

داده کاوی و بانکداری الکترونیکی

لیلا مقیمی^۱ و مهدی جوانمرد^۲

دانشجوی کارشناسی ارشد دانشگاه پیام نور: Lmoghimi93@gmail.com

استادیار دانشگاه پیام نور: Javanmard@pnu.ac.ir

خلاصه

امروزه داده‌ها قلب تپنده فرایند تجاری بیشتر شرکت‌ها تلقی می‌شوند، آن‌ها فارغ از خرد و کلان بودن نوع صنعت در تمامی صنایع نظیر ارتباطات، تولید، بیمه، کارت اعتباری و بانکداری از طریق تعاملات در سیستم‌های عملیاتی شکل می‌گیرند. لذا نیازی به ابزاری است که بتوان داده‌های ذخیره شده را پردازش کرده و اطلاعات حاصل از آن را در اختیار کاربران قرار دهد. در این راستا سازمان‌های بسیاری در حال استفاده از داده کاوی برای کمک به مدیریت تمام فازهای ارتباط با مشتری هستند. امروزه به دلیل وجود بانکداری الکترونیکی ثبت اطلاعات تراکنشی راحت‌تر صورت می‌گیرد و همین امر موجب شده است روش‌های کمی جایگزین روش‌های کیفی شود. در این تحقیق به منظور شناسایی مشتریان بانک و تدوین استراتژی مناسب برای برخورد با آنها از داده کاوی و ابزار استفاده می‌شود. بکارگیری تکنیک‌های شناسایی تقلب به منظور جلوگیری از قدمات متقلبانه در سیستم‌های بانکداری الکترونیک لازم است. عموماً روش‌های شناسایی تقلب به دو دسته اصلی تشخیص ناهنجاری و تشخیص سوء استفاده تقسیم می‌شوند. در روش تشخیص ناهنجاری، تاریخچه رفتار مشتری به عنوان یک ناهنجاری یا تقلب ثبت شود. روش سوء استفاده بر رفتارهای خاص مشتری تمرکز دارد و دقیقاً رفتارهای شناخته شده را تقلب فرض می‌کند. در این تحقیق علاوه بر مقایسه روش‌های فوق و 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 هر یک به تفصیل شرح خواهد شد.

کلمات کلیدی: داده کاوی - بانکداری الکترونیک - مدیریت ارتباط با مشتری - تقلب - تشخیص ناهنجاری - تشخیص سوء استفاده

مقدمه :

از سال ۱۹۵۰ به بعد که رایانه، در تحلیل و ذخیره سازی داده به کار گرفته شد حجم اطلاعات ذخیره شده در آن پس از حدود ۲۰ سال دو برابر شد همزمان با پیشرفت فناوری اطلاعات، حجم اطلاعات ذخیره شده، همچنان با سرعت بیشتری نسبت به گذشته، بیشتر و بیشتر می‌گردد. با وجود شبکه جهانی وب، سیستم‌های یکپارچه اطلاعاتی، سیستم‌های یکپارچه بانکی، تجارت الکترونیکی و لحظه به لحظه به حجم داده‌ها در پایگاه داده‌ها اضافه گردیده و باعث به وجود آمدن توده‌های عظیمی از داده‌ها گردید، به طوریکه ضرورت کشف و استخراج سریع و دقیق دانش از این پایگاه داده‌ها بیش از پیش احساس می‌گردد. استخراج اطلاعات مناسب از میان انبوه داده‌ها و تبدیل آنها به دانش مورد نیاز سازمان به ویژه در تصمیم‌گیری‌های سازمانی و دیدن آینده درست، نیازمند استفاده از روش‌های نوین در این حوزه است. داده کاوی (Data Mining) یکی از این ابزارها و رویکردهاست که در فضای مدیریت دانش سازمانها به کشف از پایگاه داده‌ها کمک می‌کند. استخراج اطلاعات مفید از پایگاه داده و تبدیل کردن اطلاعات به نتایج عملی، چالش اصلی است که سازمانها با آن روبرو هستند. با توجه به پیشرفت کشور در زمینه فناوری اطلاعات و نگاه‌های ویژه به دولت الکترونیک و نفوذ استفاده از سیستم‌های رایانه‌ای در صنعت و ایجاد بانک‌های اطلاعاتی بزرگ، توسط ادارات دولتی، بانکها و بخش‌های خصوصی، نیاز به استفاده از داده کاوی به طور عمیق احساس می‌شود.

