Journal of Economic Modeling Research
Print ISSN: 2228 - 6454
Online ISSN: 2538 - 4163
Homepage: https://jemr.khu.ac.ir/

Credit Rating of Real Bank Customers Based on Logistic Regression and MPLE Logistic Methods and Neural Network

Arezoo Chaghaei [1] | Seyed ahmad Ameli [2*] | Hossein Rezaee [3]

1. Graduate with a Master's degree of Social-Economic Systems (Economic Systems Planning Orientation), Faculty of Economics, Kharazmi University, Tehran, Iran
2. Corresponding Author, Assistant Professor of Resource and energy economy, Faculty of economics, Kharazmi University, Tehran, Iran, E-mail: ameli@khu.ac.ir (0000-0003-3347-9735)
3. Assistant Professor, Department of Economics, Payam Noor University, Tehran, Iran. E-mail: hrezaee@pnu.ac.ir (0000-0001-5689-9543)

Article Info

Article type:
Research Article

Article history:
Received: 03 Aug. 2025

Received in revised form:
17 Apr. 2026

Accepted: 23 May. 2026

Keywords:
Credit rating,
neural network,
logistics,
MPLE logistics,
Rock curve,
data mining

JEL:
C53-C55- G21

ABSTRACT

By designing an efficient loan management system, banks can increase efficiency and reduce the probability of non-repayment of principal and sub-loans. In this paper, the efficiency of logistic regression models, artificial neural network, was examined to predict the credit risk of real customers or in other words, applicants for microloans, which include a large group of customers in the country's banking system. Given the imbalance of the number of data, the optimal threshold was calculated using two sensitivity and detection curves, and the credit risk of each model was extracted from this method. In logistic regression, the compensated maximum likelihood method was used to estimate the coefficients considering the small number of bad customers instead of the maximum likelihood method. Finally, the accuracy and precision of each model was examined with multiple criteria. Using the Rock curve, the resolution of the models was examined, where the neural network model had the best resolution. Then, by comparing the MSE, RMSE and MAE errors, the efficiency of the methods was compared, and the performance of MPLE logistics and neural network is almost the same. Finally, considering the bank's goal in three scenarios of minimum credit risk, identifying good customers and separating customers, neural network, MPLE logistics, and in the third scenario, neural network and MPLE logistics simultaneously have been selected as the best models.

Cite this article: Chaghaei , Arezoo., Ameli, Seyed ahmad., & rezaee, hossein . (2025). Credit Rating of Real Bank Customers Based on Logistic Regression and MPLE Logistic Methods and Neural Network. *Journal of Economic Modeling Research*, 16 (59), 154-199.
DOI: 0000000000000000000

© The Author(s).
Publisher: Kharazmi University
DOI: 00000000000000000000000000
Journal of Economic Modeling Research, Vol, 16, No. 59, 2025, pp. 154-199.

Journal of Economic Modeling Research
Kharazmi University
Print ISSN: 2228 - 6454
Online ISSN: 2538 - 4163
Homepage: https://jemr.khu.ac.ir/

رتبه بندی اعتباری مشتریان حقیقی بانک مبتنی بر روش‌های رگرسیون لجستیک و لجستیک MPLE و شبکه عصبی

آرزو چقایی[۱] | سید احمد عاملی*[۲] | حسین رضائی[۳]

۱. دانش‌آموخته کارشناسی ارشد ، گروه آموزشی اقتصاد امور عمومی، دانشکده اقتصاد، دانشگاه خوارزمی، تهران، ایران.

۲. نویسنده مسئول، سید احمد عاملی، استادیار گروه اقتصاد منابع و انرژی، دانشکده اقتصاد ،دانشگاه خوارزمی،تهران،ایران.

رایانامه: ameli@khu.ac.ir (0000-0003-3347-9735)

۳. استادیار، گروه اقتصاد، دانشگاه پیام نور، تهران، ایران . رایانامه: hrezaee@pnu.ac.ir (0000-0001-5689-9543)

اطلاعات مقاله

نوع مقاله:
مقاله پژوهشی

تاریخ دریافت:
۱۴۰۴/۰۵/۱۲

تاریخ ویرایش:
۱۴۰۵/۰۱/۲۸

تاریخ پذیرش :
۱۴۰۵/۰۳/۰۲

واژه‌های کلیدی:
رتبه بندی اعتباری،
شبکه عصبی،
لجستیک،
لجستیک MPLE ،
منحنی راک،داده کاوی

طبقه‌بندی JEL:
G32, C53, C58

چکیده

بانک‌ها می‌توانند با طراحی یک سیستم کارآمد مدیریت وام، کارایی را افزایش داده و احتمال عدم برگشت اصل و فرع وام کاهش دهند.. در این مقاله کارآیی مدل‌های رگرسیون لجستیک، شبکه عصبی مصنوعی، به منظور پیش‌بینی ریسک اعتباری مشتریان حقیقی یا به عبارتی متقاضیان وام‌های خرد که گروه زیادی از مشتریان نظام بانکی کشور را شامل می‌شوند، مورد بررسی قرار گرفت. باتوجه به نامتعادل بودن تعداد داده‌ها حدآستانه بهینه با بکارگیری دو منحنی درجه حساسیت و درجه تشخیص محاسبه شده و از این روش میزان ریسک اعتباری هریک از مدل‌ها استخراج شد. در رگرسیون لجستیک برای برآورد ضرایب با توجه به تعداد اندک مشتریان بدحساب بجای روش حداکثر درستنمایی از روش حداکثر درستنمایی تاواندیده استفاده شد. در نهایت میزان صحت و دقت هر مدل با معیارهای متعدد بررسی شد. با استفاده از منحنی راک به بررسی قدرت تفکیک کنندگی مدل‌ها پرداخته که در اینجا مدل شبکه عصبی بهترین قدرت تفکیک کنندگی را دارا بود. سپس با مقایسه خطاهای MSE، RMSE و MAE کارایی سنجش روش‌ها مورد مقایسه قرار گرفت و عملکرد لجستیک MPLE و شبکه عصبی تقریبا با یکدیگر یکسان است. و در نهایت با توجه به هدف بانک در سه سناریو حداقل ریسک اعتباری، تشخیص مشتریان خوش حساب و تفکیک مشتریان به ترتیب شبکه عصبی، لجستیکMPLE و در سناریوی سوم شبکه عصبی و لجستیکMPLE بطور همزمان به عنوان مدل‌های برتر انتخاب شده اند.

استناد: چقایی ، آرزو ؛ عاملی ، سید احمد ؛ و رضائی ، حسین (۱۴۰۴). رتبه بندی اعتباری مشتریان حقیقی بانک مبتنی بر روش‌های رگرسیون لجستیک و لجستیک MPLE و شبکه عصبی. *تحقیقات مدل‌سازی اقتصادی*، ۱۶ (۵۹)، ۱۵۴-۱۹۹.
DOI: 00000000000000000000

ناشر: دانشگاه خوارزمی. © نویسندگان.

۱. مقدمه

مهمترین عملیات بانک‌ها و موسسات مالی و اعتباری، اعطای تسهیلات به متقاضیان و انجام تعهدات آنان است. بر این اساس موسسات ناچار به استقرار یک سیستم کارآمد هستند تا عملیات اعطای تسهیلات در بازارهای رقابتی کنونی هم از کارایی و سرعت لازم برخوردار باشد و هم احتمال عدم برگشت اصل و فرع تسهیلات اعطا شده به حداقل ممکن کاهش یابد. از این منظر یکی از کاربردهای اساسی سیستم‌های رتبه بندی اعتباری روشن می‌شود. استفاده از رتبه بندی اعتباری می‌تواند به بانک برای اعطای تسهیلات با اطمینان بیشتر کمک کند زمانی که بانک‌ها قصد دارند به مشتریان وام یا تسهیلاتی پرداخت کنند، وضعیت مشتری را از منظر خوش حسابی یا بدحسابی مشتری مورد نظر می‌سنجند و پس از به دست آوردن اطلاعات لازم اقدام به تصمیم‌گیری در این خصوص می‌کنند. بنابراین می‌توان اذعان داشت که اعتبار سنجی مشتریان از این جهت حائز اهمیت است که می‌تواند بر دو وظیفه اصلی بانک‌ها یعنی تجهیز و تخصیص منابع اثرگذار باشد.

اعطای تسهیلات بانک‌ها را با ریسک اعتباری مواجه می‌کند (توران، ۲۰۱۶، ص.۱)[۱]. ریسک اعتباری مهمترین نوع ریسک در تراکنش‌های مالی، تجاری و بازرگانی محسوب می‌شود. ریسک اعتباری در بیشتر مواقع به عنوان ریسک ناشی از نکول در بازپرداخت وام‌ها از سوی وام گیرندگان تعریف شده و طرف دیگر قرارداد برخلاف مفاد و شرایط قرارداد رفتار می‌کند و همین امرسبب ایجاد زیان مالی به صاحب دارایی می‌شود (اسپوچاکوا و کاگ، ۲۰۱۵، ص.۲)[۲]. نکته قابل تامل این است که بانک‌ها نیز به ویژه در برابر ریسک‌های اعتباری آسیب پذیر هستند و غالبا با انبوهی از بدهی‌های بد مواجه هستند. درنتیجه باید از روش‌های جدید برای پیشگیری و مقابله با این نوع بدهی‌ها استفاده کنند.

اعتبار سنجی مشتریان در ارائه تصویری شفاف از وضعیت و توانایی مشتری در ایفای به موقع تعهدات، بازداشتن آنان از استفاده بیش از حد منابع و سقوط به ورطه بحران مالی نقش اساسی دارد. با کمک این روش می‌توان رفتار مشتریان را در قالب افراد خوش حساب و بدحساب رتبه

---

[۱] -Turan

[۲] -Spuchakova and Cug

بندی کرد. این کار سبب کاهش ریسک اعتباری، افزایش سودآوری برای بانک و جذب مشتریان خوب می شود. روش‌های پارامتریک مثل پروبیت، لاجیت، تحلیل تمایزی و رگرسیون لاجستیک از ابتدای ظهور اعتبار سنجی مورد استفاده بوده است. به تدریج روش‌های ناپارامتریک و داده کاوی مثل درختان تصمیم گیری، شبکه‌های عصبی و سیستم‌های خبره به کار گرفته شدند.

با در نظر گرفتن اینکه اعتبار سنجی مشتریان در تسهیلات خرد ،امری اجتناب ناپذیر بوده زیرا هم از نظر تعداد ،عدد بزرگی را به خود اختصاص می‌دهند و هم از حیث شناخت فردی امکان پذیر نبوده، لذا در شبکه بانکی به طور عام و در بانک‌های مختلف بصورت خاص نیاز به طراحی سیستم اعتبار سنجی که بتواند در ضمن سود آوری متضمن کاهش ریسک اعتباری ناشی از نکول پرداخت باشد،ضروری خواهد بودو بنابراین انتظار می‌رود که مشتریان خوش حساب و بد حساب بتوانند با دقت مناسب دراین سیستم تفکیک شوند به نحوی که ریسک اعتباری ناشی از نکول به حداقل خود برسد.

در ادامه ابتدا مبانی نظری تبیین می‌شود سپس پیشینه تحقیق بیان بیان شده و در انتهای این قسمت شکاف تحقیق استخراج شده و با بیان روش تحقیق و ارائه الگوهای سنجی مناسب به اجرای مدل پرداخته و نتایج و جمعبندی حاصل از اجرای مدل بخش پایانی مقاله را تشکیل می‌دهد.

۲. مبانی نظری

ریسک اعتباری در چارچوب‌های نوین مدیریت ریسک بانکی به عنوان احتمال وقوع زیان ناشی از نکول مشتری تعریف می‌شود که در قالب شاخص احتمال نکول [۱]، زیان در صورت نکول[۲]و میزان مواجهه در زمان نکول [۳]مدل‌سازی می‌شود. در چارچوب استانداردهای نظارتی بین‌المللی

۱- Probability of Default

۲- Loss Given Default

۳- Exposure at Default

مانند کمیته نظارت بانکی بازل[۱] ، اندازه‌گیری دقیق شاخص احتمال نکول یکی از ارکان اساسی مدیریت سرمایه بانکی محسوب می‌شود [۲].

این ریسک ریشه در نظریه عدم تقارن اطلاعاتی دارد که توسط استیگلیتز و وییس[۳] (1981) تبیین شد. در چنین محیطی، بانک با دو مسئله انتخاب نامساعد و مخاطره اخلاقی مواجه است که کارایی تخصیص اعتبارات را کاهش می‌دهد. در بازارهای اعتباری، وجود عدم تقارن اطلاعاتی میان وام‌دهنده و وام‌گیرنده موجب بروز پدیده‌هایی نظیر انتخاب نامساعد و مخاطره اخلاقی شده و فرآیند تخصیص بهینه منابع مالی را با چالش مواجه می‌کند. در چنین شرایطی، بانک‌ها ناگزیر به استفاده از سازوکارهای غربالگری اطلاعاتی به منظور تفکیک متقاضیان اعتباری بر اساس سطح ریسک هستند. این فرآیند غربالگری در قالب سیستم‌های امتیازدهی اعتباری پیاده سازی می شود که هدف آن‌ها برآورد احتمال نکول مشتریان و طبقه‌بندی آن‌ها در گروه‌های ریسک متفاوت است (بایسنس،۲۰۲۱). [۴]از آنجا که امتیازدهی اعتباری مستلزم مدل سازی رابطه میان ویژگی‌های فردی و مالی مشتریان با احتمال نکول است، استفاده از مدل‌های پیش‌بینی آماری و یادگیری ماشین نظیر رگرسیون لجستیک، روش لجستیک MPLE و شبکه‌های عصبی مصنوعی به عنوان ابزارهای اصلی در توسعه سیستم‌های رتبه‌بندی اعتباری مورد توجه قرار گرفته است(لوزادا، ۲۰۲۰)[۵].

