روش ها و راهکارهای شناسایی تقلب مالی در بانکداری الکترونیک با استفاده از داده کاوی

ولی سرلک [1]، محمد گردان[2]، مجید خلجی[3]،میترا گودرزی[4]

1- دکتری مهندسی کامپیوتر،دانشگاه آزاد اسلامی واحد نجف آباد-اصفهان ایران
(Mohandesprg@gmail.com)
2- کارشناسی ارشد مهندسی کامپیوتر، دانشگاه آزاد اسلامی واحد بروجرد
(mohammad.gordan@yahoo.com)
3- دکتری مهندسی کامپیوتر،گروه کامپیوتر،دانشگاه آزاد اسلامی واحد نجف آباد-اصفهان ایران
(khalaji.eng@gmail.com)
4- کارشناسی ارشد مهندسی فناوری اطلاعات، موسسه غیر انتفاعی نور طوبی تهران
(Mitragoodrzi@gmail.com)

چکیده:

امروزه تقلب که قدیمی به اندازه زندگی بشریت دارد ، یک کسب و کار چندین میلیون دلاری در سطح دنیا محسوب شده و حجم مالی آن روز به روز در حال افزایش است .در سال های اخیر ، توسعه فناوری های جدید راه های زیادی را برای متقلبان و مجرمان باز کرده است که مرتکب تقلب شوند. ایجاد ، ممکن است فرصت های بیشتری را برای ارتکاب تقلب در اختیار مجرمان قرار دهد.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 .به دلیل هزینه های بسیار زیاد مستقیم یا غیر مستقیم تقلب،بانک ها و موسسات مالی و پولی به شدت به دنبال تسریع و سرعت عمل در شناخت فعالیت های کلاهبرداران و متقلبان می باشند.این امر به دلیل اثر مستقیم آن روی خدمت رسانی به مشتریان این موسسات ، کاهش هزینه عملیاتی وباقی ماندن به عنوان یک ارائه دهنده خدمات مالی معتبر و قابل اطمینان است.در این پژوهش به تشریح چگونگی عملکرد سازوکارهای مبتنی بر آن ،انواع تکنیکهای تشخیص تقلب در بانکداری الکترونیک ارائه و روشهای داده کاوی مورد استفاده در کشف تقلب مزایا و معایب هریک به تفصیل شرح داده خواهد شد

واژه‌های کلیدی

کشف تقلب،بانکداری الکترونیکی،تقلب مالی،داده کاوی، تشخیص ناهنجاری – تشخیص سوء استفاده

1- مقدمه

دستورالعمل هیچ تعریف پذیرفته شده جهانی از تقلب مالی وجود ندارد[1].ونگ و همکاران تقلب را این گونه تعریف کرده اند:اقدامی هدفمند برای کسب منفعت مالی غیر مجاز که برخلاف قوانین ، قواعد ،یا سیاست هاست. [2]

برای واژه تقلب در مقالات و منابع علمی ، معانی مختلفی بیان شده است ،لیکن آنچه در تمامی این تعاریف ، مشترک و یکسان می باشد ، این است که تقلب ،نوعی سوء استفاده از منابع در جهت منافع شخصی ،به عمد و کاملا غیر قانونی است تقلب در مفهوم عام ،عبارت است از تحریف حقایق با اهمیت توسط کسی که می داند مطلبش حقیقت ندارد و یا ارائه حقایق ،باکمال بی توجه به صحت آنها و به قسط فریب دیگران .

در تعریف دیگر ، واژه تقلب عبارت است از سوء استفاده از سود یک سازمان بدون اینکه لزوما به عواقب قانونی آن منجر شود .در تعریفی دیگر ،تقلب به فرایندی اشاره دارد که طی آن یک یا چند نفر ، عمدا و مخفیانه دیگران را از هر چیز با ارزشی ،به خاطر منافع شخصی خود محروم کنند[3].

امروزه با گسترش فناوری مدرن و ارتباطات جهانی ، تقلب به طرز چشمگیری در حال افزایش است و هزینه زیادی را به کسب و کارها تحمیل می کند .در نتیجه شناسایی تقلب به مساله بسیار مهمی تبدیل شده است .

انواع تقلب های گوناگون تقلبهای مالی ، مانند تقلب کارت اعتباری ، تقلب شرکتی و پولشویی ،نگرانی های بسیاری را سبب شده و نظرها را به سوی خود جلب کرده است.نگای و همکاران در یک طبقه بندی کلی ، انواع تقلب مالی را در چهار دسته تقسیم کرده اند؛تقلبهای بانکی ،تقلبهای بیمه ای ،تقلب اوراق بهادار و کالاها ،سایر تقلبهای مالی که در جدول1.مشاهده می کنید[1].

طبقات تقلب مالی	فعالیتهای متقلبانه
تقلب بانکی	تقلب رهن، پولشویی
تقلب بیمه‌ای	تقلب بیمه سلامت، تقلب بیمه خودرو
تقلب اوراق بهادار و کالاها	تقلبهای هرمی، احتلاس
سایر تقلبهای مالی	تقلبهای شرکتی، تقلب با استفاده از رسانه‌ها

جدول1. طبقه بندی تقلبهای مالی (Ngai et,al.2010)

سیستم های مالی مبتنی بر فناوری اطلاعات – به دلیل پتانسیل بالایی که در جهت امکان سرقت پولی در حجم بالا دارند – اغلب ، اهداف راحتی برای حمله کنندگان هستندکه از نقص احراز هویت های متعدد و یا نقاط ضعف موجود در مدل های امنیتی اجرا شده در سرویس ها استفاده کرده و اهداف خود را پیاده نماید . احراز هویت ضعیفی که توسط سازوکارهای امضا ، پین کد[1]، رمز عبور و کد امنیتی کارت [2] اتفاق می افتد ،باعث آسان شدن تراکنش های غیرقانونی مالی حمله کنندگان و از طریق اجرای حملات سیستمی خلاقانه می شود .

در جدول 2، مجموع زیان های مالی موسسات و بانک های انگلستان از طریق کارت های بانکی ، از سال 2004 تا 2007 و طی 4 سال نشان داده شده است .

[1] Pin code
[2] Card Security Code

در سال 2004،موسسات مالی و بانک ها به منظور کاهش آمار تقلب و کلاهبرداری از طریق کارت یک گام فعال برداشتند ؛ بدین صورت که از روش های موجود ، که براساس امضای مشتری به سمت روش احراز هویت [1] به کمک پین کد در تمامی دستگاههای POS سوئیچ کردند .