برای رقابت موثرتر در بازارهای رقابتی دنیا، بانکها بایستی درک بهتری از مشتریان و بازار داشته باشند. صنعت بانکداری در دنیا تغییرات زیادی را در نحوه انجام فعالیتهای خود متحمل شده است. بانکهای پیشرو از ابزارهای داده کاوی برای تقسیم بندی مشتریان، اعتبار سنجی و تأیید آنها، پیش بینی عدم پرداخت بدهی ها، بازاریابی و شناسایی الگوهای کلاهبرداری استفاده می کنند. در این مقاله سعی شده تا کاربرد داده کاوی در صنعت بانکداری به عنوان ابزار رقابتی مورد بررسی قرار گیرد. در این مقاله قصد داریم که استفاده از داده کاوی را در قسمتهای مختلف بانکداری الکترونیک من جمله بازاریابی و حفظ و جذب مشتریان و همچنین تشخیص کلاهبرداری و سوء استفاده را مورد بحث و بررسی قرار دهیم.

داده کاوی چیست ؟

داده کاوی فرآیند اکتشاف، تجزیه و تحلیل خودکار یا نیمه خودکار از مقادیر زیادی از داده ها به منظور کشف الگوها و قوانین معنادار می باشد. داده کاوی، فرآیند استخراج و تشخیص الگوهای پنهان یا اطلاعات از پایگاه داده می باشد. به بیان بهتر تجزیه و تحلیل ماشینی داده ها برای پیدا کردن الگوهای مفید، تازه و قابل استناد در پایگاه داده های بزرگ، داده کاوی نامیده می شود. داده کاوی نقش کلیدی در تقسیم بندی بازار، سرویس های مشتری، تشخیص تقلب و کلاهبرداری و همچنین محک زدن رفتار مشتریان دارد. داده کاوی پل ارتباطی میان علم آمار، علم کامپیوتر، هوش مصنوعی، الگو شناسی و فراگیری ماشین داده می باشد. داده کاوی به صورت یک محصول قابل خریداری نمی باشد، بلکه یک رشته علمی و فرآیندی است که بایستی به صورت یک پروژه پیاده سازی شود. در داده کاوی معمولاً به کشف الگوهای مفید از میان داده ها اشاره می شود. منظور از الگوهای مفید، مدلی در داده هاست که ارتباط میان یک زیر مجموعه از داده ها را توصیف می کند و معتبر، ساده، قابل فهم و جدید می باشد. در صنعت بانکداری الکترونیکی، با افزایش چشم گیر روش های بانکداری اینترنتی، هر روز داده های زیادی تولید و ذخیره می شوند. افزایش سرعت کامپیوترها باعث به وجود آمدن الگوریتم هایی شده است که قدرت تجزیه و تحلیل بسیار بالایی دارند بدون اینکه محدودیتی در زمینه ظرفیت و سرعت کامپیوترها داشته باشند. در داده کاوی، هوش مصنوعی، یادگیری خودکار، تئوری پایگاه داده ها و علم آمار درهم آمیخته است.

dataacademy.ir

موانع و چالش ها

با وجود مزایای فراوان که داده کاوی برای مدیریت دانش سازمانها و دست یابی به دانش برای تصمیم گیری سازمان دارد، این حوزه با موانع و چالش هایی روبروست که عبارتند از:

- قدرت محدود محاسبات ریاضی برای بدست آوردن محاسبات ریاضی مورد نیاز الگوریتم های داده کاوی
- عدم مطلوبیت و جذابیت بیشتر الگوها
- تمرکز بیش از حد بر الگوریتم ها
- وجود داده های آلوده و کسب نتایج کاملاً غلط
- هزینه نسبتاً گران سرمایه گذاری در همه حوزه ها
- عدم پوشش کامل همه حوزه ها

مراحل داده کاوی

فهم درست داده و مسأله از مهم ترین و اساسی ترین گامهای ابتدایی داده کاوی می باشد. بدون این فهم درست هیچ الگوریتمی صرفنظر از تجربه بودن آن نمی تواند نتیجه مطمئنی برای شما حاصل نماید و داده جهت کاوش آماده نموده یا نتایج را به طور صحیح تفسیر نمود. برای استفاده بهتر از داده کاوی باید یک بیان واضح از هدف داشت.

عمل داده کاوی از یک پایگاه داده دارای یازده مرحله است :

- 1- تبدیل مسأله کسب و کار و تجارت به یک مسأله داده کاوی
- 2- انتخاب داده های مناسب
- 3- شناخت داده ها
- 4- ساخت مجموعه مدل
- 5- رفع مشکلات داده ها

- ۶- تبدیل داده ها برای استخراج اطلاعات
- ۷- ساخت مدلها
- ۸- ارزیابی مدل ها
- ۹- پیاده سازی مدلها
- ۱۰- ارزیابی نتایج
- ۱۱- شروع دوباره

جمع آوری داده ها ، انتخاب و آماده سازی داده ها ، داده کاوی ، تفسیر و ارزشیابی الگوها و در نهایت ارائه ی دانش ، مراحل مختلف استخراج و کشف دانش را تشکیل می دهد.

به دلیل حجم گسترده و بسیار زیاد داده های موجود در بانک و همچنین عدم امکان ذخیره سازی آنها و نیاز به پردازش و استخراج پویای اطلاعات و دانش نهفته در داده ها ، بحث داده کاوی جریان داده ها در بانک طی سالهای اخیر بسیار مورد توجه قرار گرفته است.

امروزه به دلیل بهره گیری از فناوریهای رایانه ای و استفاده از پایگاه داده ها و داده کاوی این امکان فراهم است که بانک برای یکایک مشتریان خود برنامه ای داشته باشد و بازاریابی در مقیاس وسیعی قابل اجرا و اقتصادی در نظر گرفته شود. شناسایی تکنیک های داده کاوی در بانکداری الکترونیکی منجر به تولید سامانه های مشتری محور و امن می گردد و توجه به جزئیات داده ها منجر به خلق مزیت رقابتی پایدارتر و غیرقابل الگو برداری می شود. چهار جنبه مهم داده کاوی در بانکداری الکترونیکی شامل : بازاریابی ، مدیریت ارتباط با مشتریان ، تشخیص کلاهبرداری و مدیریت ریسک دارائنها می باشد.

از کاربردهای داده کاوی در بانک می توان به نمونه های زیر اشاره نمود :

- کشف پولشویی
 - تشخیص مشتریان ثابت و همیشگی
 - تشخیص مشتریان استفاده کننده از یک سرویس خاص
- پیش بینی الگوهای کلاهبرداری از طریق کارتهای اعتباری ، تعیین میزان استفاده از کارتهای اعتباری بر اساس گروههای اجتماعی فروش و بازاریابی تجزیه و تحلیل سهام و سرمایه ، تعیین مشتریان وفادار ، مدیریت ریسک و پیش بینی فروش بیمه ، تجزیه و تحلیل دعوی و پیشگویی میزان جریمه بیمه نامه های جدید توسط مشتریان دیگر حوزه ها ، ورزش و سرگرمی ، فضانوردی و دیگر حوزه های علم را می توان از دیگر کاربردهای داده کاوی نام برد.