امتیازدهی اعتباری به‌عنوان ابزاری برای پیش‌بینی ریسک نکول مشتری و راهنمایی در تصمیمات اعطای تسهیلات به کار گرفته می شود. در عمل، نهادهای مالی از مدل‌های کمّی برای تفکیک متقاضیان «ریسک‌پذیر» از «غیرریسک‌پذیر» استفاده می‌کنند و احتمال نکول هر مشتری را پیش‌بینی می‌نمایند. به‌عنوان مثال، Xia و همکاران (۲۰۲۲) خاطرنشان می‌کنند که مدل‌های داخلی

۱- Basel Committee on Banking Supervision

۲- BCBS, 2019

۳- Joseph Stiglitz & Andrew Weiss, 1981

۴- Baesens et al., 2021

۵- Louzada et al., 2020

رتبه‌بندی اعتباری مسأله اعطای وام را به یک «مسأله طبقه‌بندی باینری» تبدیل می‌کنند که هدف آن پیش‌بینی احتمال نکول است. در این چارچوب، امتیاز اعتباری[۱] عموماً با استفاده از اطلاعاتی نظیر ویژگی‌های دموگرافیک مشتری، سوابق اعتباری، درآمد و دارایی‌ها محاسبه شده و به‌عنوان معیاری برای سنجش احتمال نکول وام در نظر گرفته می‌شود.

در تبیین جایگاه مدل‌های انتخابی در مدل‌های کلاسیک و پیشرفته طبقه‌بندی اعتباری می توان گفت مدل‌های اولیه مانند «معامله‌گر استاندارد ز-سکور»[۲] بر تحلیل نسبت‌های مالی تمرکز داشتند تا احتمال ورشکستگی شرکت‌ها را پیش‌بینی کنند. اما در حوزه اعتباری بانکی، رگرسیون لجستیک به‌تدریج به ابزار اصلی تبدیل شد؛ از آنجا که این مدل نه تنها احتمال نکول را تخمین می زند بلکه شفافیت و تفسیرپذیری نتایج را تضمین می‌کند. مطالعات مروری و تجربی نشان می‌دهند که «لجستیک» همچنان متداول‌ترین روش در سیستم‌های سنتی امتیازدهی است. در عین حال، به‌کارگیری روش‌های یادگیری ماشین نظیر درخت تصمیم، جنگل تصادفی، بوستینگ و شبکه‌های عصبی گسترش یافته است؛ چرا که این روش‌ها در بسیاری از موارد توانایی پیش‌بینی بالاتری نسبت به رگرسیون نشان داده‌اند.

چالش مهم در اعتبارسنجی مدرن، مسئله عدم توازن کلاس‌ها است، زیرا نسبت مشتریان نکول‌کننده معمولاً بسیار کمتر از مشتریان خوش‌حساب است. مطالعات اخیر نشان می‌دهند که در چنین شرایطی، روش‌های اصلاح‌شده مانند رگرسیون لوجستیک اول یا MPLE می‌توانند تورش برآورد را کاهش دهند و پایداری مدل را افزایش دهند.(پاولو و دیمیتریو، ۲۰۲۰)[۳]

از سوی دیگر، پیشرفت‌های اخیر در یادگیری ماشین نشان داده‌اند که مدل‌های شبکه عصبی عمیق در صورت وجود روابط غیرخطی پیچیده، می‌توانند دقت پیش‌بینی بالاتری نسبت به

---

۱ -Credit Score

۲- Altman Z-score

۳- Pavlou & Dimitriou, 2020

مدل‌های سنتی ارائه دهند، هرچند مسئله تفسیرپذیری همچنان چالش اصلی آن‌ها محسوب می‌شود. (شیا و همکاران، ۲۰۱۸- لوزادا و همکاران، ۲۰۲۰)[۱]

بنابراین، استفاده همزمان از مدل‌های تفسیرپذی لجستیک و MPLE و مدل‌های با قدرت پیش‌بینی بالا (شبکه عصبی)، رویکردی نظری-کاربردی برای کاهش ریسک اعتباری در شرایط عدم تقارن اطلاعاتی محسوب می‌شود

۳. پیشینه تحقیق

ا سمیت و براون (۲۰۲۲)[۲] این مطالعه به برر سی کارایی رگر سیون لج ستیک در پیش‌بینی ری سک اعتباری پرداخته است. نویسندگان با استفاده از داده‌های واقعی مشتریان یک بانک بزرگ، عملکرد رگرسیون لجستیک را در مقایسه با مدل‌های سنتی ارزیابی کرده‌اند. یافته‌ها نشان می‌دهد که رگرسیون لجستیک در تمایز بین مشتریان با ریسک بالا و پایین، به‌ویژه در شرایطی که توزیع داده ها نامتقارن است، بسیار مؤثر است. یکی از مزایای این روش، سادگی در تفسیر نتایج و قابلیت پیاده سازی آن در محیط های مالی ا ست. همچنین، پژوه شگران به اهمیت تنظیم منا سب متغیرها و انتخاب معیارهای ارزیابی مناسب اشاره کرده‌اند تا بتوان از مزایای رگرسیون لجستیک به‌طور کامل بهره برد.

وانگ و ژائو (۲۰۲۲)[۳] این مقاله به بررسی یک روش اصلاح شده از رگرسیون لجستیک با نام MPLE(رگرسیون لجستیک با بیشینه احتمال جزئی) پرداخته است که برای بهبود ارزیابی ریسک اعتباری طراحی شده است. نویسندگان نشان می‌دهند که MPLE نسبت به رگر سیون لجستیک سنتی، دقت بالاتری در پیش‌بینی رفتار اعتباری مشتریان دارد. این مطالعه همچنین به بررسی میزان خطای مدل و معیارهای ارزیابی عملکرد پرداخته و نشان داده است که MPLE در

۱- Xia et al., 2018; Louzada et al., 2020

۲ Smith, J., & Brown, A

۳ Wang, L., & Zhao, Q

موقعیت‌های با حجم داده زیاد و پیچیدگی بالا، به‌خصوص زمانی که داده‌ها دارای همخطی[۱] هستند، عملکرد بهتری دارد. در نهایت، مقاله نتیجه‌گیری می‌کند که MPLE می‌تواند به‌عنوان یک ابزار قوی در ارزیابی ریسک اعتباری مورد استفاده قرار گیرد.

کیم و لی (۲۰۲۳)[۲] این مقاله به بررسی یک رویکرد ترکیبی برای ارزیابی ریسک اعتباری پرداخته است که شامل استفاده از شبکه‌های عصبی مصنوعی و رگرسیون لجستیک به صورت همزمان می‌باشد. نویسندگان نشان می‌دهند که این رویکرد ترکیبی می‌تواند به‌طور قابل توجهی دقت مدل های پیش‌بینی را افزایش دهد. مقاله به‌ویژه به این موضوع می‌پردازد که چگونه شبکه‌های عصبی می‌توانند الگوهای پیچیده را شناسایی کنند، در حالی که رگرسیون لجستیک امکان تفسیرپذیری بالاتری را فراهم می‌آورد. نتایج تحقیق نشان می‌دهد که این مدل ترکیبی در مقایسه با مدل‌های مجزا، دقت پیش‌بینی بیشتری دارد و می‌تواند برای شناسایی مشتریان پرخطر و کاهش نرخ نکول وام‌ها مؤثر باشد

نگوین وتران (۲۰۲۳)[۳] این مقاله به مقایسه دو روش رگرسیون لجستیک و MPLE (تخمین حداکثر شبه‌احتمال اصلاح شده) در پیش‌بینی اعتبار مشتریان پرداخته است. نویسندگان استدلال می‌کنند که MPLE در برخورد با داده های نامتوازن و پیچیده، عملکرد بهتری نسبت به رگرسیون لجستیک دارد. یافته‌های تحقیق نشان می‌دهد که MPLE قادر است انحرافات ناشی از برآورد در رگرسیون لجستیک را به حداقل برساند و دقت پیش‌بینی را بهبود دهد. در این مطالعه، چندین مجموعه داده واقعی از مشتریان بانک‌ها برای آزمون مدل‌ها استفاده شده و نتایج حاکی از برتری MPLE در شناسایی مشتریان با ریسک بالا و کاهش خطاهای طبقه‌بندی است.

پاتل و شارما (۲۰۲۴)[۴] این مطالعه به تحلیل مقایسه‌ای بین شبکه‌های عصبی و رگرسیون لجستیک در امتیازدهی اعتباری پرداخته است. نویسندگان با استفاده از چندین مجموعه داده از

---

[۱] -collinearity

[۲] Kim, S., & Lee, H

[۳] Nguyen, P., & Tran, D

[۴] Patel, R., & Sharma

بانک‌های مختلف، عملکرد دو روش را در پیش‌بینی ریسک اعتباری مقایسه کرده‌اند. یافته‌ها نشان می‌دهد که شبکه‌های عصبی در شناسایی الگوهای پیچیده و غیرخطی در داده‌ها عملکرد بهتری دارند، اما به تنظیمات دقیق‌تر و زمان بیشتری برای آموزش نیاز دارند. در مقابل، رگرسیون لجستیک با وجود سادگی، در تفسیر نتایج و پیاده‌سازی در سیستم‌های بانکی سریع‌تر و کارآمدتر است. مقاله به بانک‌ها توصیه می‌کند که بر اساس حجم داده‌ها و نوع مسئله، یکی از این دو روش یا ترکیبی از آن‌ها را به کار گیرند.

لین و ژانگ (۲۰۲۴)[۱] در این مقاله به مرور مدل‌های پیش‌بینی در ارزیابی اعتبار مشتری پرداخته‌اند. آن‌ها با بررسی مقالات و پژوهش‌های گذشته، به تحلیل نقاط قوت و ضعف هر یک از روش‌های مورد استفاده در ارزیابی ریسک اعتباری پرداخته‌اند. این مطالعه به دسته‌بندی مدل‌ها بر اساس تکنیک‌های آماری، یادگیری ماشین و یادگیری عمیق پرداخته و نتایج نشان می‌دهد که مدل‌های ترکیبی (مانند ترکیب رگرسیون لجستیک و شبکه عصبی) در شناسایی مشتریان پرریسک دقت بالاتری دارند. مقاله همچنین به چالش‌های پیاده‌سازی این مدل‌ها، از جمله کیفیت داده‌ها، پیچیدگی محاسبات و تفسیرپذیری نتایج اشاره دارد.

رابرتز و اوانز (۲۰۲۴)[۲] در این مقاله به بررسی استفاده از شبکه‌های عصبی برای بهبود امتیازدهی اعتباری پرداخته‌اند. آن‌ها با استفاده از داده‌های واقعی مشتریان، نشان داده‌اند که شبکه‌های عصبی می‌توانند به دقت بیشتری در مقایسه با مدل‌های رگرسیون سنتی دست یابند. این مطالعه به چالش‌های مهم پیاده‌سازی شبکه‌های عصبی مانند نیاز به حجم بالای داده‌ها، پیچیدگی در تنظیمات مدل و مشکل تفسیرپذیری نتایج نیز اشاره کرده است. نویسندگان توصیه می‌کنند که برای بهره‌مندی از مزایای شبکه‌های عصبی، این مدل‌ها در کنار روش‌های سنتی و به‌عنوان مکمل استفاده شوند تا بتوانند به مدیریت بهتر ریسک کمک کنند.

---

[۱] Le, T., & Vo, H

[۲] Roberts, T., & Evans

در مطالعه‌ی ژائو (۲۰۲۵)[۱]، پژوهشگران با استفاده از داده‌های واقعی مشتریان برای تحلیل عملکرد لجستیک رگرسیون در پیش‌بینی ریسک اعتباری، آن را با روش‌های دیگر مقایسه کردند. نتایج نشان داد که لجستیک رگرسیون قادر است بین مشتریان با ریسک بالا و پایین تمایز معنی‌دار ایجاد کند، به‌ویژه زمانی که داده‌ها نامتوازن هستند. یکی از مزایای کلیدی این روش سادگی تفسیر نتایج و قابلیت اجرای عملی در مؤسسات مالی است

مطالعات داخلی

مدرس و ذکاوت (۱۳۸۲) در پژوهش "مدل های ریسک اعتباری مشتریان بانک (مطالعه موردی)" برای طراحی مدل‌های ریسک اعتباری برای بانک توسعه ی صادرات ایران یک نمونه ۱۲۰ تایی از مشتریان حقوقی را انتخاب و ازروش‌های تحلیل تمایزی و رگرسیون لجستیک استفاده نمودند. متغیرهای نسبت جاری، نسبت بدهی به مجموع دارایی‌ها، نسبت حقوق صاحبان سهام به مجموع دارایی‌ها، نسبت سود قبل از کسر مالیات به حقوق صاحبان سرمایه، نسبت سود قبل از کسر مالیات به خالص فروش به عنوان پیش‌بینی کننده های نکول در مدل های تحلیل تمایزی و رگرسیون لجستیک استفاده شد. نتایج نشان داد که براساس متغیرهای مالی می‌توان مشتریان حقوقی بانک توسعه را از نظر ریسک اعتباری به خوش حساب و بد حساب دسته بندی کرد. از بین متغیرهای مالی استفاده شده در این تحقیق متغیر نسبت جاری بیشترین سهم را در تفکیک مشتریان دارا بوده است.

