



نظارت مؤثر بانکی در اعطای تسهیلات خرد با تأکید بر رتبه‌بندی اعتباری مشتریان حقیقی

غلامرضا خجسته^۱

چکیده

پژوهش حاضر با تأکید بر به‌کارگیری مدل ترکیبی رگرسیون لجستیک، شبکه‌های عصبی مصنوعی و رگرسیون سمبلیک و با هدف نظارت مؤثر بر تسهیلات خرد، عوامل اثرگذار بر ریسک اعتباری مشتریان حقیقی را در شعب منتخب بانک سپه شیراز شناسایی می‌کند. با بررسی یک نمونه ۳۵۱ تایی از مشتریان حقیقی، ۱۷ متغیر شامل میزان وام، مهلت بازپرداخت، نرخ بهره، درآمد، سن، تعداد چک برگشتی، سابقه بدهی، طول عمر حساب، نوع وثیقه، تحصیلات، جنسیت، اشتغال همسر، تأهل، وضعیت ملکی، شغل، نوع وام، تکلیفی یا غیرتکلیفی بودن، برای طبقه‌بندی مشتریان به خوش حساب و بدحساب استخراج شد. با استفاده از روش انتخاب رو به جلو والد، ۵ متغیر تأثیرگذار بر ریسک اعتباری انتخاب شد. آموزش شبکه‌های عصبی نیز با استفاده از ۳ نرون در لایه پنهان انجام گرفت. انتخاب نقطه برش بهینه، بر اساس منحنی مشخصه عملکرد سیستم انجام شد. نتایج خروجی شبکه عصبی مصنوعی روی داده‌های آزمایش نشان داد که دقت مدل ترکیبی رگرسیون لجستیک - شبکه‌های عصبی مصنوعی، در طبقه‌بندی مشتریان خوش حساب برابر ۰/۸۹ و در طبقه‌بندی مشتریان بدحساب برابر ۰/۸۳ است که از رگرسیون لجستیک و مدل ترکیبی رگرسیون لجستیک - رگرسیون سمبلیک بهتر است.

واژه‌های کلیدی: نظارت مؤثر بانکی، ریسک اعتباری، رگرسیون لجستیک، شبکه‌های عصبی مصنوعی، رگرسیون سمبلیک، منحنی مشخصه عملکرد سیستم.

طبقه‌بندی JEL: G۲۱.

۱. دکتری، گروه مدیریت دولتی، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران؛ gh_khojasteh@yahoo.com

مقدمه

طی دو دهه گذشته، سیستم بانکی در سراسر دنیا در محیط فعالیت خود تغییرات شایان توجهی را تجربه کرده و عوامل خارجی و داخلی متعددی بر ساختار و عملکرد سیستم بانکی تأثیرگذار بوده است. با وجود این، برخلاف تمامی تغییرات، سیستم بانکی همچنان میدان دار اصلی تأمین مالی فعالیت‌های اقتصادی در بسیاری از کشورها، از جمله ایران است و در انتقال منابع از پس‌اندازکنندگان به واحدهای سرمایه‌گذاری، نقش اصلی ایفا می‌کند. جذب منابع از اهداف کلیدی، اساسی و راهبردی بانک‌ها و مؤسسه‌های مالی اعتباری به شمار می‌رود و در ارائه خدمات بانک‌ها نقش ویژه‌ای ایفا می‌کند و شاخص مهمی در ارزیابی موفقیت بانک‌ها محسوب می‌شود. به طور کلی، در اقتصادهایی نظیر اقتصاد ایران که سیستم مالی آنها مبتنی بر بانک است، بانک‌ها سهم عمده‌ای در تجهیز و تخصیص منابع مالی از طریق سپرده‌پذیری و اعطای تسهیلات دارند. از این دیدگاه، تحولات حجم سپرده‌های بانکی و شناسایی عوامل مؤثر بر میزان شکل‌گیری آنها به واسطه دو کارکرد یادشده نزد پس‌اندازکنندگان، مدیران نظام بانکی و سیاست‌گذاران اقتصادی از اهمیت ویژه‌ای برخوردار است (علم‌الهدی، ۱۳۹۷).

اعطای وام، بانک را در معرض ریسک اعتباری قرار می‌دهد. ریسک اعتباری^۱ یا ریسک مرتبط با عدم بازپرداخت مبلغ بدهی، قدیمی‌ترین و رایج‌ترین شکل ریسک در بازار مالی در تاریخ بانکداری محسوب می‌شود. این نوع ریسک، مهم‌ترین نوع ریسک در تراکنش‌های مالی، تجاری و بازرگانی است (اسپوچاکوا و کاگ^۲، ۲۰۱۴).

بخشی از مشکلات امروز بانک‌ها در زمینه افزایش مطالبات معوق و سوخت‌شده، به دلیل بهره‌نگرفتن آنها از نظام اندازه‌گیری و مدیریت ریسک اعتباری است (توران^۳، ۲۰۱۶). پس اندازه‌گیری ریسک اعتباری با هدف تعیین میزان زیان بالقوه عملیات اعتباری، از ضروریات نظام بانکی است.

در حال حاضر، به دلیل حجم بالای تسهیلات بانکی، ریسک بازپرداخت آنها برای بانک‌ها، به مشکلی بزرگ تبدیل شده است. در واقع، ریسک در ذات فعالیت‌های بانکی نهفته است و در عمل، حذف ریسک از عملیات بانکداری غیرممکن به نظر می‌رسد. پس، می‌توان گفت که تنها راه‌حل، طراحی و به‌کارگیری مدل اندازه‌گیری ریسک اعتباری است. با این حال، امروزه در بیشتر بانک‌های

1. Credit Risk
2. Spuchakova & Cug
3. Turan

تجاری کشور و در بیشتر موارد برای تعیین ریسک اعتباری مشتریان از روش قضاوتی استفاده می‌شود (میرزایی، نظریان و باقری، ۱۳۹۰). بدین ترتیب، می‌توان ادعا کرد که طراحی و استقرار مدل اندازه‌گیری ریسک اعتباری در نظام بانکی، در راستای بالا بردن بهره‌وری بانک‌های کشور در تخصیص بهینه منابع، نقش کارآمدی خواهد داشت (تهرانی و فلاح شمس، ۱۳۸۴). در مجموع، یک ارزیابی درست از ریسک اعتباری می‌تواند به نوعی استفاده بسیار مؤثر از سرمایه اقتصادی تبدیل شود. وقتی بعضی از مشتریان نمی‌توانند بدهی‌های خود را بازپرداخت کنند، نوعی شکست سیستم اقتصادی برای سازمان‌های مالی وام‌دهنده رخ می‌دهد. بنابراین، اهمیت بهبود در موضوع تصمیم‌گیری در خصوص اعطای اعتبار به مشتریان بانک و درجه‌بندی اعتبار، یکی از مسائلی است که بیشتر باید به آن پرداخت. با توجه به اینکه بخش شایان توجهی از تسهیلات بانک‌ها به اشخاص حقیقی پرداخت می‌شود، تمرکز روی این گروه و شناسایی عوامل مؤثر بر بازپرداخت نشدن تسهیلات از سوی آنها و متغیرهای کلیدی مؤثر در ریسک اعتباری این مشتریان با لحاظ پارامترهایی چون احتمال نکول، میزان تعهدات و نرخ بازیافت در صورت نکول و با توجه به متغیرهای کیفی و مالی (شخصیت، ظرفیت، سرمایه، وثیقه، شرایط کلی شرایط و ضوابط اعطای تسهیلات)، در راستای کاهش ریسک اعتباری در بانک گامی مفید است. به همین سبب، انجام پژوهش در این زمینه، برای تبیین تصمیم‌های اعتباری بانک‌ها ضروری خواهد بود.

اهمیت پژوهش حاضر را می‌توان از دو بعد اهمیت موضوع ریسک اعتباری و اهمیت مدل پژوهش بررسی کرد. سید وام، قسمت عمده دارایی‌های بانک‌های تجاری را تشکیل می‌دهد. تخصص محوری این بانک‌ها، جذب سپرده از سرمایه‌گذاران و اعطای وام به متقاضیان وجوه است. سپرده‌های جذب‌شده، بانک را به پرداخت سود و اصل آن در سررسیدهای معین متعهد می‌کند و این در حالی است که وام‌های پرداخت‌شده بانک را در معرض نکول وام‌گیرندگان قرار می‌دهد. بنابراین، بررسی اعتبار متقاضیان برای تصمیم‌گیری در زمینه اعطای وام به آنها از اهمیت ویژه‌ای برخوردار است. در پژوهش‌های مختلف در زمینه رتبه‌بندی مشتریان بانکی، از مدل‌های زیادی استفاده شده است. غالب پژوهش‌ها به سمت مدل‌های ترکیبی رو آورده‌اند. فلسفه مدل‌های ترکیبی، استفاده از مزیت‌های نسبی اجرای انفرادی به منظور بهره‌برداری هم‌زمان آنها است. پژوهش حاضر نیز از مدل ترکیبی شبکه‌های عصبی و رگرسیون لجستیک و همچنین مدل ترکیبی رگرسیون سمبلیک - رگرسیون لجستیک استفاده خواهد کرد. در همین راستا، این پژوهش با هدف ارتقای نظارت مؤثر بانکی در اعطای تسهیلات و با تأکید بر به‌کارگیری رویکردهای ترکیبی لجستیک - شبکه‌های

عصبی و لجستیک - رگرسیون سمبلیک در پی شناسایی عوامل مؤثر بر ریسک اعتباری و رتبه‌بندی اعتباری مشتریان حقیقی بانک و همچنین مقایسه نتایج به‌دست‌آمده از دو رویکرد بیان‌شده برای پیش‌بینی هرچه بهتر ریسک اعتباری مشتریان و در نتیجه امکان نظارت هرچه بهتر در سیستم بانکداری است.

پیشینه پژوهش

بانک‌ها، نهادهای مالی هستند که به‌عنوان سازمان‌های واسطه‌گری در نظام بانکی نقشی بسیار مهم و حساسی ایفا می‌کنند. این نهادها از طریق جذب منابع به‌صورت سپرده‌های جزء و کلان و تخصیص آنها به‌صورت بهینه و کارشناسانه از طریق اعطای تسهیلات و سرمایه‌گذاری، در اقتصاد کشور نقشی اساسی دارند. ماهیت خدمات بانکی، مستلزم پذیرش ریسک‌های مختلف از جمله ریسک اعتباری است که این مهم، صنعت بانکداری را از بعد نظارت از سایر صنایع متمایز و با مشکلات مهم روبه‌رو کرده است. از این رو، موضوع نظارت و مدیریت ریسک در بانک‌ها از اهمیت ویژه‌ای برخوردار است. واضح است، نبود نظارت در تخصیص منابع بانک به‌ویژه با رونق اعتباری، با خطر ریسک اعتباری و افزایش مطالبات معوق مواجه می‌کند. رونق اعتبارات به گسترش شایان توجه اعتبارات به بالاتر از سطحی که به‌طور معمول طی یک سیکل تجاری تجربه می‌شود، اشاره دارد.

پیشینه پژوهش‌های داخلی

رونق اعتبارات ممکن است به این دلیل رخ دهد که مدیران ریسک‌پذیر، ویژگی‌های ریسکی یک وام‌گیرنده بالقوه را به‌دلیل یک رابطه از قبل موجود با بانک نادیده می‌گیرند. رونق اعتبارات، نتیجه نبود نظارت مؤثر بانکی و در نتیجه، ارزیابی ضعیف و خوش‌بینی بیش از حد مسئولان اعتباری بانک در زمینه درخواست‌های اعتباری وام‌گیرندگان باشد (پورعبادالهیان کویچ، اصغرپور، فلاحی و ستاره‌متی، ۱۳۹۷).

اهمیت مطالبات معوق و سررسیدشده بانک‌ها به‌اندازه‌ای است که به‌تازگی مؤسسه‌های بین‌المللی از جمله صندوق بین‌المللی پول، در خصوص آن پژوهش‌های زیادی انجام داده‌اند که محور آن تعیین روش‌ها و شیوه‌ای است که توسط آن مطالبات معوق نظام بانکی در هر کشور به‌نحوی در حساب‌های ملی آن کشور منعکس شود. بر این اساس، مطالبات معوق می‌تواند بر تمام بخش‌های اقتصادی کشور تأثیر بگذارد، اما جدی‌ترین تأثیرات آن بر مؤسسه‌های مالی، بانک‌های

تجاری و مؤسسه‌های رهنی که دارای پرتفوی وسیعی از وام‌ها هستند، خواهد بود که به‌طور غیرمستقیم بر مشتریان این مؤسسه‌ها که واسط و جوه سپرده‌گذاران و صاحبان سهام هستند، مؤثر واقع خواهد شد (اسماعیلی، ۱۳۹۷).

