فصلنامه پژوهش‌های پولی_ بانکی
سال دهم، شماره ۳۴، زمستان ۱۳۹۶
صفحات ۶۵۷_ ۶۸۰

مقایسه مدل‌های شبکه عصبی، الگوریتم ژنتیک و لاجیت در ارزیابی ریسک اعتباری مشتریان

فتح‌الله تاری [*]
سید احمد ابراهیمی [†]
سید جعفر موسوی [‡]
محمود کلانتری [§]

تاریخ دریافت: ۱۳۹۵/۰۹/۲۳ تاریخ پذیرش: ۱۳۹۷/۰۴/۰۳

چکیده

هدف پژوهش حاضر ارزیابی روش‌های رتبه‌بندی اعتباری مشتریان حقیقی (دریافت‌کنندگان اعتبارات خُرد) بانک‌ها، به‌وسیله بررسی سوابق مالی و مشخصات خصیصه‌ای فرد متقاضی می‌باشد. بررسی‌های صورت گرفته نشان می‌دهد که جهت رتبه‌بندی اعتباری مشتریان عمدتاً از سه روش؛ مدل لاجیت، شبکه عصبی و الگوریتم ژنتیک، استفاده می‌شود. در این پژوهش کارایی این روش‌ها جهت سنجش دقیق نکول مورد ارزیابی قرار می‌گیرد. بدین منظور اطلاعات و داده‌های مالی و کیفی یک نمونه تصادفی ۳۹۹ تایی از مشتریان که طی سال‌های ۸۷ الی ۹۱ تسهیلات دریافت نموده‌اند مورد بررسی قرار می‌گیرد. پس از بررسی پرونده‌های اعتباری هر یک از مشتریان، ۱۲ متغیر توضیحی شناسایی گردید که براساس آزمون لاجیت متغیرهای؛ سابقه اعتباری، معدل شش‌ماهه حساب، وضعیت اشتغال، میزان اعتبار درخواستی، اقساط ماهانه و مدت بازپرداخت تأثیر معنی‌دار بر نکول داشته‌اند. نتایج ارزیابی روش‌های رتبه‌بندی اعتباری نشان‌دهنده این است که عملکرد شبکه عصبی نسبت به مدل ژنتیک و لاجیت به مراتب بهتر بوده است چرا که درجه حساسیت ٪۸۲/۹۲ و تشخیص ٪۷۶/۹۲ می‌باشد و به‌طورکلی این مدل توانسته است ٪۸۰ نکول یا عدم نکول را درست پیش‌بینی کند. بنابراین پیشنهاد می‌شود جهت کاهش ریسک اعتباری بانک، اصلاح ساختاری مبتنی بر ایجاد سامانه اعتبار سنجی مشتریان بر اساس شبکه عصبی صورت پذیرد.

واژه‌های کلیدی: پروبیت، لاجستیک، تحلیل ممیزی، تابع توزیع تجمعی
طبقه‌بندی JEL: C58، G21

[*] دانشیار دانشکده اقتصاد، دانشگاه علامه طباطبائی (ره)؛ tarifath@gmail.com
[†] پژوهشگر مرکز تحقیقات سیاست علمی کشور؛ ahmadebrahimi20@yahoo.com (نویسنده مسئول)
[‡] مدیر بازرسی بانک حکمت ایرانیان دانشجوی دکتری کار آفرینی؛ sj_mousavi_ut@yahoo.com
[§] کارشناس ارشد اقتصاد، دانشگاه علوم اقتصادی؛ Mahmoodk67@yahoo.com

۱ مقدمه

کسب آگاهی از آینده و تلاش برای رویارویی مناسب با رویدادهای آتی از جمله مهم‌ترین دغدغه‌های بشر در طول تاریخ بوده است. جهانی‌سازی، پدید آمدن بازارهای مالی جدید، تشدید رقابت میان فعالان صنعت، تغییرات سریع اقتصادی، اجتماعی و تکنولوژیکی موجب افزایش عدم اطمینان و بی‌ثباتی در محیط‌های مالی گردیده است؛ ریسک حاصل از این شرایط فرایند تصمیم‌گیری در مسائل مالی را پیچیده‌تر و دشوارتر ساخته است. از جمله مهم‌ترین ریسک‌های چالش‌برانگیز بانک‌ها و مؤسسات اعتباری، ریسک اعتباری مشتریان می‌باشد؛ عدم ایفای تعهد مشتری در بازپرداخت اصل و فرع تسهیلات اخذشده، علاوه بر اینکه در سطح خرد با اتلاف منابع و دارایی‌های موسسه مالی موجبات ضرر و زیان بانک‌ها و ذینفعان آنان را پدید می‌آورد، در سطح کلان نیز با کاهش قدرت وام‌دهی بانک موجبات کاهش تولید ناخالص داخلی، افزایش بیکاری، اتلاف منابع کشور و ... را فراهم خواهد آورد. اعتبارسنجی مناسب مشتریان از اهمیت شایان توجهی برخوردار است، به‌گونه‌ای که می‌توان اذعان داشت در دهه اخیر، نقش ریسک اعتباری از یک فرآیند مکانیزه منفعل، به یک ابزار استراتژیک تغییر شکل یافته است (ابدوا و همکاران[1]، ۲۰۰۸). اکثر بانک‌های ایران اهمیت این موضوع را پذیرفته و اقدام به ایجاد واحدی برای کنترل و مدیریت ریسک اعتباری خود کرده‌اند.

یکی از راهبردهای اصلی مؤسسات مالی به‌منظور اتخاذ تصمیمات مناسب در اعطای اعتبار و کاهش ریسک اعتباری، طراحی و به‌کارگیری مدل‌های دارای قدرت مناسب برای پیش‌بینی ریسک اعتباری مشتریان متقاضی اعتبار می‌باشد. تاکنون مدل‌های بسیاری برای پیش‌بینی ریسک اعتباری مشتریان حقیقی و حقوقی طراحی گردیده و مورد استفاده قرارگرفته است که از جمله آنها می‌توان به مدل‌های مبتنی بر علوم آماری و مدل‌های مبتنی بر هوش مصنوعی اشاره کرد. مدل رتبه‌بندی مشتریان از دهه ۱۹۴۰، برای شناسایی ریسک درخواست اشخاص برای اعتبار و پشتیبانی تصمیمات مبنی بر رد یا قبول درخواست‌ها به کارگرفته شده است (بلوتی[2]، ۲۰۱۰). امروزه روش‌های متفاوتی برای مدل‌سازی رتبه‌بندی اعتباری مشتریان بانک‌ها وجود دارد که عموماً در امتداد مدل رتبه‌بندی z-score است که از سوی آلتمن[3] (۱۹۶۸) یکی از پیشگامان در این زمینه ارائه شده است. روش‌های قدیمی آماری که برای

[1] Abdou & et al
[2] Bellotti
[3] Altman

ایجاد مدل‌های امتیازدهی به‌کارگرفته شده‌اند عبارتند از: پروبیت، لجیت، رگرسیون خطی، تحلیل متمایزکننده خطی، تحلیل متمایزکننده نهایی، توبیت، درخت‌های صفر و یک، روش‌های کمینه، رگرسیون لجستیک می‌باشند. مدل‌های پیچیده‌تری که از شاخهٔ هوش مصنوعی است نیز، در این حوزه به کار گرفته شده‌اند که از میان آنها می‌توان به سیستم‌های شبکه‌های عصبی و الگوریتم ژنتیک اشاره کرد (کیم و آهن[1]، ۲۰۱۲). مطالعات اخیر نشان داده است روش‌های هوش مصنوعی، مثل شبکه‌های عصبی هوش مصنوعی، محاسبات تکمیلی، الگوریتم ژنتیک و ماشین بردار پشتیبانی، برای تحلیل آماری و مدل‌های بهبود ارزیابی ریسک اعتباری مناسب هستند.

در این پژوهش بر آنیم تا به‌منظور پیش‌بینی ریسک اعتباری مشتریان حقیقی، مدل‌هایی مبتنی بر لاجیت، شبکه عصبی مصنوعی و الگوریتم ژنتیک طراحی نموده و کارایی آنان را با یکدیگر مقایسه نماییم. در بخش دوم از این مقاله پیشینه تحقیق ارائه شده است و در ادامه در بخش سوم مبانی نظری شرح داده شده است. سپس یافته‌های حاصل از مطالعه موردی ارائه شده است و در انتها نتیجه‌گیری و پیشنهادات ارائه شده است.

