فصلنامهٔ پژوهش‌های پولی_ بانکی
سال یازدهم، شمارهٔ ۳۸، زمستان ۱۳۹۷
صفحات ۶۲۵_۶۵۴

پیش‌بینی ریسک تسهیلات پرداختی با استفاده از ابزارهای داده‌کاوی

مسلم نیلچی [*]
خشایار مقدم [†]
علیرضا ناصرصدرآبادی [‡]
علی فرهادیان [§]

تاریخ دریافت: ۱۳۹۷/۰۵/۱۰ تاریخ پذیرش: ۱۳۹۷/۰۹/۰۶

چکیده

یکی از شایع‌ترین علل که در ریسک اعتباری باید مورد توجه قرار گیرد عدم ایفای تعهدات مشتریان است. با پیش‌بینی رفتار اعتباری متقاضیان تسهیلات، می‌توان نرخ رشد مطالبات غیرجاری را کاهش داد. به همین منظور، این پژوهش در خصوص صاحبان کسب‌وکار متقاضی تسهیلات در یکی از بانک‌های دولتی ایران انجام گرفته است. در پژوهش حاضر، پس از گزینش شاخص‌های اساسی شکل‌دهندهٔ رفتار مشتریان، به کمک روش نمونه‌گیری طبقه‌بندی‌شده، ۵۲۱ نمونهٔ تصادفی از بین پرونده‌های تسهیلاتی مشتریان صاحبان کسب‌وکار متقاضی تسهیلات انتخاب شد؛ سپس فرایند آماده‌سازی داده‌ها با تلخیص و یکپارچه‌سازی و درون‌یابی برخی داده‌های مفقود صورت پذیرفت؛ در گام بعدی، نهایتاً ۸۵ شاخص جهت الگوسازی انتخاب و به‌منظور سنجش درجهٔ اهمیت عوامل مؤثر در رفتار اعتباری متقاضیان، از الگوریتم‌های درخت تصمیم، شبکهٔ عصبی، و ماشین بردار پشتیبان بهره گرفته شد که الگوریتم درخت تصمیم با میانگین قدر مطلق خطای ۱۴ درصد در بهترین حالت نرخ نکول را پیش‌بینی کرد. بر اساس داده‌های موجود و مطابق با نتایج به‌دست‌آمده، الگوریتم درخت تصمیم چاید و استفاده از شاخص‌های مدت قرارداد، مبلغ امهال، تعداد اقساط، سود عملیاتی به دارایی، نوع عقد، میانگین قرض‌الحسنه ۳ ماه قبل و مبلغ وام می‌تواند پیش از اعطای تسهیلات، در تصمیم‌گیری و پیش‌بینی رفتار اعتباری مشتری نقش حائز اهمیتی داشته باشد.

واژه‌های کلیدی: ریسک اعتباری، مطالبات غیرجاری، درخت تصمیم، شبکهٔ عصبی، ماشین بردار پشتیبان.
طبقه‌بندی JEL: G3 ،G2 ،E5

---

[*] دانشجوی دکتری مالی_مهندسی مالی دانشگاه یزد (نویسندهٔ مسئول)، Moslem.Nilchi@gmail.com
[†] معاون پژوهشی ادارهٔ آموزش و پژوهش، بانک توسعهٔ تعاون، khash.moghadam@gmail.com
[‡] استادیار بخش مدیریت صنعتی، دانشگاه یزد، alireza_naser@yazd.ac.ir
[§] استادیار گروه مدیریت و کارآفرینی، دانشگاه کاشان، farhadian@kashanu.ac.ir

۱ مقدمه

امروزه، بانک‌ها برای شناخت مشتریان، ارضای نیازمندی‌ها، و ارائهٔ خدمات مالی مناسب نیازمند شناسایی دقیق ویژگی‌های اعتباری آن‌ها هستند. یکی از خدمات مالی در بانک ارائهٔ تسهیلات مالی ازجمله وام به مشتریان است که خود باعث شده است فعالیت بانک‌ها با ریسک اعتباری مواجه شود. ریسک اعتباری، بنا بر تعریف کمیتهٔ بال، عبارت است از امکان بالقوهٔ اینکه قرض‌گیرنده از بانک و یا از طرف حساب وی در اجرای تعهدات خود در مقابل بانک در زمان مشخصی ناتوان شود (کمیتهٔ بال،۲۰۰۰)؛ بنابراین، می‌توان گفت ریسک اعتباری یکی از اصلی‌ترین ریسک‌هایی است که بانک با آن روبه‌روست. بحران‌های مشاهده‌شده در نظام بانکی کشورها عمدتاً ناشی از عدم کارایی در مدیریت ریسک اعتباری بوده است. عدم مدیریت صحیح این ریسک و کنترلِ آن بانک را دچار بحران و ورشکستگی می‌کند و ازآنجایی‌که بانک نهادی تأثیرگذار در نظام اقتصادی هر کشوری است (چی و لی[1]،۲۰۱۷)، این بحران‌ها کل نظام اقتصادی و اجتماعی را دچار مشکل می‌کند.

افزایش تقاضای اعتبار، افزایش رقابت، و به‌وجودآمدن کانال‌های جدید در فضای اقتصاد نوین، فرصت‌های جدیدی برای بانک‌ها و مؤسسات اعتباردهنده به‌وجود آورده و از سوی دیگر، آن‌ها را نیازمند ابزارها و روش‌های جدیدی (برای مدیریت اعتبار و ریسک‌های ناشی از آن) کرده است (امیری، بکی حسکویی، و بیگلری کامی، ۱۳۹۲). یکی از مهم‌ترین ابزارهایی که بانک‌ها برای مدیریت و کنترل ریسک اعتباری بدان نیازمندند سیستم رتبه‌بندی مشتریان است (سینکی[2]، ۱۹۹۲). بانک‌ها به دو دلیل به وجود سیستمی برای رتبه‌بندی اعتباری مشتریان خود نیازمندند. سیستم رتبه‌بندی اعتباری مشتریان این امکان را فراهم می‌سازد که با اتکا به چنین سیستمی و بر اساس نرخ‌های تکلیفی موجود ریسک پرتفوی اعتباری خود را تا حد ممکن کاهش دهند و از میان مشتریان متقاضی دریافت تسهیلات، معتبرترین و کم‌ریسک‌ترین مشتریان را گزینش کند (سفیدگران و حاله، ۱۳۹۵). در حقیقت، مؤسسهٔ اعتباردهنده با دو نوع تصمیم روبه‌روست: نخست آنکه آیا به یک متقاضی جدید اعتبار بدهد یا خیر؟ و دوم آنکه با مشتریان موجود چگونه رفتار کند؟ روش‌هایی که در مورد اول تصمیم‌گیری می‌کنند رتبه‌بندی اعتباری نامیده می‌شوند، درحالی‌که روش‌هایی که در مورد

---

[1] Chi & Li

[2] Sinkey

مشتریان موجود تصمیم‌گیری می‌کنند رتبه‌بندی رفتاری[1] نامیده می‌شوند (توماس، الدمن، و کروک[2]، ۲۰۰۲).