نوع تقلب	+/-(۰۶/۰۷) (درصد)	۲۰۰۷	۲۰۰۶	۲۰۰۵	۲۰۰٤
تلفن، اینترنت و ایمیل (تقلب‌های بدون وجود کارت)	+۳۷	۲۹۰/۵	۲۱۲/۷	۱۸۳/۲	۱۵۰/۸
جعل کارت (شامل کف‌زنی و کپی‌سازی)	+۴۶	۱۴۴/۳	۹۸/۶	۹۶/۸	۱۲۹/۷
تقلب از طریق کارت‌های گم‌شده یا به سرقت رفته	-۱۸	۵۶/۲	۶۸/۵	۸۹/۰	۱۱۴/۴
سرقت شماره شناسایی کارت	+۷	۳۴/۱	۳۱/۹	۳۰/۵	۳۶/۹
ایمیل‌های بدون رسید	-۳۴	۱۰/۲	۱۵/۴	۴۰/۰	۷۲/۹
جمع	+۲۵	۵۳۵/۲	۴۲۷/۰	۴۳۹/۴	۵۰۴/۸

جدول 2. جدول زیان های حاصل از تقلب های مالی از طریق کارت های اعتباری بانکی در انگلستان (2004 تا 2007)-منبع APAGS سال 2006 (آمار به میلیون پوند)

بطور مشابه ، در جدول 3-4 نیز ، تغییر همزمان و رشد تقلب در حوزه خدمات و برخط و بانکداری الکترونیکی طی 4 سال نشان داده شده است .طی این دوره ، تعداد حملات فیشینگ[2] انجام شده توسط کلاهبرداران از 1713 مورد در سال 2005 به 14156 مورد در سال 2007 رسیده است .این مساله باعث ایجاد سرمایه ای برای این حوزه به نام تبدیل مشتریان با دانش ضعیف در زمینه پروتکل های امنیتی بر خط به مشتریانی با دانش ضعیف در زمینه پروتکل های امنیتی بر خط به مشتریانی با اطلاعات امنیتی بر خط بالا شده است .

در حالی که افزایش سطح آگاهی مشتریان از چنین روش هایی در سال 2007 منجر به کاهش موفقیت شیادان در عملیات فیشینگ و در حوزه برخط شده ، چابکی رفتار کلاهبرداران از سال 2004 به بعد نیز ، باعث افزایش نرخ رشد تقلب در زمینه چک گردیده است.

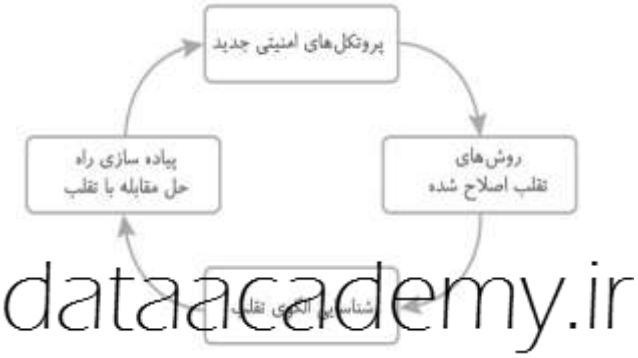
این آمار و ارقام نشان می دهند که توانایی متقلبان نه تنها به صورت خلاقانه و به سمت حمله به سیستم پیچیده تر ارتقاء یافته بلکه به صورت کاملا فعال به کمک مهندسی مجدد [3] و تطبیق روش هایشان با استقرار امنیت ، توسعه نیز یافته است .

[1] Authentication
[2] Phishing
[3] Reengineering

نوع تقلب	+/-(۰۶/۰۷) (درصد)	۲۰۰۷	۲۰۰۶	۲۰۰۵	۲۰۰٤
تقلب در بانکداری الکترونیکی (تقلب‌های آنلاین)	-۳۳	۲۲/۶	۳۳/۵	۲۳/۲	۱۲/۲
تقلب در چک‌های بانکی	+۱۰	۳۳/۵	۳۰/۶	۴۰/۳	۴۶/۲

جدول 3 .جمع زیان های حاصل از تقلب های مالی از طریق سیستم های بانکداری الکترونیکی در انگلستان (2004 تا 2007)-
منبع APAGS سال 2006 (آمار به میلیون پوند)

مطابق شکل1 ،عموما تقلب در چرخه حیات تقلب[1] ، می تواند به عنوان یک مدل به کار رود ، به گونه ای که با تحلیل و آنالیز آن، پاسخ مناسبی به کار رود، به گونه ای که با تحلیل و آنالیز آن ، پاسخ مناسبی به این تقلب داده می شود و مجددا با توسعه دانش و ارائه راه حل های و پروتکل های جدید راه بر متقلبان باز شده و روش های تقلب جدید شکل می گیرند و همچنان چرخه حیات تقلب ادامه می یابد .



شکل 1-چرخه حیات مالی

روند رو به ظهور تقلب های مالی عموما از طریق تحلیل و آنالیز و استخراج اطلاعات (داده کاوی) از بانک اطلاعاتی تراکنش های موسسات مالی ،که نشانه گذاری می گردند، تشخیص داده می شود و این امر به تدوین سیاست ها [2] و پروتکل های [3] امنیتی و احراز هویت جدید کمک می کند.

2–شناسایی تقلب

مدت هاست که روش های سنتی تجزیه و تحلیل داده ها به عنوان یک روش برای تشخیص تقلب استفاده می شود.این کار نیاز به تحقیقات پیچیده و وقت گیری دارد و نیازمند به کارگیری حوزه های مختلف دانش مانند مالی ، اقتصادی ،روش های کسب و کار و مباحث قانونی است.

به مجموعه عملیات یا اقداماتی که براساس روش ها یا متد هایی ، سعی در کشف و شناسایی تقلب های صورت گرفته و یا در حال وقوع دارند ، شناسایی تقلب گفته می شود.

موسسات مالی و پولی به شدت به دنبال سرعت عمل در شناخت فعالیت کلاهبرداران و متقلبان می باشند.این امر به دلیل اثر مستقیم آن روی خدمات رسانی به مشتریان این موسسات ، کاهش هزینه عملیاتی و باقی ماندن به عنوان یک ارائه دهنده خدمات مالی معتبر و قابل اطمینان است .

[1] Fraud Lifecycle
[2] Security Policies
[3] Security Protocols

3–انواع تقلب در بستر بانکداری الکترونیک

رویکردهای تشخیص حمله برا ساس مدل حملات به طور گسترده به دودسته تقسیم می شوند :

3–1–تشخیص سو استفاده [1]

تشخیص سوء استفاده تلاش می کند که حملات مشاهده شده قبلی را در قالب یک الگو یا امضا تشخصیص دهد .به عنوان مثال ،می توان به تغییر مداوم یک پوشه و یا تلاش متعدد به منظور خواندن یک فایل حاوی رمزهای عبور اشاره کرد [3].