روش های داده کاوی

داده کاوی به سه دسته کلی تقسیم می شود که عبارتند از : خوشه بندی ، طبقه بندی و کشف قواعد وابستگی

۱- خوشه بندی Clustering

فرآیند خوشه بندی سعی دارد که یک مجموعه داده را به چندین خوشه (گروه) تقسیم نماید به طوری که داده های قرار گرفته در یک خوشه با یکدیگر شبیه بوده و با داده های خوشه های دیگر متفاوت باشند. در حال حاضر روشهای متعددی برای خوشه بندی داده ها وجود دارد که بر اساس نوع داده ها ، شکل خوشه ها ، فاصله داده ها و غیره عمل خوشه بندی را انجام می دهند.

۲- طبقه بندی

فرآیند طبقه بندی در واقع نوعی یادگیری با ناظر می باشد که در طی دو مرحله انجام می گردد. در مرحله اول مجموعه ای از داده ها که در آن هر داده شامل تعدادی خصوصیات دارای مقدار و یک خصوصیت بنام کلاس می باشد، برای ایجاد یک مدل داده به کار می روند که این مدل داده در واقع توصیف کننده ، مفهوم و خصوصیات مجموعه داده هایی است که این مدل از روی آنها ایجاد شده است . مرحله دوم فرآیند طبقه بندی اعمال یا بکارگیری مدل داده ایجاد شده بر روی داده هایی است که شامل تمام خصوصیات داده هایی که برای ایجاد مدل بکار گرفته شده اند ، می باشد به جز خصوصیات کلاس این مقادیر که هدف از عمل طبقه بندی نیز تخمین مقدار این خصوصیات می باشد.

الگوریتم ها و روش های مختلفی برای طبقه بندی تاکنون پیشنهاد شده اند که برای مثال می توان از روشهای طبقه بندی با استفاده از شبکه های عصبی ، طبقه بندی مبتنی بر قواعد و نام برد.

برخی نرم افزارهای آماری و برخی دیگر از زاویه دید پایگاه داد ها طراحی شده اند را می توان نام برد : نرم افزارهای WEKA ،
Rapid Miner ، Mineset ، Clementine ، SQL Server 2005 ، Enterprise Miner ، Intelligent Miner

بانکداری الکترونیک

استفاده از پیشرفتهای فناوری اطلاعات و ارتباطات در زمینه کسب و کار موجب فناوری نوینی به نام بانکداری الکترونیکی گردیده است . بکارگیری متون تجارت الکترونیک از طرفی سبب ساده سازی انجام کسب و کار و کاهش بنگاه های اقتصادی گردیده و از طرف دیگر افزایش مشتریان به تبع آن افزایش سطح رضایت مندی آنان را در بردارد . تجارت الکترونیکی شامل فعالیتهای متنوعی مانند تجارت الکترونیکی کالا و خدمات ، تحویل آنلاین کالا ، انتقال پول الکترونیکی ، خدمات اینترنتی و ... می باشد .

برای بانکداری الکترونیکی تعاریف گوناگونی ارائه شده که می توان به تعاریف زیر اشاره کرد :

- ارائه کلیه خدمات بانکی به صورت الکترونیکی و بدون نیاز به حضور فیزیکی مشتریان در شعب بانک
- استفاده از فناوری های اطلاعات (IT) و ارتباطات (ICT) در ارائه خدمات بانکی
- بکارگیری نرم افزار ها و سخت افزارها مبتنی بر شبکه های ماهواره ای ، مخابراتی و یا بی سیم (Wireless) برای تبادل منابع و اطلاعات مالی

شاخه های بانکداری الکترونیک :

- بانکداری اینترنتی (با استفاده از صدور رمز یکبار مصرف Token)
- بانکداری مبتنی بر تلفن همراه و فناوری مرتبط با آن (همراه بانک و استفاده از سرویس USSD مخابرات)
- بانکداری مبتنی بر تلفن (تلفن بانک)
- بانکداری مبتنی بر دستگاه های خودپرداز
- بانکداری مبتنی بر پایانه های فروش
- بانکداری مبتنی بر شعبه های الکترونیکی (باجه هایی بدون حضور کلر و مستخدم بانک - ایستگاه های شبانه روزی)