رضا تهرانی و میر فیض فلاح شمس (۱۳۸۴) در پژوهشی با عنوان " طراحی و تبیین مدل ریسک اعتباری در نظام بانکی کشور " کارآیی مدل‌های احتمالی خطی، لجستیک و شبکه‌های عصبی مصنوعی برای پیش‌بینی ریسک اعتباری مشتریان نظام بانکی کشور، مورد بررسی قراردادند.

[۱] Zhao, F

متغیرهای پیش‌بینی کننده در این مدل‌ها، نسبت‌های مالی وام گیرندگان بوده که برای بررسی معنی داری ارتباط آنها با ریسک اعتباری از آزمون‌های آماری مناسب استفاده نمودند. در این پژوهش از داده‌های مالی و اعتباری ۳۱۶ نفر از مشتریان حقوقی بانک‌های کشور به منظور طراحی و آزمون کارآیی مدل‌های یاد شده استفاده شده است. نتیجه‌های به دست آمده بیانگر این است که ارتباط بین متغیرها در مدل پیش‌بینی ریسک اعتباری به صورت خطی نبوده و تابع‌های نمایی و سیگموئید مناسب‌ترین مدل‌های پیش‌بینی ریسک اعتباری محسوب می‌شوند. بیشترین کارآیی برای پیش‌بینی ریسک اعتباری به ترتیب مربوط به شبکه‌های عصبی مصنوعی و مدل لجستیک بوده است. انصاری(۱۳۸۸) برای اعتبارسنجی مشتریان حقوقی بانک پارسیان از شبکه ی عصبی استفاده کرده،نتایج بیانگر این بوده که متغیرهای نسبت دارایی جاری به کل دارایی، سود قبل از کسر مالیات به کل بدهی، فروش به کل بدهی، سود خالص به حقوق صاحبان سهام وکل بدهی به حقوق صاحبان سهام اهمیت بالاتری داشته و وزن‌های بالاتری را به خود اختصاص داده است. شبکه ی عصبی ۵ متغیره درداده‌های آموزش کارایی بالاتری نسبت به رگرسیون لاجیت داشت اما در داده‌های آزمایش لاجیت کارایی بیشتری نسبت به شبکه عصبی داشت.

میرزائی و دیگران (۱۳۹۰) در پژوهشی با عنوان " بررسی عوامل مؤثر بر ریسک اعتباری اشخاص حقوقی بانک‌ها (مطالعه موردی شعب بانک ملی ایران، شهر تهران)" با استفاده از روش رگرسیون لجستیک ریسک اعتباری شرکت‌های حقوقی را که در سال ۱۳۸۷ از بانک ملی ایران شعب شهر تهران تسهیلات اعتباری دریافت نموده اند، بررسی نمودند. ابتدا ۳۹ متغیر توضیح دهنده شامل متغیرهای کیفی و مالی را شناسایی کردند و درنهایت ۱۱ متغیر را که اثر معنا داری بر ریسک اعتباری و تفکیک بین دو گروه از مشتریان خوش حساب و بد حساب داشتند، انتخاب کرده و مدل نهایی را به وسیله آنها برازش کردند. نتایج نشان می‌دهد که براساس شاخص‌های آماری، این توابع از نظر ضرایب و همچنین قدرت تفکیک کنندگی معنادار بوده و اعتبار بالایی دارند.

دهمرده و دیگران (۱۳۹۱) در مقاله خود تحت عنوان "اعتبار سنجی مشتریان بانک با استفاده از رویکرد امتیازدهی اعتباری" با استفاده از روش رگرسیون لجستیک یک نمونه تصادفی از مشتریان حقیقی از شعب بانک سپه درسطح شهر زاهدان اقدام به دریافت تسهیلات نموده اند، انتخاب کرده و به وسیله ۱۵ متغیر که اثرمعناداری بر ریسک اعتباری داشته اند، مدل نهایی برازش شده است. در این مطالعه درجه حساسیت و درجه تشخیص و خطای نوع اول یعنی ریسک اعتباری،خطای نوع دوم یعنی ریسک تجاری ارزیابی شده است. همچنین آنها به منظور بررسی قدرت تفکیک کنندگی مدل از منحنی ROC استفاده نمودند. سطح زیر منحنی ROC در مدل برازش شده نشان دهنده قدرت تفکیک کنندگی بسیار بالای مدل می‌باشد. نتایج حاصل از برآورد نشان می‌دهد که براساس شاخص‌های آماری، رگرسیون لجستیک از نظر ضرایب و همچنین قدرت تفکیک کنندگی معنادار بوده و در مدیریت ریسک اعتباری بانک از اعتبار بالایی برخوردار است.

عزیزاللهی و همکاران (۱۴۰۱) در مقاله خود به اعتبار سنجی مشتریان حقیقی بانک ملت با استفاده از رویکرد دلفی فازی و رگرسیون لجستیک چندگانه پرداختند. است. بدین منظور، ابتدا به روش کتابخانه‌ای، شاخص های مؤثر بر ارزیابی ریسک اعتباری مشتریان حقیقی، شناسایی و سپس، از روش دلفی فازی، استفاده و شاخص‌ها، غربالگری گردید. جامعه این بخش را خبرگان بانکی (بانک ملت) تشکیل دادند که به روش گلوله برفی انتخاب شدند. در ادامه داده‌های نهایی مرتبط با شاخص‌ها، شامل پرونده‌های ۷۳۱۸ مشتری حقیقی بانک ملت طی سال‌های ۱۳۹۹ – ۱۳۹۳، جمع‌آوری و با روش رگرسیونی لجستیک چندگانه در چهار رده وصول به موقع، سررسید گذشته، معوق و مشکوک الوصول، تحلیل شد. نتایج نشان داد، شاخص‌های مبلغ وام، زمان بازپرداخت وام، فاصله اقساط، مبلغ هر قسط، تمدید وام، وام قبلی، وثیقه ملکی، معدل موجودی، نرخ سود تسهیلات و سطح تحصیلات، تأثیر معنی‌داری بر ریسک اعتباری مشتریان حقیقی دارند. همچنین معنی‌داری شاخص‌های جنسیت، سن و شغل، تأیید نگردید

احمدی سرتختی و همکاران (۱۴۰۲) در مقاله خود به بررسی ریسک اعتباری مؤسسه‌های مالی پرداختند. که از احتمال قصور تسهیلات‌گیرنده در بازپرداخت تسهیلاتی یا بازپرداخت به‌موقع تسهیلات شکل می‌گیرد. در این پژوهش تلاش شده است تا به‌کمک مدل شبکه عصبی مصنوعی، مدلی برای ارزیابی ریسک متقاضیان تسهیلات و ضمانت‌نامه‌ها از این صندوق طراحی شود که بیشترین قدرت پیش‌بینی احتمال نکول تسهیلات اعطایی را داشته باشد. بر اساس مدل شبکه عصبی و با سه شیوه بیزین، لونبرگ و گرادیان مزدوج و با ۱ تا ۳۰ نرون در لایه پنهان، بهترین مدل‌ها استخراج شدند. بهترین مدل با کلیه متغیرهای مستقل (یعنی ۶۹ متغیر مستقل)، پیش‌بینی‌ای با۹۶/۲ درصد دقت داشته است که از مدل اقتصادسنجی پروبیت بیشتر است. بر اساس نتایج حاصل از این پژوهش، می‌توان ضمن تفکیک مشتریان خوش‌حساب از بدحساب و بر اساس رتبه اعتباری مشتریان، میزان وثایق اخذ شده از مشتریان را متناسب با وضعیت اعتباری گروه‌های اعتباری تنظیم کرد؛ بدین معنا که چنانچه مشتری جزء مشتریان خوش‌حساب و با رتبه اعتباری خوب باشد، می‌توان وثایق کمتری از ایشان گرفت و اگر مشتری ریسک اعتباری بالاتری داشته باشد، متناسب با آن، وثایق بیشتری را به‌عنوان تضمین تعهد بازپرداخت درخواست کرد. در این صورت، ضمن اندازه‌گیری ریسک اعتباری و تفکیک مشتریان، می‌توان به مدیریت بهینه ریسک و پرتفوی اعتباری صندوق ضمانت صادرات ایران اقدام کرد.

کاظمی و همکاران (۱۴۰۳) در پژوهشی، داده‌های واقعی مشتریان بانک برای مقایسه پیش‌بینی ریسک اعتباری با استفاده از شبکه عصبی مصنوعی و رگرسیون لجستیک بررسی کرده اند. در این پژوهش نشان می‌دهد که شبکه عصبی و رگرسیون لجستیک هردو توانایی در رتبه‌بندی مشتریان دارند و می‌توانند مشتریان را از نظر ریسک اعتباری به گروه‌های مختلف تقسیم کنند.

احمدی و همکاران (۱۴۰۳) در این مطالعه به بررسی شاخص‌های مؤثر در اعتبارسنجی و رتبه‌بندی مشتریان خرد در بانک خاورمیانه می‌پردازد و با تحلیل آماری داده‌های مشتریان واقعی، عوامل مالی و غیرمالی را تعیین می‌کند که می‌توانند در مدل‌های رتبه‌بندی اعتباری به‌کار روند.

این تحقیق زمینه مفهومی قوی برای اســتفاده از مدل‌های آماری و یادگیری ماشــین در رتبه‌بندی مشتریان فراهم می‌کند، زیرا شناخت صحیح متغیرهای اعتباری پایه هر مدل پیش‌بینی است.

سعی می‌شود کارآیی مدل‌های لجستیک و شبکه‌های عصبی برای پیش‌بینی ریسک اعتباری مشتریان نظام بانکی در شعب بانکهای خصوصی بررسی شود. علاوه بر این تلاش می‌شود ارتباط بین پارامترهای مربوط به حساب سپرده‌ها، ویژگی‌های مشتریان بانکی و وضعیت اعتباری آنها ارزیابی شود.

٤. مدل تحقیق و روش برآورد

شبکه عصبی

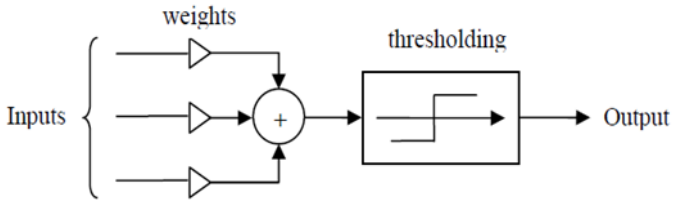
یک سیستم عصبی مصنوعی در حقیقت، فرآیند یادگیری انسانی را شبیه سازی می‌کند. سیستم شبکه‌ی عصبی مصنوعی با تقلید از سیستم عصبی و مغزی انسان می‌کوشد که ارتباط بین داده‌ها (جنسیت، تاهل، مبلغ تسهیلات و...) و ستاده‌ها (وضعیت اعتباری وام گیرنده) را از راه تکرار نمونه-برداری از مجموعه‌ی داده‌های گذشته داده/ ستاده یاد گیرد. شبکه‌ی عصبی دارای یک برتری اساسی نسبت به سیستم خبره بوده و آن این است که هنگامی که داده‌ها کامل نبوده و یا دارای پارازیت باشند، از راه آموخته‌های گذشته حدس منطقی از داده‌ها می‌سازد. یک شبکه‌ی عصبی بر اساس سه ویژگی، داده‌های ورودی، وزنها و لایه‌های پنهان مشخص می‌شود. (ساندرز و آلن،۲۰۰۲)[1]

در ســاختار ایــن شــبکه، تعــداد نــرونهــای لایــهٔ ورودی $(x_1,x_2,\ldots..,x_d)$ با تعداد خصوصیات موجود برای تصمیم‌گیری در خصوص هر یک از نمونه‌هـای داده‌هابرابر است.همچنین تعداد نرون‌های لایـهٔ خروجـی شـبکه$(z1,z2,\ldots..zc)$ برابـر است با تعداد کلاسهای داده‌ها و هر نود در لایهٔ خروجی، متناظر اسـت با یک کلاس خاص در مجموعه داده‌هـــا. بخـــش دیگر ساختار این شبکه، لایهٔ پنهان یا لایهٔ میانی آن ($y_1,y_2,\ldots,y_c$) است. تعداد لایه‌های پنهان معمولاً

[1] Saunders. Anthony,, , Allen, Linda

یـک لایـه در نظر گرفته میشود. تعداد بیشتر لایه‌های پنهان، مسئله را پیچیده و اجرای آن را زمانبر می‌کند. (احمدی سرتختی و همکاران،۱۴۰۲)

در یک نگاه ساده، مدل یک عصب باید شامل ورودیهایی باشد که در نقش سیناپس انجام وظیفه کنند. این ورودیها در وزنهایی ضرب می‌شوند تا قدرت سیگنال را تعیین کنند. نهایتاً یک عملگر ریاضی تصمیم‌گیری می‌کند که آیا نورون فعال شود یا خیر و اگر جواب مثبت باشد، میزان خروجی را مشخص می‌سازد. بنابراین شبکه عصبی مصنوعی با استفاده از مدل ساده شده عصب واقعی به پردازش اطلاعات می‌پردازد. لذا می‌توان مدل ساده ای برای توصیف یک نورون (یک گره در شبکه عصبی مصنوعی) پیشنهاد کرد. این مدل در شکل زیر نشان داده شده است.



