به نام آنکه جانرا فکرت آموخت*چراغ دل به نور جان برافروخت

بانک مرکزی جمهوری اسلامی ایران

دوره آموزشی :
کشف تقلب در بانک ها

مدرس :
سید مرتضی ذکاوت
پائیز ۱۴۰۱

فهرست مطالب

۱. فصل اول :تشخیص، پیشگیری و تجزیه و تحلیل تقلب

۱-۱. مقدمه

این فصل اول، ابتدا با تعریف و توصیف تقلب شروع می شود و انواع مختلف تقلب را مورد بحث قرار می گیرد. در مرحله بعد، کشف و پیشگیری از تقلب به عنوان وسیله ای برای رسیدگی و محدود کردن میزان و تأثیر کلی تقلب مورد بحث قرار می گیرد. کلان داده ها و تجزیه و تحلیل ها ابزارهای قدرتمندی را ارائه می دهند که ممکن است سیستم تشخیص تقلب سازمان را بهبود بخشد. ما به تفصیل درباره چگونگی و چرایی این ابزارها مکمل رویکردهای سنتی کشف تقلب مبتنی بر متخصصین صحبت می کنیم. متعاقباً، مدل فرآیند تجزیه و تحلیل تقلب معرفی می‌شود که یک نمای کلی در سطح بالا از مراحلی که در توسعه و پیاده‌سازی یک سیستم تشخیص تقلب مبتنی بر داده دنبال می‌شود، ارائه می‌کند. این فصل با بحث در مورد ویژگی‌ها و مهارت‌های یک دانشمند داده کلاهبرداری خوب به پایان می‌رسد. ، به دنبال یک دیدگاه علمی در
موضوع.

۱-۲. تعریف و مصادیق تقلب

در این قسمت ابتدا به تعریف تقلب و مصادیق آن می پردازیم:
از آنجایی که یک بحث یا تحقیق کامل مستلزم تعاریف روشن و دقیق از موضوع مورد علاقه است، این بخش ابتدا با تعریف تقلب و با برجسته کردن تعدادی از ویژگی های اساسی آغاز می شود. متعاقباً، یک مدل مفهومی توضیحی معرفی می‌شود که بینش عمیق‌تری را در محرک‌های اصلی متقلبان و کلاهبرداران، یعنی افرادی که مرتکب تقلب وکلاهبرداری می‌شوند، ارائه می‌کند. بینش در زمینه کاربرد - یا به عبارت دیگر، دانش تخصصی - برای استفاده موفقیت آمیز تجزیه و تحلیل در هر محیطی بسیار مهم است و در نهایت به اندازه مهارت فنی اهمیت دارد. دانش یا بینش متخصص در مورد مشکل پیش رو به تحلیلگر کمک می کند تا اطلاعات مناسب را به شیوه ای درست جمع آوری و پردازش کند و داده ها را سفارشی کند تا تکنیک های تحلیلی به بهترین نحو ممکن در تشخیص تقلب عمل کنند. فرهنگ لغت آکسفورد تقلب را چنین تعریف می کند[1]:

فریبکاری مجرمانه با هدف منفعت مالی یا شخصی.

علی رغم اینکه این تعریف ماهیت کلاهبرداری را در بر می گیرد و اشکال و انواع مختلفی از کلاهبرداری را پوشش می دهد. لیکن بسیار دقیق نیست چرا خصوصیات دقیق تقلب و چگونگی آنرا توضیح نمی دهد و لذا از

[1] Baesens et al. 2015

آن الزامات یک سیستم تشخیص تقلب را استنباط نمود. شایان ذکر است که تقلب قطعاً یک پدیده مربوطه دوران جدید نیست که منحصر به جامعه مدرن باشد و حتی منحصر به نوع بشر نیست. گونه‌های جانوری نیز در فعالیت‌هایی شرکت می‌کنند که می‌توان آن‌ها را متقلبانه نامید، اگرچه شاید بهتر باشد رفتارهای حیواناتی مانند آفتاب‌پرست را به جای فعالیت‌های متقلبانه ، به‌عنوان رفتارهای دستکاری طبقه‌بندی کنیم، رفتارهای نادرست و مجرمانه صرفاً برای انسان اطلاق می شود در مورد حیوانات صادق نیست. در واقع، غیرقانونی یا مجرمانه بودن فعالیت‌ها به قوانین یا قوانین قابل اجرا بستگی دارد، که به طور صریح و رسمی این دسته‌بندی‌ها را تعریف می‌کند تا بتوان رفتار را به عنوان متقلبانه طبقه‌بندی کرد. توصیف دقیق تر و دقیق تری از پدیده چند وجهی تقلب توسط ون ولاسلر و همکاران (۲۰۱۵)ارائه شده است.:

کلاهبرداری یک جرم غیر معمول و بخوبی طراحی و سنجیده شده و اغلب سازماندهی شده ایست که به طور نامحسوس و پنهان، در حال تکامل در انواع مختلفی ظاهر می شود.

این تعریف پنج ویژگی را برجسته می‌کند که با چالش‌های خاص مربوط به توسعه یک سیستم کشف تقلب، که در این نوشتار مطرح می شود، مرتبط است. اولین ویژگی و چالش که مورد تاکید است این واقعیت است کهد که تقلب غیر پدیده ای غیر معمول است. مستقل از تنظیمات یا کاربرد دقیق، فقط تعداد کمی از جمعیت درگیر پرونده‌ها معمولاً به کلاهبرداری مربوط می‌شوند، علاوه بر این، فقط تعداد محدودی از آنها به تقلب شناخته می‌شوند. این امر تشخیص تقلب را دشوار می‌کند زیرا موارد کلاهبرداری توسط موارد غیر متقلبانه پوشش داده می‌شود، و به دلیل اینکه نمونه‌های کمی در دسترس است ،یادگیری از پرونده‌های تاریخی برای ساختن یک سیستم تشخیص تقلب قدرتمند را بسیار سخت می کند.

بواقع متقلبان و کلاهبرداران سعی می‌کنند رفتار خود را با رفتار افراد عادی ترکیب کنند و رفتاری متفاوت از دیگران نداشته باشند تا مورد توجه قرار نگیرند. ارتکاب تقلب قطعاً لحظه ای و بدون برنامه نیست، زیرا اگر چنین بود، تشخیص آن بسیار آسان‌تر بود. متقلبان مرتباً روش‌های خود را تطبیق داده و اصلاح می‌کنند تا شناسایی نشوند. سیستم های تشخیص تقلب بهبود می یابند و با مثال یاد می گیرند. بنابراین، تکنیک‌ها و ترفندهایی که کلاهبرداران به کار می‌گیرند به مرور زمان همراه با مکانیسم‌های کشف تقلب تکامل می‌یابند. این بازی موش و گربه بین متقلبان و مبارزان با کلاهبرداری و تقلب ممکن است یک بازی بی پایان به نظر برسد، اما تاکنون راه حل جایگزینی وجود ندارد. با اتخاذ و توسعه روش ها و سازوکارهای پیشرفته تشخیص و پیشگیری از تقلب سازمان ها موفق به کاهش زیان های ناشی از آن می شوند. کلاهبرداران، مانند سایر مجرمان، تمایل دارند به دنبال راه آسان باشند و به دنبال فرصت‌های ساده‌تر می گردند. بنابراین، مبارزه با تقلب با ساختن سیستم‌های تشخیص پیشرفته و قدرتمند، قطعاً تلاشی بی‌معنی نیست، اما مسلماً، به احتمال زیاد تلاشی بدون پایان است. کلاهبرداری

اغلب یک جنایت با دقت سازمان‌دهی شده است، به این معنی که کلاهبرداران اغلب به طور مستقل عمل نمی‌کنند و در بسیاری از موارد با سایر کلاهبرداران هماهنگی و ارتباط دارند و حتی ممکن است از روش های سایر متقلبان کپی برداری نمایند. علاوه بر این، چندین نوع روش تقلب و کلاهبرداری مانند پول شویی و کلاهبرداری چرخ و فلک[1] شامل ساختارهای پیچیده ای است که به منظور ارتکاب کلاهبرداری به صورت سازماندهی شده ایجاد شده است. این باعث می‌شود که کلاهبرداری یک رویداد مجزا نباشد و به همین دلیل برای شناسایی تقلب باید زمینه وقوع آن(به عنوان مثال، شبکه اجتماعی کلاهبرداران) در نظر گرفته شود. همانطور که در مطالعه موردی فرار مالیاتی شرکت توسط ون ولاسلر و همکاران (۲۰۱۵) صورت پذیرفت نشان می دهد[2] که شرکت های متقلب در واقع بیشتر با سایر شرکت های کلاهبردار مرتبط هستند تا با شرکت های غیر متقلب. تجزیه و تحلیل شبکه های اجتماعی برای کشف تقلب با استفاده هوشمندانه از اطلاعات متنی توصیف کننده شبکه یا محیط فعالیت یک فرد، ابزار قدرتمندی برای آشکارسازی کلاهبرداری باشد. یکی از عناصر دیگر در تعریف تقلب ارائه شده توسط ون ولاسلر و همکاران (۲۰۱۵) نشان دهنده انواع قالب هایی مختلفی است که در آن تقلب رخ می دهد. براین اساس تقلب با استفاده از مجموعه گسترده ای از تکنیک ها و رویکردها و همچنین به بسیاری از تنظیمات مختلف همچنین فعالیت های اقتصادی مستعد صورت می پذیرد. جدول۱ یک نمای کلی و توصیفی غیر جامع از تعدادی از انواع مهم کلاهبرداری ارائه می دهد.

جدول۱- ۱ شرح و مصادیق انواع تقلب ها

نوع تقلب	شرح و مصادیق
تقلب کارت اعتباری	در کلاهبرداری کارت اعتباری، برداشت غیرمجاز از اعتبار دیگری وجود دارد. برخی از انواع متداول کلاهبرداری کارت اعتباری عبارتند از: کارت های اعتباری جعلی (تعریف تقلبی، در زیر ارائه شده است)، استفاده از کارت های گم شده یا دزدیده شده، یا به دست آوردن اعتبار از طریق پست (تعریف ازdefinitions.uslegal.com). همانطور که توسط بولتون و هند (۲۰۰۲) توصیف شده است، می توان دو زیرگروه را شناسایی کرد: (۱) تقلب در برنامه، شامل افرادی است که با استفاده از اطلاعات شخصی جعلی، کارت‌های اعتباری جدید را از شرکت‌های صادرکننده دریافت می‌کنند و سپس تا حد امکان در مدت زمان کوتاهی هزینه می‌کنند. (۲) کلاهبرداری رفتاری، که در آن جزئیات کارت ها قانونی وجود دارد لیکن به طور تقلبی به دست آمده و فروش بر اساس "صاحب کارت حاضر نیست" انجام می شود. این لزوماً به سرقت کارت فیزیکی نیاز ندارد، فقط به سرقت اعتبار کارت نیاز

