دسترسی در سایتِ http://jnrm.srbiau.ac.ir
سال هشتم، شماره سی و پنجم، فروردین و اردیبهشت ۱۴۰۱
شماره شاپا: ۲۵۸۸X-۵۸۸

دانشگاه آزاد اسلامی، واحد علوم و تحقیقات

پژوهش‌های نوین در ریاضی

ارائه یک متدولوژی مبتنی بر نقشه‌های خود سازمانده و شبکه‌های عصبی چندلایه برای رخدادهای مشکوک به پول شویی در سطح شعب بانک‌ها

حمید مهدوی کوچکسرایی[۱]، محمدرضا شهریاری[*۲]، فریدون رهنمای رودپشتی[۳]، سید عبدالله سجادی جاغرق[۴]

(۱) گروه مدیریت منابع انسانی، دانشگاه آزاد اسلامی، واحد امارات، دبی، امارات متحده عربی
(۲) گروه مدیریت صنعتی، واحد تهران جنوب، دانشگاه آزاد اسلامی، تهران، ایران
(۳و۴) گروه مدیریت مالی، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران

تاریخ ارسال مقاله: ۱۳۹۹/۰۵/۰۵ تاریخ پذیرش مقاله: ۱۳۹۹/۰۶/۰۱

چکیده

با توجه به اهمیت سیستم‌های بانکداری و سوء استفاده از این بستر برای مقاصد پولشویی، نیاز مبرم به پیاده‌سازی سیستم‌های ضد پولشویی از طرف دولت‌ها و موسسات سیاست‌گذار در امور اقتصادی مورد توجه است. همچنین با توجه به رشد تروریسم و تقلب‌های سازماندهی شده و از طرفی تصویب قوانین متعدد علیه این موارد نیاز به این سیستم‌ها در حال افزایش است. از سوی دیگر، پیچیدگی رفتارهای مشکوک به پولشویی به گونه‌ای است که بدون ابزاری هوشمند و داده محور نمی‌توان در کشف پولشویی اقدام قابل توجهی انجام داد. نکته مهم و شاید کاربردی در ایران نزدیکی این سیستم‌ها با سیستم‌های ضد رشوه خواری، تقلب، تخلف و سیستم‌های بازرسی است که می‌تواند به عنوان ابزاری کارآمد برای واحد بازرسی بانک تلقی گردد. در این مقاله رویکردی مبتنی بر آنالیز و پردازش داده‌ها پیشنهاد می‌شود. در این رویکرد با استفاده از نقشه‌های خودسازمانده شعب بانک براساس رفتارهای مشابه خوشه‌بندی می‌شوند سپس با استفاده از یک شاخص خطی فرایند برچسب‌گذاری شعب صورت می‌گیرد. در مرحله بعد با استفاده از آموزش یک شبکه عصبی چند لایه، الگویی جهت شناسایی شعب بانک که در آنها فرایندهای مشکوک پولشویی صورت می‌گیرد معرفی می‌شود.

واژه‌های کلیدی: پولشویی، بانک، نقشه‌های خودسازمانده، شبکه عصبی چند لایه.

*. عهده‌دار مکاتبات: Email: shahriari.mr@gmail.com

۱- مقدمه

در دو دهه گذشته، داده‌کاوی، شبکه‌های عصبی، الگوریتم‌های ژنتیکی، و غیره به‌طور گسترده‌ای برای مدل‌سازی مالی مورد استفاده قرار گرفته‌اند. با این حال، تحقیقات کمی در استفاده از این روش‌ها در زمینه تشخیص الگوهای پولشویی وجود دارد.

نویسندگان در [۱] بیان داشته‌اند، درخت تصمیم یکی از پرکاربردترین روش‌های استنباط استقرایی از سال ۱۹۶۰ تا کنون است. در این مقاله روش درخت تصمیم برای تعیین میزان ریسک پولشویی، بر اساس مشخصات مشتری بکار گرفته شده است. بدین صورت که یک نمونه ۲۸تایی با چهار مشخصه برای القای یک درخت تصمیم بکار گرفته شده است. نتایج، کارایی درخت تصمیم را در ساختن قواعد ضدپولشویی از مشخصات گروه‌های مشتریان نشان می‌دهد. رویکرد مقاله به روش درخت تصمیم برای تفکیک مشتریان بدین صورت است که، برنامه "شناخت مشتری" در یکی از بانک‌های چین برای شناسایی افرادی که از بانک‌ها (محصولات و خدمات بانکی) برای پولشویی استفاده می‌کنند، اجرا شده است.

در [۲] نویسندگان برای تشخیص رفتارهای غیر معمول مشتری ماشین بردار پشتیبان[2] [۳] را توسعه داده‌اند. آن‌ها ترکیبی از الگوریتم‌های نظارت شده و بدون ناظر ماشین بردار پشتیبان را ارائه داده‌اند. ماشین بردار پشتیبان یک رده‌ای [۳] یک رویکرد یادگیری بدون ناظر مورد استفاده برای شناسایی نقاط پرت[3] براساس مجموعه داده‌های آموزشی بدون برچسب است که برای مجموعه آموزش پولشویی بسیار مناسب است. مزیت این روش آن است که میتواند با مجموعه داده‌های ناهمگون کار کند. با این حال ارزیابی عملکرد آن براساس مجموعه داده‌های شبیه‌سازی شده برای موارد مشکوک می‌باشد.

پژوهشگران در[۴] فرآیند گسسته را برروی مجموعه داده، برای ساختن خوشه‌ها اعمال کرده است. آن‌ها در ابتدا کل جدول زمانی را به نمونه‌های زمانی مختلف گسسته‌سازی نمودند. از این رو، هر تراکنش به عنوان یک گره در یک فضای جدول زمانی یک بعدی مشاهده شده است. آن‌ها برای تشکیل یک هیستوگرام تمام تراکنشات مشتری را به وسیله سرجمع تراکنش‌ها و فرکانس آن‌ها بر روی محور جدول زمانی طرح‌ریزی کردند، سپس خوشه‌ها را براساس بخش‌های هیستوگرام ایجاد نمودند. همچنین برای تشخیص الگوهای مشکوک تحلیل همبستگی محلی و سراسری را اعمال نمودند. این رویکرد با کاهش مسئله خوشه‌بندی به مسئله بخش‌بندی، پیچیدگی را کاهش می دهد.[۵] بعلاوه، آزمودن تراکنشات برای تشخیص رفتارهای مشکوک مرتبط با تپه‌های غیر طبیعی در هیستوگرام به منظور تحلیل کردن رفتارهای فردی و رفتارهای گروهی مناسب است. اگرچه، از آنجائیکه باید تراکنشات و مشتریان بیشمار را در طول یک دوره زمانی طولانی تحلیل کنیم از آنجائی‌که در هیستوگرام تعداد بسیار کم و یا هیچ تپه اوج وجود ندارد، تشخیص موارد مشکوک مشکل می‌باشد. بنابراین در ابتدا تحلیل سراسری دیگری نیاز است و سپس می‌توان این روش را برای تحلیل‌های آتی در این مورد اعمال نمود.