نمودار ۱- مدل شبکه عصبی

توابع مورد استفاده برای آستانه گذاری بسیار متنوع می‌باشند، از جمله مشهورترین این توابع می‌توان به تابع‌هایی نظیر arcsin, arctan, sigmoid اشاره کرد. (صادقی،سهرابی وفاو سلمانی، ۱۳۹۳)

شمار لایه‌ها و شمار نورونها در هر لایه مخفی به طور معمول به وسیله روش آزمون و خطا مشخص می‌شود، نورونهای لایه‌های مجاور در شبکه به طور کامل با هم در ارتباط هستند. به طور کلی شبکه‌های عصبی به دو نوع شبکه‌های پیشخور[۱] و پسخور[۲] تقسیم می‌شوند تفاوت آنها در این است که در شبکه‌های پسخور، حداقل یک سیگنال برگشتی از یک نورون به همان نورون یا نورونهای همان لایه و یا لایه قبل وجود دارد. در برخی موارد، شبکه‌های عصبی پسخور می‌توانند

[۱] Feed Forward Network(FFN)

[۲] Feed Back Network(FBN)

بسیار مفید واقع شوند ولی با این حال، در ۸۰ درصد کاربردها از شبکه‌های عصبی پیشخور استفاده می‌شود. (منهاج ،۱۳۸۴)

شبکه‌های پرسپترون چند لایه[۱]

این شبکه شامل سه لایهٔ ورودی، مخفی و خروجی است که تعداد سلولهای هر لایه به روش سعی و خطا مشخص می‌گردد. سیگنالهای ورودی به وسیلهٔ ضریب‌های بهنجار کننده به مقدار یک نرمالیزه شده و بعد از محاسبات، خروجی به مقدار واقعی برگردانده می‌شود. همچنین مقادیر اولیهٔ وزنها به صورت اتفاقی در نظر گرفته شده‌اند. این شبکه بر مبنای الگوریتم پس انتشار خطا آموزش می‌بیند. بدین ترتیب که خروجیهای واقعی با خروجیهای دلخواه مقایسه می‌شوند و وزنها به وسیله الگوریتم پس انتشار، به صورت تحت نظارت تنظیم می‌گردند تا الگوی مناسب بوجود آید. برای الگوی ورودی p ام، مربع خطای خروجی برای تمامی سلولهای لایه خروجی شبکه به صورت زیر در می‌آید:

$$E_p = \frac{1}{2}(d^P - y^P)^2 = \frac{1}{2}\sum_{j=1}^{s}(d_j^p - y_j^p)^2 \qquad (۱)$$

که در آن d_j^p خروجی دلخواه برای j امین سلول در لایه خروجی و y_j^p خروجی واقعی برای j امین سلول در لایهٔ خروجی، s ابعاد بردار خروجی، y^p بردار خروجی واقعی و d^p بردار خروجی دلخواه هستند. مربع خطای کل E برای Pالگو بصورت زیر درمی آید:

$$E = \sum_{p=1}^{p} E_p = \frac{1}{2}\sum_{p=1}^{p}\sum_{j=1}^{s}(d_j^p - y_j^p)^2 \qquad (۲)$$

وزنها با هدف کاهش تابع هزینه E به مقدار مینیمم به روش گرادیان نزولی تنظیم می‌گردند . معادلهٔ به روز درآوردن وزنها به صورت زیر است:

$$w_{ij}(t+1) = w_{ij}(t) + \eta\Delta w_{ij}(t) + \alpha\Delta w_{ij}(t-1) \qquad (۳)$$

[۱] Multi Layer Perceptron(MLP)

که در آن $\Delta w_{ij}(t) = -(\frac{\partial E_p}{\partial w_{ij}(t)})$ ، η ضـریب یادگیری، α ضـریب لحظه ای $w_{ij}(t+1)$ وزن جدید و $w_{ij}(t)$ وزن قبلی می‌با شد. همچنین در این روش ، وزنها به طور مکرر برای تمامی الگوهای یادگیری به روز درآورده می شوند. روند یادگیری هنگامی متوقف می شود که مجموع کل خطا E برایp الگو از مقدار آ ستانه تعیین شده کمتر شود یا تعداد کل دوره تعلیم به پایان برسد. (هایکین،۱۹۹۹)[۱]

لازم به ذکر اســت که روش تعلیم مذکور، روش تعلیم پس انتشــار خطا با ترم لحظه ای می‌باشــد و احتمال همگرایی در مینیمم‌های محلی را نسـبت به روش پس انتشــار خطا کاهش می‌دهد.

روند شبیه سازی مسئله

با توجه به بحث پیشــین، انتخاب تعداد لایه‌ها محدود به ۳ لایه می‌شــود. در لایه ورودی باید به تعداد ابعاد هر الگوی ورودی، نورون قرار دهیم. بنابراین سایز لایه ورودی را بعد داده‌های ورودی تعیین می کند. در لایه خروجی نیز به و ضوح باید به تعداد کلا سها، نورون دا شته با شیم. در حالت ایده آل، با آمدن ورودی مربوط به یک کلاس انتظار داریم نورون مربوط به آن کلاس مقدار ۱ و مابقی نورونها مقدار ۰ را به خود بگیرند. اما در عمل با توجه به تابع سیگموئید مورد استفاده، مقدار خروجی شــبکه عددی بین صــفر و یک اســت. که خروجی بیانگر مشــتری خوب و مشــتری بد می‌باشد، شبکه مقدار یک را به متقاضی خوب و مقدار صفر را به متقاضی بد اختصاص می‌دهد در مورد تعداد نورونهای لایه میانی، مبنای خاصــی وجود ندارد و معمولاً با ســعی و خطا به نحوی انتخاب می گردد که شبکه جواب معقولی در اختیار بگذارد.

[۱] Haykin, Simon

قدم بعدی در شبیه‌سازی، انتخاب نوع تابع خروجی نورون است که عمدتاً تابع سیگموئید استفاده می‌شود.

در اینجا ۸۰ درصد دادها به منظور آموزش و ۲۰ درصد مابقی برای تست مدل استفاده گردیده است و مدل را ۱۰۰ بار اجرا نموده و در نهایت خطای میانگین را محاسبه می‌کنیم.پس ازآنکه شبکه توسط این داده‌ها آموزش دید، وزنها مقدار نهایی خود را یافته‌اند به نحوی که شبکه برای داده‌های آموزش، کمترین خطا را بدست می‌دهد. مابقی داده‌ها که در آموزش نقشی نداشته اند به عنوان ورودی به شبکه داده شده و پاسخ شبکه با پاسخ مطلوب (برحسب آنها) مقایسه می‌گردد و بدین ترتیب راندمان شبکه آموزش دیده محک زده می‌شود.

رگرسیون لجستیک

رگرسیون لجستیک یک روش آماری است که برای مدل‌سازی و تحلیل متغیرهای وابسته دو مقداری (باینری) به کار می‌رود. این روش در اعتبارسنجی بانک‌ها به منظور ارزیابی ریسک اعتباری مشتریان و پیش‌بینی احتمال نکول وام‌ها استفاده می‌شود. با استفاده از رگرسیون لجستیک، بانک‌ها می‌توانند تصمیمات بهتری در مورد اعطای تسهیلات اتخاذ کرده و ریسک‌های مرتبط با نکول را کاهش دهند.

رگرسیون لجستیک نوعی از رگرسیون است که متغیرهای مستقل می‌تواند هم در مقیاس کمی و هم در مقیاس کیفی باشد، ولی متغیر وابسته کیفی دو سطحی است. این دو مقوله به گونه‌ای معمول به عضویت یا عدم عضویت در یک گروه اشاره دارد. این مدل رگرسیون، شبیه رگرسیون معمولی است با این تفاوت که روش تخمین ضرایب یکسان نمی باشد و به جای حداقل کردن مجذور خطاها (که در رگرسیون معمولی انجام میشود)، احتمالی که یک واقعه رخ می‌دهد را حداکثر می‌نماید. در رگر سیون لجستیک از مفهومی به نام بخت برای مقدار متغیر وابسته ا ستفاده می شود در ا صطلاح آماری بخت به معنی احتمال رخداد یک پیشامد (Pi) بر احتمال عدم رخداد

$(1 - P_i)$ آن می‌باشد، احتمال بین ۰و ۱ تغییر می‌کند، در حالی که بخت ممکن است بیش از یک باشد. واژه ی کلیدی در تحلیل رگرسیون لجستیک سازه ای به نام لوجیت[۱] است که لگاریتم طبیعی بخت می‌باشد. رگرسیون لجستیک به صورت معادله ۳-۱ تعریف می‌شود:

$$L = \ln\left(\frac{P(Y=1)}{1-P(Y=1)}\right) = \beta_0 + \sum_{i=1}^{n} \beta_i X_i \qquad (۴)$$

$$P(Y=1) = \frac{1}{1+e^{-z}} \quad \text{where} \quad z = \beta_0 + \sum_{i=1}^{n} \beta_i X_i \qquad (۵)$$

در این مدل $\beta 0$ عرض از مبدا و β_i ضریب مربوط به متغیر مستقل i ام را نشان می‌دهد. (تهرانی و همکاران،۱۳۸۴)

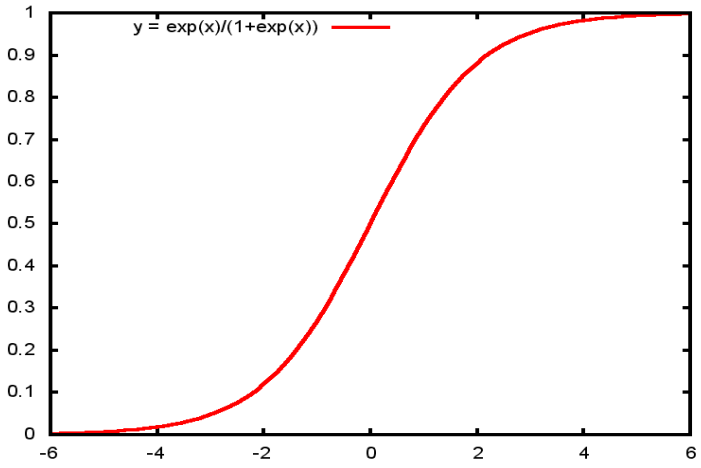
L (لگاریتم نسبت شانس[۲]) ، نسبت به پارامترها خطی است و مقدار آن پس از برآورد ضرایب از روش حداکثر درستنمایی (MLE) به دست می‌آید. بدین ترتیب با برآورد ضرایب، مدلی حاصل می‌شود که هریک از ضرایب آن مقدار متغیر L (لگاریتم شانس به نفع قصور در بازپرداخت) را به ازای یک واحد تغییر متغیر مستقل نشان می‌دهد و سپس با محاسبه احتمال قصور در بازپرداخت از معادله ۳-۲ می‌توان مشتریان بانک را طبقه‌بندی نمود.

شکل کلی این تابع مانند زیر خواهد بود:

---

[۱] Logit

[۲] Odd Ratio



نمودار ۲: مدل لاجیت

مأخذ: نتایج انجام تحقیق

در برازش رگرسیون لجستیک معنی‌داری ضرایب رگرسیون از طریق تابع درستنمایی و آزمون والد سنجیده می‌شود. تابع درستنمایی یک روش برای آزمون معنی داری ضرایب مدل رگرسیون است. مقایسه‌ی مقدار مشاهده شده با مقادیر پیش‌بینی شده با استفاده از تابع درستنمایی صورت می‌گیرد و آزمون والد یکی از راه‌هایی است که معمولاً برای آزمون معنی‌داری ضرایب در رگرسیون لجستیک به کار برده می‌شود.(نیکولیک و دیگران،۲۰۱۳)[۱]

روند شبیه سازی مسئله

ابتدا یک نمونه به حجم ۸۹۳ از میان متقاضیان دریافت وام انتخاب و اطلاعات مربوط به آنها را جمع آوری می کنیم.[۲] به منظور بکارگیری مدل لاجیت در تحقیق حاضر ابتدا در یک ستون وضعیت اعتباری افراد را اعم از خوش حساب و بد حساب با کد صفر و یک مشخص و سپس در مقابل هر

[۱] Nikolic et al

[۲]-در ابتدا با توجه به نوع داده و بانک و محدودیت پژوهش نمونه اولیه ۱۰۴۰ داده دریافت شد در مرحله اول ۸۰ نمونه بر اساس بررسی اشتباهات تایپی حذف شد و به ۹۶۰ داده رسید و در مرحله بعد داده های پرت حذف شده و به عدد۸۹۳ داده نهایی شد.

