرخ یادگیری مدل پیشنهادی ما ۱۰۰ epoch است؛ بنابراین پیچیدگی محاسباتی راه‌حل پیشنهادی ما حداکثر از مرتبه n^2 خواهد بود.

۶ جمع‌بندی

با توجه به رشد روزافزون کسب و کارهای اینترنتی و گسترش خدمات پرداخت الکترونیکی، باید مراقبت نمود که مجرمان پولشویی در پوشش



شکل ۱۰. نمودار ROC-AUC مدل RBM-CBR

وب‌سایت‌های صوری از خدمات پرداخت الکترونیکی شرکت‌های ارائه کننده خدمات پرداخت برای تطهیر پول‌های غیرقانونی بهره‌برداری ننمایند. در این مقاله با ترکیبی از یادگیری عمیق مبتنی بر ماشین بولتزمن محدود و استدلال مبتنی بر مورد (سوابق پولشویی) و با روش تحلیل درون گروهی توانستیم پولشویی الکترونیکی را در تراکنش‌های تامین کنندگان خدمات پرداخت با دقت ۹۹٪ تشخیص دهیم.

بهبود عمده ما نسبت به راه‌حل ارائه شده در [۱] استفاده از پارامتر تناسب حجم بازدیدکنندگان وب‌سایت با حجم تراکنش‌های ارجاع شده از آن برای تشخیص الگوهای خارج از صنف بوده است. برای این منظور ابتدا با ترکیب رتبه‌های گوگل و الکسای وب‌سایت‌های فروشندگان، تعداد بازدیدکنندگان آنها را تخمین زدیم؛ سپس با توجه به دهک رتبه‌های بازدید، ۱۰ زیرگروه تجمعی از وب‌سایت‌های صنف مربوطه ایجاد نمودیم و در هر زیرگروه وب‌سایت‌های مشکوک به پولشویی را در پنجره لغزان جاری با استفاده ماشین بولتزمن محدود و روش استدلال مبتنی بر سوابق پولشویی تشخیص دادیم. در نهایت با ترکیب نتایج زیرگروه‌ها، نمره خطر پولشویی وب‌سایت‌های مشکوک را محاسبه نموده و وب‌سایت‌های با خطر پولشویی بالا را استخراج کردیم. نتایج ارزیابی‌ها نشان داد که راه‌حل پیشنهادی ما نرخ اعلام‌های مثبت نادرست راه‌حل ارائه شده در [۱] را به میزان ۱۲ درصد کاهش داده و میزان دقت تشخیص آن را به میزان ۱۱ درصد بهبود داده است. امتیاز F1 سیستم پیشنهادی ما ۰/۹۲ است که در مقایسه با امتیاز ۰/۵۲ راه‌حل قبلی ارتقای قابل توجهی را نشان می‌دهد. برای بهبود این راه‌حل می‌توان به کارهای آتی زیر اشاره نمود:

- تشخیص وب‌سایت‌های پوششی برای اهداف پولشویی با تشخیص نامتناسب بودن حجم فروش وب‌سایت با نرخ بروزرسانی محصولات جدید.
- تکمیل بردار رفتار فروشندگان با توجه به ویژگی‌های قابل استخراج از وب‌سایت نظیر عرضه محصولات غیر مجاز، طول عمر دامنه، کشور میزبان دامنه، کشور و شرکت گواهینامه SSL، اعتبارنامه دولتی تجارت الکترونیکی نماد و غیره.
- تشخیص دقیق‌تر حد آستانه سطح انرژی آزاد ماشین بولتزمن

محدود با روش‌های یادگیری ماشین.

سپاسگزاری

مجموعه دادگان مورد استفاده در این پژوهش مجموعه دادگانی است که در پژوهش‌های [۱، ۲] نیز مورد استفاده قرارگرفته است. در این راستا از آقای مهندس محمدحسین حجتی به جهت همکاری صمیمانه‌شان تشکر می‌نمائیم. همچنین از مجموعه پرداخت الکترونیک سامان کیش به جهت تهیه و تدارک مجموعه دادگان مورد استفاده تشکر می‌گردد.

مراجع

[۱] حجتی، محمدحسین و صادقیان، بابک. تحلیل فورنسیک دیجیتال در تراکنش‌های مالی شرکت‌های psp به کمک روش‌های ترکیبی. در کارشناسی ارشد، تهران، ۱۳۹۸. دانشگاه صنعتی امیرکبیر.