در مطالعات گذشته، اغلب از روش‌های آماری مانند الگوهای رگرسیونی لاجیت و پروبیت و روش تحلیل ممیزی برای این منظور استفاده می‌شد، ولی در سال‌های اخیر با گسترش فنّاوری‌های نوین اطلاعاتی و کاربردهای وسیع آن در پردازش و طبقه‌بندی اطلاعات، الگوهای مبتنی بر هوش مصنوعی و روش‌های ابتکاری (اخباری، ۱۳۸۷) به‌دلیل قابلیت طبقه‌بندی، تعمیم، و یادگیری الگوها نسبت به سایر روش‌ها از انعطاف‌پذیری بالاتری برخوردار است و در سال‌های اخیر مورد توجه بیشتری قرار گرفته‌اند (کاظمی، قاسمی، و زندیه، ۱۳۹۰).

پژوهش حاضر بر آن است تا به پرسش‌های زیر پاسخ دهد: ۱ـ کدام شاخص‌های مالی و غیرمالی در ریسک اعتباری مشتریان تأثیرگذار است؟ ۲ـ کدام‌یک از ابزارهای داده‌کاوی پیش‌بینی بهتری از رفتار بازپرداخت مشتریان حقوقی خواهد داشت؟ اهمیت موضوع باعث شده است در مقالهٔ حاضر با مطالعهٔ الگوهای مختلف مبتنی بر یادگیری و درنظرگرفتن قابلیت هریک از الگوها، به شناسایی شاخص‌ها و عوامل مالی و غیرمالی تأثیرگذار در ریسک اعتباری و نکول تسهیلات مشتریان پرداخته شود و در نهایت مناسب‌ترین روش برای پیش‌بینی رفتار مشتریان ارائه شد. با توجه به توضیحات، پس از شناسایی و گزینش شاخص‌های اساسی شکل‌دهندهٔ رفتار اعتباری مشتریان، اقدام به اخذ نمونهٔ تصادفی از بین پرونده‌های تسهیلاتی مشتریان صاحبان کسب‌وکار متقاضی تسهیلات شد، سپس فرایند آماده‌سازی داده‌ها با تلخیص و یکپارچه‌سازی و درون‌یابی برخی داده‌های مفقود با الگوریتم‌های شبکهٔ عصبی، درخت تصمیم، و ... صورت گرفت. در گام بعدی به‌منظور سنجش درجهٔ اهمیت عوامل مؤثر در رفتار اعتباری متقاضیان، از الگوریتم‌های درخت تصمیم، شبکهٔ عصبی، و ماشین بردار پشتیبان بهره گرفته و جهت ارزیابی و سنجش روش‌های دسته‌بندی و تعیین درجهٔ اهمیت شاخص‌ها از شاخص «میانگین قدرمطلق خطا» استفاده شد.

در ادامه، مبانی نظری و پیشینهٔ پژوهش توضیح داده و روش‌شناسی پژوهش بر اساس الگوی پنج‌گانهٔ استاندارد داده‌کاوی انجام شده و در پایان به جمع‌بندی و نتیجه پرداخته شده است.

---

[1] Behavioral Scoring

[2] Thomas, Eldman, & Crook

۲ مبانی نظری

۱.۲ ریسک اعتباری

هریک از ریسک‌های بانکی مفهوم و شرایطی دارند، اما باوجود ریسک‌های متعدد و طبقه‌بندی‌های مختلف با دیدگاه‌های متفاوت در حوزهٔ بانکی، برخی ریسک‌ها از نظر ناظران بین‌المللی ـ ازجمله کمیتهٔ بال ـ مستلزم توجه خاصی بوده است که در کشور ما نیز این ریسک‌ها مصداق عینی دارد و لازم است به آن‌ها توجه شود. این ریسک‌ها عبارت‌اند از: ریسک اعتباری، ریسک نقدینگی، ریسک عملیاتی، و ریسک بازار. ریسک اعتباری از منظر شدت احتمال و قدمت در مقایسه با سایر ریسک‌ها در اولویت است. ریسک اعتباری ریسکی است که از نُکول/ قصور طرف قرارداد، یا در حالتی کلی‌تر ریسکی که از «اتفاقی اعتباری» به‌وجود می‌آید. به‌طور تاریخی، این ریسک معمولاً در مورد اوراق قرضه واقع می‌شد، بدین‌صورت که قرض‌دهنده‌ها از بازپرداخت وامی که به قرض‌گیرنده داده بودند، بی‌اطمینانی داشتند. به همین سبب، گاهی اوقات ریسک اعتباری را «ریسک نکول» هم می‌گویند (مهرآرا، موسایی، تصوری، و حسن‌زاده، ۱۳۸۸). ریسک اعتباری را می‌توان به‌عنوان ضرر محتمل که براثر یک رخداد اعتباری اتفاق می‌افتد بیان کرد و ریسک اعتباری اصلی‌ترین عامل ورشکستگی بانک‌ها محسوب می‌شود و تمرکز اعطای تسهیلات با حجم بالا به افراد حقیقی و حقوقی از عوامل افزایش‌دهندهٔ این ریسک خواهد بود (سفیدگران و حاله، ۱۳۹۵؛ محمدی، شاکری، اسکندری، و کریمی، ۱۳۹۶). بانک برای بقای خود و جلوگیری از ورشکستگی نیازمند یک سیستم مدیریت ریسک جامع است تا قادر باشد زیان ناشی از عملیات بانکی را پیش‌بینی و بموقع از آن جلوگیری کند. یکی از ابزارهای مدیریت ریسک اعتباری سیستم رتبه‌بندی و امتیازدهی اعتباری است که یک سیستم تحلیلی مبتنی بر آمار است و به‌منظور تعیین درجهٔ ریسک متقاضیان تسهیلات اعتباری به‌کار می‌رود و به وام‌دهندگان در شناسایی مشتریان خوش‌حساب و بدحساب کمک می‌کند و از این طریق می‌توان ریسک اعتباری را مدیریت کرد. الگوهای مختلفی برای رتبه‌بندی و امتیازدهی اعتباری ارائه شده است که الگوهایی همچون الگوهای رگرسیون لجستیک، تحلیل ممیزی، شبکه‌های عصبی، و غیره نمونه‌هایی از این روش‌ها هستند (سینگ، مورتی، و اِستفز[1]، ۲۰۱۳).