۲ ادبیات پژوهش

پژوهش‌های نخستین بر مدل‌های تک متغیره همچون یک نسبت مالی تمرکز داشتند؛ یکی از قدیمی‌ترین نسبت‌های مالی که برای ارزیابی وضعیت اعتباری در سال ۱۸۷۰ مورد استفاده قرار گرفت نسبت جاری بود. از نخستین تلاش‌های پژوهشگران در طراحی مدلی برای اندازه‌گیری و درجه‌بندی ریسک اعتباری می‌توان به کاوش جان موری بر روی اوراق قرضه در سال ۱۹۰۹ اشاره نمود(گلانتز[2]،۲۰۰۳). برای نخستین بار آلتمن اثر ترکیب‌های مختلف از نسبت‌های مالی را برای پیش‌بینی ورشکستگی شرکت‌ها مورد بررسی قرار داد. وی در این مطالعه الگوی تحلیل ممیزی چندگانه[3] (MDA) را به کار گرفت. مدلی که او به دست آورد و به مدل امتیاز Z (Z-score)معروف است هنوز هم به‌عنوان شاخصی برای پیش‌بینی ورشکستگی شرکت‌ها مورد استفاده قرار می‌گیرد. فرضیه اصلی آلتمن(۱۹۶۸) این بود که مدل پیش‌بینی ورشکستگی او که از ۵ نسبت مالی تشکیل می‌شد می‌تواند برای تشخیص شرکت‌های ورشکسته از غیر ورشکسته مورد استفاده قرار گیرد. با توجه به اینکه عدم بازپرداخت وام

[1] Kim & Ahn

[2] Glantz

[3] Multiple Discriminant Analysis

عمدتاً مربوط به شرکت‌هایی است که در آینده دچار درماندگی مالی می‌شوند بنابراین امکان پیش‌بینی ریسک اعتباری با استفاده از این مدل امکان‌پذیر خواهد بود؛ ازاین‌رو ساندرز و آلن[1] (۲۰۰۲) از مدل Z آلتمن برای پیش‌بینی ریسک اعتباری شرکت‌هایی که از بانک‌ها وام دریافت کرده بودند بهره بردند و بررسی‌های آنان مشخص نمود که این مدل از قابلیت بالایی در پیش‌بینی ریسک اعتباری برخوردار است. مین و لی[2] (۲۰۰۵) با استفاده از ماشین بردار پشتیبان اقدام[3] (SVM) برای طراحی مدلی برای پیش‌بینی ورشکستگی شرکت‌ها نمودند. نتیجه پژوهش آنان حاکی از آن بود که SVM نسبت به مدل‌های آماری سنتی از عملکرد بهتری برخوردار است. مین، لی و هان[4](۲۰۰۶) به‌صورت هم‌زمان الگوریتم ژنتیک و ماشین بردار پشتیان را به کار گرفته و آن را مدل GA-SVM نامیدند. نتایج پیش‌بینی آنها حاکی از دقت پیش‌بینی ۸۶ درصدی در مجموعه آموزشی و ۸۰ درصدی در نمونه آزمایشی در یک سال پیش از ورشکستگی بود.

در داخل کشور نیز پژوهش‌هایی در چند سال اخیر صورت گرفته است، کشاورز و آیتی گازار (۱۳۸۶) به مقایسه کارکرد مدل لاجیت و روش درخت‌های طبقه‌بندی[5] (CART) و رگرسیونی لاجیت در فرایند اعتبارسنجی متقاضیان حقیقی برای استفاده از تسهیلات پرداختند و به این نتیجه رسیدند که روش غیرپارامتری (CART) از دقت بالاتری در پیش‌بینی مشتریان خوب و بد دارد. مهر آرا و همکاران (۱۳۹۰) به بررسی رتبه‌بندی اعتباری مشتریان حقوقی بانک پارسیان با استفاده رگرسیون لاجیت و پروبیت و مدل شبکه‌های عصبی هوشمند پرداختند. نتایج این پژوهش نشانگر آن بود که در پیش‌بینی عملکرد صحیح، شبکه‌های عصبی به مراتب بهتر از الگوی لاجیت و پروبیت می‌باشد. کریمی و همکاران (۱۳۹۴) به بررسی عوامل مؤثر بر ریسک اعتباری مشتریان با روش لاجستیک برای بانک‌های تجاری پرداختند نتایج حاصل از این تحقیق نشان داد که مدت تسهیلات، نرخ تسهیلات و نوع وثیقه و نوع تسهیلات تأثیر معنی‌داری بر وصول مطالبات بانکی دارد. دیگر پژوهش‌ها عبارتند از: نبوی و همکاران (۱۳۸۹)، اخباری و رفیعی (۱۳۸۹)، تهرانی و فلاح (۱۳۹۰)، فیروزیان و همکاران (۱۳۹۰).

---

[1] Saunders & Allen
[2] Min & Lee
[3] Support Vector Machine
[4] Min, Lee & Han
[5] Classification And Regression Tree

این پژوهش حاوی سه نوآوری نسبت به مطالعات قبلی است اولاً شناسایی متغیرهای تأثیرگذار بر اعتبار مشتریان حقیقی که متغیرهای هیستوریکال (وضعیت چک، معدل حساب و...)، متغیرهای دموگرافیکال (جنسیت، سن، تحصیلات و...) و انتخاب مشتریانی که از تسهیلات خرد استفاده کرده‌اند با میانگین مبلغ تسهیلات در حدود ۱۵ میلیون تومان، مورد آزمون قرار گرفته‌اند، ثانیاً مقایسه هر سه مدل لاجیت، شبکه عصبی و الگوریتم ژنتیک ثالثاً محاسبه حساسیت و تشخیص برای هر سه مدل می‌باشد.

۱.۲ ریسک اعتباری

ریسک اعتباری ریسکی است که از نکول/قصور طرف قرارداد یا در حالت کلی‌تر از اتفاقی اعتباری پدید می‌آید. از نظر تاریخی این ریسک معمولاً در مورد اوراق قرضه در نظر گرفته می‌شد، بدین ترتیب که قرض دهندگان از بازپرداخت وام اعطایی به قرض گیرنده اطمینان نداشتند؛ ازاین‌رو از این ریسک با نام ریسک نکول یاد می‌شود. ریسک اعتباری ناشی از ناتوانی و یا عدم تمایل دریافت‌کننده تسهیلات در بازپرداخت آن می‌باشد. این عدم ایفای تعهدات می‌تواند ناشی از رکود شرایط کسب‌وکار یا دیگر عوامل اقتصادی باشد که دریافت‌کننده تسهیلات با آن مواجه است. بدین ترتیب ریسک اعتباری عبارت است از احتمال کاهش ارزش یا بی‌ارزش شدن برخی از دارایی‌های بانک یعنی تسهیلات اعطایی آن در اثر عدم ایفای تعهدات دریافت‌کننده تسهیلات به بازپرداخت اصل و فرع آن. با توجه به اینکه رقم سرمایه بانک‌ها در قیاس با کل ارزش دارایی‌های آنها کم است حتی اگر درصد کمی از وام‌های اعطایی قابل وصول نباشند، بانک با خطر ورشکستگی مواجه خواهد شد. فرآیند مدیریت ریسک اعتباری به معنی شناسایی، ارزیابی، تجزیه‌وتحلیل و واکنش مناسب به ریسک اعتباری و نیز نظارت مستمر بر آنان با توجه به شرایط متغیر محیط اعم از شرایط اقتصادی، سیاسی، اجتماعی، تکنولوژیکی و ... می‌باشد. مهم‌ترین ابزاری که بانک‌ها برای مدیریت و کنترل ریسک اعتباری بدان نیازمندند سیستم رتبه‌بندی اعتباری یا اعتبارسنجی مشتریان می‌باشد (مهرآرا و همکاران، ۱۳۹۰).