در مقاله دیگر یک روش تشخیص ناهنجاری هیبرید ارائه می‌گردد، که از تکنیک خوشه‌بندی برای مبنا قرار دادن رفتارهای نرمال مشتری، و از روش‌های آماری برای بیان انحراف از معیار یک تراکنش خاص از رفتار گروه مرتبط با آن، استفاده می‌کند. این

[2] Support Vector Machine
[3] Outlier

رویکرد برای دسته‌بندی مشتریان در خوشه‌های متفاوت از تئوری اقلیدسی[۴] استفاده می کند [۶].
یک ترکیب ازخوشه بندی و پرسپترون چند لایه[۵] توسط [۷] پیشنهاد شده است. نویسندگان یک تکنیک خوشه‌بندی مبتنی بر مرکز ساده برای تشخیص موارد مشکوک پولشویی را ارائه کرده‌اند. این تکنیک براساس دو ویژگی مهم است که به عنوان ورودی فرآیند آموزش پرسپترون چند لایه استفاده شده‌اند. نتایج ابتدائی آن‌ها نشان می‌دهد که رویکرد آنان کارا می‌باشد. اگرچه تعداد ویژگی‌ها و تعداد الگوهای آموزش کاملا کوچک است و این مسئله دقت کلی را متاثر نماید.
در مقاله دیگر مجموعه داده آموزش جدید، برای تبدیل داده بانک به داده‌های مناسب برای الگوریتم CLOPE که در خصوص تکنیک خوشه‌بندی برای داده اسمی (مقادیر رشته‌ای) می‌باشد. به منظور تشخیص موارد پولشوئی، ارائه گردیده است. نتایج آزمایشی نشان می‌دهد که CLOPE یک الگوریتم مناسب برای تشخیص موارد پولشوئی می‌باشد. اما این سیتم نمی‌تواند به تنهائی، به‌طور کامل، اجرا گردد. و باید از توانائی تحلیل گران در تجزیه و تحلیل داده‌ها، و ارائه مجموعه‌ای از قوانین (معیارهای تعیین شده) برای اعتبار بخشی به خوشه‌ها پس از عمل خوشه‌بندی، استفاده نمود[۸].
به منظور شناسایی موارد مشکوک به پولشویی لازم است بانک‌ها و مؤسسات اعتباری دارای یک سیستم نظارت بر گردش عملیات بانکی مشتری باشند. در این راستا و از آنجا که شناسایی و کنترل گردش عملیات مشتری، فرایندی مستمر است؛ لازم است مؤسسات مذکور با نصب نرم افزارهای ویژه مبارزه با پولشویی، اطلاعات اخذ شده از مشتریان را مدیریت کنند. از جمله ویژگی‌های نرم افزارهای فوق، قابلیت جمع‌آوری، تجزیه و تحلیل، نمایش و عرضه اطلاعات و شناسایی موارد مشکوک است. نرم افزارهای مذکور به گونه‌ای طراحی شده‌اند که امکان بررسی و کنترل حساب‌ها و گردش عملیات بانکی مشتریان به صورت مستمر و در کوتاه‌ترین زمان را فراهم می‌سازند [۹].
داده‌کاوی عبارت است از تحلیل داده‌ها به‌منظور کشف روابطِ از قبل نامعلومی که اطلاعات مفیدی ارائه می‌کنند. کشف تقلب به یکی از جا افتاده‌ترین کاربردهای داده‌کاوی چه در صنعت و چه در دولت تبدیل شده است. هر چند بسته‌های نرم‌افزاری تجاری مختلفی با محیط تعاملی آسان برای کاربران، در حال حاضر در دسترس هستند که انجام کارهای پیچیده داده‌کاوی را به‌ظاهر آسان می‌کنند و الگوریتم‌های داده‌کاوی بسیاری برای کشف تقلب مورد استفاده قرار گرفته‌اند، اما کاربرد آن‌ها، همچنان از چارچوب سنتی داده‌کاوی یعنی انتخاب ویژگی، نمایندگی، گرداوری و مدیریت داده‌ها، پیش‌پردازش، داده‌کاوی، پس‌پردازش و ارزیابی عملکرد پیروی می‌کند. پرکاربردترین روش‌های داده‌کاوی برای کشف تقلب‌های مالی عبارتند از مدل‌های رگرسیون لجستیک (رایج‌ترین)، شبکه‌های عصبی، شبکه استنباط بیزین و درخت تصمیم که همه آن‌ها راه‌حل‌های قابل توجهی را برای مشکلات ذاتی در کشف و طبقه‌بندی داده‌های تقلبی ارائه می‌کنند [۱۰].
داده‌کاوی این توان را دارد که سناریوهای جدیدی برای کشف نمونه‌های پولشویی ارائه دهد. داده‌کاوی به عنوان فرایندی خودکار برای استخراج اطلاعات و الگوهای جالب توجه، بارز، ضمنی، از قبل ناشناخته و بالقوه مفید از انبار داده‌های حجیم، شناخته می‌شود. هنگامی که الگوریتم‌ها و تکنیک‌های

[4] Euclidean Adaptive Resonance Theory(TEART)
[5] Multi Layer Perceptron

داده‌کاوی روی چنین تراکنش‌هایی اعمال می‌شوند، الگوهای پنهان از جریان وجوه را پرده برداری می‌کنند [۱۱].

همراه شدن اطلاعات مشتری و دانش فرد خبره منجر به آن می‌گردد که معاملات مشکوک به طور همروند با زمان اتفاق افتاد نشان شناسائی شوند. سیستم ضد پولشویی قابلیت اجرا با الگوریتم‌های نظارت شده و بدون نظارت را به طور هم‌زمان دارد. در این سیستم علاوه بر به کارگیری الگوریتم‌های رایج در داده‌کاوی، شاخص‌هایی محاسبه می‌شوند که می‌توانند بخوبی ملاکی برای میزان مشکوک بودن عملکرد یک مشتری باشند و براساس آن شاخص‌ها، داده‌های پرت (خارج از محدوده) تفکیک و شناسائی شوند.

از جمله مطالعاتی که در سال‌های اخیر جهت شناسایی مشتریان مشکوک به پولشویی بر اساس داده‌کاوی صورت گرفته است شامل مقاله جمشیدی و همکاران [۱۲] می‌شود. آنها یک رویکرد چند هدفه مبتنی بر سیستم استنتاج نورو دیفیوژن سازگار برای شناسایی پولشویی بانکی و مبادله ارز ارائه دادند. مگومدوو و همکاران [۱۳]، دورفیو و همکاران [۱۴] و پلاکسی و همکاران [۱۵] از روش-های یادگیری ماشین مبتنی بر هوش مصنوعی جهت طراحی و نظارت بر سیستم‌های ضدپولشویی بهره برده‌اند. لیت و همکاران [۱۶]، ریدر [۱۷] و تیاواری و همکاران [۱۸] در مقالات مروری مجموعه‌ای غنی از تحقیقات صورت گرفته مبتنی بر داده‌کاوی و هوش مصنوعی جهت مقابله با پولشویی و سایر جرائم بانکی را گردآوری کرده‌اند. کاماد و همکاران در سال ۲۰۱۹ کاربرد تکنیک‌های داده کاوی در سیستم‌های بانکی را با هدف کشف تقلب و مدیریت ارتباط با مشتری مورد بررسی قرار دادند [۱۹]. همچنین در سال ۲۰۲۰ کانچانا و همکاران با استفاده از تکنیک‌های طبقه‌بندی در داده‌کاوی به بررسی حملات فیشینگ به سایت‌های بانکی پرداختند [۲۰]. شهبازی در مقاله‌ی دیگری با استفاده از الگوریتم‌های طبقه بندی مبتنی بر درخت تصمیم یک مدل رتبه‌بندی اعتباری برای مشتریان بانک ارائه دادند [۲۱]. دینسر وهمکاران یک روش مبتنی بر داده کاوی جهت ارزیابی رضایت مشتریان از نرم‌افزارهای موبایلی در صنعت بانکداری ارائه دادند [۲۲]. کور و همکاران نیز از داده کاوی برای پیشبینی ریسک اعتباری مشتریان بانک استفاده کردند [۲۳]. مطالعه دیگری از حسنی و همکارن منتشر شده است که در آن دیجیتال سازی و داده کاوی در سیستم‌های بانکی مورد مطالعه قرار گرفته است [۲۴].

۲- بیان مسئله

با توجه به اهمیت سیستم‌های بانکداری و سوء استفاده از این بستر برای مقاصد پولشویی، نیاز مبرم به پیاده سازی سیستم‌های ضد پولشویی از طرف دولت‌ها و موسسات سیاست‌گذار در امور اقتصادی مورد توجه است. با توجه به رشد تروریسم و تقلب‌های سازماندهی شده و از طرفی تصویب قوانین متعدد علیه این موارد نیاز به این سیستم‌ها در حال افزایش است. از سوی دیگر، پیچیدگی رفتارهای مشکوک به پولشویی به گونه‌ای است که بدون ابزاری هوشمند و داده محور نمی‌توان در کشف پولشویی اقدام قابل توجهی انجام داد. نکته مهم و شاید کاربردی در ایران نزدیکی این سیستم‌ها با سیستم‌های ضد رشوه خواری، تقلب، تخلف و سیستم‌های بازرسی است که میتواند به عنوان ابزاری کارآمد برای واحد بازرسی بانک تلقی گردد.

