مجموعه مقالات بیست‌وهفتمین همایش سالانه سیاست‌های پولی و ارزی
ثبات مالی بستر رشد اقتصادی پایدار
صفحات ۱۸۱ـ۲۰۳

اعتبارسنجی مشتریان در تسهیلات قرض‌الحسنه بانک صادرات ایران

سمانه افتخاری مهابادی [*] **نیلوفر ابراهیمی** [†]
حسن کیائی [‡]

یکی از مشکلات اساسی سیستم بانکداری کشور، مشکل مطالبات معوق و تسهیلات وصول نشده آنها است. تدوین مدل جهت سنجش ریسک اعتباری در نظام بانکی نقش کارآمدی در راستای افزایش بهره‌وری بانک‌ها و موسسات مالی و اعتباری خواهد داشت. هدف از این تحقیق شناسایی عوامل موثر بر ریسک اعتباری مشتریان بانک صادرات ایران و مدل‌سازی علل معوق شدن تسهیلات قرض‌الحسنه است. به این منظور، تاثیر مجموعه‌ای از متغیرهای مستقل شامل ارزش وثیقه، زمینه‌ی استفاده از تسهیلات، مدت زمان بازپرداخت، مبلغ تسهیلات و برخی خصوصیات فردی مشتریان یکی از شعب بانک صادرات ایران در دوره‌ای پنج ساله از سال ۱۳۹۰ تا ۱۳۹۴ مورد بررسی قرار گرفته‌است. برای مدل‌سازی علل معوق شدن تسهیلات از دو رویکرد مدلسازی آماری به روش رگرسیون لوژستیک و روش‌های مختلف یادگیری آماری برای طبقه بندی (classification) استفاده نموده و جهت تدوین مناسب ترین مدل به مقایسه‌ی نتایج آن‌ها با یکدیگر می‌پردازیم. نتایج تحلیل رگرسیون لوژستیک نشان می‌دهند که متغیرهای وضعیت تاهل، وضعیت مالکیت مسکن و ارزش وثیقه بر ریسک اعتباری متقاضیان تسهیلات قرض‌الحسنه اثر معناداری دارند. همچنین از میان روش‌های مدل‌سازی مختلف، شبکه‌های عصبی و پس از آن روش بوستینگ، بیشترین دقت در پیش‌گویی وضعیت اعتباری فرد را کسب نموده‌اند.

[*] عضو هیئت علمی دانشکده ریاضی، آمار و علوم کامپیوتر دانشگاه تهران، تهران؛ seftekhari@ut.ac.ir
[†] دانشجوی کارشناسی دانشکده ریاضی، آمار و علوم کامپیوتر دانشگاه تهران؛
niloofarebrahimi812@gmail.com
[‡] عضو هیئت علمی دانگاه امام صادق (ع)؛ kiaee@isu.ac.ir

۱ مقدمه

در سراسر جهان، صنعت بانکداری یکی از ارکان بسیار مهم اقتصاد به شمار می‌رود و به دلیل ارائه‌ی خدمات متنوع مالی و اعتباری، نقش تعیین کننده‌ای در توسعه و رشد اقتصادی ایفا می‌کند. به طوری‌که می‌توان از آن به عنوان نیروی محرکه، شتاب دهنده، متعادل کننده و سامان بخش اقتصاد یاد کرد. نگاهی به تاریخچه‌ی شکل گیری بانک مؤید این است که این نهاد در داد و ستدهای درونی و برونی، مسئولیت مبادلات مالی و پولی را به عهده داشته و از بدو شکل گیری هم امین شهروندان و هم آسان کننده‌ی مبادلات پولی بوده و تأثیر بسزایی در اقتصاد داشته است؛ بنابراین توسعه و بهبود فعالیت‌های بانکی بویژه اعطای تسهیلات به همراه نظامی کارآمد، نقش عمده‌ای در توسعه و پیشرفت صنعت بانکداری و اقتصاد خواهد داشت. بانک به عنوان یک نهاد مالی با اعطای تسهیلات موجب جریان پول و سرمایه در جامعه می‌شود. با توجه به اهمیت بانک‌ها، به کارگیری معیارهای دقیق برای ارزیابی عملکرد آنها در راستای بالابردن استانداردهای مالی ضرورت می‌یابد. بانک‌ها در طول حیات خود با ریسک‌های مختلفی از جمله ریسک نقدینگی، اعتباری، تجاری، مالی، عدم توانایی در پرداخت، نرخ ارز، نرخ بهره، تورم و ... روبرو هستند. که در بین آنها ریسک اعتباری جایگاه ویژه‌ای دارد، چرا که به اولین و مهمترین نقش بانک در بازارهای مالی یعنی گردآوری سپرده و اعطای وام اشاره می کند. کاهش حاشیه سود بانک‌ها عمدتاً ناشی از عدم کارایی در مدیریت ریسک اعتباری بوده و بانک‌ها را متحمل فشارهایی جهت کاهش هزینه ها می نماید. بانک‌ها و مؤسسات مالی به دنبال هر سرمایه گذاری و اعطای وام ناچار به پذیرش ریسک‌هایی هستند که بازپرداخت وام‌ها را تحت تأثیر خود قرار می دهند. عواملی همچون تغییرات تکنولوژی، تغییر در نوع و سلیقه مصرف کنندگان، تغییرات نرخ ارز، جابجایی مدیریت و ... ممکن است با افزایش میزان ریسک اعطای تسهیلات، بخشی از منابع بانک را به خطر بیاندازند. احتمال عدم برگشت اصل و فرع تسهیلات اعطا شده را ریسک اعتباری می‌نامند.

در ایران، بانک‌ها بزرگترین نهاد مالی بوده و اعتبارآن‌ها رگ حیات سیستم بانکی و تمامی واحدهای اقتصادی است. همچنین توزیع اعتبارها برای هر دو بخش متأثر از سیاست‌های کلان اقتصادی، بخصوص در حوزه مالی و پولی است. از آنجا که بانک‌های کشور از لحاظ سرمایه بیشتر دولتی هستند، در نتیجه اولین توجه به منظور تأمین منابع به سوی بانکها معطوف می‌شود. افزایش مطالبات غیرجاری، موجب عدم تحقق اهداف ارائه‌ی تسهیلات بانکی در جهت رشد اقتصادی کشور و اتلاف منابع می‌گردد. در ایران نیز با توجه به ساختار اقتصادی کشور و به دلایلی همچون عدم توسعه بازارهای سرمایه و سایر شبکه های غیر بانکی

و قراردادی، تأمین مالی بخش‌های واقعی اقتصاد بر عهده‌ی شبکه بانکی کشور است. متأسفانه این بخش نیز در رسیدن به اهداف خود چندان موفق نبوده است. هم اکنون تداوم فعالیت‌ها و بقای بیشتر بانک ای کشور، ناشی از حمایت‌های دولتی است. بالا بودن ذخایر بانک ها و تسهیلات اعطایی سوخت شده و یا معوقه‌ی بانک‌ها، گویای نبود مدل‌های مناسب اندازه گیری ریسک اعتباری و سیستم های مدیریت ریسک در شبکه بانکی است (تهرانی و فلاح شمس، ۱۳۸۴: ۴۵).

در تحقیق حاضر برای اولین بار برای ایجاد مدل‌های پیش بینی عوامل مؤثر بر ریسک اعتباری مشتریان بانک از الگوریتم NN شبکه ی عصبی مصنوعی، رگرسیون لوژستیک و روش‌های دیگر طبقه‌بندی نظیر انواع درخت‌های تصمیم استفاده شده است و سعی شده تا مدلی برای پیش‌بینی احتمال معوق شدن تسهیلات قرض‌الحسنه از سوی مشتریان بر اساس ویژگی‌های فردی، اجتماعی و اقتصادی ثبت شده برای آن‌ها طراحی شود.