۳ روش‌شناسی پژوهش

۱.۳ تابع لاجیت

مدل لاجیت یکی از مدل‌های رگرسیونی است که در حالتی که متغیر وابسته باینری یا موهومی باشد مورد استفاده قرار می‌گیرد. در این حالت Y به‌عنوان متغیر وابسته تنها می‌تواند مقادیر ۰ (متناظر با بازپرداخت به‌موقع وام توسط مشتری) و ۱ (نکول در بازپرداخت) را اختیارکند. اگر متغیرهای مستقل $X_{1i}, \dots, X_{ni}$ عوامل تعیین‌کننده احتمال نکول باشند احتمال این‌که متغیر Y مقدار ۱ را بپذیرد برابر است با:

$$p_i = E(y=1|X_i, i=1,2,\dots,n) = \frac{1}{1+e^{-(\beta_0+\beta_1 X_{1i}+\dots+\beta_n X_{ni})}} \quad (۳)$$

که در آن p_i احتمال نکول مشتری و متغیرهای X_{1i} تا X_{ni} عوامل تعیین‌کننده احتمال یاد شده می‌باشند. این تابع را تابع توزیع تجمعی لجستیک[1] می‌نامند. این تابع را می‌توان به شکل زیر دوباره‌نویسی کرد:

$$p_i = \frac{1}{1+e^{-Z_i}}$$
$$Z_i = \beta_0 + \beta_1 X_{1i} + \dots + \beta_n X_{ni} \quad (۴)$$

همچنان که Z_i بین -∞ تا +∞ تغییر می‌کند p_i بین ۰ و ۱ مقادیر خود را اختیار خواهد کرد. در این رابطه p_i به‌طور غیرخطی به Z_i (یعنی X_i) مربوط است. اما از تابع p_i ملاحظه می‌شود p_i نه‌تنها برحسب X بلکه برحسب βها هم غیرخطی است. اما به راحتی می‌توان نشان داد که می‌توان معادله p_i را به‌صورت رابطه‌ای خطی برحسب پارامترها تبدیل نمود.

[1] Logistic Distribution Function

$$1-p_i=\frac{1}{1+e^{Z}} \quad (۵)$$

$$\frac{p_i}{1-p_i}=\frac{1+e^{Z}}{1+e^{-Z}}=e^{Z} \quad (۶)$$

حال به‌طور ساده $\frac{P_i}{1-P_i}$ نسبت احتمال حادثه مورد نظر بر آلترناتیو آن است که در اینجا بیانگر میزان برتری احتمال وقوع حادثه بر عدم آن می‌باشد.

حال چنانچه از تابع مذکور لگاریتم بگیریم نتیجه زیر به دست می‌آید:

$$Z_i=\beta_0+\beta_1X_{1i}+...+\beta_nX_{ni}=\ln(\frac{p_i}{1-p_i}) \quad (۷)$$

یعنی Z که لگاریتم نسبت برتری یا مزیت است نه‌تنها برحسب X بلکه (نکته مهم از نظر تخمین) برحسب پارامترها نیز، خطی است.

۲.۳ مدل پروبیت

مدل پروبیت نیز مانند مدل لاجیت یک تابع توزیع تجمعی[1] (CDF) می‌باشد، با این تفاوت که در مدل لاجیت تابع توزیع لوجیستیکی و در مدل پروبیت نرمال می‌باشد. اگر I_i شاخص وقوع حادثه باشد، این شاخص به‌صورت زیر بیان می‌شود:

$$I_i=\beta_0+\beta_1X_{1i}+...+\beta_nX_{ni} \quad (۸)$$

از طرفی دیگر I_i^* را مقدار آستانه شاخص فوق می‌نامیم، بدین‌صورت که اگر $I_i>I_i^*$ حادثه به وقوع می‌پیوندد. احتمال وقوع حادثه به‌صورت زیر بیان می‌شود:

[1] Cumulative distribution function

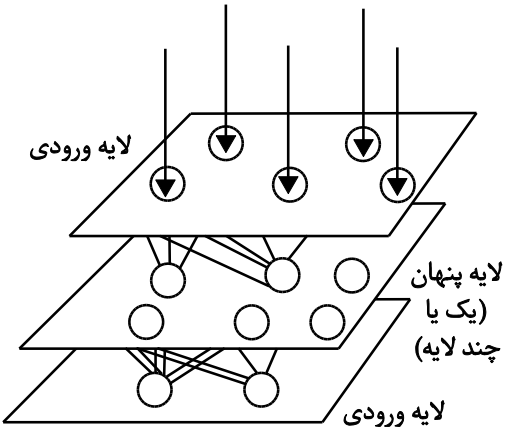
(۹)

$$p_i = p_r(Y=1) = p_r(I_i \geq I_i^*) = F(I_i \geq I_i^*) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{T_i} e^{\frac{-t^2}{2}} dt = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\beta_0+\beta_1 X_{1i}+\ldots+\beta_n X_{ni}} e^{\frac{-t^2}{2}} dt$$

که در آن t متغیر نرمال استاندارد است یعنی $t \sim \mathbb{N}(0,1)$ برای تخمین پارامترهای این مدل از روش حداکثر راست نمایی استفاده می‌شود.

۳.۳ مدل‌های پیش‌بینی مبتنی بر شبکه عصبی

شبکه عصبی مصنوعی تقلیدی بسیار ساده از سیستم عصبی بیولوژیکی و مغز انسان است، این تقلید بر اساس یک پیکربندی ریاضی می‌باشد، به‌طوری‌که متشکل از چندین لایه و همچنین چندین نرون در هر لایه است. نرون‌ها کوچک‌ترین واحدهای سازنده شبکه عصبی مصنوعی می‌باشند و حکم سلول‌های مغز انسان را دارند و هر شبکه از یک لایه ورودی و یک لایه خروجی و یک یا چند لایه میانی تشکیل شده است. شکل زیر یک شبکه عصبی با چنین ساختمانی را نشان می‌دهد (هرتز و همکاران[1]، ۱۹۹۱).



شکل ۱. نحوه کارکرد شبکه عصبی

شبکه‌های عصبی مصنوعی سیستم‌های یادگیری ماشینی هستند که بر اساس شبیه‌سازی اصول کاری مغز انسان ابداع گردیده‌اند تا طبقه جدیدی از مدل‌های رگرسیون غیرخطی که

[1] Hertz, J., A. Krogh, and R.G. Palmer

دارای قدرت تفکیک‌کنندگی را ایجاد نمایند. به همان طریقی که شبکه عصبی بیولوژیکی[۱] به‌منظور انجام یک عمل شناختی (نظیر تشخیص چهره‌ها یا یادگیری یک مفهوم) خود را تغییر می‌دهد، شبکه‌های عصبی مصنوعی نیز پارامترهای درونی خویش را به‌منظور انجام وظیفه محوله تعدیل می‌نمایند. وظایف معمولی که شبکه‌های عصبی به‌گونه‌ای اثربخش و کارا انجام می‌دهند عبارتند از:

۱) طبقه‌بندی[۲]
۲) شناسایی الگوهای حاضر در مجموعه داده‌ها[۳]
۳) پیش‌بینی (آنجلینی، دی‌تولو و رولی[۴]،۲۰۰۸)

از شبکه‌های عصبی در بسیاری از مسائل مالی از جمله پیش‌بینی ریسک اعتباری و درماندگی مالی استفاده گردیده است. اودو معتقد است که ویژگی‌ها و ظرفیت‌های شبکه‌های عصبی مصنوعی آنها را به یکی از جذاب‌ترین ابزارها به‌منظور پیش‌بینی درماندگی مالی تبدیل کرده است (اودو[۵]،۱۹۹۳).

یک شبکه عصبی مصنوعی از گروهی از نرون‌ها[۶] که با توپولوژی (ساختار) از پیش تعیین‌شده‌ای به یکدیگر ارتباط دارند تشکیل شده است. عموماً براساس وظیفه‌ای که شبکه می‌بایستی انجام دهد برخی توپولوژی‌های محتمل وجود دارند؛ عموماً توپولوژی شبکه ثابت نگاه داشته می‌شود اما در پاره‌ای کاربردها (برای نمونه در رباتیک) خود توپولوژی می‌تواند به‌عنوان یک پارامتر تلقی گردیده و به‌گونه‌ای پویا تغییر کند.

اتصالات (ارتباطات) میان نرون‌ها با یک «وزن» ایجاد گردیده‌اند که نوع و شدت اطلاعاتی را که ردوبدل می‌شوند، تعیین می‌کنند. نرون‌ها واحدهای محاسباتی پایه‌ای شبکه هستند؛ هر نرون، ورودی‌هایی را از دیگر نرون‌ها دریافت کرده و خروجی‌ای که به نرون‌های مقصد منتقل می‌گردد پدید می‌آورد.