پولشویی فرآیندی است که طی آن افراد متخلف با استفاده از امکانات مؤسسات مالی و بانک‌ها درآمد نامشروع خود را تطهیر نموده و آنرا قانونی جلوه می‌دهند. به منظور شناسایی این فرآیند، دولت‌ها همواره گزارش‌های متعددی از بانک‌های خود طلب

می‌نمایند و بانک‌ها موظف هستند موارد مشکوک به پولشویی را در قالب چارچوب مورد درخواست بانک مرکزی تهیه و به‌طور مستمر ارائه نمایند. از سوی دیگر، پیشرفت بانکداری الکترونیکی به تدریج شناسایی روش‌ها و رفتارهای مشکوک به پول‌شویی را پیچیده‌تر نموده است. به‌طوری که شناسایی این رفتارها با توجه به حجم انبوه اطلاعات در یک بانک بدون استفاده از سیستم‌های ضد پولشویی میسر نمی‌باشد.

سیستم ضد پولشویی با جمع‌آوری اطلاعات پراکنده در سطح بانک عملیات پردازش و شناسایی افراد و رفتارهای مشکوک به پول‌شویی را در زمان مناسب انجام داده و گزارش‌های مورد نیاز را تولید و ارائه می‌نماید.

به منظور شناسایی موارد مشکوک به پولشویی لازم است بانک‌ها و مؤسسات اعتباری دارای یک سیستم نظارت بر گردش عملیات بانکی شعب باشند. در این راستا و از آنجا که شناسایی و کنترل گردش عملیات شعب، فرایندی مستمر است؛ لازم است مؤسسات مذکور با نصب نرم افزارهای ویژه مبارزه با پولشویی، اطلاعات اخذ شده از شعب را مدیریت کنند. از جمله ویژگی‌های نرم افزارهای فوق، قابلیت جمع‌آوری، تجزیه و تحلیل، نمایش و عرضه اطلاعات و شناسایی موارد مشکوک است. نرم افزارهای مذکور به گونه‌ای طراحی شده باشند که امکان بررسی و کنترل حساب‌ها و گردش عملیات بانکی شعب به صورت مستمر و در کوتاه ترین زمان را فراهم سازند. از جمله مواردی که در بررسی موارد مشکوک به بررسی دقیق‌تر و کسب اطلاعات بیشتری از شعبه نیاز دارد، می‌توان به موارد زیر اشاره کرد:

- ✓ معاملات کلان پیچیده و غیرمتعارف فاقد توجیه اقتصادی و قانونی
- ✓ وجوه انتقالی فاقد اطلاعات متقاضی (نام، نشانی و شماره حساب) در سند انتقال وجه (حواله)
- ✓ عملیات بانکی و معاملات انجام شده در مناطق فاقد قوانین استاندارد برای مبارزه با پولشویی
- ✓ نگهداری غیرمنطقی حساب از طریق اتباع خارجی
- ✓ معاملات انجام شده از طریق روابط کارگزاری
- ✓ معاملات غیرحضوری از طریق بانکداری الکترونیکی و اینترنتی

۳- متدلوژی تحقیق

دستیابی به هدف‌های تحقیق زمانی میسر می‌شود که جستجوی شناخت یا روش‌شناسی به شکلی درست انجام پذیرد. از این رو اتخاذ روش علمی تنها راه دستیابی به دستاوردهای قابل قبول و علمی خواهد بود. به عبارت دیگر، روش تحقیق مجموعه‌ای از قواعد، ابزارها و راه‌های معتبر و نظام یافته برای بررسی واقعیت‌ها، کشف مجهولات و دستیابی به راه حل مشکلات است.

به منظور گردآوری داده‌ها و تجزیه و تحلیل آنها از مطالعات کتابخانه‌ای و میدانی استفاده خواهد شد. تحقیق حاضر به لحاظ هدف از نوع کاربردی و از نظر روش توصیفی و از شاخه میدانی است.

برای گردآوری اطلاعات در زمینه مبانی نظری و ادبیات موضوع، از منابع کتابخانه‌ای، مقالات، کتاب‌ها و اینترنت استفاده شده است. به منظور جمع‌آوری داده‌ها و اطلاعات برای تجزیه و تحلیل، از داده‌های بانک اطلاعاتی سازمان استفاده خواهد شد.

تحقیق حاضر از لحاظ نوع هدف جزء پژوهش‌های کاربردی است.زیرا اولا این پژوهش نظریه‌های آزموده شده در جوامع دیگر را به منظور شناسایی و حل مسئله پیش‌رو مورد بررسی و تجزیه و تحلیل قرارداده است، و ثانیاً تاکید آن بر شناسایی موثرترین اقدامات بوده و کمتر به چرایی و چیستی علل توجه کرده است. این تحقیق از نظر نوع روش توصیفی (غیر آزمایشی) است. یعنی هدف آن

توصیف شرایط یا پدیده مورد بررسی است. نباید از یاد برد که تحقیق توصیفی می‌تواند صرفاً براساس شناخت بیشتر شرایط موجود با یاری دادن به فرآیند تصمیم‌گیری انجام پذیرد. این تحقیق از نوع تحقیق توصیفی- پیمایشی بوده و به مطالعه ویژگی‌ها و صفات عناصر/اشیاء پرداخته و وضعیت فعلی جامعه را در قالب چند صفت مورد بررسی قرار می‌دهد. هدف نهایی از انجام پژوهش حاضر مدل ارائه مدل ضد پولشویی است که با جمع آوری اطلاعات پراکنده در سطح بانک عملیات پردازش و شناسایی افراد و رفتارهای مشکوک به پولشویی را در زمان مناسب انجام داده و گزارش‌های مورد نیاز را تولید و ارائه مینماید. برای این منظورر ابتدا با مرور دقیق پیشینه تحقیق و اسناد و مدارک آرشیوی (اعم از قوانین و مقررات، آئین نامه‌ها و دستورالعمل‌ها) گام‌های ضروری برای اجرای پروژه برداشته شد.

در این مرحله و براساس سوالات کسب وکار (تجاری) ایجاد شده در بخش بالا به دنبال پیشنهاد و تست راه‌های داده کاوانه جهت ارائه راه حل در دو فاز هستیم.

فاز یک: فاز یک خود به سه زیر بخش اصلی تقسیم گردید که شامل موارد زیر هستند:

۱) شناسایی موارد به کلی مشکوک با ایجاد شاخص‌های کلیدی سنجش مانند موارد با واریز نقدی خیلی بالا

۲) پیشنهاد یک شاخص خطی مناسب جهت دسته بندی شعب بانک و تشخیص داده‌های پرت

۳) پیشنهاد یک روش خوشه‌بندی با استفاده از نقشه‌های خود سازمانده جهت کشف داده‌های پرت با استفاده از روش‌های غیرخطی

پس از انجام سه مرحله فوق و با ترکیب نتایج سه روش مرحله می‌توان موارد مشکوک و پرت را براساس اطلاعات یکساله بانک تشخیص داد.

فاز دو: در فاز ۲ با استفاده از خوشه‌بندی و با فرض راهنمایی شبکه توسط داده‌های پرت فاز یک سعی در کشف الگویی از افراد پولشو براساس مشخصات موجود در انبار داده شعب بانک خواهیم داشت. این فاز از تحقیق به کمک شبکه‌های عصبی چند لایه صورت می‌گیرد. همچنین در فاز دو پروژه به دنبال آن بودیم که آیا شک بیشتر به شعب و بررسی سوابق و مشخصات دقیق‌تر آن‌ها می‌تواند با توجه به مشخصات عمومی آن‌ها از الگویی تبعیت کند یا خیر.

