

اعتبارسنجی مشتریان موسسات مالی با استفاده از شبکه عصبی مصنوعی توسعه یافته با کمک الگوریتم ژنتیک (مطالعه موردی: بانک ملی مرکزی ایلام)

ساسان سلطانی^۱

^۱ کارشناسی ارشد مهندسی کامپیوتر، گرایش نرم افزار، دانشگاه باختر ایلام (نویسنده مسئول)

چکیده

ریسک‌پذیری اغلب برای سودآوری و رفتار مالی دیده می‌شود. با این حال، ارزیابی تأیید اعتبار در بانک‌های کاملاً متمایز است مدیریت اعتبار بانک‌ها برای پیشرفت دقت پیش‌بینی شده در مورد قضاوت‌های مربوط به ریسک اعتباری، باید مدل‌های عملیاتی تری را پیش ببرند. در این مقاله یک مدل جدید به منظور رتبه‌بندی مشتریان بانک مبتنی بر ترکیب شبکه عصبی و الگوریتم ژنتیک ارائه شده است. در مدل پیشنهادی، داده‌ها پس از پردازش توسط یک الگوریتم استخراج مبتنی بر تبدیل موجک، ویژگی به ضرایب موجک تبدیل شده و بارزترین ضرایب به عنوان ویژگی‌های استخراج شده مورد استفاده قرار می‌گیرند. با استفاده از ویژگی‌های استخراج شده مدل شبکه عصبی مصنوعی آموزش داده شده و مدل طبقه‌بندی تولید می‌شود. در روش پیشنهادی، به منظور آموزش شبکه عصبی و تعیین بردار وزن بهینه نرون‌های آن، از الگوریتم ژنتیک استفاده می‌شود. به منظور ارزیابی عملکرد روش پیشنهادی از یک پایگاه داده شامل اطلاعات ۱۰۸ مشتری حقیقی و حقوقی بانک ملی شعبه مرکزی شهر ایلام در طی سال ۱۳۹۶ استفاده شده است. مدل پیشنهادی از جنبه‌های مختلف مورد ارزیابی قرار گرفته و دقت آن در رتبه‌بندی مشتریان بانک آزمایش شد. نتایج آزمایشات نشان داد که مدل پیشنهادی می‌تواند رتبه‌بندی مشتریان بانک را با میانگین دقت ۹۳٫۶۴٪ انجام دهد. مقایسه نتایج بدست آمده از آزمایش روش پیشنهادی با الگوریتم‌های مشابه نشان می‌دهد که مدل ارائه شده دارای دقت و کارایی بالاتری در رتبه‌بندی مشتریان بانک‌ها می‌باشد.

واژه‌های کلیدی: ریسک اعتباری، اعتبارسنجی، بهینه‌سازی، شبکه عصبی مصنوعی توسعه یافته، الگوریتم ژنتیک

مقدمه

امروزه بانک‌ها برای شناخت مشتریان، ارضای نیازمندی‌ها و ارائه خدمات مالی مناسب نیازمند شناسایی دقیق ویژگی‌های اعتباری آن‌ها هستند (سیلوا و مورتیرا^۱، ۲۰۰۹). ایجاد تعادل میان عرضه و تقاضا در منابع و تسهیلات بانکی، مدیریت و کاهش مطالبات معوق و رهایی از سیستم وثیقه محوری از جمله مسائلی است که ضرورت و نیاز به پیاده سازی نظام امتیازدهی و اعتبارسنجی را در سیستم بانکی بیش از سایر مسائل نمایان می‌سازد. یکی از خدمات مالی در بانک، اعطای تسهیلات مالی از جمله وام به متقاضیان اعتباری است. در این بین مشتریانی وجود دارند که علیرغم خوش حساسی مانند سایر مشتریان با آن‌ها برخورد می‌شود و ضمانت‌های زیادی برای اعطای تسهیلات از آن‌ها دریافت می‌شود و حتی گاهی مشتریان بد حساب تر تسهیلات درخواستی خود را در مدت زمانی کوتاه تر از مشتریان خوش حساب دریافت می‌کنند. تکنیک‌های داده‌کاوی می‌توانند با ارائه یک الگو یا مدل به کشف دانش پنهان در حجم زیادی از داده‌های تراکنش‌های اعتباری مشتریان بانک‌ها کمک کنند (دهمرد و همکاران، ۱۳۹۶). کاهش و کنترل ریسک به عنوان یکی از عوامل موثر بر بهبود فرایند اعطای اعتبار و در نتیجه بر عملکرد بانک‌ها مطرح است و نقش اساسی در تداوم ارائه تسهیلات بقای بانک‌ها و موسسات مالی دارد. از مهمترین دلایل اهمیت سنجش این ریسک می‌توان به موارد زیر اشاره نمود: الف- اکنون مهمترین عامل ورشکستگی بانک‌ها ریسک اعتباری است. منظور از ریسک اعتباری همان احتمال عدم برگشت اصل وام اعطایی از سوی مشتریان می‌باشد. در این زمینه باید گفت که در برخی از مواد دریافت کنندگان وام توان ارزیابی منافع حاصل از سرمایه گذاری خود را ندارند و درنهایت امر متضرر می‌شوند و به تبع آن نمی‌توانند تسهیلات دریافتی را پرداخت نمایند. اگر مشتری به موقع تعهدات خود را باز پرداخت نکند، این تسهیلات به صورت مطالبات معوق بانکی در می‌آید و این امر موجب اختلال در توزیع اعتبارات بانکی و در نتیجه اختلال در اقتصاد کشور می‌شود؛ ب- اندازه گیری ریسک اعتباری با پیش بینی زیانهای عدم بازپرداخت اعتبارات و ایجاد رابطه منطقی بین ریسک و بازده، امکان بهینه سازی ترکیب پرتفوی اعتباری، قیمت گذاری دارایی‌ها و تعیین سرمایه اقتصادی بانک‌ها را به منظور کاهش هزینه‌های سرمایه ای و حفظ توان رقابتی فراهم و نوعی مزیت نسبی برای بانک‌ها و موسسات اعتباری ایجاد می‌نماید.

بانک‌ها به عنوان سازمان های مالی بسیار ضروری در نظر گرفته می‌شوند که با ارائه خدمات مختلف به مشتریان در حالی که با انواع مختلف ریسک مواجه هستند، به دنبال سود هستند. این مستلزم ارزیابی فیزیکی مشخصات فردی، تحمیل تعصبات متشکل از دانش، بینش شخصی و گزینه مدیر اجرایی اعتبار است. این روش در تعداد محدودی از بانک‌ها با استفاده از مدل‌های امتیازدهی اعتبار جایگزین شده است تا تصمیمات اعتباری مناسبی اتخاذ شود. برعکس، بانک‌ها داده های مربوط به مشتریان خود را در انبارهای داده نگهداری می‌کردند که می‌تواند به عنوان دارایی دانش پنهان باشد که می‌تواند از طریق ابزار داده کاوی خوانده و اعمال شود. علاوه بر این، این امر علاوه بر صرفه جویی در وقت و هزینه تحقیق، قضاوت اعتبار را ارتقا داده و مسئولیت های موسسات مالی را کنترل می‌کند. اعتبار، رفتار اساسی در بخش بانکی را تشکیل می‌دهد زیرا رفتار اعتباری بر موفقیت و استحکام یک بانک تأثیر می‌گذارد. بنابراین، احکام وام برای سازمان های مالی قابل توجه است زیرا خطر اعتباری را دور می‌کند.

اولوکویو^۲ (۲۰۱۱) تأکید می‌کند که وام در هسته اصلی تجارت بانکی است. غالباً، مدیران بانک‌ها در حالی که امکان عدم پرداخت را کاهش می‌دهند، با مشکل خسته کننده شدن برای افزایش حجم اعتبار روبرو هستند. از طرف دیگر، مدل‌های امتیازدهی اعتباری مدیران امور بانکی را قادر می‌سازد حساب‌هایی را که احتمالاً دارای اعتبار هستند و یا احتمالاً پیش فرض (خطاهای اعتباری بد) را براساس ویژگی‌های متقاضی گرفته شده از فرم درخواست، تعیین کنند.

^۱Silva & Murteira^۲Olokoyo

امروزه، آینده تجارت بانکی به شدت درگیر ظرافت های مدیریت ریسک است. بانک ها به دنبال ابزارهای مدیریت ریسک موثرتر و مدل های پشتیبانی تصمیم گیری هستند که با تکنیک های تحلیلی برای دوام در محیط های تجاری نامشخص همراه هستند. اساس مدیریت ریسک ایجاد چارچوبی است که به تصویب وام، سیستم رتبه بندی ریسک اعتباری، سیستم قیمت گذاری تعدیل شده با ریسک و سیستم گزارشگری جامع می پردازد (آرانکومار و کوتشوار^۳، ۲۰۰۶). صنعت اعتبار در دهه گذشته پیشرفت و توسعه جدی داشته است. کارشناسان بانک ها برخی از مدل های امتیازدهی اعتبار را برای تکمیل روش های سنتی به منظور طبقه بندی برنامه های وام به خوب یا بد بر اساس ویژگی های متقاضی مانند سابقه اشتغال، سابقه اعتبار قبلی، سن و غیره ایجاد کرده اند. محققان دائما در جستجوی الگوریتم های جدید برای دقت شبیه سازی امتیازبندی اعتبار را افزایش دهید.

طبق گفته هانگ^۴ و همکاران (۲۰۰۷)، پیشرفت حتی در کسری از درصد صحت اعتبار منجر به پس انداز قابل توجهی خواهد شد. بنسیچ، سارلیجا و زکیچ سوساک (۲۰۰۵) در مطالعه خود نشان دادند که مدل احتمالی شبکه عصبی بهترین نتایج را به دست می آورد. آنها خواستار گسترش تجزیه و تحلیل روش با افزودن الگوریتم های شبکه عصبی بیشتر مانند طبقه بندی کننده های بدون نظارت و همچنین کاوش سایر روش های آماری پیشرفته با تکنیک های هوش مصنوعی مانند الگوریتم های ژنتیک در مدل سازی امتیازدهی اعتبار شدند (لیمسام بونیچ و لی^۵، ۲۰۰۵) مدل تصمیم گیری برای وام را ایجاد کردند و نشان دادند که مدل PNN به طور موثری در گروه بندی و غربالگری وام کشاورزی در تایلند مورد استفاده قرار گرفت. رقوندرا و سیمام^۶ (۲۰۱۰) از الگوریتم های انتخاب ویژگی داده کاوی در استرالیا، آلمانی و ژاپنی، داده های اعتباری استفاده کردند تا بتوانند مجموعه ایده آل ویژگی ها را برای مدل طبقه بندی مشخص کنند. از الگوریتم انتخاب ویژگی و دقت طبقه بندی برای اندازه گیری عملکرد مدل پیش بینی با شبکه عصبی برای طبقه بندی ریسک استفاده شد. نتایج نشان می دهد که دقت طبقه بندی و تعداد ویژگی های انتخاب شده الگوریتم ها با شبکه عصبی در مقایسه با سایر روش ها کارآمدتر بوده است. لاسانا^۷ و همکاران (۲۰۱۰) اشاره کردند که برای دستیابی به حتی کمی پیشرفت در دقت امتیازدهی اعتبار، از روشهای محاسبات نرم برای کمک به روشهای موجود استفاده می شود. بنابراین، شبکه های عصبی ابزاری پاداش دهنده جایگزین برای هوش تجاری هستند که می توانند در مدل های امتیازدهی اعتبار به کار روند زیرا دقت طبقه بندی بهتری را ارائه می دهند.