یک سایر متغیرها را اعم از رسته‌ای و کمی مشخص و در ستونهای دیگر مرتب می‌کنیم.به منظور بررسی عملکرد مدل رگرسیون، ۸۰ درصد داده‌ها (تعداد ۷۱۵ متقاضی دریافت تسهیلات) را به منظور آموزش مدل و ۲۰ درصد باقیمانده داده‌ها یعنی اطلاعات مربوط به ۱۷۸ مشتری به منظور سنجش اعتبار مدل به صورت تصادفی جداسازی شده به جهت مقایسه و تحلیل بهتر مدل، مدلسازی با ۱۰۰ مرتبه تکرار انجام شده و مقدار خطای متوسط برای این روش رگرسیون محاسبه شد.

۱- داده‌ها و نتایج مدل

متغیرهای مورد مطالعه در این تحقیق عبارتند از:متغیر وابسته: وضعیت اعتباری (y).

متغیرهای مستقل: سطح تحصیلات (e)، جنسیت (g)، هدف از دریافت وام (tf)، نوع وثیقه (co)، مبلغ تسهیلات (t)، سود تسهیلات (p)، مانده حساب مشتری (b) و مبلغ وثیقه (c)[۱]

شبکه عصبی

در این تحقیق از شبکه عصبی پیشخور تعمیم یافته با یک لایه ورودی، یک لایه خروجی و سه لایه پنهان (با استفاده از آزمون سعی و خطا)، جهت طراحی شبکه عصبی استفاده شده است. گره‌های ورودی در حقیقت همان متغیرهای تحقیق است که معافی شده است.

بررسی قدرت پیش بینی مدل

نتایج اجرای مدلسازی شبکه عصبی پرسپترون با حد آستانه ۰/۵ در جدول زیر نشان داده شده است.بیشتر از آستانه خوش حساب و کمتر از آن بد حساب تلقی می‌شوند.

جدول ۱: بررسی قدرت پیش بینی مدل شبکه عصبی برای داده‌های آموزش در حد آستانه ۰,۵

مشاهدات برآورد	$Y = 0$	$Y = 1$	کل
$p(y) \leq 0.5$	۳	۱	۴

[۳] – این تحقیق دارای ۹ متغیر می‌باشد که متغیر وضعیت اعتباری به عنوان متغیر هدف در نظر گرفته شده است. ۸ متغیر دیگر نیز به عنوان متغیرهای پیش بینی کننده مدل در نظر گرفته شده است. از ۸ متغیر پیش بینی کننده انتخاب شده در این تحقیق ۵ متغیر بر اساس مرور ادبیات و مطالعات پیشین انتخاب گردیده و ۳ متغیر دیگر براساس اطلاعات به دست آمده از بانک مورد مطالعه انتخاب شده است.

$p(y) > 0.5$	۷۴	۶۳۷	۷۱۱
کل	۷۷	۶۳۸	۷۱۵
درست	۳	۶۳۷	۶۴۰
درصد درست	٪3/9	99/84٪	۸۹٪/۵۱
درصد نادرست	٪96/1	0/16٪	10/49٪

مأخذ: محاسبات تحقیق

جدول 2: بررسی قدرت پیش‌بینی مدل شبکه عصبی برای داده‌های تست در حد آستانه ۰/۵

مشاهدات برآورد	$Y = 0$	$Y = 1$	کل
$p(y) \leq 0/5$	۰	۲	۲
$p(y) > 0/5$	۱۹	۱۵۷	۱۷۶
کل	۱۹	۱۵۹	۱۷۸
درست	۰	۱۵۷	۱۵۷
درصد درست	٪۰	٪98/74	٪88/2
درصد نادرست	٪۱۰۰	٪1/26	٪1/18

مأخذ: محاسبات تحقیق

با درنظر گرفتن مقادیر محاسباتی در جدول (۱) ، نسبتی از مشاهدات Y=1 که به درستی پیش بینی شده اند را "درجه حساسیت" می‌نامند که در اینجا برای داده‌های آموزش برابر 99/84٪ و برای داده‌های تست برابر99/74 و نسبتی از مشاهدات Y=0 که به درستی پیش بینی شده اند، "درجه تشخیص" یا می‌نامند که در اینجا برای داده‌های آموزش برابر 3/9٪ و برای داده‌های تست برابر ۰٪ بوده است.همچنین با توجه به جدول (۱) خطای نوع اول (عدد یک منهای درجه حساسیت) یعنی ریسک تجاری در مدل برازش شده برابر 0/16درصد و خطای نوع دوم 96/1٪ (عدد یک منهای درجه تشخیص) یعنی ریسک اعتباری می‌باشد.با در نظر گرفتن بزرگی خطای نوع دوم و بیان این که در مواقعی که نسبت داده‌ها به صورت ۵۰-۵۰ نمی‌باشد باید به دنبال حد آستانه ی بهینه گشت

تا قدرت برازش مدل را بهبود بخشد و خطا را در حد متعادل قرار دهد.(کوهن و جانسون،۲۰۱۳)[۱] به دنبال بدست آوردن حدآستانه بهینه خواهیم بود.

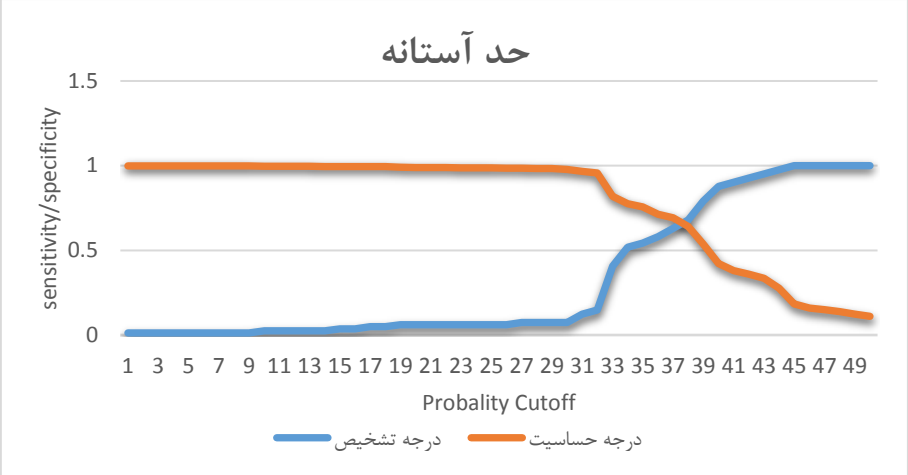
حد آستانه بهینه شبکه عصبی

طبق تعریف، حد آستانه بهینه برای مدل (یا احتمالی که در آن متوسط ریسک اعتباری و تجاری مدل به حداقل می‌رسد) برابر احتمالی است که در آن یکی از حالات زیر رخ دهد نخست: مجموع درجه حساسیت و درجه تشخیص مدل ماکزیمم شود؛ دوم: درصورتیکه درجه حساسیت مدل بیش از ۸۰ درصد باشد، مجموع درجه حساسیت و درجه تشخیص، ماکزیمم شود؛ سوم: مقدار مینیمم بین دو مقدار برای تمام مشاهدات ماکزیمم شود؛ چهارم: درصورتیکه اهمیت درجه حساسیت مدل α برابر درجه تشخیص آن باشد، حاصل رابطه (درجه تشخیص*۱+ درجه حساسیت*α) ماکزیمم شود.(کرشلم،۲۰۰۴)[۲].مقدار عددی حد آستانه به طورمستقیم قابل محاسبه نیست؛ ولی با توجه به مقادیر درجه حساسیت و تشخیص و نیز احتمال وقوع پیامد مورد نظر (دراینجا ریسک اعتباری)، برای کل مشاهدات نمونه قابل محاسبه است. مقدار این حد از طریق یکی از چهار روش بالا و به شکل دستی قابل محاسبه خواهد بود. در اینجا از روش سوم استفاده می‌نماییم و محل تلاقی منحنی درجه حساسیت و درجه تشخیص می‌باشد. برای این منظور از روش اعتبار سنجی ضربدری استفاده می‌نماییم و با ۱۰۰ بار مدل را اجرا نمودیم و با استفاده از میانگین گرفتن از درجه تشخیص و درجه حساسیت‌های حاصله از اجرای مدل‌های مختلف، منحنی درجه حساسیت و درجه تشخیص را رسم نمودیم و تلاقی دو منحنی همان حد آستانه بهینه خواهد بود.

---

[۱] Kuhn, Max., Johnson, Kjell.

[۲] Korsholm, Lars



نمودار ۳: منحنی درجه حساسیت و درجه تشخیص برای شبکه عصبی

مأخذ: نتایج انجام تحقیق

ملاحظه می‌شود که در x=37.6289 درجه حساسیت و درجه تشخیص برابر ۰/۶۶۰۶۸۶۴ می-باشد. برای به دست آورد حدآستانه بهینه یک حلقه تعریف می کنیم که از حد آستانه ابتدایی را برابر ۰,۵ قرار می‌دهد و با گام ۰/۰۱ حدآستانه را مرتبا افزایش می‌دهد تا محل تلاقی دو منحنی بدست آید بنابراین حد آستانه بهینه با قرار دادن x در معادله زیر که در مدل تعریف نمودیم بدست خواهد آمد.

$$Th = 0.5 + (x * 0.01) \qquad (۶)$$

لذا با توجه به رابطه فوق در اینجا حد آستانه بهینه برابر ۰/۸۷۶۳ می‌باشد[۱]. حال مدل را با حد آستانه ۰/۸۷۶۳ اجرا می کنیم.

[۱]- است بدین معنا که در صورتی که احتمال پیش بینی شده بیشتر از۰/۸۷۶۳ متغیر پاسخ در گروه مشتریان خوش حساب قرار می گیرد و در صورتی که احتمال پیش بینی شده کمتر از ۰/۸۷۶۳ باشد مشتری بد حساب تلقی می‌شود.

برازش مجدد مدل شبکه عصبی

با در نظر گرفتن مقدار جدید حد آستانه مدل مجددا برازش شده و نتایج در جدول زیر ارائه می شود.

جدول ۳: بررسی مجدد قدرت پیش بینی مدل شبکه عصبی برای داده‌های آموزش در حد آستانه ۰/۸۷۶۳

مشاهدات برآورد	$Y=0$	$Y=1$	کل
$p(y) \leq 0.8763$	۵۰	۲۳۱	۲۸۱
$p(y) > 0.8763$	۲۴	۴۱۰	۴۳۴
کل	۷۴	۶۴۱	۷۱۵
درست	۵۰	۴۱۰	۴۶۰
درصد درست	۶۷/۰۵۷٪	۹۶/۰۶۳٪	۶۴/۰۳۴٪
درصد نادرست	۳۲/۰۴۳٪	۳۶/۰۰۴٪	۳۵/۰۶۴٪

مأخذ: محاسبات تحقیق

باتوجه به جدول فوق ریسک اعتباری در این مدل برابر ۳۲/۴۳٪ می‌باشد و در حالت کلی ۶۴/۳۴ درصد از رفتار متقاضیان تسهیلات بانکی به درستی پیش‌بینی شده است. همچنین خطای نوع اول در مدل برازش شده برابر ۳۶/۰۴ درصد و خطای نوع دوم ۳۲/۴۳ درصد می‌باشد.

جدول٤: بررسی پیش بینی مدل شبکه عصبی برای داده‌های تست در حد آستانه ۰,۸۷۶۳

مشاهدات برآورد	$Y=0$	$Y=1$	کل
$p(y) \leq 0/8763$	۱۵	۶۹	۸۴
$p(y) > 0/8763$	۷	۸۷	۹۴
کل	۲۲	۱۵۶	۱۷۸
درست	۱۵	۸۷	۱۰۲
درصد درست	۶۸/۰۲٪	۵۵/۰۷۷٪	۵۷/۰۳٪
درصد نادرست	۳۱/۰۸٪	۴۴/۰۲۳٪	۴۲/۰۷٪

مأخذ: محاسبات تحقیق

نمونه ی خارج از برازش به تأیید اعتبار مدل کمک می‌نماید که در آن ۵۷/۳ درصد از مشتریان به درستی پیش‌بینی شده‌اند و همچنین رفتار ۶۸/۲ درصد از مشتریان بد حساب با استفاده از مدل شبکه عصبی به درستی پیش‌بینی گردیده است. ریسک اعتباری یا خطای نوع دوم برابر ۳۱/۸ درصد در داده‌های تست است.