مدیریت ارتباط با مشتری (CRM) (Customer Relationship management)

با بهره گیری از CRM ، ارتباط مشتریان با سازمان و نیازمندیهای آنان مورد بررسی و تجزیه و تحلیل اصولی قرار می گیرد. در واقع CRM فرآیندی جهت گردآوری و یکپارچه سازی اطلاعات به منظور بهره برداری موثر و هدف دار از آنها می باشد. این اطلاعات می تواند در برقراری ارتباط با مشتریان ، فروش ، بازاریابی موثر ، حساسیت و یا نیازهای بازار باشد. مدیریت ارتباط با مشتری ، یک راهبرد کسب و کار است که هدف آن طبقه بندی مشتریان و مدیریت آنها به منظور بهینه سازی ارزش مشتری در دراز مدت و بهره گیری سازمان از آن است. مدیریت ارتباط با مشتری ، در واقع فرآیندهای پیدا کرن مشتری ، نزدیک شدن به آن ، مدیریت و ایجاد رضایت در مشتریان و نگهداری آنهاست . مدیریت روابط مشتری تحلیلی شامل استفاده ی داده های مشتری برای آنالیز ، مدل سازی و ارزیابی می باشد. همچنین روشهایی مانند داده کاوی (OLAP) (Online Analytical Processing) و نرم افزارهای مدیریت روابط مشتری را اجرا می کند. داده کاوی می تواند مدلی را ارائه نماید که قادر به بهره برداری از رفتار مشتریان بوده و این رفتار را پیش بینی می کند .

پیش بینی رویگردانی

وفادارای مشتریان به شکلی است که سازمانهای ارتباطاتی بایستی به این مهم رسیدگی کنند. این موضوع می تواند ده برابر جذب یک مشتری به جای حفظ مشتری موجود، هزینه در برداشته باشد. از سوی دیگر، هزینه ی حفظ مشتری کمتر از هزینه ی برگرداندن آن پس از ترک سازمان است. به همین دلیل است که داده کاوی با هدف پیش بینی رویگردانی مورد استفاده قرار گرفته است. با استفاده از داده کاوی، تولید لیستی از مشتریان با احتمال بالای ترک سازمان امکان پذیر است. تکنیک های داده کاوی به سازمانهای ارتباطاتی جهت شناسایی الگوهای رفتاری مشتریان در رویگردانی از سازمان کمک میکند که توسط رقبا فریب خورده و اقدام به ترک سازمان می کنند. در این روش ، آنها اقدام به برنامه ریزی عملیاتی خاص، جهت پیش بینی رویگردانی در سازمانها میکنند. چندین تکنیک متفاوت برای هدف پیش بینی رویگردانی مورد استفاده قرار گرفته است. اما معروف ترین آنها موارد زیر هستند:

- ✓ درخت تصمیم گیری
- ✓ شبکه های عصبی
- ✓ رگرسیون

پیشنهادات

- ۱- افزایش اطلاعات مدیران از تکنولوژی داده کاوی و مزایای آن
- ۲- وجود زیر ساخت اطلاعاتی یکپارچه و پایگاه اطلاعاتی نرمال
- ۳- آگاهی برنامه نویسان بانک های اطلاعاتی از نیازمندی بانک به برخی اطلاعات ضروری جهت اجرا تکنولوژی جدید
- ۴- ثبت کامل تمامی اطلاعات در فیلد بانک اطلاعاتی
- ۵- افزایش آگاهی مدیران نسبت به سود آوری حاصل از ارزیابی
- ۶- توجه به مسائل کیفیت عملکرد و نحوه گزینش مدیریت برای پست سازمانی
- ۷- افزایش دوره های نظرسنجی از مشتریان و احترام به درخواست و نیاز مشتری