در این مقاله با بهبود الگوریتم شبکه عصبی مصنوعی به بررسی عوامل موثر در اعتبار سنجی مشتریان و همچنین پیش بینی خوش حساب یا بد حساب بودن مشتریان جدید با همین شبکه بهبود یافته خواهیم پرداخت و تصویری شفاف از وضعیت مشتری در ایفای به موقع تعهدات، ارائه می دهیم و نتایج کار ما منجر به بازداشتن آنان از استفاده بیش از حد منابع و سقوط به ورطه ی بحران مالی می شود و ریسک اعتباری^۸ را در سیستم بانکی کشور، به حد قابل توجهی کاهش دهیم. در این مقاله با در نظر گرفتن خصوصیات متقاضیات وام و با استفاده از داده کاوی نسبت به اعتبار سنجی مشتریان بانک اقدام خواهد شد. اهداف کلی مقاله عبارتند از:

^۳Arun Kumar & Kotreshwar

^۴Huang

^۵Limsombunchai & Lee

^۶Raghavendra & Simha

^۷Lahsana

^۸Credit Risk

- ۱- بهبود شبکه‌های عصبی مصنوعی به کمک الگوریتم ژنتیک به منظور توسعه‌ی موضوعات داده‌کاوی در این پژوهش و پژوهش‌های آتی.
- ۲- مقایسه روش پیشنهادی با سایر روش‌های موجود در زمینه پیش‌بینی و سنجش اعتبار مشتریان در موسسات مالی و ...
- ۳- آنالیز و تحلیل ویژگی‌های موثر در تعیین اعتبار مشتریان و رتبه‌بندی این ویژگی‌ها در پژوهش حاضر.
- ۴- ارائه راهکارهای پیشنهادی در جهت بهبود وضعیت عملکردی بانک‌ها و موسسات مالی و اعتباری.

مروری بر مطالعات پیشین

امروزه اکثر تحقیقات و کاربردها در اعتبارسنجی مبتنی بر دو روش هستند: ۱- روش‌های سنتی (آماري) ۲- روش‌های داده‌کاوی (هوش مصنوعی). می‌توان مدل‌های اعتبار سنجی را به مدل‌های پارامتریک و نا پارامتریک تقسیم کرد. روش‌های پارامتریک مثل پروبیت^۱ و لوجیت^۲، تحلیل تمایزی خطی و رگرسیون خطی که از ابتدای ظهور اعتبارسنجی مورد استفاده قرار گرفتند و سپس روش نا پارامتریک و داده‌کاوی مثل درختان تصمیم‌گیری، شبکه‌های عصبی و ماشی بردار پشتیبان که امروزه بیشتر به کار گرفته می‌شود (صالحی و کرد کتولی، ۱۳۹۶).

البرزی و همکاران (۱۳۹۵) در مقاله‌ای تحت عنوان به کارگیری الگوریتم ژنتیک در بهینه‌سازی درختان تصمیم‌گیری برای اعتبار سنجی مشتریان بانک‌ها، یک مدل مناسب مبتنی بر تکنیک‌های خوشه‌بندی، انتخاب ویژگی‌ها، درختان تصمیم‌گیری و الگوریتم ژنتیک جهت اعطای تسهیلات اعتباری متناسب با هر طبقه ارائه کردند. در ساخت این مدل فرآیند توسعه در شناخت الگو و فرایند CRISP برای اعتبار سنجی مشتریان به کار رفته شده است. نتایج نشان می‌دهد که دقت طبقه‌بندی مدل پیشنهادی به طور تقریبی از تمام مدل‌های درخت تصمیم‌گیری مقایسه شده در این مقاله بالاتر است. همچنین تعداد برگ‌ها و اندازه‌ی درخت تصمیم‌گیری و در نتیجه‌ی پیچیدگی آن از همه کمتر است.

یوباس^۳ و همکاران (۲۰۱۱) در پژوهشی تحت عنوان رویکرد مدیریت ریسک اعتباری و رویکرد هوش تجاری بخش بانکی در اردن، در این پژوهش نشان داده شد که یک مدل شبکه عصبی مصنوعی را به عنوان یک سیستم پشتیبانی تصمیم برای ارزیابی اعتبار در بانک‌های تجاری اردن بر اساس ویژگی‌های متقاضی ایجاد کند. مدل پیشنهادی می‌تواند برای کمک به کارشناس اعتبار در تصمیم‌گیری بهتر هنگام ارزیابی برنامه‌های وام در آینده مورد استفاده قرار گیرد. برای توسعه مدل عصبی مصنوعی، از یک برنامه اعتباری در دنیای واقعی از موارد درخواست‌های رد شده و رد نشده از بانک‌های مختلف اردن استفاده شد. نتایج تجربی نشان داد که شبکه‌های عصبی مصنوعی نویدبخش افزودن به روشهای طبقه‌بندی موجود را دارند.

اورسکی^۴ و همکاران (۲۰۱۲) بررسی کردند که مجموعه داده‌های متعلق به یک بانک تا چه اندازه می‌تواند برای پیش‌بینی توانایی وام‌گیرنده در بازپرداخت به موقع وام مؤثر باشد. آن‌ها یک تکنیک انتخاب ویژگی با الگوریتم ژنتیک و شبکه عصبی برای یافتن زیر مجموعه ویژگی‌های بهینه که دقت طبقه‌بندی را افزایش می‌دهد، ارائه دادند. آزمایش‌ها برای ارزیابی دقت روش آن‌ها بر روی داده‌های یک بانک کرواتان انجام شد. در پژوهش ذکرشده از اورسکی و همکاران، با استفاده از الگوریتم ژنتیک و شبکه عصبی به انتخاب ویژگی‌های بهینه پرداخته و با سایر روش‌های انتخاب ویژگی مقایسه شده است؛ اما کاربردی از روش‌های طبقه‌بندی دیگر برای تعیین تأیید صحت طبقه‌بندی دیده نمی‌شود. در پژوهش خود روشی مبتنی بر الگوریتم رقابت

^۱Probit

^۲Logit

^۳Yobas

^۴Oreski

استعماری و شبکه عصبی ارائه داده ایم که عملکرد بهتری نسبت به مدل اورسکی و همکاران داشته است و برای تصدیق ویژگی های انتخابی از چهار روش طبقه بندی دیگر نیز استفاده کرده ایم.

تقوی و همکاران (۲۰۱۶) در پژوهشی تحت عنوان مقایسه ریسک اعتباری بین شبکه عصبی مصنوعی و مدل های رگرسیون لجستیک در بانک توسعه تعاون گیلان، در این تحقیق، تلاش های زیادی برای بررسی کارایی مدل های رگرسیون لجستیک و شبکه عصبی برای تصمیم گیری اعتباری متقاضیان طبیعی دریافت وام های اقساطی برای فروش انجام شده است. در بانک توسعه تعاون گیلان، به همین دلیل، مشتریانی که از ابتدای سال ۱۳۸۸ (۲۰۰۹) تا پایان سال ۱۳۹۲ (۲۰۱۳) درخواست وام داده بودند و همچنین دارای پرونده اطلاعات کامل بودند، ۳۷۶ مورد بودند و بر اساس متغیرهای مستقل این تحقیق از جمله درآمد متقاضی، سود تسهیلات، دوره بازپرداخت، میزان وام ضامن و نوع اطمینان گرفته شده. نتیجه این بررسی نشان میدهد که مدل های رگرسیون لجستیک و شبکه عصبی برای پیش بینی ریسک اعتباری متقاضیان بسیار کارآمد هستند، اما مقایسه این دو مدل نشان می دهد شبکه عصبی کارایی و دقت بیشتری دارد.

عبدو^۳ (۲۰۱۴) در مقاله خود تحت عنوان برنامه نویسی ژنتیک برای امتیازدهی اعتباری، مطالعه موردی بانک های دولتی مصر به معرفی GP4 به عنوان یکی از مهمترین تکنیک های مورد استفاده در طبقه بندی و امتیازدهی پرداخته است. در این مقاله دو تکنیک مختلف امتیازدهی آماری GP، PA5 برای پیش بینی کیفیت وام گیرنده، مورد بحث و مقایسه قرار گرفته است.

دانگ^۴ و همکاران (۲۰۱۶) یک مدل رگرسیون لاجیت با ضرایب تصادفی را برای ایجاد کارت امتیازدهی اعتباری پیشنهاد کردند. آنها برای ارزیابی عملکرد الگوریتم پیشنهادی، داده های اعتباری آلمان را مورد بررسی قرار دادند. نتایج تجربی نشان داد که مدل پیشنهادی میتواند بدون از بین بردن ویژگی های مطلوب خود، دقت پیش بینی مدل رگرسیون لاجیت با ضرایب ثابت را بهبود بخشد.

تانگ^۵ و همکاران (۲۰۱۷) در پژوهشی تحت عنوان مدل شبکه عصبی احتمالی در تجزیه و تحلیل طبقه بندی اعتبار امروزه، مدل های طبقه بندی اعتباری به طور گسترده ای مورد استفاده قرار می گیرند زیرا می توانند به تصمیم گیرندگان مالی در مدیریت مسائل طبقه بندی اعتبار کمک کنند در این مقاله، یک شبکه عصبی احتمالی (PNN) را پیشنهاد دادند و آن را برای حل مشکل طبقه بندی اعتبار با استفاده از مجموعه داده های اعتباری شناخته شده استرالیا و ژاپن اعمال شده. این مدل از غیرخطی بودن سیناپسی یک درخت دندریتیک در یک مدل عصبی بیولوژیکی الهام گرفته شده است. و توسط الگوریتم انتشار برگشت خطا آموزش داده می شود. علاوه بر این، ما از مدارهای منطقی (LC) برای شبیه سازی ساختار دندریتیک با موفقیت استفاده می کنیم که باعث می شود PNN به طور موثر بر روی سخت افزار پیاده سازی شود. نتایج آماری آزمایشات تأیید کرده که PNN از نظر دقت و کارایی محاسباتی در مقایسه با الگوریتم های کلاسیک عملکرد برتر را به دست می آورد.