۱-۱ پیشینه تحقیق

به احتمال عدم بازپرداخت یا پرداخت با تأخیر اصل و فرع تسهیلات اعطایی بانک‌ها و سایر بدهی‌ها از سوی مشتری، ریسک اعتباری گویند. طراحی مدلی برای اندازه‌گیری و درجه بندی ریسک اعتباری برای نخستین بار در سال ۱۹۰۹ توسط «جان موری» بر روی اوراق قرضه انجام شد. مشابهت زیاد تسهیلات اعتباری بانک‌ها به اوراق قرضه سبب شد که درجه بندی ریسک اعتباری تسهیلات بانک ها، یعنی اندازه‌گیری ریسک عدم بازپرداخت اصل و بهره وام ها از سوی برخی از محققان مورد توجه قرار گیرد. از جمله مطالعات اولیه انجام شده در این زمینه می توان به کارهای دیکین[1] (۱۹۶۴) در زمینه طراحی مدل نمره دهی، مطالعات آلتمن (Altman, E. I. ۱۹۶۸) در طراحی مدل نمره Z، مورگان[2] (۱۹۹۴) در طراحی مدل اعتبارسنجی و کارهای تریسی[3] (۱۹۹۸) در طراحی مدل ارزش در معرض خطر[4] برای تخمین تابع چگالی احتمال عدم بازپرداخت اشاره نمود. در سال‌های دهه‌ی ۱۹۷۰ تحولات بازارهای مالی و ورشکستگی چندین بانک معتبر جهانی همچون «فرانکلین نشنال» آمریکا و «هرشتات» آلمان، مقامات بانکداری را به تأمل در مورد چالش های فراروی آنها و استفاده از مدل‌هایی برای اندازه گیری ریسک بانک به‌طورعام و ریسک اعتباری به‌طورخاص وادار

[1] Deaken
[2] Morgan
[3] Treasy
[4] value at risk

نمود (Glant, 2003:9). هم اکنون همگام با مطالعات کمیته بال (کمیته نظارت بر مقررات بانکی، بانک تسویه بین‌المللی)، تحقیقات بسیاری از سوی مؤسسات اعتباری برای طراحی مدل دقیق اندازه‌گیری ریسک اعتباری و معوق شدن مطالبات انجام می‌شود. همچنین مدل‌های بسیاری با استفاده از روش‌های آماری و شبکه‌های عصبی و منطق فازی برای اندازه گیری و پیش‌بینی مطالبات معوق در بانک ها مورد استفاده قرار می‌گیرد. درادامه به برخی از این تحقیقات اشاره مختصری می نماییم.

۱-۱-۱ تحقیقات داخلی

نورا و همکاران (۱۳۹۰) در مطالعه‌ای به بررسی مفهوم مدیریت ریسک و تأثیر آن بر عملکرد بانکها پرداختند. آنها برای ارزیابی کارایی از مدل تحلیل پوشش داده‌های شبکه‌ای استفاده کردند و مدل مدنظر را در دو حالت مورد تحلیل قرار دادند. بدین ترتیب که در مدل اول عملکرد بانک‌ها بدون اعمال ریسک و در مدل دوم عملکرد بانک‌ها با اعمال ریسک بررسی شد. سپس اثر و اهمیت ریسک را با بکارگیری مدل‌های محدودیت وزنی بررسی نمودند. یافته‌های این مقاله نشان می‌دهد که مدلی مبتنی بر محدودیت وزنی، کارایی منطقی‌تری نسبت به مدل بدون ریسک و مدل با احتساب ریسک ایجاد می‌کند. پناهیان و ابیاک (۱۳۹۲) در مطالعه‌ای به شناسایی و بررسی تأثیر ریسک‌های اعتباری، عملیاتی و بازار بر کارایی سه بانک ملی، صادرات و ملت طی سالهای ۱۳۸۷ ـ ۱۳۸۹ پرداختند. آنها برای این منظور از یک روش دو مرحله‌ای استفاده کردند. در مرحله‌ی اول با استفاده از روش DEA و انتخاب تعداد شعب، حجم سپرده بانک و مجموعه هزینه‌ها به عنوان ورودی و همچنین حجم تسهیلات و مجموع درآمد به عنوان خروجی به ارزیابی کارایی این بانک‌ها پرداختند. سپس در مرحله دوم با در اختیار داشتن کارایی از مرحله اول، یک مدل رگرسیونی را که متغیر وابسته آن کارایی و متغیر مستقل آن انواع ریسک بوده‌است، تصریح نمودند و این مدل را با استفاده از روش داده‌های تابلویی برآورد کردند. که نتایج حاکی از آن است که نسبت تسهیلات نکول شده به کل تسهیلات اعطایی به عنوان شاخص ریسک اعتباری ارتباط معنی‌داری با کارایی دارد. عزیزی و همکاران (۱۳۹۲) به بررسی نقش مدیریت ریسک بر کارایی شعب بانک انصار استان گیلان پرداختند. نتایج نشان داد که مدیریت ریسک عاملی مهم بر کارایی است و بر عملکرد بانک مؤثر است و هرگاه واحدهای ناکارا به متغیر مدیریت ریسک توجه زیادی داشتند، تأثیر مستقیمی بر کارایی‌شان داشته و بر تعداد واحدهای کارا افزوده شده است.

در تحقیق پورزمانی و همکاران (۱۳۸۸) نسبت های مالی شرکت‌های ناموفق در مقابل شرکت‌های موفق بورس مورد بررسی قرارگرفته و به دنبال پیش بینی موفقیت یا عدم موفقیت

شرکت‌ها، نتیجه شده که بر اساس داده های یکسال قبل، روش تحلیل ممیزی و بر اساس داده های دو سال قبل، روش رگرسیون لوژستیک پیش بینی بهتری دارند ولی در مجموع این دو روش اختلاف معناداری با یکدیگر ندارند. در تحقیقی که در سال ۱۳۸۷ توسط مؤسسه عالی بانکداری ایران با عنوان «طراحی و تبیین مدل ریسک اعتباری در نظام بانکی کشور» صورت پذیرفت، به منظور رسیدن به مدلی کارا برای تبیین ریسک اعتباری نظام بانکی کشور، از سه مدل احتمالی خطی، لوجستیک و شبکه عصبی استفاده شد. متغیرهای پیشگو در هرکدام از این مدلها، نسبت‌های مالی مشتریان بانک‌ها بود. نتایج این تحقیق نشان می‌دهد که ارتباط بین متغیرها در مدل‌های پیش بینی ریسک اعتباری به صورت خطی نبوده و توابع نمایی و سیگموئید مناسبترین مدلها برای پیش بینی ریسک اعتباری می باشند. در این تحقیق، بیشترین قدرت پیش بینی به ترتیب مربوط به شبکه عصبی و رگرسیون لوژستیک بوده است. جعفر برادری (۱۳۸۶) در پایان نامه اش تحت عنوان «بررسی وضعیت و عوامل مؤثر بر پیدایش مطالبات معوق و ارائه راهکارهای مطلوب پیشگیری آن در بانک صادرات ایران بر اساس مدل Moral Hazard (مطالعه موردی بانک صادرات استان تهران)» به این نتیجه رسید که از میان عوامل رفتاری مورد بررسی، متغیرهای رشد فعالیت متقاضی، نوع مالکیت، مدرک تحصیلی، نوع وثیقه، نوع عقد، مقطع اخذ تسهیلات، سابقه اعتباری متقاضی، داشتن چک برگشتی، نسبت سرمایه به مبلغ تسهیلات دریافتی، با متغیر وابسته (پیدایش مطالبات معوق) رابطه معنی دار داشته است. همچنین وجود اختلاف معنادار بین نرخ سود بانکی و نرخ بازار و منافع حاصل از عدم بازپرداخت ازسوی متقاضیان، به عنوان یکی از عوامل عمده در بروز مطالبات معوق، مورد تأیید قرارگرفته است. مدرس و ذکاوت (۱۳۸۲) در مقاله خود با عنوان «مدلهای ریسک اعتباری مشتریان بانک (مطالعه موردی مشتریان بانک توسعه صادرات ایران)»، به بررسی ریسک اعتباری برای مشتریان حقوقی این بانک می پردازد. آنها در این تحقیق برای مدل سازی ریسک اعتباری مشتریان بانک، از دو روش تحلیل تشخیصی و تحلیل رگرسیون لوژستیک استفاده نمودند. نتایج این تحقیق عبارتند از: ۱ـ مدل های معنی دار آماری بر اساس تحلیل ممیزی و رگرسیون لوژستیک به منظور تعیین وضعیت ریسک اعتباری شرکت‌های مشتری بانک توسعه صادرات وجود دارد. ۲ـ بر اساس متغیرهای مالی، شرکت‌های مشتری بانک توسعه صادرات ایران را می‌توان از نظر ریسک اعتباری دسته بندی کرد. ۳ـ از بین متغیر های مالی استفاده شده در این تحقیق، متغیر نسبت جاری بیشترین سهم را در تفکیک مشتریان به دو گروه خوش حساب و بد حساب دارد. ۴ـ روش‌های تحلیل ممیزی و رگرسیون لوژستیک در رابطه با دسته بندی شرکت

های مشتری بانک توسعه صادرات ایران ازنظر ریسک اعتباری، نتایج تقریباً مشابهی را ارائه می دهند.