۴- مفروضات تحقیق

در یک پروژه پولشویی یکی از مهمترین موارد رسیدن به سوال کسب وکاری است که باید با داده‌کاوی بدان پاسخ داد با توجه به مطالعات و مشاوراتی که با کارشناسان بانکی انجام گرفت سناریوها و پرسش‌های مختلفی مطرح شد که و هر یک از آن‌ها از جهات مختلفی مانند قابلیت اجرا، دسترسی به داده‌ها، مفید بودن روش، قابل فهم و اجرایی بودن در سیستم بانکی و ... مورد بررسی قرار گرفتند. تا به یک سناریو و سوال مناسب که کلیه شرایط بالا را داشته باشند دست یافتیم.

یکی از محدودیت‌های این تحقیق عدم ارائهٔ آمار و گزارشات دقیق در زمینهٔ پولشویی می‌باشد. بانک‌ها و بانک مرکزی به دلیل برخی مسائل از بیان چنین گزارشاتی خودداری می‌نمایند و با توجه به نیاز این پروژه برای شناسایی سناریوهای قبلی که به عنوان ورودی برای الگوریتم‌های نظارت‌شده لازم می‌باشد، ناچار به استفاده از تجربیات و سوابق مدیران با تجربه در زمینهٔ بازرسی بانک‌ها شده و نتایج این همکاری به گونه‌ای بود که تا حد زیادی این محدودیت مرتفع گردید.

با توجه به مدل‌های بین المللی که در حال حاضر در زمینهٔ پولشویی استفاده می‌شود لزوم برقراری ارتباط بین اطلاعات بانک‌های کشور حائز اهمیت

می‌باشد، ولی با توجه با در اختیار نبودن چنین گزارشاتی صرفاً به بررسی داده‌های یک بانک اکتفا شده است. به‌طور کلی مفروضات تحقیق حاضر به صورت زیر می‌باشد.

- ✓ مابین جرائم بانکی و تحلیل داده های مربوط به رفتار شعب یک بانک وابستگی معنادار وجود دارد.
- ✓ تمامی داده‌های تحقیق مشخص و قطعی فرض شده‌اند.
- ✓ توزیع آماری داده‌های مربوط به تحلیل رفتار شعب ناپارامتریک است.

۵- پیاده‌سازی مدل ضدپولشویی مبتنی بر داده‌کاوی

برای پیاده سازی مدل نهایی بر روی ۱۶۷۰ شعبه یک بانک در سراسر کشور متمرکز شده‌ایم. در این مرحله اطلاعات به ورودی مورد نیاز سیستم تبدیل گشتند جهت ورودی فاز یک این اطلاعات استخراج گشتند:

- امتیاز پرسنل
- سود پرداختی
- مطالبات معوق
- تسهیلات
- جمع چهار سپرده
- سور دریافتی
- کارمزد دریافتی
- سایر منابع
- کد شعبه

با توجه به اهمیت شاخص‌ها در سیستم‌های مبتنی بر داده‌کاوی پس از تعریف و پیاده سازی یک شاخص با هدف شناسایی شعب مشکوک طبق مراحل زیر عمل نمودیم. به این منظور شاخص پولشویی طراحی و اجرا گردید. این شاخص در این تحقیق به صورت قدر مطلق اختلاف مابین سود دریافتی و سود پرداختی تعریف شد. اکنون می‌توان یک حد آستانه بالا برای شناسایی شعب مشکوک بر اساس شاخص فوق تعریف نمود که در این تحقیق این حد آستانه به صورت رابطه ۱ محاسبه می‌شود:

$$Cr = mean\left(MLI\right) + \left(1.64 \times \sqrt{\mathrm{var}(MLI)}\right) \quad (۱)$$

پس شعب مشکوک، از نظر ما آن شعبی هستند که شاخص پولشویی برای آن‌ها بیش از این آستانه باشد. که در این تحقیق این مقدار برابر با ۵۹۶۷۷۱۰۵۰۴ می‌باشد.

در این بخش بنا داریم تا ابتدا با استفاده از شاخص دیویس بولدین تعداد مناسب خوشه‌ها را جهت انجام فرایند خوشه‌بندی داده‌ها تعیین نماییم. سپس براساس تعداد مناسب خوشه‌ها با استفاده از نقشه‌های خود سازمانده اقدام به خوشه بندی داده‌های در دست نماییم.

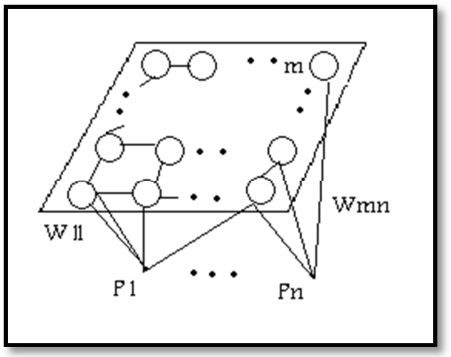
پس از طی مراحل فوق خوشه‌ای که دارای بیشترین میانگین شاخص پولشویی بر اساس شاخص تعریف شده باشد خوشه دارای شعب مشکوک لقب می‌گیرد. بر این اساس تمام شعب موجود در این خوشه تحت عنوان شعب مشکوک برچسب گذاری می‌شوند. نکته حائز اهمیت این است که ممکن است برخی شعب موجود در این خوشه دارای شاخص پولشویی بیش از حد آستانه معرفی شده در رابطه۱ نباشند ولی به دلیل تشابه رفتاری این شعب با شعب مشکوک که دارای شاخص پولشویی بالاتر از حد آستانه هستند آنها نیز تحت عنوان شعب مشکوک برچسب گذاری می‌شوند.

همچنین در سایر خوشه‌ها هر یک از شعبه‌ای که دارای شاخص پولشویی بیشتر از حد آستانه باشند برچسب شعبه مشکوک را می‌گیرند. با سپری شدن مراحل فوق دو گروه از شعب دارای برچسب شعبه مشکوک می‌شوند. گروه اول شعبی که دارای شاخص پولشویی بیش از حد آستانه هستند و گروه دوم شعبی که به دلیل تشابه رفتاری با شعب مشکوک در خوشه شعب مشکوک قرار گرفته‌اند.

هر پروژه داده کاوی شرایط منحصر بفرد خود را دارد و ممکن است نتایج داده کاوی که دو پژوهشگر بر روی یک مجموعه داده مشابه انجام می‌دهند با هم یکسان نباشند. اما اگر در فرآیند داده کاوی از استانداردها پیروی کرده باشند، باید نتایج نزدیک به هم باشد. گاهی نیز نتایج به ابزار پردازشی مورد استفاده وابستگی دارد. برای نمونه برای مقایسه زمان ساختن مدل‌ها، هرچه ابزار پردازشی از توانمندی بیشتری برخوردار باشد، زمان ساخت مدل کوتاهتر می‌شود و یا هرچه ابزار پردازشی توانایی بیشتری داشته باشد، می‌توان از حجم بیشتری از داده‌ها برای مرحله یادگیری بهره برد. از سویی دیگر بخش مهمی از فرآیند داده کاوی مربوط به پیش پردازش داده‌هاست و کارهایی که در این مرحله بر روی داده‌ها انجام می‌گیرد، تا حد زیادی بر روی کیفیت نتایج حاصل موثر است. بنابراین برای اطمینان هر چه بیشتر از نتایج، تکرار چندباره مراحل فرآیند داده کاوی لازم است. این کار موجب بالا رفتن پایداری نتایج حاصل می‌شود. در این پژوهش نیز چندین و چند مرتبه مدل‌ها ساخته و ارزیابی شده و بر اساس نتایج حاصل، برای رسیدن به نتایج مطلوب تر، پارامترهای الگوریتم‌ها بهینه سازی شده‌اند. در تنظیم پارامترهای هر الگوریتم، تلاش شده تاثیر تک تک پارامترهای یک الگوریتم ارزیابی شود. گرچه به دلیل زیاد بودن شمار پارامترهای برخی الگوریتم‌ها، آزمایش تمام حالت‌ها امکان‌پذیر نیست. به هر حال اعتقاد بر این است در روند کسب نتایج، دقت لازم به خرج داده شده است و نتایج ارائه شده به اندازه لازم از اطمینان بالایی برخوردارند. در ادامه این مقاله نتایج حاصل از داده کاوی بوسیله الگوریتم‌های طبقه‌بندی بر روی مجموعه داده موجود، ارائه و ارزیابی شده‌اند.