ایجاد خروجی توسط نرون در ۲ گام صورت می‌گیرد: نخست مجموع وزنی ورودی‌ها ارزیابی می‌گردد، بدین ترتیب که هر ورودی به نرون در وزن اتصال مربوطه ضرب می‌شود و

---

[۱] مغز موجود زنده
[۲] تصمیم‌گیری درباره اینکه یک نمونه داده‌شده به کدام طبقه اختصاص دارد
[۳] نظیر تشخیص یک بیماری از برخی نشانه‌ها یا تشخیص علل به‌محض اینکه آثار مشاهده شوند
[4] Angelini, Di Tollo&Roli
[5] Udo
[6] neuron

مجموع وزنی حاصل با تورش[1] نرون جمع می‌شود؛ در گام دوم با اعمال تابع فعال‌سازی نرون به حاصل این جمع، خروجی پدید می‌آید (آنجلینی، دی‌تولو و رولی ،۲۰۰۸).

سازوکار یادگیری شبکه‌های عصبی عبارتند از:

- **یادگیری تحت نظارت[2]:** این یادگیری با استفاده از مجموعه آموزشی صورت می‌گیرد که شامل زوج‌هایی از ورودی و خروجی‌های مطلوب متناظر است که تفاوت میان خروجی شبکه و خروجی مطلوب (خطای ایجادشده) برای تعدیل اوزان مورد استفاده قرار می‌گیرد. این نوع از یادگیری در کاربردهایی بکار گرفته می‌شود که شبکه می‌بایستی آموزش بیند تا مثال‌های داده‌شده را تعمیم دهد. کاربرد معمول، طبقه‌بندی داده‌ها می‌باشد.
- **یادگیری بدون نظارت[3]:** در این الگوریتم شبکه بدون بهره‌گیری از خروجی مطلوب متناظر و تنها با مجموعه‌ای از ورودی‌ها تغذیه خواهد شد. الگوریتم، شبکه را راهنمایی می‌نماید تا خود سازمان‌یافته باشد و اوزانش را تعدیل نماید. این نوع از یادگیری برای اعمالی نظیر داده‌کاوی که می‌بایست برخی قواعد در میان انبوهی از داده یافت شوند مورد استفاده قرار می‌گیرد.
- **یادگیری تقویتی[4]:** در این الگوریتم شبکه با ارائه پاداش و جریمه به‌عنوان تابع پاسخ شبکه آموزش می‌بینند؛ در واقع این پاداش و خطاها برای تعدیل اوزان مورد استفاده قرار می‌گیرد. برای نمونه الگوریتم‌های یادگیری تقویتی برای آموزش سیستم‌های تطابق پذیر که وظیفه‌ای متشکل از سلسله‌ای از اعمال را انجام می‌دهند به کار برده می‌شوند.

در کوشش برای تنظیم بهینه پارامترهای شبکه عصبی و الگوریتم یادگیری، تعداد زیادی آزمایش انجام گردیده است. علیرغم سال‌ها پژوهش در حوزه شبکه عصبی، هنوز تنظیم بهینه پارامترهای شبکه عصبی و الگوریتم یادگیری، فاقد چارچوب نظری بوده و فاز طراحی شبکه به‌صورت تجربی و بر مبنای حساسیت، تجربه و دقت طراح صورت می‌گیرد (آنجلینی، دی‌تولو و رولی ،۲۰۰۸).

---

[1] bias

[2] Supervised learning

[3] Unsupervised learning

[4] Reinforced learning

۴.۳ مدل‌های پیش‌بینی مبتنی بر الگوریتم ژنتیک

الگوریتم‌های ژنتیک ابزارهای قدرتمند بهینه‌سازی هستند که اعمال آنان به دامنه وسیعی از حوزه‌های کاربرد سودمند و موفقیت‌آمیز بوده است؛ این الگوریتم که در طی دهه ۶۰ و ۷۰ میلادی توسط هالند[1] ابداع گردیده است مبتنی بر مفهوم تکامل طبیعی داروین و قانون انتخاب طبیعی[2] می‌باشد؛ بر اساس قانون انتخاب طبیعی، قوی‌ترین یا بلندترین یا سریع‌ترین موجود نیست که بقا می‌یابد بلکه متناسب‌ترین است که بقا می‌یابد؛ بر اساس قانون انتخاب طبیعی، تناسب[3] با نیازهایی که توسط دنیای خارج تحمیل می‌گردد عامل اساسی است که موجب می‌گردد هر چه امکان تطبیق یک موجود بیشتر باشد بقای آن امکان‌پذیرتر باشد و احتمال تولیدمثل بیشتری بیابد؛ مفهوم ضمنی تکامل نیز، ایده بهبود مستمر گونه‌های زیستی به نسبت نسل‌های قبلی است. از الگوریتم ژنتیک در بررسی موارد مربوط به مقوله مدیریت مالی از جمله پیش‌بینی درماندگی مالی و ورشکستگی بهره برده شده است؛ شین و لی پس از به‌کارگیری این مدل در پیش‌بینی درماندگی مالی عنوان کردند که علاوه بر اینکه الگوریتم ژنتیک مدلی مناسب برای پیش‌بینی درماندگی مالی می‌باشد؛ درک منطق پیش‌بینی آن نیز برای استفاده‌کنندگان ساده است (شین و لی[4]،۲۰۰۲)(Shin & Lee, 2002). الگوریتم ژنتیک اصول تکامل طبیعی را در طریقی خاص تکرار می‌نماید؛ این مدل هنگام اعمال بر یک مسئله بر روی یک راه‌حل خاص کار نمی‌کند بلکه بر روی مجموعه‌ای از راه‌حل‌ها که جمعیت خوانده می‌شود کار می‌کند. این ابزار اصل بقای مناسب‌ترین را برای انتخاب راه‌حل مسئله تحت بررسی به کار می‌گیرد. الگوریتم ژنتیک برای حل یک مسئله، مجموعه بسیار بزرگی از راه‌حل‌های ممکن را پدید می‌آورد که هر یک از این راه‌حل‌ها با استفاده از یک تابع تناسب مورد ارزیابی قرار می‌گیرند؛ بر مبنای شماری از بهترین راه‌حل‌ها، راه‌حل‌هایی جدید پدید می‌آیند که تکامل‌یافته راه‌حل‌های قبلی می‌باشند؛ در واقع راه‌حل‌ها به‌گونه‌ای تکامل می‌یابند که به راه‌حل مطلوب دست یابند (وارتو[5]،۱۹۹۸). برای موفقیت عملکرد یک الگوریتم ژنتیک دو موضوع اساسی می‌بایستی مورد توجه قرار گیرند:

---

[1] Holland
[2] Natural selection
[3] fitness
[4] Shin & Lee
[5] Varetto

- **نمایش و ارائه مناسب راه‌حل‌های احتمالی:** الگوریتم‌های ژنتیک به جای کار کردن مستقیم با پارامترها یا متغیرهای مسئله، با شکل کدشده آنان سروکار دارند، و می‌بایستی راه‌حل‌هایی که ما قصد یافتن و بهینه‌سازی آنان را داریم با کدی که الگوریتم توانایی دست‌کاری آن را دارد نمایش داده شود. این کدگذاری ژنتیک می‌تواند در اشکال متنوعی انجام پذیرد اما عموماً کدگذاری دودویی یا باینری که با توالی ۰ و ۱ انجام می‌پذیرد مورد استفاده قرار می‌گیرد. رشته[1] یا دنباله‌ای از بیت‌ها که به‌عنوان شکل کدشده یک جواب از مسئله مورد نظر می‌باشد کروموزوم خوانده می‌شود؛ علت نمایش راه‌حل‌های احتمالی به‌صورت رشته‌ای از بیت‌ها این است که با این کار اعمال ژنتیکی بر روی آنان ساده‌تر صورت می‌گیرد.
- **تعریف تابع برازندگی صحیح:** تابع برازندگی[2] برای ارزیابی عملکرد کروموزوم‌های حاضر در یک جمعیت (راه‌حل‌ها) طراحی می‌گردد؛ این تابع، برازندگی ــ کیفیتــ راه‌حل‌های ایجادشده توسط الگوریتم را مورد ارزیابی قرار داده و مقداری عددی را به آن منتسب می‌سازد که این عدد احتمال انتخاب کروموزوم مورد نظر را برای مشارکت در تولید جمعیت بعدی تعیین خواهد کرد. مکانیزم مورد استفاده توسط الگوریتم ژنتیک به‌گونه‌ای طراحی گردیده است که رشته‌هایی که عدد برازندگی بالاتری دارند احتمال بیشتری برای ترکیب و تولید رشته‌های جدید داشته و در مرحله جایگزینی نسبت به دیگر رشته‌ها مقاوم‌ترند. بدین ترتیب درمی‌یابیم که تابع برازندگی، معیاری برای رتبه‌بندی راه‌حل‌های مسئله است که موجب می‌شود راه‌حل‌های بهتری برای تولید نسل بعدی جمعیت انتخاب گردند.