---

[1] Singh, Murthi, & Steffes

۲.۲ داده‌کاوی

داده‌کاوی عبارت است از اقتباس یا استخراج دانش از مجموعه‌ای از داده‌ها (ادوارد و میشکین[1]، ۱۹۹۵). به بیان دیگر، داده‌کاوی فرایندی است که با استفاده از تکنیک‌های هوشمند، دانش را از مجموعه‌ای از داده‌ها استخراج می‌کند. دانش استخراج‌شده در قالب مدل‌ها، الگوها یا قواعد ارائه می‌شود. این الگوها، مدل‌ها و قواعد اشکال مختلفی از ارائهٔ دانش استخراج شده‌اند. این دانش می‌تواند ملاک تصمیم‌گیری‌های آتی، عملکردهای بعدی، و یا تغییرات لازم در سیستم قرار گیرد (مروج، ۱۳۸۳). داده‌کاوی علمی میان‌رشته‌ای و ترکیبی ار علم آمار، یادگیری ماشین، و هوش مصنوعی است (کوتو و دشپاند[2]، ۲۰۱۴). از تکنیک‌های کلاسیک علم آمار از قبیل الگوریتم‌های بیزین، رگرسیون لجستیک، تحلیل تشخیص، و ... به فراوانی در این‌گونه تحقیقات استفاده شده است؛ اما در این پژوهش از ابزارهای جدید و معروف هر حوزه و پرکاربرد در پژوهش‌هایی با داده‌های پیوسته استفاده شده است. به همین منظور از حوزهٔ علم آمار، درخت تصمیم که مبتنی بر جداول توافقی است، از حوزهٔ یادگیری ماشین، ابزار ماشین بردار پشتیبان، و از حوزهٔ هوش مصنوعی، ابزار شبکه‌های عصبی مصنوعی مورد توجه قرار گرفت.

در ادامه، سه روش داده‌کاوی که مورد استفاده در پژوهش حاضر بوده، معرفی شده است.

۱.۲.۲ درخت تصمیم

درخت‌های تصمیم روشی برای نمایش یک‌سری از قوانین‌اند که منتهی به یک رده یا مقدار می‌شوند. درخت تصمیم یک روش کارآمد طبقه‌بندی داده‌هاست که به‌دلیل برخورداری از ویژگی‌هایی مانند سادگی، قابل‌فهم‌بودن، دقت و سرعت الگوریتم، بسیار پرکاربرد است. یک درخت تصمیم از مجموعهٔ نمونه‌های ورودی و خروجی ایجاد شده است، از روش یادگیری با ناظر استفاده می‌کند و با بیان یکسری قوانین ضمنی تعمیم‌پذیر، اقدام به پیش‌بینی رفتار متغیر وابسته (برچسب دسته) می‌کند (اولیواس[3]، ۲۰۰۷). درخت‌های تصمیم از طریق جداسازی متوالی داده‌ها به گروه‌های مجزا ساخته می‌شوند و هدف در این فرایند افزایش فاصله بین گروه‌ها در هر جداسازی است. از معروف‌ترین روش‌های درخت تصمیم می‌توان به

---

[1] Edward & Mishkin

[2] Kotu & Deshpande

[3] Olivas

CHAID[1]، CART[2]، ID3[3]، C4.5، C5، QUEST[4]، و AID[5] اشاره کرد که با توجه به نوع گسسته یا پیوسته‌بودن داده‌ها، هریک با توجه به نوع پژوهش انتخاب خواهد شد.

۲.۲.۲ شبکه‌های عصبی مصنوعی[6]

به‌طورکلی، می‌توان گفت شبکهٔ عصبی روشی است که قصد دارد با استفاده از الگوهای ریاضی و توان کامپیوتری، به شبیه‌سازی فرایند یادگیری انسان بپردازد؛ به‌عبارت دیگر، شبکهٔ عصبی مصنوعی یک «پردازندهٔ توزیع‌شدهٔ موازی[7]» است که میل طبیعی برای ذخیرهٔ دانش تجربی و کاربردی‌کردن آن را دارد. شبکهٔ عصبی مصنوعی از دو جهت به مغز انسان شباهت دارد.

- دانش از طریق یک فرایند یادگیری توسط شبکه کسب می‌شود.
- قدرت ارتباطی که به‌عنوان وزن‌های سیناپسی[8] شناخته می‌شود، برای ذخیرهٔ دانش مورد استفاده قرار می‌گیرد.

۳.۲.۲ ماشین بردار پشتیبان[9]

استفاده از بردارهای پشتیبان خطی در مسائل دسته‌بندی رویکرد جدیدی است که در چند سال اخیر مورد توجه بسیاری قرار گرفته است. ماشین‌های بردار پشتیبان دارای یک‌سری خواص‌اند که در زیر به تعدادی از آن‌ها اشاره می‌شود:

- طراحی دسته‌بندی‌کننده با حداکثر تعمیم،
- رسیدن به بهینهٔ سراسری تابع هزینه،
- تعیین خودکار ساختار و توپولوژی بهینه برای طبقه‌بندی‌کننده.

الگوریتم‌های مبتنی بر ماشین‌های بردار پشتیبان الگوریتم‌هایی‌اند که سعی می‌کنند برای دسته‌بندی داده‌های برچسب یک حاشیه[10] را بیشینه کنند.

---

[1] Chi-Squared Automatic Interaction Detection
[2] Classification and Regression Trees
[3] Iterative Dichotomizer3rd
[4] Quick, Unbiased, Efficient, & Statistical Tree
[5] Automatic Interaction Detection
[6] artificial neural networks
[7] parallel distributed processor
[8] Synaptic weights
[9] Support Vector Machine
[10] margin

۳ پیشینهٔ پژوهش

پژوهش‌ها و کاربردهای متعددی در حوزهٔ اعتبارسنجی مشتریان بانک‌ها صورت پذیرفته است. روش قضاوتی در اعتبارسنجی به‌دلیل خطا و زمان زیاد به‌تدریج جای خود را به روش‌های پارامتریک و ناپارامتریک داده است. روش‌های پارامتریک مانند لاجیت، پروبیت، و رگرسیون لجستیکی که در ابتدای ظهور اعتبارسنجی مورد استفاده قرار گرفتند و سپس استفاده از روش‌های ناپارامتریک و داده‌کاوی مانند درخت تصمیم، شبکه‌های عصبی و ... مورد توجه محققان قرار گرفت (مهرآرا و همکاران، ۱۳۸۸). در این مورد، می‌توان به مطالعهٔ فیشر[1] (۱۹۳۶) به‌عنوان اولین سیستم ارزیابی تقاضای اعتبار و مطالعهٔ دوراند[2] (۱۹۴۱) که با استفاده از تحلیل تشخیص و با تکیه بر نتایج فیشر انجام گرفت، اشاره کرد.