۲-۱-۱ تحقیقات خارجی

آبیالا و اولاسی[1] (۲۰۱۴) به بررسی اثر ریسک اعتباری بر عملکرد بانک‌های تجاری در نیجریه پرداختند. آنها برای برآورد مدل مدنظر از اطلاعات گزارش مالی ۷ بانک تجاری طی ۷ سال ۲۰۰۵ الی ۲۰۱۱ و روش داده‌های تابلویی استفاده کردند. برای این منظور آنها از بازده حقوق صاحبان سهام و بازدهی دارایی به عنوان شاخص‌های عملکرد بانکها و از نسبت کفایت سرمایه و وام‌های غیرعملیاتی به عنوان شاخص‌های ریسک اعتباری استفاده کردند. نتایج حاکی از آن است که مدیریت ریسک اعتباری اثر معنی‌دار و مثبتی بر عملکرد بانک‌های تجاری در نیجریه دارد. چارلز و کنس[2] (۲۰۱۳) به بررسی اثر مدیریت ریسک اعتباری و کفایت سرمایه بر عملکرد مالی بانک‌های تجاری در نیجریه پرداختند. برای این منظور مدلی بیان شده است که متغیر وابسته آن بازدهی دارایی و متغیرهای مستقل آن کفایت سرمایه و شاخص‌های مربوط به مدیریت ریسک اعتباری است. در سال ۲۰۱۰ دانگ و دیگران تحقیقی با عنوان کارت امتیازدهی اعتباری بر اساس رگرسیون لجوژستیک با استفاده از ضرایب تصادفی انجام دادند. در این تحقیق، برای ارزیابی اعتباری مشتریان بانک از روش رگرسیون لوژستیک و برای بالا بردن دقت پیش بینی از ضرایب تصادفی استفاده و بر اساس نتایج، مشخص شد که مدل به‌دست آمده قادر است قدرت پیش بینی اعتباری مشتریان را نسبت به مدل رگرسیون لوژستیک با ضرایب ثابت افزایش دهد (Dong, G. etal. ۲۰۱۰). تحقیقی دیگر در سال ۲۰۱۰ با عنوان «پیش بینی نکول وام های بانکی» توسط جوآ ای بستوس انجام شد. در این مطالعه، عملکرد رگرسیون پاسخ برگشتی ۸ پارامتری و مدل درخت رگرسیون پیش بینی اتلاف اعتبار وام بانکی را ارزیابی کرده است. مدل درخت رگرسیون برای افق زمانی کوتاه مدت ۱۲ و ۲۴ ماهه بهتر عمل می کند و رگرسیون پاسخ برگشتی برای افق زمانی بلندمدت‌تر، نتیجه بهتری می دهد (.2010.Joas A. Bastose)

۲-۱ هدف تحقیق

هدف تحقیق حاضر طراحی مدلی برای پیش‌بینی احتمال معوق شدن تسهیلات قرض‌الحسنه از سوی مشتریان بر اساس ویژگی‌های فردی، اجتماعی و اقتصادی ثبت شده برای آن‌هاست.

---

[1] Abiola and Olausi

[2] Charls and Kens

این امر باعث افزایش توانایی بانک، جهت مدیریت بهتر اعطای تسهیلات قرض الحسنه به متقاضیان این تسهیلات خواهد شد. در این مقاله برای ایجاد مدل‌های پیش بینی عوامل مؤثر بر ریسک اعتباری مشتریان بانک صادرات ایران از الگوریتم NN شبکه ی عصبی مصنوعی، رگرسیون لوژستیک و روش‌های دیگر طبقه‌بندی نظیر انواع درخت‌های تصمیم استفاده می نماییم. به طور کلی مدل‌های برگرفته از تکنیک های داده کاوی از دقت بالایی برخوردار بوده و با افزایش حجم داده‌ها، پیش‌بینی‌های دقیق‌تری را برای نمونه‌های جدید فراهم می‌آورد.

در بخش ۲ پس از معرفی جامعه‌ی آماری، نحوه‌ی جمع آوری داده‌ها و معرفی متغیرهای به کار رفته در این پژوهش، به توضیح و تحلیل توصیفی داده‌ها و سپس تشریح مدل‌های به کاررفته در پژوهش می پردازیم. سپس در بخش ۳ نتایج تحل‌ی‌ل‌های پ‌ی‌اده شده برای داده‌های مورد نظر گزارش می‌شوند.

۲ توضیح و تحلیل توصیفی داده‌ها

۲-۱ توضیح داده‌ها

- جامعه‌ی آماری در این پژوهش، شامل تمامی مشتریان یکی از شعب بانک صادرات ایران در شهر تهران است که در بازه زمانی ۱۳۹۰ تا انتهای ۱۳۹۴، از تسهیلات قرض الحسنه‌ی این بانک استفاده نموده‌اند. مشتریان این شعبه طی ۵ سال مورد بررسی شامل ۳۰۸ مشتری است که از این تعداد، تسهیلات اعطایی مربوط به ۱۲۸ مشتری در گروه معوق شده جای می‌گیرد. در این پژوهش مشتریان را از نظر وضعیت اعتباری و بازپرداخت تسهیلات می‌توانیم به پنج گروه تقسیم نماییم:
- مشتریانی که تسهیلات اعتباری دریافتی شامل اصل و سود تسهیلات را در موعد مقرر تسویه نموده‌اند (تسهیلات تسویه شده)
- مشتریانی که تسهیلات اعتباری از بانک دریافت نموده و با وجود فرا رسیدن زمان سررسید تسهیلات دریافتی تا حداکثر ۲ ماه
- پس از سررسید نسبت به باز پرداخت وام های خود اقدام ننموده اند (مطالبات جاری)
- مشتریانی که با وجود گذشت فاصله زمانی ۲ تا ۶ ماه از زمان سررسید تسهیلات آنها، نسبت به ایفای تعهدات خود عمل ننموده اند (مطالبات سررسید گذشته)
- مشتریانی که ۶ تا ۱۸ ماه پس از سررسید هنوز بدهی خود را تسویه ننموده اند (مطالبات معوق).

- مشتریانی که با گذشت بیش از ۱۸ ماه ازسررسید، اقدامی جهت باز پرداخت بدهی خود نکرده‌اند (مطالبات مشکوک الوصول)

به طور منطقی بانک‌ها درصدد آن هستند که منابع خود را به گروه‌هایی تخصیص دهند که در موعد مقرر نسبت به بازپرداخت تسهیلات دریافت شده‌ی خود اقدام نمایند (مشتریانی که در گروه اول قرار می‌گیرند). البته همواره این امکان وجود دارد که مشتری به دلایلی تعهدات خود را با تأخیر تسویه نماید، لیکن این امر به معنی قصور در بازپرداخت و یا عدم بازپرداخت نیست. به همین دلیل پس از سررسید شدن اقساط وام، بانک‌ها مدتی را به منظور مماشات با مشتری جهت تسویه تعهدات وی در نظر می‌گیرند و پس از گذشت مدت ۲ ماه، مبلغ تسهیلات ارائه شده را به سرفصل سررسید گذشته انتقال می‌دهند و با گذشت مدت معینی پس از این تاریخ اگر تعهدات مربوطه تسویه نگردد به سر فصل معوق و سپس مشکوک‌الوصول انتقال می‌یابند. لذا تعیین معیاری برای مشتریانی که در بازپرداخت تعهدات خود قصور می‌نماید از اهمیت ویژه‌ای در مدیریت ریسک اعتباری بانک برخوردار است.