[1] carousel fraud

[2] Baesens et al. 2015

نوع تقلب	شرح و مصادیق
	دارد. کلاهبرداری رفتاری بیشتر به کلاهبرداری کارت اعتباری مربوط می شود هرچند کلاهبرداری از کارت بدهی رخ نیز می دهد، ولی کمتراست. کلاهبرداری از کارت اعتباری شکلی از سرقت هویت است که در زیر توضیح داده خواهد شد.
تقلب در بیمه	کلاهبرداری و تقلب در حوزه بیمه به طور گسترده در دسته بندی مربوط به هر نوع بیمه، هم از طرف خریدار یا فروشنده یک قرارداد بیمه می تواند صورت پذیرد. کلاهبرداری بیمه ای از سوی صادرکننده (فروشنده) شامل بیمه نامه های فروش از شرکت های غیرموجود، نقص(عمدی) در درج حق بیمه و نیز شرایط ابطال بیمه برای ایجاد کمیسیون بیشتر است. کلاهبرداری خریدار شامل ادعاهای اغراق آمیز (بیمه اموال: دریافت مبلغی بیشتر از ارزش اموال تخریب شده)، سابقه پزشکی جعلی (بیمه مراقبت های بهداشتی: صدمات جعلی)، بیمه نامه های تاریخ گذشته، مرگ جعلی، آدم ربایی یا قتل (کلاهبرداری بیمه عمر)، و خسارت جعلی[1] (صحنه سازی تصادف) .
فساد (مالی)[2]	فساد عبارت است از سوء استفاده از قدرت محول شده (توسط میراث، تحصیل، ازدواج، انتخاب، انتصاب یا هر چیز دیگری) برای کسب منافع شخصی. این تعریف شبیه به تعریف تقلب ارائه شده توسط فرهنگ لغت آکسفورد است که قبلاً به آن اشاره شد، زیرا هدف هردو منفعت شخصی است. لیکن از این جهت متفاوت است که بر استفاده نادرست از قدرت محول شده تمرکز دارد. این تعریف طیف وسیعی از انواع فرعی فساد را در بر می گیرد، بنابراین نه تنها فساد توسط یک سیاستمدار یا یک کارمند دولتی، بلکه مواردی همچون، توسط مدیر عامل یا مدیر مالی یک شرکت، سردفتر اسناد رسمی، رهبر تیم، مدیر یا مسئول پذیرش در یک مدرسه یا بیمارستان خصوصی، مربی یک تیم فوتبال، و غیره را نیز پوشش می دهد. (مأخذ www.corruptie.org).
جعل[3]	جعل تقلیدی از یک سند یا شی با ارزش است که قرار است با تقلب یا فریبکاری به عنوان واقعی معرفی شود. تقلب معمولاً در مورد اشیاء با ارزش، کارت های اعتباری، کارت های شناسایی، محصولات محبوب، پول و غیره صورت می پذیردمأخذ: www.dictionary.com
تقلب در گارانتی و ضمانت محصول[4]	گارانتی محصول نوعی ضمانت است که سازنده یا طرف مشابه در رابطه با وضعیت محصول خود می دهد و همچنین به شرایط و شرایطی اشاره دارد که در صورت عدم عملکرد محصول مطابق توضیحات اولیه، تعمیر یا تعویض و یا هراقدامی که در ضمانت آمده انجام می شود.

[1] staged collision
[2] corruption
[3] counterfeit
[4] product warranty fraud

نوع تقلب	شرح و مصادیق
	به طور مثال هنگامی که محصولی نتواند عملکردهای توصیف شده را ارائه دهد یا ویژگی ها یا رفتارهای انحرافی را نشان دهد که نتیجه فرآیند تولید است و نتیجه سوء استفاده مشتری نیست، وی می تواند از سازنده یا ارائه دهنده غرامت یا پاداش دریافت نماید. زمانی که شرایط محصول به دلیل استفاده مشتری از محصول تغییر کرده باشد، گارانتی اعمال نمی شود. ادعای عمدی نادرست غرامت یا پاداش بر اساس گارانتی محصول را تقلب در گارانتی محصول می نامند.
تقلب در مراقبت های بهداشتی[1]	کلاهبرداری در مراقبت های بهداشتی شامل ارائه ادعاهای غیر صادقانه مراقبت های بهداشتی به منظور کسب سود است. برخی از این طرح‌های متقلبانه عبارتند از: • افرادی که قرص‌های نسخه‌ای یارانه‌ای یا کاملاً تحت پوشش را دریافت می‌کنند که در واقع غیرضروری هستند و سپس آن‌ها را برای کسب سود در بازار سیاه می‌فروشند. • صورت‌حساب توسط پزشکان برای مراقبتی که هرگز انجام نداده‌اند. • ثبت ادعاهای تکراری برای همان خدمات ارائه شده؛ صورت‌حساب یک سرویس غیر تحت پوشش به عنوان یک سرویس تحت پوشش؛ اصلاح سوابق پزشکی و غیره اعضا می توانند با ارائه اطلاعات نادرست هنگام درخواست برنامه ها یا خدمات، جعل یا فروش داروهای نسخه ای، قرض دادن[2] کارت بیمه یا استفاده از کارت بیمه دیگران و غیره مرتکب کلاهبرداری در مراقبت های بهداشتی شوند (مأخذ: www.law.cornell.edu).
تقلب مخابراتی[3]	کلاهبرداری مخابراتی شامل سرقت خدمات مخابراتی (تلفن، تلفن همراه، رایانه و غیره) و یا استفاده از خدمات مخابراتی برای انجام سایر اشکال کلاهبرداری است (تعریف برگزیده از itlaw.wikia.com). یک مثال مهم به کلاهبرداری شبیه سازی (یعنی شبیه سازی یک شماره تلفن و اعتبار تماس مربوطه توسط یک کلاهبردار) مربوط می شود، که نمونه ای از کلاهبرداری ترکیبی[4] است که در آن استفاده متقلبانه بر استفاده مشروع از یک حساب اضافه می شود (فاوست و پروست ۱۹۹۷).
پول شوئی[5]	فرآیند برداشت عواید حاصل از فعالیت مجرمانه و قانونی جلوه دادن آنها. پول شویی به مجرمان اجازه می دهد تا سود به دست آمده غیرقانونی را به وجوه به ظاهر قانونی تبدیل کنند.

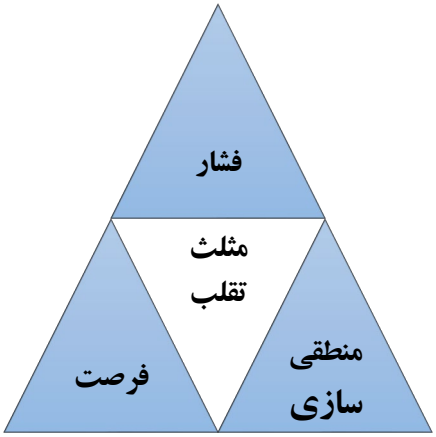
[1] healthcare fraud
[2] loaning
[3] telecommunications fraud
[4] superposition fraud
[5] money laundering

نوع تقلب	شرح و مصادیق
	این یک مشکل جهانی است و سالانه ۳۰۰ میلیارد دلار در ایالات متحده این فرآیند را طی می کند (مأخذ: legal-dictionary.thefreedictionary.com)[1].
کلاهبرداری کلیک[2]	کلاهبرداری کلیک یک عمل غیرقانونی است که زمانی اتفاق می‌افتد که افراد بر روی تبلیغات کلیکی یک وب‌سایت (اعم از تبلیغات بنری یا لینک‌های متنی پولی) کلیک می کنند تا تعداد کلیک‌های قابل پرداخت به تبلیغ کننده را افزایش دهند. کلیک‌های غیرقانونی می تواند با کلیک کردن دستی شخصی بر روی لینک‌های تبلیغاتی یا با استفاده از نرم‌افزارهای خودکار یا ربات‌های آنلاینی که برای کلیک بر روی این تبلیغات بنر و پیوندهای تبلیغات متنی با پرداخت به ازای کلیک برنامه‌ریزی شده‌اند انجام شود (مأخذ: www.webopedia.com)
سرقت هویت[3]	در سرقت هویت اطلاعات شخصی یا مالی شخص دیگری برای استفاده از هویت وی به منظور انجام معاملات یا خرید به سرقت می رود. برخی از سارقان هویت در سطل های بازیافت کامپیوتر ها به دنبال صورت حساب بانکی و کارت اعتباری می گردند. سایر روش‌های پیشرفته‌تر شامل دسترسی به پایگاه‌های اطلاعاتی شرکت‌ها برای سرقت فهرست‌های اطلاعات مشتری است (مأخذ: www.investopedia.com).
فرار مالیاتی[4]	فرار مالیاتی عمل یا عمل غیرقانونی عدم پرداخت بدهی مالیاتی است. در مشاغل، فرار مالیاتی می تواند در ارتباط با مالیات بر درآمد، مالیات شغل، مالیات فروش و مالیات غیر مستقیم و سایر مالیات های فدرال، ایالتی و محلی رخ دهد. نمونه‌هایی از اقداماتی که فرار مالیاتی تلقی می‌شوند عبارتند از گزارش نکردن آگاهانه درآمد یا گزارش کم درآمد (یعنی ادعای درآمد کمتر از آنچه واقعاً از یک منبع خاص دریافت کرده‌اید) (مأخذ: biztaxlaw.about.com).
سرقت ادبی[5]	دزدی ادبی توسط فرهنگ لغت آنلاین مریام وبستر به عنوان دزدیدن و به اشتراک گذاشتن (ایده ها یا کلمات دیگری) به عنوان متعلق به خود، استفاده از (تولید دیگری) بدون ذکر منبع، ارتکاب سرقت ادبی و ارائه یک ایده جدید و بدیع یا محصول مشتق شده از منبع موجود، تعریف شده است.. این امر شامل دزدیدن کار شخص دیگری و هم دروغ گفتن در مورد آن پس از آن است (مأخذ: از www.plagiarism.org).

در پایان، همانطور که در تعریف کلاهبرداری ارائه شده توسط فرهنگ لغت آکسفورد تأکید شده است، فعالیت های متقلبانه به منظورکسب به سود یا منافع برای کلاهبردار است. سود یا منفعت بالقوه، معمولاً پولی، در اکثر

[1] Baesens et al. 2015
[2] click fraud
[3] identity theft
[4] tax evasion
[5] plagiarism

موارد عامل اصلی ارتکاب کلاهبرداری است. به اصطلاح مثلث تقلب همانطور که در شکل ۱-۱ نشان داده شده است توضیح مفصل تری برای انگیزه ها یا محرک های اساسی ارائه می دهد.



شکل ۱- ۱ مثلث تقلب

مأخذ: (Baesens et al. 2015)

در پایان، همانطور که در تعریف کلاهبرداری ارائه شده توسط فرهنگ لغت آکسفورد تأکید شده است، فعالیت های متقلبانه منتج به سود یا منافع برای کلاهبردار است. سود یا منفعت بالقوه، معمولاً پولی، در اکثر موارد عامل اصلی ارتکاب کلاهبرداری است. به اصطلاح مثلث تقلب همانطور که در شکل ۱-۱ نشان داده شده است توضیح مفصل تری برای انگیزه ها یا محرک های اصلی برای ارتکاب کلاهبرداری شغلی ارائه می دهد. مثلث تقلب از فرضیه ای نشأت می گیرد که توسط دونالد آر. کرسی در کتابش در سال ۱۹۵۳ با نام «پول دیگران: مطالعه روانشناسی اجتماعی اختلاس» مطرح شد:

افراد معتمد زمانی به ناقض اعتماد تبدیل می شوند که خود را دارای مشکل مالی غیرقابل اشتراک تصور کنند، بدانند که این مشکل را می توان مخفیانه با نقض جایگاه اعتماد مالی حل کرد، و می توانند در آن موقعیت به رفتار خود اعمال کنند. استفاده از درازگویی و زبان شفاهی آنها را قادر می سازد تصورات خود را از خود به عنوان افراد مورد اعتماد با تصورات خود از خود به عنوان استفاده کنندگان از وجوه مورد اعتماد تطبیق دهند.

این مدل مفهومی اساسی عواملی را توضیح می‌دهد که با هم باعث می‌شوند یا محرک‌های یک فرد برای ارتکاب کلاهبرداری شغلی را توضیح می‌دهند، اما بینش مفیدی را در مورد پدیده کلاهبرداری از دیدگاه گسترده‌تری نیز ارائه می‌دهد. این مدل دارای سه پایه است که با هم رفتار متقلبانه را ایجاد می کنند:

۱. **فشار** اولین قدم است و به انگیزه اصلی مربوط می شود

ارتکاب کلاهبرداری فردی مرتکب کلاهبرداری می شود زیرا فشار یا مشکلی از نظر مالی، اجتماعی یا هر چیزی دیگری توسط تجربه می شودو نمی توان آن را در مراجع مجاز حل و فصل کرد یا از بین برد.

۲. **فرصت**، پای دوم مدل و پیش شرط مطرح بر اینکه یک فرد مرتکب تقلب و فعالیت های کلاهبرداری شود. تقلب تنها زمانی قابل انجام است که این فرصت برای فرد وجود داشته باشد که مشکلی که در اثر فشار تجربه شده است را به روش غیر مجاز و به صورت پنهان حل نماید.