چی، تَن، وگوه[3] (۲۰۰۶) در پژوهشی در خصوص ساختار رتبه‌بندی اعتباری، از روش‌های درخت تصمیم، شبکهٔ عصبی، و رگرسیون لجستیک در ساخت الگوی داده‌کاوی استفاده کردند و نتایج به‌دست‌آمده از تحقیقات حاکی از آن است که تلفیق الگوها نتایج بهتری را نسبت به نتایج فردی هریک از الگوها دربر خواهد داشت. نظری و علیدادی (۲۰۱۳) در پژوهش خود از تکنیک شبکهٔ عصبی مصنوعی برای سنجش ریسک اعتباری متقاضیان استفاده کردند. نتایج نشان داد که تعداد دفعات وام و میزان وام مهم‌ترین تأثیر و همچنین وضعیت حساب بانکی مشتری، سابقهٔ ارتباط مشتری با بانک و خدمات دریافتی، در شناسایی معیارهای طبقه‌بندی مشتریان خوب و بد تأثیر کمتری دارد. دومپاس و زورپونیدیس[4] (۲۰۱۴) با استفاده از روش پرومتی۲[5] به رتبه‌بندی بانک‌ها بر اساس معیارهای مدیریت ریسک مالی پرداختند. آن‌ها علت استفاده از این روش را توانایی بررسی جزئیات بیشتر گزینه‌ها در مقایسه با یکدیگر عنوان کردند. چی و لی (۲۰۱۷) در مقالهٔ خود به بررسی رابطهٔ بین عدم قطعیت سیاست‌های اقتصادی با ریسک اعتباری بانک‌ها و تصمیمات وام‌دهی آن‌ها پرداختند. نتایج پژوهش حاکی از آن است که عدم قطعیت سیاست‌های اقتصادی ریسک اعتباری بانک‌ها را افزایش می‌دهد و تأثیر منفی در میزان وام، به‌ویژه حقوق صاحبان سهام بانک دارد.

---

[1] Fisher
[2] Durand
[3] Chye, Tan, & Goh
[4] Doumpos & Zopounidis
[5] Promethee2

شریعت پناهی و هاشمی برکادهی (۱۳۸۸) در پژوهشی نشان دادند پنج شاخص نسبت بدهی به سرمایه، نسبت سود قبل از بهره و مالیات به خالص فروش، نسبت سود قبل از بهره و مالیات به حقوق صاحبان سهام، نسبت سرمایهٔ در گردش به کل دارایی‌ها، و نسبت بدهی جاری به کل دارایی‌ها قدرت بالایی در تفکیک گروه‌های شرکت‌های دارای نکول و بدون نکول دارا هستند و همچنین دریافتند نسبت بازده دارایی با احتمال نکول رابطهٔ معکوس دارد.

البرزی، خان‌بابایی، و محمدپور زرندی (۱۳۹۲) بیان کردند استفاده از الگوی ترکیبی در بهینه‌سازی درختان تصمیم‌گیری توسط تکنیک الگوریتم ژنتیک منجر به افزایش دقت طبقه‌بندی نسبت به بسیاری از الگوریتم‌های مقایسه‌شدهٔ پژوهش آن‌ها شده است؛ ولی پیچیدگی الگوریتم الگوی ترکیبی پیشنهادی از برخی الگوریتم‌های طبقه‌بندی مقایسه‌شده در این مقاله بیشتر است. سهرابی، رئیسی وانانی، و زارع میرک‌آباد (۱۳۹۵) در پژوهش خود با هدف کاستن مشکلات ارائهٔ تسهیلات مناسب به مشتریان، به طراحی سیستم توصیه‌گر تسهیلات بانکی با استفاده از داده‌های قبلی مربوط به تسهیلات ثبت‌شده در یک بانک منتخب و بزرگ کشور پرداختند. به‌منظور طراحی راهکار توصیه‌گر، روش‌های داده‌کاوی مناسب به‌همراه رویکرد اعتبارسنجی متناسب با هر روش‌انتخاب کردند.

بررسی حاصل از مطالعات فوق نشان می‌دهد پژوهشگران در خصوص رتبه‌بندی اعتباری مشتریان از روش‌ها و شاخص‌های گوناگونی استفاده نمودند که در این پژوهش هدف شناسایی شاخص‌های مهم براساس سه روش پرکاربرد در حوزه داده‌کاوی بوده است که در ادامه به بررسی آن ها پرداخته شده است.

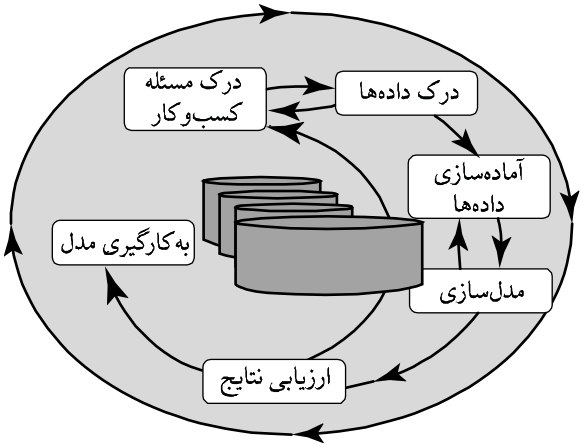
۴ روش‌شناسی پژوهش

با توجه به ماهیت تحقیق که استفاده از داده‌کاوی جهت طبقه‌بندی مشتریان بانک است، این تحقیق از نوع داده‌محور[1] می‌باشد. پایهٔ اصلی تحقیق حاضر بر کشف دانش از پایگاه داده‌های بانک مورد مطالعه نهاده شده است. از این‌رو، استاندارد جهانی CRISP-DM[2] جهت انجام فرایند تحقیق مورد استفاده قرار گرفته است؛ که در شکل ۱ این مراحل که شامل درک مسئلهٔ کسب‌وکار، درک داده‌ها، آماده‌سازی داده‌ها، الگوسازی، ارزیابی نتایج، و به‌کارگیری الگوست، به‌همراه ارتباط بین مراحل مشخص است.

---

[1] data oriented

[2] CRoss Industry Standard Process for Data Mining



شکل ۱. فرایند داده‌کاوی الگو CRISP-DM

مأخذ: چاپمن و همکاران[1](۱۹۹۹)

بر این اساس پژوهش حاضر برای بررسی فرضیه‌های اشاره شده مراحل درک مسئله کسب و کار، درک داده‌ها، ساختار و آماده‌سازی داده‌ها، مدل‌سازی و ارزیابی نتایج را طی می‌کند که شرح این مراحل در زیر آورده شده است.

۱.۴ درک مسئلهٔ کسب‌وکار

در این فاز از فرایند پژوهش، نخست اهداف اصلی کسب‌وکار تعیین شد که اصلی‌ترین هدف کسب‌وکار (بانک مورد مطالعه) در این تحقیق پیش‌بینی وضعیت اعتباری مشتریان صاحبان کسب‌وکار بانک است و این کار بر اساس کشف الگوهای پنهان موجود بین خصیصه‌های مشتریانی است که در گذشته تسهیلات اعتباری به آن‌ها اعطا شده است. بر همین اساس، موقعیت موجود محیط که بخش اعتبارات بانک است، مورد ارزیابی قرار گرفت تا فرایند اعطای اعتبار به مشتریان و بررسی‌های کارشناسی خبرگان اعتباری بررسی شود. همچنین، نحوهٔ اخذ اطلاعات از مشتریان مورد بررسی دقیق قرار گرفت تا شناخت مناسبی از وضعیت کلی طبقه‌بندی به مشتریان در حالت فعلی کسب شود.

[1] Chapman, et al.