در روش تشخیص سوء استفاده ،زمانی که تراکنشی انجام می شود ، این تراکنش با نمونه امضاهای قبلی و حملات شناخته شده قبلی مقایسه می شود و در صورت تشخیص شباهت ،این تراکنش به عنوان یک حمله شناسایی می گردد .روال کار رویکردهای تشخیص سوء استفاده مشاهده مشابه سازوکاری است که نرم افزارهای آنتی ویروس رایانه ها با آن عمل می کنند .در کلیه نرم افزارهای آنتی ویروس ، بانک اطلاعاتی از کلیه امضاهای ویروس ها وجود دارد زمانی که فایل مورد تجزیه و تحلیل قرار می گیرد ، فایل با امضای ویروس های شناخته شده مقایسه می شود و در صورت وجود شباهت ، به عنوان یک تهدید شناسایی می گردد.

تشخیص سوء استفاده به کارگیری حملات شناخته شده قبلی و عملات گذاری الگوی قابل تطبیق به منظور شناسایی تقلب های آتی می باشد .در این روش رفتار حمله کاملا شناخته شده است و مطابق روال شناخته شده عمل می شود .

دقت بالا از مزایای این روش می باشد ، اما بدیهی است که حملات جدیدی که قبلا توسط سیستم شناسایی نشده اند را شامل نمی شود . لذا سازوکار بسیار امنی تلقی نمی گردد.

رویکردهای تشخیص سوء استفاده شامل سیستم خبره [2]،استدلال بر پایه مدل [3]،تجزیه و تحلیل عبور حالت [4] ومونیتورینگ پویایی ضربه کلید [5] می باشد[4].

تشخیص سوء استفاده از روش تشخیص ناهنجاری بسیار ساده تر است هر چند یک اشکال اساسی که به این روش وارد است،این است که در این روش همه حملات قابل پیش بینی نیستند و این امر هم به دلیل الزام شناخت الگوهای سوء استفاده از قبل می باشد. لذا این به عنوان یک ضعف روش تشخیص سوء استفاده مطرح بوده که باید مدنظر قرار گیرد.

با توجه به اینکه در تشخیص سوء استفاده از قواعد و ویژگی های رفتاری شناخته شده استفاده می شود،به راحتی می توان رفتارهای شناخته شده مشکوک مشتریان را تشخیص داد.یک تحلیل تجربی که روی مجموعه ای از تراکنش های واقعی صورت گرفته ، آشکار نموده است که بیشتر تقلب ها دارای ویژگی های رفتاری می باشند.

[1] Misuse Detection
[2] Expert Systems
[3] Model-Based Reasoning
[4] State Transition Analysis
[5] Keystroke Dynamics Monitoring

به عنوان مثال ،برخی از رفتارهایی که نشانه ای از تقلب محسوب می شوند ،عبارتند از:

- حجم زیادی از حساب های متعدد که توسط یک مشتری یکسان دسترسی پیدا کرده اند؛
- تراکنش های که حاوی مبالغ کوچک و از حساب های متعدد و زیاد می باشد؛
- تراکنش های پرداخت بیش از حد معمول در یک حساب خاص؛
- افزایش دفعات ثبت رمز ورود با شکست ، قبل از وقوع تقلب .

این گونه رفتارهای مشتری می توانند به عنوان رفتارهای مشکوک در نظر گرفته شوند و به محض مشاهده مجدد،تقلب منظور گردند[5].

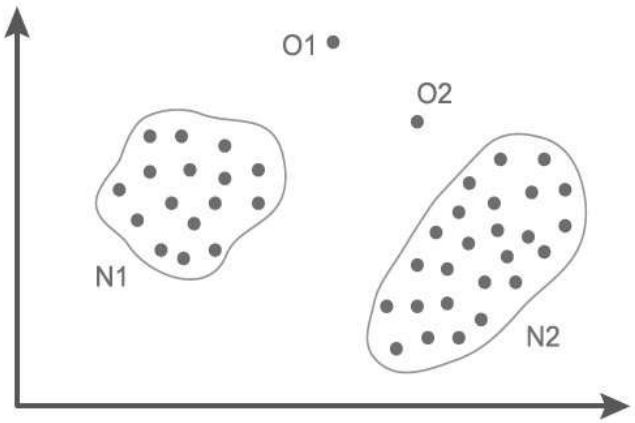
3-2-تشخیص ناهنجاری[1]

در روش تشخیص ناهنجاری تلاش می شود تا یک مشخصه[2] از تاریخچه عملکرد برای هر کاربر ایجاد گردیده و سپس از استخراج هرگونه انحراف[3] به قدر کافی بزرگ در مشخصه کاربر ،بروز یک حمله پی برده شود[6].

اگر بخواهیم تشخیص ناهنجاری را تعریف کنیم ، شاید بهترین تعریف ، تشخیص انحراف از آنچه انتظار داریم و یا انحراف از رفتار نرمال باشد .به دلیل محدود نبودن این روش ، توانایی تشخیص حملات جدید از مزایای آن می باشد .این روش در حقیقت تشخیص تلاش های بدون مجوز به منظور دسترسی به سیستم است .در این روش رفتار معمولی تعریف شده و هر رفتاری دیگری،غیر نرمال توصیف شود .

در نمودار 1،چگونگی دسته بندی اطلاعات بر مبنای رفتار غیر نرمال مشخص شده است .همان طور که مشاهده می شود ، داده های نرمال ، که به علت ماهیت رفتار مشتری ، چگالی بیشتری خواهند داشت ،کاملا در یک دسته مشخص شده اند و داده های غیرنرمال پراکندگی بیشتری از داده های عادی خواهند داشت .

مطابق نمودار1 ، ناحیه N1 و N2 رفتار نرمال را نشان می دهند و نقاط O1 و O2 نیز رفتارهای غیر عادی و ناهنجاری را نمایان می سازند.



نمودار 1.چگونگی دسته بندی داده ها براساس رفتارعادی

روش تشخیص ناهنجاری بر خلاف روش تشخیص سوء استفاده ، مبتنی بر راهکارها و امضاهای از پیش شناخته شده ای نیست بلکه سازوکار آن مبتنی بر تجربه و تحلیل رفتار مشتریان می باشد .