نقش داده کاوی در مدیریت ریسک صنعت بانکداری

داده کاوی به طور گسترده برای مدیریت ریسک در صنعت بانکداری مورد استفاده است. مدیران اجرایی بانکها لازم است بدانند مشتریانی که با آنها معامله می کنند مورد اعتماد هستند یا نه. ارائه کارتهای اعتباری به مشتریان جدید، گسترش مشتریان اعتباری موجود و تائید پرداخت وام به مشتریانی که اطلاع کافی از درآمد ماهیانه آنها در دسترس نیست، ریسک بالایی برای پرداخت بانکها به همراه دارد. داده کاوی می تواند با شناسایی مشتریانی که احتمال تأخیر در پرداخت اقساط ایشان وجود دارد، در کاهش ریسک به بانکهایی که کارتهای اعتباری صادر می کنند و یا تسهیلات پرداخت می نمایند، کمک کند. این عمل به منظور تحلیل رفتار تراکنشی مشتریان در حسابهای سپرده ایشان و تعیین احتمال تأخیر در بلزپرداخت وام آنها صورت می گیرد. در حقیقت رتبه بندی اعتباری، اولین ابزار توسعه یافته مدیریت ریسک مالی می باشد. رتبه بندی اعتباری می تواند برای واه دهندگان در صنعت بانکداری با ارزش باشد و وام دهندگان نباید بدون ابزار ارزیابی ریسک صحیح، عینی و قابل کنترل وامی پرداخت نمایند. داده کاوی همچنین می تواند رفتار اعتباری وام گیرندگان را با توجه به خصوصیات نظیر پیشینه اعتباری، طول مدت اشتغال و طول مدت اقامت تعیین نمود. مشتریان با سابقه طولانی و همچنین داشتن دستمزد بالا، از موقعیت بهتری نسبت به مشتریان جدیدی که پیشینه ای در بانک نداشته و درآمد بالایی نیز ندارند برخوردار هستند و احتمال دریافت تسهیلات برای مشتریان گروه اول بسیار بیشتر و با ریسک کمتر برای بانک عامل می باشد. با احتمال تأخیر هر مشتری، بانک در موقعیت بهتری برای کاهش ریسک قرار دارد.

نقش داده کاوی در تشخیص کلاهبرداری در صنعت بانکداری

خسارت های غیر مستقیمی که متقلبان به صنعت بانکداری و بیمه وارد می سازند، بسیار بالاتر از رقمی است که این سازمان ها به طور مستقیم متضرر می شوند. تقلب در بانکداری الکترونیک در بستر خدمات الکترونیک و به صورت برخط (online) اتفاق می افتد و حاصل آن، انتقال پول الکترونیکی از یک حساب به حساب دیگر، به صورت نامشروع و غیر قانونی می باشد. امروزه حجم زیادی از معاملات و نقل و انتقالات پولی و مالی در سطح اینترنت و در بستر الکترونیکی انجام می شود و رشد روز افزون این خدمات و تراکنش ها از یک طرف و همچنین ناشناس ماندن مجرمان در بستر اینترنت از طرف دیگر، باعث تشویق و تحریک متقلبان و شیادان به ورود به این حوزه می گردد. به دلیل عدم حضور فیزیکی مشتریان در بستر خدمات الکترونیک، لزوم تشخیص هویت در ارائه این خدمات از دیدگاه موسسات مالی و پولی بسیار حیاتی و با اهمیت است و شاید بتوان ادعا کرد که محدودیت اصلی در ارائه خدمات گسترده تر و وسیع تر بانکی، لزوم تشخیص هویت افراد می باشد بانکهای فراهم آورنده خدمات اینترنتی، روشهای مختلفی را برای تشخیص جرم و غربال نمودن تراکنشهای مشتریان به کار می برند. روشهایی که اخیراً مورد استفاده می شود شامل:

- تصدیق نشانی (AVS) Address Verification Systems
- روش تصدیق کارت (CVM) Card Verification Method
- شماره شناسایی شخصی (PIN) Personal Identification Number
- روشهای زیست سنجی Biometrics

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.