۳. **عقلانی سازی** مکانیسم روانی است که توضیح می دهد چرا کلاهبرداران از ارتکاب کلاهبرداری خودداری نمی کنند و فکر می کنند رفتار آنها قابل قبول است.

مقاله‌ای توسط دافیلد و گرابوسکی (۲۰۰۱) اساس انگیزشی تقلب را از دیدگاه روان‌شناختی بررسی می‌کند و نتیجه می گیرد که تعدادی از عوامل روانی ممکن است در افرادی که مرتکب تقلب می شوند وجود داشته باشد، اما این عوامل همچنین با اشکال کاملاً مشروع تلاش انسانی مرتبط هستند. و بنابراین کلاهبرداران را نمی توان صرفاً بر اساس ویژگی ها یا الگوهای روانشناختی از افراد غیر کلاهبردار متمایز کرد. کلاهبرداری یک پدیده اجتماعی است به این معنا که منافع بالقوه برای کلاهبرداران به ضرر قربانیان است. این قربانیان افراد، شرکت ها یا دولت و به عنوان یک کل جامعه هستند. برخی از اعداد اخیر نشان دهنده اندازه تخمینی و تأثیر مالی تقلب هستند:

- یک سازمان سالانه به طور متوسط ۵ درصد از درآمد خود را به دلیل تقلب از دست می دهد (www.acfe.com).
- هزینه کل تقلب بیمه ای (بیمه غیر بهداشتی) در ایالات متحده بیش از ۴۰ میلیارد دلار در سال تخمین زده می شود(www.fbi.gov).
- تقلب سالانه ۷۳ میلیارد پوند برای بریتانیا ضرر دارد (مرجع ملی تقلب).
- شرکت‌های کارت‌های اعتباری به ازای هر صد دلار تراکنش تقریباً هفت سنت به دلیل تقلب از دست می‌دهند (اندروشریج، مالی شخصی خرابکاران پول، ۲۰۱۲).
- اندازه متوسط اقتصاد غیررسمی نسبت به GNI در سال ۲۰۰۰ در کشورهای در حال توسعه ۴۱ درصد است. این نسبت در کشورهای در حال گذار ۳۸ درصد و در کشورهای OECD، ۱۸ درصد است (اشنایدر ۲۰۰۲).

حتی اگر این اعداد تخمین های تقریبی باشند که از اندازه گیری‌ها بر اساس شواهد بدست آمده اند، نشان‌دهنده اهمیت و تأثیر پدیده تقلب و کلاهبرداری و در نتیجه نیاز سازمان‌ها و دولت‌ها به مبارزه فعال و جلوگیری از تقلب با تمام ابزارهایی است که در اختیار دارند. از سوی دیگراین اعداد همچنین نشان می‌دهند که احتمالاً سرمایه‌گذاری در سیستم‌های کشف تقلب و بازیابی تقلب ارزشمند است، زیرا می‌توان بازده مالی قابل توجهی را در سرمایه‌گذاری آن می توان کسب نمود.اهمیت و نیاز به سیستم‌های موثر کشف تقلب و پیشگیری از تقلب به‌علاوه توسط اشکال یا انواع مختلف کلاهبرداری برجسته می‌شود که تعدادی از آنها در این موارد خلاصه شده است.جدول ۱-۱، که البته جامع نیست نشان دهنده وقوع گسترده تقلب در صنایع مختلف و بخش های محصول

و خدمات است. دسته بندی های گسترده کلاهبرداری که در جدول مذکور فهرست شده و به اختصار تعریف شده اند را می توان به زیرگروه های خاص تری تقسیم کرد لیکن بدلیل اینکه ما را بیش از حد به جزئیات هر یک از این اشکال تقلب می برد از حوصله این نوشتار خارج است.

۳-۱. کشف و پیشگیری از تقلب

دو مؤلفه که بخش‌های اساسی تقریباً هر استراتژی مؤثر برای مبارزه با تقلب هستند، مربوط به کشف تقلب و پیشگیری از آن است. کشف تقلب به توانایی شناسایی یا کشف فعالیت های متقلبانه اشاره دارد، در حالی که پیشگیری از تقلب به اقداماتی اشاره دارد که می توان برای جلوگیری یا کاهش تقلب انجام داد. تفاوت بین هر دو واضح است. اولی یک رویکرد بعد از وقوع[۱] است در حالی که دومی یک رویکرد قبل از وقوع [۲] است. هر دو ابزار ممکن است و احتمالاً باید به شکلی مکمل برای پیگیری هدف مشترک، یعنی کاهش تقلب استفاده شوند. با این حال، همانطور که در ادامه با جزئیات بیشتر مورد بحث قرار خواهد گرفت، اقدامات پیشگیرانه استراتژی های تقلب و در نتیجه قدرت تشخیص تاثیر را تغییر خواهند داد. نصب یک سیستم تشخیص باعث می شود که کلاهبرداران سازگار شوند و رفتار خود را تغییر دهند و بنابراین خود سیستم تشخیص در نهایت قدرت تشخیص خود را مختل می کند. بنابراین اگر چه مکمل هستند، اما کشف و پیشگیری از تقلب مستقل نیستند و بنابراین باید در یک راستا قرار گیرند و به صورت یک کل در نظر گرفته شوند. رویکرد کلاسیک برای کشف تقلب، یک رویکرد مبتنی بر متخصص است، به این معنی که بر تجربه، شهود، و دانش تجاری یا حوزه تحلیلگر تقلب استوار است. چنین رویکردی مبتنی بر متخصص معمولاً شامل بررسی دستی یک مورد مشکوک است که ممکن است به عنوان مثال، توسط مشتری شکایت از دریافت هزینه برای تراکنش‌هایی که انجام نداده است، علامت‌گذاری شده باشد. چنین تراکنش مورد مناقشه ای ممکن است نشان دهنده یک مکانیسم کلاهبرداری جدید باشد که توسط کلاهبرداران کشف یا توسعه یافته است، و بنابراین نیاز به بررسی دقیق برای سازمان برای درک و متعاقباً رسیدگی به مکانیسم جدید دارد. درک مکانیسم یا الگوی تقلب امکان گسترش مکانیسم تشخیص و پیشگیری از تقلب را فراهم می کند که اغلب به عنوان یک قانون یا موتور به کار می رود، یعنی در قالب مجموعه ای از قواعد اگر-آنگاه [۳]، مکانیسم تقلب تازه کشف شده را توصیف می کند. این قوانین، همراه با قوانینی که الگوهای تقلب شناسایی شده قبلی را توصیف می‌کنند، در موارد یا معاملات آتی اعمال می‌شوند و زمانی که کلاهبرداری با استفاده از این مکانیسم انجام می‌شود، هشدار یا سیگنالی را ایجاد می‌کنند. یک مثال

[1] Ex post
[2] Ex ante
[3] If-then

ساده و در عین حال احتمالاً بسیار مؤثر از یک قانون کشف تقلب در یک تنظیم تقلب خسارت بیمه به شرح زیر است:

اگر:

مبلغ ادعا بالاتر از آستانه است، **یا**[1]

تصادف شدید، اما پلیس گزارشی ندارد، **یا**

آسیب شدید، اما هیچ گزارشی از پزشک وجود ندارد، **یا**

مدعی دارای چندین نسخه از تصادف است، یا

رسیدهای متعدد ارسال شده است

آنگاه:

ادعا را به عنوان مشکوک پرچم گذاری کنید، **و**[2]

به افسر تحقیق درباره تقلب هشدار دهید.

چنین رویکرد کارشناسی از تعدادی معایب رنج می برد. ساخت مبانی و قواعد و یا به عبارت دیگر موتورها معمولاً گران است، زیرا نیاز به ورودی دستی پیشرفته توسط متخصصان تقلب دارند و اغلب نگهداری و مدیریت آنها نیز دشوار است. قوانین باید به‌روز نگه داشته شوند و فقط یا در بیشتر مواردی که تقلب واقعی رخ داده است مورد استفاده قرارگیرند، زیرا هر مورد علامت‌گذاری شده نیاز به پیگیری و بررسی انسانی دارد. بنابراین، چالش اصلی اینگونه سیستم ها کارا و موثر نگه داشتن قواعد و یا قوانین است. به عبارت دیگر، تصمیم گیری اصلی در مورد زمان و زمان اضافه کردن، حذف، به روز رسانی یا ادغام کردن قوانین می باشد. درک این نکته مهم است که کلاهبرداران می توانند، به عنوان مثال، از طریق آزمون و خطا، قوانین کسب و کار که آنها را مسدود یا افشا می کند یاد بگیرند و راه حل های ابتکاری برای مقابله با آن ابداع نمایند. از آنجایی که قوانین در سیستم تشخیص مبتنی بر قانون مبتنی بر تجربه گذشته است، الگوهای جدید تقلب در حال ظهور به طور خودکار علامت گذاری یا هشدار داده نمی شوند. کلاهبرداری یک پدیده پویا است و بنابراین نیاز به ردیابی مداوم دارد. در نتیجه، سیستم تشخیص و پیشگیری از تقلب نیز باید به طور مداوم نظارت شود، بهبود یابد و به روز شود تا موثر بماند. یک سیستم تشخیص کلاهبرداری مبتنی بر متخصص به ورودی، ارزیابی و نظارت متخصص انسانی متکی است و لذا نیاز مند مقدار زیادی مداخله انسانی است. از سوی دیگر یک رویکرد خودکار برای ساخت و نگهداری یک سیستم تشخیص تقلب، که به مشارکت کمتر انسانی نیاز دارد، می‌تواند منجر به یک سیستم کارآمدتر و مؤثرتر برای کشف تقلب شود. در ادامه و قسمت های بعدی برخی از رویکردهای جایگزین برای سیستم‌های خبره معرفی

---

[1] OR [2] AND

می شود که استفاده از حجم عظیمی از داده‌ها که امروزه با هزینه بسیار کم قابل جمع‌آوری و پردازش است، به شیوه ای خودکارتر و کارآمدتر امکان توسعه، نظارت و به‌روزرسانی یک سیستم تشخیص تقلب رافراهم می آورد. شایان ذکر است که این رویکردهای جایگزین هنوز نیاز به دانش و ورودی متخصص دارند و بر پایه آن استوار است، که برای ساختن یک سیستم موثر حیاتی است.

۱-۴. داده های بزرگ برای کشف تقلب

زمانی که فعالیت‌های متقلبانه شناسایی و تأیید شد که به طور مؤثر به کلاهبرداری مربوط می‌شود، معمولاً دو نوع اقدام انجام می‌شود:

۱- **اقدامات اصلاحی**، با هدف حل تقلب و اصلاح عواقب نادرست، به عنوان مثال از طریق پیگیری استرداد یا جبران خسارات وارده. این اقدامات اصلاحی ممکن است شامل اقداماتی به صورت گذشته نگر نیز باشد. براین اساس موارد تقلب مشابهی را که از همان مکانیسم یا حفره‌های موجود در *سیستم تشخیص و پیشگیری از تقلب* که سازمان در اختیار دارد، شناسایی کرده و متعاقباً رسیدگی کنید.

۲- **اقدامات پیشگیرانه**، شامل اقداماتی است که هدف آنها جلوگیری از کلاهبرداری در آینده توسط کلاهبردار دستگیر شده یا کلاهبردار بالقوه دیگر است. این اقدامات به طور مثال می تواند شامل فسخ قرارداد با مشتری، و همچنین اقداماتی که با هدف جلوگیری از کلاهبرداری از همان نوع توسط افراد دیگر باشد. هنگامی که یک رویکرد مبتنی بر سیستم خبره اتخاذ می شود، یک اقدام پیشگیرانه نمونه می تواند گسترش موتور قانون با ترکیب قوانین اضافی است که امکان شناسایی و جلوگیری از مکانیسم تقلب کشف نشده در آینده را می دهد. یک پرونده کلاهبرداری باید به طور کامل مورد بررسی قرار گیرد تا بتوان مکانیسم زیربنایی کشف نموده و دانش تخصصی موجود را گسترش داد و این امکان فراهم شود که بتوان در آینده با قوی‌تر کردن سازمان و آسیب‌پذیری کمتر در برابر تقلب از وقوع آن جلوگیری نمود.