[1] Anomaly Detection
[2] Profile
[3] Deviation

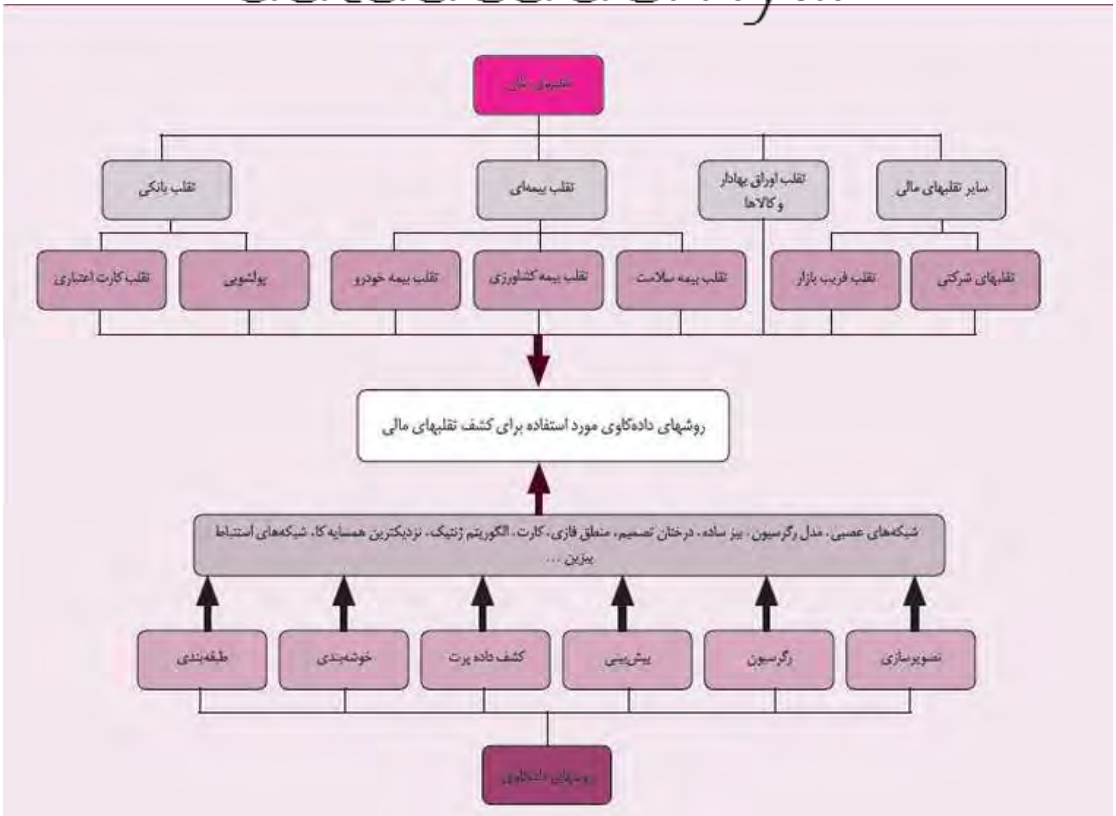
به این گونه که رفتار و تاریخچه عملکرد مشتری و تراکنش های وی مورد تجزیه و تحلیل قرار می گیرد و در صورتی که تراکنش جدیدی از سمت مشتری صادر شود، به نحوی که با تاریخچه عملکرد وی متفاوت باشد ، این تراکنش می تواند به عنوان یک تقلب شناسایی گردد.گرچه این سازوکار ،بخش بزرگ تری از تقلب ها را پوشش می دهد و از بابتی می تواند یک مزیت نسبت به روش تشخیص سوء استفاده تلقی شود ، لیکن به دلیل آنکه هر انحرافی را می تواند به عنوان یک حمله شناسایی کند ، دقت پایینی دارد و چه بسا بسیاری از رفتارهای عادی مشتریان را نیز می تواند به عنوان یک تقلب تلقی نماید .

4–روشهای داده کاوی در مورد استفاده در تحقیقات کشف تقلبهای مالی

برای طبقات / وظایف مختلف داده کاوی یادشده در بالا ، روشهای بسیاری ایجاد شده اند که از علوم محتلفی مانند هوش مصنوعی ، الگوشناسی[1] ، یادگیری ماشینی، آمار برگرفته شده اند.در تحقیقات مختلف ،26 روش داده کاوی در کشف تقلبهای مالی به کار رفته اند[1].

شکل2 ، در یک تقسیم بندی کلی ، طبقات تقلب مالی و 6 گروه کاربرد داده کاوی مورد استفاده برای کشف این تقلبها را نشان می دهد.

در جدول 4، می توان انواع مختلف روشهای داده کاوی مورد استفاده برای کشف تقلبهای شزکتی (شامل تقلب در صورت های مالی) را مشاهده کرد.پرکاربردترین روشهای مورد استفاده برای کشف تقلبهای مالی عبارتند از مدلهای رگرسیون لجستیک (رایج ترین) ،شبکه های عصبی ، شبکه استنباط بیزین[2] و درختان تصمیم که همه آنها راه حل های با اهمیی برای مشکلاتی ذاتی در کشف و طبقه بندی داده های متقلبانه ارائه می کند[1].



شکل 2.روشهای داده کاوی استفاده شده برای کشف انواع تقلبهای مالی (Ngai et,al.2010)

[1] Pattern Recognition
[2] The Bayesian Belief Network

طبقات تقلب مالی	فعالیتهای متقلبانه	نوع داده‌کاوی مورد استفاده	روشهای داده‌کاوی
سایر تقلبهای مالی	تقلبهای شرکتی	طبقه‌بندی	شبکه‌های عصبی، درختان تصمیم، شبکه استنباط عصبی، **کمک تصمیم چندمعیاره**[20]، (UTADIS)[21]، **الگوریتم‌های تکاملی**[22]، منطق فازی، مدل لجستیک، کارت، نزدیکترین همسایه کا، (RIPPER)[23]، ماشین بردار پشتیبان، **روش نقطه جدایی**[24]
		خوشه‌بندی	بیز ساده
		پیش‌بینی	شبکه‌های عصبی
		رگرسیون	مدل لجستیک

جدول 4.اهداف اصلی تحقیقتهای انجام شده در مورد کشف تقلبهای شزکتی از سال 1997 تا 2008 (Ngai et.al ,2010)

این 4 روش ، همگی در گروه (طبقه بندی)قرار می گیرند که در ادامه به شکل مشروحتری مورد بحث قرار گرفته اند.