در این پژوهش مشتریان حقوقی تسهیلات بانک را به دو گروه زیر طبقه‌بندی می‌نماییم:

۱. **مشتریان خوش حساب:** افرادی که تسهیلات دریافتی خود را حداکثر تا ۲ ماه پس از سررسید، تسویه می‌نمایند (شامل تسهیلات تسویه شده و مطالبات جاری).
۲. **مشتریان بدحساب:** افرادی که تسهیلات دریافتی آنها با گذشت ۲ ماه از سررسید به سر فصل تسهیلات سررسید گذشته، معوق یا مشکوک الوصول منتقل گردیده ولی هنوز نسبت به تسویه تعهدات خود اقدام ننموده‌اند. (شامل مطالبات سررسید گذشته، معوق و مشکوک الوصول)

در واقع، وضعیت بازپرداخت تسهیلات دریافتی مشتریان را به عنوان متغیر وابسته‌ی دو سطحی در این پژوهش درنظر می‌گیریم. این متغیر مشتریان را به دو گروه مشتریان خوش‌حساب و بدحساب بر اساس تعاریف بالا تقسیم می‌نماید. فرض می‌کنیم این متغیر، مقدار صفر را برای مشتریانی که قصور در بازپرداخت ندارند (که در ادامه تحت عنوان «تسهیلات جاری» از آن یاد می‌شود) و مقداریک را برای مشتریانی که قصور در باز پرداخت دارند (که در ادامه تحت عنوان «مطالبات» یاد می‌شود)

از آنجا که هدف تحقیق حاضر پیش‌بینی میزان ریسک اعتباری مشتری پیش از اعطای تسهیلات به اوست، علاقه‌مندیم که از ویژگی‌های فردی وی که به سادگی در دسترس بانک قرار می‌گیرد، به عنوان متغیرهای مستقلو پیشگوی ریسک اعتباریاستفاده نماییم. از جمله ویژگی‌های فردی مورد نظر برای مشتریان در این پژوهش می‌توان به جنسیت (مرد و زن)، محل تولد (تهران و شهرستان)، نوع شغل (دولتی و آزاد)، نوع مسکن (مالک و مستاجر)،

وضعیت تاهل (متاهل و مجرد) و سن وی اشاره نمود. دور از ذهن نیست که به جز عوامل شخصی و اجتماعی فرد، نحوه و شرایط اعطال تسهیلات نیز به عنوان عوامل تاثیرگذار بالقوه در بازپرداخت تسهیلات در نظر گرفته شوند. به همین دلیل از اطلاعات مربوط به وام‌های داده شده از جمله، سال اخذ وام، تعداد اقساط، نوع وام، دوره پرداخت اقساط و کمینه مبلغ وثیقه به عنوان دیگر متغیرهای پیشگو بهره خواهیم برد.

۲-۲ تحلیل توصیفی داده ها

در این بخش به منظور کشف روابط حاکم میان اطلاعات موجود، از روش های آمار توصیفی همچون رسم نمودار و ارائه‌ی شاخص‌های خلاصه سازی در کنار روش‌های استنباط آماری از جمله آزمون های فرض استفاده می‌کنیم.

به منظور مقایسه‌ی ویژگی‌های نمونه‌ای دو گروه مشتریانی که تسهیلات آنها جاری و یا تبدیل به مطالبات شده است از آزمون‌های فرض آماری کمک می‌گیریم. جدول ۱، نتایج انجام آزمون‌های معناداری را برای تمامی متغیرهای پیشگو نشان می‌دهد. در این جدول به منظور بررسی رابطه‌ی پیشگوهای کیفی و کمی با متغیر وابسته، به ترتیب فراوانی و درصد فراوانی مربوط به رده‌های متغیر پیشگو و انحراف معیار ± میانگین به تفکیک تسهیلات جاری و مطالبات گزارش است.

همچنین نسبت احتمال ایجاد مطالبات غیرجاری در مقایسه‌ی سطوح متغیرهای پیشگوی کیفی با عنوان ریسک نسبی[1] محاسبه شده است. برای آزمون فرض استقلال متغیرهای پیشگوی کیفی و کمی از متغیر وابسته، به ترتیب p ـ مقدار مربوط به آزمون کای

[1] Risk Ratio

جدول ۱

مقایسات فراوانی و درصد متغیرها برای دو ویژگی تسهیلات جاری و مطالبات معوق

متغیرهای پیشگو	تسهیلات جاری	مطالبات	p-value	p-value ناپارامتری	ریسک‌نسبی اطمینان) (فاصله
جنسیت					
زن	۹۷ (%۵۴)	۸۱ (%۶۳)	۰/۱۲		۱/۲۶(۰/۹۵،۱/۶۶)
مرد	۸۳ (%۴۶)	۴۷ (%۳۷)			
محل تولد					
تهران	۸۶ (%۴۸)	۶۲ (%۴۸)	۱		۱/۰۱(۱/۳۲،۰/۷۷)
شهرستان	۹۴ (%۵۲)	۶۶ (%۵۱)			
شغل					
آزاد	۱۰۱ (%۵۶)	۸۳ (%۶۴)	۰/۱۵۵		۱/۲۴ (۱/۶۴،۰/۹۷)
دولتی	۷۹ (%۴۳)	۴۵ (%۳۵)			
مسکن					
مالک	۱۰۸ (%۶۰)	۶۰ (%۴۷)	۰/۰۰۱<		۰/۷۳ (۰/۹۵،۰/۵۶)
مستاجر	۷۲ (%۴۰)	۶۸ (%۵۳)			
تاهل					
متاهل	۱۵۰ (%۸۳)	۷۳ (%۵۷)	۰/۰۰۱<		۰/۵۰ (۰/۶۴،۰/۳۹)
مجرد	۳۰ (%۱۶)	۵۵ (%۴۳)			
سال اخذ وام					
۱۳۹۰	۱۸ (%۱۰)	۱۲ (%۹)	۰/۴۰		۱/۱۰ (۱/۹۵،۰/۶۱)
۱۳۹۱	۲۸ (%۱۵)	۱۶ (%۱۲)			۱/۱۲ (۱/۹۱،۰/۶۶)
۱۳۹۲	۴۷ (%۲۶)	۲۶ (%۲۰)			۰/۸۰ (۱/۳۰،۰/۴۹)
۱۳۹۳	۴۱ (%۲۳)	۴۱ (%۳۲)			۰/۹۵ (۱/۵۹،۰/۵۷)
۱۳۹۴	۴۶ (%۲۵)	۳۳ (%۲۶)			
نوع وام					
خرید لوازم خانگی	۹۹ (%۵۵)	۶۱ (%۴۷)	۰/۲۱		۰/۹۱ (۱/۳۰،۰/۶۳)
خرید خودرو	۳۵ (%۱۹)	۲۵ (%۱۹)			۰/۷۲ (۰/۹۸،۰/۵۳)
تعمیرمسکن	۳۲ (%۱۷)	۳۵ (%۲۷)			۱/۱۴ (۲/۱۶،۰/۶۰)
ازدواج	۱۴ (%۷)	۷ (%۵)			
مبلغ وثیقه	۱۰۰/۰۵±۱۲۵/۶۶	۱۲۲/۰۸±۱۵۵/۶۷	۰/۰۰۱<	۰/۰۰۱<	
سن	۱۱/۸۰±۴۴/۳۵	۱۱/۶۱±۴۲/۱۴	۰/۰۰۱<	۰/۰۰۱<	
مبلغ وام	۲۱/۹۲±۱۳/۳۸	۲۳/۰۶±۱۵/۸۲	۰/۰۰۱<	۰/۰۰۱<	
درآمدبرحسب میلیون تومان	۱/۷۰±۱/۴۷	۲/۲۷±۱/۹۹	۰/۰۰۱<	۰/۰۰۱<	

منبع: نتایج نویسندگی

دوی پی‌یرسون[1] و آزمون t دو نمونه‌ای مستقل[2] گزارش شده است. لازم به ذکر است که پیش از انجام آزمون t فرض برابری واریانس‌ها با استفاده از آزمون بارتلت[3] بررسی شده است. علاوه بر این برای بررسی فرض برابری میانگین‌ها در متغیرهای پیشگوی پیوسته p ـ مقدار مربوط به آزمون ناپارامتری ویلکاکسون[4] نیز که نیازی به فرض نرمال بودن ندارد برای مقایسه نتایج گزارش شده است.

بر اساس نتایج جدول ۱ و با توجه به p-value آزمون‌های پارامتری و ناپارامتری، از نظر متغیرهای مبلغ وثیقه، سن، مبلغ وام و درآمد، دو گروه تسهیلات جاری و مطالبات یکسان نبوده‌اند ($p < ۰/۰۵$).