۵.۳ معیار ارزیابی مدل

کارایی مدل‌های رتبه‌بندی اعتباری با مقایسه پیش‌بینی داده‌های شاهد (که معمولاً ۲۰٪ کل نمونه می‌باشد) با مقادیر واقعی صورت می‌گیرد و همچنین دو شاخص ارزیابی نیز مشخص می‌گردد که این شاخص‌ها عبارتند از:

- درجه حساسیت[3]: نسبتی از مشتریان بدحساب که نتیجه امتیازدهی آنان نیز در مدل رتبه‌بندی به‌عنوان بدحساب پیش‌بینی شده‌اند.

---

[1] String

[2] Fitness function

[3] Sensitivity

- درجه تشخیص[1] (ویژگی): نسبتی از مشتریان خوش‌حساب که نتیجه امتیازدهی آنان نیز در مدل رتبه‌بندی به‌عنوان خوش‌حساب پیش‌بینی شده‌اند.

۴ یافته‌های پژوهش برای مطالعه موردی

۱.۴ گردآوری داده‌ها

داده‌های مورد استفاده از یک نمونه تصادفی ۳۹۹ نفری از مشتریان دریافت‌کننده تسهیلات مورد بررسی قرارگرفته‌اند؛ اعضای نمونه از پایگاه داده مشتریان شعب تهران یکی از بانک‌های خصوصی در حد فاصل سال‌های ۹۱ـ۱۳۸۷ استخراج گردیده است.[2]

۲.۴ انتخاب پارامترها

متغیرهای الگو بر اساس ادبیات ریسک اعتباری، مطالعات مشابه (روزباخ[3] ۱۹۹۸، کشاورز و آیتی گازار ۱۳۸۶، مهرآرا ۱۳۹۰ و...) و مجموعه رهنمودهای بانک مرکزی برای مدیریت مؤثر ریسک اعتباری و با توجه به محدودیت دسترسی به داده‌ها انتخاب شده‌اند. محرمانه بودن و یا در دسترس نبودن اطلاعات مشتریان مهم‌ترین محدودیت در انتخاب متغیرها به‌حساب می‌آید. بر همین اساس متغیرها به شرح زیر انتخاب و تعریف گردیدند:

---

[1] Specificity

[2] دوره بررسی این مقاله در سال ۱۳۹۵ بوده است ولی پرونده‌های تسهیلاتی که در سال‌های ۸۷ الی ۹۱ تسهیلات دریافت کرده‌اند و در زمان پژوهش دوره بازپرداخت آن تسهیلات باید به پایان رسیده باشد و وضعیت آنها ممکن است در حال حاضر معوقً مشکوک‌الوصول و سوخت شده و یا استمهال شده باشد.

[3] Roszbach, 1998

جدول ۱

متغیرهای مورد استفاده در تحقیق

	ردیف	متغیر	توضیحات
مشخصات خصیصه‌ای فرد متقاضی	X_1	جنسیت	درصورتی‌که مشتری مرد باشد ۱ و درصورتی‌که زن باشد ۰
	X_2	سن	زیر ۳۰ سال عدد ۰، ۳۰ـ۳۵ عدد ۱، ۳۵ـ۴۰ عدد ۲ و بالای ۴۰ عدد ۳
	X_3	وضعیت تأهل	درصورتی‌که مشتری متأهل باشد ۱ و در غیر این صورت ۰
	X_4	تحصیلات دانشگاهی	درصورتی‌که مشتری دارای تحصیلات دانشگاهی باشد ۲ و درصورتی‌که مدرک دیپلم داشته باشد عدد ۱ و در غیر این صورت ۰
شرایط متقاضی	X_5	سابقه اعتباری	درصورتی‌که مشتری سابقه چک برگشتی داشته باشد ۱ و در غیر این صورت ۰
	X_6	معدل ۶ ماهه حساب	اگرکمتر از ۱۰۰ هزار باشد عدد ۰، بین ۱۰۰ هزار تا یک میلیون عدد ۱، بیشتر از یک میلیون عدد ۲
	X_7	وضعیت تملک مسکن	درصورتی‌که مشتری ملکی را در تملک خویش داشته باشد ۱ و در غیر این صورت ۰
وضعیت شغلی	X_8	وضعیت اشتغال	درصورتی‌که مشتری شاغل رسمی باشد ۱ و در غیر این صورت ۰
	X_9	سابقه کاری	اگرکمتر از ۳ سال باش ۰، ۳ الی ۵ سال عدد ۱، بیش از ۵ سال عدد ۲
شرایط اعتبار درخواستی	X_{10}	میزان اعتبار درخواستی	۵ میلیون و کمتر عدد ۰، بین ۵ الی ۱۰ میلیون عدد ۱ و بین ۱۵-۱۰ میلیون عدد ۲
	X_{11}	اقساط ماهیانه	کمتر از ۳۰۰ هزار عدد ۰، بین ۳۰۰ الی ۶۰۰ عدد ۱، بین ۶۰۰ الی ۹۰۰ عدد ۲ و بالاتر از آن عدد ۳
	X_{12}	مدت بازپرداخت	کمتر از یک سال عدد ۰، کمتر از ۲۴ ماه عدد ۱ و کمتر از ۳۶ ماه عدد ۲ و بیشتر از ۳۶ ماه عدد ۳

۳.۴ برآورد مدل لاجیت

ابتدا تخمین مدل با ۱۲ متغیرکه در قسمت قبل بیان گردید، و متغیر وابسته وضعیت نکول (نکول یا بدحساب ۱ و عدم نکول یا خوش‌حساب ۰) برای ۳۱۹ مشتری حقیقی که تسهیلات دریافت کرده بودند و سررسید آنها به اتمام رسیده بود با استفاده از مدل لاجیت برآورد گردید، سپس با استفاده از الگو برآورد شده، متغیر وابسته (نکول یا عدم نکول) را برای ۸۰ پرونده دیگر (داده‌های شاهد) که از اطلاعات این ۸۰ پرونده برای تخمین استفاده نشده است، پیش‌بینی می‌گردد. اطلاعات واقعی در خصوص متغیر وابسته برای داده‌هایی شاهد در دسترس هست و بر اساس آن می‌توان عملکرد پیش‌بینی الگو در خارج از دوره تخمین (نسبت

پیش‌بینی‌های صحیح در خصوص نکول یا عدم نکول در خصوص ۸۰ پرونده) را ارزیابی کرد. و سپس در ادامه عملکرد پیش‌بینی مذکور را با عملکرد پیش‌بینی مدل شبکه عصبی و الگوریتم ژنتیک مقایسه می‌کنیم.

نتایج حاصل از تخمین ضرایب مبتنی بر الگو لاجیت با استفاده از ۳۱۹ مشاهده در جدول ذیل ارائه شده است.