در خصوص مدیریت ریسک اعتباری دو مسئله عمده وجود دارد. یکی، متغیرهای اثرگذار و پیش‌بینی‌کننده این ریسک و دیگری، مدل‌های استفاده‌شده برای ترکیب متغیرهای پیش‌بینی‌کننده به‌منظور رسیدن به دقت مناسب در طبقه‌بندی. بدین ترتیب، شناسایی عوامل مؤثر و متغیرهای دخیل بر ریسک اعتباری بسیار حائز اهمیت است.

در پژوهش‌های بسیاری نظیر پژوهش میرزایی و همکاران (۱۳۹۰) و بیک‌زاده، آقازاده و آقازاده (۱۳۹۳)، متغیرهای کلیدی مؤثر در ریسک اعتباری مشتریان بانکی بر مبنای مدل ۵C شناسایی و بررسی شده‌اند که شامل متغیرهای کیفی و مالی، شخصیت، ظرفیت، سرمایه، وثیقه، شرایط کلی شرایط و ضوابط اعطای تسهیلات است.

به‌لحاظ تجربی باید به این مورد توجه کرد که بیشتر پژوهش‌های انجام‌شده در زمینه شناسایی عوامل مؤثر بر ریسک اعتباری، به شاخص‌های ۵C و به‌تازگی با اضافه شدن عاملی با عنوان «شرایط و ضوابط تسهیلات (متقاضی چه میزان تسهیلات برای چه مقصودی و برای چه مدتی درخواست کرده)» به آن، از شاخص ۶C استفاده کرده‌اند (قاسمی و دنیایی، ۱۳۹۴). پژوهش حاضر در خصوص انتخاب متغیرهای کاندید در زمینه توضیح ریسک اعتباری، از ۱۷ متغیر مالی و غیرمالی استفاده می‌کند. در پژوهش‌های مختلف در زمینه ریسک اعتباری بانک‌ها (چه در زمینه مشتریان حقیقی و چه در زمینه مشتریان)، از متغیرهای توضیح‌دهنده متفاوت و مدل‌های مختلفی استفاده شده است. برای نمونه می‌توان به موارد زیر اشاره کرد:

حاجی‌ها و قربانی (۱۴۰۰) در پژوهشی با عنوان بررسی تأثیر ریسک عدم بازپرداخت بدهی‌ها بر رتبه‌بندی اعتباری با توجه به نقش تعدیلی محدودیت مالی به ریسک اعتباری پرداخته و به‌منظور آزمون فرضیه‌های پژوهش از الگو داده‌های ترکیبی استفاده کرده‌اند. نتایج آزمون فرضیه‌های پژوهش نشان داد که ضعیف عمل کردن در بازپرداخت اصل و بهره تعهدات مالی، باعث بی‌اعتمادی اعتباردهندگان می‌شود که رتبه اعتباری را کاهش می‌دهد. همچنین، نتایج پژوهش نشان داد که محدودیت مالی بر رابطه بین ریسک عدم بازپرداخت بدهی‌ها و رتبه‌بندی اعتباری، دارای تأثیر مثبت است. بدین ترتیب، افرادی که دارای محدودیت مالی هستند، ریسک عدم بازپرداخت بدهی‌ها بالایی

دارند و در نتیجه آن، بی‌اعتمادی اعتباردهندگان افزایش می‌یابد که به کاهش رتبه اعتباری منجر می‌شود.

محقق نیا، دهقان دهنوی و بائی (۱۳۹۸) پژوهشی با عنوان تأثیر عوامل درونی و بیرونی صنعت بانکداری بر ریسک اعتباری بانک‌ها در ایران انجام دادند. این پژوهش، تأثیر عوامل درونی و بیرونی صنعت بانکداری بر ریسک اعتباری بانک‌ها را در ایران بررسی کرده است. برای شناسایی تأثیر عوامل مستقل بر ریسک اعتباری از مدل داده‌های جدولی استفاده شده است. جامعه آماری پژوهش مشتمل بر ۳۱ بانک بوده و نتایج نشان می‌دهد که از میان متغیرهای درون بانکی، اندازه و سرمایه اثر مثبت و توسعه تأمین اعتبارات اثر منفی و از میان متغیرهای برون بانکی متغیر تمرکز، نرخ رشد نقدینگی و رشد نرخ ارز اثر مثبت و متغیر توسعه بخش بانکی و نرخ رشد اقتصادی اثر منفی بر ریسک اعتباری داشته است.

محتشمی و محتشمی (۱۳۹۶) در پژوهشی با استفاده از مدل شبکه عصبی - فازی به رتبه‌بندی اعتباری مشتریان بانک تجارت پرداختند و نتایج به‌دست‌آمده را با مدل شبکه عصبی مقایسه کردند. برای دستیابی به هدف، ۷۰۰ مشاهده از بانک تجارت استخراج ارزیابی شد. برای تحلیل داده‌ها از دو تکنیک شبکه عصبی - فازی و لجستیک استفاده شد. نتایج، گویای برتری تکنیک شبکه‌های عصبی فازی بود.

کریمی، اسدی گرجی، گیلک حکیم آبادی و اسدی (۱۳۹۵) پژوهشی در زمینه ریسک اعتباری مشتریان حقیقی روی ۲۵۴۵ پرونده مشتریان بانک تجارت شهرستان نکا طی سال‌های ۱۳۸۱ تا ۱۳۹۰ با رهیافت رگرسیون لجستیک انجام دادند. نتایج حاصل از این پژوهش نشان می‌دهد که مدت تسهیلات، نرخ تسهیلات و نوع وثیقه و نوع تسهیلات بر وصول مطالبات بانکی تأثیر معناداری دارد و متغیرهای تکلیفی یا غیرتکلیفی بودن تسهیلات و میزان تسهیلات بر احتمال نکول اثر معناداری ندارد.

حسینی و زیبایی (۱۳۹۴) پژوهشی با استفاده از مدل شبکه عصبی و با هدف شناسایی عوامل مؤثر بر ریسک اعتباری و ارائه مدلی برای پیش‌بینی ریسک اعتباری و رتبه‌بندی مشتریان حقوقی متقاضی تسهیلات اعتباری انجام دادند. نتیجه پژوهش با استفاده از ۱۷ متغیر توضیح‌دهنده نشان داد که مدل شبکه عصبی توانسته است با درصد صحت پیش‌بینی ۹۵/۵ درصدی، مشاهده‌ها را منطبق بر واقعیت برآورد کند که نشانگر توانایی بالای شبکه عصبی در پیش‌بینی ریسک اعتباری مشتریان است.

حاجی کرد، اصغرزاده زعفرانی و امام‌دوست (۱۳۹۵) با هدف تدوین مدلی به‌منظور ارزیابی ریسک اعتباری مشتریان حقوقی بانک از ماشین بردار پشتیبان و الگوریتم ژنتیک بهره‌گرفته‌اند. بدین منظور، روی متغیرهای مالی ۲۸۲ شرکت که طی سال‌های ۱۳۸۷ تا ۱۳۹۰ از بانک تجارت تسهیلات دریافت کرده‌اند، پژوهشی انجام شده است. در این پژوهش برای بهینه‌سازی ورودی‌های ماشین بردار پشتیبان از الگوریتم ژنتیک بهره‌گرفته شده است. در مدل هیبریدی، الگوریتم ژنتیک داده‌های ورودی مدل ماشین بردار پشتیبان را بهینه می‌کند. یافته‌های پژوهش نشان می‌دهد که مدل هیبریدی در مقایسه با مدل بردار پشتیبان، در شناسایی مشتریان خوش حساب و بد حساب و پیش‌بینی ریسک اعتباری مشتریان، عملکرد بهتری دارد.

عبدلی و فرد حریری (۱۳۹۴) در پژوهشی با استفاده از یک نمونه تصادفی ۳۳۰ تایی از مشتریانی که در سال‌های ۱۳۹۱ و ۱۳۹۲ از شعب بانک رفاه در سراسر کشور تسهیلات اعتباری دریافت کرده بودند و با به‌کارگیری روش رگرسیون لاجیت عوامل مؤثر بر ریسک اعتباری، این مشتریان را شناسایی کردند. نتایج این پژوهش نشان می‌دهد که متغیرهای میانگین موجودی (معدل حساب در ۶ ماه گذشته)، نسبت بازده فروش، نسبت سود خالص به فروش خالص و دارایی جاری به بدهی جاری اثر معکوس و متغیرهای تعداد چک برگشتی و نسبت مبلغ معوق به دارایی جاری اثر مستقیم بر ریسک اعتباری دارند.

قاسمی و دنیایی (۱۳۹۴) در پژوهشی ۲۹ شاخص مؤثر در ارزیابی ریسک اعتباری مشتریان حقوقی نظیر متغیرهایی چون تعداد چک برگشتی مشتری، سابقه بدهی بانکی، طول عمر حساب مشتری را در کنار شاخص‌های مالی ۶C شناسایی و طبقه‌بندی کردند و با استفاده از مدل شبکه عصبی به تعیین خوشه‌های بهینه پرداختند. در نهایت، وزن نسبی هر یک از شاخ‌های دخیل در ارزیابی ریسک اعتباری را مشخص کردند.

بهارلو، امین بیدختی و محقق نیا (۱۳۹۴) در پژوهشی عوامل مؤثر بر ریسک اعتباری مشتریان حقیقی را در ۱۴ رده مختلف شامل خوش حسابی، سابقه بدهی بانکی، جنسیت، سطح تحصیلات، سن، متوسط درآمد ماهیانه، وضعیت تأهل، شغل، وضعیت ملکیت منزل، شغل همسر، مبلغ تسهیلات مورد درخواست، نرخ بهره، ارزش وثیقه و تعداد اقساط در نظر گرفتند. نتایج هر دو مدل برازش لجستیک چندگانه و باینری تهیه‌شده، نشان‌دهنده اهمیت نسبی متغیرهای سطح تحصیلات و سن مشتری حقیقی نسبت به سایر متغیرهای مستقل بود. به نظر آنها، نتایج یکسان هر دو مدل می‌تواند

نشان دهنده تأثیر کم افزایش تعداد رده بندی متغیر وابسته (اعتبار مشتری) یا توزیع نامناسب تعداد مشتریان برای رده های مختلف اعتبار مشتری باشد.

بیک زاد و همکاران (۱۳۹۳) در پژوهشی، عوامل مؤثر و اولویت بندی معیارهای اعتبارسنجی و امتیازدهی اعتباری مشتریان بانکها را شناسایی کردند. یافته های پژوهش ایشان نشان داد که بین معیارهای امتیازدهی اعتباری بر مبنای مدل C6 تفاوت و اولویت وجود داشته و در این بین، عامل وثیقه، بالاترین اولویت را دارد و پس از آن، به ترتیب سرمایه، شخصیت، ظرفیت، شرایط کلی و شرایط و ضوابط اعطای تسهیلات در اولویت های بعدی قرار می گیرند.

پیشینه پژوهش های خارجی

خماخم و بوجلبنس^۱ (۲۰۱۸) به مقایسه مدل هایی برای ارزیابی ریسک اعتباری مشتریان حقیقی در سیستم بانک های تجاری تونس در بازه ۲۰۱۰ تا ۲۰۱۶ پرداختند. برای این منظور، از یک نمونه ۸۲۳ تایی از پرونده های وام اشخاص حقیقی استفاده شد و ۲۶ متغیر توضیح دهنده نیز به کار گرفته شد. مدل های رگرسیون لجستیک، شبکه های عصبی و رگرسیون بردار پشتیبان برای امتیازدهی اعتباری استفاده شد. نتیجه پژوهش نشان می دهد که مدل رگرسیون بردار پشتیبان در معیار مساحت زیر منحنی مشخصه عملکرد دارای بالاترین دقت است.

بخت^۲ (۲۰۱۸) به مقایسه دو مدل رگرسیون لجستیک و شبکه عصبی پایه شعاعی در نظام بانک های تجاری اردن پرداخت. دوره بررسی شده ۲۰۰۶ تا ۲۰۱۱ است و از ۴۹۲ پرونده به کمک ۲۰ متغیر توضیح دهنده در راستای مدل سازی استفاده شده است. نتیجه پژوهش نشان می دهد که دقت مدل شبکه های عصبی در حدود ۸۵ درصد است و این در حالی است که دقت رگرسیون لجستیک در حدود ۷۰ درصد است. از این رو، عملکرد شبکه های عصبی دارای دقت بالاتری است.

لی، تینگ، چا و زون فنگ^۳ (۲۰۱۸) در پژوهشی، ریسک اعتباری اشخاص حقیقی را در یک نمونه ۶۴۰ تایی از پرونده های متقاضی وام در آلمان بررسی کردند. برای محاسبه امتیاز اعتباری از سه رویکرد شامل رگرسیون معمولی، رگرسیون لجستیک و رگرسیون لجستیک بر پایه الگوریتم ژنتیک استفاده شد. نتیجه پژوهش نشان داد که رویکرد الگوریتم ژنتیک در محاسبه اوزان بهینه رگرسیون لجستیک نسبت به روش های آماری متداول دارای بیشترین قدرت تفکیک است.