باتاچارجی^۶ و همکاران (۲۰۱۷) در پژوهش تحت عنوان یک مدل گروه مبتنی بر شبکه عصبی مصنوعی برای ارزیابی و استقرار ریسک اعتباری به عنوان یک رابط کاربر گرافیکی، به این نتایج دست یافتند که ریسک اعتبار یک تهدید مشترک برای

^۳Abdou

^۴Dang

^۵Tang

^۶Bhattacharjee

صنعت مالی است زیرا مدیریت نادرست ریسک اعتبار منجر به خسارات مالی زیادی در بخشهای بانکی و غیر بانکی می شود. در گذشته از روش های داده کاوی برای ارزیابی ریسک اعتبار استفاده شده است. این مطالعه با استفاده از مجموعه داده های اعتباری آلمان که از مخزن یادگیری ماشین UCI برای تولید یک مدل یادگیری گروهی مبتنی بر شبکه عصبی مصنوعی برای ارزیابی ریسک اعتبار استفاده شده است. یازده الگوریتم داده کاوی در ابزار منبع باز Weka برای انجام رتبه بندی اعتبار در مجموعه داده های اعتبار آلمان با استفاده از روش یادگیری نظارت شده استفاده شده است. نتایج پیش بینی شده از پنج الگوریتم برتر رتبه بندی شده با استفاده از یک بسته 'nnet' در R به یک شبکه عصبی مصنوعی پیشخوان تغذیه می شود. مدل گروه مبتنی بر شبکه عصبی مصنوعی به دقتی ۹۸,۹۸٪ دست یافت و عملکرد بهتری نسبت به الگوریتم های مولفه منفرد داشت. بر اساس این مدل گروه مبتنی بر ANN، یک رابط کاربر گرافیکی تعاملی بیشتر در R توسعه داده شد.

در مقاله حاضر، پیشنهادی ابتدا یک مجموعه داده آموزش به منظور ساخت مدل طبقه بندی مورد استفاده قرار می گیرد. این مجموعه داده شامل مشخصات مالی مشتریان بانک می باشد. این داده ها پس از پردازش توسط یک الگوریتم استخراج مبتنی بر تبدیل موجک، ویژگی به ضرایب موجک تبدیل شده و بارزترین ضرایب به عنوان ویژگی های استخراج شده مورد استفاده قرار می گیرند. با استفاده از ویژگی های استخراج شده مدل شبکه عصبی مصنوعی آموزش داده شده و مدل طبقه بندی تولید می شود. در روش پیشنهادی، به منظور آموزش شبکه عصبی و تعیین بردار وزن بهینه نرون های آن، از الگوریتم ژنتیک استفاده می شود. پس از تولید مدل یادگیر، در فاز آزمایش ابتدا اطلاعات مشتری مورد آزمون جمع آوری می شود. سپس با استفاده از تبدیل موجک ویژگی های مربوط به مشتری استخراج شده و توسط الگوریتم طبقه بندی ارزیابی می شوند. در نهایت، خروجی سیستم تشخیص مبنی بر خوش حساب یا بدحساب بودن مشتری تولید می شود. برای اولین بار در این مقاله از ترکیب شبکه عصبی مصنوعی و الگوریتم ژنتیک برای ارزیابی ریسک اعتباری مشتریان بانکی استفاده شده است.

پیاده سازی

در ایران به شکل تاریخی بانک ها به ترکیبی از تجربیات کارشناسان خود و تحلیل های کمی و کیفی برای تصمیم گیری در مورد پرداخت تسهیلات و میزان اعتبار مشتری تکیه می کنند. در این مطالعه به منظور بررسی عوامل مؤثر بر رفتار اعتباری مشتریان حقوقی بانکی، بجز سرمایه در گردش که اختلاف بدهی های جاری و دارایی های جاری است، در سایر موارد از نسبتهای مالی مناسب استفاده شده است. شایان ذکر است که این نسبتهای بنا بر پیشنهاد اداره کل تحقیقات کنترل ریسک بانک مورد بررسی، تهیه و ارائه شده اند. در ادامه این نسبتهای که همان متغیرهای مستقل مدل نیز به شمار می روند، معرفی و مورد بررسی قرار می گیرد.

با توجه به اینکه هدف این مقاله کلاس بندی مشتریان به دو دسته خوش حساب و بدحساب است، جامعه آماری از نظر وضعیت اعتباری به دو دسته مشتریان خوش حساب (مشتریانی که تسهیلات دریافتی خود را در سررسید مربوطه قبل از اینکه تسهیلات سررسید گذشته شود، تسویه می کند) و مشتریان بدحساب (مشتریانی که پس از دریافت تسهیلات، در سررسید مربوطه نسبت به تسویه تعهدات خود شامل اصل و سود تسهیلات اقدام نمی کند) تقسیم شده است.

پس از مشخص شدن ابعاد داده های ورودی (متغیر مستقل) اکنون زمان گردآوری داده ها فرا می رسد. داده های مربوط به مشتریان مختلف، در طی سال های مختلف جمع آوری و در پایگاه داده ذخیره می شود. برای این منظور داده های مربوط به ۱۰۸ مشتریان حقیقی و حقوقی بانک ملی شعبه مرکزی شهر ایلام (۸۲ نمونه خوش حساب و ۲۶ نمونه بد حساب) مختلف در طی بازه های زمانی ۱۳۹۶ جمع آوری شدند. این داده ها، خام و مربوط به مشتریانی است که برای بانک از اهمیت بالاتری نسبت به سایر مشتریان برخوردارند. کیفیت داده ها در پیش بینی همواره یک موضوع کلیدی است. برای بهبود نتایج، پیش از اینکه داده -

ها برای آموزش به مدل داده شوند، پردازش‌هایی بر روی آنها انجام می‌گیرد. متغیرهای پژوهش به دودسته مستقل و وابسته تقسیم می‌شوند:

۱- **متغیر وابسته:** با توجه به اینکه مشتریان اعتباری به دو دسته خوش حساب و بد حساب تقسیم شده‌اند، این متغیر می‌تواند دو حالت yes یا no را به خود اختصاص دهد. اگر مشتری در بازپرداخت تسهیلات نکول کرده باشد تسهیلات سررسید گذشته یا معوق داشته باشد $t=yes$ در نظر گرفته می‌شود و چنانچه بازپرداخت تسهیلات در سررسید انجام شده باشد، $t=no$ در نظر گرفته می‌شود.

۲- **متغیرهای مستقل:** شامل آن دسته از متغیرهایی می‌شوند که بر ریسک اعتباری مشتری و به عبارتی دیگر بر متغیر وابسته مدل تأثیر گذارند. بنابراین با توجه به بررسی های انجام شده، مجموعه متغیرهایی که به نوعی می‌توانند بر متغیر وابسته مدل تأثیر داشته باشند. تمام ویژگی‌های مستقل از نوع عددی بوده و مقادیر مثبت یا منفی دارند.

معرفی مجموعه داده

منظور از لیست مشخصات مشتری، اطلاعات ذخیره شده برای هر نمونه در پایگاه داده می‌باشد. پایگاه داده مورد استفاده دارای اطلاعات مشتریان بانک ملی شعبه مرکزی شهر ایلام می‌باشد. این پایگاه داده حاوی اطلاعات مالی سال ۱۳۹۶ مربوط به ۱۰۸ مشتری حقیقی و حقوقی می‌باشد. براساس اظهارات کارشناس مالی بانک ارائه کننده داده‌ها، ۸۲ مشتری در دسته خوش حساب و ۲۸ مورد نیز در دسته مشتریان بدحساب قرار می‌گیرند. لیست اطلاعات مشتریان و نحوه محاسبه آنها در جدول ۱ نمایش داده شده است.

جدول ۱- لیست اطلاعات مشتریان در پایگاه داده مورد استفاده

ردیف	عنوان متغیر	نحوه محاسبه	نوع متغیر	میانگین	حداقل	حداکثر
۱	سرمایه در گردش	$D_j - B_j$	عددی (مستقل)	-۲.۴۵۲۶۰+۰.۸	-۹.۰۶۷۵۵۵۶۰	۵۸۹۹۶۸۱۲۵
۲	نسبت بدهی جاری	$\frac{B_j}{D}$	عددی (مستقل)	۰.۱۸۷۲	۸.۰۲۰۰۰-۰.۵	۱.۱۶۰۲
۳	نسبت مالکانه	$\frac{H}{D}$	عددی (مستقل)	۰.۰۰۴۰	۲.۰۸۰۰۰-۰.۵	۰.۰۲۱۵
۴	نسبت دارایی	$\frac{D_j}{D}$	عددی (مستقل)	۰.۳۰۶۴	۰.۰۰۹۶	۰.۸۶۵۵
۵	نسبت سریع (آنی)	$\frac{D_j - M_k}{B_j}$	عددی (مستقل)	-۳۵.۱۶۷۳	-۳.۳۷۸۷۰+۰.۳	۱۴۷.۹۱۳۷
۶	نسبت گردش نقد	$\frac{M_n}{B_j}$	عددی (مستقل)	۴۶.۳۲۴۲	۰.۰۱۷۲	۳.۳۴۳۸۰+۰.۳
۷	موجودی نقد به کل دارایی‌ها	$\frac{M_n}{D}$	عددی (مستقل)	۰.۳۴۷۵	۰.۰۰۷۵	۰.۹۲۹۷