در روش تشخیص ناهنجاری تلاش می شود تا یک مشخصه (profile) از تاریخچه عملکرد برای هر کاربر ایجاد گردیده و سپس از استخراج هرگونه انحراف Deviation به قدر کافی بزرگ در مشخصه کاربر، بروز یک حمله پی برده شود.

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 بودن 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.

تکنیک تشخیص کلاهبرداری و تقلب

تکنیک های متعددی به منظور شناسایی تقلب طراحی و اجرا شده است:

۱- سیستم خبره

سیستم های خبره به گونه ای از سیستم های محاسباتی اطلاق می شود که توانایی ارائه و استدلال در برخی از حوزه های غنی دانش با نگاه حل مشکلات و ارائه راهکار داشته باشد. سیستم های خبره، به کمک قانون $X \rightarrow Y$ مشخص می کنند در چه حالتی، چه اتفاقی باید بیفتد.

۲- برون هشته ای Outlier Detection

سازوکار برون هشته ای به معنی مشاهده و استخراج انحراف های است که تفاوت هایی را با دیگر مشاهدات، تعیین می کند. این روش به دو نوع بدون نظارت Unsupervised Learning Approach و با نظارت Supervised Learning Approach تقسیم می شود. در روش بدون نظارت نیازی به دانش قبلی و تاریخچه اتفاقات و تراکنش های قبلی در پایگاه های داده نیست، لیکن در روش تشخیص تغییرات در رفتار تراکنش های غیر عادی امکان پذیر می باشد و می توان هرگونه تغییری که منجر به تقلب می شود را شناسایی نمود.

در تکنیک های با نظارت، مدل هایی طراحی می شوند که می توانند بین رفتارهای تقلب گونه (مشکوک) و رفتار های عادی و واقعی تفاوت قائل شوند. این روش ها به شناسایی دقیق از تراکنش های تقلب گونه در تاریخچه بانک اطلاعاتی نیاز دارند. به عبارتی دیگر، جهت استفاده از این روش ها حتما باید تاریخچه ای از اطلاعات در بانک اطلاعاتی داشت تا بتوان، با مقایسه این داده ها، رفتارهای غیر عادی را شناسایی کرد. لذا این روش فقط و فقط می تواند تقلب هایی را شناسایی نماید که حداقل یک بار در گذشته رخ داده و تاریخچه آنها نیز در بانک اطلاعاتی موجود است.

۳- شبکه عصبی Neural Network

یک شبکه عصبی، مجموعه ای از گره های node به هم متصل شده است که با تقلید از کارکرد مغز انسان طراحی شده اند. هر گره ارتباطات وزن داری Weighted Connection به چندین گره دیگر در لایه های مجاور دارد.

در شبکه های عصبی، به صورت نرم افزاری، ساختار داده ای طراحی می شود که می تواند همانند نورون عمل نماید؛ به این ساختار داده ها، گره گفته می شود. سپس با ایجاد شبکه ای بین این گره ها و اعمال یک الگوریتم آموزشی به آن، شبکه را آموزش می دهند. در این حافظه یا شبکه عصبی گره ها دارای دو حالت فعال (روشن یا ۱) و غیر فعال (خاموش یا ۰) می باشند و هر یال (ارتباط بین گره ها) دارای یک وزن است. یال های با وزن مثبت، موجب تحریک یا فعال کردن گره غیر فعال بعدی می شوند و یال های با وزن منفی، گره متصل بعدی را غیر فعال (در صورتی که فعال بوده باشد) می کنند. اگر این سیستم یک تراکنش تقلب گونه را روی کارت تشخیص دهد، بلافاصله با دارنده کارت یک تماس تلفنی گرفته خواهد شد و اگر دارنده کارت، تقلب روی کارت را تایید کند، به منظور جلوگیری از وقوع تقلب، کارت، بلافاصله بلوکه خواهد شد.