1. Khemakhem and Boujelbene
2. Bekhet
3. Li, Tingting, Chao and Zong fang

عبدو^۱ (۲۰۱۸) با استفاده از برنامه‌نویسی ژنتیک، ریسک اعتباری را در یک نمونه ۱۲۶۲ تایی از پرونده‌های وام مربوط به افراد حقیقی در نظام بانکی مصر بررسی کرد. در این نمونه آماری، ۸۵۱ پرونده سالم و باقی دارای پرداخت‌های معوق بوده‌اند. برای مدل‌سازی، از رویکرد برنامه‌نویسی ژنتیک با ۲۶ متغیر استفاده شد که نتیجه پژوهش حاکی از دقت بالاتر رویکرد پژوهش نسبت به رگرسیون لجستیک و روش مجموع موزون رویدادها است.

لی و چن^۲ (۲۰۱۸) به ارزیابی ریسک اعتباری در سیستم بانک‌های تجاری تایلند با استفاده از یک نمونه ۵۱۰ تایی از پرونده‌های وام پرداختند. در این نمونه، ۴۵۹ پرونده سالم و ۵۱ پرونده ناسالم وجود دارد. روش پژوهش، ترکیب دو روش رگرسیون چندمتغیری اسپیلانی و شبکه‌های عصبی است. روش کار به این صورت است که مهم‌ترین متغیرهای تأثیرگذار توسط رویکرد رگرسیونی انتخاب می‌شوند و در ادامه، توسط شبکه‌های عصبی برای محاسبه امتیازدهی اعتباری به کار می‌روند. نتیجه پژوهش روی نمونه نشان می‌دهد که دقت روش ترکیبی از روش‌های انفرادی بیشتر است. دقت طبقه‌بندی مشتریان سالم در حدود ۸۷ درصد و مشتریان ناسالم برابر ۸۱ درصد است.

رگرسیون لجستیک از خواص مدل‌های خطی بهره می‌برد و رگرسیون سمبلیک توانایی خوبی در مدل‌سازی غیرخطی دارد. مدل‌های ترکیبی با هدف استفاده از قوت مدل‌های انفرادی و به‌منظور رسیدن به مدل‌های کارآمدتر و با دقت بیشتر (نسبت به مدل‌های انفرادی) به وجود آمده‌اند. از این رو، پژوهش حاضر به دنبال ارزیابی مدل ترکیبی رگرسیون لجستیک - سمبلیک در ارزیابی ریسک اعتباری بانک‌هاست. علاوه بر مدل ترکیبی، مدل ریسک اعتباری در خصوص مورد مطالعه پژوهش، یعنی بانک سپه برای نخستین بار در پژوهش حاضر انجام می‌شود. اضافه بر اینکه مدل ترکیبی روی داده‌های تست نیز ارزیابی می‌کند تا قابلیت‌های آن در موقعیت‌های واقعی سنجش شود.

روش‌شناسی و تبیین مدل پژوهش

پژوهش حاضر از نوع توصیفی - تحلیلی و از نظر هدف یک پژوهش کاربردی است. جامعه آماری پژوهش را کلیه مشتریان حقیقی شعب بانک سپه شهرستان شیراز تشکیل می‌دهد و مجموع تمام داده‌هایی که بر اساس یک فرایند اداری و بخش نامه‌ای در اختیار پژوهشگر قرار گرفت، شامل ۳۵۱ پرونده از ۱۶ شعبه بانک سپه شیراز است که به‌عنوان نمونه آماری استفاده شد. متغیرهای مستقل

1. Abdou

2. Lee & Chen

استفاده شده در الگو که بر اساس مطالعات کتابخانه در ادبیات مربوطه و نیز استفاده از نظر خبرگان این حوزه انتخاب شده‌اند، عبارت‌اند از: میزان، مدت و نرخ بهره تسهیلات، نوع وثیقه، متوسط درآمد ماهیانه، سن، تعداد چک برگشتی مشتری، سابقه بدهی بانکی، طول عمر حساب مشتری، نوع تسهیلات اعطایی و تکلیفی یا غیر تکلیفی بودن تسهیلات، خوش‌حسابی، سطح تحصیلات، شغل، وضعیت ملکیت منزل، وضعیت تأهل و وضعیت اشتغال همسر. برای تبیین مدل بهتر است به‌طور خلاصه مدل‌های استفاده شده در پژوهش حاضر را تشریح کنیم.

رگرسیون لجستیک

رگرسیون لجستیک یک مدل آماری رگرسیون برای متغیرهای دودویی مانند بیماری یا سلامت، خوش‌حسابی و بدحسابی و... است. این مدل را می‌توان به‌عنوان مدل خطی تعمیم یافته‌ای که از تابع لجوجیت به‌عنوان تابع پیوند استفاده می‌کند و خطایش از توزیع چندجمله‌ای پیروی می‌کند، به حساب آورد. مدل رگرسیون لجستیک، بر اساس فرض‌های کاملاً متفاوتی (درباره رابطه متغیرهای وابسته و مستقل) از رگرسیون خطی است. تفاوت مهم این دو مدل می‌تواند در دو ویژگی رگرسیون لجستیک دیده شود. نخست، توزیع شرطی $y | x$ یک توزیع برنولی به‌جای یک توزیع گوسی است (زیرا متغیر وابسته دودویی است). دوم مقادیر پیش‌بینی احتمالاتی است و محدود بین صفر و ۱ و به‌کمک تابع توزیع لجستیک به دست می‌آید. در رگرسیون لجستیک رابطه بین متغیرهای مستقل و متغیر وابسته که دارای دو مقدار صفر و ۱ است، به‌صورت زیر تعریف می‌شود:

$$p = p(y = 1|x) = \frac{1}{1 + e^{-\alpha - \beta x}} \quad \text{رابطه (۱)}$$

در این صورت به‌راحتی می‌توان دید که:

$$\ln\left(\frac{p}{1-p}\right) = \ln\left(\frac{1}{e^{-\alpha - \beta x}}\right) = \alpha + \beta x \quad \text{رابطه (۲)}$$

برای برآورد پارامترهای α و β از روش برآورد حداکثر درست‌نمایی استفاده می‌شود. خروجی رگرسیون لجستیک بین صفر و ۱ است و به‌صورت متداول مقدار ۰/۵ به‌عنوان نقطه برش^۱ تلقی می‌شود. بنابراین، در صورتی که مقدار حاصل از رگرسیون از نیم بزرگ‌تر یا کوچک‌تر باشد، کلاس داده ورودی متفاوت خواهد بود. در حالت کلی، مقدار ۰/۵ نقطه برش کارا نیست و تعیین این نقطه به

1. Cutoff point

بهینه‌سازی نیاز دارد. ملاک مناسب دیگر برای سنجش کیفیت و کارایی مدل‌های برازش‌شده و تعیین توان پیش‌بینی آنها استفاده از مساحت زیر منحنی مشخصه عملکرد (ROC)^۱ است. این منحنی عبارت است از حساسیت برحسب یک ویژگی و مساحت زیر آن که عددی است بین صفر و ۱، به‌عنوان ملاکی برای سنجش توانایی پیش‌بینی مدل‌ها استفاده می‌شود. هرچه این عدد نزدیک به ۱ باشد، نشان‌دهنده توانایی و کارایی بیشتر مدل پیش‌بینی است. از این رو، از ابزارهای پرکاربرد در مدل‌سازی رتبه‌بندی اعتباری، مدل آماری رگرسیون لجستیک است (خماخم و بوجلینس، ۲۰۱۸). این روش، نوعی مقایسه تصویری است و می‌توان به‌راحتی این منحنی را در نرم‌افزار اس‌پی‌اس‌اس ترسیم کرد.

مدل شبکه عصبی مصنوعی

یکی دیگر از ابزارهای پرکاربرد در مدل‌سازی رتبه‌بندی اعتباری، شبکه‌های عصبی است. شبکه‌های عصبی مفهومی الهام‌گرفته از مغز انسان است که از حدود ۱۰۰ بیلیون واحد بسیار ریز به نام نورون ساخته شده است. نرون‌ها با هم در ارتباط هستند و این ارتباط از طریق سیگنال‌های الکتروشیمیایی یا کانال یونی انجام می‌شود. سیگنال‌هایی که به یک نورون می‌رسند از طریق اتصالاتی که سیناپس نام دارد، دریافت می‌شوند. این اتصالات در انتهای هسته سلول عصبی که منشعب می‌شود، دندریت نام دارد. نورون به‌طور پیوسته از این ورودی‌ها سیگنال می‌گیرد. هر نرون ورودی‌های دریافتی خود را با سیستم مخصوص به خود ترکیب می‌کند و چنانچه نتیجه نهایی از یک مقدار آستانه بیشتر شود، نورون برانگیخته می‌شود و یک ولتاژ ایجاد می‌کند و سیگنالی را در امتداد جسمی که آکسون نام دارد، می‌فرستد (کاسابوو^۲، ۲۰۱۰). در این شبکه‌ها اگر یک سلول آسیب ببیند، بقیه سلول‌ها می‌توانند نبود آن را جبران کرده و نیز در بازسازی آن سهیم باشند. این شبکه‌ها قادر به یادگیری‌اند. برای مثال، با اعمال سوزش به سلول‌های عصبی لامسه، سلول‌ها یاد می‌گیرند که به‌طرف جسم داغ نروند و با این الگوریتم سیستم می‌آموزد که خطای خود را اصلاح کند. یادگیری در این سیستم‌ها به‌صورت تطبیقی انجام می‌شود، یعنی با استفاده از مثال‌ها وزن سیناپس‌ها به‌گونه‌ای تغییر می‌کند که در صورت دادن ورودی‌های جدید، سیستم پاسخ درستی تولید کند. فلسفه اصلی شبکه عصبی مصنوعی، مدل کردن ویژگی‌های پردازشی مغز انسان برای تقریب زدن روش‌های معمول محاسباتی با روش پردازش زیستی است. به بیان دیگر، شبکه عصبی مصنوعی روشی است که دانش ارتباط بین چند

1 . Receiver Operating Characteristic (ROC)

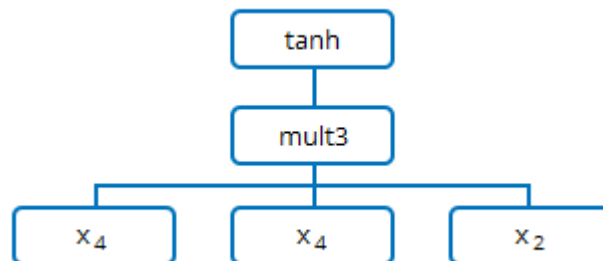
2. Kasabov

مجموعه داده را از طریق آموزش فراگرفته و برای استفاده در موارد مشابه ذخیره می کند (فقیهی نژاد و مینایی، ۱۳۹۷).

رگرسیون سمبلیک (برنامه ریزی ژنتیک)

برنامه ریزی ژنتیک، یکی از شاخه های مهم هوش مصنوعی است که از ترکیب برنامه نویسی کامپیوتری با ایده های الگوریتم ژنتیک برای حل مسائل بهره می برد. در برنامه ریزی ژنتیک برای حل یک مسئله در ابتدا تعدادی جواب در نظر گرفته می شود. این جواب ها، به طور لزوم بهترین جواب یا راه حل بهینه مسئله نیستند و اغلب، به صورت تصادفی تولید می شوند. سپس، این جواب ها که همگی به اصطلاح یک نسل را تشکیل می دهند، بر اساس میزان شایستگی خود در حل مسئله، به رقابت و ترکیب شدن با یکدیگر می پردازند تا نسل بعد جواب ها که ممکن است دارای جواب بهتر یا کارآمدتری باشند، تشکیل شود (کوزا^۱، ۱۹۹۲).