ردیف	عنوان متغیر	نحوه محاسبه	نوع متغیر	میانگین	حداقل	حداکثر
۸	نسبت جاری	$\frac{D_j}{B_j}$	عددی (مستقل)	۴۰.۴۱۲۸	۰.۰۷۵۴	۲.۸۲۵۸۰+۰.۳
۹	نسبت بدهی	$\frac{B}{D}$	عددی (مستقل)	۰.۴۱۱۱	۰.۰۰۱۷	۲.۳۱۹۷
۱۰	حاشیه سود	$\frac{S}{F}$	عددی (مستقل)	۰.۰۸۹۲	۰.۰۸۰۰	۰.۱۰۰۰
۱۱	بازده دارایی	$\frac{S}{D}$	عددی (مستقل)	۰.۰۰۱۵	۱.۱۷۰۰-۰.۵	۰.۰۰۵۹
۱۲	سود انباشته به کل دارایی‌ها	$R_s \times D$	عددی (مستقل)	۰.۰۱۱۳	۵.۸۶۰۰-۰.۵	۰.۰۴۸۳
۱۳	نسبت بدهی به حقوق صاحبان سهام	$\frac{B}{H}$	عددی (مستقل)	۳۵۰.۹۶۵۸	۰.۵۱۹۶	۱.۴۴۰۲+۰.۴
۱۴	کل وام‌های بانکی به کل دارایی‌ها	$\frac{V}{D}$	عددی (مستقل)	۰.۴۵۲۱	۰.۰۱۲۵	۲.۴۴۳۴
۱۵	کل وام‌های بانکی به کل بدهی‌ها	$\frac{V}{B}$	عددی (مستقل)	۱.۴۳۳۳	۱.۰۰۱۵	۲۰.۶۰۹۰
۱۶	فروش خالص به کل دارایی‌ها	$\frac{F}{D}$	عددی (مستقل)	۰.۰۱۶۱	۱.۱۷۲۰-۰.۴	۰.۰۵۹۸
۱۷	موجودی کالا به فروش خالص	$\frac{M_k}{F}$	عددی (مستقل)	۱۰۵.۶۶۶۸	۱۰۰.۱۵۸	۲.۸۱۲۶+۰.۳
۱۸	بدهی‌های جاری به فروش خالص	$\frac{B_j}{F}$	عددی (مستقل)	۹۲.۸۶۳۹	۰.۰۰۳۷	۳.۰۱۱۵+۰.۳
۱۹	خوش حساب/بد حساب	-	اسمی (وابسته)	-	-	-

B: بدهی کل، B_j: بدهی جاری، D: دارایی کل، D_j: دارایی جاری، H: حقوق صاحبان سرمایه، M_k: موجودی کالا،

M_n: موجودی نقدی، S: سود خالص، F: فروش، R_s: نرخ سود انباشته، V: کل وام‌های بانکی

در جدول ۱ تمامی متغیرهای ذکر شده در ردیف ۱ تا ۱۸، متغیرهای مستقل بوده و براساس روابط ذکر شده در جدول محاسبه گردیده‌اند. این متغیرها بصورت مستقیم بر متغیر وابسته پایگاه داده (خوش حساب یا بد حساب بودن مشتری) تاثیرگذار می‌باشند. به منظور محاسبه مقادیر متغیرهای مستقل برای هر مشتری، تمامی اطلاعات ذکر شده در انتهای جدول برای هر مشتری جمع‌آوری شده و سپس با استفاده از روابط موجود در جدول فوق محاسبه شده‌اند. اولین گام در روش پیشنهادی، پیش پردازش اطلاعات مالی مشتریان می‌باشد.

پیش‌پردازش داده‌ها

مرحله پیش‌پردازش داده‌ها اولین گام در روش پیشنهادی بوده و به منظور آماده سازی پایگاه داده برای پردازش در مراحل بعد بکار می‌رود. این مرحله شامل دو گام اصلی جهت نرمالسازی داده‌ها می‌باشد. در روش پیشنهادی، از تبدیل موجک برای حذف نویز از داده‌ها استفاده می‌شود. توابع موجک برای رفع نواقص تبدیل فوریه ارایه شده است. یکی از مزایای تبدیل موجک نسبت به روش‌های فوریه سنتی، توان بالای تحلیل آنها در شرایطی است که سیگنال‌ها با گسستگی و جهش‌های سریع همراه هستند. استفاده از روش موجک در داده‌های مالی مزایای عمده‌ای دارد که از جمله آنها می‌توان به مطالعه مستقیم سری‌های زمانی نامانا و بررسی مدل‌ها و رفتارهای مالی در مقیاس‌های متفاوت اشاره نمود. تبدیل موجک پیوسته معمولاً حاوی حجم بالایی از اطلاعات است که نیازی به همه آنها در بازسازی تابع و مشاهده و تحلیل پیوستگی‌ها نیست و حتی می‌توان با نمونه‌برداری، ضرایب موجک را به نحوی به دست آورد که تابع اصلی را از آنها بازسازی نمود. در مدل پیشنهادی که در آن تبدیل موجک به عنوان یک روش نویززدایی به کار می‌رود، فرآیند نویز زدایی در سه مرحله صورت خواهد گرفت. ابتدا داده‌های اولیه به وسیله تبدیل موجک به لایه‌هایی تجزیه می‌شود. تبدیل موجک را می‌توان به صورت تجزیه تکه به تکه اطلاعات در یک سری زمانی در نظر گرفت. در این فرآیند ابتدا داده اولیه به وسیله تبدیل موجک به لایه‌هایی تجزیه می‌شوند که به این عمل تجزیه یا تحلیل تابع می‌گویند. هر بخش بردار داده‌ی تجزیه شده را می‌توان به صورت یک ضریب موجک و یک ضریب مقیاسگر در نظر گرفت. در واقع در تجزیه موجک داده‌ها به دو سری داده‌های پرفراوانی و کم فراوانی تقسیم می‌شوند. داده‌های کم فراوانی که از طریق اعمال موجک پدر بر سری اصلی حاصل می‌گردند، نشان دهنده ویژگی‌های اصلی سری است. داده‌های پرفراوانی نیز از طریق اعمال موجک مادر بر سری اصلی حاصل می‌گردند که اغلب نویز نامیده می‌شوند. در واقع هدف اصلی تجزیه موجک جداسازی ویژگی‌های اصلی سری از نویز است.

در روش پیشنهادی، از تابع موجک هار^۴ استفاده شده است. تابع موجک هار ساده ترین نوع تابع موجک بوده و به عنوان اولین نوع تابع موجک شناخته می‌شود. این تابع گسسته بوده و رفتاری مشابه با توابع پله‌ای دارد. با داشتن ورودی t ، تابع موجک مادر برای مدل هار را می‌توان بصورت رابطه زیر تعریف نمود:

$$\psi(t) = \begin{cases} 1 & 0 \leq t < 0.5 \\ -1 & 0.5 \leq t < 1 \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

تابع موجک تحت تاثیر فیلتر گذر از باند بوده و گذر در هر سطح موجب نصف شدن پهنای باند آن می‌شود. این مشکل موجب عدم پوشش تمام طیف خواهد شود. چرا که برای پوشش تمام طیف به بی‌نهایت سطح نیاز خواهیم داشت. برای حل این مشکل از تابع مقیاس گذاری (تابع پدر) استفاده شده است. این تابع با فیلتر نمودن پایین‌ترین سطح تبدیل از پوشش تمام طیف اطمینان حاصل می‌کند. در روش پیشنهادی از تجزیه گسسته موجک بدون اتلاف استفاده می‌کنیم. بدین منظور ابتدا هر رکورد پایگاه داده را در سطح چهارم و با استفاده از تابع هار تجزیه می‌کنیم تا لایه‌های تجزیه داده‌های مشتریان حاصل شود. سپس عملیات بازسازی داده بوسیله تبدیل معکوس موجک انجام خواهد شد تا داده‌های نویززدایی شده حاصل شوند. این داده‌های بعنوان ورودی گام بعدی روش پیشنهادی مورد استفاده قرار خواهند گرفت. در گام بعدی روش پیشنهادی، یک زیرمجموعه بهینه از ویژگی‌ها انتخاب خواهد شد.

^۴ Haar wavelet

انتخاب ویژگی با استفاده از روش پیشنهادی

پس از انجام عملیات پیش پردازش و استخراج ضرایب موجک از ویژگی‌های پایگاه داده، ضرایبی که دارای بیشترین فاصله با توزیع نرمال باشند به عنوان ویژگی‌های نهایی انتخاب می‌شوند. در روش پیشنهادی قصد داریم ویژگی‌هایی را انتخاب کنیم که به بهترین نحو دسته‌ها را از هم متمایز کنند. یکی از راهکارهای مرسوم برای این کار استفاده از معیار واریانس می‌باشد. هرچند ویژگی‌هایی که بزرگترین واریانس را داشته باشند، الزاما منجر به تمایز دقیق دسته‌ها نخواهند شد. برای رسیدن به بهترین حالات تفکیک دسته‌ها می‌توان ویژگی‌هایی را انتخاب کرد که دارای بیشترین انحراف از حالت نرمال باشند. توزیع نرمال یا گاوسی توصیف کننده یک توزیع استاندارد از داده‌های تصادفی است و نمونه‌های این توزیع دارای میانگین و انحراف معیار صفر می‌باشند. داده‌های تصادفی در واقع فاقد هرگونه اطلاعات قابل پردازش توسط الگوریتم‌های یادگیری ماشین می‌باشند. لذا حذف این دسته از داده‌ها موجب افزایش دقت پیش‌بینی در الگوریتم طبقه‌بندی خواهد گردید. به همین دلیل، در فرآیند انتخاب ویژگی در روش پیشنهادی، ویژگی‌هایی انتخاب خواهد گردید که دارای بیشترین فاصله ممکن از توزیع گاوسی باشند. براین اساس، اگر با فرض داشتن مجموعه ویژگی‌های انتخاب شده x ، توزیع گاوسی این داده‌ها را $G(x)$ بنامیم و تابع توزیع انباشته ویژگی‌ها را با $F(x)$ نمایش دهیم، آنگاه انحراف داده x از حالت نرمال بصورت زیر محاسبه خواهد شد:

$$D_x = \max(|F(x) - G(x)|) \quad (2)$$

در رابطه فوق، توزیع گاوسی $G(x)$ بصورت زیر محاسبه می‌گردد:

$$G(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad (3)$$

که در رابطه فوق x نشان دهنده داده‌های یک ویژگی در پایگاه داده می‌باشد. همچنین در رابطه ۲، تابع توزیع انباشته $F(x)$ نشان دهنده احتمال وجود یک نمونه با مقدار کمتر از مقادیر x در مجموعه مشاهدات داده بوده و بصورت زیر توصیف می‌شود:

$$F_X(x) = \Pr(X \leq x) \quad (4)$$

با توجه به رابطه ۲، D_x مجموعه ویژگی‌هایی خواهد بود که دارای بیشترین انحراف از حالت نرمال بوده و به عنوان ورودی مدل طبقه‌بندی استفاده می‌شوند.