برای متغیر جنسیت در دو گروه تسهیلات جاری و مطالبات p-value آزمون کای دو ($p = ۰/۱۱۲$) نشان می‌دهدکه زنان و مردان در دو این گروه یکسان و همچنین فاصله اطمینان ریسک نسبی عدد ۱را شامل میشود پس جنسیت در سطح معناداری ۵ درصد، درنحوه پرداخت اقساط تاثیری ندارد. به همین ترتیب ارتباط متغیر وابسته با محل تولد، شغل، سال اخذ وام و نوع وام از معناداری قابل قبولی برخوردار نیست.

نتیجه آزمون کای دو برای وضعیت مسکن و تاهل و همچنین فاصله اطمینان ریسک نسبی آنها، نشان میدهد که دو گروه از نظر این متغیرها یکسان نبوده‌اند ($p < ۰/۰۵$). در واقع بر اساس نتایج، مالکان مسکن ریسک اعتباری کمتری نسبت به متقاضیان مستاجر و افراد متاهل نیز ریسک کمتری نسبت به مجردها دارند.

با استفاده از هر دو رویکرد پارامتری و ناپارامتری تفاوت معناداری در میانگین سن، درآمد، مبلغ وثیقه و مبلغ وام در دو گروه مشتریان خوش حساب و بد حساب وجود دارد.

به منظور بررسی تغییرات ریک اعتباری مشتریان در گذر زمان، شکل ۱ نمودار پروفایل فراوانی مشتریان دارای تسهیلات جاری یا مطالبات معوق را در بازه‌ی ۵ ساله تحقیق نشان می‌دهد. همچنین در شکل ۲ نمودار میله ای نسبت پرداخت یا عدم پرداخت قسط در زمان مقرر در در طول زمان رسم شده است. بااستفاده از این نموداها می‌توانیم مشاهده کنیم که فراوانی تسهیلات جاری در هر سال همواره بیشتر از فراوانی مطالبات غیر جاری است. البته درسال ۱۳۹۳ میزان تسهیلات جاری و مطالبات تقریبا یکسان بوده و نسبت مطالبات معوق

---

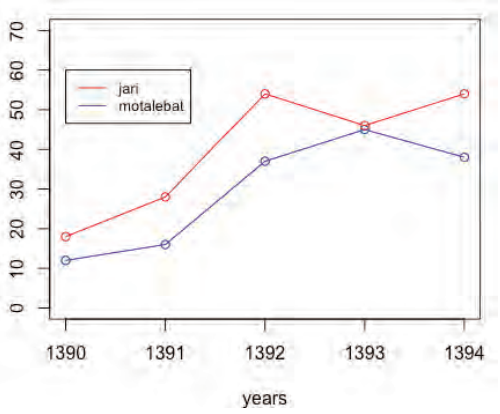
[1] Pearson Chi-squared

[2] Two Independent sample's t-test

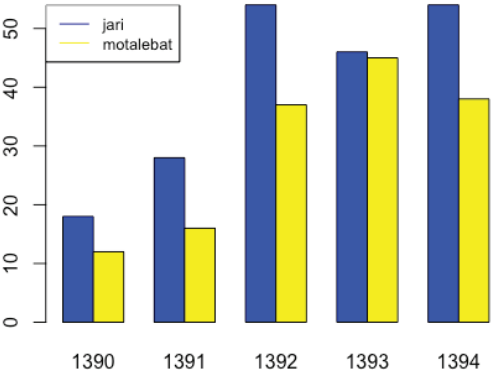
[3] Bartlett's test

[4] Wilcoxon test

به بیشترین میزان خود در این دوره‌ی زمانی رسیده است. همچنین می‌توان مشاهده کرد که تعداد تسهیلات اعطایی در شعبه‌ی مورد بررسی تا سال ۱۳۹۲ رشدی صعودی داشته و پس از آن کاهش یافته است.



شکل ۱. نمودار پروفایل پرداخت یا عدم پرداخت قسط در زمان مقرر در طول زمان



شکل ۲. نمودار میله ای نسبت پرداخت یا عدم پرداخت قسط در زمان مقرر در در طول زمان

۳ مدلسازی داده‌ها و نتایج تحقیق

همان‌طور که اشاره شد، برای بررسی عوامل مؤثر بر ریسک اعتباری مشتریان بانک، ۱۱ ویژگی جنسیت، محل تولد، شغل، مسکن، وضعیت تأهل، سال اخذ وام، نوع وام، مبلغ

وثیقه، سن، مبلغ وام و درآمد شخص وام‌گیرنده به‌عنوان متغیر های پیشگو در نظر گرفته شده‌اند.

در ابتدا لازم به ذکر است که مسأله ی مهم در ارزیابی دقت مدل‌های پیش‌بینی، نحوه‌ی تفکیک داده‌ها، به دو بخش جهت آموزش و سپس تست کفایت مدل است.

به منظور ارزیابی کیفیت مدل، به جای استفاده از یک زیر نمونه گیری ساده برای تشکیل دو مجموعه داده‌ی آموزش و آزمایش از روش اعتبارسنجی K – Fold cross validation با مقدار ۱۰ برای k استفاده می‌نماییم. این روش، مجموعه داده‌ی مورد نظر را به ۱۰ قسمت با حجم تقریباً مساوی تقسیم نموده، به‌گونه‌ای که نسبت داده‌های دو کلاس (متغیر وابسته) در هر قسمت، برابر نسبت آن‌ها در کل داده‌ها باشد. بر اساس این روش در هر اجرا، ۹ قسمت از ۱۰ قسمت داده‌ها جهت آموزش شبکه و یک قسمت باقی‌مانده برای آزمودن شبکه استفاده می‌شود. این عمل برای هر ۱۰ قسمت انجام می‌شود.

در این پژوهش از الگوریتم‌های داده‌کاوی متفاوت NN (Neural Network)[1]، درخت تصمیم[2]، Bagging و Boosting و همچنین روش آماری رگرسیون لوژستیک برای تولید مدل پیش‌بینی مناسب استفاده شده است. بعلاوه با استفاده از آزمون‌های آماری مختلف و معیارهای ارزیابی شامل دقت، حساسیت (Sensitive) ویژگی (Specificity) و سطح زیر منحنی ROC (Receive Operating Characteristic) کارآیی مدل‌های مورد ارزیابی قرار گرفته است.

۳-۱ روش رگرسیون لوژستیک

رگرسیون لوژستیک، یک روش آماری پراستفاده و نوعی تحلیل رگرسیونی است که برای پیش‌بینی برآمد یک متغیر پاسخ با تعداد محدودی از مقادیر ممکن، بر اساس متغیرهای پیشگو مستقل استفاده می‌شود.

هرچند در رگرسیون لوژستیک همانند رگرسیون خطی، ضرایب پارامتری متغیرهای مستقل برآورد می‌شوند، لیکن نحوه‌ی برازش در این دو مدل کاملاً متفاوت است. در رگرسیون های خطی از روش حداقل مربعات استفاده می‌شود. در این روش مجموع مجذورات اختلاف بین مقادیر واقعی و مقادیر پیش‌بینی‌شده متغیر وابسته حداقل می‌گردد. در مقابل به منظور برازش مدل لوژستیک که یک مدل خطی تعمیم یافته است، از روش حداکثر

---

[1] شبکه های عصبی

[2] Decision tree

درستنمایی استفاده می‌گردد که به گونه‌ای ضرایبی مدل را تعیین می‌نماید که احتمال وقوع داده‌های حاضر حداکثر شود.

نتایج برازش رگرسیون لوژستیک بر داده‌های پژوهش در جدول ۲ گزارش شده است. مانند رگرسیون معمولی صفر بودن ضریب هر متغیر مستقل را می‌توان با استفاده از آماره "Wald" به جای آماره t در رگرسیون خطی آزمون نمود. آماره والد برای هر یک از ضرایب همچون رگرسیون معمولی نشان‌دهنده معنی دارای ضریب مربوطه است. این آماره دارای توزیع x^2 با درجه‌ی آزادی ۱ می‌باشد. همان‌طور که در جدول ۲ ملاحظه می‌شود، سطح معنی‌داری برای برخی ضرایب کمتر از ۰/۰۵ است. که این بدان معناست که فرضیه صفر برای آن ضرایب در سطح اطمینان ۹۵٪ رد شده و لذا پیشگوهای متناظر انها تاثیر معناداری بر احتمال وقوع مطالبات غیر جاری دارند.