۲.۴ درک داده‌ها

در این مرحله، گام‌های زیر انجام شده است:

۱.۲.۴ گزینش شاخصه‌های اساسی شکل‌دهندهٔ رفتار مشتریان

به‌منظور انتخاب و گزینش شاخصه‌های اساسی شکل‌دهندهٔ رفتار مشتریان، نتایج پژوهش‌های انجام‌شده در نظام بانکی (داخلی و خارجی) مطابق جدول پیوست شمارهٔ ۱ و سیستم‌های انتخاب و گزینش شاخص مؤثر در رفتار اعتباری متقاضیان تسهیلات اعتباری شامل 6C[1]، LAPP[2]، و 5P[3] مورد استفاده قرارگرفت و نظر کارشناسان مجرب بانکی نیز در این نتایج تأثیر داده شد. در نهایت بر اساس ادبیات موضوع، تعداد ۸۵ متغیر مستقل مؤثر در رفتار اعتباری گیرندگان تسهیلات انتخاب شد که در گروه متغیرهایی شامل ویژگی‌های پروندهٔ تسهیلاتی، مشخصات شرکت متقاضی، وضعیت تضمینات اخذشده، وضعیت صورت‌های مالی و سابقهٔ اعتباری مشتریان مورد پژوهش دسته‌بندی شدند.

شایان ذکر است در این پژوهش شاخص نرخ نکول (نسبت ماندهٔ تسهیلات غیرجاری به اصل و سود تسهیلات) به‌عنوان متغیر وابسته مدنظر قرارگرفته و همچنین همهٔ این متغیرها در ساخت الگو دخالت داده شدند، اما تشخیص متغیرهای مهم بر عهدهٔ الگو گذاشته شد و پژوهشگران دخالتی در این امر نداشتند.

۲.۲.۴ نمونه‌گیری و جمع‌آوری داده‌ها

جامعهٔ آماری این پژوهش پرونده‌های تسهیلاتی مشتریان صاحبان کسب‌وکار متقاضی تسهیلات بالای ۱,۰۰۰,۰۰۰,۰۰۰ ریال بانک مورد پژوهش از سال ۱۳۹۰ تا پایان سال ۱۳۹۵ است.

برای انتخاب نمونه‌های تصادفی از روش نمونه‌گیری طبقه‌بندی‌شده با فاصلهٔ اطمینان ۹۵ درصد بهره گرفته شد که به انتخاب ۶۰۹ پرونده از بین ۱۸۲۹ پرونده منجر شد که اطلاعات آماری این پرونده‌ها از طریق اطلاعات ثبت‌شده در نظام بانکداری متمرکز بانک و مکاتبه با شعب جمع‌آوری گردید.

---

[1] Six C of Credit

[2] Liquidity Activity Profitability Potential

[3] Five P of Credit

۳.۴ ساختار و آماده‌سازی داده‌ها

از بین ۶۰۹ نمونهٔ اخذشده، تعداد ۷۳ پرونده به‌دلیل تسویه‌شدن و ۱۵ پرونده به‌دلیل عدم ارائهٔ اطلاعات از سوی شعب از پژوهش کنار گذاشته شد. به‌منظور افزایش کیفیت داده‌های جمع‌آوری‌شده بنا به برخی دلایل ازجمله عدم ثبت، ورود ناقص، یا اشتباه برخی داده‌های پرونده‌های اعتباری در سامانهٔ متمرکز الکترونیک و نبود یا وجود تناقض در داده‌های مربوط به شاخص‌های مالی پرونده‌های اشخاص حقوقی، فرایند آماده‌سازی داده‌ها با عاری‌ساختن آن‌ها از داده‌های ناقص[۱]، مغشوش[۲] و ناسازگار[۳]، تلخیص و یکپارچه‌سازی و درون‌یابی برخی داده‌های مفقود با الگوریتم‌های شبکهٔ عصبی و درخت تصمیم صورت پذیرفت. از بین ۵۲۱ پروندهٔ باقی‌مانده تنها ۱۰ درصد پرونده‌ها دارای اطلاعات کامل همهٔ شاخص‌ها بود. با ادغام برخی شاخص‌های مؤثر در رفتار اعتباری گیرندگان تسهیلات، نهایتاً ۷۶ شاخص جهت الگوسازی انتخاب شد. از بین شاخص‌های منتخب تنها تقریبا ۴۱ درصد شاخص‌ها دارای اطلاعات کامل در ۵۲۰ پرونده از ۵۲۱ پرونده مورد بررسی بوده است. (مطابق جدول پیوست شمارهٔ ۲).

به‌منظور افزایش کیفیت داده‌های جمع‌آوری‌شده بنا بر برخی دلایل ازجمله عدم ثبت، ورود ناقص یا اشتباه برخی داده‌های پرونده‌های اعتباری در سامانهٔ متمرکز الکترونیک و نبود داده‌های مربوط به شاخص‌های مالی پرونده‌های تسهیلاتی، فرایند آماده‌سازی داده‌ها با عاری‌ساختن آن‌ها از داده‌های ناقص، پرت، و ناسازگار، تلخیص، یکپارچه‌سازی، و درون‌یابی برخی داده‌های مفقود و ... صورت پذیرفت. به این منظور از الگوریتم درخت تصمیم استفاده شد(برخی از الگوهای درخت تصمیم فقط می‌توانند مقادیر متغیرهای گسسته را پیش‌بینی کنند، درحالی‌که در این پژوهش متغیرهای منتخب، پیوسته هستند؛ بر همین اساس از بین الگوهای مختلف درخت تصمیم، درخت چاید[۴] استفاده شده است) در ادامه، به بررسی هریک از اختلالات یادشده به شرح زیر می‌پردازیم.

۱ـ داده‌های ناقص

مطابق با جدول ۲، ۷۶ شاخص با کمبود داده مواجه بود که از این میان برخی شاخص‌هامانند سود خالص به دارایی ثابت، سود به فروش، سود خالص به فروش، با کمبود در حدود ۶۰

---

[۱] ناقص (incomplete): نمونه‌های ناکافی، کمبود برخی مقادیر مشخصه‌ها و...

[۲] مغشوش (noisy):داده‌های دارای مقادیر خطا و نامفهوم

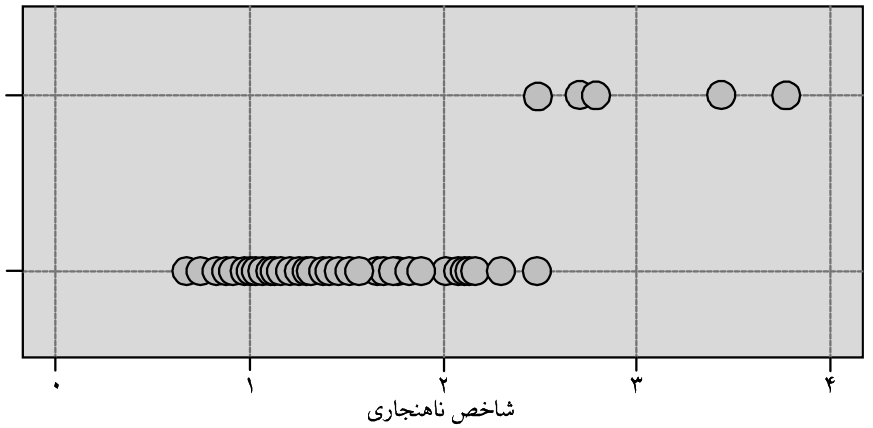
[۳] ناسازگار (inconsistent): داده‌های دارای تناقض

[4] CHAID

درصدی داده مواجه بودند. برای جای‌گذاری داده‌های مفقود[1] از تکنیک‌های مختلفی استفاده شد؛ برای نمونه برای جای‌گذاری شاخص سود خالص به دارایی ثابت، سود به فروش، سود خالص به فروش به‌ترتیب از درخت تصمیم کارت، شبکهٔ عصبی و شبکهٔ عصبی استفاده شد.