رگرسیون لجستیک

رگرسیون لجستیک به صورت رابطه زیرتعریف می‌شود:

$$Z_i = ln\left(\frac{P_i}{1-P_i}\right) = \beta_0 + \sum_{i=1}^{n} \beta_i X_i \qquad (۷)$$

برآورد مدل رگرسیون لجستیک

ابتدا با استفاده از آزمون والد وثیقه منقول را با سایر وثایق ادغام می‌کنیم. با شبیه سازی مدل اطلاعات مربوط به جدول زیر به دست می‌آید:

جدول ۵ :خروجی رگرسیون لجستیک

نام متغیر	β	St d.Error	Z	P>|z|
مقدار ثابت	۱/۶۵۹۲۵	۰/۳۱۲۱۶	۵/۳۱۵	0/000000106
c	۱۶/۲۳۴۴۶	۷/۱۸۳۵۳	۰/۲۶	۰/۰۲۳۸
f	۰/۶۴۳۳	۱/۱۱۰۵۴	۰/۵۷۹	۰/۵۶۲۴
b	۱۷/۲۸۷۶۸	۷/۸۱۰۶	۲/۲۱۳	۰/۰۲۶۹
p	۱/۵۳۶۷۳	۰/۶۲۹۹۷	۲/۴۳۹	۰/۰۱۴۷
(g)1	۰/۸۵۹۸۸	۰/۳۵۹۵۲	۲/۳۹۲	۰/۰۱۶۸
(co)1	−۲/۴۳۲۶۹	۰/۹۶۰۶۵	−−۲/۵۳۲	۰/۰۱۱۳
(co)2	−۱/۱۴۱۰۶	۰/۴۹۶۳۴	−۲/۲۹۹	۰/۰۲۱۵
(tf)1	−۰/۶۳۴۹۸	۰/۳۱۱۲۱	−۲/۰۴	۰/۰۴۱۳
(tf)2	−۰/۱۵۵۰۲	۰/۳۰۱۲	−۰/۵۱۵	۰/۶۰۶۸
(e)1	−۰/۰۴۴۰۲	۰/۳۳۹۵	−۰/۱۳	۰/۸۹۶۸
(e)2	۰/۲۰۶۷	۰/۲۹۵۴۵	۰/۷	۰/۴۸۴۲
(e)3	۰/۸۳۱۳۹	۰/۷۵۴۰۵	۱/۱۰۳	۰/۲۷۰۲
(e)4	۱۳/۵۱۲۱۴	۴۷۶/۸۶۰۹۲	۰/۰۲۸	۰/۹۷۷۴

مأخذ: محاسبات تحقیق

با استفاده از آماره والد برای آزمون معناداری ضرایب متغیرهای مدل لاجیت و حذف متغیرهایی که معنی داری لازم را در سطح پنج درصد ندارند، مدل را دوباره اجرا می‌کنیم و نتایج آن را در جدول بعدی ملاحظه می‌کنیم.

جدول ٦ : خروجی رگرسیون لجستیک پس از غربال

نام متغیر	β	Std.Error	Z	P>|z|	Exp(β)
مقدار ثابت	۱/۸۲۶۴	۰/۲۵۶۴	۷/۱۲۳	۰/۰۰	۶/۲۱۱۸
c	۱/۰۱۶۵	۷/۰۲۲۱	۲/۴۲۳	۰/۰۱۵۴	۲۴۵۵۷۳۹۰
b	۱۷/۰۴۹۸	۷/۷۶۶۸	۲/۱۹۵	۰/۰۲۸۱	۲۵۳۸۹۱۶۰
p	۱/۵۵۲۳	۰/۶۲۲۷	۰,۰۱۲۷ ۲,۴۹۳	۰/۰۱۲۷	۷/۷۲۲۲
(g)1	۰/۸۵۳	۰/۳۵۶۹	۲/۳۹	۰/۰۱۶۸	۲/۳۴۶۶
(co)1	-۲/۴۰۵۷	۰/۹۵۶۷	-۲/۵۱۴	۰/۰۱۱۹	۰/۰۹۰۲
(co)2	-۱/۱۳۹۸	۰/۴۹۲۷	-۲/۳۱۳	۰/۰۲۰۷	۰/۳۱۹
(tf)1	۰/۶۴۲۵	۰/۳۱	۰	۰/۰۳۸۲	۰/۵۲۶۰
(tf)2	-۰/۱۷۵۵	۰/۲۹۱۹	-۰/۶۰۱	۰/۵۴۷۷	۰/۸۳۹۱

مأخذ: محاسبات تحقیق

در جدول فوق ملاحظه می‌شود که در سطح ۰/۰۵ بجز یک متغیر سایر داده‌ها معنادار هستند. لذا شکل کلی تابع لاجیت به صورت زیر در خواهد آمد:

$$Z_i = \ln\left(\frac{P_i}{1-P_i}\right) = 1/8264 + 17/0165\, c + 17/0498\, b + 1/5523\, p + 0/853\, (g)1 - 2/4057\, (co)1 - 1.1398\, (co)2 - 0/6425\, (tf)1 - 0/1755\, (tf)2 \qquad (۸)$$

بدین ترتیب با برآورد ضرایب، مدلی حاصل می‌شود که هر یک از ضرایب آن مقدار متغیر (لگاریتم شانس به نفع قصور در بازپرداخت) را به ازای یک واحد تغییر متغیر مستقل نشان می‌دهد با محاسبه احتمال قصور در بازپرداخت از رابطه زیر می‌توان مشتریان بانک را طبقه بندی نمود

$$P_i = \frac{1}{1 + e^{-Z_i}} \qquad (۹)$$

.

در رگرسیون لاجیت از شاخص‌های مندرج در جدول به منظور بررسی نکویی برازش مدل استفاده می‌شود:

جدول ۷ : معناداری و نیکویی برازش

شاخص	مقدار	احتمال
LR(8df)	۳۰/۰۹۴۴	۰/۰۰۰۲۰۳
Lemeshow chi-Hosmer(8)	۹/۲۶۷۸	۰/۳۲۰۲

مأخذ: محاسبات تحقیق

آماره تابع آزمون LR: این آماره دارای توزیع کای دو با درجه آزادی ۸ (تعداد متغیرهای مستقل) می‌باشد. و برای آزمون فرضیه بی اثر بودن متغیرهای مستقل بر متغیر وابسته به کار می‌رود. به بیان دیگر، در این آماره فرضیه مساوی صفر بودن تمامی ضرایب متغیرهای مستقل در مقابل مخالف صفر بودن آن‌ها آزمون می‌شود. همانطور که ملاحظه می‌شود احتمال این آماره برابر۰/۰۰۰۲ و در نتیجه کمتر از ۰/۰۵ می‌باشد و نشان می‌دهد که که فرض صفر مبنی بر بی اثر بودن تمام متغیرهای مستقل رد شده و در نتیجه رگرسیون معنادار است.

آماره هاسمر لمشو[۱]: در این آماره مشاهدات به j=10 گروه مساوی تقسیم می‌شود و آماره دارای توزیع کای دو با j-2=8 درجه آزادی می‌باشد. این آزمون مقادیر برازش شده مورد انتظار را با مقادیر واقعی هرگروه مقایسه می‌کند. اگر اختلافات بزرگ باشد مدل رد می‌شود. (نظر دهمرده و دیگران،۱۳۹۱). همانطور که ملاحظه می‌شود مقدار این آماره برابر ۲/۲۶۷۸ و احتمال آن بزرگتر از ۰/۰۵ ۰محاسبه شده است. لذا در اینجا فرضیه صفر پذیرفته می‌شود که حاکی از صحت تعیین مدل است.

[۱] Hosmer , Lemeshow

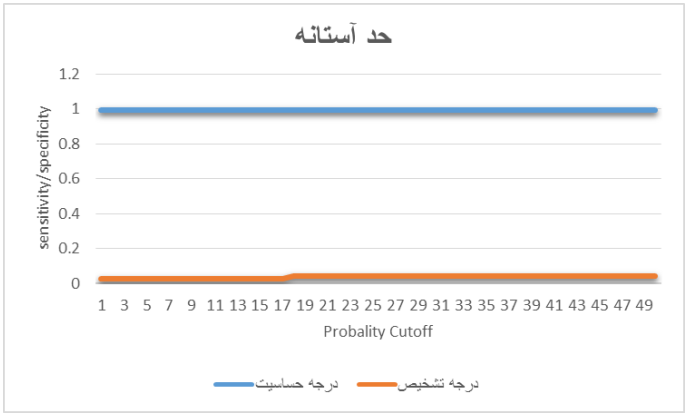
بررسی قدرت پیش بینی مدل

جدول ۸: پیش بینی مدل رگرسیون لجستیک در حدآستانه ۰/۵

مشاهدات برآورد	$Y = 0$	$Y = 1$	کل
$p(y) \leq 0/5$	2	2	4
$p(y) > 0/5$	68	643	711
کل	70	645	715
درست	2	643	645
درصد درست	%۹/۲	%۹۹/۷	%۹۰/۷
درصد نادرست	%۹۹/۱	%۰/۳	%۹/۳

مأخذ: محاسبات تحقیق

بنابراین درجه حساسیت با توجه به جدول فوق برابر 99/7 درصد و درجه تشخیص برابر 2/9درصد می‌باشد. همچنین خطای نوع اول (عدد یک منهای درجه حساسیت) یعنی ریسک تجاری در مدل برازش شده برابر 0/3 درصد و خطای نوع دوم 97/1 درصد (عدد یک منهای درجه تشخیص) یعنی ریسک اعتباری می‌باشد.بنابراین با توجه به مطالب گفته شده در قسمت شبکه عصبی برای رسیدن به نتیجه مطلوب درمورد رگرسیون لجستیک نیز باید حد آستانه بهینه محاسبه شود. با روش اعتبارسنجی ضربدری و ۱۰۰ بار اجرای مدل و در نهایت میانگین گرفتن ازدرجه حساسیت و درجه تشخیص مدل‌های مختلف، نمودار درجه حساسیت و درجه تشخیص به صورت زیر ترسیم می گردد.



نمودار٤: منحنی درجه حساسیت و درجه تشخیص برای رگرسیون لجستیک

مأخذ: نتایج انجام تحقیق

اما ملاحظه می‌شود که در هیچ نقطه ای منحنی درجه تشخیص و درجه حساسیت مدل یکدیگر را قطع نمی‌نمایند. چرا که برای مدل به دست آمده درجه حساسیت به شدت بالا و به یک همگرا می‌باشد اما درجه تشخیص به صفر همگرا می‌باشد لذا نمی‌توان نقطه ای را یافت تا این دو منحنی یکدیگر را قطع کنند.

در واقع علت این امر آن است که بانک مورد بررسی در مقایسه با سایر بانک‌ها عملکرد بسیار مناسبی در خصوص معوقات خود دارد و در این زمینه پیشتاز است. بنابراین در اینجا می‌توان بیان داشت که داده‌های مورد مطالعه از نوع داده‌های نادر می‌باشند چرا که تنها درصد کمی از مطالبات معوق می‌باشند.

آیا استفاده از رگرسیون لجستیک معمولی برای داده‌های نادر مجاز می‌باشد یا خیر.در این حالت مشکل آن است که برآورد ماکسیمم درستنمایی مدل لجستیک است که به خوبی مشخص شده به سمت نمونه کوچک تمایل ندارد. (کینگ و زنگ،۲۰۰۱)[۱]

[۱] King, Gary., Zeng, Langche

روش حداکثر درستنمایی جریمه شده[۱] یک رویکرد عمومی برای کاهش این تمایل در برآورد حداکثر درستنمایی می‌باشد. بدین ترتیب از روش برآورد حداکثر درستنمایی تاوان‌دیده به منظور به دست آوردن پارامترهای باثبات‌تر در مدل رگرسیون لجستیک استفاده می‌شود. به جای به حداکثر رساندن احتمال درستنمایی، MPLE احتمال درستنمایی، جریمه مدل را حداکثر می‌نماید، که این احتمال توسط یک عامل جریمه تنظیم می‌شود.

$$Log\ L - 0/5\ \lambda\ \sum(\ s_i\beta_i)^2 \qquad (۱۰)$$

که در معادله فوق، L حداکثر درستنمایی مدل برازش شده ، λ عامل جریمه، β ضرایب رگرسیون برازش شده برای هر متغیر مستقل i در مدل می‌باشد و s_i یک پیمایش برای بی واحد نمودن β_i می‌باشد. معمولا انحراف استاندارد یکی از متغیرهایی است که برای بی واحد نمودن پارامتر رگرسیون مربوطه استفاده می‌شود.

یکی از مسائل مهم در استفاده از MPLE نحوه انتخاب بهینه عامل جریمه λ می‌باشد. بسیاری از محققان از روش اعتبارسنجی ضربدری[۲]به منظور برآورد عامل جریمه استفاده می‌کنند. با این حال حداکثر سازی معیار AIC [۳] نیز یک روش کارآمد برای این منظور می‌باشد.(هارل،۲۰۰۱)[۴]

برآورد مدل رگرسیون با روش MPLE

برای برآورد ماکسیمم درستنمایی جریمه نیز ۸۰٪ داده‌ها به طور تصادفی برای طراحی مدل و ۲۰٪ داده‌های تصادفی دیگر را به عنوان داده‌های شاهد در نظر می‌گیریم تا پس از برآورد با نتایج واقعی مقایسه شود.