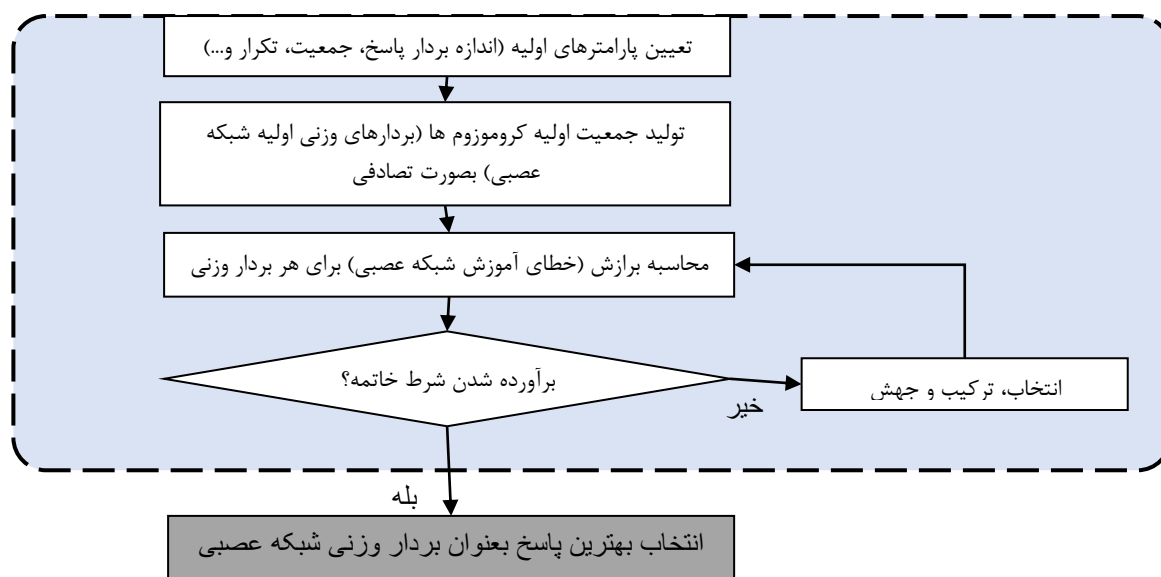
طبقه‌بندی ویژگی‌های انتخاب شده توسط روش پیشنهادی

پس از انتخاب ویژگی‌ها توسط روش پیشنهادی می‌توان عمل طبقه‌بندی داده‌ها را انجام داد. در روش پیشنهادی، عمل طبقه‌بندی توسط ترکیب شبکه عصبی مصنوعی و الگوریتم ژنتیک صورت می‌گیرد.

در این بخش به جزئیات مدل طبقه‌بندی پیشنهادی برای رتبه‌بندی مشتریان براساس خصوصیات آنها می‌پردازیم. این شبکه عصبی، یک شبکه پروسپترون با یک لایه مخفی می‌باشد. لایه مخفی این شبکه، دارای ۷ نرون بوده و تابع انتقال آن از نوع سیگموید لگاریتمی^۸ تعیین شده است. همچنین تعداد نرون‌های لایه ورودی برابر با تعداد ویژگی‌های استخراج شده برای هر نمونه (P) تعیین شده و تعداد نرون‌های لایه خروجی برابر با تعداد دسته‌های خروجی برابر با دو می‌باشد. در مدل شبکه عصبی، پیش از آغاز عمل آموزش ویژگی‌های استخراج شده هر نمونه نرمال‌سازی می‌شوند. به منظور نرمال‌سازی داده‌ها، تمامی مقادیر ورودی با استفاده از رابطه زیر تبدیل می‌شوند.

$$N_i = \frac{X_i - \min_i}{\max_i - \min_i} \quad (5)$$

در رابطه‌ی بالا، N_i مقدار نرمال شده صفت X_i مقادیر صفت \min_i و \max_i مقادیر ماکزیمم و مینیمم برای این صفت در پایگاه داده هستند. این رابطه تمامی مقادیر ورودی را به بازه‌ی [۰،۱] نگاشت می‌کند. پس از نرمال‌سازی داده‌ها، آموزش شبکه عصبی انجام می‌شود. در این گام از روش پیشنهادی و به منظور تعیین وزن‌های بهینه برای شبکه عصبی ذکر شده از الگوریتم ژنتیک استفاده می‌شود. در این روش، از راهکار تغییر اوزان ارتباطات بین نرون‌های شبکه عصبی و سپس محاسبه خطای پیش‌بینی استفاده می‌شود. در ادامه به تشریح روند بهینه‌سازی وزن‌ها در شبکه عصبی با استفاده از الگوریتم ژنتیک خواهیم پرداخت. شکل ۱ دیاگرام بلوک روش پیشنهادی برای آموزش شبکه عصبی با استفاده از الگوریتم بهینه‌سازی ژنتیک را نشان می‌دهد.



شکل ۱: دیاگرام بلوک روش پیشنهادی برای آموزش شبکه عصبی مصنوعی با استفاده از الگوریتم ژنتیک

مطابق شکل ۱، پس از تشکیل جمعیت اولیه کروموزوم‌ها از الگوریتم ژنتیک و عملگرهای ژنتیکی آن برای یافتن جواب بهینه استفاده می‌شود. الگوریتم ژنتیک یک تکنیک برنامه‌نویسی است که از تکامل ژنتیکی به عنوان یک الگوی حل مسئله استفاده می‌کند. میزان نزدیک شدن به جواب بهینه مسئله‌ای که باید حل شود از طریق تابع برازش^۱ محاسبه می‌شود. این تابع هر راه حل کاندید را ارزیابی می‌کند که اکثر آنها به صورت تصادفی انتخاب می‌شوند. یک راه‌حل برای مسئله مورد نظر، با یک لیست از پارامترها نشان داده می‌شود که به آنها کروموزوم یا ژنوم می‌گویند. کروموزوم‌ها عموماً به صورت یک رشته ساده از داده‌ها نمایش داده می‌شوند. در ابتدا از مقادیر تصادفی برای ایجاد نسل اول استفاده می‌شود. هر کروموزوم می‌تواند یک جواب احتمالی برای رسیدن به بهینگی باشد.

هدف الگوریتم ژنتیک در روش پیشنهادی، بهینه‌سازی بردار وزنی شبکه عصبی مصنوعی می‌باشد. به بیانی دیگر، بردار پاسخ در الگوریتم ژنتیک مورد استفاده در این بخش، تعیین کننده وزن ارتباطات بین نرون‌ها و همچنین بایاس‌های شبکه عصبی می‌باشد. بنابراین برای یک شبکه عصبی با I نرون ورودی، H نرون مخفی و P نرون خروجی، طول هر بردار پاسخ در الگوریتم ژنتیک برابر با $H \times (I + 1) + P \times (H + 1)$ خواهد بود.

^۱fitness

در طول هر نسل، هر کروموزوم با استفاده از تابع برازش ارزیابی می‌شود. بنابراین، در الگوریتم ژنتیک بهترین پاسخ با استفاده از تابع برازش تعیین می‌شود. پس از تعیین اوزان شبکه عصبی توسط بردار پاسخ، خروجی‌های شبکه عصبی برای نمونه‌های آموزش تولید شده و با مقادیر هدف واقعی مقایسه می‌شود. سپس از معیار مجموع مربعات خطا (SSE) به منظور ارزیابی عملکرد شبکه عصبی و بهینگی پاسخ تولید شده استفاده می‌شود.

$$SSE = \sum_{i=1}^N (T_i - Z_i)^2 \quad (6)$$

که در رابطه فوق، N نشان دهنده تعداد نمونه‌های آموزش بوده و T_i مشخص کننده مقدار هدف برای نمونه آموزشی i ام می‌باشد. همچنین Z_i نشان دهنده خروجی تولید شده توسط شبکه عصبی برای نمونه آموزش i ام است.

گام بعدی ایجاد نسل بعدی از جامعه است که بر پایه فرایندهای انتخاب، تولید مثل و جهش انجام می‌شود. برای هر فرد، یک جفت والد انتخاب می‌شود. انتخاب‌ها به گونه‌ای‌اند که مناسبترین عناصر انتخاب شوند تا حتی ضعیف‌ترین عناصر هم شانس انتخاب داشته باشند تا از نزدیک شدن به جواب محلی جلوگیری شود. بدین منظور در روش پیشنهادی از الگوریتم انتخاب چرخ رولت برای انتخاب والد‌ها استفاده می‌شود. فرض می‌کنیم که اندازه جمعیت P برابر با N_p می‌باشد. نرخ ترکیب را برابر با P_s و نرخ جهش را بصورت P_m در نظر می‌گیریم. در الگوریتم پیشنهادی، کروموزوم‌ها که هر یک حاوی یک پاسخ فرضی می‌باشند با استفاده از تابع برازش ارزیابی شده و میزان برازندگی آنها محاسبه می‌شود. سپس تمامی پاسخ‌های ممکن در جمعیت جاری به شیوه‌ای نزولی بر مبنای مقدار برازندگی خود طبقه‌بندی می‌شوند. اگر مقدار برازندگی یک کروموزوم مانند \bar{F}_i برابر با $FV(\bar{F}_i)$ باشد؛ آنگاه نسل جدید P_n به اینصورت تولید می‌شود:

- انتخاب: به شکل تصادفی اولین $N_p * p_s$ از P انتخاب می‌شود و سپس به P_n اضافه می‌شود.
- ترکیب (تقاطع): به طور تصادفی $(1 - P_s) \times N_p / 2$ از جمعیت P بر اساس $\Pr(F_i)$ انتخاب می‌شود. برای هر جفت کروموزوم، دو فرزند با اعمال عملگر تقاطع و اضافه کردن آنها به P_n تولید می‌شود. احتمال انتخاب $\Pr(\bar{F}_i)$ بصورت زیر تعریف می‌شود.

$$\Pr(\bar{F}_i) = \frac{FV(\bar{F}_i)}{\sum_{j=1}^{N_p} FV(\bar{F}_j)} \quad (7)$$

- جهش: هر کروموزوم با احتمال $P_m \times N_p$ از جمعیت P (بصورت تصادفی با توزیع یکنواخت) انتخاب می‌شود. برای هر کروموزوم، یک مکان (مرکز خوشه) انتخاب شده تصادفی با یک مقدار تصادفی در بازه مجاز متناظر با آن مکان در کروموزوم جایگزین می‌شود.