جدول ۲

نتایج حاصل از تخمین الگو به روش‌های لاجیت

متغیر	ضرایب	انحراف معیار	آماره T	سطح معنی‌داری
عرض از مبدأ	۲/۲۹	۲/۵۸	۰/۸۹	۰/۳۸
جنسیت	۰/۱۷-	۰/۲۸	۰/۶۲-	۰/۵۳
سن	۰/۰۲-	۰/۰۱	۱/۴۸-	۰/۱۴
وضعیت تأهل	۰/۰۶-	۰/۰۹	۰/۶۳-	۰/۵۳
تحصیلات دانشگاهی	۰/۰۷-	۰/۲۹	۰/۲۵-	۰/۸۰
سابقه اعتباری	۰/۳۷	۰/۱۱	۳/۴۸	۰/۰۰
معدل ۶ ماهه حساب	۰/۱۲-	۰/۰۴	۳/۱۶-	۰/۰۰
وضعیت تملک مسکن	۰/۱۶-	۰/۲۷	۰/۵۸-	۰/۵۶
وضعیت اشتغال	۰/۷۰-	۰/۲۵	۲/۷۵-	۰/۰۱
سابقه کاری	۰/۰۰	۰/۰۲	۰/۱۹-	۰/۸۵
میزان اعتبار درخواستی	۰/۰۳	۰/۰۱	۲/۵۱	۰/۰۱
اقساط ماهیانه	۰/۰۵	۰/۰۲	۲/۳۲	۰/۰۲
مدت بازپرداخت	۰/۸۸-	۰/۳۱	۲/۸۸-	۰/۰۰

منبع: محاسبات پژوهش

نتایج نشان‌دهنده این است که متغیرهای: سابقه اعتباری، معدل ۶ ماهه حساب، وضعیت اشتغال، میزان اعتبار درخواستی، اقساط ماهانه، مدت بازپرداخت معنی‌دار بوده‌اند و تفسیر نتایج بدین‌صورت است که اشخاصی که دارای سابقه چک برگشتی هستند ریسک اعتباری آنها زیاد می‌باشد (مشتری بدحساب)، هرچقدر معدل ۶ ماه حساب متقاضی تسهیلات بیشتر باشد ریسک اعتباری آنها کمتر است (مشتری خوش‌حساب)، درصورتی‌که مشتری شاغل رسمی باشد ریسک اعتباری آن کمتر است (مشتری خوش‌حساب)، مبلغ تسهیلات و اقساط ماهانه هرچقدر بیشتر باشد ریسک اعتباری آن بیشتر است (مشتری بدحساب). و در نهایت چنانچه مدت باز بازپرداخت طولانی‌تر باشد ریسک اعتباری کمتر

است. درنهایت، پس از آزمون ترکیب‌های مختلفی از ۱۲ متغیر منتخب، ۶ متغیر که به لحاظ آماری معنادار نبودند، از مدل حذف شدند و مدل نهایی با ۶ متغیر اثرگذار بر ریسک اعتباری تصریح شد.

۴.۴ تعیین میزان نیکویی برازش مدل برآوردشده

در این پژوهش به دلیل اینکه متغیرهای باینری هستند، برای بررسی خوبی برازش مدل از آمارهٔ کای دو، آماره R^2 مک فادن و آمارهٔ هاسمر[1]ـ لمشو[2] استفاده گردیده است که خلاصه نتایج آن در جدول ذیل مشاهده می‌شود. آمارهٔ LR که شبیه آمارهٔ F در مدل رگرسیون خطی است، دارای توزیع کای دو با K درجه آزادی است. در این پژوهش مقدار این آماره ۲۱/۸۹ می‌باشد و احتمال آماره LR کمتر از ۰/۰۵ است بنابراین فرض صفر مبنی بر صفر بودن کلیهٔ ضرایب متغیرهای مستقل در سطح اطمینانی ۹۵٪ رد شده، درنتیجه رگرسیون معنی‌دار است. آماره R^2 مک‌فادن که مقدار آن بین صفر و یک تغییر می‌کند، خوبی برازش مدل اندازه‌گیری می‌شود. هرچه این شاخص به یک نزدیک‌تر باشد، میزان تطابق مدل با واقعیت بیشتر و به عبارتی نیکویی برازش بیشتر است؛ برعکس، هرچه مقدار شاخص به صفر نزدیک‌تر باشد، نیکویی برازش کمتر خواهد بود. آمارهٔ R^2 مک‌فادن در مدل تخمین زده شده ۰/۵۲ است؛ که این عدد برای رگرسیون لاجیت قابل قبول است. آماره هاسمرـ لمشو دارای توزیع χ^2 با Z-J درجهٔ آزادی است (J تعداد گروه‌ها است). مقدار این آماره با درجه آزادی ۸ مقدار آن ۷/۱۹ و احتمال آن ۰/۵۱ می‌باشد که بزرگ‌تر از ۱۰٪ به‌دست‌آمده است. (مقدار آماره هاسمرـ لیمر کمتر از ۱۶ باشد نشان از مناسب بودن مدل است). بنابراین، فرض صفر که بیانگر نیکویی برازش است، پذیرفته می‌شود (رد نمی‌شود) (هاسمر و لمشو، ۱۹۸۹.) پس متغیرهای مستقل، قدرت توضیح دهندگی میزان ریسک اعتباری را دارند.

[1] Hasmer, W. D
[2] Lemeshow, S.

جدول ۳

آزمون معنی‌داری و نکویی برازش مدل

معیار	مقدار آماره	سطح معنی‌داری
آماره LR	۲۱/۸۹	۰/۰۳
آماره هاسمر ـ لمشو	۷/۱۹	۰/۰۰
مقدار مک فادن	۰/۵۲	

منبع: محاسبات پژوهش

پس از بررسی آزمون‌های مختلف جهت معنی‌داری کلی مدل لاجیت به سنجش دقت و حساسیت پیش‌بینی درست مشتریان پرداخته می‌شود بر همین اساس از داده‌های شاهد جهت پیش‌بینی عملکرد مدل استفاده می‌شود. اطلاعات واقعی در خصوص متغیر وابسته برای این پرونده‌های تسهیلاتی در دسترس هست و بر اساس آن می‌توان عملکرد پیش‌بینی الگو در خارج از دوره تخمین را ارزیابی کرد.

پیش‌بینی‌های مورد انتظار برای داده‌های شاهد در جدول ذیل مشاهده می‌شود. این ۸۰ پرونده تسهیلاتی به‌طور واقعی دارای ۴۳ پرونده بدحساب و ۳۷ پرونده خوش‌حساب می‌باشد. که بر اساس پیش‌بینی مدل لاجیت از بین ۴۳ پرونده بدحساب ۲۸ پرونده را درست پیش‌بینی‌شده (۶۵/۱۱٪) و از بین ۳۷ پرونده خوش‌حساب ۲۵ پرونده درست پیش‌بینی کرده است (۶۷/۵۶٪). درجه حساسیت این مدل ۷۰٪ و تشخیص ۶۲/۵٪ می‌باشد. بنابراین در مجموع از ۸۰ پرونده ۵۳ مورد را درست پیش‌بینی کرده است لذا در مجموع درصد پیش‌بینی درست این مدل ۶۶/۲۵ درصد می‌باشد.

جدول ۵

نتایج سنجش کارایی مدل لاجیت در داده‌های شاهد

گروه	موارد تشخیص به‌عنوان بدحساب	موارد تشخیص به‌عنوان خوش‌حساب	کل (واقعی)
بدحساب (۱)	۲۸	۱۵	۴۳
خوش‌حساب (۰)	۱۲	۲۵	۳۷
مجموع (مدل)	۴۰	۴۰	۸۰
درجه حساسیت		٪۷۰/۰۰	
درجه تشخیص		٪۶۲/۵۰	
درصد پیش‌بینی درست بدحساب		٪۶۵/۱۱	
درصد پیش‌بینی درست خوش‌حساب		٪۶۷/۵۶	
درصد پیش‌بینی درست		٪۶۶/۲۵	

منبع: محاسبات پژوهش

۵.۴ طراحی مدل شبکهٔ عصبی

با توجه به توانایی مدل شبکه عصبی در شناسایی متغیرهای زائد و انتخاب متغیرهای مهم در فرایند مدل‌سازی، تمام متغیرها (۱۲ متغیر) را در مدل‌سازی مورد استفاده قرار می‌دهیم. در این پژوهش برای پیش‌بینی ریسک اعتباری؛ یک شبکهٔ عصبی تغذیه به جلوی (پیشخور) ۳ لایه با یک لایهٔ مخفی و الگوریتم یادگیری پس انتشار خطا را که با متغیرهای پیش‌بین شناسایی‌شده توسط مدل لاجیت تغذیه می‌شود به کار می‌گیریم. برای تعیین شمار بهینهٔ نرون‌ها در لایهٔ مخفی آموزش (۳۱۹ مشاهده) را با یک نرون در لایهٔ مخفی آغاز نموده و در هر بار تکرار یادگیری، یک نرون به لایهٔ مخفی اضافه می‌نماییم و این رویه را تا زمانی ادامه می‌دهیم که شاخص تعیین‌شده برآورده گردد. محاسبات برای داده‌های آموزشی نشان داد که ۲ نقطهٔ ۰/۵۱ و ۰/۵۲ می‌توانند به‌عنوان نقطهٔ بهینهٔ مدل برگزیده شوند که در این نقاط خطای طبقه‌بندی کلی مدل کمینه است. با دسته‌بندی انجام‌شده نقطه بهینه ۰/۵۱ در نظر گرفته شده است. و مدل شبکه عصبی طراحی گردید جهت سنجش کارایی مدل فوق در تشخیص دقیق نکول از داده‌های شاهد[1] استفاده شده است.