در مسئله کلی رگرسیون خطی، برای توضیح یک متغیر وابسته از ترکیب خطی تعدادی متغیر مستقل استفاده می شود. رگرسیون سمبلیک نوعی از رگرسیون است که بر اساس برنامه نویسی ژنتیک به ترکیب خطی چند تابع غیرخطی از متغیرهای مستقل می پردازد. الگوریتم ژنتیک در تکرارهای خود، ساختارهای توابع غیرخطی یادشده را هدایت می کند و ضرایب خطی حاصل از ترکیب این توابع از کمینه سازی مجموع مربعات خطا محاسبه می شود. در یک رگرسیون سمبلیک تعدادی تابع پایه در نظر گرفته می شود که هر تابع غیرخطی از ترکیب آنها به وجود می آید. برای نمونه، این توابع می تواند عملگرهای محاسباتی معمولی مانند جمع، تفریق، ضرب، تقسیم یا توابع نمایی و لگاریتمی یا توابع مثلثاتی سینوس و کسینوس، تانژانت هایپربولیک و... باشند. نخستین مفهوم مهم در رگرسیون سمبلیک با برنامه ریزی ژنتیک، مفهوم ژن است. ساختار یک ژن ضمن معرفی ویژگی های درخت نمایش دهنده آن مشخص می شود. در اینجا هر درخت از تعدادی توابع ریاضی مانند جمع و تفریق و سینوس و... و d متغیر حرفی $x_1, x_2, x_3, \dots, x_d$ تشکیل شده است. یکی دیگر از ویژگی های درخت نمایش عمق درخت است که تعداد سطوح درخت را نشان می دهد. گره های پایانی درخت را ترمینال می نامند که همواره متغیر حرفی است و نمی تواند از جنس تابع باشد. برای نمونه، نمونه ژن های زیر درخت های از عمق چهار است که هر یک، یک تابع را محاسبه می کنند.



$$gene = \tanh(x_2 \times x_4 \times x_4)$$

شکل ۱. نمایش درختی یک ژن در برنامه‌نویسی ژنتیک

در الگوریتم رگرسیون سمبلیک، در ابتدا یک جمعیت تصادفی از کروموزم‌ها شکل می‌گیرد. هر کروموزم از تعدادی ژن تشکیل شده است. در پژوهش حاضر، طول هر کروموزم برابر ۴ اختیار شده که ساختار کروموزم آن مطابق رگرسیون معمولی از یک ساختار خطی برخوردار است. تعداد کروموزم در هر نسل یا جمعیت یکی از پارامترهای قابل تنظیم این الگوریتم است. بنابراین، فرم نمایش یک کروموزم m ژنی خطی به صورت زیر است:

$$crom = C_0 + C_1 Gene1 + C_2 Gene2 + \dots + C_m Gene_m \quad (\text{رابطه ۳})$$

بعد از شکل‌گیری یک جمعیت تصادفی ابتدایی از کروموزم‌ها، برای هر کروموزم مقادیر C_0, C_1, \dots, C_m به روش مینیمم‌سازی حداقل مربعات معمولی به کمک داده‌های آموزشی محاسبه می‌شود. پس از محاسبه ضرایب، معکوس مجموع مربعات خطا به‌عنوان تابع برازندگی انتخاب می‌شود که مقادیر بزرگ‌تر آن مطلوب است. بر این اساس، شایستگی کروموزمی برای حضور در نسل بعد بیشتر است که دارای برازندگی $Fitness = \frac{1}{SSE}$ بالاتری باشد.

بعد از محاسبه تابع برازندگی، کروموزم‌ها وارد مرحله انتخاب می‌شوند تا کروموزم‌های کاندید برای نسل بعد به دست آیند. سازوکار انتخاب کروموزم در پژوهش حاضر، روش انتخاب تورنمنت^۱ است. در این روش، یک عدد به نام اندازه مسابقه مشخص می‌شود و سپس به آن اندازه، کروموزم به صورت تصادفی انتخاب شده و کروموزم‌ها در مقدار برازندگی با هم رقابت می‌کنند و بهترین کروموزم کاندید ورود به نسل بعد می‌شود. پس از مرحله انتخاب، الگوریتم وارد مرحله تقاطع و جهش

می‌شود تا نسل بعد ساخته شوند. عمل تقاطع روی تعدادی کروموزم انجام خواهد شد که درصد تقاطع که از پارامترهای قابل تنظیم مدل است، مقدار آن را مشخص خواهد کرد. در مرحله تقاطع، دو کروموزم به‌عنوان والدین با هم ترکیب شده و دو فرزند تولید می‌کنند. همان‌طور که در طبیعت جهش‌های ژنتیکی با احتمال کم اتفاق می‌افتد در این الگوریتم نیز این موضوع شبیه‌سازی شده است. برای این منظور، درصدی از کروموزم‌های انتخاب‌شده تحت عملگر جهش قرار می‌گیرند. درصد جهش از پارامترهای قابل تنظیم مدل است. پس از مشخص شدن نسل جدید، الگوریتم دوباره وارد مرحله محاسبه مقدار برازندگی می‌شود و مراحل گفته‌شده ادامه می‌یابند. معیار اتمام الگوریتم می‌تواند تعداد تکرار مشخصی باشد که الگوریتم با رسیدن به آن تعداد خاتمه یابد. معیار دیگر زمانی است که تابع برازندگی از مقدار معینی بیشتر شود و الگوریتم با رسیدن به آن مقدار پایان می‌یابد. در پایان الگوریتم بهترین کروموزم به‌عنوان مدل رگرسیون سمبلیک پیش‌بینی‌کننده انتخاب می‌شود. به‌طور کلی، الگوریتم‌های ژنتیک یک روش جست‌وجوی مؤثر در فضاهای بسیار وسیع و بزرگ است که در نهایت، به جهت‌گیری به‌سمت پیدا کردن یک جواب منجر می‌شود. الگوریتم‌های ژنتیک با یک سری متغیرهای کدشده کار می‌کنند. مزیت کار با متغیرهای کدشده در این است که اصولاً کدها قابلیت تبدیل فضای پیوسته به فضای گسسته را دارند (شیتا و همکاران^۱، ۲۰۱۵).

با معرفی اجمالی رگرسیون لجستیک، شبکه‌های عصبی و رگرسیون سمبلیک مدل ترکیبی پژوهش به‌صورت زیر عمل می‌کند. گام نخست: به کمک رگرسیون لجستیک متغیرهای کاندید در زمینه پیش‌بینی ریسک اعتباری با رویکرد جست‌وجوی رو به جلو والد^۲ پالایش شده و فهرستی از متغیرهای مهم تأثیرگذار بر حسب عملکرد رگرسیون لجستیک مشخص می‌شود. بنابراین، در صورتی که متغیرهای اولیه برابر X_1, X_2, \dots, X_p باشند، خروجی این مرحله متغیرهای مؤثر $X_1^*, X_2^*, \dots, X_k^*$ است.

در گام دوم، متغیرهای اثرگذار که از مرحله قبل تعیین شده‌اند، به‌عنوان ورودی شبکه‌های عصبی مصنوعی با خروجی خوش‌حسابی و بدحسابی استفاده می‌شوند. شبکه‌های عصبی مصنوعی به‌عنوان مدلی غیرخطی، ارتباط بین ورودی‌ها و خروجی‌ها را در داده‌های آموزشی برقرار می‌کند. برای محاسبه نقطه بهینه برش در رگرسیون سمبلیک نیز از منحنی مشخصه عملکرد استفاده می‌شود، سپس، دقت مدل طراحی‌شده روی داده‌های تست ارزیابی می‌شود.

1. Sheta et al
2. Walh-forward method

در گام سوم، از متغیرهای اثرگذار خروجی رگرسیون لجستیک به‌عنوان ورودی رگرسیون سمبلیک با خروجی خوش‌حسابی و بدحسابی استفاده می‌شود. رگرسیون سمبلیک با کارایی غیرخطی بالا ارتباط بین ورودی‌ها و خروجی‌ها را در داده‌های آموزشی برقرار می‌کند. برای محاسبه نقطه بهینه برش در رگرسیون سمبلیک نیز از منحنی مشخصه عملکرد استفاده می‌شود. سپس، دقت مدل طراحی شده روی داده‌های تست ارزیابی می‌شود.

یافته‌های پژوهش

فهرست ۱۷ متغیر استفاده‌شده پژوهش در راستای پیش‌بینی خوش‌حسابی یا بدحسابی اشخاص حقیقی متقاضی تسهیلات بانکی به‌صورت جدول ۱ است.

جدول ۱. معرفی متغیرهای پژوهش

نماد	عنوان
X۱	میزان وام (بر حسب ریال)
X۲	مهلت بازپرداخت (بر حسب ماه)
X۳	نرخ بهره سالیانه تسهیلات اعطاشده
X۴	درآمد ماهیانه (بر حسب ریال)
X۵	سن (بر حسب سال)
X۶	تعداد چک‌های برگشتی
X۷	سابقه بدهی بانکی: عدد ۱ برای نداشتن سابقه بدهی و عدد صفر برای سابقه‌داران بدهی بانکی
X۸	طول عمر حساب (بر حسب ماه)
X۹	نوع وثیقه: عدد ۱ برای سپرده و رهن و عدد صفر برای چک و سفته (اسناد تجاری)
X۱۰	سطح تحصیلات: عدد ۱ برای تحصیلات دانشگاهی و عدد صفر برای تحصیلات زیر دانشگاه
X۱۱	جنسیت: عدد ۱ برای مردان و عدد صفر برای خانم‌ها
X۱۲	وضعیت اشتغال همسر: عدد ۱ برای صاحبان شغل و عدد صفر برای خانه‌داران
X۱۳	وضعیت تأهل: عدد ۱ برای متأهلان و عدد صفر برای مجردان
X۱۴	وضعیت ملکی: عدد ۱ برای مالکان و عدد صفر برای مستأجران
X۱۵	نوع شغل: عدد ۱ برای شغل‌های دولتی و عدد صفر برای مشاغل آزاد
X۱۶	نوع وام: عدد ۱ برای تسهیلات مرابحه و عدد صفر برای قرض‌الحسنه
X۱۷	تکلیفی یا غیرتکلیفی بودن: عدد ۱ برای غیرتکلیفی و عدد صفر برای تکلیفی بودن وام

متغیر وابسته نیز متغیر مجازی Y است که $Y = 1$ برای مشتریان خوش حساب و $Y = 0$ برای مشتریان بدحساب استفاده خواهد شد. افراد بدحساب، افراد حقیقی‌ای هستند که در بازپرداخت تسهیلات اعطایی بانک با مشکل مواجه شده‌اند و دارای ۶ و بیشتر قسط معوقه هستند. تعداد کل پرونده‌های استخراج شده برابر ۳۵۱ پرونده در بازه ۱۳۹۰ تا ۱۳۹۶ است که از این تعداد ۳۳۰ پرونده مرتبط با اشخاص خوش حساب و ۲۱ پرونده مرتبط با بدحسابان است. بنابراین، نزدیک به ۰/۹۴ از آمار مشتریان نمونه را خوش حساب‌ها و ۰/۰۶ باقی مانده را افراد بدحساب تشکیل داده است. آمار توصیفی مربوط به متغیرهای پژوهش در جدول ۲ ارائه شده است.

جدول ۲. آمار توصیفی متغیرهای پژوهش

شاخص آماری	طول عمر حساب	تعداد چک برگشتی	سن	درآمد ماهیانه	نرخ بهره	مدت بازپرداخت	مبلغ تسهیلات
میانگین	۳/۱۶	۰/۸	۴۱/۳	۲۵،۰۰۰،۰۰۰	۱۵/۶۴۹۶	۴۲/۶۴	۲۴۱،۰۰۰،۰۰۰
میانه	۲	۰	۳۹	۲۱،۰۰۰،۰۰۰	۱۸	۳۶	۱۷۰،۰۰۰،۰۰۰
بیشینه	۱۷	۲۱	۷۵	۲۰۰،۰۰۰،۰۰۰	۳۰	۶۰	۳۲۰،۰۰۰،۰۰۰
کمینه	۰	۰	۲۱	۰	۴	۱۲	۱۲۰،۰۰۰،۰۰۰
انحراف معیار	۳/۰۲	۱/۱۵	۱۱/۱	۱۸،۰۰۰،۰۰۰	۵/۳۴۵۲	۳۱/۷۲	۲۷۷،۰۰۰،۰۰۰
چولگی	۱/۷۸	۱۷/۲۸	-/۴۷	۴/۹۱	۱/۷۲۲۵	۱۲/۴۹	۶/۰۸
کشیدگی	۷/۰۲	۳۱۰/۸۶	۲/۴۳	۳۹/۲۸۱۲	۱۷/۱۳	۲۰۳/۱۸	۵۵/۵۶
آماره چارک - برا	۴۲۳/۰۳	۱۴۰۳۶۲۳	۱۷/۷۸	۲۰۶۶۲/۷	۳۰۹۴/۳۸	۵۹۵۲۱۸	۴۲۵۸۳/۳
مقدار احتمال	۰	۰	۰	۰	۰	۰	۰

مقدار آماره چارک برا برای تمامی متغیرهای جدول ۲ حاکی از این است که توزیع هیچ یک از متغیرها در سطح اطمینان ۰/۹۵ از توزیع نرمال پیروی نمی‌کند. همچنین، توصیف آماری متغیرهای مجازی از بایت میانگین در جدول ۳ ارائه شده است.