4–1–مدل رگرسیون

در ادبیات پژوهشهای داده کاوی برای کشف تقلب ،رگرسیون رایج ترین روش مورد استفاده است .مدلهای رگرسیون استفاده شده عبارتند از لوجیت (LOGIT) ،لجستیک گام به گام ، روش کمک تصمیم چند متغیره و بتا 2 تعمیم یافته نمایی (EGB2)[22].

مدل لجستیک ، رایج ترین مدل مورد استفاده است.مدل لجستیک ، یک مدل خطی تعمیم یافته[1] است که برای رگرسیون دوگانه ای استفاده می شود که در آن متغیرهای پیش بینی کننده می توانند کمی یا کیفی باشند.این مدل اساسا برای حل مسائل مطرح در تقلب بیمه خودرو و تقلبهای شرکتی استفاده می شود[1].

ایده پشتیوانه رگرسیون این است که با استفاده از نسبتهای مالی شرکتها ، مدلی به دست آید تا مسخص شود کدام نسبتها با صورتهای متقلبانه و صورتهای مالی غیر متقلبانه ، می توان فهمید که کدان عوامل به شکل معنیداری بر شرکتهای دارای صورتهای مالی متقلبانه اثر می گذارد و سپس می توان بر این اساس معادله را صورت بندی کرد .مدل ، بر مبنای نسبتهای صورتهای مالی که در مرحله آموزش به عنوان نشانگرهای تقلب مستند شده اند،شرکتها را به گروه های متقلبانه و غیر متقلبانه طبقه بندی خواهد کرد [22].

4–2–شبکه های عصبی مصنوعی

شبکه عصبی روشی است که با استفاده از مجموعه ای از گره های به هم مرتبط ، از کارکرد مغز انسان تقلید می کند.این روش مبتنی است بر مدلهای رایانه ای از نورونهای زیستی .یک شبکه عصبی چند لایه در برگیرنده تعداد زیادی واحد (نورون) به هم مرتبط در الگویی از ارتباطات است[23].

این روش به شکل گسترده ای در طیقه بندی و خوشه بندی استفاده شده است و پس از رگرسیون ، پرکاربردترین روش داده کاوی مورد استفاده در کشف تقلبهای مالی است[22].

[1] Generalized Linear Model

نخست با استفاده از مجموعه ای از داده های زوجی برای ترسیم ورودیها و خروجیها آموزش داده می شود .سپس وزن ارتباطات بین نورون ها تثبیت می شود و شبکه برای تعیین طبقه بندی های مجموعه ای جدید از داده ها مورد استفاده قرار می گیرد[23].مزایای این روش از قرارند ،نخست اینکه این روش انطباق پذیر است .دوم اینکه این روش ،مدلهای دارای پایانی[1] ایجاد می کند و سوم اینکه اگر وزنهای آموزشی تغییر کنند ، فرایند طبقه بندی را نیز می توان اصلاح کرد .شبکه های عصبی بیشتر برای تقلبهای کارت اعتباری ،بیمه خودرو تقلب های شرکتی به کار می روند[1].

چن و دو با استفاده از شبکه های عصبی مصنوعی 68 شرکت فعال در بورس تایوان را مطالعه قرار دادند . آنان با استفاده از داده های مالی و غیر مالی ، یک مدل بحران مالی تدوین کردند .نتایج مطالعه آنان نشان می دهد که شبکه های عصبی مصنوعی بهتر از روشهای سنتی آماری ،بحران مالی را پیش بینی می کنند.

4–3–شبکه استنباط بیزین

شبکه استنباط بیزین نشاندهنده مجموعه ای از متغیرهای تصادفی و استقلال مشروط آنها با استفاده از یک نمودار غیر چرخه ای هدایت شده[2] است که در آن گره ها نشاندهنده متغیرهای تصادفی اند و استقلال مشروط بین متغیرها را تعیین می کند [21].

شبکه استنباط بیزین ، اغلب در کشف تقلب کارت اعتباری ،بیمه خودرو ،و تقلبهای شرکتی مورد استفاده قرار می گیرد[1].

4–4–درختان تصمیم

درختان تصمیم ، ابزار پشتیبان تصمیم پیش بینی کننده ای هستند که تصویری از مشاهدات برای پیامدهای ممکن را ایجاد می کنند[19].درختان تصمیم ، درختانی هستند که موضوعها را براساس مقادیر صفتها طبقه بندی می کنند.برگ نماد پیش بینی ها هستند ، هر گره در یک درخت تصمیم نماینده یک صفت در یک موضوع مورد طبقه بندی است و هر شاخه نماینده مقداری است که یک گره می تواند اختیاز کند و در واقع اشتراک ویژگیها را نشان می دهد [24].

می توان از طریق الگوریتم های مبتنی بر یادگیری ماشینی از قبیل کارت[3](CART)، آی دی تری[4] و الگوریتم سی 4/5[5] (C 4.5) ،این درختان را کاشت.درختان تصمیم به طور معمول در تقلب کارت اعتباری ،بیمه خودرو و تقلبهای شرکتی استفاده می شوند[1].

کرکاس و همکاران در مطالعه خود همزمان سه روش را به کار بردند که عبارت بودند از شبکه عصبی ، درخت تصمیم و بیزین .مطالعه آنان سودمندی این مدلها را شناسایی صورتهای مالی متقلبانه بررسی و مقایسه می کند .بردارهای ورودی [6] ، از نسبتهای مالی استخراج شده از صورتهای مالی تشکیل شده است.این سه مدل از جهت عملکردشان مقایسه شده اند .نمونه آنان از 76 شرکت تولیدی یونانی تشکیل شده بود که 38 شرکت به عنوان متقلب و 38 شرکت به عنوان غیر متقلب طبقه بندی شدند.معیار طبقه بندی به عنوان متقلب ، به طور عمده گزارشهای حسابرسان و مقامات مالیاتی نسبت به تلاش شرکت برای فرار مالیاتی با انجام دستکاری های با اهمیت در صورتهای مالی ،قرار گرفتن در فهرست شرکتهای تحت نظارت در بورس آتن ، تعلیق معاملات سهام شرکت به دلایل مرتبط با دستکاری داده های مالی شرکت و وجود پروندهایی در دادگاه مرتبط با موضوع صورتهای مالی