به طور معمول، هرچه اقدامات اصلاحی زودتر انجام شود و در نتیجه هر چه زودتر تقلب شناسایی شود، چنین اقداماتی مؤثرتر خواهد بود و می توان از ضررهای بیشتری جلوگیری کرد یا آنرا جبران نمود. از سوی دیگر، تشخیص تقلب هر چه بیشتر می‌گذرد، به دلایل خاص آسان‌تر می‌شود. وقتی مکانیسم یا مسیر تقلب وجود داشته باشد – به معنای خلأ در سیستم تشخیص و پیشگیری یک سازمان – تعداد دفعات این مسیر وجود دارد. دنبال خواهد شد (یعنی مکانیسم تقلب مورد استفاده) در زمان رشد می کند و بنابراین تعداد وقوع این نوع خاص از کلاهبرداری نیز افزایش می یابد. هرچه مسیر تقلب بیشتر باشد، آشکارتر می‌شود و معمولاً، در واقع از نظر آماری،

شناسایی آسان‌تر می‌شود. می توان انتظار داشت که تعداد وقوع یک نوع خاص از کلاهبرداری افزایش یابد زیرا به نظر می رسد بسیاری از کلاهبرداران مجرمان مکرر هستند. همانطور که گفته می شود، "یک بار یک دزد، همیشه یک دزد[1]." علاوه بر این، مکانیسم تقلب ممکن است توسط چندین نفر یا با دانش مشترک بین کلاهبرداران کشف و ابداء شود. مطمئناً برخی از انواع کلاهبرداری این قابلیت دارند که به صورت ویروسی پخش شوند و آنچه را که اثرات شبکه اجتماعی نامیده می شود به وقوع بپیوندد. این خود می تواند دلیل دیگر بر این مدعا باشد که کلاهبرداران دانش خود را در مورد نحوه ارتکاب کلاهبرداری به اشتراک می گذارند. این اثر نیز منجر به تعداد فزاینده‌ای از رخدادها و در نتیجه خطر یا شانس بالاتر،(البته بسته به دیدگاه فرد)، برای تشخیص می‌شود. پس از افشای یک مورد از نوع خاصی از کلاهبرداری، این امر منجر به افشای موارد مشابه کلاهبرداری خواهد شد که در گذشته اتفاق افتاده و از همان مکانیسم استفاده شده است. به طور معمول، یک غربالگری گذشته نگر برای ارزیابی اندازه یا تأثیر نوع تقلب تازه کشف شده و همچنین برای حل و فصل موارد تقلب (با اقدامات اصلاحی، رجوع کنید به بالا) انجام می شود. به این ترتیب، تشخیص تقلب با گذشت زمان آسان‌تر می‌شود، زیرا موارد تقلب مشابه بیشتری در زمان رخ می‌دهد و احتمال کشف نوع خاص کلاهبرداری را افزایش می‌دهد و همچنین به این دلیل که متقلبینی که کلاهبرداری مکرر انجام می‌دهند، خطر فردی آنها را افزایش می‌دهد. قرار گرفتن در معرض ریسک فردی هر چه متقلبی که بیشتر مرتکب کلاهبرداری شود به دلیل اینکه شانس جلب توجه بیشتر می‌شود فزایش می‌یابد. دلیل نهایی اینکه چرا کشف تقلب با گذشت زمان آسان تر می شود این است که تکنیک های تشخیص بهتر در حال توسعه هستند، به راحتی در دسترس هستند و توسط تعداد فزاینده ای از سازمان ها اجرا و اعمال می شوند. یک محرک مهم برای بهبود با توجه به تکنیک های تشخیص، در دسترس بودن داده ها در حال رشد است. اطلاعات و دیجیتالی شدن تقریباً هر جنبه از جامعه و زندگی روزمره منجر به فراوانی داده های موجود می شود. این به اصطلاح داده های بزرگ را می توان برای طیف وسیعی از اهداف از جمله کشف تقلب[2] با هزینه بسیار کم مورد کاوش و بهره برداری قرار داد.

۱-۵. تحلیل گرهای[3] تقلب

در این بخش به تحلیل فعالیت متقلبانه می پردازیم. ابتدا لازم است که زمینه وقوع تقلب و اکوسیستم رفتارهای متقلبانه مورد بررسی قرار گیرد. بر این اساس به خصوصیات رفتاری افراد متقلب، همچنین شخصیت افرادی که به منظور تقلب اقدام به حمله و نفوذ در سیستم ها می نمایند پرداخته می شود. نکته مهم در زمینه تحلیل فعالیت های

[1] once a thief, always a thief
[2] Baesens 2014
[3] fraud analytics

متقلبانه بررسی نحوه تفکر و نیز عقاید[1] افراد متقلب[2] است. قطعاً افراد متقلب دارای شخصیت نامطلوب برای سیستم ها می باشند لیکن به هرتقدیر انسان هستند و دارای شخصیت انسانی اند هر چند برای جامعه و سیستمی که در آن زندگی می کنند شخصیت غیر قابل قبول[3] باشند. این افراد مانند هر انسان دیگری دارای نیازهای انسانی هستند لذا شناخت اینکه آنها از کجا آمده اند می تواند به فعالیت های کشف تقلب کمک نماید. در مرحله بعد هر محقق و پژوهشگری در زمینه شناخت الگوهای تقلب نیاز به ابزار تحلیلی دارد که بر اساس آن بتواند بر موارد غیرمعمول[4] کشف کند. لذا روش ها و تکنیک های ریاضی و آماری و بویژه روش های آماری برای شناسایی الگوها[5] نقش مهمی برای کشف و شناسایی رفتارهای متقلبانه ایفا می نمایند. بواقع بدون داشتن ابزارتحلیلی عملاً مدیریت ریسک تقلب امکانپذیر نیست.

۶-۱. تشخیص تقلب مبتنی بر داده[6]

اگرچه رویکردهای کلاسیک و مبتنی بر متخصص در تشخیص تقلب همانطور که قبلاً مورد بحث قرار گرفت هنوز در حال استفاده گسترده هستند و قطعاً نقطه شروع و ابزار مکمل خوبی برای یک سازمان برای توسعه یک سیستم موثر کشف تقلب و پیشگیری است، تغییری به سمت داده‌ها در حال وقوع است. روش‌های شناسایی تقلب مبتنی بر آمار به سه دلیل زیر مورد استقبال قرار گرفته اند:

۱. **دقت**[7]. روش‌های تشخیص تقلب مبتنی بر آمار، قدرت تشخیص بیشتری در مقایسه با رویکردهای کلاسیک دارند. با پردازش حجم عظیمی از اطلاعات، الگوهای کلاهبرداری می توان کشف نمود که به اندازه کافی برای چشم انسان آشکار نیستند. توجه به این نکته مهم است که قدرت بهبود یافته رویکردهای داده محور بر پردازش انسانی را می توان در برنامه های مشابه مانند امتیازدهی اعتبار[8] یا پیش بینی ریزش مشتری[9] مشاهده کرد. بیشتر سازمان ها از ظرفیت محدودی برای بررسی پرونده ها توسط بازرس برای تأیید اینکه آیا پرونده به طور مؤثر مربوط به کلاهبرداری است یا خیر، برخوردارند. هدف یک سیستم تشخیص تقلب می تواند استفاده بهینه از ظرفیت محدود بازرسی موجود یا به عبارت دیگر به حداکثر رساندن نسبت پرونده های تقلبی در میان پرونده های بازرسی شده و همچنین احتمالاً افزایش میزان رخدادهای تقلب کشف شده، باشد. یک سیستم با دقت بالاتر، همانطور که توسط روش‌های مبتنی بر داده ارائه می‌شود، به نسب روش بازرسی تعداد بیشتری از موارد آشکار شده را دراختیار می گذارد.

---

[1] mindset
[2] fraudster
[3] persona non grata
[4] anomaly
[5] pattern recognition
[6] data-driven fraud detection
[7] precision
[8] credit scoring
[9] customer churn prediction

۲. **کارایی عملیاتی**[1]. در موارد خاص، افزایش تعداد مواردی که باید تجزیه و تحلیل شوند، نیازمند یک فرآیند خودکار است که توسط روش‌های تشخیص تقلب مبتنی بر داده ارائه می شود. علاوه بر این، در چندین برنامه، ممکن است الزامات عملیاتی وجود داشته باشد که محدودیت‌های زمانی را برای رسیدگی به یک پرونده تحمیل می نماید. برای مثال، هنگام ارزیابی تراکنش با کارت اعتباری، تصمیم تقریباً فوری در رابطه با تأیید یا مسدود کردن تراکنش به دلیل سوء ظن مورد نیاز است. مثال دیگر مربوط به کشف تقلب برای گمرک در بندر است، که در آن باید در یک بازه زمانی محدود تصمیم گیری شود که آیا اجازه داده شود کانتینر عبور کند و به داخل کشور وارد شود، یا اینکه نیاز به بازرسی بیشتر است که ممکن است باعث تاخیر شود. رویکردهای خودکار مبتنی بر داده می تواند به تسهیل و تسریع امور مزبور کمک نموده و با الزامات عملیاتی سختگیرانه نیز مطابقت داشته باشد.

۳. **کارایی هزینه**[2]. همانطور که قبلاً در بخش قبل ذکر شد، توسعه و حفظ یک سیستم تشخیص تقلب کارآمد و ناب مبتنی بر متخصصان و سیستم های خبره چالش برانگیز بوده و نیازمند کار فشرده است. یک رویکرد خودکارتر و کارآمدتر برای توسعه و حفظ یک سیستم تشخیص تقلب روش‌های مبتنی بر داده‌ها ارائه می‌شود، علاوه بر کارایی عملیاتی که در قسمت قبل به اجمال به آن پرداخته شد از نظر هزینه ای و نیز بازگشت سرمایه نیز دارای کارایی بیشتری است.

۱-۷. تکنیک های کشف تقلب

کلاهبرداران استراتژی‌های پیشرفته‌ای را برای پوشاندن زیرکانه ردپای خود توسعه می‌دهند تا از کشف نشدن جلوگیری کنند. کلاهبرداران حداکثر تلاش را انجام می دهند تا با محیط اطراف ترکیب شوند. چنین رویکردی یادآور تکنیک های استتار است که توسط ارتش یا حیواناتی مانند آفتاب پرست و حشرات چوبی استفاده می شود. این بدیهی است که تقلب بر اساس فرصت نیست، بلکه با دقت برنامه ریزی شده است، که تشخیص آنها نیازمند تکنیک های جدیدی است که قادر به شناسایی و رسیدگی به الگوهایی هستند که در ابتدا به نظر می رسد با رفتار عادی مطابقت دارند، اما در واقع فعالیت های متقلبانه را ایجاد می کنند. مکانیسم های تشخیص مبتنی بر تکنیک‌های یادگیری بدون نظارت[3] یا تجزیه و تحلیل توصیفی[4]، معمولاً به دنبال یافتن رفتاری است که از رفتار عادی منحرف می‌شود، یا به عبارت دیگر تشخیص ناهنجاری‌ها. این تکنیک‌ها از مشاهدات تاریخی یاد می گیرند، و بدون نظارت نامیده می‌شوند، زیرا نیازی به برچسب زدن این مشاهدات به عنوان نمونه‌های جعلی یا غیر متقلبانه

[1] operational efficiency
[2] cost efficiency
[3] unsupervised learning techniques
[4] descriptive analytics