[۲] حجتی، محمدحسین و صادقیان، بابک. دسته‌بندي مشتريان شركت‌هاي ارائه دهنده سرويس‌هاي پرداخت با استفاده از تارنماي فروشگاهی آنان به کمک روش‌هاي داده‌كاوي. در شانزدهمین کنفرانس بین المللی انجمن رمز ایران، ۱۳۹۸.

[3] Mubalaike, Aji Mubarek and Adali, Esref. Deep learning approach for intelligent financial fraud detection system. in *2018 3rd International Conference on Computer Science and Engineering (UBMK)*, pp. 598–603. IEEE, 2018.

[4] Do, Kien, Tran, Truyen, and Venkatesh, Svetha. Energy-based anomaly detection for mixed data. *Knowledge and Information Systems*, 57(2):413–435, 2018.

[5] Weibing, Peng. Research on money laundering crime under electronic payment background. *Journal of Computers*, 6(1):147–154, 2011.

[6] Basel AML Index 2017 Report. [Online]. Available: https://baselgovernance.org/sites/default/files/2020-06/2017_report.pdf.

[7] He, Ping. A typological study on money laundering. *Journal of Money Laundering Control*, 2010.

[8] Fischer, Asja and Igel, Christian. An introduction to restricted boltzmann machines. in *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications: 17th Iberoamerican Congress, CIARP 2012, Buenos Aires, Argentina, September 3-6, 2012. Proceedings 17*, pp. 14–36. Springer, 2012.

[9] Abutair, Hassan YA and Belghith, Abdelfettah. Using case-based reasoning for phishing detection. *Procedia Computer Science*, 109:281–288, 2017.

[10] Filipkowski, Wojciech. Cyber laundering: An analysis of typology and techniques. *International Journal of Criminal Justice Sciences*, 3(1), 2008.

[11] Bolton, Richard J, Hand, David J, et al. Unsupervised profiling methods for fraud detection. *Credit scoring and credit control VII*, pp. 235–255, 2001.

[12] Ferdousi, Zakia and Maeda, Akira. Unsupervised outlier detection in time series data. in *22nd International Conference on Data Engineering Workshops (ICDEW'06)*, pp. x121–x121. IEEE, 2006.

[13] Kim, Yoonseong and Sohn, So Young. Stock fraud detection using peer group analysis. *Expert Systems with Applications*, 39(10):8986–8992, 2012.

MonadiAfta-Bianual Scientific Journal for Cyberspace Security
2022, Volume 20, Number 2 (pp. 11–21)
https://monadi.isc.org.ir

Presented at the ISCISC 2022 in University of Giulan, Rasht, Iran

Electronic Money Laundering Detection in Transactions of Payment Service Providers★

Ali Nazari * and Babak Sadeghiyan

Department of Computer Engineering, Amirkabir University of Technology, Tehran, Iran

ARTICLE INFO.

Keywords:
electronic money laundering
illegal money
abnormal transaction
restricted Boltzmann machine
case-based reasoning
crime investigation

dor: 20.1001.1.24763047.1401.11.2.2.1

Type: research paper

ABSTRACT

Under the coverage of legitimate commerce, criminals money-launder their illicit incomes through the payment gateways provided by Payment Service Providers (PSP). In order to do money-laundering forensics in transactions of PSP companies, a new method was proposed by Hojati et al which is done through detecting deviations from class behavior based on peer group analysis (PGA) method. Our experiments showed that using the proposed method for money laundering detection leads to a false positive rate of about 13%. In this paper, we improved the proposed method and reduced the false positive rate to less than 1%. To achieve this, we analyzed the amount of financial transactions of sellers along with the number of visitors to their websites in PGA. Based on the number of visitors, we estimated the volume of transactions for each seller. If the volume of sales was much higher than expected, we considered it abnormal. We achieved a higher detection accuracy by using a restricted Boltzmann machine to separate out-of-class transactions. We also reduced rate of false negative alarms by the help of CBR method. Our proposed system detects money laundering online using a four-week sliding window. The experimental results confirmed the detection accuracy of 99% for our proposed system.

★ The ISCISC 2022 Program Committee effort is highly acknowledged for reviewing this paper.
* Corresponding author
Email addresses: nazariali@aut.ac.ir (Ali Nazari), basadegh@aut.ac.ir (Babak Sadeghiyan)