این سیستم قادر است با استفاده از شبکه های عصبی الگوی خرج کردن صاحب کارت را آموزش دیده و هرگونه مغایرتی در روش و چگونگی پرداخت پرداخت وجه را تشخیص داده و به عنوان تقلب تلقی نماید. در طراحی و توسعه سیستم پیش بینی فالکون، تکنیک ها و فناوری های یادگیری ماشینی Machine Learning، تشخیص الگوی تطبیقی Adaptive Pattern Recognition، شبکه های عصبی و مدل های آماری Statistical Models نقش داشته اند.

۴- استدلال بر پایه مدل

استدلال بر پایه مدل یک تکنیک تشخیص سوء استفاده که حملات را از طریق فعالیت های قابل مشاهده ای که از طریق یک امضای حمله Attack Signature استنتاج می شود، تشخیص می دهند. برای این منظور، به یک بانک اطلاعاتی از سناریوی حملات یا دنباله ای از رفتار حملات نیاز است

دقیقا مشابه روال کار نرم افزارهای آنتی ویروس که از روی امضای هر ویروس روی فایل ها ، پی به وجود ویروس می برند ، این تکنیک نیز از طریق امضا و بانک اطلاعاتی که در اختیار دارد ، حمله را شناسایی می کند.

۵- رویکرد مبتنی بر قواعد Rule-Based Approach

این روش ترکیبی است از کاربرد های تجزیه و تحلیل مطلق و تفاضلی در تجزیه و تحلیل تفاضلی ، یک سری معیارهای قابل انعطافی می توانند پیاده سازی شوند تا هر گونه تغییری در جزئیات تاریخچه رفتار یک کاربر را شناسایی نمایند.

مدیریت کردن این روش کاری بسیار دشوار است و این مساله به دلیل این است که پیکربندی مناسبت قواعد ، نیازمند برنامه نویسی زمان بر ، دقیق و پر زحمتی برای هر امکان تقلب قابل تصور می باشد. یکی دیگر از این روش ها ، الگوریتم ژنتیک Genetic Algorithm است که منظور تشخیص حملات مخرب و جداسازی آنها از استفاده های عادی و نرمال به کار می رود. الگوریتم ژنتیک روشی از هوش مصنوعی است با تکیه بر حل مساله که بر اساس فرضیه تکامل داروین عمل کرده و در ریاضیات کاربرد وسیعی دارد. این الگوریتم به گونه ای است که در آن ، هر فردی به عنوان یک مدار رفتاری ممکن عمل می کند لذا این رویکرد یک نرخ تشخیص بالا و همچنین یک نرخ هشدار اشتباه پایین را فراهم می کند.

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 بر اساس امضای می توان رفتار جاری مشتریان را بررسی نمود طبیعتا 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.

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. تاریخچه از بحث فنی ذکر این نکته نیز در اینجا بسیار ضروری به نظر می رسد که با توجه به رشد روز افزون خدمات مالی بانک ها و موسسات مالی و اعتباری به صورت الکترونیکی در سطح کشور و افزایش ضریب نفوذ استفاده کاربران از خدمات بانکداری الکترونیکی ؛ رویکرد کلاهبرداران و متقلبان به سمت بانکداری الکترونیکی نیز رو به افزایش است. بدین ترتیب نگرانی های بسیاری را سبب شده و توجه زیادی را به سوی خود جلب کرده 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. روش های اصلی مورد استفاده برای کشف تقلب های مالی عبارتند از مدل های 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 متقلبان ارائه می کنند.

منابع و مراجع

1. Data Mining Concepts and Techniques/mehdiesmaeli /20012
2. An Introduction to Data Mining: <http://www.theartling.com/>, retrieved on Mar 2, 2007.
3. Data Mining: Efficient Data Exploration and Modeling: <http://research.microsoft.com/dmx/DataMining/>, retrieved on Mar 2, 2007.
4. Christine Gertisio and Alan Dussauchoy, "Knowledge Discovery from Industrial Data base", Journal of Intelligent Manufacturing, 15, 29-37, 2004.