همچنین میزان دقت، حساسیت و قابلیت تشخیص مدل رگرسیون لوژستیک در جدول ۳ ارائه شده است. میزان دقت به دست آمده از مدل رگرسیون لجستیک ۰/۶۲۳ است یعنی احتمال تشخیص درست وضعیت اعتباری مشتری بر پایه ویژگی های ورودی گفته شده حدود ۶۲ درصد است.

جدول ۲

نتایج برازش مدل رگرسیون لوژستیک به داده‌های ریسک اعتباری دریکی از شعب بانک صادرات

نام متغیر Variable	براورد پارامتر ها Estimate	انحراف معیار Std. Error	آماره z z-Statistic	آماره Wald	درجه آزادی df	معنی داری Prob.
جنسیت	۰/۳۵۰-	۰/۲۵۹	۱/۳۵۰-	۱/۸۵۲	۱	۰/۱۷۲
محل تولد	۰/۰۴۲	۰/۲۸۳	۰/۱۴۹	۰/۰۲۳	۱	۰/۴۳۰
شغل	۰/۰۳۲-	۰/۲۷۱	۱/۱۹۶-	۰/۷۶۲	۱	۰/۳۸۳۴
مسکن	۰/۶۶۰	۰/۲۵۸	۲/۵۵۵	۴/۹۰۱	۱	۰/۰۲۷
تاهل	۱/۲۱	۰/۳۰۰	۴/۰۴۰	۱/۰۵۵	۱	۰/۰۰۱
سال‌اخذ وام	۰/۰۴۸	۰/۱۰۶	۰/۴۵۴	۰/۰۱۱	۱	۰/۸۰۵
نوع وام	۴/۳۶۶	۱۷/۵۷	۰/۲۴۸		۱	۰/۸۰۳
مبلغ وثیقه	۰/۰۰۹	۰/۰۱۹	۰/۵۰۴	۱۰/۶۳	۱	۰/۰۱۸
سن	۰/۰۲-	۰/۰۱۲	۱/۰۰-	۱/۴۷۵	۱	۰/۱۲۸
مبلغ وام	۶/۸۳۵-	۱۹/۹۸	۰/۳۲۸-	۰/۵۰۳	۱	۰/۹۵۴
درآمد	۰/۰۰۰	۰/۰۰	۱/۶۷۸	۲/۸۹۱	۱	۰/۰۹۰

منبع: نتایج نویسندگان

جدول ۳

معیارهای دقت، حساسیت، قابلیت تشخیص برای رگرسیون لوژستیک

Model	Performance measures	
	Accuracy	۰/۶۲۳
	%95 CI	(۰/۹۴۸،۰/۷۴۳)
glm	Sensitivity	۰/۷۵۰
	Specificity	۰/۴۸۲
	Positive predict value	۰/۶۱۵
	Negative predict value	۰/۶۳۶

منبع: نتایج نویسندگان

در رگرسیون لوژستیک برای مقایسه‌ی کارآیی چند مدل مختلف با یکدیگر و یا نیکویی برازش یک مدل خاص آماره‌های متعددی مورداستفاده قرار می‌گیرد که ازجمله آن‌ها می‌توان به موارد زیر اشاره نمود.

۳-۱-۱ آماره کای ۲

برای بررسی میزان نیکویی برازش مدل برآورد شده روشی مشابه رگرسیون خطی چندگانه به کار می‌رود. برای آزمون یک صحت یک مدل، مقدار تابع لگاریتم درست نمایی تحت مدل مور نظر به‌صورت 2LnL- در نظرگرفته می‌شود. آماره 2LnL- دارای توزیع x^2 با n-q درجه‌ی آزادی است که در آن n تعداد مشاهدات و q تعداد پارامترهای برآورد شده در مدل است.

بر این اساس تأثیر ورود متغیرهای مستقل در بهبود مدل سنجیده می‌شود. در این صورت اگر آماره مزبور در سطح ۵% در ناحیه رد قرار نگیرد، فرضیه صفر مبنی بر معناداری متغیرهای مستقل در مدل مورد نظر پذیرفته خواهد شد. بدین ترتیب در هر مرحله، معنی‌دار بودن ضرایب به‌وسیله آزمون X^2 برای تعیین متغیرهای مهم مشخص می‌گردد، مشروط بر این‌که متغیرها بتوانند از مدل حذف شوند.

آماره LR که شبیه آماره F در مدل رگرسیون خطی است، دارای توزیع کای دو با ۱۱k = (k: تعداد متغیرهای مستقل مدل) درجه آزادی است. مقدار این آماره برابر ۴/۳۰ به‌دست‌آمده است.

احتمال آماره LR که مقدار آن برابر ۰/۸۲۰ است و نشان‌دهنده‌ی پذیرفته شدن فرض صفر مبنی بر بهتر بودن مدل بهینه رگرسیون با ضرایب تاهل مسکن و مبلغ وثیقه در سطح اطمینان ۹۵% می‌باشد. درنتیجه مدل رگرسیون مورد نظر معنی‌دار است.

۳-۱-۲ آماره شبه R^2 یا R^2 مک فادن[1]

آماره شبه R^2 یا R^2 مک فادن برای رگرسیون لوژستیک، مشابه R^2 در رگرسیون معمولی است. این آماره از رابطه‌ی زیر محاسبه می‌شود:

$$R^2 Logit = \frac{\text{Mode } (-2\text{Ln}) - \text{mull } (-2\text{Ln})}{\text{mull } (-2\text{Ln})}$$

مقدار این آماره بین صفر و یک تغییر می‌کند و خوبی برازش مدل را اندازه‌گیری می‌نماید. هر چه این شاخص نزدیک به یک باشد، میزان تطابق مدل با واقعیت بیشتر بوده و به عبارتی نیکویی برازش بیشتر است و بالعکس هرچه مقدار شاخص به صفر نزدیک‌تر باشد نیکویی برازش کمتر خواهد بود.

آماره R^2 مک فادن در مدل تخمین زده شده برابر ۰/۲۳ به‌دست‌آمده است. که برای رگرسیون لوجیت عدد قابل قبولی می‌باشد.

۳-۱-۳ آماره هاسمر – لمشو[2]

در این روش با استفاده از گروه‌بندی مشاهدات، مقادیر پیش‌بینی‌شده توسط مدل با مقادیر واقعی مشاهدات مقایسه می‌شوند. اگر اختلاف‌ها بزرگ باشند، مدل رد شده و نشان‌دهنده‌ی این است که مدل به‌خوبی برازش نشده است و در غیر این صورت مدل پذیرفته می‌شود.

معمولاً و در اکثر نرم‌افزارها مشاهدات به ۱۰ گروه مساوی تقسیم می‌شوند. استفاده مناسب از این روش مستلزم تعداد کافی مشاهدات است، بطوریکه در هر گروه حداقل باید ۵ مشاهده وجود داشته باشد. آماره هاسمر–لمشو دارای توزیع x^2 با j-2 درجه آزادی می‌باشد (j تعداد گروه‌ها است). خاطرنشان می‌نماییم که توزیع x^2 نسبت به تعداد نمونه حساس است. لذا زمانی که تعداد نمونه خیلی بزرگ باشد تفاوت‌های کمی بین مقادیر مشاهده‌شده و پیش‌بینی‌شده از طریق این روش نشان داده می‌شود و آزمون معتبر می‌باشد.

مقدار این آماره در مدل برآورد شده دارای توزیع X^2 با ۱۱ درجه آزادی و برابر ۸/۸۷ می‌باشد و p ـ مقدار آن بزرگ‌تر از ۰/۰۵ و برابر ۰/۳۵ به‌دست‌آمده است. بنابراین فرض صفر که برابر نکویی برازش است، پذیرفته می‌شود (رد نمی‌گردد). پس متغیرهای مستقل مدل قدرت توضیح دهندگی میزان ریسک اعتباری را دارا می‌باشند.

---

[1] Mc Fadden R-Square.

[2] Hosmer and Lemeshow, 1989.