ندارند. نمونه‌ای از رفتاری که با رفتار عادی در تنظیم تقلب اشتراک مخابراتی مطابقت ندارد، توسط مجموعه داده‌های تراکنش با سوابق جزئیات تماس یک مشترک خاص که در جدول ۲ نشان داده شده است، ارائه می‌شود[1] توجه داشته باشید که تماس هایی که جعلی هستند (ستون آخر جدول نشان دهنده راهزن) به خودی خود مشکوک نیستند. با این حال، آنها از رفتار عادی برای این مشترک خاص منحرف می شوند. تکنیک‌های تشخیص پرت ارزش زیادی دارند و امکان شناسایی بخش قابل توجهی از موارد تقلبی را فراهم می‌کنند. به طور خاص، آنها ممکن است امکان کشف تقلبی را فراهم کنند که ماهیت متفاوتی با تقلب تاریخی دارد، یا به عبارت دیگر تقلبی که از مکانیسم‌های جدید و ناشناخته استفاده می‌کند و منجر به الگوی تقلب جدید می‌شود. این الگوهای جدید توسط سیستم‌های خبره کشف نمی‌شوند، و به همین دلیل تحلیل‌های توصیفی ممکن است برای اولین بار باشد ابزار تکمیلی که توسط یک سازمان به منظور بهبود سیستم تشخیص تقلب مبتنی بر قوانین متخصص خود اتخاذ می شود. با این حال، تکنیک‌های توصیفی دقیقاً با استراتژی‌های تقلب استتاری که قبلاً مورد بحث قرار گرفت، مستعد فریب هستند. بنابراین، سیستم تشخیص را می توان با تکمیل آن با ابزاری بهبود بخشید که می تواند کلاهبردارانی را که تکنیکی شبیه استتار را اتخاذ می کنند، بپوشاند. بنابراین، در فصل ۴، نوع دومی از تکنیک ها معرفی شده است. هدف تکنیک های یادگیری نظارت شده یا تجزیه و تحلیل پیش بینی، یادگیری از اطلاعات یا مشاهدات تاریخی به منظور بازیابی الگوهایی است که امکان تمایز بین رفتار عادی و متقلبانه را فراهم می کند. هدف این تکنیک‌ها یافتن آلارم‌های بی‌صدا، بخش‌هایی از مسیرهای آنهاست که کلاهبرداران نمی‌توانند آن‌ها را بپوشانند. یادگیرندگان تحت نظارت را می توان برای پیش بینی یا کشف تقلب و همچنین برای تخمین میزان تقلب استفاده کرد. تجزیه و تحلیل پیش‌بینی‌کننده نیز محدودیت‌هایی دارد، احتمالاً مهم‌ترین آنها این است که آنها به نمونه‌های تاریخی برای یادگیری نیاز دارند (به عنوان مثال، مجموعه داده برچسب‌گذاری شده از رفتار تقلبی مشاهده شده تاریخی). این امر قدرت تشخیص آنها را با توجه به انواع کلاهبرداری به شدت متفاوت با استفاده از مکانیسم‌ها یا روش‌های جدید کاهش می‌دهد، که تاکنون شناسایی نشده‌اند و بنابراین در پایگاه داده تاریخی موارد کلاهبرداری که مدل پیش‌بینی‌کننده از آن آموخته شده است، گنجانده نشده‌اند. همانطور که قبلاً بحث شد، تجزیه و تحلیل توصیفی ممکن است با توجه به شناسایی چنین مکانیزم‌های کلاهبرداری جدید بهتر عمل کند، حداقل اگر یک مکانیسم تقلب جدید منجر به انحراف قابل تشخیص از حالت عادی شود. این امر مکمل بودن روش‌های تحت نظارت و بدون نظارت را نشان می‌دهد و انگیزه استفاده از هر دو نوع روش را به عنوان ابزارهای مکمل در توسعه یک سیستم قدرتمند کشف تقلب و پیشگیری ایجاد می‌کند. نوع سوم از ابزار تکمیلی مربوط به تجزیه و تحلیل شبکه های اجتماعی است که با یادگیری و شناسایی ویژگی های رفتار متقلبانه

[1] Fawcett and Provost1997

در شبکه ای از موجودیت های مرتبط، توانایی های سیستم کشف تقلب را بیشتر گسترش می دهد. تجزیه و تحلیل شبکه های اجتماعی جدیدترین ابزار در جعبه ابزار ما برای مبارزه با تقلب است، و شواهد نشان می دهد که این دسته از روش ها ابزار بسیار قدرتمندی هستند که با استفاده از روابط بین نهادها می تواند در کشف الگوهای خاصی که نشان دهنده تقلب است، کمک کنند. لازم به تأکید است که این سه نوع مختلف تکنیک ممکن است مکمل یکدیگر باشند زیرا بر جنبه های مختلف تقلب تمرکز می کنند و نباید به عنوان جایگزین انحصاری در نظر گرفته شوند. یک سیستم موثر شناسایی و پیشگیری از تقلب از این ابزارهای مختلف استفاده کرده و بر حسب نیاز آنها را ترکیب می‌کند و از امکانات و محدودیت‌های متفاوت آنها در یک مجموعه و به نحوی که یکدیگر را تقویت نمایند، سود می برد. هنگام توسعه یک سیستم کشف تقلب، یک سازمان احتمالاً از ترتیبی که ابزارهایی مختلفی که معرفی شده اند استفاده می کند. به عنوان گام اول ممکن است یک موتور قانون مبتنی بر خبرگی ایجاد شود، در مرحله دوم ممکن است با تجزیه و تحلیل توصیفی، و متعاقباً با تجزیه و تحلیل شبکه های اجتماعی و پیش بینی تکمیل شود. توسعه یک سیستم تشخیص تقلب به این ترتیب به سازمان اجازه می دهد تا تخصص و بینش را به صورت گام به گام به دست آورد و بدین وسیله مرحله بعدی را تسهیل می کند. با این حال، ترتیب دقیق اتخاذ تکنیک‌های مختلف ممکن است به ویژگی‌های نوع تقلبی که یک سازمان با آن مواجه است بستگی داشته باشد.

جدول ۱-۲ سوابق جزئیات تماس یک مشتری با موارد پرت نشان دهنده فعالیت مشکوک در اشتراک مشتری
(رفتار انحرافی که در یک لحظه خاص در زمان شروع می شود) (فاوست و پروست ۱۹۹۷)

زمان						
1/01	10:05:01	Mon	13 mins	Brooklyn, NY	Stamford, CT	
1/05	14:53:27	Fri	5 mins	Brooklyn, NY	Greenwich, CT	
1/08	09:42:01	Mon	3 mins	Bronx, NY	White Plains, NY	
1/08	15:01:24	Mon	9 mins	Brooklyn, NY	Brooklyn, NY	
1/09	15:06:09	Tue	5 mins	Manhattan, NY	Stamford, CT	
1/09	16:28:50	Tue	53 sec	Brooklyn, NY	Brooklyn, NY	
1/10	01:45:36	Wed	35 sec	Boston, MA	Chelsea, MA	
1/10	01:46:29	Wed	34 sec	Boston, MA	Yonkers, MA	
1/10	01:50:54	Wed	39 sec	Boston, MA	Chelsea, MA	
1/10	11:23:28	Wed	24 sec	White Plains, NY	Congers, NY	
1/11	22:00:28	Thu	37 sec	Boston, MA	East Boston, MA	
1/11	22:04:01	Thu	37 sec	Boston, MA	East Boston, MA	

مأخذ: (Baesens et al. 2015,p.20)

۸-۱. چرخه تقلب[1]

شکل ۲ چرخه تقلب را معرفی می کند و چهار فعالیت اساسی را نشان می دهد:

- تشخیص تقلب: استفاده از مدل‌های تشخیص در موارد جدید و تعیین خطر تقلب برای هر مشاهده.
- بررسی تقلب: یک متخصص انسانی اغلب مورد نیاز است تا موارد مشکوک و نشاندار شده را با توجه به پیچیدگی آن به طور دقیق بررسی نماید.
- تایید تقلب: تعیین اینکه موارد مشکوک واقعاً تقلب هستند این موضوع احتمالاً شامل تحقیقات میدانی برای این منظور است.
- پیشگیری از تقلب: این امر به منظور جلوگیری از تقلب درآینده صورت می گیرد و حتی ممکن است منجر به کشف تقلب قبل از آن نیز شود حتی قبل از اینکه کلاهبردار بداند که می خواهد در آینده کلاهبرداری نماید.

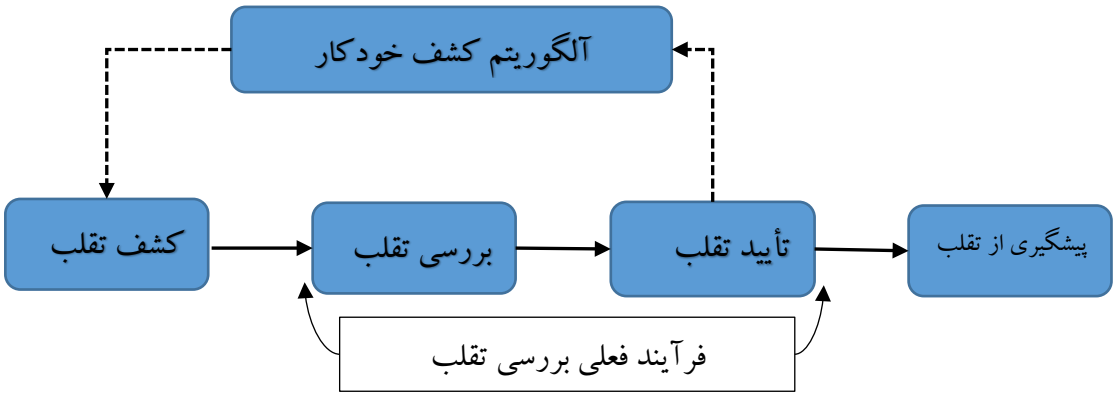
این موضوع دقیقاً فرضیه داستان کوتاه علمی تخیلی گزارش اقلیت فیلیپ کی دیک در سال ۱۹۵۶ است. حلقه بازخورد در شکل ۲ را از فعالیت تأیید تقلب به سمت فعالیت کشف تقلب توجه کنید. موارد تازه کشف شده باید (در اسرع وقت!) به پایگاه داده پرونده های تقلب تاریخی اضافه شود که برای یادگیری یا القای مدل تشخیص استفاده می شود. مدل کشف تقلب ممکن است هر بار که یک مورد جدید تقلب مشاهده می شود دوباره آموزش داده نشود. با این حال، با توجه به ماهیت پویای تقلب و اهمیت کشف تقلب در اسرع وقت، به‌روزرسانی منظم مدل توصیه می‌شود. فرکانس مورد نیاز برای بازآموزی یا به روز رسانی مدل تشخیص به عوامل مختلفی بستگی از جمله موارد زیردارد:

- نوسانات رفتار کلاهبرداری
- قدرت تشخیص مدل فعلی که مربوط به نوسانات رفتار کلاهبردارانه است
- تعداد موارد (مشابه) تایید شده در حال حاضر در پایگاه داده در دسترس است
- میزان تایید موارد جدید
- تلاش لازم برای بازآموزی مدل

بسته به نیاز جدید به بازآموزی که توسط این عوامل تعیین می شود، و همچنین عوامل اضافی احتمالی، یک رویکرد خودکار مانند یادگیری تقویتی می تواند در نظر گرفته شود که به طور مداوم مدل تشخیص را با یادگیری و با توجه به جدیدترین مشاهدات به روز نماید.

---

[1] fraud cycle



شکل ۱- ۲چرخه تقلب

۱-۹. بررسی موردی: یادگیری تحت نظارت و بدون نظارت برای کشف تقلب در کارت اعتباری

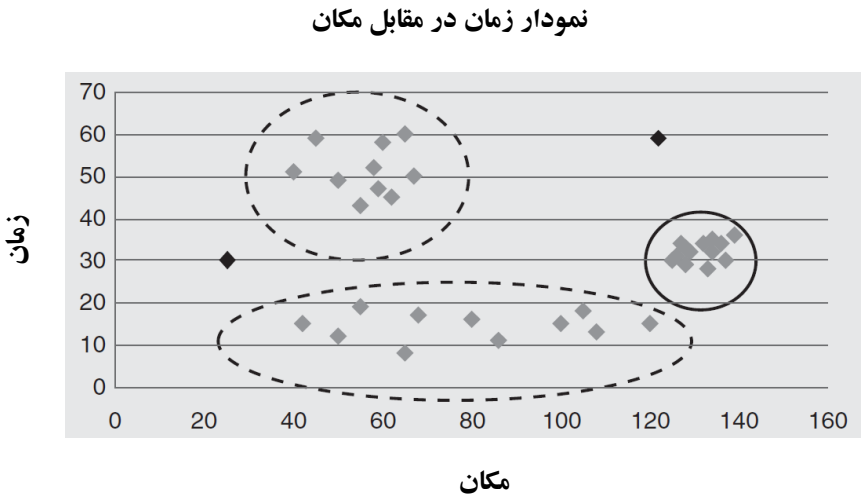
به منظور مبارزه با تقلب و با توجه به در دسترس بودن داده‌های فراوان، شرکت‌های کارت اعتباری از جمله اولین پذیرندگان رویکردهای کلان داده برای توسعه سیستم‌های موثر تشخیص و پیشگیری تقلب بوده‌اند. یک تراکنش معمولی با کارت اعتباری در سیستم های شرکت کارت اعتباری با درج تا صد یا بیشتر مشخصه که جزئیات یک تراکنش را توصیف می کند، ثبت می شود. جدول ۳ برای اهداف توضیحی تعدادی از این ویژگی ها یا متغیرهایی را ارائه می دهد که در حال ثبت هستند (Hand 2007).