نتایج سنجش کارایی مدل شبکه عصبی در داده‌های شاهد در جدول ذیل مشاهده می‌شود. بر اساس عملکرد پیش‌بینی مدل شبکه عصبی از بین ۴۳ پرونده بدحساب ۳۴ مورد را درست پیش‌بینی‌شده (٪۷۹/۰۶) و از بین ۳۷ پرونده خوش‌حساب ۳۰ پرونده درست

[1] تعداد داده‌های شاهد ۸۰ پرونده تسهیلاتی می‌باشد که در هر سه مدل این پرونده‌ها یکسان است.

پیش‌بینی کرده است (۸۱/۰۸٪). درجه حساسیت این مدل ۸۲/۹۲٪ و درجه تشخیص ۷۶/۹۲٪ می‌باشد. بنابراین در مجموع از ۸۰ پرونده ۶۴ مورد را درست پیش‌بینی کرده است بنابراین در مجموع درصد پیش‌بینی درست این مدل ۸۰/۰۰٪ درصد می‌باشد.

جدول ۷

نتایج سنجش کارایی مدل شبکه عصبی در داده‌های شاهد

گروه	موارد تشخیص به‌عنوان بدحساب	موارد تشخیص به‌عنوان خوش‌حساب	کل
نکول (۱)	۳۴	۹	۴۳
خوش‌حساب (۰)	۷	۳۰	۳۷
مجموع	۴۱	۳۹	۸۰
درجه حساسیت		۸۲/۹۲٪	
درجه تشخیص		۷۶/۹۲٪	
درصد پیش‌بینی درست بدحساب		۷۹/۰۶٪	
درصد پیش‌بینی درست خوش‌حساب		۸۱/۰۸٪	
درصد پیش‌بینی درست		۸۰/۰۰٪	

منبع: محاسبات پژوهش

۶.۴ طراحی مدل برنامه‌ریزی ژنتیک

در اینجا هدف طراحی مدل ژنتیکی است که قادر باشد به‌طور خودکار، بر اساس متغیرهای پیش‌بین معنی‌دار تشخیص داده شده توسط مدل لاجیت ــ سابقه اعتباری، معدل شش‌ماهه حساب، وضعیت اشتغال، میزان اعتبار درخواستی، اقساط ماهانه و مدت بازپرداخت ــ، نسبت به دسته‌بندی مشتریان به دو گروه خوش‌حساب و بدحساب اقدام نماید. با توجه به اینکه متغیرهای پیش‌بین و وابسته، متغیرهایی دودویی می‌باشند لذا با استفاده از مدل برنامه‌ریزی ژنتیک به پیاده‌سازی یک درخت تصمیم می‌پردازیم، تابع تناسب به کار گرفته شده، صحت پیش‌بینی مدل می‌باشد، حاصل برازش مدل مجموعه‌ای از جملات شرطی بر اساس شش متغیر معنی‌دار که خوش‌حسابی یا بدحسابی مشتری را نتیجه خواهند داد. در اجرای برنامه‌ریزی ژنتیک نیز همچون رویه اتخاذشده در بنای دو مدل قبلی پژوهش، برازش مدل که منجر به شکل‌گیری کروموزوم نهایی می‌گردد با استفاده از داده‌های مدل آموزشی (۳۱۹ پرونده تسهیلاتی) صورت گرفت و برای آزمون مدل نیز، کروموزوم حاصل به داده‌های شاهد (۸۰ پرونده) اعمال گردید.

نتایج سنجش کارایی مدل ژنتیک در داده‌های شاهد در جدول ذیل مشاهده می‌شود بر اساس عملکرد پیش‌بینی مدل شبکه عصبی از بین ۴۳ پرونده بدحساب ۳۲ نفر را درست

پیش‌بینی‌شده (۷۴/۴۱٪) و از بین ۳۷ پرونده خوش‌حساب ۲۷ پرونده درست پیش‌بینی کرده است (۷۲/۹۷٪) درجه حساسیت این مدل ۷۴/۰۰٪ و درجه تشخیص ۷۱/۰۵٪ می‌باشد. بنابراین در مجموع از ۸۰ پرونده ۵۹ مورد را درست پیش‌بینی کرده است بنابراین در مجموع درصد پیش‌بینی درست این مدل ۷۳/۷۵ درصد می‌باشد.

جدول ۷

نتایج سنجش کارایی مدل ژنتیک در داده‌های شاهد

گروه	موارد تشخیص به‌عنوان بدحساب	موارد تشخیص به‌عنوان خوش‌حساب	کل
نکول (۱)	۳۲	۱۱	۴۳
خوش‌حساب (۰)	۱۰	۲۷	۳۷
مجموع	۴۲	۳۸	۸۰
درجه حساسیت		۷۴/۰۰٪	
درجه تشخیص		۷۱/۰۵٪	
درصد پیش‌بینی درست بدحساب		۷۴/۴۱٪	
درصد پیش‌بینی درست خوش‌حساب		۷۲/۹۷٪	
درصد پیش‌بینی درست		۷۳/۷۵٪	

منبع: محاسبات پژوهش

۵ نتیجه‌گیری و پیشنهادات

سیستم‌های بانکی یکی از مهم‌ترین عوامل مؤثر بر رشد اقتصادی یک کشور می‌باشند؛ بانک‌ها به‌عنوان بخش اصلی نظام مالی، بار اصلی تأمین مالی مشتریان حقیقی و حقوقی را به دوش می‌کشند؛ در سلسله عملیات بانک، چنانچه بانک‌ها بتوانند منابع جمع‌آوری‌شده را به نحوی مطلوب به دریافت‌کنندگان تسهیلات اعطا نمایند، علاوه بر کمک به رشد و توسعه اقتصادی، موجب برگشت صحیح منابع، سوددهی و موفقیت خویش خواهند گردید. سنجش صحیح ریسک اعتباری به بانک‌ها امکان می‌بخشد تراکنش‌های وام‌دهی آتی خویش را به‌گونه‌ای مهندسی نمایند که مشخصه‌های ریسک و بازده هدف‌گذاری شده را تحقق بخشند؛ همین امر موجب گردیده است که ریسک اعتباری به موضوعی مهم و به صورتی گسترده مورد مطالعه قرارگرفته توسط پژوهشگران و مدیران بانک‌ها تبدیل گردد. تاکنون الگوهای متنوعی برای پیش‌بینی ریسک اعتباری مشتریان بانکی ارائه گردیده است که از جمله آنان می‌توان به مدل‌های آماری مبتنی بر تحلیل ممیزی تک متغیره، تحلیل ممیزی چند متغیره، رگرسیون خطی، تحلیل رگرسیون لاجیت و پروبیت اشاره نمود؛ همچنین با پیشرفت قابل ملاحظه

علوم رایانه و ریاضی، و ابداع و گسترش حوزه هوش مصنوعی الگوهای پیشرفته‌ای نظیر شبکه‌های عصبی مصنوعی، الگوریتم ژنتیک، ماشین بردار پشتیبان و... ارائه گردیده و به کار گرفته شده‌اند. پژوهش پیش‌رو به‌منظور بررسی کارایی (جهت سنجش دقت پیش‌بینی صحیح) سه مدل لاجیت، شبکه عصبی و الگوریتم ژنتیک در تعیین اعتبار سنجی مشتریان ارائه شده است. با توجه به جدول ذیل مدل شبکه عصبی دقت پیش‌بینی بهتری نسبت به مدل لاجیت و ژنتیک دارد و همچنین مدل ژنتیک قدرت پیش‌بینی بهتری نسبت به مدل لاجیت دارد. همچنین براساس آزمون لاجیت متغیرهای؛ سابقه اعتباری، معدل شش‌ماهه حساب، وضعیت اشتغال، میزان اعتبار درخواستی، اقساط ماهانه و مدت بازپرداخت تأثیر معنی‌دار بر نکول داشته‌اند و تفسیر نتایج بدین‌صورت است که اشخاصی که دارای سابقه چک برگشتی، مبلغ بالای تسهیلات و افزایش میزان اقساط هستند ریسک اعتباری آنها زیاد می‌باشد و هر چقدر معدل ۶ ماه حساب متقاضی تسهیلات بیشتر، مشتری شاغل رسمی و یا مدت بازپرداخت طولانی باشد ریسک اعتباری آنها کمتر است.