۵-۱- خوشه‌بندی شعب براساس نقشه‌های خودسازمانده

معمولاً دانش پژوهان این شبکه را یکی از سخت‌ترین شبکه‌های تک لایه می‌دانند.کوهنن شبکه‌ای طراحی کرد که تنها پارامتر معلوم آن نرون‌های ورودی است در حالی که وزن‌ها و نرون‌های خروجی به عنوان پارامترهای مجهولی هستند که باید پیدا شوند. مهمترین خصوصیت این شبکه خودسازمانده بودن آن است. روش کار این شبکه به این صورت است که برای تعداد نرون‌های خروجی عددی را انتخاب می‌کند و از یک منطق ساده فاصله هندسی الگو را به دست می‌آورد. نرون‌های ورودی و خروجی با مقادیر باینری مقداردهی می‌شوند. اساس کار شبکه بر مبنای کم کردن فاصله خود از الگوهای ورودی است. مقدار وزن‌ها با تکرار بدست می‌آید و شبکه به صورت غیرخطی عمل می‌کند. مدل کوهنن یک مدل بدون ناظر است. در این مدل تعدادی سلول عصبی که معمولاً در یک توپولوژی مسطح کنار یکدیگر چیده می‌شوند، با رفتار متقابل روی یکدیگر وظیفه شبکه خود سازمانده را ایفا می‌کنند. برداری را که هر یک از درایه‌های آن دارای چگالی احتمال است، در نظر بگیرید. در این فضای چگالی نمونه‌هایی را به تناوب و تصادف انتخاب کرده به شبکه اعمال می‌کنیم بر اساس موقعیت بردار ورودی در فضا، وزن‌های سلول‌ها طبق الگوریتمی تغییر می‌کنند. این تغییر به نحوی انجام می‌گیرد که در نهایت، بردارهای وزن مربوط به سلول‌ها به‌طور یکنواخت در فضای چگالی احتمال ورودی توزیع می‌شوند و بدین ترتیب شبکه با پراکندن سلول‌های خود در فضای ورودی، چگالی احتمال آن را تخمین می‌زند. پراکندگی سلول‌ها در فضای احتمال ورودی به نوعی می‌تواند فشرده‌سازی اطلاعات محسوب شود، زیرا اکنون هر سلول مطابق شکل ۱ مبین تقریبی از یک محدوده مشخص در فضای R^2 است.

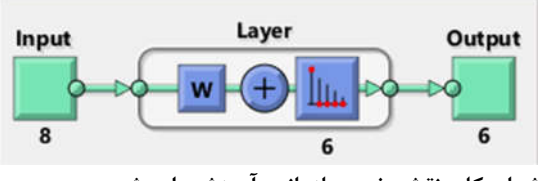


شکل ۱- شبکه عصبی خود سازمانده مدل کوهنن

از آنجائیکه این تحقیق به دنبال خوشه‌بندی شعب می‌باشد تا ریسک پولشویی هر شعبه را براساس خوشه‌ای که به آن تعلق دارد بسنجد باید در هریک از خوشه‌های بدست آمده نماینده‌ای از آن خوشه وجود داشته باشد تا معرف آن خوشه باشد. از این رو روش‌های خوشه‌بندی براساس نقشه‌های خود سازمانده نیز از آن دسته است که می‌تواند مفید باشد. شمای کلی نقشه خود سازمانده آموزش داده شده بر روی مجموعه داده در این تحقیق در شکل ۲ نشان داده شده است.

جدول۱ اطلاعات مربوط به ۶ خوشه بدست آمده از فرایند خوشه بندی داده‌ها با نقشه‌های خودسازمانده را نشان می‌دهد. همانطور که در این جدول مشخص است خوشه ۵ با بیشترین شاخص پولشویی خوشه شعب مشکوک قلمداد می‌شود. شعب سایر خوشه‌ها که دارای شاخص پول‌شویی بیشتر از حد آستانه باشند مانند شعب خوشه ۵ تحت عنوان شعب مشکوک برچسب گذاری می‌شوند. براین اساس به طور کلی ۲۳۰ شعبه از میان ۱۶۷۰ شعبه برچسب مشکوک می‌گیرند.

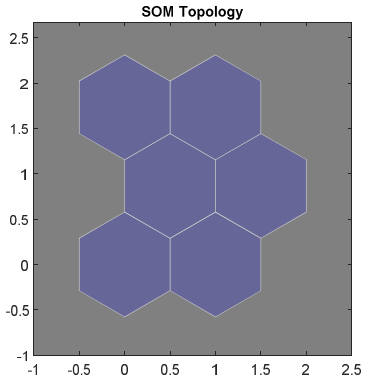
در شکل ۳ توپولوژی نقشه خود سازمانده آموزش داده شده، در شکل ۴ نحوه‌ی اتصال نرون‌ها، در شکل ۵ شدت فاصله مابین نرون‌ها و در شکل ۶ میزان حساسیت هر یک از نرون‌ها به هر یک از ابعاد مجموعه داده، نشان داده شده است. با توجه به قابلیت الگوریتم خوشه‌بندی مورد استفاده در این تحقیق در شناسایی الگوهای پیچیده نسبت به الگوهای خطی که با کمک شاخص‌ها استخراج می‌گردد می‌توان گفت که تلفیق این دو روش اجماع خوبی روی داده‌ها ایجاد خواهد نمود. با توجه به اینکه داده واقعی در مورد موارد پولشویی در دسترس نیست از این خروجی‌ها به عنوان برچسب برای فازبعدی که الگوریتم نظارت‌شده است استفاده می‌کنیم. در صورتی که بر روی این خروجی یک مرحله از خبرگی و بررسی بازرسان بانک انجام شود مسلماً خروجی بسیار بهتری استنتاج خواهد شد.



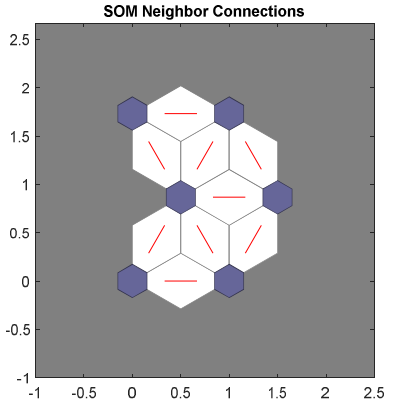
شکل ۲- شمای کلی نقشه خود سازمانده آموزش داده شده بر روی مجموعه داده

جدول ۱- اطلاعات مربوط به ۶ خوشه بدست آمده از فرایند خوشه‌بندی داده ها با نقشه‌های خودسازمانده