جدول ۱- ۳ فیلدهای داده تراکنش کارت اعتباری

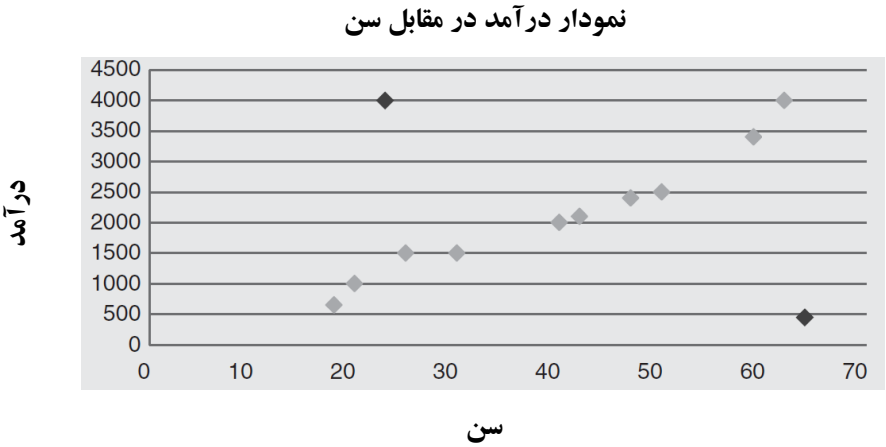
با ثبت این اطلاعات در یک دوره زمانی، یک مجموعه داده در حال ایجاد می شود که امکان استفاده از تجزیه و تحلیل توصیفی را فراهم می کند. این شامل تکنیک‌های تشخیص داده های پرت[1] می‌شود که امکان تشخیص رفتار و/یا ویژگی‌های غیرعادی یا غیرعادی را در مجموعه داده‌ها فراهم می‌کند. داده های پرت می تواند فعالیت

[1] outlier detection techniques

های مشکوک را نشان دهد. پرت بودن داده ممکن است در سطح آیتم داده یا سطح مجموعه داده رخ دهد. شکل ۳ تصویری از نقاط پرت را در سطح آیتم داده ارائه می دهد، در این مثال تراکنش هایی که از رفتار عادی مشتری منحرف می شوند. نمودار پراکندگی به وضوح سه دسته از انواع معمولی و مکرر را نشان می دهد که با بعد زمانی و مکانی معاملات برای یک مشتری خاص مشخص می شود، و همچنین دو تراکنش انحرافی که با رنگ سیاه مشخص شده اند. این موارد پرت مشکوک هستند و احتمالاً مربوط به معاملات متقلبانه هستند و بنابراین ممکن است برای تحقیقات بیشتر انسانی علامت گذاری شوند. نقطه پرت در سطح مجموعه داده به این معنی است که رفتار یک فرد یا نمونه با رفتار کلی مطابقت ندارد. شکل ۱.۵ ویژگی های سن و درآمد مشتریان را که هنگام درخواست کارت اعتباری ارائه شده است، نشان می دهد. دو نقطه پرت که با رنگ سیاه در نمودار مشخص شده اند ممکن است به اصطلاح تقلب در اشتراک را نشان دهند (به جدول ۱.۱، تعریف کلاهبرداری کارت اعتباری مراجعه کنید)، زیرا این ترکیبات سن و درآمد به شدت از رفتار عادی منحرف می شوند.



شکل ۱- ۳ تشخیص داده های پرت در سطح رفتاریک فرد(آیتم)



شکل ۱- ۴ تشخیص داده های پرت در سطح مجموعه ای از داده ها

افزودن یک فیلد در مجموعه داده‌های تراکنش موجود که نشان می‌دهد آیا یک تراکنش تقلبی بوده است، امکان استفاده از تجزیه و تحلیل پیش‌بینی‌کننده را فراهم می‌کند تا مدلی را ارائه دهد که یک نمونه را به عنوان تقلبی یا غیرتقلبی پیش‌بینی یا طبقه‌بندی می‌کند. چنین مدل های مبتنی بر داده های پرت می تواند برای درک الگوهای رفتاری کلاهبرداری کارت اعتباری کمک نماید که درجای خود به پیش بینی اینکه آیا یک تراکنش ممکن است تقلبی باشد، منجر می گردد. نمونه هایی از چنین الگوهایی به شرح زیر است:

- خرید کوچک و به دنبال آن یک خرید بزرگ
- خرید اینترنتی زیاد در مدت زمان کوتاه
- خرج کردن هر چه سریعتر
- خرج کردن مقادیر کمتر، در طول زمان توزیع می شود

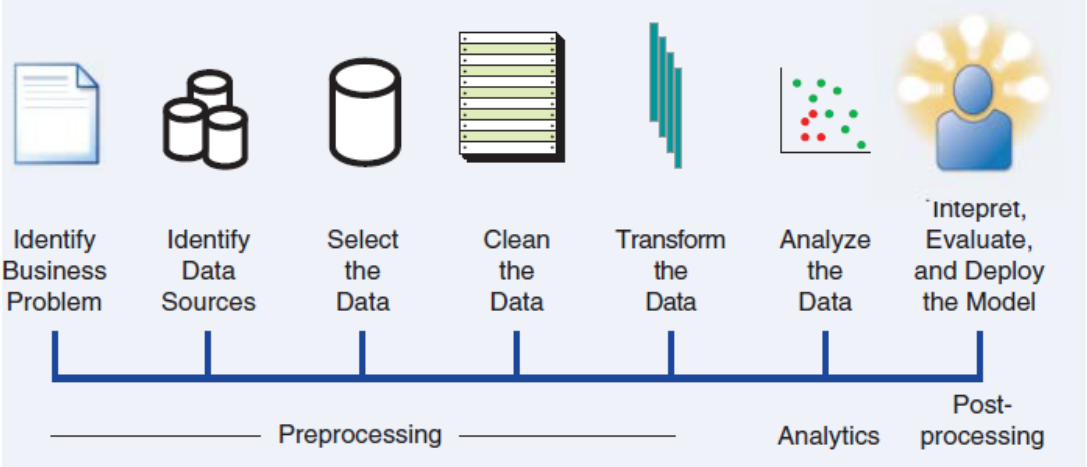
شناسایی چنین الگویی ممکن است سخت تر باشد و به روش های پیشرفته ای که توسط کلاهبرداران اتخاذ شده و دقیقاً برای جلوگیری از شناسایی توسعه داده شده اند نگران باشد.

۱۰-۱. مدل فرآیندی تحلیل تقلب

شکل ۵ نمای کلی سطح بالایی از فرآیند تحلیل را ارائه می دهد.مدل (هان و کامبر ۲۰۱۱؛ هند، مانیلا و اسمیت ۲۰۰۱؛ تان، استینباخ و کومار ۲۰۰۵). به عنوان اولین قدم، یک تعریف کامل از مشکل کسب و کار مورد نیاز است تا با تجزیه و تحلیل حل شود. در مرحله بعد، تمام داده های منبع باید شناسایی شوند که می توانند مورد علاقه بالقوه باشند. این یک گام بسیار مهم است، زیرا داده‌ها جزء کلیدی هر تمرین تحلیلی هستند و انتخاب داده‌ها تأثیر قطعی بر مدل‌های تحلیلی خواهد داشت که در مرحله بعدی ساخته می‌شوند. سپس تمام داده ها در یک منطقه مرحله بندی جمع آوری می شوند که می تواند یک دیتا مارت(داده گاه)[1] یا انبار داده[2] باشد.

[1] data mart

[2] data warehouse



شکل ۱- ۵ مدل فرآیند تحلیل تقلب

مأخذ: (Baesens et al. 2015)

برخی از تحلیل‌های اکتشافی اولیه را می‌توان در اینجا با استفاده از امکانات OLAP[1] (پردازش تحلیلی آنلاین) برای تجزیه و تحلیل داده‌های چند بعدی (مانند جمع‌آوری، جزءکاوی[2]، برش دادن[3] و جزء به جزء کردن[4]) در نظر گرفت. این یک مرحله پاکسازی داده ها برای خلاص شدن از همه ناسازگاری ها، مانند مقادیر از دست رفته و داده های تکراری دنبال می شود. ممکن است تبدیل‌های اضافی مانند تبدیل داده های عددی به داده های رسته ای[5] است که در آمار و علوم داده به آن تظریف(ظرف بندی، سبد بندی)[6] گویند، همچنین کدگذاری الفبایی به عددی، تجمیع جغرافیایی و غیره در نظر گرفته شود. در مرحله تجزیه و تحلیل، یک مدل تحلیلی بر روی داده های از پیش پردازش شده و تبدیل شده تخمین زده می شود. در این مرحله، مدل واقعی کشف تقلب ساخته می شود. در نهایت، زمانی که مدل ساخته شد، توسط کارشناسان تقلب تفسیر و ارزیابی خواهد شد. اما مسلماً، مسئله کلیدی یافتن الگوهای ناشناخته و در عین حال جالب و قابل اجرا (که گاهی به عنوان الماس دانش نیز نامیده می شود) است که می تواند بینش و قدرت تشخیص بیشتری را ارائه دهد. هنگامی که مدل تحلیلی به طور مناسب اعتبار سنجی شد وتایید شده است، می توان آن را به عنوان یک برنامه تحلیلی (به عنوان مثال، سیستم پشتیبانی تصمیم، موتور امتیازدهی) در خط تولید قرار داد. نکته مهمی که در اینجا باید در نظر گرفته شود این است که چگونه خروجی مدل را به روشی کاربرپسند نشان دهیم، چگونه آن را با سایر برنامه‌ها ادغام کنیم (به عنوان مثال، سیستم تشخیص و پیشگیری، موتورهای ریسک)، و چگونه مطمئن شویم که مدل تحلیلی می‌تواند به طور مناسب نظارت شود و به طور مداوم مورد آزمایش قرار گیرد. توجه به این نکته مهم است که مدل فرآیندی که در شکل۵

[1] OnLine Analytical Processing)
[2] drill down
[3] slicing
[4] dicing
[5] categorical data
[6] binning