جدول ۸

جمع‌بندی نتایج پژوهش

	لاجیت	شبکه عصبی	ژنتیک
درجه حساسیت	٪۷۰/۰۰	٪۸۲/۹۳	٪۷۴/۰۰
درجه تشخیص	٪۶۲/۵۰	٪۷۶/۹۳	٪۷۱/۰۵
درصد پیش‌بینی درست	٪۶۶/۲۵	٪۸۰/۰۰	٪۷۳/۷۵

منبع: محاسبات پژوهش

با توجه به نتایج پژوهش وجود بانک اطلاعاتی یکپارچه برای بخش تسهیلات (حقیقی و حقوقی) ضروریست که پیشنهاد می‌گردد بانک مرکزی با کمک سایر بانک‌ها چنین سیستم راه‌اندازی کنند و ضروریست که بانک مرکزی موظف است به این سیستم کنترل داشته باشد و به شفافیت اطلاعات و نظارت اقدام کند، زیرا که وجود چنین بانک اطلاعاتی می‌تواند به اجرای صحیح اعتبارسنجی کمک کند، پیشنهاد اصلی این پژوهش این است که سیستم اعتبار سنجی بر اساس مدل شبکه‌های عصبی طراحی شود، و همواره مدل‌های مختلفی که با پیشرفت علم تدوین می‌شوند مورد آزمون قرارگیرند لازمه چنین آزمون‌هایی داشتن اطلاعات شفاف و به‌روز می‌باشد که وجود بانک اطلاعاتی پیش از پیش ضروریست. همچنین با توجه به نتایج آزمون لاجیت که افزایش مبلغ تسهیلات احتمال ریسک اعتباری را زیاد می‌کند پیشنهاد می‌گردد که سقف تسهیلاتی برای مشتریان حقیقی در نظر گرفته شود و براساس

سایر متغیرهای تأثیرگذار بر ریسک اعتباری از جمله معدل شش‌ماهه، وضعیت اشتغال و... میزان تسهیلات و مدت بازپرداخت تغییر کند. لازم به ذکر است ازآنجاکه نتایج حاصل از این تحقیق به دلیل محدودیت جامعه آماری و متغیرهای استفاده‌شده پژوهش، قابل تعمیم به کل سیستم بانکی کشور نیست، (البته هدف پژوهش حاضر مقایسه سه مدل لاجیت، شبکه عصبی و الگوریتم ژنتیک بود، و اینکه اثبات شود که گسترش دانش و تحقیقات در زمینه مدیریت ریسک اعتباری، خدمات شایسته و ارزشمندی را برای بانک‌ها و مؤسسات مالی و اعتباری به همراه خواهد داشت. انجام این تحقیق کاربردی هم گام کوچکی در این جهت می‌باشد) پیشنهاد می‌شود در تحقیقات آینده بر اساس الگوهای ارائه شده، مدل‌های مشابهی در سطح گسترده‌تر برآورد شود و دامنه وسیع‌تری از متغیرها، جهت بالا بردن سطح اطمینان و صحت پیش‌بینی، مورد برازش قرار گیرد.

فهرست منابع

اخباری، م.، و مخاطب رفیعی، ف. (۱۳۸۹). کاربرد سیستم‌های استدلال عصبی-فازی در رتبه‌بندی اعتباری مشتریان حقوقی بانک‌ها، *مجله تحقیقات اقتصادی*. شماره ۹۲، ۲۱-۱

تهرانی، ر.، و فلاح شمسی، م. (۱۳۸۴). طراحی و تبیین مدل ریسک اعتباری در نظام بانکی کشور، *مجله علوم اجتماعی و انسانی دانشگاه شیراز*. شماره ۴۳، ۴۵-۶۰

فیروزیان، م.، جاوید، د.، و نجم‌الدینی، ن. (۱۳۹۰). کاربرد الگوریتم ژنتیک در پیش‌بینی ورشکستگی و مقایسه آن با مدل Z آلتمن در شرکت‌های پذیرفته‌شده در بورس اوراق بهادار تهران، *فصلنامه علمی-پژوهشی بررسی‌های حسابداری و حسابرسی*. ۱۸(۶۵)، ۹۹-۱۱۴.

کریمی، ز. و همکاران (۱۳۹۴). عوامل مؤثر بر ریسک اعتباری مشتریان بانک‌های تجاری (مطالعه موردی: بانک تجارت شهر نکا- استان مازندران)، *دوفصلنامه اقتصاد پولی و مالی (دانش و توسعه سابق)*. دوره جدید، سال بیست و دوم، شماره ۱۰، پاییز و زمستان ۱۳۹۴.

کشاورز حداد، غ.، و آیتی گازار، ح. (۱۳۸۶). مقایسه کارکرد مدل لاجیت و روش درخت‌های طبقه‌بندی و رگرسیونی در فرآیند اعتبار سنجی متقاضیان حقیقی برای استفاده از تسهیلات بانکی، *فصلنامه علمی پژوهشی پژوهش‌های اقتصادی (رشد و توسعه پایدار)*. ۷(۴)، ۷۱-۹۷.

مهرآرا، م.، موسایی، م.، تصوری، م.، و حسن‌زاده، آ. (۱۳۹۰) رتبه‌بندی اعتباری مشتریان حقوقی بانک پارسیان، *فصلنامه مدل‌سازی اقتصادی*. سال سوم، ش. ۳، ۱۲۱-۱۵۰

نبوی چاشمی، س. ع.، احمدی، م.، و مهدوی فرح‌آبادی، ص. (۱۳۸۹). پیش‌بینی ورشکستگی شرکت‌ها با استفاده از مدل لاجیت. *مهندسی مالی و مدیریت اوراق بهادار*. ۲(۵)، ۵۵-۷۹.

Abdou, H., Pointon, J., & Masry, E. I. (2008). A. Neural nets versus conventional techniques in credit scoring in Egyptian banking. *Expert Systems with Applications*. 35, 1275-1292.

Altman, E. I. (1968). Financial ratios discriminate analysis and the prediction of corporate bankruptcy. *The Journal of Finance*. 23, 589-609.

Angelini, E., di Tollo, G., & Roli, A. (2008). A neural network approach for credit risk valuation. *The quarterly review of economics and finance*. 48(4), 733-755.

Bellotti, T. (2010). A simulation study of Basel II expected loss distributions for a portfolio of credit cards, *Journal of Financial.*

Glantz, M. (2003). *Managing Bank Risk: An introduction to broad-base credit engineering* (Vol. 1). Academic press.

Hasmer, W. D., & Lemeshow, S. (1989). *Applied Logistic Regression*. Wiley.

Kim, K. J., & Ahn, H. (2012). A corporate credit rating model using multi-class support vector machines with an ordinal pairwise partitioning approach, *Computers & Operations Research*. 39(8), 1800-1811.

Hertz, J., Krogh, A., & Palmer, R. G. (1991). *Introduction to the Theory of Neural Computation*. Addison-Wesley, Redwood City, CA,

Roszbach, K. (1998). *Bank Lending Policy, Credit Scoring and the Survival of the Loan*, Doctoral Thesis School of Business and Economics, Stokholm Uni.

Saunders, A., & Allen, L. (2002). *Credit Risk Measurement*. Second Edition, New York: John Wiley & Sons.

Shin, K., & Lee, Y. (2002). A genetic algorithm application in bankruptcy prediction modelling, *Expert Systems with Applications*. Vol. 23 No.3, 321-8

Udo, G. (1993). Neural network performance on the bankruptcy classification problem, *Computers and Industrial Engineering*. 27(1), 445-448, vol. 54, no. 1, 78-94

Varetto, F. (1998). Genetic algorithms applications in the analysis of insolvency risk, *Journal of Banking & Finance*. 22(10), 1421-1439.