۲ـ داده‌های پرت

درصورتی‌که دادهٔ پرت را داده‌های دور از مرکز تجمع داده‌ها بدانیم، طبق جدول ۳، تعداد داده‌های پرت ۲۱۰ مورد است که حذف همهٔ آن‌ها به ازدست‌دادن اطلاعات این گروه از داده منجر می‌شود. برای حل این مشکل از یک الگو با قابلیت کشف مغایرت در اطلاعات گروه‌های غالب[2] به‌شکل سیستمی شکل ۱ استفاده شد. این الگو ابتدا پرونده‌ها را بر اساس فاصله اقلیدسی از مرکز خوشه‌ها به صورت سیستمی به چند گروه دسته‌بندی کرده و سپس پرونده‌هایی که در هیچ‌یک از گروه‌ها جای نگرفته است، به‌عنوان دادهٔ پرت معرفی می‌کند (جدول پیوست ۳).



شکل ۲. نمایش داده‌ای پرت به‌صورت سیستمی

در شکل ۲ شاخص نابهنجاری در محور افقی و گروه‌های دوگانه (نقاط به‌هم‌پیوسته) در محور عمودی نمایش داده شده است. برحسب شاخص نابهنجاری، آن گروه از نقاطی که بیشترین فاصله را با یکی از گروه‌ها دارد از مجموع نقاط حذف می‌کنیم (مقدم و همکاران، ۱۳۹۲). در

[1] NULL
[2] peer group

این میان، دو نقطه فاصلهٔ محسوسی با گروه‌های دوگانه دارند و در نتیجه آن دو از الگو حذف شدند. شاخص نابهنجاری در اینجا بزرگ‌تر از ۳ در نظر گرفته شده است.

۳ـ داده‌های ناسازگار

به آن دسته از داده‌های یک شاخص اطلاق می‌شود که با داده‌های سایر شاخص‌ها متناقض باشد؛ به‌طور مثال در شاخص تعداد حساب‌های قرض‌الحسنه، مشتریان بدون داشتن حساب، دارای موجودی ماندهٔ حساب قرض‌الحسنه هستند.

با توجه به وجود ناسازگاری شدید داده‌ها در سه شاخص ۱ـ تعداد حساب‌های بلندمدت، ۲ـ تعداد حساب‌های کوتاه‌مدت، ۳ـ تعداد حساب‌های قرض‌الحسنه با میانگین موجودی حساب‌های مربوطه، از این شاخص‌ها صرف‌نظر شد.

۱.۳.۴ شناسایی شاخص‌های کم‌اثر

پس از آماده‌سازی داده‌ها، کلیهٔ شاخص‌ها در فرایند الگوسازی یک‌بار مورد استفاده گرفتند. نتایج نشان داد تعداد زیادی از شاخص‌ها نتوانستند الگوهایی با میانگین قدر مطلق خطای پایین تولید کنند. بنابراین پس از تکمیل داده‌ها، ازآنجاکه همهٔ شاخص‌ها اهمیت یکسانی ندارند، برای جداسازی و حذف شاخص‌های کم‌اهمیت از سایر شاخص‌ها از الگوریتم‌های انتخاب خصیصه[1] استفاده شد (همان). این الگوریتم‌ها برای غربال کردن داده‌ها در دو سطح اقدام به ارزش‌گذاری داده‌ها می‌کنند که این دو سطح عبارتند از الف‌ـ برحسب درصد مقادیر گمشده، پراکندگی پایین متغیرها و ...، ب‌ـ آزمون‌های همبستگی با ضریب همبستگی کرامر[2]، پیرسون[3]، و ....

چکیدهٔ نتایج حاصل از این الگوریتم‌ها در قالب جدول ۴ نشان داده شده است. در این جدول، شاخص‌ها برحسب ارزش[4] به سه گروه مهم، سرمرز، و غیرمهم تقسیم شده‌اند. به‌شکل معمول، متغیرهای مهم و حاشیه‌ای حفظ و متغیرهای غیرمهم از فرایند پژوهش حذف شدند. شایان ذکر است براساس تجربه پژوهشگران ، دوشاخص غیر مهم شامل مبلغ وام و جمع تضامین جهت بررسی بیشتر در الگو باقی ماند.

---

[1] Feature Selection

[2] Cramer

[3] Pearson

[4] value

جدول ۱
درجهٔ اهمیت متغیرها

شاخص	درجهٔ اهمیت	ارزش
مدت قرارداد	مهم	۱
نوع عقد	مهم	۱
نوع اقساط	مهم	۱
تعداد اقساط	مهم	۱
نرخ سود	مهم	۱
تعداد کوتاه‌مدت	سر مرز	۹۹/۹۵
تعداد پرونده در بانک توسعه	سر مرز	۹۹/۹۴
کد نوع عقد فنّاوری	سر مرز	۹۹/۹۰
مبلغ چک برگشتی در نظام بانکی	سر مرز	۹۹/۷۲
مبلغ امهال	سر مرز	۹۹/۲۷
بخش اقتصادی	سر مرز	۹۸/۸۵
شغل ضامن دوم	سر مرز	۹۸/۳۳
نوع شرکت	سر مرز	۹۸/۲۶
سود عملیاتی به دارایی	سر مرز	۹۷/۲۹
تعداد کل چک‌های برگشتی قبل از دریافت وام در بانک توسعه	سر مرز	۹۷/۲۷
سود خالص به دارایی ثابت	سر مرز	۹۷/۰۲
تعداد چک برگشتی در نظام بانکی	سر مرز	۹۶/۷۴
میانگین قرض‌الحسنهٔ ۶ ماه قبل	سر مرز	۹۶/۴۴
میانگین قرض‌الحسنهٔ ۳ ماه قبل	سر مرز	۹۵/۷۷
آیا صورت مالی شرکت حسابرسی شده است؟	سر مرز	۹۵/۷۳
علت پرداخت تسهیلات	سر مرز	۹۲/۳۱
مبلغ وام	غیر مهم	۸۵/۹۰
جمع تضمینات	غیر مهم	۷۷/۵۹

۴.۴ الگوسازی

در گام بعدی، پس از آماده‌سازی داده‌ها به‌منظور سنجش درجهٔ اهمیت عوامل مؤثر در رفتار اعتباری متقاضیان، از الگوریتم‌های درخت تصمیم، شبکهٔ عصبی، و ماشین بردار پشتیبان به کمک نرم‌افزار کلمنتاین[1] استفاده شد. همچنین به‌منظور ارزیابی و سنجش روش‌های

[1] Clementine

دسته‌بندی و تعیین درجهٔ اهمیت شاخص‌ها، از شاخص «میانگین قدر مطلق خطا» استفاده شد.