مشخص شده است ماهیت تکراری دارد به این معنا که ممکن است در طول تمرین مجبور شوید به مراحل قبلی بازگردید. به عنوان مثال، در طول مرحله تجزیه و تحلیل، ممکن است نیاز به داده های اضافی شناسایی شود که نیاز به تمیز کردن، تبدیل و غیره اضافی داشته باشد. زمان برترین مرحله معمولاً مرحله انتخاب داده و پیش پردازش است که معمولاً حدود ۸۰ درصد از کل تلاش های مورد نیاز برای ساخت یک مدل تحلیلی را به خود اختصاص می دهد. یک مدل کشف تقلب قبل از به کارگرفتن باید به طور کامل ارزیابی شود. بسته به تنظیم دقیق و استفاده از مدل، ممکن است لازم باشد جنبه های مختلف در طول ارزیابی ارزیابی شود تا از قابل قبول بودن مدل برای اجرا اطمینان حاصل شود. جدول ۴ چندین ویژگی کلیدی مدل های تجزیه و تحلیل تقلب موفق را بررسی می کند که بسته به کاربرد دقیق ممکن است ملاک بررسی قرار گیرند تعدادی از چالش های خاص ممکن است در هنگام توسعه و اجرای یک مدل کشف تقلب ظاهر شوند که احتمالاً منجر به مشکلات در دستیابی به اهداف می شود. همانطور که با ویژگی های بحث شده در جدول ۴ بیان شده است. اولین چالش کلیدی مربوط به ماهیت پویای تقلب است. کلاهبرداران به طور مداوم سعی می کنند با توسعه استراتژی ها و روش های جدید، سیستم های تشخیص و پیشگیری را شکست دهند. بنابراین، مدل‌های تحلیلی تطبیقی و سیستم‌های تشخیص و پیشگیری، به منظور کشف و رفع تقلب در اسرع وقت مورد نیاز است. همانطور که قبلاً بحث شد، شناسایی هر چه زودتر تقلب بسیار مهم است. بدیهی است که تشخیص تقلب تا آنجا که ممکن است باید با دقت بسیار زیاد بتواند تقلب را آشکار نماید به نحوی که بسیاری از موارد کلاهبرداری، به ویژه در موارد کلاهبرداری که دارای مقدار زیادی یا تأثیر مالی هستند از نظر پنهان نمانند. هزینه از غافل شدن از یک مورد کلاهبرداری یا مکانیزم کلاهبرداری ممکن است قابل توجه باشد. در ارتباط با داشتن قدرت تشخیص خوب، در کنار نرخ تشخیص صحیح موارد درست تقلب، برخورداری از نرخ هشدار نادرست پایین بسیار مهم است، زیرا پرهیز از آزار و اذیت مشتریان خوب همچنین مسدود شدن بی مورد حساب ها یا تراکنش ها از اهمیت بالایی برخوردار است. در توسعه مدل های تحلیلی با قدرت تشخیص خوب ونرخ هشدار نادرست پایین، یک مشکل اضافی مربوط به چولگی بودن [1] داده ها است، به این معنی که ما معمولاً نمونه های تاریخی زیادی از پرونده های غیر متقلبانه داریم، اما فقط تعداد محدودی از پرونده های متقلبانه در دسترس قراردارد. به عنوان مثال، در تنظیمات کلاهبرداری با کارت اعتباری، معمولا کمتر از ۰.۵ درصد از تراکنش ها جعلی هستند. چنین مشکلی معمولاً به عنوان یک مشکل سوزن در انبار کاه نامیده می شود و ممکن است باعث شود یک تکنیک تحلیلی در یادگیری یک مدل دقیق با مشکل مواجه شود. بسته به کاربرد دقیق، کارایی عملیاتی نیز ممکن است یک نیاز کلیدی باشد، به این معنی که سیستم تشخیص تقلب ممکن است فقط زمان محدودی برای تصمیم گیری و اجازه دادن به یک تراکنش داشته باشد. به عنوان

[1] skewness

مثال، در تنظیم تشخیص کلاهبرداری از کارت اعتباری، زمان تصمیم گیری معمولاً باید کمتر از هشت ثانیه باشد. چنین نیازی به وضوح بر طراحی سیستم‌های فناوری اطلاعات عملیاتی و همچنین طراحی مدل تحلیلی تأثیر می‌گذارد. ارزیابی مدل تحلیلی همچنین جمع آوری یا محاسبه اطلاعات یا متغیرهایی که توسط مدل استفاده می شود نباید زیاد طول بکشد. به عنوان مثال، محاسبه متغیرهای روند در زمان واقعی، ممکن است از منظر عملیاتی امکان پذیر نباشد، زیرا این کار زمان زیادی را صرف می کند. این موضوع یکی از چالش عمده در ارتباط با حجم عظیم داده‌های موجود و نیاز به پردازش آنها برای تشخیص تقلب می باشد.

جدول۱- ۴ ویژگی های کلیدی مدل های تحلیل تقلب موفق

شاخص	شرح
دقت آماری	این شاخص به قدرت تشخیص و صحت مدل آماری در موارد علامت گذاری به عنوان مشکوک اشاره می کند. چندین معیار ارزیابی آماری وجود دارد و ممکن است برای ارزیابی این جنبه اعمال شود، مانند نرخ برخورد با هدف[1]، منحنی بهبود[2]، AUC[3]، و غیره. دقت آماری همچنین ممکن است به معنی داری آماری[4] اشاره داشته باشد. به این معنی که الگوهایی که در داده ها یافت شده اند باید واقعی باشند و نتیجه تصادفی نباشند. به عبارت دیگر، ما باید مطمئن شویم که مدل به خوبی تعمیم می یابد و بیش از حد به مجموعه داده های تاریخی تطبیق داده نمی شود.
تفسیر پذیری	هنگامی که به درک عمیق‌تری از الگوهای تقلب کشف‌شده نیاز است، به عنوان مثال برای اعتبارسنجی مدل قبل از استفاده از آن، لازم است مدل کشف تقلب تفسیر پذیر باشد. این جنبه مستلزم درجه خاصی از ذهنیت گرایی است، زیرا تفسیرپذیری ممکن است به دانش کاربر بستگی داشته باشد. تفسیرپذیری یک مدل به قالب آن بستگی دارد، که به نوبه خود توسط تکنیک تحلیلی اتخاذ شده تعیین می شود. مدل‌هایی که به کاربر اجازه می‌دهند دلایل اساسی را بفهمند که چرا مدل به یک مورد مشکوک است، مدل‌های جعبه سفید نامیده می‌شوند، در حالی که مدل‌های پیچیده ریاضی غیرقابل درک اغلب به عنوان مدل‌های جعبه سیاه نامیده می‌شوند. هرچند در برخی محیط ها استفاده از مدل های جعبه سیاه برای کشف تقلب قابل قبول هستند، لیکن در بیشتر نظام های کشف تقلب، سطحی از درک و در واقع اعتبارسنجی که با تفسیرپذیری تسهیل می‌شود، برای اطمینان از مدیریت و امکان عملیاتی‌سازی مؤثر مورد نیاز است.

[1] hit rate
[2] lift curve
[3] Area Under Curve (in ROC Curve)
[4] statistical significance

شاخص	شرح
کارایی عملیاتی	کارایی عملیاتی به زمان مورد نیاز برای ارزیابی مدل یا به عبارت دیگر زمان مورد نیاز برای ارزیابی مشکوک بودن یا نبودن یک مورد اشاره دارد. هنگامی که موارد نیاز به ارزیابی در زمان واقعی دارند، به عنوان مثال برای نشان دادن کلاهبرداری احتمالی کارت اعتباری، کارایی عملیاتی بسیار مهم است و در طول ارزیابی عملکرد مدل نگرانی اصلی است. کارایی عملیاتی همچنین مستلزم جمع‌آوری و پیش پردازش داده‌ها، ارزیابی مدل، نظارت و آزمون بک تست(آزمون پشتیبان)[1] مدل، و برآورد مجدد آن در صورت لزوم است.
هزینه اقتصادی	توسعه و اجرای یک مدل کشف تقلب هزینه قابل توجهی برای یک سازمان دارد. هزینه کل شامل هزینه های جمع آوری، پیش پردازش و تجزیه و تحلیل داده ها و هزینه های قرار دادن مدل های تحلیلی به دست آمده در خط تولید است. علاوه بر این، هزینه های نرم افزار و همچنین منابع انسانی و محاسباتی باید در نظر گرفته شود. احتمالاً داده های خارجی نیز باید خریداری شود تا داده های داخلی موجود غنی شود. واضح است که انجام یک تجزیه و تحلیل کامل هزینه و فایده در شروع پروژه و به دست آوردن بینشی در مورد عوامل تشکیل دهنده بازده سرمایه گذاری ساخت یک سیستم پیشرفته تشخیص تقلب دارای اهمیت است.
تطبیق با مقررات	بسته به زمینه ممکن است مقررات و قوانین داخلی یا سازمانی و خارجی وجود داشته باشد که برای توسعه و کاربرد یک مدل اعمال شود. بدیهی است که یک مدل تشخیص تقلب باید مطابق با تمام مقررات و قوانین قابل اجرا باشد، به عنوان مثال در مورد حفظ حریم خصوصی، استفاده از کوکی ها در یک مرورگر وب و غیره.

۱۱-۱. متخصصین علوم داده های تقلب [2]

در حالی که در بخش قبل ویژگی‌های یک مدل تشخیص تقلب خوب را مورد بحث قرار دادیم، در این پاراگراف ویژگی‌های کلیدی یک متخصص خوب در زمینه علوم داده های تقلب از دیدگاه مدیر استخدام را توضیح

[1] back test

[2] fraud data scientists

خواهیم داد. این مبتنی بر تجربه مشاوره و تحقیقی[1] است که در رابطه با بسیاری از شرکت‌ها در سراسر جهان در زمینه داده‌های بزرگ، تجزیه و تحلیل و کشف تقلب انجام شده است.

۱-۱۱-۱. یک دانشمند داده های کلاهبرداری باید مهارت های کمی داشته باشد

بدیهی است که یک دانشمند علوم داده های کلاهبرداری باید پیشینه کاملی در آمار، یادگیری ماشین و/یا داده کاوی داشته باشد. تمایز بین این رشته های مختلف بیشتر و بیشتر مبهم می شود و در واقع چندان مرتبط نیست. همه آنها مجموعه ای از تکنیک های کمی را برای تجزیه و تحلیل داده ها و یافتن الگوهای مربوط به کسب و کار در یک زمینه خاص مانند کشف تقلب ارائه می کنند. یک دانشمند داده باید بداند که کدام تکنیک را می توان در چه زمانی و چگونه به کار برد. او نباید بیش از حد روی جزئیات اساسی ریاضی (مثلاً بهینه سازی) تمرکز کند، بلکه باید درک خوبی از مسئله تحلیلی که یک تکنیک می تواند آنرا حل کند و همچنین چگونگی تفسیر نتایج آن، داشته باشد. در این زمینه، آموزش مهندسان در علوم کامپیوتر ویا کسب و کار و مهندسی صنایع باید با هدف کسب یک دیدگاه یکپارچه و چند رشته‌ای صورت پذیرد و فارغ‌التحصیلان باید توانمندی در زمینه استفاده از تکنیک‌ها و همچنین هوش تجاری لازم برای به ثمر رساندن تلاش‌های جدید برخورد ار باشند . همچنین مهم است که زمان کافی برای اعتبارسنجی نتایج تحلیلی به‌دست‌آمده صرف شود تا از موقعیت‌هایی که اغلب به عنوان ماساژ داده‌ها[2] ویا شکنجه داده‌ها[3] نامیده می‌شود، که در آن داده‌ها (عمداً) اشتباه ارائه می‌شوند ویا تمرکز بیش از حد صرف بحث در مورد همبستگی‌های جعلی می‌شود، اجتناب شود. هنگام انتخاب روش کمی بهینه، دانشمند متخصص درعلوم داده های تقلب باید ویژگی های زمینه ای و مشکل یا برنامه تشخیص تقلب را در نظر بگیرد. الزامات معمولی برای مدل‌های کشف تقلب در بخش قبلی مورد بحث قرار گرفته‌اند و دانشمند داده‌های تقلب باید درک و احساس اساسی برای همه آنها داشته باشد. بر اساس ترکیبی از این الزامات، دانشمند داده باید قادر به انتخاب بهترین تکنیک تحلیلی برای حل مشکل تجاری خاص باشد.

۲-۱۱-۱. یک متخصص درعلوم داده های کلاهبرداری باید یک برنامه نویس خوب باشد

طبق تعریف، دانشمندان داده با داده ها کار می کنند. این شامل فعالیت‌های زیادی مانند نمونه‌برداری و پیش‌پردازش داده‌ها، تخمین مدل و پردازش پس از آن است (به عنوان مثال، تجزیه و تحلیل حساسیت، استقرار مدل، آزمون‌های برگشتی، اعتبارسنجی مدل). اگرچه امروزه بسیاری از ابزارهای نرم افزاری کاربرپسند برای

[1] Baesens et al. 2015
[2] data massage
[3] data torture

خودکارسازی و پشتیبانی از این وظایف در بازار موجود است، هر مورد تحلیلی لازم است تا طی مراحلی متناسب با ویژگی های یک مشکل و تنظیمات تجاری خاص تبدیل شود. برای انجام موفقیت آمیز این مراحل باید برنامه نویسی انجام شود. از این رو، یک دانشمند داده خوب باید دارای مهارت های کافی در زمینه برنامه نویسی داشته باشد (به عنوان مثال، SAS، R، Python و غیره). زبان برنامه نویسی به خودی خود اهمیت چندانی ندارد، تا زمانی که او با مفاهیم اولیه برنامه نویسی آشنا باشد و بداند چگونه از آنها برای خودکارسازی کارهای تکراری یا انجام روال های خاص استفاده کند.