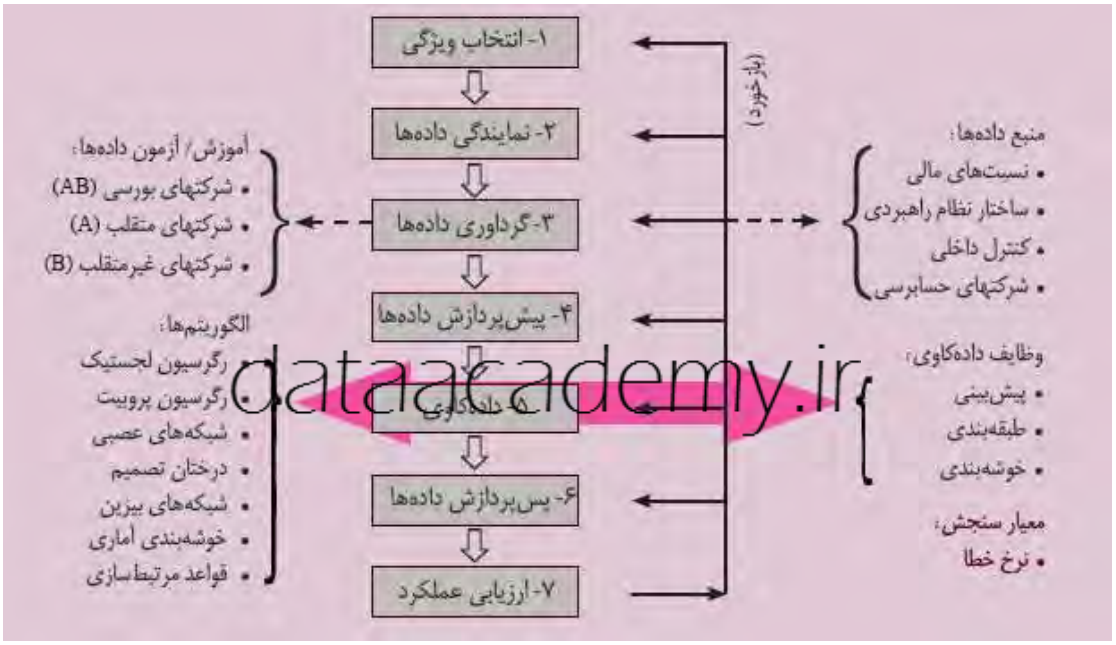
[1] Robust Models
[2] Directed Acyclic Graph (DAG)
[3] Classification and Regression Trees (CART)
[4] Iterative Dichotomizer3 (ID3)

[6] Input Vector

متقلبانه ، از مواردی هستند که به عنوان نشانه های تقلب یک شرکت در نظر گرفته شده اند.کرکاس و همکاران (2007)گزارش کرده اند که پس از آموزش مدلها ، در مرحله آزمون تقسیم نمونه ،مدل در خت تصمیم با 96/2 در صد دقت ، مدل شبکه استنباط بیزین با 94/7 درصد دقت توانسته اند شرکتهای متقلب را از شرکتهای غیر متقلب بازشناسی کنند[21].

4-5-یک چارچوب کلی برای الگورتیم های داده کاوی

هرچند الگوریتم های داده کاوی بسیاری برای کشف تقلب مورد استفاده قرار گرفته اند ،اما کاربرد آنها ،همچنان از الگوی سنتی داده کاوی انتخاب ویژگی نمایندگی[1] گرد آوری و مدیریت داده ها ، پیش پردازش ،داده کاوی ،پس پردازش و ارزیابی عملکرد پیروی می کند . یوئه و همکاران (2007)،ویژگیهای فنون داده کاوی مورد استفاده برای هدف خاص کشف تقلبهای مالی را در یک چارچوب کلی خلاصه کرده اند (شکل 5).



شکل 5. چارچوب کلی کشف تقلبهای مالی با استفاده از کارایی (Yue et,al .2007)

بر مبنای توزیع داده ها ، الگوریتم های کشف تقلبهای مالی را نخست می توان به دوگروه عمده تقسیم کـرد؛ داده های گردآوری شده از شرکتهای متقلب و غیر متقلب و همچنـین ،داده هـای حسابرسـی .در پژوهشـهای گذشـته ،تمرکز بیشتر بر تلاش برای کشف تقلب در مجموعه ای از داده های متقلبانه و غیر متقلبانه بوده است.

5-راه آینده چالشهای پیش رو

نگای و همکاران اظهار می دارند که یک دلیل برای محدود بودن تعداد مقاله های مرتبط با موضوع کشف تقلبهای مالی (49 تا بین سالهای 1997 تا 2008)،سختی به دست آوردن داده های تحقیق مناسب است .مشکل آن است که پیش از هر کاری ، و برای شروع آموزش مدل ،باید مجموعه ای از صورتهای مالی را به دو گروه متقلب و غیرمتقلب تقسیم کرد.چالش شناسایی صورتهای مالی متقلبانه ،موانع بسیاری در سر راه تحقیقات کشف تقلب مالی قرار می دهد . هرچند روشهای داده کاوی ذکر شده در بالا عموما نشان داده اند که در کشف تقلب صورتهای مالی اثر بخشی بوده اند ، اما کاربرد آنها برای کشف تقلب در صورتهای مالی ،معایب و محدودیت های کاربردی بسیاری داشته است .در پس عمده روشهای داده کاوی موجود برای کشف تقلب در صورتهای مالی ، دامنه کاربرد خاص و محدودیت های ویژه ای وجود دارند[25]. برای مثال ، هرچند که این

[1] Representation

روش ها به خوبی برای مدل سازی پیش بینی کننده توسعه یافته اند،اما آنها برای ارزیابی اثر به خوبی توسعه پیدا نکرده اند . به طور مشخص ، هنوز برای برخی از روشهای داده کاوی آمارهای آزمونی ساخته نشده است که با آن بتوان به ارزیابی اثرهای متغییرهای مستقل بر متغیرهای وابسته پرداخت[25].

یک نکته دیگر که باید به آن توجه شود، این است که اغلب روشهای داده کاوی نقاط پرت را به عنوان استثنا یا اختلال[1] کنار می گذارند؛در حالی که در کشف تقلب ، رویدادهای نادر می توانند جالب تر از رویدادهای معمول و مکرر باشند.بنابراین ،تحلیل نقاط پرت برای کشف الگوهای متقلبانه باید بیش از پیش مورد توجه قرار گیرد.البته نبود تحیقیقاتی در مورد کاربرد روشهای کشف داده های پرت برای کشف تقلبهای مالی ممکن است به خاطر سختی کشف داده های پرت باشد.در واقع کشف داده های پرت وظیفه پیچیده ای است که بی شباهت به جستن سوزن در انبار کاه نیست .برخلاف دیگر روشهای داده کاوی ، روشهای کشف داده پرت متمرکز بر یافتن الگوهای نادر مرتبط با اشیایی داده ، بسیار اندکند[20].