---

[۱] Maximum Penalized Likelihood (MPLE)

[۲] cross-validation

[۳] Akaike's information criterion

[۴] FE Jr, Harrell. (2001)

با شبیه سازی مدل با روش اعتبار سنجی ضربدری و ۱۰۰ بار اجرای مدل و در نهایت میانگین گرفتن از عامل جریمه، اطلاعات مربوط به جدول زیر به دست می‌آید:

جدول ۹ : خروجی روش MPLE

نام متغیر i	β	St d.Error	Chi sq	p
مقدار ثابت	1/81656	0/2799724	۴۸/۶۲۸۵۶۷۷	0/00000
c	16/3521	7/4628211	4/594277	0/0320789
b	20/5442	9/5094267	7/5079631	0/0061427
p	1/29466	0/6577706	4/3438177	0/0371436
(g)1	0/83277	0/3847143	5/4182856	0/0199269
(co)1	-2/6575	0/9997499	5/7789386	0/0162193
(co)2	-1/1294	-0/5206847	3/8242365	0/0505164
(tf)1	-0/6246	0/3435549	3/2258798	0/0724828
(tf)2	-0/2226	0/3115911	0/5041324	0/5041324

مأخذ: محاسبات تحقیق

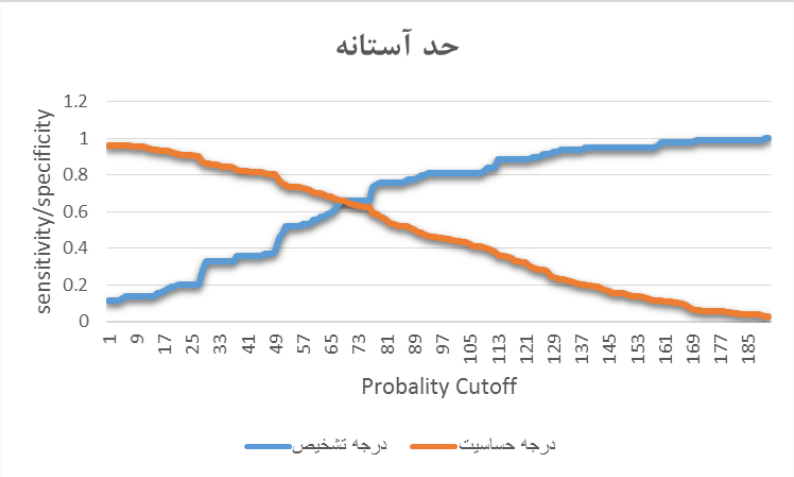
تفسیر ضرایب در اینجا نیز مشابه برازش مدل با روش MLE می‌باشد.(کرشلم،۲۰۰۴)[۱]

در اینجا فرضیه صفر بدین معناست که متغیر مذکور هیچگونه اثری بر متغیر وابسته ندارد لذا با توجه به جدول ملاحظه می‌گردد که سطح معنا داری برای کلیه متغیرها بجز متغیر(tf)2,(tf)1 کمتر از 0/05 می‌باشد که نشان دهنده معنا دار بودن آن ضرایب است.

حد آستانه بهینه مدل MPLE

در اینجا از روش اعتبار سنجی ضربدری استفاده می‌نماییم و مدل را ۱۰۰بار اجرا می‌نماییم و در نهایت میانگین گرفتن از درجه حساسیت و درجه تشخیص مدل‌های مختلف، نمودار درجه حساسیت و درجه تشخیص به صورت زیر ترسیم می‌شود.

[۱] Korsholm, Lars



نمودار ۵ : منحنی درجه تشخیص و درجه حساسیت برای روش MPLE

مأخذ: نتایج انجام تحقیق

باتوجه به نمودار ملاحظه می‌شود که دو نمودار منحنی تشخیص و منحنی حساسیت در نقطه x=67/9293 یکدیگر را قطع نموده و برابر با 0/6564 می‌باشند. برای به دست آورد حدآستانه بهینه کافی است x را در معادله زیر که در مدل تعریف نمودیم قرار دهیم. در اینجا مقدار اولیه را برابر 0/8 قرار دادیم و با گام 0/001 جلو می‌رویم لذا داریم:

$$th = 0/8 + (x * 0/001) \qquad (۱۱)$$

بنابراین در اینجا حد آستانه بهینه برابر ۰/۸۶۷۹۲ می‌باشد و نتایج مدل را با این حد آستانه گزارش می‌نماییم.

جدول۱۰ : پیش بینی مدل MPLE با حد آستانه 0/8679برای داده‌های آموزش

مشاهدات برآورد	$Y=0$	$Y=1$	کل
$p(y) \leq 0/8679$	۵۱	۲۰۶	۲۵۷
$p(y) > 0/8679$	۲۱	۴۳۷	۴۵۸
کل	۷۲	۶۴۳	۷۱۵
درست	۵۱	۴۳۷	۴۸۸
درصد درست	٪70/83	٪67/96	٪68/25
درصد نادرست	٪29/17	٪32/04	٪31/75

مأخذ: محاسبات تحقیق

باتوجه به جدول فوق خطای نوع دوم مدل برابر 29/17 درصد برازش شده یعنی ریسک اعتباری در این مدل برابر 29/17 در صد می‌با شد و درجه حساسیت در این روش یعنی متقا ضیان خوش حسابی که به درستی پیش بینی شده اند برابر67/96 درصد و خطای نوع اول یا ریسک تجاری برابر 32/04 درصد به دست آمده است.

جدول ۱۱ : پیش بینی مدل MPLE با حد آستانه ۰,۸٦۷۹ برای داده‌های تست

مشاهدات برآورد	$Y=0$	$Y=1$	کل
$p(y) \leq 0/8679$	۱۹	۷۰	۸۹
$p(y) > 0/8679$	۵	۸۴	۸۹
کل	۲۴	۱۵۴	۱۷۸
درست	۱۹	۸۴	۱۰۲
درصد درست	79/17٪	٪54/54	57/3٪
درصد نادرست	٪20/83	٪45/46	42/7٪

مأخذ: محاسبات تحقیق

با توجه به جدول ۱۱ خطای نوع اول یا ریسک تجاری 45/46 درصد و ریسک اعتباری یا خطای نوع دوم برابر 20/83 درصد در داده‌های تست می‌باشد.

منحنی ROC و قدرت تفکیک کنندگی

به منظور بررسی قدرت تفکیک کنندگی دو گروه (مشتریان خوش حساب و مشتریان بدحساب) از منحنی ROC[۱] استفاده می‌نماییم. این منحنی در صفحه مختصاتی که محور افقی آن (عدد یک منهای درجه تشخیص) و محور عمودی آن درجه حساسیت مدل می باشد, از نقطه (۰و۰) در گوشه پایین سمت چپ به سمت (۱و۱) در گوشه بالا، سمت راست کشیده می شود.

منحنی ROC نشان دهنده چند نکته اساسی می‌باشد:

۱. این منحنی نشان دهنده موازنه بین درجه حساسیت و درجه تشخیص می‌باشد (هر گونه افزایش در حساسیت همراه با کاهش در ویژگی خواهد بود).
۲. هرچه این منحنی به سمت گوشه سمت چپ بالا (۱و۰) نزدیکتر باشد، نشان دهنده قدرت بیشتر مدل در تفکیک کنندگی بین دو گروه خواهد.
۳. هرچه به خط مورب ۴۵ درجه نزدیک‌تر باشد، نشان دهنده قدرت کمتر مدل می‌باشد.
۴. سطح زیر منحنی به منظور اندازه گیری دقت مدل است.

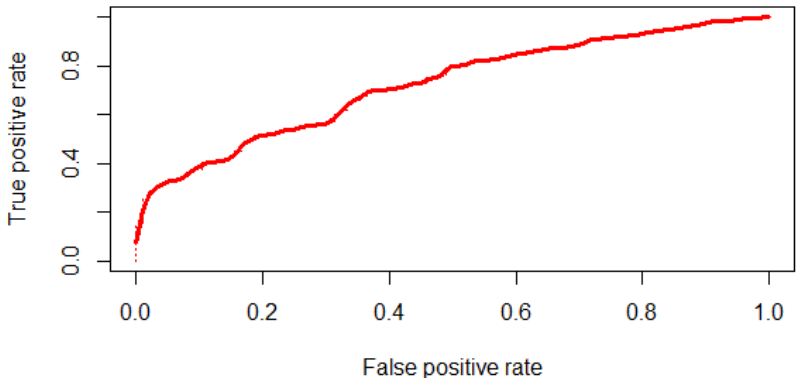
مدلی که بیشترین قدرت تفکیک کنندگی بین دو گروه (از مشتریان خوش حساب و بدحساب) را دارد، سطح زیر منحنی ۱ و مدلی که قدرت تفکیک کنندگی آن صفر است دارای سطح زیر منحنی ۰,۵ می‌باشد(متز و دیگران،۱۹۹۸)[۲] .

---

[۱] Receiver Operating Characteristic, ROC Curve

[۲] Metz et al, 1998

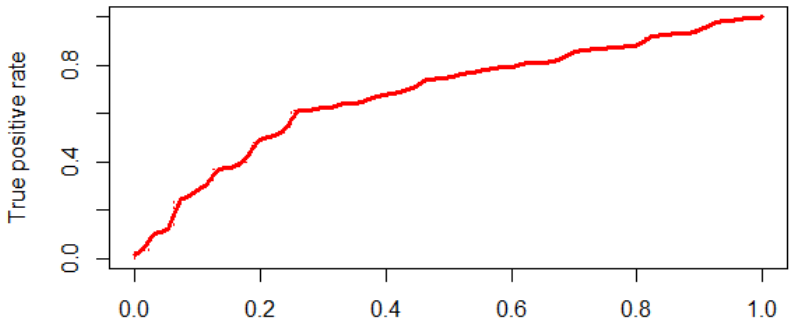
در زیر نمودار ROC مربوط به مدل‌های ارائه شده در تحقیق را رسم نموده ایم و مساحت زیر منحنی راک را برای هر یک از مدل‌ها محاسبه نمودیم.



نمودار ٦: منحنی ROC برای مدل رگرسیون لجستیک

مأخذ: نتایج انجام تحقیق

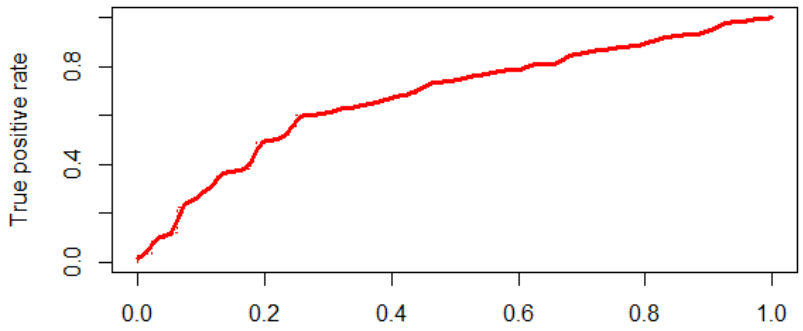
در مدل شبکه عصبی سطح زیر منحنی ROC برابر 0/752122می‌باشد که نشان از قدرت تفکیک کنندگی بالای مدل می‌باشد.



نمودار ۷ : منحنی ROC برای مدل شبکه عصبی

مأخذ: نتایج انجام تحقیق

با توجه به نمودار ۷ مساحت زیر منحنی برای مدل رگرسیون لجستیک برابر0/6816055 می-باشد.



نمودار ۸ : منحنی ROC برای روش حداکثر درستنمایی تاواندیده در رگرسیون لجستیک

مأخذ: نتایج انجام تحقیق

در این روش سطح زیر منحنی ارائه شده در نمودار فوق برابر 0/69768 می‌باشد.

ملاحظه می‌شود که در بین مدل‌های بررسی شده در مطالعه ی حاضر مدل شبکه عصبی از قدرت تفکیک پذیری بیشتری برخوردار می‌باشد. بعد از آن روش حداکثر درستنمایی تاواندیده برازش گردید با مساحت زیر منحنی 0/7 و سپس مدل رگرسیون لجستیک است.

کارایی سنجش روش‌ها

دراین بخش میزان خطای پیش‌بینی روند متغیر در روش‌های ارائه شده، با معیارهای میانگین مربعات خطا (MSE)، ریشه میانگین مربعات خطا[۱](RMSE)، میانگین قدرمطلق خطا[۲] (MAE) سنجیده می‌شود:

$$MES = \frac{1}{\mathrm{T}}\sum(P-A)^2 \qquad (۱۱)$$

$$\mathrm{RMSE}=\sqrt{\frac{1}{\mathrm{T}}\sum(P-A)^2} \qquad (۱۲)$$

$$\mathrm{MAE}=\frac{1}{T}\sum|P-A| \qquad (۱۳)$$

۱- Root Mean Squared Error

۲- Mean Absolute Error

در قسمت‌های قبلی داده‌های آموزش و داده‌های تست به صورت تصادفی انتخاب می‌شدند، در این قسمت به بررسی کارایی مدل‌های ارائه شده با داده‌های یکسان می‌پردازیم:

جدول ۱۲: مقایسه میزان خطای مدل‌ها

مدل	MSE	RMSE	MAE
شبکه عصبی	0/09027	0/3	0/181
برای رگرسیون لجستیک MPLE	0/092714	0/3045	0/188518

مأخذ: محاسبات تحقیق

براساس خطای روش‌های ارائه شده در جدول فوق روش شبکه عصبی با اختلاف اندکی از روش MPLE برای برآورد رگرسیون برتر بوده و با هم رقابت می‌کنند.

جدول ۱۳: مقایسه کارایی مدل‌ها

وضعیت واقعی				
مدل	پیش بینی	Y = ۰	Y = ۱	
	Y = ۰	۶۰	۲۴۲	
	Y = ۱	۳۶	۵۵۵	
شبکه عصبی	درجه	۶۹/۶۴	خطای نوع اول	۳۰/۳۶ درصد
	درجه	۵/۶۲	خطای نوع دوم	۳۷/۵۰ درصد
	صحت	۶۸/۸۷	خطای کلی	۳۱/۱۳ درصد
	Y = ۰	۰	۱	
	Y = ۱	۹۶	۷۹۶	
رگرسیون لجستیک	درجه	۹۹/۸۷	خطای نوع اول	۰/۱۳ درصد
	درجه	۰۰۰۰۰/	خطای نوع دوم	۱۰۰ درصد
	صحت	۸۹/۱۴	خطای کلی	۱۰/۸۶۰ درصد
	Y = ۰	۵۴	۲۳۵	
	Y = ۱	۴۲	۵۶۲	
MPLE	درجه	70/50	خطای نوع اول	۲۹/۴۹ درصد
	درجه	56/25	خطای نوع دوم	۴۳/۷۵ درصد
	صحت	69	خطای کلی	۳۱

مأخذ: محاسبات تحقیق

- ممکن است هدف یافتن مشتریان بد حساب باشد تا بدین طریق ضرر وارده به بانک را کمینه شود.