برای این منظور، داده‌ها به دو مجموعهٔ تصادفی داده‌های آموزشی[1]، و آزمون[2] تقسیم شد که از داده‌های آموزشی جهت الگوسازی و از داده‌های آزمون جهت بررسی ثبات و تکرارپذیری نتایج الگو بهره گرفته شد.

در این مرحله، به برازش داده‌ها با هدف شناسایی الگوهای برتر و شناسایی شاخص‌های غیرمؤثر اقدام شد. همچنین به‌منظور داوری دقیق‌تر دربارهٔ داده‌ها، آن‌ها به دو گروه آموزش و آزمون تقسیم شدند. برای آموزش از ۵۰ الی ۷۰ درصد داده‌ها استفاده گردید در نتیجه باقی مانده داده‌ها درگروه آزمون قرا گرفتند. لازم به ذکر است تغییر در نسبت داده‌های آموزش و آزمون به‌منظور بررسی ثبات مدل بوده است. سپس به حل سه الگوی درخت تصمیم چاید، شبکهٔ عصبی، و ماشین بردار پشتیبان اقدام شد[3] (نتایج جدول ۵).

با توجه به نتایج جدول ۵ و بر پایهٔ میانگین قدر مطلق خطای به‌دست‌آمده برای داده‌های تست الگو و میزان ثبات و تکرارپذیری، نتایج الگوی درخت تصمیم چاید به‌عنوان بهترین الگو برای پیش‌بینی نکول برگزیده شد که در این میان درخت تصمیم چاید ۱ به‌عنوان الگوی منتخب با میانگین قدر مطلق خطای ۱۴ درصد بهترین وضعیت را برای پیش‌بینی نرخ نکول متقاضیان تسهیلات دارد.

---

[1] train

[2] test

[3] هریک از الگوریتم‌های داده‌کاوی ذیل، با نسبت‌های مختلف داده‌های آموزش و آزمون حل شد. به همین دلیل در هر بار حل، یک شماره به هر الگوریتم حل‌شده نسبت داده شد؛ مانند چاید ۱، چاید ۲، و...

جدول ۲

میانگین قدر مطلق خطای پیش‌بینی

میانگین	میانگین قدر مطلق خطا		الگوریتم
	دادهٔ آزمون	دادهٔ آموزش	
۰.۱۴	۰.۱۴	۰.۱۴	CHAID1
۰.۱۶	۰.۱۶	۰.۱۵	CHAID2
۰.۱۶	۰.۱۷	۰.۱۵	CHAID3
۰.۱۵	۰.۱۶	۰.۱۵	میانگین
میانگین	میانگین قدر مطلق خطا		الگوریتم
	دادهٔ آزمون	دادهٔ آموزش	
۰.۱۹	۰.۱۹	۰.۱۸	ANN1
۰.۲۰	۰.۲۱	۰.۲۰	ANN2
۰.۲۰	۰.۲۰	۰.۲۱	ANN3
۰.۲۰	۰.۲۰	۰.۱۹	میانگین
میانگین	میانگین قدر مطلق خطا		الگوریتم
	دادهٔ آزمون	دادهٔ آموزش	
۰.۱۹	۰.۲۰	۰.۱۹	SVM1
۰.۱۹	۰.۲۱	۰.۱۸	SVM2
۰.۲۰	۰.۲۰	۰.۲۱	SVM3
۰.۲۰	۰.۲۰	۰.۱۹	میانگین

در ادامه، متغیرهای مؤثر در نتایج حاصل از الگوریتم مذکور به شرح جدول ۶ ارائه شده است.

جدول ۳
اندازهٔ اهمیت شاخص‌ها

شاخص	اهمیت
مدت قرارداد	۰/۵۸
مبلغ امهال	۰/۴۶
تعداد اقساط	۰/۲۴
سود عملیاتی به دارایی	۰/۰۵
نوع عقد	۰/۰۳
میانگین قرض‌الحسنهٔ ۳ ماه قبل	۰/۰۳
مبلغ وام	۰/۰۱

۵.۴ ارزیابی نتایج

مطابق نتایج حاصل از الگو، مدت قرارداد، مبلغ امهال، تعداد اقساط، سود عملیاتی به دارایی، نوع عقد، میانگین قرض‌الحسنهٔ ۳ ماه قبل، و مبلغ وام اهمیت بیشتری نسبت به سایر متغیرها دارند. همان‌طور که مشخص است، مهم‌ترین متغیر مدت قرارداد با ضریب ۰/۵۸ است که می‌توان گفت دورهٔ بازپرداخت تسهیلات باید تناسب لازم را با دورهٔ درآمدزایی طرح‌های اقتصادی داشته باشد تا احتمال نکول کاهش یابد. به‌عبارت دیگر، یکی از معضلات اساسی نظام بانکی در حوزهٔ اعطای تسهیلات بی‌توجهی به این اصل کلیدی است. در ادامه به توضیح سایر متغیرها پرداخته شده است:

مبلغ امهال: از لحاظ نظری، امهال به‌منظور ایجاد فرصت برای بنگاه اقتصادی برای بازپرداخت تسهیلات دریافتی متناسب با شرایط و توانایی بنگاه است. متأسفانه در سال‌های اخیر، بانک‌ها از این روش به‌صورت گسترده به‌منظور کاهش صوری مطالبات غیرجاری خود استفاده می‌کنند که این مسئله بانک‌ها را با دارایی‌های موهومی مواجه ساخته است. در نتیجه بنگاه‌ها را بدهکارتر می‌کند.

تعداد اقساط: امکان متناسب‌سازی مقاطع کسب درآمد بازپرداخت بخشی از تسهیلات را فراهم می‌آورد یا با تدریجی‌کردن بازپرداخت تسهیلات احتمال بازگشت اصل و سود تسهیلات را ممکن می‌سازد. درصورتی‌که تعداد اقساط با ظرفیت‌های مالی بنگاه هماهنگ نباشد، بانک با مشکل عقب‌افتادن اقساط مواجه خواهد شد.

سود عملیاتی به دارایی: این نسبت مالی از توانایی مدیران در به‌کارگیری دارایی‌های خود نشان دارد. در واقع، این نسبت همواره به هنگام اعطای تسهیلات به بنگاه‌ها باید مورد توجه قرارگیرد، زیرا اگر این نسبت برای بنگاه متقاضی تسهیلات کوچک‌تر از متوسط صنعت بنگاه

مدنظر باشد، به‌راحتی می‌توان دریافت که مدیران این بنگاه در به‌کارگیری دارایی‌های خود به‌خوبی عمل نکرده‌اند، در نتیجه در اعطای تسهیلات باید محتاطانه عمل کرد، زیرا ریسک نکول این‌گونه بنگاه‌ها به‌نسبت بالاست.