جدول ۳. میانگین متغیرهای مجازی

میانگین	توضیح
۰/۹۲۳	نداشتن سابقه بدهی
۰/۴۶۱	وثیقه سپرده و رهن
۰/۶۰۶	تحصیلات دانشگاهی
۰/۶۶۶	جنسیت مرد
۰/۴۱۸	همسران شاغل
۰/۷۷۲	متاهل
۰/۵۷۵	مالکان مسکونی
۰/۲۸۷	مشاغل دولتی
۱	تکیفی بودن
۰/۹۸	تسهیلات مرابحه
۰/۹۴۰۱۷۱	خوش حسابان

با توجه به اینکه تمام تسهیلات ارائه‌شده غیرتکلیفی هستند، متغیر مجازی تکلیفی بودن از جمع متغیرهای پژوهش خارج شد. ۳۵۱ پرونده نمونه به دو دسته داده‌های آموزشی و داده‌های تست تقسیم شدند. ۱۵ پرونده سالم (از بین ۳۳۰ پرونده) و ۱۵ پرونده بدحساب (از بین ۲۱ پرونده و معادل ۰/۲۵ داده‌ها) به صورت تصادفی به عنوان داده‌های آموزشی انتخاب شدند. در واقع از آنجا که تعداد مشتریان بدحساب در مقایسه با تعداد مشتریان خوش حساب به مراتب کمتر است، برای افزایش قدرت مدل در پیش‌بینی افراد بدحساب (درجه تشخیص مدل) دو نمونه مساوی در نظر گرفته شدند. در ادامه، به کمک رویکرد انتخاب رو به جلوی والد^۱ در نرم‌افزار اس‌پی‌اس‌اس، متغیرهای X۴: متوسط درآمد ماهیانه، X۶: تعداد چک‌های برگشتی؛ X۷: سابقه بدهی بانکی؛ X۸: طول عمر حساب و X۹: نوع وثیقه برای ورود به رگرسیون لجستیک انتخاب شدند که تقریب رگرسیون در جدول ۴ ارائه شده است.

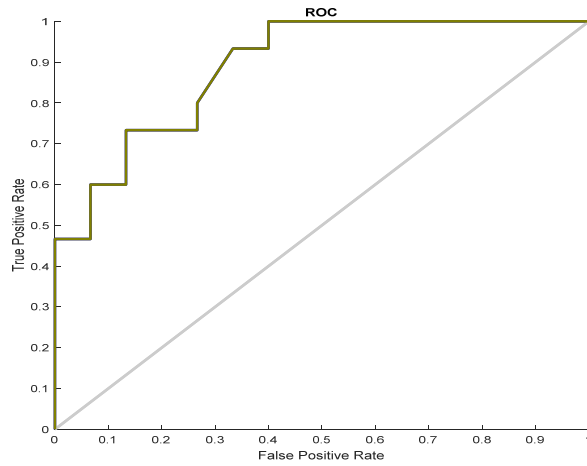
جدول ۴. تقریب مدل رگرسیون لجستیک

متغیر	مقدار احتمال	درجه آزادی	آماره والد	خطای استاندارد	مقدار	e^B
درآمد ماهانه	۰/۰۰۸	۱	۷/۰۵۹	۰/۰۰۰۱	۰/۰۰۱	۱/۰۰۰
تعداد چک برگشتی	۰/۰۴۴	۱	۰/۵۸۶	۱/۰۴۷	-۰/۸۰۱	-۰/۴۴۹
نداشتن سابقه بدهی	۰/۰۰۲	۱	۹/۶۵۸	۰/۷۲۲	۲/۲۴۴	۹/۴۳۵
طول عمر حساب مشتری	۰/۰۱۰	۱	۶/۵۷۷	۰/۱۴۹	۰/۳۸۲	۱/۴۶۵
نوع وثیقه	۰/۰۱۳	۱	۶/۱۴۰	۰/۷۱۱	۱/۷۶۳	۵/۸۲۸
C	۰/۸۱۹	۱	۰/۰۵۳	۰/۷۵۶	۰/۱۷۳	۱/۱۸۹

ستون آخر مقدار افزایش یا کاهش احتمال را به ازای یک واحد افزایش در متغیر مستقل نشان می‌دهد. بنابراین مدل رگرسیونی لجستیک به صورت زیر است:

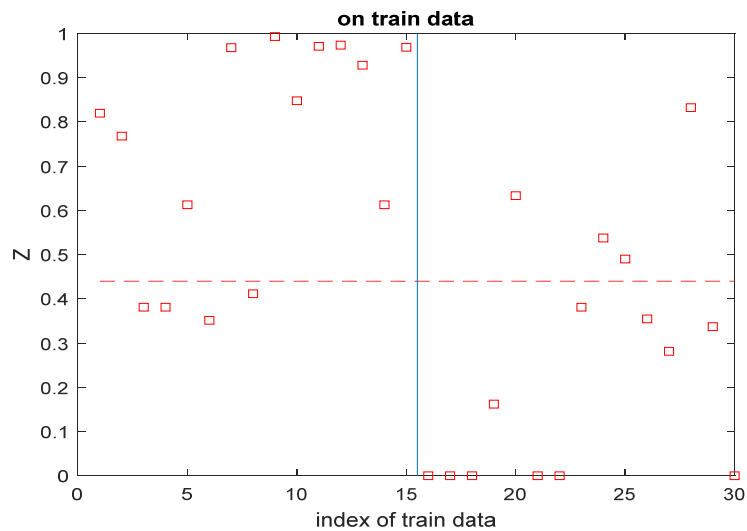
$$Z = \frac{1}{1 + \exp(-(0.173 + 0.001x_4 - 0.80x_6 + 2.24x_7 + 0.382x_8 + 1.76x_9))}$$

بر اساس اطلاعات جدول ۴ تمام متغیرهای موجود در مدل دارای اثر معناداری بر متغیر وابسته هستند. در این بین، اثر تعداد چک‌های برگشتی منفی و معنادار است و این بدان معناست که افزایش تعداد چک‌های برگشتی باعث کاهش متغیر وابسته و افزایش احتمال قرار گرفتن آن در دسته بدحسابان می‌شود. سابقه بدهی بانکی با وجود انتظار دارای اثر مثبت و معناداری است. به صورت متداول برای رگرسیون لجستیک نقطه برش برابر ۰/۵ در نظر گرفته می‌شود. در پژوهش حاضر برای این منظور از منحنی عملکرد مشخصه یا ROC مطابق نمودار شکل ۱ استفاده شد و برش بهینه متناظر با نقطه‌ای انتخاب شد که به نقطه ایدئال (۰,۱) نزدیک‌تر باشد. در واقع، نقطه (۰,۰) دارای بیشترین قدرت تشخیص (شناسایی افرا بدحساب) و کمترین درجه حساسیت (شناسایی خوش حساب‌ها) است و نقطه (۱,۰) عکس این مطلب است.



شکل ۱. منحنی ROC برای محاسبه نقطه بهینه برش

بر اساس منحنی ROC، نقطه برش بهینه برابر $0/44$ در داده‌های آموزشی محاسبه شد. عملکرد مدل لجستیک در طبقه‌بندی داده‌های آموزشی مطابق نمودار شکل ۲ است. ۱۵ داده اول خوش حساب و ۱۵ داده دوم مشتریان بدحساب است.



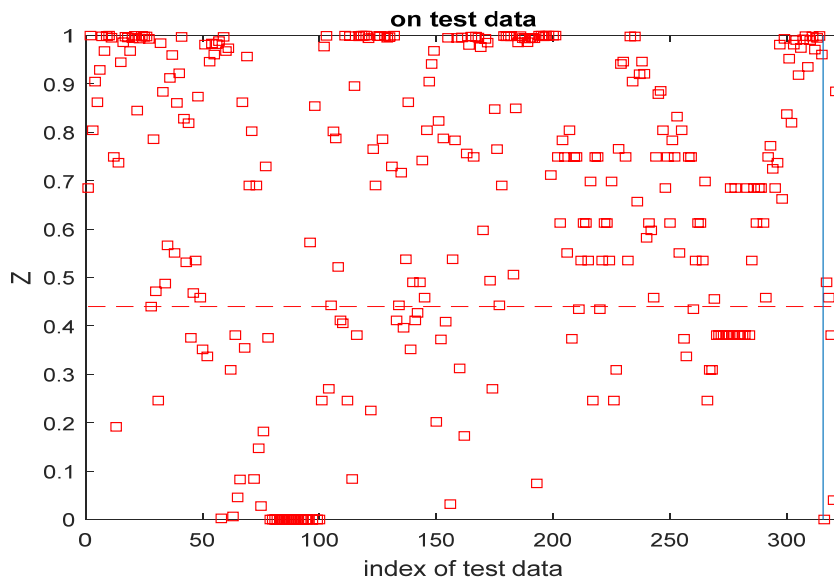
شکل ۲. عملکرد مدل لجستیک در داده‌های آموزشی

بر اساس شکل ۲:

الف. احتمال طبقه‌بندی درست یک مشتری خوش حساب 0.73 است.

ب. احتمال طبقه‌بندی درست یک مشتری بدحساب برابر 0.73 است.

در ادامه، عملکرد مدل رگرسیونی روی ۳۱۵ داده تست از پرونده‌های سالم (۳۲۱ داده اول در نمودار شکل ۲) و ۶ داده از پرونده‌های بدحساب (۶ داده پایانی نمودار شکل ۳) ارائه شده است.



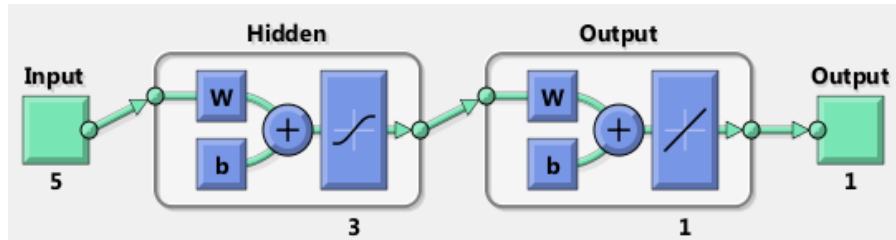
شکل ۳. عملکرد مدل لجستیک در داده‌های تست

بر اساس نتایج به‌دست‌آمده از نمودار شکل ۳:

الف. احتمال طبقه‌بندی درست یک مشتری خوش حساب برابر 0.73 است.

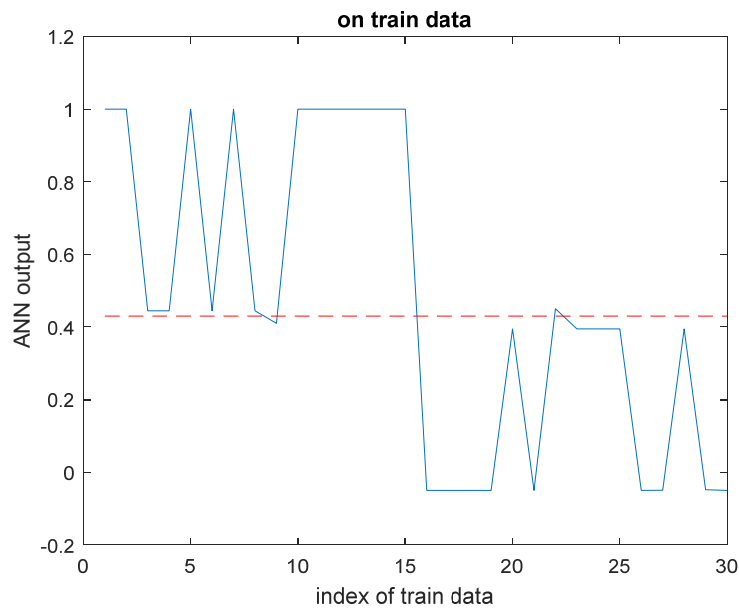
ب. احتمال طبقه‌بندی درست یک مشتری بدحساب برابر 0.50 است.

همان‌طور که مشاهده می‌شود، رگرسیون لجستیک دارای عملکرد مناسبی در داده‌های تست نیست. در ادامه، برای طبقه‌بندی مشتریان، از یک شبکه عصبی با سه نرون در لایه پنهان استفاده خواهد شد که ورودی‌های آن ۵ متغیر انتخاب‌شده در عمل‌گزینش رو به جلوی والد در رگرسیون لجستیک است. نمای شبکه عصبی استفاده‌شده در شکل ۴ نشان داده شده است.



شکل ۴. شمای شبکه عصبی پژوهش

به کمک کدنویسی در نرم‌افزار متلب شبکه عصبی به کمک روش پس انتشار خطا روی همان مجموعه داده‌های آموزشی مدل لجستیک آموزش شد. عملکرد شبکه عصبی در طبقه‌بندی مشتریان در داده‌های آموزشی در نمودار شکل ۵ ارائه شده است. در این نمودار، مانند گذشته ۱۵ داده اول خوش حساب و ۱۵ داده بعدی بدحساب هستند و نقطه برش به کمک منحنی عملکرد مشخصه برابر ۰/۴۳ محاسبه شد.