بانک مورد مطالعه در اعطای تسهیلات خرد بسیار موفق عمل کرده به طوریکه کمتر از ۱۰ درصد مطالبات معوق در این نوع تسهیلات داشته است لذا ما در اینجا باید به دنبال روشی باشیم که بتوانیم همین تعداد معدود مشتری بدحساب را شناسایی نماییم تا در اعطای تسهیلات عملکرد بهتری از قبل داشته باشیم. در نتیجه باید مدلی را بیابیم که درپیش بینی مشتریان بد حساب به درستی عملکرد بهتری داشته باشد و این بدان معناست که باید ازمدل با ریسک اعتباری پایین‌تر استفاده کنیم بنابراین مدلی که خطای نوع دوم پایین‌تری دارد، هدف می‌باشد. با توجه به مدل‌های ارائه شده و نتایج به دست آمده مدل رگرسیون لجستیکی که با روش حداکثر درستنمایی برآورد شده است دارای بیشترین خطای نوع دوم با مقدار ۱۰۰ درصد می‌باشد که مدلی کاملا ناکارا می‌باشد. با این رویکرد می‌توان مدل‌ها را براساس نحوه عملکرد به صورت زیر طبقه‌بندی نمود:

جدول ١٤: طبقه‌بندی براساس ریسک اعتباری یا درجه تشخیص

ردیف	مدل	درجه تشخیص	خطای نوع دوم یا ریسک اعتباری
۱	شبکه عصبی	٪62/5	٪37/5
۲	MPLE	٪56/25	٪43/75
۳	رگرسیون لجستیک	٪0000	٪100

مأخذ: محاسبات تحقیق

- ممکن است بانک تنها بخواهد مشتریان خوش حساب را شناسایی نماید و تنها به مشتریان خوش حساب وام اعطا نماید بنابراین مدلی که خطای نوع اول پایین‌تری دارد، هدف است. بعد از روش رگرسیون لجستیک که به لحاظ معنایی در گروه داده‌های مورد استفاده در این پژوهش به دلیل نادر بودن داده‌ها ناکارا برآورد گردید، مدل رگرسیون لجستیک با برآورد حداکثر

درستنمایی تاواندیده (MPLE) با درجه حساسیت ۷۰/۵۱ درصد و خطای نوع اول ۲۹/۴۹ درصد مدل مناسبی با اختلاف اندک از شبکه عصبی است.

جدول ۱۵: طبقه‌بندی براساس درجه حساسیت یا خطای نوع اول

ردیف	مدل	درجه حساسیت	خطای نوع اول
۱	MPLE	۷۰٪۵۱	٪49/29
۲	شبکه عصبی	64/69%	٪36/30

مأخذ: محاسبات تحقیق

- ممکن است هدف تفکیک مشتریان به درستی باشد یعنی در اینجا مدلی باید انتخاب شود که مشتریان خوش حساب و بد حساب را با دقت بیشتری شناسایی نموده است یعنی مدلی که دارای صحت کلی بالاتر و خطای کلی کمتری باشد، تا بدین طریق ضرر وارده به بانک را کمتر گردد. روش MPLE و شبکه عصبی با اختلاف اندکی با یکدیگر رقابت می‌کنند.

جدول ۱٦: طبقه‌بندی براساس صحت کلی یا خطای کلی مدل‌ها

ردیف	مدل	درجه تشخیص	خطای نوع دوم یا ریسک اعتباری
۱	MPLE	٪69	٪31
۲	شبکه عصبی	٪78/68	٪31/13

مأخذ: محاسبات تحقیق

٦. نتیجه‌گیری

با توجه به مدلسازی و نتایج بدست آمده از الگوهای سه گانه بکار رفته در مقاله موارد زیر را می توان ذکر کرد.

۱- هیچ یک از مدل‌ها به دلیل نامتقارن بودن داده‌های مربوط به مشتریان خوش حساب و بد حساب در حد آستانه ۰/۵ دارای کارایی لازم نبودند لذا به منظور یافتن حد آستانه بهینه از محل تلاقی دو منحنی درجه حساسیت و درجه تشخیص استفاده شد.

۲- برای مدل شبکه عصبی حد آستانه جدید ۰/۸۷۶۳ بدست آمد برای مدل رگرسیون لجستیک با توجه به ویژگی‌های مدل عملا حد آستانه بهینه جدید نمی‌توانست تعیین شود لذا مدل رگرسیون لجستیک MPLE به مدلسازی اضافه شده و بر اساس شاخص‌های تعریف شده حد آستانه این مدل ۰/۸۶۷۹ بدست آمده است.

۳- به منظور برازش ضرایب مدل رگرسیون لجستیک از روش حداکثر درستنمایی تاواندیده و برای بررسی قدرت تفکیک کنندگی مدل‌ها از منحنی راک استفاده شد. با رسم منحنی راک و به دست آوردن مساحت زیر منحنی در نهایت مدل‌ها براساس قدرت تفکیک‌کنندگی به ترتیب شبکه عصبی، رگرسیون لجستیک MPLE و در نهایت رگرسیون لجستیک طبقه‌بندی شد که در این طبقه‌بندی مدل شبکه عصبی دارای بیشترین قدرت تفکیک می‌باشند.

۴- کارایی روش ها با معیارهای میانگین مربعات خطا (MSE)، ریشه میانگین مربعات خطا[۱](RMSE)، میانگین قدرمطلق خطا[۲] (MAE) مورد بررسی قرار گرفت. که شبکه عصبی در اینجا دارای کمترین میانگین مربعات خطا و بعد از آن رگرسیون لجستیک MPLE دارای خطایی نزدیک به آن است.

۵- در این تحقیق یافته‌های مطالعات وانگ و ژائو(۲۰۲۲) و نگوین و تران(۲۰۲۳) که مقایسه عملکرد دو روش رگرسیون لجستیک و لجستیک MPLE را بررسی می‌کرد، مورد تایید قرار

۱- Root Mean Squared Error

۲- Mean Absolute Error

گرفت و نشان داده شد در داده‌های نامتقارن قدرت تفکیک کنندگی و کارائی روش MPLE بهتر از لجستیک است.

۶- در این مقاله یافته‌های تحقیق در نتایج مقالات پاتل و شارما(۲۰۲۴) و رابرتز و اوانز(۲۰۲۴) در مقایسه دو روش شبکه عصبی و رگرسیون سنتی مورد تایید قرار گرفت و روش شبکه عصبی نسبت به رگرسیون لجستیک برتری دارد.البته بعضا برای بانک‌ها روش رگرسیون لجستیک با توجه به تفسیر ساده ضرایب مورد کاربرد خواهد بود خصوصا زمانی که نتایج دو روش تفاوت قابل اغماضی داشته باشند.

سناریوهای ممکن جهت تعیین رویکرد بانک‌ها در استفاده از مدلسازی:

۱- اگر هدف بانک یافتن مشتریان بدحساب باشد تا زیان بانک را کمینه کنند بایستی از مدل‌هایی با کمترین خطای نوع دوم یا کمترین ریسک اعتباری ا ستفاده کنند که بر اساس گزارش مقاله بهترین عملکرد را روش شبکه عصبی دارد.

۲- اگر هدف یافتن مشتریان خوش حساب است و فقط به آنها وام بدهد در این صورت بایستی از مدلی با کمترین خطای نوع اول استفاده نماید که بر اساس نتایج این تحقیق رگرسیون لجستیک MPLE به عنوان بهترین مدل این مقاله توصیه می‌شود.

۳- در آخرین سناریو اگر تفکیک دقیق مشتریان هدف اصلی بانک باشد در این صورت مدلی که دارای صحت کلی بالاتر و خطای کلی کمتری باشد، توصیه می شود تا بدین طریق ضرر وارده به بانک را کمتر گردد.در این حالت روش MPLE و شبکه عصبی با اختلاف اندکی با یکدیگر رقابت می‌کنند.

References

Adabi firouzjaee B, Mehrara M, Mohammadi S. (2016). Estimation and Evaluation of Tehran Stock Exchange Value at Risk Based on Window Simulation Method. *Journal of Economic Modeling Research*. 6 (23), 35-73. (in Persian)

Andersen, T.G., Bollerslev, T., Christo¤ersen, P., Diebold, F.X. (2006). Volatility and Correlation Forecasting, in (ed.s) G. Elliott, C.W.J. Granger, and A. Timmermann, Handbook of Economic Forecasting, Vol. 1. Elsevier, Oxford.

Asayesh K, Fallahshams M, Jahangirnia H, Gholami Jamkarani R. (2020) Explaining the Systemic Risk Model Using the Marginal Expected Shortfall Approach (MES) for the Banks Listed on the Tehran Stock Exchange. JPBUD; 25 (2) :115-134. (in Persian)

Babalooyan, S., & Nikoomaram, H., & Vakilifard, H., & Rahnamay Roodposhti, F. (2020). Evaluating Value at Risk and Expected Shortfall for Tehran and International Stock Markets (Conditional Extreme Value Theory). JOURNAL OF FINANCIAL ECONOMICS (FINANCIAL ECONOMICS AND DEVELOPMENT), 14(52), 55-80. (in Persian)

Bu, D., Liao, Y., Shi, J., & Peng, H. (2019). Dynamic expected shortfall: A spectral decomposition of tail risk across time horizons. *Journal of Economic Dynamics & Control*, 108 (2019) 103753.

Creal, D.D., S.J. Koopman, and A. Lucas (2013). Generalized Autoregressive Score Models with Applications, *Journal of Applied Econometrics*, 28(5), 777-795.

Creal, D.D., S.J.Koopman, A. Lucas, & M. Zamojski (2015). Generalized Autoregressive Method of Moments, Tinbergen Institute Discussion Paper, TI 2015-138/III.

Diebold, F.X. and R.S. Mariano. (1995). Comparing predictive accuracy, *Journal of Business & Economic Statistics*,13(3), 253.263.

Engle, R.F. and S. Manganelli (2004a). CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles, *Journal of Business & Economic Statistics*, 22, 367-381.

Fallahshams, M., Saghafi, A., Naserpoor, A. (2016). Futures Contracts Margin Setting by General Pareto Distribution VaR, *Journal of Securities Exchange*, 9(33), 25-45. (in Persian)

Fissler, T., and J. F. Ziegel. (2016) Higher order elicitability and Osband.s principle, Annals of Statistics, 44(4), 1680-1707.

Gerlach,R., Wang,C. (2020). Semi-parametric dynamic asymmetric Laplace models for tail risk forecasting, incorporating realized measures, *International Journal of Forecasting*, 36, (2), 489-506.

Gneiting, T. (2011). Making and evaluating point forecasts. *Journal of the American Statistical Association*, 106(494), 746–762.

Gregoriou, Greg.N. (2009). The VaR Implementation Handbook, Volue I, McGraw-Hill,Inc.

Heidari, H., K. Haddad, G. (2017). Ranking of Parametric Value at Risk Models with Consideration of Trader Position (Application of Asymmetric Distribution Functions in GARCH Models). *Economics Research*, 17(66), 151-178. (in Persian)

Hallin, Marc & Trucíos, Carlos . (2020). Forecasting Value-at-Risk and Expected Shortfall in Large Portfolios: a General Dynamic Factor Approach, Working Papers ECARES 2020-50, Universite Libre de Bruxelles.

Meharani, A., Najafi Moghadam, A., Baghani, A. (2021). Estimation value at risk (VAR) and conditional value at risk (CoVaR) at Tehran Stock Exchange by approach to using Fréchet distribution (FD). Financial Engineering and Protfolio Management, 12(46),449-475. (in Persian)

Naderi Nooreini, M. (2018). The Best Methodology of Estimation of Value-at-Risk in Iranian Mutual Funds. *Asset Management and Financing*, 6(1), 159-180. (in Persian)

Naseri S A, Jabal Ameli F, Barkhordary Dorbash S. (2020). Investigating the Correlation of Selected Banks with Dynamic Conditional Correlation (DCC) Model and Identifying Systemically Important Banks with Conditional Value at Risk and Shapley Value Method. *Journal of Economic Modeling Research*. 11 (41) :145-196. (in Persian)

Nikola Radivojevi , Milena Cvjetkovi ,Saša Stepanov. (2016). The new hybrid value at risk approach based on the extreme value theory, Estudios the Economia.43, 29-52.

Patton, A.J. , Ziegel, J.F. , Chen, R. (2018). Dynamic semiparametric models for expected shortfall (and Value-at-Risk). J. Econom. 211 (2), 388–413

R. Roodposhti, F., Klantari Dehaghi, M. (2014). Investigation of Multifractaly Models in Finance. *Financial Knowledge of Securities Analysis*, 7(24), 25-47. (in Persian)

Saranj, A., Nourahmadii, M. (2016). Estimating of value at risk and expected shortfall by using conditional extreme value approach in Tehran Securities Exchange. *Financial Research Journal*, 18(3), 437-460. (in Persian)