بطور کلی، عملگر ترکیب می‌تواند یکی از حالات تقاطع یک نقطه‌ای، دو نقطه‌ای و یکنواخت باشد. در روش پیشنهادی از ترکیب یک نقطه‌ای برای تولید کروموزوم‌های فرزند استفاده می‌شود. این فرایند باعث به وجود آمدن نسل جدیدی از کروموزوم‌هایی می‌شود، که با نسل قبلی متفاوت است. کل فرایند برای نسل بعدی هم تکرار می‌شود، جفت‌ها برای ترکیب انتخاب می‌شوند، جمعیت نسل‌های بعدی به وجود می‌آیند این فرایند تکرار می‌شود تا این که به یکی از شرایط خاتمه برسیم.

شرایط خاتمه الگوریتم ژنتیک در مورد طول زمان جستجو و احتمالاً بهینگی راه حل‌های پیدا شده تصمیم گیری می‌کند. شرایط خاتمه مورد استفاده در روش پیشنهادی عبارتند از:

- به تعداد ثابتی از نسل‌ها برسیم.
 - یک جواب پیدا شود که کمترین مقدار برازش (خطای ۰) را برآورده کند.
 - به بیشترین تعداد نسل‌های بدون بهبود برسیم. در این حالت فرض می‌شود که جستجوی بیشتر، نتایج بهتری را در بر نخواهد داشت.
- شبه کد جستجوی ژنتیک برای آموزش شبکه عصبی مصنوعی بصورت زیر می‌باشد.

Genetic Algorithm

begin

Choose initial population using FCM Clustering;

repeat

Evaluate the individual fitness of a certain proportion of the population;

Select pairs of best-ranking individuals using Roulette Wheel;

Apply Single point crossover operator;

Apply mutation operator;

Construct new Generation;

until (terminating condition);

end.

براین اساس، در روش پیشنهادی ویژگی‌های استخراج شده به دو دسته داده‌های آموزش و آزمون تقسیم‌بندی می‌شوند. داده‌های آموزش برای ساخت مدل شبکه عصبی ارائه شده بکار گرفته می‌شوند.

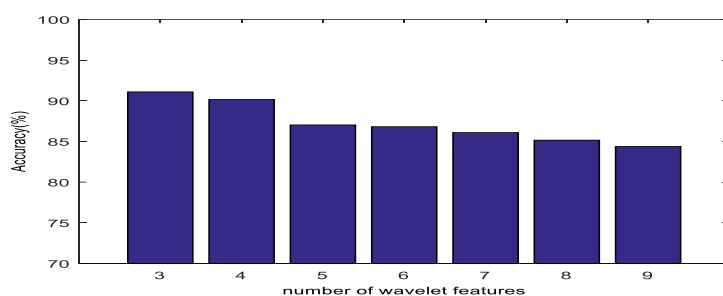
نتایج عملی

در روش پیشنهادی از روش ترکیب تک نقطه‌ای برای ترکیب کروموزوم‌های استفاده شده است. انتخاب کروموزوم برای عمل جهش نیز با توزیع گاوسی انجام می‌شود. به منظور انتخاب ویژگی‌ها ابتدا از تجزیه گسسته مویک به منظور تبدیل ویژگی‌ها استفاده می‌شود. تابع مویک مورد استفاده از نوع هار بوده و عمل تجزیه در چهار سطح انجام می‌شود. به منظور بالا رفتن دقت نتایج آزمایش، آزمایشات را ۱۰ بار تکرار می‌کنیم. در هر بار تکرار ۹۰٪ داده‌ها برای آموزش مدل و ۱۰٪ باقی برای آزمایش عملکرد آن استفاده می‌شوند. در هر تکرار داده‌های جدیدی برای آزمون مدل آموزش یافته استفاده می‌شود؛ بنابراین پس از ۱۰ بار تکرار آزمایش‌ها، تمامی داده‌ها در طی سناریو مورد آزمایش قرار می‌گیرند.

ارزیابی عملکرد روش پیشنهادی در رتبه‌بندی مشتریان

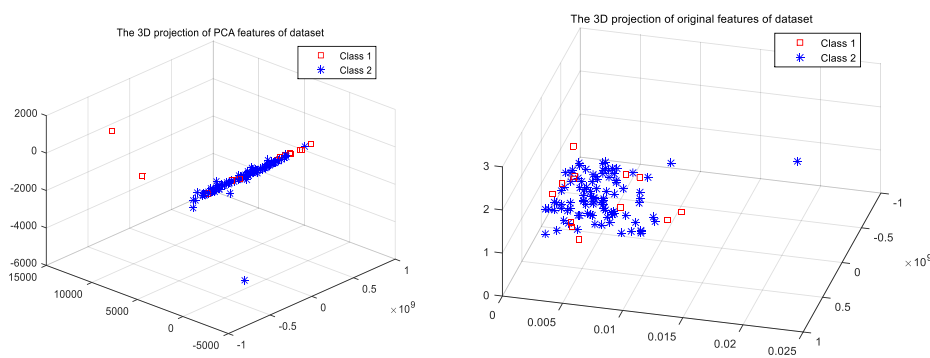
به منظور ارزیابی عملکرد روش پیشنهادی، نتایج بدست آمده را با الگوریتم استخراج ویژگی PCA و الگوریتم انتخاب ویژگی متوالی (SQFS) مقایسه کرده‌ایم. همچنین در تمامی حالات، نتایج بدست آمده را با حالت بدون استفاده از الگوریتم‌های استخراج ویژگی مقایسه کرده‌ایم. تعداد ویژگی‌های مناسب برای کاهش ابعاد داده با استفاده از انجام آزمایش و تکرار آن به ازای تعداد ویژگی‌های مختلف بدست آمده است. نتایج این آزمایشات در شکل ۴-۲ نمایش داده شده است. در این نمودار، محور افقی نشان دهنده تعداد ویژگی‌های استفاده شده در تجزیه مویک و محور عمودی دقت در رتبه‌بندی صحیح مشتریان با استفاده از این تعداد ویژگی توسط شبکه عصبی و الگوریتم ژنتیک را مشخص می‌کند. لازم به ذکر است که این نمودار میانگین دقت به ازای تعداد ویژگی‌های مورد آزمایش و با استفاده از الگوریتم طبقه‌بندی بحث شده را نمایش می‌دهد.

^۱SeQuential Feature Selection



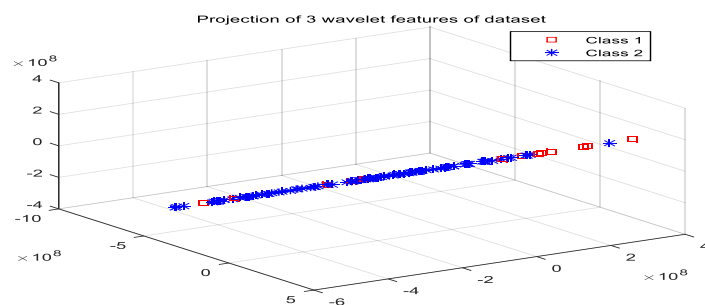
شکل ۲: میانگین دقت رتبه بندی مشتریان به ازای تعداد ویژگی‌های استخراج شده مختلف با تجزیه گسسته موجک در روش پیشنهادی

همانطور که در شکل ۲ نمایش داده شده، استفاده از ۳ ویژگی بیشترین دقت در طبقه‌بندی را به همراه خواهد داشت. مزیت استفاده از الگوریتم‌های تبدیل و استخراج ویژگی، توصیف پایگاه داده در ابعاد کوچک‌تر و بارزتر کردن اختلاف بین داده‌هایی است که در کلاس‌های مختلف قرار می‌گیرند.



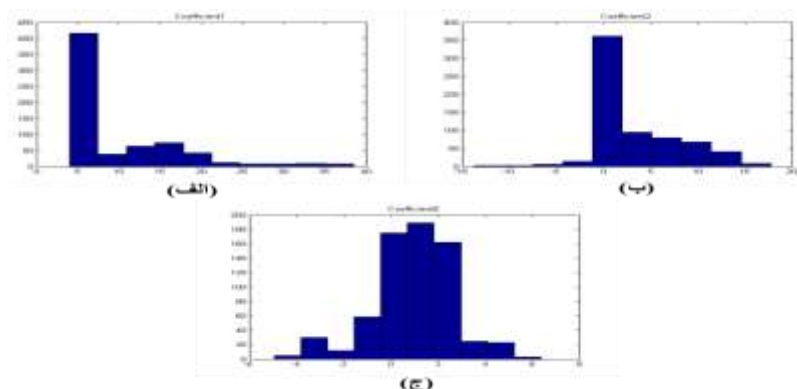
شکل ۳: توزیع دسته‌های خروجی در داده‌های اصلی پایگاه داده برای سه ویژگی انتخابی توسط الف: الگوریتم SQFS ب: الگوریتم PCA

همانطور که در شکل ۳ نمایش داده شده است، ویژگی‌های انتخاب شده توسط الگوریتم SQFS بخوبی نمی‌توانند دسته‌ها را از هم تفکیک کنند. این شکل نشان می‌دهد که استفاده از داده‌های اصلی بدون تبدیل آن ممکن است موجب طبقه‌بندی نادرست ویژگی‌های پایگاه داده شود. همچنین به منظور مقایسه کارایی روش پیشنهادی، با استفاده از الگوریتم PCA ابعاد ویژگی‌های پایگاه داده به سه ویژگی کاهش یافته است. نشان داد که دسته‌های خروجی توسط ویژگی‌های استخراج شده توسط الگوریتم PCA بصورت مناسب‌تری از هم تفکیک شده‌اند.



شکل ۴: توزیع دسته‌های خروجی در سه ویژگی استخراج شده توسط روش پیشنهادی

همانطور که در شکل ۴ نمایش داده شده است، روش پیشنهادی می‌تواند ویژگی‌ها را بصورتی انتخاب کند که از نظر دیداری نیز دسته‌ها قابل تفکیک باشند. با توجه به اینکه الگوریتم‌های رتبه‌بندی براساس فواصل و موقعیت داده عمل دسته‌بندی را انجام می‌دهند، این ویژگی منجر به دسته‌بندی داده‌ها با دقتی بالاتر از سایر روش‌های بررسی شده خواهد شد.



شکل ۵ نمودار فراوانی مقادیر مربوط به سه بردار ویژگی انتخاب شده توسط روش پیشنهادی برای پایگاه داده

در جدول ۲ زمان صرف شده برای پردازش داده‌های پایگاه داده و کاهش ابعاد ویژگی‌ها با استفاده از الگوریتم‌های SQFS، PCA و تجزیه مویک در روش پیشنهادی با هم مقایسه شده است.