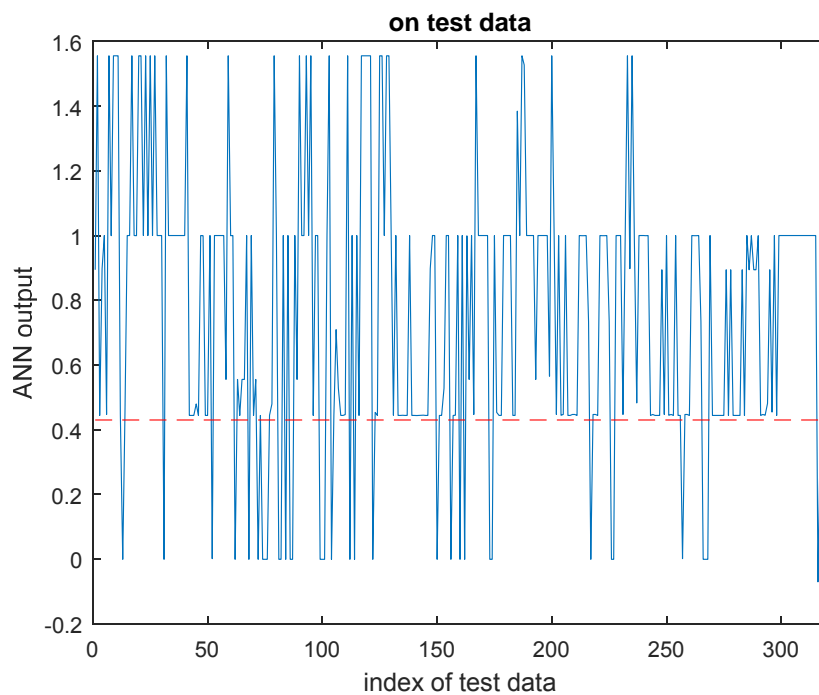
شکل ۵. عملکرد شبکه عصبی روی داده‌های آموزشی

بر اساس شکل ۵:

الف. احتمال طبقه‌بندی درست یک مشتری خوش حساب برابر $0/933$ است.

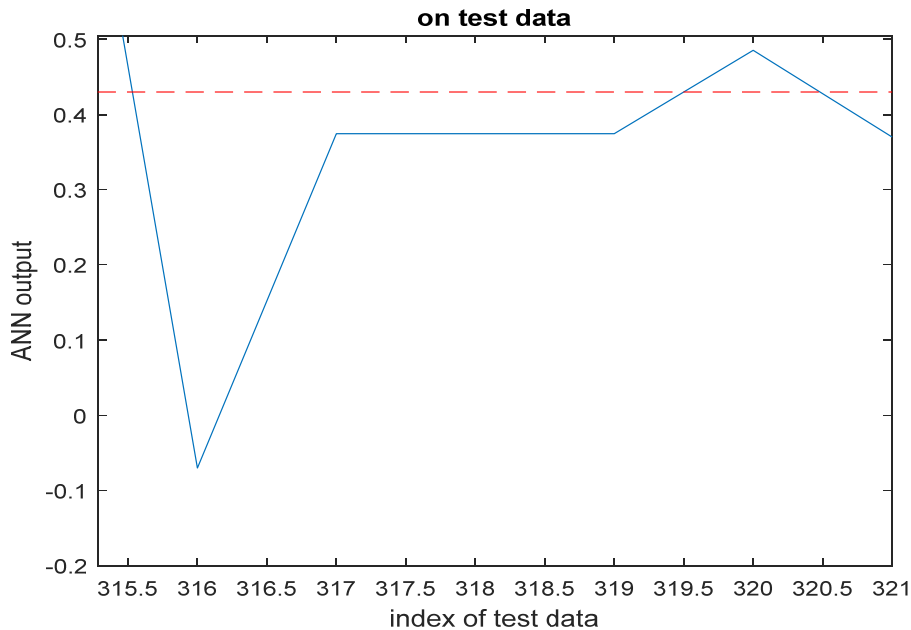
ب. احتمال طبقه‌بندی درست یک مشتری بد حساب برابر $0/933$ است.

نتایج حاصل از شبکه عصبی در داده‌های آموزشی در مقایسه با رویکرد لجستیک کاملاً بهتر است. حال باید داده‌های تست نیز بررسی شوند تا دقت مدل در شرایط واقعی مشخص شود. مانند گذشته عملکرد مدل رگرسیونی روی ۳۱۵ داده تست از پرونده‌های سالم (۳۲۱ داده اول نمودار شکل ۶) و ۶ داده از پرونده‌های بد حساب (۶ داده پایانی نمودار شکل ۶) ارائه شده است.



شکل ۶. عملکرد شبکه عصبی روی داده‌های تست

عملکرد شبکه عصبی در طبقه‌بندی مشتریان بد حساب در ۶ داده آخر نمودار شکل ۶ موجود است که برای مشاهده بهتر به صورت بزرگ‌نمایی در نمودار شکل ۷ ارائه شده است:



شکل ۷. عملکرد شبکه عصبی روی داده‌های تست بدحساب

بر اساس نتایج به‌دست آمده از شبکه عصبی روی داده‌های تست:

الف. احتمال طبقه‌بندی درست یک مشتری خوش حساب برابر $0/۸۹$ است.

ب. احتمال طبقه‌بندی درست یک مشتری بدحساب برابر $0/۸۳۳$ است.

بنابراین، عملکرد مدل ترکیبی شبکه عصبی - رگرسیون لجستیک توانسته در مقایسه با مدل لجستیک، نتایج بهتری ارائه کند و هر دو احتمال را افزایش دهد که این مطلب معادل کاهش هم‌زمان خطای نوع اول (طبقه‌بندی اشتباه یک مشتری خوش حساب) و خطای نوع دوم (طبقه‌بندی اشتباه یک مشتری بدحساب) نسبت به مدل متداول لجستیک است.

همان‌طور که مشاهده شد، رویکرد شبکه‌های عصبی مصنوعی دقت پیش‌بینی بالایی داشته و در مقایسه با رگرسیون لجستیک در این زمینه بسیار بهتر عمل می‌کند. به‌منظور بررسی بیشتر و مقایسه رویکردهای مختلف، در ادامه از یک رگرسیون سمبلیک با مشخصات جدول ۵ برای طبقه‌بندی مشتریان استفاده خواهد شد. ورودی‌های سیستم ۵ متغیر انتخاب‌شده در عمل گزینش رو به جلوی والد در رگرسیون لجستیک است.

جدول ۵. مشخصات ساختاری رگرسیون سمبلیک

پارامتر	مقدار
تعداد ژن‌ها در هر کروموزم	۴
عمق درخت	۴
تعداد ورودی	۴
تعداد جمعیت	۲۵۰
حداکثر تکرار	۱۵۰
اندازه تورنمنت	۵۰
درصد نخبه‌گرایی	۰/۷
نرخ تقاطع	۰/۸۴
نرخ جهش	۰/۱۴
نوع پیچیدگی	عبارتی
توابع استفاده‌شده	TIMES MINUS PLUS TANH MULT3 ADD3

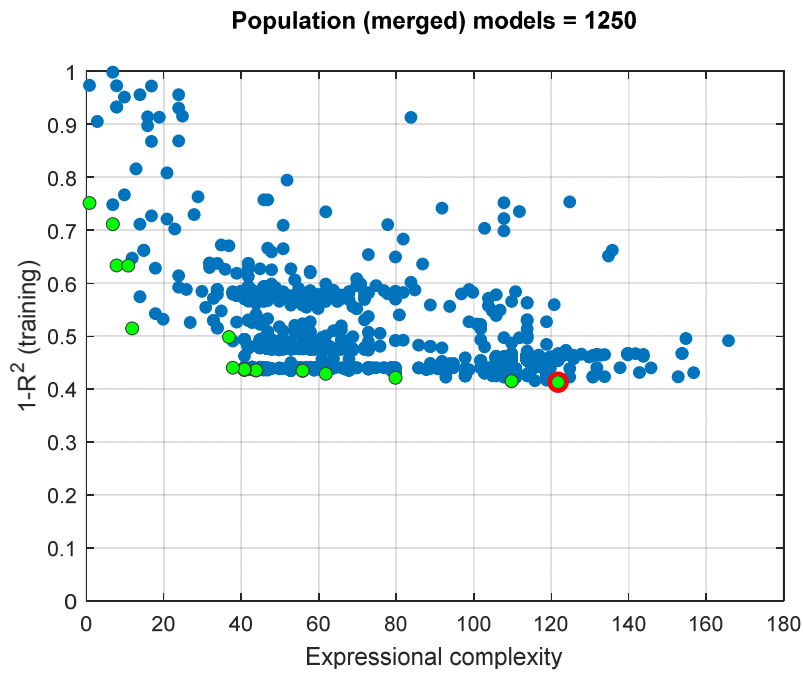
که توابع استفاده‌شده به صورت زیر روی متغیرهای مستقل عمل می‌کنند:

$$add: (x, y) \rightarrow x + y \quad times: (x, y) \rightarrow x \times y$$

$$minus: (x, y) \rightarrow x - y \quad mult3: (x, y, z) \rightarrow xyz$$

$$add3: (x, y, z) \rightarrow x + y + z \quad tanh: x \rightarrow \tanh(x)$$

به کمک کدنویسی در نرم‌افزار متلب رگرسیون سمبلیک روی همان مجموعه داده‌های آموزشی مدل لجستیک آموزش شد. نمودار شکل ۸ مرز پارتو را نسبت به دو هدف میزان ضریب تعیین و میزان پیچیدگی عبارتی رابطه رگرسیونی محاسباتی نشان می‌دهد. نقطه قرمز رنگ دارای بیشترین مقدار ضریب تعیین است و در عین حال در میان تمام کروموزم‌های با این مقدار ضریب تعیین کمترین مقدار پیچیدگی محاسباتی را دارد.



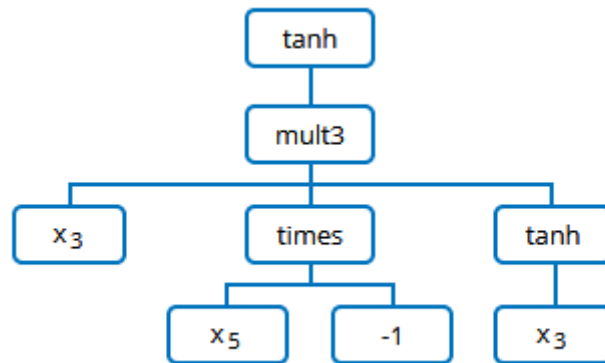
شکل ۸. مرز پارتو برای دو تابع هدف ضریب نیکویی برازش و میزان پیچیدگی مدل

ژن‌های نهایی تشکیل‌دهنده بهترین کروموزوم در جدول ۶ ارائه شده است:

جدول ۶. ژن‌های کروموزوم بهینه

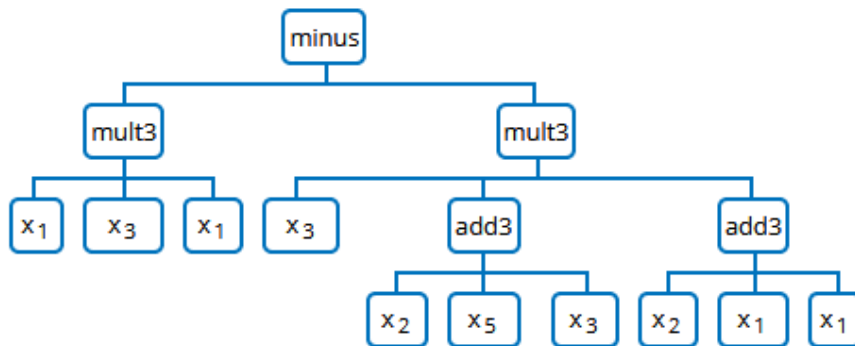
ژن	مقدار
Gene 1	$\tanh(x_3 x_5 \tanh(x_3))$
Gene 2	$x_1^2 x_3$
Gene 3	$x_3(2.0 x_1 + x_2)(x_2 + x_3 + x_5) - 5.23 x_1^2 x_3$
Gene 4	$x_1^2 \tanh(x_3)(x_5 - 4.0)(x_2 + x_3 + x_4)$

برای نمونه درخت متناظر با ژن اول به صورت زیر است.



شکل ۹. درخت متناظر با ژن بهینه اول

یا درخت متناظر با ژن سوم به صورت زیر است.