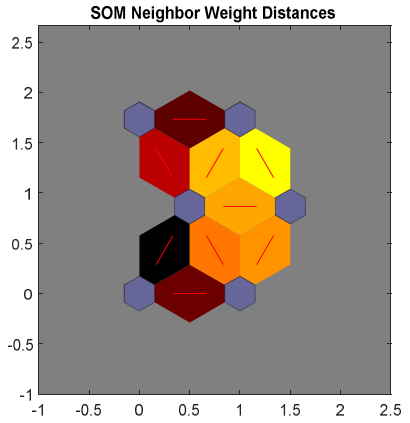
شماره خوشه	۱	۲	۳	۴	۵	۶
شاخص پولشویی	۶۱۰۸۵۶۷۰۷۸	۴۹۹۲۸۸۵۶۵۸	۱۲۸۴۰۹۸۷۰۴۳	۵۳۵۱۷۷۸۷۰	۵۴۱۶۴۵۵۱۴۸۳	۱۹۶۵۲۳۱۶۷۸
میانگین نرمال شده امتیاز پرسنل	۰.۱۹۸۸	۰.۰۷۸۷	۰.۳۲۸۰	۰.۰۶۳۱	۱.۰۰۰۰	۰.۱۵۸۷
میانگین نرمال شده سود پرداختی	۰.۰۳۱۲	۰.۰۱۸۳	۰.۲۴۴۸	۰.۰۱۱۱	۱.۰۰۰۰	۰.۰۳۹۱
میانگین نرمال شده مطالبات معوق	۰.۰۱۰۴	۰.۰۰۷۴	۰.۲۳۰۶	۰.۰۰۳۰	۱.۰۰۰۰	۰.۰۱۵۶
میانگین نرمال شده تسهیلات	۱.۰۰۰۰	۰.۱۷۰۲	۰.۰۲۷۸	۰.۰۰۹۱	۰.۰۲۳۶	۰.۰۱۶۸
میانگین نرمال شده جمع چهار سپرده	۱.۰۰۰۰	۰.۱۱۴۰	۰.۰۴۴۷	۰.۰۰۸۳	۰.۰۳۱۱	۰.۰۱۷۵
میانگین نرمال شده سور دریافتی	۰.۲۴۱۸	۰.۱۶۰۹	۰.۰۲۵۸	۰.۰۰۵۵	۰.۰۰۱۲	۰.۰۱۱۵
میانگین نرمال شده کارمزد دریافتی	۱.۰۰۰۰	۰.۲۳۹۳	۰.۰۲۷۲	۰.۰۰۴۰	۰.۰۵۷۲	۰.۰۱۱۰
میانگین نرمال شده سایر منابع	۰.۹۵۱۸	۰.۳۵۶۳	۰.۰۳۷۵	۰.۰۰۴۴	۰.۰۵۸۴	۰.۰۱۳۵



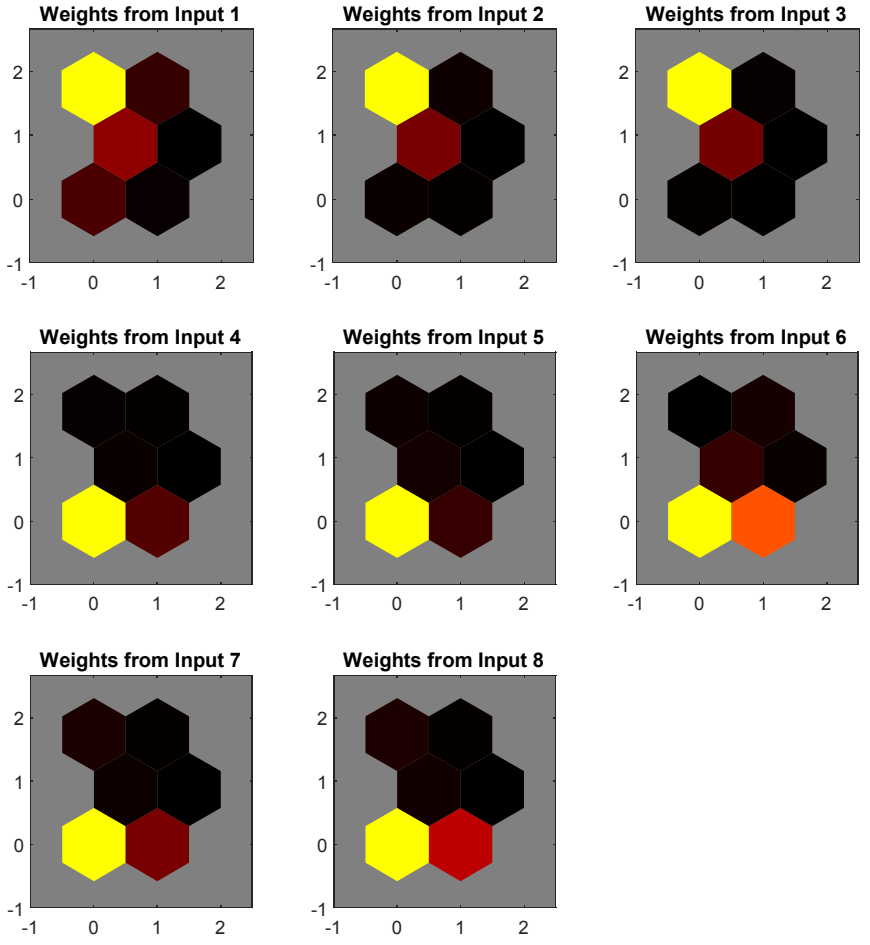
شکل ۳- توپولوژی نقشه خود سازمانده آموزش داده شده



شکل ۴- نحوه‌ی اتصال همسایگی‌ها در نقشه خود سازمانده آموزش داده شد



شکل ۵- شدت فاصله همسایگی‌ها در نقشه خود سازمانده آموزش داده شده



شکل ۶- میزان حساسیت هر یک از نرون‌ها به هر یک از ابعاد داده در نقشه خود سازمانده آموزش داده شده

۵-۲- طبقه‌بندی داده‌های شعب با استفاده از شبکه عصبی چند لایه

شبکه‌های عصبی مصنوعی الگویی برای پردازش اطلاعات می‌باشند که با تقلید از شبکه‌های عصبی بیولوژیکی مثل مغز انسان ساخته شده‌اند. عنصر کلیدی این الگو ساختار جدید سیستم پردازش اطلاعات آن می‌باشد و از تعداد زیادی عناصر (نرون) با ارتباطات قوی داخلی که هماهنگ با هم برای حل مسائل مخصوص کار می‌کنند تشکیل شده‌اند. شبکه‌های عصبی مصنوعی با پردازش روی داده‌های تجربی، دانش یا قانون نهفته در ورای داده‌ها را به ساختار شبکه منتقل می‌کند که به این عمل یادگیری می‌گویند. اصولاً توانایی یادگیری مهمترین ویژگی یک سیستم هوشمند است. سیستمی که بتواند یاد بگیرد منعطف‌تر است وساده‌تر برنامه‌ریزی می‌شود، بنابراین بهتر می‌تواند در مورد مسایل و معادلات جدید پاسخگو باشد. انسان‌ها از زمان‌های بسیار دور سعی بر آن داشتند که بیوفیزیولوژی مغز را دریابند چون همواره مسئله هوشمندی انسان و قابلیت یادگیری، تعمیم، خلاقیت، انعطاف‌پذیری و پردازش موازی در مغز برای بشر جالب بوده و بکارگیری این قابلیت‌ها در ماشین‌ها بسیار مطلوب می‌نمود. روش‌های الگوریتم یک برای پیاده‌سازی این خصایص در ماشین‌ها مناسب نمی‌باشند در نتیجه می‌بایست روش‌ها مبتنی بر همان مدل های بیولوژیکی باشد. درست مثل انسان‌ها با استفاده از مثال‌ها آموزش می‌بینند. به عبارت دیگر شبکه عصبی مصنوعی یک سامانه پردازشی داده‌ها است که از مغز انسان ایده گرفته و پردازش داده‌ها را به عهده پردازنده‌های کوچک و بسیار زیادی سپرده که به صورت شبکه‌ای به هم پیوسته و موازی با یکدیگر رفتار می‌کنند تا یک مسئله را حل نمایند. در این شبکه‌ها به کمک دانش برنامه‌نویسی، ساختار داده‌ای طراحی می‌شود که می‌تواند همانند نورون عمل کند. که به این ساختارداده گره گفته می‌شود. بعد با ایجاد شبکه‌ای بین این گره‌ها و اعمال یک الگوریتم آموزشی به آن، شبکه را آموزش می‌دهند. در این حافظه یا شبکه عصبی گره‌ها دارای دو حالت فعال (روشن یا ۱) و غیرفعال (خاموش یا ۰)اند و هر یال (سیناپس یا ارتباط بین گره‌ها) دارای یک وزن می‌باشد. یال‌های با وزن مثبت، موجب تحریک یا فعال کردن گره غیر فعال بعدی می‌شوند و یال‌های با وزن منفی، گره متصل بعدی را غیر فعال یا مهار (در صورتی که فعال بوده باشد) می‌کنند.