همچنین روشهای تصویرسازی نیز توانایی درخور توجه در شناسایی و ارائه بی قاعدگی ها در دادها دارند.این ویژگی می تواند شناسایی و کمی سازی طرحهای تقلب را بسیار آسانتر کند[1].

سخن آخر اینکه در زمان کنونی ،تقلبهای مالی همواره در حال تغییر شکل و تکامل هستند؛ پس سازوکارهای ماشینی کشف تقلب نیز باید با استفاده از آگاهی های تخصصی در دسترس ،اثر بخشی و کارایی خود را به بطور مستمر افزایش دهند .همان گونه که ژو و کاپور به خوبی تذکر می دهند ، کشف تقلب مالی با استفاده از روشهای کشف فعلی ، به طور روز افزون مشکل می شود .یک مدیر عامل آگاه به همه مسائل که اراده کرده است جرمی مرتکب شود،متابع کافی برای دور زدن سیستم را به راحتی در اختیاز دارد و قادر است که هر نوع سازوکار کشفی را خنثی کند.ژو و کاپور روشهای کشف تقلب مالی مبتنی بر داده کاوی (مانند رگرسیون ، درخت تصمیم ،شبکه های عصبی ،و شبکه های بنزین)را مورد بررسی قرار داده اند .آنان به ویژه ، اثر بخشی و محدودیت های این روشهای داده کاوی را در هنگام پدید آمدن شگردهای جدید تقلب صورتهای مالی که خود را با این روشهای کشف انطباق داده اند ، به نقد کشیده اند .نویسندگان سپس یک روش نوین را پیشنهاد می کنند ؛یک برنامه کشف فعال که پیش از متقلبان بالقوه تکامل می یابد .توانمند کردن یک سیستم کشف هوشمند برای پیش بینی ،پیش از اینکه هر گونه تقلب ناشناخته ای در آینده اتفاق افتد.این توان را به وجود می آورد که انواع جدید تقلبهای صورتهای مالی به طور اثر بخش کشف گردند .البته چنانکه این دو نویسنده خود ادغان می دارند ، تحقیقهای بیشتری در آینده نیاز است تا برنامه کشف فعالی طراحی شود که هم اثر بخش و هم کارا باشد.

6–نتیجه گیری

از آنچه که در بررسی های به عمل آمده در خصوص شناسایی تقلب در روش های تشخیص سوء استفاده و تشخیص ناهنجاری بیان شد،این نکته استنتاج می گردد که تکنیک های مبتنی بر رویکرد تشخیص سوء استفاده زمانی به کار گرفته می شوند که تشخیص تقلب به صورت از پیش شناخته شده بوده و براساس امضای می توان رفتار جاری مشتریان را برسی نمود طبیعا به دلیل شناخت کامل رفتار قبلی مشتریان ،دقت شناسایی تقلب در این روش بسیار بالاست .اما نقطه ضعف این روش ها ، عدم پوشش دهی کامل محدوده تقلب می باشد،بدین معنی که فقط و فقط تقلب هایی شناسایی و کنترل می شوند که حداقل یک بار رخ داده و یا امضای آن به سیستم تشخیص تقلب ارائه شده باشد.

[1] Noise

اما در مقابل ، رویکردهای مبتنی بر تشخیص ناهنجاری ، سعی در پیش بینی رفتار آتی مشتری داشته و با منظور ، تاریخچه رفتار وی را مورد بررسی قرار می دهند .در این گونه روش ها ،هیچ قاعده ثابتی جهت تعریف نمی شود ،بلکه رفتار عادی و نرمال مشتری به سیستم تشخیص تقلب آموخته شده و هر گونه انحراف از آن ، به معنی تقلب فرض می گردد.

به منظور شناسایی رفتار عادی مشتریان نیز از تاریخچه تراکنش های مشتری استفاده شده و رفتار عادی وی تلقی می گردد.این روش نسبت به رویکرد سوء استفاده ، دقت بالایی ندارد و ممکن است تراکنش های عادی به صورت تقلب فرض شوند . لذا دقت این روش نسبت به رویکرد سوء استفاده بسیار کمتر می باشد اما مزیت این روش این است که گستره بیشتری از حملات و تراکنش های غیرقانونی را پوشش می دهد و امکان پیش بینی تقلب های مشاهده نشده از مزیت های این روش است .

عموما در سیستم های تجاری ،سیستم های ترکیبی که شامل هر دو روش تشخیص سوء استفاده و ناهنجاری باشد،بهترین نتیجه را از نظر عملکردی در پی دارد .لذا سیستم های پیاده سازی شده به گونه ای طراحی شده است که با ترکیب این دو رویکرد علاوه بر دقت بالا،امکان پیش بینی رفتار مشتریان را نیز داشته باشند و تقلب های ناشی از رفتار غیر نرمال را نیز شناسایی نمایند.

فارغ از بحث فنی ،ذکر این نکته نیز در اینجا بسیار ضروری به نظر می رسد که با توجه به رشد روز افزون حدمات مالی بانک ها و موسسات مالی و اعتباری به صورت الکترونیکی در سطح کشور و افزایش ضریب نفوذ استفاده کاربران از خدمات بانکداری الکترونیک ؛ رویکرد کلاهبرداران و متقلبان به سمت بانکداری الکترونیک نیز رو به افزایش است . بدین ترتیب نگرانی های بسیاری را سبب شده و توجه زیادی را به سوی خود جلب کرده است . البته حوزه کشف تقلب مالی نیز تحول هایی چشمگیری را شاهد بوده است . به طور مشخص ،داده کاوی نظرها را به شکل گسترده ای به خود جلب کرده است و محبوبیت فزاینده ای در جهان مالی به دست آورده است .کاربرد های موفقیت آمیزی از داده کاوی گزارش شده است و تحقیقات نشان داده اند که داده کاوی در میزان کاربرد و اثر بخشی گسترش یافته است .سازمان های حرفه ای حسابداری نیز داده کاوی را به عنوان یک فناوری مهم برای سده جدید شناخته اند [25].روش های اصلی مورد استفاده برای کشف تقلبهای مالی عبارتند از مدلهای رگرسیون لجستیک ، شبکه های عصبی ، شبکه استنباط بیزین و درختان تصمیم که همه آنها راه حل های با اهمیتی را برای مشکلات ذاتی در کشف و طبقه بندی داده های متقلبانه ارائه می کنند.