جدول ۲ مقایسه زمان صرف شده برای کاهش ابعاد داده

الگوریتم	SQFS	PCA	تجزیه مویک (روش پیشنهادی)
زمان پردازش (ثانیه)	۴۵	۰,۸۵	۰,۵۱

آزمون زمان پردازش در جدول ۲ با استفاده از یک سیستم با پردازنده intel core i7 2.4Ghz و حافظه ۸GB انجام شده. پیاده سازی آزمایش زمان پردازش با استفاده از MATLAB 2015b و در سیستم عامل ویندوز نسخه ۷ انجام شده است. همانطور که نتایج جدول ۲ نشان می‌دهد روش SQFS دارای بیشترین زمان پردازش می‌باشد. همچنین روش پیشنهادی در ۰,۵۱ ثانیه می‌تواند ویژگی‌های پایگاه داده را انتخاب کند. در نتیجه روش پیشنهادی دارای زمان پردازشی کمتر نسبت به روش‌های SQFS و PCA است. دلیل بیشتر بودن زمان پردازش در روش PCA نسبت به روش پیشنهادی را می‌توان عملیات ضرب ماتریسی در آن نسبت داد. چرا که انجام ضرب ماتریسی موجب پیچیدگی محاسباتی بالا و تحمیل بار پردازشی بر سیستم خواهد شد.

به منظور ارزیابی روش پیشنهادی در رتبه‌بندی صحیح مشتریان بانک‌ها، عملکرد روش پیشنهادی را با دو الگوریتم طبقه‌بندی دیگر مقایسه می‌کنیم:

- ماشین بردار پشتیبان با تابع هسته خطی
- درخت تصمیم^{۲۲} CART

در این آزمایشات، هر یک از این الگوریتم‌ها را از نظر معیارهای: درستی^{۲۳}، حساسیت^{۲۴}، ویژگی^{۲۵} و سطح زیر نمودار ROC^{۲۶} مقایسه می‌کنیم. لازم به ذکر است که نتایج نمایش داده شده در ادامه این فصل میانگین حاصل از ۱۰ بار تکرار آزمایشات می‌-

^{۲۲}Classification and Regression Tree

^{۲۳}Accuracy

^{۲۴}Sensitivity

^{۲۵}Specifity

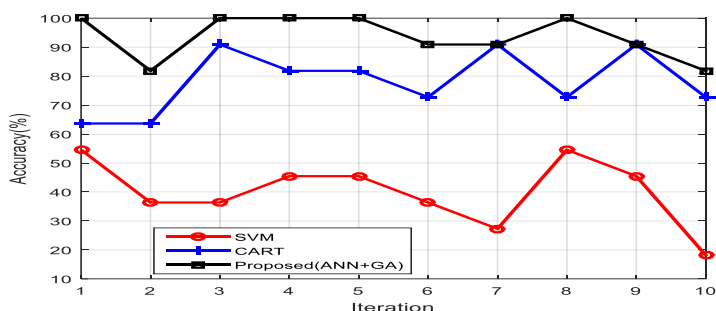
^{۲۶}Receiver Operating Characteristic

باشد. در هر بار تکرار، ۹۰٪ نمونه‌ها به عنوان نمونه‌های آموزش و ۱۰٪ باقیمانده به عنوان نمونه‌های آزمون در نظر گرفته شده‌اند. در جدول ۳ میانگین درستی مربوط به رتبه بندی صحیح مشتریان هر یک از الگوریتم‌های ذکر شده با استفاده از سه حالت مختلف ویژگی‌ها نمایش داده شده است.

جدول ۳- میانگین دقت در رتبه بندی صحیح مشتریان بانک

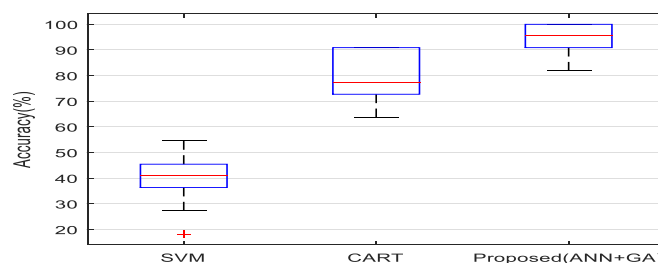
تجزیه موجک (روش پیشنهادی)	PCA	داده های اصلی	
۷۸.۱۸۱۸	۷۴.۳۷	۷۶.۵۶	درخت تصمیم CART
۴۰.۰۰	۳۴.۰۶	۳۸.۱۲	ماشین بردار پشتیبان
۹۳.۶۴	۸۹.۳۸	۸۸.۱۲	ANN+GA (روش پیشنهادی)

نتایج جدول ۳ نشان می‌دهد که با استفاده از روش پیشنهادی یعنی ترکیب تجزیه موجک برای انتخاب ویژگی‌ها و ترکیب شبکه عصبی مصنوعی و الگوریتم ژنتیک برای رتبه‌بندی مشتریان می‌توان به بالاترین دقت دست یافت. نتایج مربوط به دقت رتبه بندی مشتریان با استفاده از روش پیشنهادی و دو الگوریتم مقایسه شده در ۱۰ تکرار آزمایش در شکل ۲ نمایش داده شده است.



شکل ۲- دقت رتبه بندی مشتریان در ۱۰ بار تکرار آزمایش

همانطور که نتایج حاصل از بکارگیری روش پیشنهادی در شکل ۲ نشان می‌دهد، استفاده از ترکیب شبکه عصبی و الگوریتم ژنتیک می‌تواند موجب بهبود حداقل ۱۵.۵ درصدی دقت در رتبه بندی مشتریان نسبت به سایر الگوریتم‌های مقایسه شده - شود. در شکل ۳ نمودار جعبه‌ای^۷ درستی رتبه‌بندی مشتریان برای هر یک از الگوریتم‌های مقایسه شده نمایش داده شده است.

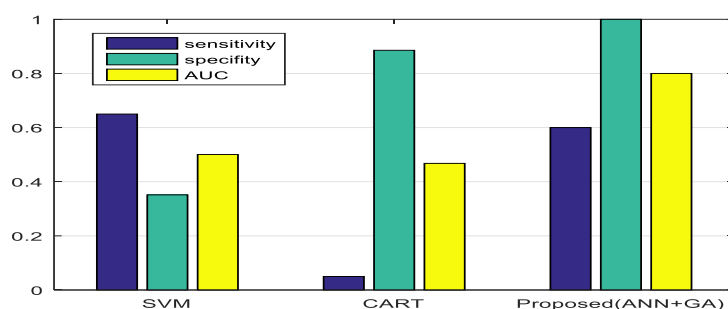


شکل ۳ : نمودار جعبه ای درستی رتبه‌بندی مشتریان برای هر یک از الگوریتم‌های مقایسه شده

^۷Box Plot

در نمودار جعبه‌ای، روشی از اعتماد بالاتری برخوردار خواهد بود که علاوه بر داشتن میانگین درستی بیشتر، دارای بازه تغییرات محدودتری باشد. در شکل ۳ نشان داده شده که الگوریتم پیشنهادی علاوه بر داشتن میانگین درستی بیشتر، از بازه تغییرات محدودتری نسبت به دو الگوریتم دیگر برخوردار است.

در شکل ۴ نتایج حاصل از محاسبه معیارهای حساسیت، ویژگی و سطح زیر نمودار ROC برای روش پیشنهادی و دو الگوریتم مقایسه شده نمایش داده شده است.



شکل ۴ مقایسه معیارهای حساسیت، ویژگی و سطح زیر نمودار ROC در رتبه بندی مشتریان

معیار حساسیت برای اندازه گیری نسبت کل افراد خوش‌حسابی است که بطور صحیح طبقه بندی شده‌اند و بصورت زیر محاسبه می‌شود:

$$\text{Sensitivity} = 100 \times \frac{TP}{TP+FN} \quad (8)$$

که در رابطه فوق، TP تعداد افراد خوش‌حسابی است که بصورت صحیح تشخیص داده شده‌اند و FN تعداد افراد خوش‌حسابی که اشتباه تشخیص داده شده‌اند. معیار ویژگی برای اندازه‌گیری افراد بدحسابی که بصورت صحیح طبقه‌بندی شده‌اند بکار می‌رود. این معیار بصورت زیر محاسبه می‌شود:

$$\text{Specificity} = 100 \times \frac{TN}{TN+FP} \quad (9)$$

در رابطه فوق، TN تعداد افراد بدحسابی است که بصورت صحیح تشخیص داده شده‌اند و FP تعداد افراد بدحسابی که اشتباه تشخیص داده شده‌اند. همانطور که از نتایج ارائه شده در شکل ۵ مشخص است، استفاده از ترکیب شبکه عصبی و الگوریتم ژنتیک در روش پیشنهادی بیشترین میزان حساسیت و سطح زیر نمودار ROC را در پی خواهد داشت. این نتایج نشان می‌دهد که الگوریتم پیشنهادی در تشخیص نمونه‌های TP و TN بهتر از سایر الگوریتم‌های مقایسه شده عمل می‌کند. در شکل ۶ ماتریس درهم‌ریختگی حاصل از طبقه‌بندی نمونه‌های آزمون توسط روش پیشنهادی نمایش داده شده است. در ماتریس در هم ریختگی نمایش داده شده، عدد ۱۰ واقع در سطر و ستون اول نشان دهنده تعداد مشتریان بدحسابی است که بدرستی توسط الگوریتم طبقه‌بندی تشخیص داده شده‌اند. این عدد در ماتریس درهم ریختگی با نام TN شناسایی می‌شود. عدد ۷ واقع در سطر دوم و ستون اول نشان دهنده تعداد مشتریان بدحسابی است که بصورت اشتباه توسط الگوریتم، خوش-حساب تشخیص داده شده‌اند. این عدد در ماتریس درهم ریختگی با نام FP شناسایی می‌شود. سطر دوم و ستون دوم ماتریس درهم ریختگی که عدد ۹۳ را نمایش می‌دهد، مشتریان خوش‌حسابی را مشخص می‌کند که بدرستی توسط الگوریتم طبقه‌بندی شده‌اند و بعنوان نمونه‌های TP شناسایی می‌شوند. همچنین عدد ۰ واقع در سطر اول و ستون دوم نشان دهنده تعداد

مشتريان خوش حسابی است که بصورت اشتباه توسط الگوریتم طبقه‌بندی بدحساب تشخیص داده شده‌اند. این عدد در ماتریس درهم ریختگی با نام FN شناسایی می‌شود.