در این مرحله به وسیله خروجی بدست آمده از فاز قبل (نقشه خودسازمانده) به عنوان ورودی، شبکه عصبی چند لایه را آموزش داده و در پایگاه دانش ذخیره می‌نمائیم. هدف از اجرای این روش در مدل پیشنهادی، پیش‌بینی کلاس یاطبقه شعب بر مبنای فرایند خوشه‌بندی است. روش‌های طبقه‌بندی داده‌ها، از رایج‌ترین تکنیک‌های داده‌کاوی برای ساخت مدل‌های پیش‌بینی خوشه شعب براساس ویژگی‌های موجود می‌باشد.

در این مرحله به منظور طبقه‌بندی شعب داده‌های موجود به دو مجموعه آموزش (۸۰ درصد داده‌ها) و آزمایش با (۲۰ درصد داده‌ها) تقسیم شدند. از مجموعه داده‌های آموزش جهت یادگیری و مدل‌سازی و از مجموعه آزمایش برای ارزیابی مدل و دقت پیش‌بینی استفاده گردید. در این قسمت شبکه عصبی چند لایه که زیر مجموعه‌ای از روش‌های هوش‌مصنوعی به شمار می‌آیند بکار گرفته شده است.

تعداد گره ورودی و همچنین خروجی در شبکه عصبی به وسیله نگاشتی که به شبکه ارائه می‌شود قابل تعیین است در این تحقیق به دلیل وجود ۸ متغیر مستقل و یک متغیر وابسته، تعداد گره‌های لایه ورودی ۸ و تعداد گره‌های لایه خروجی یک می‌باشد. اما تنها راه تعیین تعداد لایه‌های پنهان و تعداد گره در هر لایه و همچنین نوع تابع تبدیل، سعی و خطا می‌باشد، به گونه‌ای که اگر تعداد

لایه‌های پنهان و تعداد نرون‌های هر لایه کافی نباشد، شبکه نمی‌تواند به‌طور مناسب به یک جواب بهینه، همگرا شود و اگر تعداد آن‌ها بیش از حد لازم باشد، شبکه دچار بی‌ثباتی می‌شود.

برای تعیین بهترین شبکه، طراحی‌های بسیار زیادی مورد آزمون واقع شد، ولی شبکه مطابق شکل زیر با طراحی ۴ لایه (سه لایه پنهان) و تعداد ۱۵ گره در لایه پنهان اول، ۱۰ گره در لایه پنهان دوم و ۵ گره در لایه پنهان سوم بهترین جواب را به ما ارائه داد، بدین معنی که شبکه با طرح 15-10-5-1 مطابق شکل ۷ مورد استفاده قرار گرفت. همچنین انواع توابع تبدیل در لایه‌های مختلف مورد آزمون واقع شد و بهترین توابع تبدیل، تابع تانژانت هیپربولیک برای لایه‌های مختلف بدست آمد.

خطای فاز آموزش با شبکه عصبی چند لایه ۰.۰۲۱ می‌باشد. جدول ۲، دقت و خطای پیش‌بینی شعبه مشکوک به پولشویی را توسط شبکه عصبی در داده‌های فاز آزمایش را نشان می‌دهد.

پیش از برشمردن معیارهای ارزیابی باید مفهوم ماتریس در هم‌ریختگی روشن شود. این ماتریس چگونگی عملکرد الگوریتم رده‌بندی را با توجه به مجموعه داده ورودی به تفکیک انواع رده‌های مسئله نشان می‌دهد.

مفاهیم ماتریس درهمریختگی به شرح زیر تعریف می‌شوند:

تعداد منفی‌های صحیح[6] (TN): تعداد رکوردهایی که رده واقعی آن‌ها منفی بوده و الگوریتم رده‌بندی نیز آن‌ها را بدرستی منفی تشخیص داده است.

تعداد مثبت‌های نا صحیح[7] (FP): تعداد رکوردهایی که رده واقعی آن‌ها منفی بوده ولی الگوریتم رده بندی آن‌ها را به اشتباه مثبت تشخیص داده است.

تعداد منفی‌های ناصحیح[8] (FN): تعداد رکوردهایی که رده واقعی آن‌ها مثبت بوده ولی الگوریتم رده بندی آن‌ها را به اشتباه منفی تشخیص داده است.

تعداد مثبت‌های صحیح[9] (TP): تعداد رکوردهایی که رده واقعی آن‌ها مثبت بوده ولی الگوریتم رده بندی آن‌ها را به درستی مثبت تشخیص داده است.

شکل ۷- شمای کلی شبکه عصبی چند لایه آموزش داده شده بر روی مجموعه داده

جدول ۲- دقت و خطای پیش‌بینی شعبه مشکوک توسط شبکه عصبی چند لایه در داده‌های فاز آزمایش

پیش‌بینی / واقعی	۰	۱
۰	۲۹۴	۳
۱	۳	۳۴

[6] True Negative
[7] False Positive
[8] False Negative
[9] True Positive

مهم‌ترین معیار برای تعیین کارایی تکنیک دسته‌بندی معیار دقت[10] است. این معیار دقت کل یک دسته‌بندی را محاسبه می‌نماید این معیار نشان دهنده این حقیقت است که دسته‌بندی طراحی شده چند درصد از کل مجموعه رکوردهای آزمایشی را به درستی دسته‌بندی کرده است. دقت دسته بندی با استفاده از رابطه ۲ بدست می‌آید که بیان می‌کند دومقدار TPوTN مهمترین مقادیری هستند که در یک مسئله دو دسته‌ای باید بیشینه شوند.

مشکل اصلی هم نامتعادل بودن داده‌ها و تفاوت معنی‌دار تعداد نمونه‌های هر دسته است که باعث می‌شود یک مدل متمایل به دسته پرتعداد، دقت کلی را بالا نشان دهد. بنابراین نیاز به معیاری دقیق‌تر برای سنجش دقت و کارآیی الگوریتم‌های پیشنهادی دسته‌بندی هستیم. که در رابطه ۳ نمایش آمده است.

گاهی بازخوانی ما به خاطر ضعیف بودن مدل پیشنهادی، بالاست. این ضعیف بودن را با معیار دیگری باید اندازه بگیریم. برای حل این مشکل، در کنار معیار بازخوانی معیار دیگری را به نام صحت، برابر تعداد نمونه‌های تشخیصی درست مثبت به کل نمونه‌های مثبت اعلام شده به صورت رابطه ۴ تعریف می‌کنیم تا میزان مثبت‌های اشتباه را هم در نظر گرفته باشیم.

$$\text{Accuracy}=\frac{tp+tn}{tp+tn+fp+fn} \quad (۲)$$

$$\text{Recall}=\frac{tp}{tp+fn} \quad (۳)$$

$$\text{Presition}=\frac{tp}{tp+fp} \quad (۴)$$

همانطور که گفته شد مهم‌ترین معیار برای تعیین کارایی تکنیک دسته‌بندی معیار Accuracy است. این معیار دقت کل یک دسته‌بندی را محاسبه می‌نماید این معیار نشان‌دهنده این حقیقت است که دسته‌بندی طراحی شده چند درصد از کل مجموعه رکوردهای آزمایشی را به درستی دسته‌بندی کرده است.

نتایج جدول ۳ نشان می‌دهد که شبکه عصبی چند لایه طراحی شده برای طبقه‌بندی اعضاء ۹۸.۲۰ درصد داده‌های آزمایشی را به درستی تشخیص داده است. و با توجه به بالا بودن درصد سه معیار دقت، بازخوانی و صحت نتیجه می‌گیریم این روش کارامد می‌باشد.

۶- نتیجه‌گیری

خروجی‌های این مقاله شامل مستندات علمی، روش‌های داده محور برای شناسایی پولشویی، ارائه یک مدل برای پیاده‌سازی سیستم ضد پولشویی در بانک و در نهایت بهره‌گیری از تکنولوژی‌های نرم افزاری و پیاده‌سازی کل جریان پروژه می‌باشد. بررسی داده‌ها و اعلام نظر روی آن‌ها از دو روش کلی تبعیت می‌نماید که در اکثر سیستم‌های ضد پولشویی نیز صادق می‌باشد.