آماره های مورد استفاده در تعیین میزان نیکویی برازش رگرسیون لجستیک و مقدار بدست آمده برای هریک را میتوان به شرح جدول ۴ خلاصه نمود:

جدول ۴

آزمون معنی داری و نیکویی برازش مدل رگرسیون لوژستیک

LR statistic	۴/۳۰	Probability (LR stat)	۰/۰۲۸
McFadden R-squared	۰/۲۳		
H-L Statistic	۸/۸۷	Prob.Chi-Sq	۰/۶۳

منبع: نتایج نویسندگان

۲-۳ درخت تصمیم

درخت تصمیم یکی از مشهورترین و قدیمی‌ترین روش‌های ساخت مدل طبقه بندی است. در الگوریتم‌های طبقه‌بندی مبتنی بر درخت تصمیم دانش خروجی به‌صورت یک درخت از حالات مختلف مقادیر ویژگی‌ها ارائه می‌شود. نمایش دانش به شکل درخت سبب شده است که دسته‌بندی‌های مبتنی بر درخت تصمیم کاملاً قابل تفسیر باشند.

بیشتر الگوریتم‌های یادگیری که برای درختان تصمیم توسعه یافته‌اند از یک الگوریتم پایه مشتق شده‌اند که یک جستجوی حریصانه‌ی از بالا به پایین را در فضای درختان تصمیم ممکن، بکار می‌گیرد. این روش توسط الگوریتم ID3I نشان داده می‌شود.

الگوریتم پایه‌ی ID3: این الگوریتم درختان تصمیم را از بالا (ریشه) به پایین (برگ) می‌سازد و با طرح این سؤال که چه صفتی باید در ریشه درخت آزمایش شود آغاز می‌کند. برای پاسخ به این سؤال، با استفاده از یکی از انواع آزمایش‌های آماری برای تعیین مناسب‌ترین صفت برای دسته‌بندی نمونه‌های آموزشی، تصمیم بر اساس هر ویژگی نمونه را ارزیابی می‌کند. سپس بهترین ویژگی را انتخاب کرده و درگره ریشه درخت استفاده می‌کند. برای هر مقدار ممکن ویژگی قرار داده‌شده در ریشه، یک گره متناظر ایجادشده و نمونه‌های آموزشی بر اساس مقادیر این ویژگی، بین این گره‌ها افراز می‌شوند. تمام فرایند ذکرشده، با استفاده از نمونه‌های آموزشی، برای انتخاب بهترین ویژگی برای ایجاد شاخه در هرگره‌ی درخت تکرار می‌شود.

در جدول ۵ نتایج حاصل از مدل درخت تصمیم بطورخلاصه آمده است که طبق آن دقت به دست آمده از این روش ۰/۶۴ است.

جدول ۵

مقادیر پیش‌بینی شده برای مدل درخت تصمیم

Model	Performance measures	
Decision tree	Accuracy	۰/۶۴۰
	Positive predict value	۰/۱۶۴
	Negative predict value	۰/۹۵۰

Bagging ۱-۲-۳

در این روش از رویکرد بوت استرپ برای بهبود یادگیری‌های آماری مثل درخت های تصمیم استفاده می‌شود. درخت های تصمیم که در بخش قبل توضیح داده شده‌اند از واریانس زیاد رنج می برند به این معنی که اگر داده های آموزشی را به صورت تصادفی به دو قسمت تقسیم کنیم و یک درخت تصمیم به هر کدام از آن ها برازش بدهیم آن گاه نتایجی که می گیریم می تواند کاملاً متفاوت باشد. در مقابل یک روش با واریانس کم می تواند نتایج مشابهی را تولید کند، اگر روی مجموعه داده های مختلف به n به طور مکرر تکرار بشود.

بگینگ یا مجموعه بوت استرپ، یک فرآیند کلی و عمومی است که برای کاهش واریانس روش‌های یادگیری آماری استفاده می شود. برای پیاده کردن بگینگ روی درخت های رگرسیونی، با استفاده از چندین مجموعه آموزشی بوت استرپ شده، تعدادی درخت رگرسیونی تشکیل شده و پیش بینی نهایی به عنوان میانگین پیش بینی ها در نظر گرفته می‌شود. بدین ترتیب هم اریبی و هم واریانس روش کاهش می‌یابد.

در جدول ۷ نتایج حاصل ازاین مدل به طورخلاصه آمده است که طبق آن دقت به دست آمده از این روش ۰/۶۸۸ میباشد.

جدول ۷

مقادیر پیش‌بینی شده برای مدل Bagging

Model	Performance measures	
Bagging	Accuracy	۰/۶۸۸
	% CI۹۵	(۰/۵۹۱،۰/۸۲۸)
	Sensitivity	۰/۷۸۹
	Specificity	۰/۶۰۸
	Positive predict value	۰/۷۶۹
	Negative predict value	۰/۶۳۶

Boosting ۲-۲-۳

روش دیگری که به بهبود پیش بینی های یک درخت تصمیم کمک می کند، روش بوستینگ است. مشابه روش بگینگ، روش بوستینگ هم می تواند روی بسیاری از روش های یادگیری آماریاز جمله درخت‌های رگرسیوین و طبقه بندی پیاده سازی شود. بگینگ تعداد زیاد و متعددی نمونه از روی مجموعه داده ی آموزشی با استفاده از روش بوت استرپ کپی برداری می نماید و سپس یک درخت تصمیم روی هر کدام از این نمونه های کپی شده برازش دهد و در انتها از همه ی این درخت، یک مدل پیش بینی کننده منفرد ایجاد کند. قابل توجه است که هر کدام از این درخت های تصمیم که با استفاده از داده های بوت استرپ ساخته می شود از سایر درخت ها مستقل هستند. روش بوستینگ به یک شکل مشابهی کار می کند به جز اینکه درخت ها به صورت دنباله ای گسترش پیدا می کنند به این معنی که هر درختی با استفاده از اطلاعات درخت های قبلی گسترش پیدا می کند.

بوستینگ با مساله نمونه گیری بوت استرپ درگیر نیست و به جای آن هر درخت روی یک زیرمجموعه‌ای از مجموعه داده ی اصلی برازش می شود و با ترکیب تعداد زیادی از درخت ها به نتیجه نهایی می‌رسیم.

در جدول ۶ نتایج حاصل از این روش بطورخلاصه آمده است که طبق آن دقت به دست آمده از این روش ۰/۷۰۵ میباشد. که تا این قسمت از روش های دیگر از دقت بیشتری برخوردار است.

جدول ۶

مقادیر پیش‌بینی شده برای مدل Boosting

Model	Performance measures	
Boosting	Accuracy	۰/۷۰۵
	95% CI	(۰/۵۷،۰/۸۱۴)
	Sensitivity	۰/۶۹۲
	Specificity	۰/۷۲۷
	Positive predict value	۰/۸۱۸
	Negative predict value	۰/۵۷۱

۳-۳ مدل شبکه عصبی

این مدل شامل شبکه ای از اجزای به هم مرتبط است که آن را شبکه ی عصبی می نامند. این اجزاء از مطالعات سیستم های عصبی زیستی الهام گرفته شده است. به عبارت دیگر شبکه‌های عصبی کوششی است تا با استفاده از اجزایی که شبیه سلول های عصبی زیستی رفتار می‌کنند، ماشین هایی ایجاد شوند که مانند مغز انسان کار می‌کنند. نورون ها در چنین شبکه ای در قالب لایه های چندگ انه (Multi-Layer Perception) سازماندهی می‌شوند.

هر NN حداقل از سه لایه شامل لایه‌های ورودی، میانی و خروجی تشکیل شده است. نورون‌های لایه‌های میانی به صورت کامل به نورون‌های دو لایه دیگر متصل می‌شوند. هرکدام از اتصالات بین نورون ها دارای وزنی است که مقدار این وزن قدرت تاثیر نورون را روی نورون‌های بعدی نشان می‌دهد. مقادیر متغیرهای مستقل به نورون های لایه‌ی ورودی متصل شده و مقدار متغییر وابسته در لایه‌ی خروجی مشخص می شود. جریان اطلاعات از لایه‌ی ورودی به سمت لایه‌های میانی حرکت و در نهایت منجر به تعیین مقدار خروجی در لایه‌ی خروجی می‌شود و مدل را آموزش می دهد. جدول ۸ بطورخلاصه نتیجه حاصل از مدل شبکه عصبی را شرح می‌دهد که طبق آن مشاهده می‌شود که دقت آن نسبت به دیگر مدل ها بیشتر بوده و دارای بالاترین دقت است.