Proposed(ANN+GA)

Output Class	1	10 9.1%	0 0.0%	100% 0.0%
	2	7 6.4%	93 84.5%	93.0% 7.0%
		58.8% 41.2%	100% 0.0%	93.6% 6.4%
		1	2	
		Target Class		

شکل ۶: ماتریس درهم ریختگی روش پیشنهادی برای ۱۰ تکرار آزمایش

در ماتریس‌های درهم‌ریختگی نمایش داده شده، عدد ۱ نمایش دهنده دسته مشتریان بدحساب و عدد ۲، مشخص کننده دسته مشتریان خوش حساب می‌باشد. همچنین ماتریس درهم ریختگی مربوط به سایر الگوریتم‌های مقایسه شده در شکل ۷ نمایش داده شده است.

CART

Output Class	1	1 0.9%	11 10.0%	8.3% 91.7%
	2	13 11.8%	85 77.3%	86.7% 13.3%
		7.1% 92.9%	88.5% 11.5%	78.2% 21.8%
		1	2	
		Target Class		

(ب)

SVM

Output Class	1	10 9.1%	62 56.4%	13.9% 86.1%
	2	4 3.6%	34 30.9%	89.5% 10.5%
		71.4% 28.6%	35.4% 64.6%	40.0% 60.0%
		1	2	
		Target Class		

(الف)

شکل ۷ ماتریس درهم ریختگی الگوریتم‌های مقایسه شده برای ۱۰ تکرار آزمایش: (الف) SVM، (ب) درخت تصمیم

نتیجه گیری و تحقیقات آتی

در این تحقیق یک مدل تطبیقی جدید به منظور رتبه بندی مشتریان بانک مبتنی بر ترکیب شبکه عصبی و الگوریتم ژنتیک ارائه شد. در مدل پیشنهادی، ابتدا یک مجموعه داده آموزش به منظور ساخت مدل طبقه‌بندی مورد استفاده قرار گرفت. این مجموعه داده شامل مشخصات مالی مشتریان بانک می‌باشد. این داده‌ها پس از پردازش توسط یک الگوریتم استخراج مبتنی بر تبدیل موجک، ویژگی به ضرایب موجک تبدیل شده و بارزترین ضرایب به عنوان ویژگی‌های استخراج شده مورد استفاده قرار می‌گیرند. با استفاده از ویژگی‌های استخراج شده مدل شبکه عصبی مصنوعی آموزش داده شده و مدل طبقه بندی تولید می‌شود. در روش پیشنهادی، به منظور آموزش شبکه عصبی و تعیین بردار وزن بهینه نرون‌های آن، از الگوریتم ژنتیک استفاده می‌شود. پس از تولید مدل یادگیر، در فاز آزمایش ابتدا اطلاعات مشتری مورد آزمون جمع‌آوری می‌شود. سپس با استفاده از تبدیل موجک ویژگی‌های مربوط به مشتری استخراج شده و توسط الگوریتم طبقه‌بندی ارزیابی می‌شوند. در نهایت، خروجی سیستم تشخیص مبنی بر خوش حساب یا بدحساب بودن مشتری تولید می‌شود. به منظور ارزیابی عملکرد روش پیشنهادی از نرم افزار MATLAB استفاده شده است. بدین منظور از یک پایگاه داده شامل اطلاعات ۱۰۸ مشتری حقیقی و حقوقی بانک ملی شعبه مرکزی شهر ایلام در طی سال ۱۳۹۶ استفاده شده است. مدل پیشنهادی از جنبه‌های مختلف مورد ارزیابی قرار گرفته و دقت آن در رتبه بندی مشتریان بانک آزمایش شد. نتایج آزمایشات نشان داد که مدل پیشنهادی می‌تواند رتبه بندی

مشتریان بانک را با میانگین دقت ۹۳٫۶۴٪ انجام دهد. مقایسه نتایج بدست آمده از آزمایش روش پیشنهادی با الگوریتم‌های مشابه نشان می‌دهد که مدل ارائه شده دارای دقت و کارایی بالاتری در رتبه بندی مشتریان بانک‌ها می‌باشد. با توجه به نتایج پژوهش، پیشنهادهای کاربردی به شرح زیر قابل ارائه است:

- پیشنهاد می‌شود یک دیتابیس حاوی داده های مالی و مدیریتی متعلق به مشتریان بانک ملی شهر ایلام که متغیرهای پیش بینی کننده خطر اعتباری را ذخیره میکند ایجاد شود. این دیتابیس بایستی بر اساس داده های جدید برای تخمین میزان خطر اعتباری مشتریان در هر زمان ممکن به روز شود.
- پیشنهاد می‌شود یک سیستم نرم افزاری برای مدل های الگوریتم ژنتیک و شبکه عصبی مصنوعی برای پیش بینی خطر اعتباری مشتریان طراحی ایجاد گردد.
- پیشنهاد می‌شود استفاده از سایر الگوریتم‌های بهینه سازی مانند الگوریتم‌های بهینه سازی چندهدفه در جهت تعیین بردار وزنی بهینه شبکه عصبی بررسی شود.
- استفاده از سایر توابع تجزیه گسسته موزون (مانند تابع موزون دوگانه) می‌تواند در جهت بهبود روند استخراج ویژگی موثر باشد.
- ارزیابی کارایی الگوریتم پیشنهادی توسط سایر الگوریتم‌های طبقه بندی می‌تواند موضوع کارهای آتی باشد.
- پیشنهادی می‌شود کارایی روش پیشنهادی در رتبه‌بندی مشتریان بانک‌ها برای سایر پایگاه داده‌های فراهم شده توسط بانک‌ها و مؤسسات مالی مورد ارزیابی قرار گیرد.
- پیشنهادی می‌شود برای پیش بینی خطر اعتباری مشتریان در یک سیستم خبره و ایجاد رتبه بندی اعتباری مشتریان بر اساس این مدل‌ها از ترکیب روش های کمی و کیفی استفاده گردد
- پیشنهادی می‌شود در مدل پیش بینی ریسک اعتباری بر اساس بازخورد مداوم نتایج تجدید نظر گردد
- پیشنهادی می‌شود بین واحدهای درگیر در فرآیند اعتباردهی، مانند مدیریت ارزیابی، مدیریت اعتبار، مدیریت ریسک اعتباری و فرآیندهای اداری هماهنگی صورت گیرد.

منابع

- البرزی، م؛ خان بابایی، م؛ محمدپور زرنندی، م؛ (۱۳۹۵)، بکارگیری تکنیک‌های خوشه بندی و الگوریتم ژنتیک در بهینه سازی درختان تصمیم گیری برای اعتبارسنجی مشتریان بانک‌ها، فصلنامه آینده پژوهی مدیریت، سال اول / شماره اول / زمستان ۱۳۹۵.
- دهمرد، ن؛ شهرکی، ج؛ سیف الدین پور، س؛ اسفندیاری، م؛ (۱۳۹۶). اعتبار سنجی مشتریان بانک با استفاده از رویکرد امتیازدهی اعتباری. پژوهش های مدیریت عمومی ۱۸: ۱۵۲-۱۳۵
- صالحی، مجتبی؛ کرد کتولی، علیرضا (۱۳۹۶). انتخاب ویژگی های بهینه به منظور تعیین ریسک اعتباری مشتریان بانکی، فصلنامه مطالعات مدیریت کسب و کار هوشمند سال ششم، شماره ۲۲، زمستان ۹۶، صفحات ۱۲۹ تا ۱۵۴
- Arun Kumar, R., & Kotreshwar, G. (2006). Risk management in commercial banks: A case study of public and private sector banks. Indian institute of capital markets, 9th capital markets conference. Retrieved from SSRN <http://ssrn.com/abstract=877812>
- Bi-plab Bhattacharjee, Amulyashree Sridharm, Muhammad Shafi (2017) An artificial neural network-based ensemble model for credit risk assessment and deployment as a graphical user interface, Int. J. Data Mining, Modelling and Management, Vol. 9, No. 2, 2017

- Gang Dang, Kin Keung Lai, Jerome Yen, (2016). "Credit scorecard based on logistic regression with random coefficients". *Procedia Computer Science* 1: 2463–2468.
- Huang, C. L., Chen, M. C., & Wang, C. J. (2007). Credit scoring with a data mining approach based on support vector machines. *Experts Systems with Applications*, 33, 847–856. doi:10.1016/j.eswa.2006.07.007
- Hussein A. Abdou, (2014). "Genetic programming for credit scoring: The case of Egyptian public sector banks". *Expert Systems with Applications* 36: 11402–11417.
- Jha, S., Sisodia, R., Khandelwal, Y., Dube, N., & Gunjal, A. (2017). Financial Analysis Using Data Mining Techniques and Modelling. *Asian Journal For Convergence In Technology (AJCT)*, 3.
- Jiao Tang, Junkai Ji, Shangce GAO, Hongwei Dai, Yang Yu, and Yuki Todo(۲۰۱۷) A Pruning Neural Network Model in Credit Classification Analysis, *Computational Intelligence and Neuroscience* Volume 2018, Article ID 9390410, 22 pages <https://doi.org/10.1155/2018/9390410>
- Kamesh, V., Karthick, M., Kavin, K., Velusamy, M., & Vidhya, R. (2019). Real-Time Fraud Anomaly Detection in E-banking Using Data Mining Algorithm.
- Lahsana, A., Ainon, R., & Wah, T. (2010). Credit scoring models using soft computing methods: A survey. *The International Arab Journal of Information Technology*, 7(2), 129–139.
- Limsombunchai, G. V. C., & Lee, M. (2005). Lending decision model for agricultural sector in Thailand, *American Journal of Applied Science*, 2(8), 1198–1205.
- Olokoyo, F. (2011). Determinants of commercial banks lending in Nigeria. *International of Financial Research*, 2(2). doi:10.5430/ijfr.v2n2p61
- Raghavendra, B. K., & Simha, J. (2010). Evaluation of feature selection methods for predictive modeling using neural networks in credits scoring. *The International Journal of Advanced Networking and Applications*, 2(3), 714–718.
- S.M.Taghavi Takyar, R.Aghajan Nashtaei*, E.Chirani (2014) The Comparison of Credit Risk between Artificial Neural Network and Logistic Regression Models in Tose-Taavon Bank in Guilan, *International Journal of Applied Operational Research* Vol. 5, No. 1, pp. 63-72, Winter 2015
- Santos Silva, J.M.C., Murteira, J.M.R., (2009). "Estimation of default probabilities using incomplete contracts data". *Journal of Empirical Finance* 16, 457–465.
- Yobas M.B., Crook J.N., Ross P., (2011). "Credit scoring using neural and evolutionary techniques".