شکل ۱۰. درخت متناظر با ژن بهینه سوم

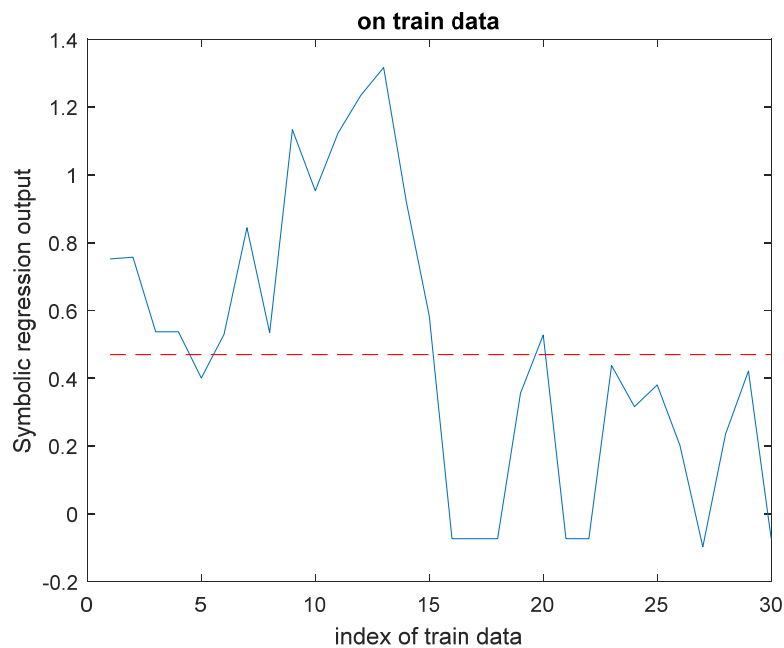
مدل نهایی رگرسیون سمبلیک به صورت زیر ارائه شده است.

$$y = 5.23 x_3(2.0 x_1 + x_2) (x_2 + x_3 + x_5) - 200.0 x_1^2 x_3 -$$

$$0.488 \tanh(x_3 x_5 \tanh(x_3)) -$$

$$22.7 x_1^2 \tanh(x_3) (2.0 x_5 - 8.0) (x_2 + x_3 + x_4) - 0.023$$

عملکرد رگرسیون سمبلیک در طبقه‌بندی مشتریان در داده‌های آموزشی در نمودار شکل ۱۱ ارائه شده است. در این نمودار، مانند گذشته ۱۵ داده اول خوش‌حساب و ۱۵ داده بعدی بدحساب هستند و نقطه برش به کمک منحنی عملکرد مشخصه برابر $0/47$ محاسبه شد.



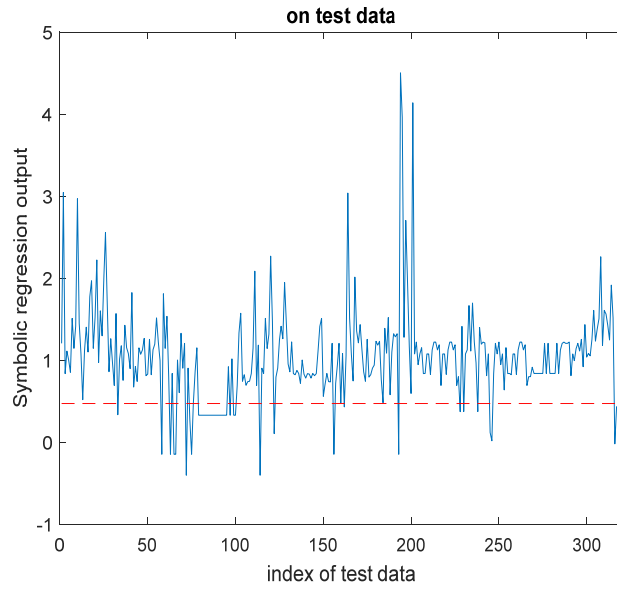
شکل ۱۱. عملکرد رگرسیون سمبلیک روی داده‌های آموزشی

بر اساس نمودار شکل ۱۱:

الف. احتمال طبقه‌بندی درست یک مشتری خوش‌حساب برابر $0/93$ است.

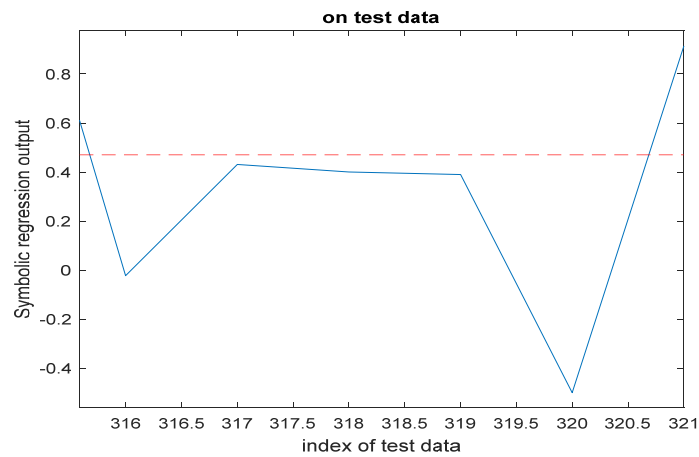
ب. احتمال طبقه‌بندی درست یک مشتری بدحساب برابر $0/93$ است.

نتایج حاصل از رگرسیون سمبلیک در داده‌های آموزشی در مقایسه با رویکرد لجستیک کاملاً بهتر است. حال باید روی داده‌های تست نیز بررسی شود. مانند گذشته، عملکرد مدل رگرسیونی روی ۳۱۵ داده تست از پرونده‌های سالم (۳۲۱ داده اول نمودار شکل ۱۲) و ۶ داده از پرونده‌های بدحساب (۶ داده پایانی نمودار شکل ۱۲) ارائه شده است.



شکل ۱۲. عملکرد رگرسیون سمبلیک روی داده‌های تست

عملکرد رگرسیون سمبلیک در طبقه‌بندی مشتریان بدحساب در ۶ داده آخر نمودار شکل ۱۳ موجود است که برای مشاهده بهتر به صورت بزرگ‌نمایی در نمودار شکل ۱۳ ارائه شده است:



شکل ۱۳. عملکرد رگرسیون سمبلیک روی داده‌های تست بدحساب

بر اساس نتایج به‌دست‌آمده از رگرسیون سمبلیک روی داده‌های تست:

الف. احتمال طبقه‌بندی درست یک مشتری خوش‌حساب برابر $0/88$ است.

ب. احتمال طبقه‌بندی درست یک مشتری بدحساب برابر $0/83$ است.

بنابراین عملکرد مدل ترکیبی رگرسیون لجستیک - سمبلیک توانسته در مقایسه با مدل لجستیک نتایج بهتری ارائه کند و هر دو احتمال را افزایش دهد که این مطلب معادل کاهش هم‌زمان خطای نوع اول (طبقه‌بندی اشتباه یک مشتری خوش‌حساب) و خطای نوع دوم (طبقه‌بندی اشتباه یک مشتری بدحساب) نسبت به مدل متداول لجستیک است.

نتیجه‌گیری

نظارت در سیستم بانکی دارای سه مرحله الف. نظارت بر اعطای تسهیلات، ب. نظارت بر نحوه مصرف منابع و ج. نظارت بر بازگشت منابع بوده و چنانچه نظارت بر اعطای تسهیلات به‌عنوان نخستین مرحله نظارت به‌درستی انجام شود، مراحل بعدی نظارت با مشکل کمتری مواجه خواهد بود. پژوهش حاضر با هدف ارائه یک مدل ترکیبی با رویکرد رگرسیون لجستیک و رگرسیون سمبلیک در طبقه‌بندی مشتریان خوش‌حساب و بدحساب در مورد مطالعاتی بانک سپه با تمرکز بر نظارت بر اعطای تسهیلات از طریق شناسایی متغیرهای اثرگذار بر ریسک اعتباری انجام شد. بدین منظور، ۱۷ متغیر مالی و غیرمالی شامل میزان، مدت و نرخ بهره تسهیلات، نوع وثیقه، متوسط در آمد ماهیانه، سن، تعداد چک برگشتی مشتری، سابقه بدهی بانکی، طول عمر حساب مشتری، نوع تسهیلات اعطایی و تکلیفی یا غیرتکلیفی بودن تسهیلات، خوش‌حسابی، سطح تحصیلات، شغل، وضعیت ملکیت منزل، وضعیت تأهل و وضعیت اشتغال همسر در راستای پیش‌بینی خوش‌حسابی یا بدحسابی اشخاص حقیقی متقاضی تسهیلات بانکی استفاده شد. افراد بدحساب افراد حقیقی هستند که در بازپرداخت تسهیلات اعطایی بانک با مشکل مواجه شده دارای ۶ و بیشتر قسط معوقه هستند. تعداد کل پرونده‌های استخراج‌شده برابر ۳۵۱ پرونده بوده که از این تعداد ۳۳۰ پرونده مرتبط با اشخاص خوش‌حساب و ۲۱ پرونده مرتبط با بدحسابان است. بنابراین، نزدیک به $0/94$ از آمار مشتریان نمونه را خوش‌حساب‌ها و $0/06$ باقی‌مانده را افراد بدحساب تشکیل داده است. ۳۵۱ پرونده نمونه به ۲ دسته داده‌های آموزشی و داده‌های تست تقسیم شدند. ۱۵ پرونده سالم (از بین ۳۳۰ پرونده) و ۱۵ پرونده بدحساب (از بین ۲۱ پرونده و معادل $0/25$ داده‌ها) به‌صورت تصادفی به‌عنوان داده‌های آموزشی انتخاب شد.

به کمک رویکرد انتخاب رو به جلوی والد^۱ در نرم افزار اسپاس، متغیرهای متوسط درآمد ماهیانه، تعداد چک‌های برگشتی، سابقه بدهی بانکی، طول عمر حساب و نوع وثیقه برای ورود به رگرسیون لجستیک انتخاب شدند. بر اساس منحنی ROC، نقطه برش بهینه برابر ۰/۴۴ در داده‌های آموزشی محاسبه شد. نتیجه مدل لجستیک حاکی از احتمال طبقه‌بندی درست یک مشتری خوش حساب ۰/۷۳ و احتمال طبقه‌بندی درست یک مشتری بدحساب برابر ۰/۷۳ است. عملکرد مدل رگرسیونی روی ۳۱۵ داده تست از پرونده‌های سالم نشان می‌دهد که احتمال طبقه‌بندی درست یک مشتری خوش حساب برابر ۰/۷۳ و احتمال طبقه‌بندی درست یک مشتری بدحساب برابر ۰/۵۰ است. در ادامه، از یک شبکه عصبی با سه نرون در لایه پنهان برای طبقه‌بندی مشتریان استفاده شد که ورودی‌های آن ۵ متغیر انتخاب شده در عمل‌گزینش رو به جلوی والد در رگرسیون لجستیک است. نقطه برش به کمک منحنی عملکرد مشخصه برابر ۰/۴۳ محاسبه شد و احتمال طبقه‌بندی درست یک مشتری خوش حساب برابر ۰/۹۳۳ و احتمال طبقه‌بندی درست یک مشتری بدحساب برابر ۰/۹۳۳ است. بنابراین، نتایج حاصل از شبکه عصبی در داده‌های آموزشی نسبت به رویکرد لجستیک کاملاً بهتر است.

بر اساس نتایج به دست آمده از شبکه عصبی روی داده‌های تست احتمال طبقه‌بندی درست یک مشتری خوش حساب برابر ۰/۸۹ و احتمال طبقه‌بندی درست یک مشتری بدحساب برابر ۰/۸۳۳ است. بنابراین، عملکرد مدل ترکیبی شبکه عصبی - رگرسیون لجستیک توانسته در مقایسه با مدل لجستیک نتایج بهتری ارائه کند و هر دو احتمال را افزایش دهد که این مطلب معادل کاهش هم‌زمان خطای نوع اول (طبقه‌بندی اشتباه یک مشتری خوش حساب) و خطای نوع دوم (طبقه‌بندی اشتباه یک مشتری بدحساب) نسبت به مدل متداول لجستیک است.

مانند رویکرد ترکیبی لجستیک - شبکه‌های عصبی، رویکرد لجستیک - رگرسیون سمبلیک نیز در زمینه ارزیابی ریسک اعتباری مشتریان حقیقی پژوهش شد. ورودی رگرسیون سمبلیک نیز متغیرهای پالایش شده از رویکرد لجستیک است. برای مدل‌سازی رگرسیون سمبلیک از یک برنامه‌نویسی چهارژنه با عمق درختی برابر ۴ و ۶ تابع مانند جمع، تفریق و ضرب استفاده شد. نقطه برش به کمک منحنی عملکرد مشخصه برابر ۰/۴۷ محاسبه شد و احتمال طبقه‌بندی درست یک مشتری خوش حساب برابر ۰/۹۳ و احتمال طبقه‌بندی درست یک مشتری بدحساب برابر ۰/۹۳ است. نتایج حاصل از رگرسیون سمبلیک در داده‌های آموزشی نسبت به رویکرد لجستیک کاملاً بهتر است.