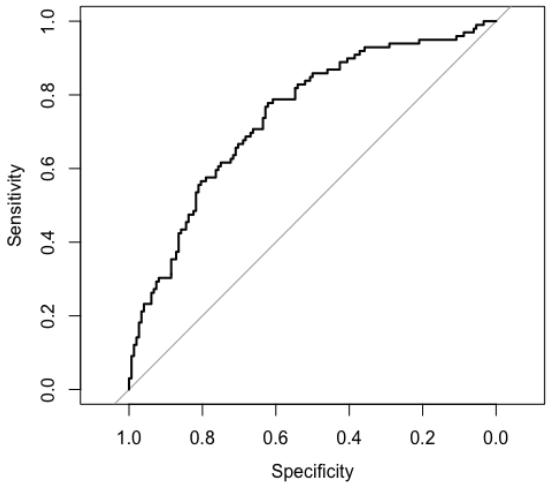
جدول ۸

مقادیر پیش‌بینی شده برای مدل سیستم شبکه عصبی

Model	Performance measures	
Neural Networking	Accuracy	۰/۷۲۱
	95% CI	(۰/۵۹۱،۰/۸۲۸)
	Sensitivity	۰/۹۱۴
	Specificity	۰/۴۶۱
	Positive predict value	۰/۶۹۵
	Negative predict value	۰/۸۰۰

۳ـ۳ـ۱ بررسی قدرت تفکیک کنندگی مدل

پس از محاسبه درجه حساسیت و درجه تشخیص مدل به منظور بررسی قدرت تفکیک کنندگی دو گروه (در اینجا مشتریان خوش حساب و مشتریان بد حساب) از یکدیگر از منحنی به نام ROC استفاده میشود (شکل ۳). این منحنی در صفحه مختصاتی که محور افقی آن یک منهای درجه تشخیص و محور عمودی آن درجه حساسیت مدل می‌باشد، از نقطه (۰ و ۰) در گوشه پایین سمت چپ به سمت نقطه (۱ و ۱) در گوشه بالا سمت راست، کشیده می‌شود. هر چه این منحنی به سمت گوشه سمت چپ بالا (۱ و ۰) نزدیک‌تر باشد، نشان‌دهنده‌ی قدرت بیشتر مدل و تفکیک‌کنندگی بین دو گروه خواهد بود. در نقطه (۱ و ۰) درجه حساسیت و تشخیص مدل هر دو در بیشترین مقدار خود و برابر یک هستند. شاخص خوبی مدل، سطح زیر منحنی است. مدلی که بیشترین قدرت تفکیک کنندگی بین دو گروه را داراست، دارای سطح زیر منحنی ۱ و مدلی که قدرت تفکیک کنندگی آن صفر است، دارای سطح زیر منحنی ۰/۵ است. همانطور که از شکل ۳ مشخص است سطح زیر منحنی ROC در مدل برازش شده برابر با ۰/۷۷۴ می‌باشد و لذا نشان دهنده‌ی این است که مدل برازش شده از قدرت تفکیک کنندگی بالایی بین دو گروه برخوردار است.



شکل ۳. منحنی ROC. سطح زیر منحنی ROC: ۰/۷۷۴

۴ نتیجه گیری

این مطالعه با هدف شناسایی عوامل مؤثر براحتمال عدم بازپرداخت بدهی‌های تسهیلاتی مشتریان حقوقی بانک‌ها انجام گرفته است. در این راستا با استفاده از روش‌هایی که در بخش‌های قبل ذکر شد به تعیین مدل بهینه جهت سنجش ریسک اعتباری مشتریان حقوقی بانک صادرات ایران پرداختیم. به این منظور بر اساس نظریه‌های اقتصادی و مالی ۱۱ ویژگی به عنوان متغیرهای اثرگذار برریسک اعتباری انتخاب شدند که از بین آنها ۳ متغیر وضعیت تاهل، مبلغ وثیقه و وضعیت مسکن جهت تبیین مدل نهایی انتخاب گردیدند. بهترین عملکرد از روش شبکه های عصبی با بیشترین دقت در مقایسه با روش های دیگر انتخاب شد. همچنین معنی داری ضرایب این متغیرها در رگرسیون لوژستیک با استفاده از آماره wald و معنی داری کل رگرسیون و خوبی برازش، با استفاده از آماره LR و هاسمر-لمشو در سطح اطمینان ٪۹۵ مورد تأیید قرار گرفتند.

معنادار بودن و علامت ضرایب متغیرهای مستقل مدل، حاکی از تأیید نظریه های اقتصادی و مالی در زمینه عوامل تعیین کننده ریسک اعتباری می باشند. از میان متغیرهای منتخب مدل، به ترتیب میزان اهمیت، وضعیت تاهل، مبلغ وثیقه و سپس وضعیت مسکن اثر مستقیم بر ریسک اعتباری دارند.

فهرست منابع

تهرانی، رضا و فلاح شمس، میر فیض. (۱۳۸۴). طراحی و تبیین مدل ریسک اعتباری در نظام بانکی کشور؛ مجله علوم اجتماعی و انسانی دانشگاه شیراز، دوره بیست و دوم، شماره دوم، تابستان.

پناهیان، حسین، ابیاک، حسن (۱۳۹۲). تبیین اثرات ریسک بر کارایی بانک ها با استفاده از محاسبه ی کارایی به روش DEA (تحلیل پوششی داده ها.) تحقیقات حسابداری و حسابرسی، سال چهارم، شماره ۸۵-۶۸: ۱۷.

پور زمانی، زهرا و توانگر، افسانه وکیارسی، آوا (۱۳۸۸). بررسی کارآیی الگوی لجیت و تحلیل تمایزی چند متغیره در پیش بینی وضعیت مالی شرکت های بورس اوراق بهادار تهران؛ پژوهش نامه حسابداری مالی و حسابرسی.:۹۵- ۱۲۲

برادری، جعفر. (۱۳۸۶). بر رسی وضعیت و عوامل مؤثر بر پیدایش مطالبات معوق و ارائه راهکارهای مطلوب پیشگیری آن در بانک صادرات بر اساس مدل Moral Hazard. پایان نامه کارشناسی ارشد، مؤسسه عالی بانکداری.

نورا، عباسعلی، و حسین زاده، فرانک، سیف پناه، محمد، صالحی، آزاد. (۱۳۹۰). بررسی اثر ریسک بر کارایی با استفاده از تحلیل پوششی داده های شبکه ای در حضور خروجی نامطلوب. سومین همایش ملی تحلیل پوششی داده ها، فیروزکوه: دانشگاه آزاد اسلامی.

مدرس، احمد و ذکاوت، سید مرتضی. (۱۳۸۲). مدل های ریسک اعتباری مشتریان بانک (مطالعه موردی: بانک توسعه صادرات ایران)؛ فصلنامه حسابرس، شماره ۱۹، بهار. ۵۴:

حیدری فر، سوده. (۱۳۹۰). بررسی عوامل مؤثر بر عدم بازگشت به موقع تسهیلات اعطایی بانک ملی ایران؛ پایان نامه کارشناسی ارشد، دانشکده مدیریت دانشگاه تهران، بهار

شیرین بخش، شمس اله و یوسفی، ندا و جهانگیر، قربان زاد (۱۳۹۰). بررسی عوامل موثر بر احتمال عدم باز پرداخت تسهیلات اعتباری. فصلنامه علمی پژوهشی دانش مالی تحلیل اوراق بهادار، شماره دوازدهم. زمستان

Abiola, I., & Olausi Awoyemi.S. (2014). the Impact of Credit Risl Management on the Comercial BanksPerformanc in Nigeria. International Journal of Management and Sustainability, 3(5): 295-306.

Altman, E. I. (1968) Financial Ratio, Discriminant Analysis and the Prediction of Corporate Bankruptcy; Journal of Finance: 589-609.

Glantz, M. (2003) Managing Bank Risk: An Introduction to Broad-Base Credit Engineering; Academic Press; 1 edition 2-13.

McFadden, D. (1974). Conditional Logit analysis of qualitative choice behaviour, in Zarembka, P. (ed. by), Frontiers in econometrics, Academic Press, New York, pp. 105-142.