جدول ۳- دقت، یادآوری و صحت مدل های بدست آمده از شبکه عصبی چند لایه

الگوریتم طبقه‌بندی	دقت	بازخوانی	صحت
شبکه عصبی چند لایه	٪۹۸.۲۰	٪۹۵.۴۴	٪۹۵.۴۴

[10] Accuracy
[11] Precision

اول استفاده از شاخص‌ها با استفاده از روش‌های خطی و آماری و دوم استفاده از الگوریتم‌های داده‌کاوی که آن هم به دو صورت نظارت‌شده و بدون ناظر تقسیم شده و در این پروژه از هر دو روش استفاده شده است. نکتهٔ دیگر اینکه استفاده از روش‌های مختلف می‌تواند بر اساس ابزارهای مختلف انجام شود که در نهایت هر ابزار روش مورد نظر را با استفاده از یک الگوریتم خاص اجرا می‌نماید و این تنوع تاثیر متفاوتی در نتایج و نحوهٔ اجرا بر جا می‌گذارد. مهمترین تفاوت این الگوریتم‌ها در سرعت و کارایی آن‌ها می‌باشد. ضمن اینکه با توجه به مقدار و نوع داده‌های ورودی از نظر عددی یا متنی یا دیگر ویژگی‌ها انتخاب الگوریتم می‌تواند تغییر نماید. بر همین اساس در این مقاله به عنوان مثال برای پیاده سازی روش خوشه بندی با هدف شناسایی داده‌های غیرمتعارف از نقشه‌های خود سازمانده استفاده گردید. هدف از اجرای این مدل ارائه گزارشات مفید برای واحد بازرسی بانک به صورت گزارشات عددی و نموداری است که در اختیار ایشان قرار می‌گیرد.

در این تحقیق در هر دو فاز خوشه‌بندی و طبقه‌بندی داده‌های به‌دست آمده از رفتار شعب بانک از رویکردهای قطعی استفاده شد با توجه به توسعه و مزیت‌های موجود در روش‌های فازی پیشنهاد می‌شود از متدولوژی‌های مبتنی بر مدل‌های فازی برای این منظور استفاده شده و نتایج آن‌ها با رویکردهای قطعی مقایسه شود. همچنین در این تحقیق تأثیر متغیرهای خارجی مانند نرخ تورم، نرخ مالیات، سود سپرده‌های بانکی و غیره در پیش‌بینی رفتار شعب دیده نشده است حال آنکه ممکن است با در نظر گرفتن این متغیرهای خارجی به پیش‌بینی‌های دقیق‌تری از رفتار شعب دست پیدا کنیم. پیشنهاد می‌شود در مطالعات آتی تاثیر متغیرهای اقتصادی بر رفتار شعب تجزیه‌وتحلیل شود. مورد بعدی که در نظر گرفتن آن جالب توجه می‌باشد تشخیص شاخص کلیدی عملکرد، نقدینگی شعب در دهک‌های مختلف سپرده‌گذاران و پیش بینی نقدینگی شعب می‌باشد.

فهرست منابع

[1]. SU-NAN WANG, JIAN-GANG YANG, A Money Laundering Risk Evaluation Method Based On Decision Tree, Sixth International Conference on Machine Learning and Cybernetics(2007) 283-286

[2]. J. Kingdon, AI Fights Money Laundering, IEEE Transactionson Intelligent Systems, (2004), 87-89.

[3]. B. Scholkopf, A short tutorial on kernels. Microsoft Research, MSR-TR-200-6t,(2000).

[4]. 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, 50(3), (2010), 559-569

[5]. D.R Wilson and T. R. Martinez, Improved Heterogeneousdistance functions. Journal of Artificial,(1997), 1-34

[6]. Asma S. Larik, Sajjad Haider, Clustering based Anomalous Transaction Reporting, Procedia Computer Science 3 (2011) 606–610

[7]. N-A. Le-Khac, S. Markos and M-T. Kechadi, Towards a newData Mining-based approach for Anti Money laundering inan international investment bank. International Conference onDigital Forensics & Cyber Crime (ICDF2C), Springer VerlagLNIST, 10(2009)

[8]. Dang Khoa Cao, Phuc Do, Applying Data Mining in Money Laundering Detection for the Vietnamese Banking Industry, Springer, (2012), 207–216

[9]. Chen W.S. and Y.K. Du, Using Neural Networks and Data Mining Techniques for The Financial Distress Prediction Model, Expert Systems with Applications, 36, (2009), 4075–4086

[10]. Nhien An Le Khac, M-Tahar Kechadi, Application of Data Mining for Anti-Money Laundering Detection: A Case Study, IEEE International Conference on Data Mining Workshops, (2010)

[11]. Gao. Zengan, Application of Cluster-Based Local Outlier Factor Algorithm in Anti-Money Laundering, IEEE, 978-1-4244-4639-1,(2009)

[12]. Jamshidi, M.B.; Gorjiankhanzad, M.; Lalbakhsh, A.; Roshani, S. A Novel Multiobjective Approach for Detecting Money Laundering with a Neuro-Fuzzy Technique. In Proceedings of the 2019 IEEE 16th International Conference on Networking, Sensing and Control (ICNSC), Ban_, AB, Canada, 9–11 May 2019; pp. 454–458.

[13]. Magomedov, G.S.; Dobrotvorsky, A.S.; Khrestina, M.P.; Pavelyev, S.A.; Yusubaliev, T.R. Application of Artificial Intelligence Technologies for the Monitoring of Transactions in AML-Systems Using the Example of the Developed Classification Algorithm. Int. J. Eng. Technol. 2018, 7, 76–79.

[14]. Dorofeev, D.; Khrestina, M.; Usubaliev, T.; Dobrotvorskiy, A.; Filatov, S. Application of Machine Analysis Algorithms to Automate Implementation of Tasks of Combating Criminal Money Laundering. In Digital Transformation and Global Society, DTGS 2018, Communications in Computer and Information Science; Alexandrov, D.,

Boukhanovsky, A., Chugunov, A., Kabanov, Y., Koltsova, O., Eds.; Springer: Cham, Switzerland, 2018.

[15]. Plaksiy, K.; Nikiforov, A.; Miloslavskaya, N. Applying Big Data Technologies to Detect Cases of Money Laundering and Counter Financing of Terrorism. In Proceedings of the 2018 6th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW), Barcelona, Spain, 6–8 August 2018; pp. 70–77.

[16]. Sobreira Leite, Gleidson, Adriano Bessa Albuquerque, and Plácido Rogerio Pinheiro. "Application of Technological Solutions in the Fight Against Money Laundering—A Systematic Literature Review." Applied Sciences 9.22 (2019): 4800.

[17]. Ryder, Nicholas. "Money laundering: A review of the UKs anti-money laundering policy." (2020).

[18]. Tiwari, Milind, Adrian Gepp, and Kuldeep Kumar. "A review of money laundering literature: the state of research in key areas." Pacific Accounting Review (2020).

[19]. Kamath, D., Pavithra, K., & Pujari, K. (2019). Data mining techniques applied in banking sector-A review. International Journal of Social and Economic Research, 9(3), 358-365.

[20]. Kanchana, M., Chavan, P., & Johari, A. (2020). Detecting Banking Phishing Websites Using Data Mining Classifiers (No. 2855). EasyChair.

[21]. Shahbazi, F. (2020). Using Decision Tree Classification Algorithm to Design and Construct the Credit Rating Model for Banking Customers. IOSR Journal of Business and Management, 21(3, Series 2), 24-28.

[22]. Dinçer, H., Yüksel, S., Canbolat, Z. N., & Pınarbaşı, F. (2020). Data Mining-Based Evaluating the Customer Satisfaction for the Mobile Applications: An Analysis on Turkish Banking Sector by Using IT2 Fuzzy DEMATEL. In Tools and Techniques for Implementing International E-Trading Tactics for Competitive Advantage (pp. 320-339). IGI Global.

[23]. Kaur, M., Bhaddal, P., & Singh, G. (2019). Calculation of client credit risk prediction in banking sector using data mining.

[24]. Hassani, H., Huang, X., & Silva, E. (2018). Digitalisation and big data mining in banking. Big Data and Cognitive Computing, 2(3), 18.