کاربرد روشهای داده کاوی بر روی نسبتهای مالی استخراج شده از صورتهای مالی شرکتهای و نیز دیگر اطلاعات در دسترس ، می تواند به حسابرسان در کشف تقلب کمک کند ؛به طوری که آنان می تواند از نتایج این تحلیل ها به عنوان یک علامت اولیه هشداردهنده نسبت به وقوع احتمالی تقلب صورتهای مالی استفاده کند.نشانگرهای تقلب در صورتهای مالی ،اثری با اهمیت بر تعیین تقلب صورتهای مالی دارد.

همچنین ،انواع تقلب و الگوهای تقلب در صنایع مختلف در طول زمان تغییر کرده است .درک اینکه طرحهای تقلب چگونه متحول شده اند مهم است .همچنین ،پیش بینی جهت تغییر این تقلبها با هر وسیله ممکن و به روز نگاه داشتن روشهای ماشینی کشف تقلب اهمیت دارد.پژوهش در این راستا ممکن است نتایج با اهمیتی داشته باشد که برای تدوین فرایندهای تجاری قویتر و نیز سازوکار های کشف تقلب انطباق پذیر برای مدیریت/پیشگیری /کشف خطر تقلب ،سودمند باشد.

7–مراجع

[1] Ngai E.W.T., Yong Hu, Y.H. Wong, Yijun Chen, Xin Sun, The Application of Data Mining Techniques in Financial Fraud Detection: A Classification Framework and an Academic Review of literature; Decision Support Systems, vol. 50(3), 2010, pp. 559-569.

[2] Wang J., Y. Liao, T. Tsai, G. Hung, Technology-based Financial Frauds in Taiwan: Issue and Approaches, IEEE Conference on: Systems, Man and Cyberspace, 2006, pp. 1120–1124.

[3] Clifton Phua and et al. (2005).A Comprehensive Survey of Data Mining –based Fraud Detection Research, from http://www.arxiv.org/pdf/1009.6119.

[4] Spafford, S. Kumar & Eugene. H. (1994).A Pattern Matching Model for Misuse Intrusion Detection. 17th Notional Computer Security Conference, From www. docs. lib.purdue,edu.

[5] Winslow, E. & Smaha, J.(1994). Misuse Detection Tools. Computer Security Journal, 3, 39 – 49

[6] Stephan, Kovach & Wilson, Vicente Ruggiero.(2011).Online Banking Fraud Detection Based on Local and Global Behavior. ICDS, from www.thinkmind.org/ download php ? Article = icds-2011-6-40.

[7] Ghosh, A.K., Schwartzbard, A & Sehatz, M. (1999). A Study in Using Neural Networks for Anomaly and Misuse Detection. 8 th USENIX Security Symposium, from . www. acm.org/citation.cfm?d=1251433.

[8] Lunt, T.F and et al. (1990). A Real-Time intrusion Detection Expert System (IDES) - Final Technical Report. Technical Report SRI Computer Science Laboratory, SRI International, from http://www.wenke.gtisc.gatech.edu.

[9] Anderson, D., Frivold,T., Tamaru, A & Valdes,A.(1994). Next generation intrusion detection expert system (NDES). software user's manual,beta-update release. Technical Report SRIXSL-9547. Computer Science Laboratory, SRI International, from www.thc.org/root/ docs/intrusion-detection/...NIDES-summary.pdf.

[10] Ghosh, S and et al.(1994).Credit card fraud detection with a neural-network. 27th Annual Hawaii International Conference on System Science. Los Alami, CA: IEEE Computer Society.

[11] Brause, R., Langsdorf, T., Hepp, M.(1999). Credit Card Fraud Detection by Adaptive Neural Data Mining. 11 th IEEE International Conference on Tools with Artificial Intelligence. (pp.103-106). Los Alami, CA: IEEE International Conference on tools with Artificial Intelligence.

[12] Hassibi, K. (Ed.).(2000). Detecting Payment Card Fraud With Neural Networks. Singopore: World Scientific.

[13] Dorronsoro., J, Ginel, E & Sanchez, C.(1997). Neural Fraud Detection in Credit Card Operations, From http:// www. ieeexplore.ieee.org

[14] Ryan,J., Lin, M.J & Miikkulainen,R.(1998).Intrusion detection with neural networks. in M. J. Keams, and S. A. Solla M. I. Jordan. (Eds.) Advances in Neural Information Pmcessing Systems. Cambridge: The MIT Press.

[15] Burge,P and et al.(1999). Fraud Detection and Management in MobileTelecommunicationsNetworks. London:Royal Holloway University.

[16] Ilgun, K.(1993). USTAT A Real-time intrusion detection system for UNIX. IEEE Symposium on Research in Security and Privacy. (pp.16-28). Oakland, CA: IEEE Symposium on Research in Securiry and Privacy.

[17] Chittur, A.(2001).Model Generation for an Intrusion Detection System Using Genetic Algorithms. Ossining High school Honors Thesis.

[18] Stolfo, W.L & San, S.(1998). Data Mining Approaches for Intrusion Detection. TX 7th USENIX Security Symposium. Antonio, TX: USENIX Security Symposium.

[19] Turban E., J.E. Aronson, T.P. Liang, R. Sharda, Decision Support and Business Intelligence Systems, Eighth ed,Pearson Education, 2007.

[20] Han J., M. Kamber, Data Mining: Concepts and Techniques (Second ed), Morgan Kaufmann Publishers, 2006, pp. 285–464.

[21] Zhang D. and L. Zhou, Discovering Golden Nuggets: Data Mining in Financial Application, IEEE Transactions on Systems, Man and Cybernetics, Vol. 34(4), 2004 pp.513-522.

[22] Kerkaus E., C. Spathis, Y. Manolopoulos, Data Mining Techniques for the Detection of Fraudulent Financial Statements, Expert Systems with Applications, Vol.32,

[23] Yue D., X. Wu, Y. Wang, Y. Li and C. Chu, A Review of Data Mining-based Financial Fraud Detection Research, International Conference on Wireless Communications, Networking and Mobile Computing, 2007, pp.5519–5522.

[24] Yamanishi K., J. Takeuchi, G. Williams and P. Milne , On-Line Unsupervised Outlier Detection Using Finite Mixtures with Discounting Learning Algorithms, Data Mining and Knowledge Discovery, Vol. 8, 2004, pp.

[25] Phua C., V. Lee, K. Smith, R. Gayler, A Comprehensive Survey of Data Mining-based Fraud Detection research, Clayton School of Information Technology, Monash University, 2005

[26] Zhou W., G. Kapoor, Detecting Evolutionary FinancialStatement Fraud, Decision Support Systems, Vol. 50(3), 2011, pp. 570-576