بر اساس نتایج به‌دست‌آمده از رگرسیون سمبلیک روی داده‌های تست احتمال طبقه‌بندی درست یک مشتری خوش‌حساب برابر $0/88$ و احتمال طبقه‌بندی درست یک مشتری بدحساب برابر $0/83$ است. بنابراین، عملکرد مدل ترکیبی رگرسیون لجستیک - سمبلیک توانسته نتایج بهتری نسبت به مدل لجستیک ارائه کند و هر دو احتمال را افزایش دهد که این مطلب معادل کاهش هم‌زمان خطای نوع اول (طبقه‌بندی اشتباه یک مشتری خوش‌حساب) و خطای نوع دوم (طبقه‌بندی اشتباه یک مشتری بدحساب) نسبت به مدل متداول لجستیک است. با وجود اینکه عملکرد هر دو مدل ترکیبی روی داده‌های آموزشی یکسان است، اما رویکرد شبکه‌های عصبی در داده‌های تست، دقت بالاتری دارد. در پژوهش‌های داخلی با استفاده از مدل‌های مختلف و ترکیبی، ریسک اعتباری مورد مطالعه قرار گرفته است و دقت‌های متفاوتی از مدل‌های استفاده‌شده، گزارش شده است. برای نمونه جامعی و همکاران (۱۳۹۴) دقت $0/91$ را برای طبقه‌بندی مشتریان خوش‌حساب و $0/93$ برای مشتریان بدحساب گزارش دادند. در پژوهش عرب مازار و رویین تن این مقادیر به ترتیب برابر $0/86$ و $0/86$ است. عبدلی و فرد حریری (۱۳۹۴) دقت $0/90$ را برای طبقه‌بندی مشتریان خوش‌حساب و $0/86$ را برای مشتریان بدحساب گزارش دادند. در پژوهش مهرآرا، موسایی، تصوری و حسن‌زاده (۱۳۸۸)، این مقادیر به ترتیب برابر $0/86$ و $0/75$ و در پژوهش محتشمی و محتشمی (۱۳۹۶) این مقادیر به ترتیب برابر $0/94$ و $0/82$ است.

از طرفی، نتایج پژوهش نشان می‌دهد که ترکیب کردن روش‌ها یعنی رگرسیون لجستیک و شبکه‌های عصبی و رگرسیون لجستیک و برنامه‌نویسی ژنتیک توانسته دقت مدل انفرادی لجستیک را افزایش دهد که این نتیجه با پژوهش لی و چن^۱ (۲۰۱۸) و لینگ لین^۲ (۲۰۰۹) سازگار است. لی و چن (۲۰۱۸) به ریسک اعتباری در سیستم بانک‌های تجاری تایلند را با استفاده از یک نمونه ۵۱۰ تایی از پرونده‌های وام ارزیابی کردند. در این نمونه ۴۵۹ مورد پرونده سالم و ۵۱ مورد ناسالم است. روش استفاده‌شده به صورت ترکیب دو روش رگرسیون چندمتغیری اسپیلانی و شبکه‌های عصبی است. روش کار به این صورت است که متغیرهای مهم تأثیرگذار توسط رویکرد رگرسیونی انتخاب می‌شوند و در ادامه، توسط شبکه‌های عصبی برای محاسبه امتیازدهی اعتباری به کار می‌رود. نتیجه پژوهش روی نمونه نشان می‌دهد که دقت روش ترکیبی از روش‌های انفرادی بیشتر است. دقت طبقه‌بندی مشتریان خوش‌حساب در حدود ۸۷ درصد و مشتریان بدحساب برابر ۸۱ درصد است.

1. Lee and Chen
2. Ling lin

لینگ لین (۲۰۰۹) به بررسی یک مدل ترکیبی برای سنجش و طبقه‌بندی ریسک اعتباری مشتریان حقوقی ۳۷ بانک تایوان در بازه ۲۰۰۰ تا ۲۰۰۴ و با استفاده از ۲۶ متغیر مالی پرداخت. مدل ترکیبی شامل بخش رگرسیونی لجستیک برای محاسبه متغیرهای مهم تأثیرگذار و بخش شبکه‌های عصبی برای مدل‌سازی امتیاز اعتباری بر اساس متغیرهای استخراج‌شده از رگرسیون لجستیک است. ترکیب این دو مدل خود به سه گروه تقسیم شده که اساس این تقسیم بر پایه نحوه انتخاب نقطه برش است. نتیجه پژوهش نشان می‌دهد که اولاً روش ترکیبی از سایر روش‌ها بهتر است و ثانیاً انتخاب مناسب نقطه برش می‌تواند دقت کار را تا حدود ۵ درصد بهبود بخشد.

شایان ذکر است، فرایند طبقه‌بندی در پژوهش حاضر را از لحاظ درصد پایین مشتریان بدحساب در نمونه، می‌توان مسئله سختی قلمداد کرد. با وجود این، مدل ترکیبی توانسته با شناسایی یک مرز حساس به نتایج مناسبی در داده‌های تست دست یابد، هرچند این مرز بسیار حساس است و داده‌های تعدادی از مشتریان دو گروه به مرز کاملاً نزدیک است. بر اساس نتایج به‌دست‌آمده، استفاده از این سیستم در پیش‌بینی ریسک اعتباری به‌صورت کاربردی توصیه می‌شود. همچنین، متغیرهای اثرگذار در این پژوهش درآمد، تعداد چک برگشتی، سابقه نداشتن بدهی، طول عمر حساب مشتری، نوع وثیقه مشخص شدند که برای اعمال صحیح نظارت بانکی در مرحله نخست این متغیرها بایستی بیشتر به آنها توجه شود.

منابع و مأخذ

الف. فارسی

- اسماعیلی، بابک (۱۳۹۷). نقش وقوع سیکل‌های تجاری در مطالبات معوق بانک‌های کشور با استفاده از فیلترهای میان‌گذر. *فصل‌نامه اقتصاد مالی*، (۴۴)، ۱۶۱-۱۸۸.
- بهارلو، ناهید؛ امین بیدختی، علی اکبر و محقق‌نیا، جواد (۱۳۹۴). مقایسه مدل بهینه رگرسیون لجستیک چندگانه و باینری برای رتبه‌بندی اعتباری مشتریان حقیقی بانک رفاه کارگران. *فصل‌نامه پژوهش‌نامه اقتصادی*، (۶۳)، ۱۴۷-۱۶۶.
- بیک زاد، جعفر؛ آقازاده، غلامرضا و آقا زاده، محمدرضا (۱۳۹۳). بررسی عوامل مؤثر بر ریسک اعتباری و اولویت‌بندی معیارهای امتیازدهی اعتباری (۶C) مشتریان بانکی با تکنیک AHP (مطالعه موردی: بانک ملی آذربایجان غربی). *فصل‌نامه روند*، (۶۸)، ۱۲۱-۱۵۰.

- پورعبادالهیان کویچ، محسن؛ اصغرپور، حسین؛ فلاحی، فیروز و ستارهمتی، همت (۱۳۹۷). اندازه‌گیری شکنندگی سیستم بانکی ایران بر اساس شاخص BSFI، *فصل‌نامه اقتصاد مالی*، (۴۵)، ۱-۲۶.
- تهرانی، رضا و فلاح شمس، میرفیض (۱۳۸۴). طراحی و تبیین مدل ریسک اعتباری در نظام بانکی کشور. *مجله علوم اجتماعی و انسانی دانشگاه شیراز*، ۲۲(۲)، ۴۵-۶۰.
- حاجی کرد، امین؛ اصغر زاده زعفرانی، ذملیحه و امام دوست، مصطفی (۱۳۹۵). بررسی ریسک اعتباری مشتریان حقوقی با استفاده از مدل ماشین بردار پشتیبان و مدل هیبریدی الگوریتم ژنتیک (مطالعه موردی: بانک تجارت). *مهندسی مالی و مدیریت اوراق بهادار*، ۷(۲۷)، ۳-۱۷.
- حاجی‌ها، زهره و قربانی، ایوب (۱۴۰۰). بررسی تأثیر ریسک عدم بازپرداخت بدهی‌ها بر رتبه‌بندی اعتباری با توجه به نقش تعدیلی محدودیت مالی. *حسابداری مدیریت*، ۱۴ (۵۰)، ۱-۱۳.
- حسینی، عبدالخالق و زیبایی، منصور (۱۳۹۳). مدیریت ریسک اعتباری در بانک کشاورزی شهرستان ممسنی با استفاده از مدل شبکه عصبی. *مجله اقتصاد کشاورزی*، ۲(۲)، ۱۰۳-۱۱۹.
- عبدلی، قهرمان و فرد حریری، علیرضا (۱۳۹۴). الگوسازی سنجش ریسک اعتباری مشتریان حقوقی بانک رفاه. *فصل‌نامه نظریه‌های کاربردی اقتصاد*، ۲(۱)، ۱-۲۴.
- علم‌الهدی، سید سجاد (۱۳۹۷). بانکداری اسلامی و ریسک، یک تحلیل تطبیقی. *فصل‌نامه اقتصاد مالی*، (۹۷)، ۶۹-۸۶.
- فقیهی نژاد، محمدتقی و مینایی، بهروز (۱۳۹۷). پیش‌بینی رفتار بازار سهام بر اساس شبکه‌های عصبی مصنوعی با رویکرد یادگیری جمعی هوشمند. *مدیریت صنعتی*، ۱۰(۲)، ۳۱۵-۳۳۴.
- قاسمی، احمدرضا و دنیایی، طاهره (۱۳۹۴). اندازه‌گیری ریسک اعتباری مشتریان با رویکرد شبکه عصبی در یکی از بانک‌های دولتی. *مهندسی مالی و مدیریت اوراق بهادار*، ۲(۲۷)، ۱۵۵-۱۸۱.
- کریمی، زهرا؛ اسدی گرجی، حسین؛ گیلک حکیم آبادی، محمد تقی و اسدی، نوراله (۱۳۹۴). عوامل مؤثر بر ریسک اعتباری مشتریان بانکهای تجاری (مطالعه موردی: بانک تجارت شهر نکا - استان مازندران). *فصل‌نامه اقتصاد پولی و مالی*، (۱۰)، ۲۰۵-۲۳۴.
- محتشمی، فاطمه و محتشمی، علی (۱۳۹۶). کاربرد شبکه‌های عصبی فازی در رتبه‌بندی اعتباری مشتریان بانک تجارت. *کنفرانس ملی تحقیقات علمی جهان در مدیریت، حسابداری، حقوق و علوم اجتماعی، شیراز، دانشگاه آزاد*. اسلامی واحد زرقان - دانشگاه علمی کاربردی شوشتر.

محقق نیا، محمدجواد؛ دهقان دهنوی، محمدعلی و بائی، محیا (۱۳۹۸). تأثیر عوامل درونی و بیرونی صنعت بانکداری بر ریسک اعتباری بانکها در ایران. *اقتصاد مالی (اقتصاد مالی و توسعه)*، ۱۳ (۴۶)، ۱۲۷-۱۴۴.

مهرآرا محسن، موسایی؛ میثم، تصوری؛ مهسا و حسن زاده، آیت (۱۳۸۸). رتبه‌بندی اعتباری مشتریان حقوقی بانک پارسیان. *مدل‌سازی اقتصادی*، ۳ (۴)، ۱۲۱-۱۵۰.

میرزایی، حسین؛ نظریان، رافیک و باقری، رعنا (۱۳۹۰). بررسی عوامل مؤثر بر ریسک اعتباری اشخاص حقوقی بانکها (مطالعه موردی: شعب بانک ملی ایران، شهر تهران). *فصل‌نامه روند پژوهش‌های اقتصادی*، ۱۹ (۵۸)، ۶۷-۹۸.

ب. انگلیسی

Abdo, H. (2018). Genetic programming for credit scoring: The case of Egyptian public sector banks. *Expert Systems with Applications*, 36 (2009), 11402-11417.

Bekhet, H. (2018). Credit risk assessment model for Jordanian commercial banks: Neural scoring approach. *Review of Development Finance*, 4(1), 20-28.

Kasabov, N. K. (2010). *Foundations of Neural Networks, Fuzzy Systems, and Knowledge Engineering*, The MIT Press. ISBN 0-262-11212-4.

Khemakhem, S., Boujelbene, Y. (2018), Artificial Intelligence for Credit Risk Assessment: Artificial Neural Network and Support Vector Machines. *ACRN Oxford Journal of Finance and Risk Perspectives*, 6(2), 1-17.

Koza, J. (1992). *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. MIT Press, Cambridge, MA, USA, 1992.

Li, S., Tingting, L., Chao, X. & Zongfang, Z. (2018). ????. *Management Science and Engineering*, 7(3), 81-85.

Lin, L.Sh. (2009). A new two-stage hybrid approach of credit risk in banking industry. *Expert Systems with Applications*, 36(4), 8333-8341.

Sheta, A., Ahmed, S., Faris, H. (2015). Evolving stock market prediction models using multi-gene symbolic regression genetic programming. *Artificial Intelligence and Machine Learning (AIML) Journal*, (12), 11-20.

Spuchlakova, E., & Cug, J. (2014). *Education, Business, and Information Science*, Shanghai, China, EDUGait Press, Canada, PP. 12-15.

Turan, H. (2016). The Weighting of Factors Affecting Credit Risk in Banking. *Istanbul Conference of Economics and Finance*, ICEF 2015, 22-23 October 2015, Istanbul, Turkey.