۳-۱۱-۱. یک دانشمند داده کلاهبرداری باید در مهارت های ارتباطی و تجسمی بالایی داشته باشد

چه بخواهیم چه نخواهیم، تجزیه و تحلیل یک تمرین تکنیکی است. هم اکنون شکاف بزرگی بین مدل های تحلیلی و کاربران تجاری وجود دارد. برای پر کردن این شکاف، امکانات ارتباطی و تجسمی بسیار کلیدی است. از این رو، دانشمندان داده باید بدانند که چگونه مدل‌های تحلیلی و آماری و گزارش‌های همراه آن‌ها را به روش‌های کاربرپسند با استفاده از رویکردهای چراغ راهنمایی[1]، امکانات OLAP (پردازش تحلیلی آنلاین)، قوانین تجاری If-then و غیره نشان دهند. آنها باید قادر باشند مقدار مناسبی از اطلاعات را بدون گم شدن در جزئیات پیچیده (مثلاً آماری) منتقل کنند، بدون اینکه مانع از استقرار موفقیت آمیز یک مدل می شود. با انجام این کار، کاربران تجاری ویژگی ها و رفتار در داده های (کلان) خود را بهتر درک خواهند کرد که نگرش آنها را نسبت به مدل های تحلیلی و پذیرش آن ها بهبود می بخشد. مؤسسات آموزشی باید تعادل را بیاموزند، زیرا بسیاری از دوره های تحصیلی در برگیرنده دانش‌آموزانی هستندکه به دانش تحلیلی یا عملی بیش از حد تمایل دارند.

۴-۱۱-۱. یک دانشمند داده تقلب باید درک تجاری خوبی داشته باشد

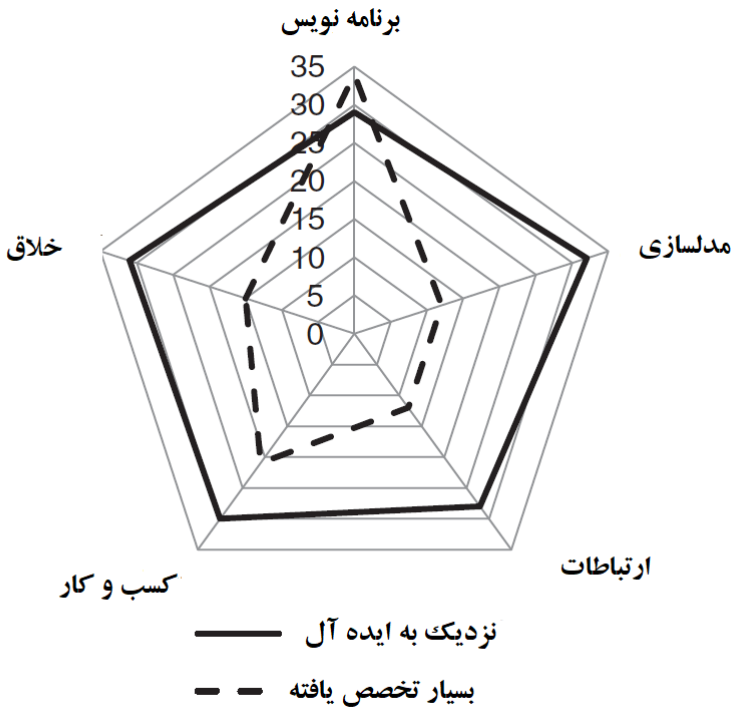
اگرچه این ممکن است واضح باشد، ما شاهد (بیش از حد) بسیاری از پروژه های علم داده بوده ایم که شکست خورده اند زیرا تحلیلگر مربوطه مشکل تجاری را درک نکرده است. در اینجا منظور "کسب و کار" به حوزه کاربردی مربوطه است. چندین نمونه از چنین حوزه های کاربردی تکنیک های کشف تقلب در جدول۱-۱ خلاصه شده است. هر یک از این زمینه ها ویژگی های خاص خود را دارند که برای دانشمند داده های تقلب مهم است که آنها را بداند و درک کند تا بتواند یک سیستم تشخیص تقلب سفارشی طراحی و پیاده سازی کند. هرچه

[1] traffic-light approaches

سیستم تشخیص با محیط هماهنگ تر باشد، همانطور که در هر یک از ابعادی که قبلاً مورد بحث قرار گرفت ، عملکرد آن بهتر خواهد بود.

۵-۱۱-۱. دانشمند داده های کلاهبرداری باید خلاق باشد

یک دانشمند داده حداقل در دو سطح به خلاقیت نیاز دارد. اول، در سطح فنی، مهم است که در انتخاب ویژگی، تبدیل داده ها، و تمیز کردن داده خلاق باشد. این مراحل فرآیند تجزیه و تحلیل استاندارد باید با هر برنامه خاص تطبیق داده شود، و اغلب "حدس درست" می تواند تفاوت بزرگی ایجاد کند. دوم، داده های بزرگ و تجزیه و تحلیل در یک زمینه به سرعت در حال توسعه است. مشکلات، فناوری‌ها و چالش‌های جدید به‌طور مداوم ظاهر می‌شوند. علاوه بر این، کلاهبرداران نیز بسیار خلاق هستند و تاکتیک ها و روش های خود را به طور مداوم تطبیق می دهند. بنابراین بسیار مهم است که یک دانشمند داده‌های کلاهبرداری با این تحولات و فناوری‌های جدید همراه باشد و خلاقیت کافی برای ایجاد فرصت‌های جدید داشته باشد. شکل ۱-۶ صوصیات و نقاط قوت کلیدی را که نمایه دانشمند علوم داده تقلب ایده آل را تشکیل می دهند، خلاصه می کند.



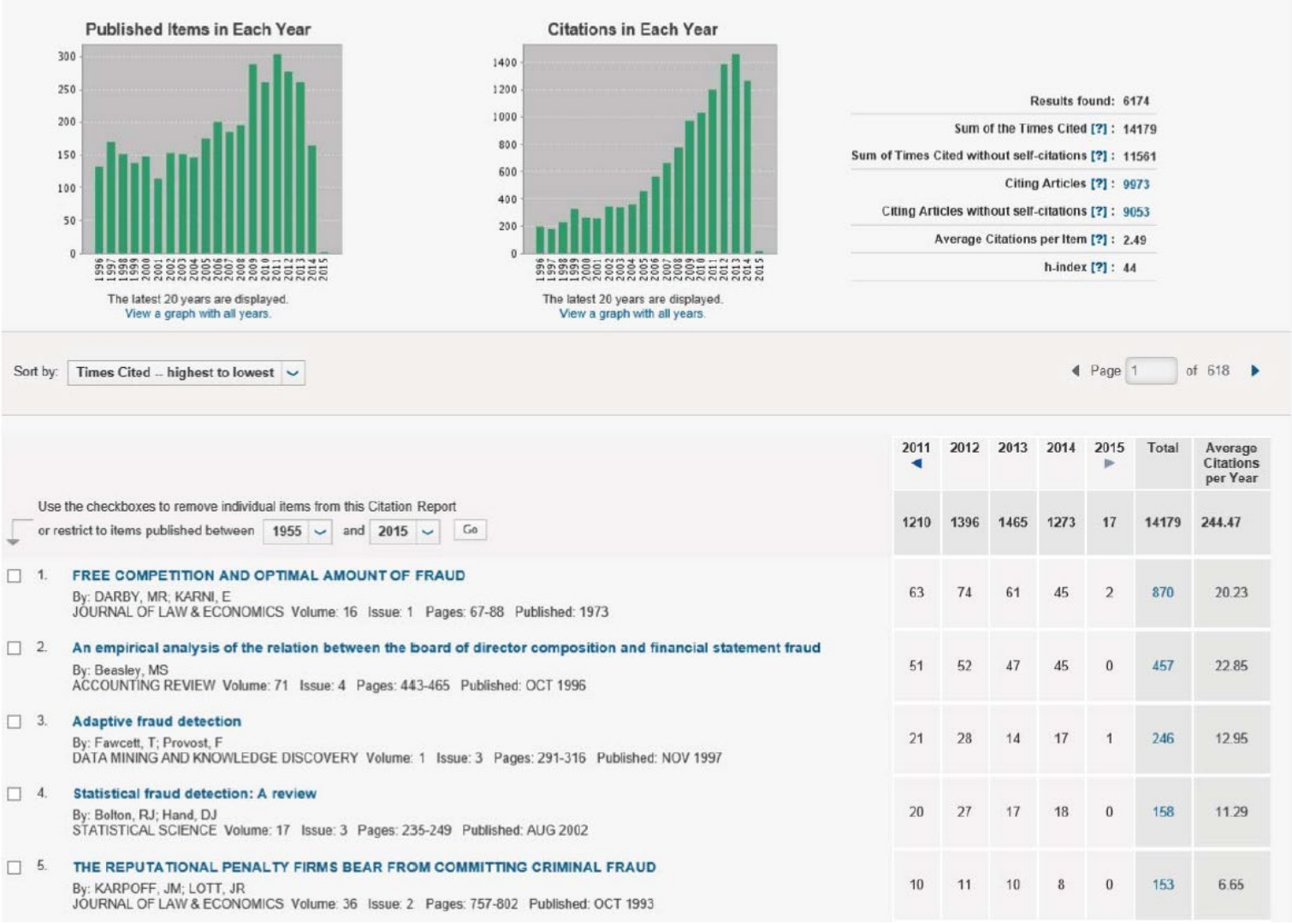
شکل ۱- ۶ مشخصات یک متخصص علوم داده کلاهبرداری

۱۲-۱. دیدگاهی علمی در مورد تقلب

در پایان این فصل، یک دیدگاه علمی از تحقیقات در مورد تقلب ارائه می نماییم. شکل۷-۱ تصویری از تار نمای آمار علم[1] را نشان می دهد که مشتمل بر آمار علم مربوط به همه انتشارات علمی بین ۱۹۹۶ و ۲۰۱۴ برای واژه کلیدی کلاهبرداری است. این شامل آمار مقالات منتشر شده در هر سال، تعداد استنادها و پنج مورد برتر مقالات پر استناد به شرح زیر می باشد:

- ۶۱۷۴ مقاله علمی با موضوع تقلب در طول دوره گزارش منتشر شده است.
- شاخص h ۴۴ است، یعنی که حداقل ۴۴ مقاله با ۴۴ نقل قول در مورد تقلب منتشر شده است.
- تعداد نشریات به طور پیوسته در حال افزایش است که نشان دهنده علاقه روزافزون جامعه دانشگاهی و تحقیقات به موضوع تقلب می باشد.
- نقل قول ها به طور تصاعدی در حال رشد هستند که مرتبط است با افزایش تعداد انتشارات
- دو مقاله از پنج مقاله ذکر شده به مطالعه استفاده از تجزیه و تحلیل برای کشف تقلب، به وضوح نشان دهنده توجه روزافزون در زمینه برای رویکردهای داده محور برای کشف تقلب می باشد.

[1] Web of Scienc



شکل ۱- ۷ تصویری از Web of Science Statistics برای انتشارات علمی در مورد تقلب بین سال های ۱۹۹۶ و ۲۰۱۴

مأخذ: (Baesens et al. 2015)

۲. فصل دوم : جمع آوری داده ها، نمونه برداری و پیش پردازش

۳. منابع و مآخذ:

Baesens, Bart (2014): Analytics in a big data world. The essential guide to data science and its applications / Bart Baesens. Hoboken, New Jersey: Wiley (Wiley & SAS business series).

Baesens, Bart; van Vlasselaer, Veronique author; Verbeke, Wouter author (2015): Fraud analytics using descriptive, predictive, and social network techniques. A guide to data science for fraud detection / Bart Baesens, Veronique Van Vlasselaer and Wouter Verbeke. 1st. Hoboken, New Jersey: John Wiley & Sons (Wiley & SAS business series).