نوع عقد: به‌سبب ماهیت متفاوت عقود اسلامی، نوع عقد می‌تواند در احتمال نکول اثرگذار باشد.

میانگین قرض‌الحسنهٔ ۳ ماه قبل: به‌عنوان شاخص اساسی مراودهٔ مشتری با بانک در نظر گرفته می‌شود. هرچه این نسبت بیشتر باشد، حاکی از کاهش ریسک نکول مشتری خواهد بود.

مبلغ وام: درصورتی‌که با اندازه و توانایی‌های بنگاه همخوانی نداشته باشد، بانک با مشکل دیرکرد تعهدات بنگاه مواجه خواهد شد.

لذا، توجه به این شاخص‌ها قبل از اعطای تسهیلات به مشتریان صاحبان کسب‌وکار می‌تواند در تصمیم‌گیری و پیش‌بینی رفتار اعتباری مشتری نقش بسیار حائز اهمیتی داشته باشد. این شاخص‌ها مبنایی برای طراحی و تدوین فرم‌های اعتبارسنجی مشتریان خواهند بود. همچنین، الگوریتم درخت تصمیم چاید ۱، با میانگین قدر مطلق خطای ۱۴ درصد، قابلیت برآورد احتمال نکول تسهیلات دریافتی مشتریان را نسبت به سایر تکنیک‌ها دارد.

۵ نتیجه‌گیری

موضوع سنجش ریسک اعتباری به‌منظور پیش‌بینی احتمال عدم بازپرداخت وام، یکی از سازوکارهای مهم مدیریت ریسک اعتباری است. ازاین‌رو، بانک‌ها برای مدیریت اعتباراتشان از یک روش خاص برای رتبه‌بندی اعتباری مشتریان خود استفاده می‌کنند. ازآنجایی‌که تصمیمات در خصوص اعطای تسهیلات از حساسیت بالایی برخوردار است، لذا انتخاب نادرست متقاضی تبعات جبران‌ناپذیری برای بانک در پی خواهد داشت. در این پژوهش سعی شده است با مطالعهٔ اغلب پژوهش‌های صورت‌پذیرفته در خصوص سنجش ریسک اعتباری مشتریان در داخل و خارج از کشور، تمام شاخص‌ها شناسایی و مورد ارزیابی قرار گیرد و سپس از میان این شاخص‌ها، آن‌هایی که بیشترین تأثیر را در ریسک اعتباری مشتریان دارند، شناسایی شود. از طرفی دیگر با مطالعهٔ پژوهش‌های صورت‌گرفته، متوجه آن شدیم که پژوهشگران اغلب از داده‌های گسسته استفاده کرده‌اند، لذا نویسندگان برای ارزیابی بهتر، از داده‌های پیوسته استفاده کردند. همچنین برای مقایسهٔ بهتر شاخص‌ها، از تکنیک‌های درخت تصمیم، شبکهٔ عصبی، و ماشین بردار پشتیبان استفاده شدکه نتایج حاصل از پژوهش نشان می‌دهد درخت تصمیم درصد خطای کمتری برای پیش‌بینی دارد و در پایان متغیرهای

مدت قرارداد، مبلغ امهال، تعداد اقساط، سود عملیاتی به دارایی، نوع عقد، میانگین قرض‌الحسنهٔ سه ماه قبل، و مبلغ وام اهمیت بیشتری نسبت به سایر متغیرها دارند. بنابراین، توجه به این شاخص‌ها پیش از اعطای تسهیلات به مشتریان صاحبان کسب‌وکار می‌تواند در تصمیم‌گیری و پیش‌بینی رفتار اعتباری مشتری نقش بسیار حائز اهمیتی داشته باشد.

از مزایای این تحقیق بررسی تمام متغیرهای به‌کاربرده‌شده در حوزهٔ سنجش ریسک اعتباری و به‌تبع آن شناسایی متغیرهای مؤثر در رفتار اعتباری و طبقات مشتریان با استفاده از ابزارهای داده‌کاوی، استفاده از داده‌های پیوسته (به‌جای تمرکز صرف بر داده‌های گسسته)، و همچنین کاهش خطاهای انسانی و جلوگیری از فرایند قضاوتی کارشناسان خبره به‌وسیلهٔ به‌کارگیری الگوهای کشف‌شدهٔ این تحقیق در یک سیستم هوشمند و پاسخ‌گویی سریع‌تر به متقاضیان تسهیلات اعتباری ماست. ازجملهٔ محدودیت‌های این تحقیق نبودِ اطلاعات منسجم در یک سیستم اطلاعاتی رایانه‌ای، نبودِ وحدت رویه در ثبت اطلاعات مشتریان و ثبت داده‌ها در یک سری زمانی مشخص، و وجود اطلاعات ناقص و غلط در داده‌های مورد بررسی بود که داده‌ها را با نویز بسیاری مواجه ساخت.

منابع و مآخذ

اخباری، م. (۱۳۸۷). رتبه‌بندی اعتباری مشتریان حقوقی بانک‌ها با رویکرد هوش مصنوعی (پایان‌نامهٔ کارشناسی ارشد). دانشگاه صنعتی اصفهان، اصفهان.

احمدی‌زاده، ک. (۱۳۸۵). لزوم تأسیس مراکز اعتبارسنجی و رتبه‌بندی. *فصلنامهٔ حسابرس*، ۳۴.

البرزی، م.، محمدپور زرندی، م. ا.، و خان‌بابایی، م. (۱۳۹۲). به‌کارگیری الگوریتم ژنتیک در بهینه‌سازی درختان تصمیم‌گیری برای اعتبارسنجی مشتریان بانک‌ها. *مدیریت فنّاوری اطلاعات*، ۲(۴)، ۳۸ـ۲۳.

اکرامی، م.، و راهنما اسکی، آ. (۱۳۸۸). بررسی عوامل مؤثر در مطالبات سررسید گذشته و معوق بانک. *پژوهشنامهٔ اقتصادی*، ویژه‌نامهٔ بانک، ۶،۸۵ـ۱۰۷.

امیری، م.، بکی حسکوئی، م.، و بیگلری کامی، م. (۱۳۹۲). رتبه‌بندی شرکت‌های تولیدی در بورس اوراق بهادار تهران با استفاده از مدل‌های تصمیم‌گیری با معیارهای چندگانه و شبکهٔ عصبی مصنوعی. *فصلنامهٔ دانش سرمایه‌گذاری*، ۷، ۷۳ـ۸۴.

پرویزیان، ک.، ذکاوت، م. و محمدیان، م.(۱۳۸۸). رتبه بندی داخلی مشتریان بانک ها با استفاده از مدل های رگرسیونی لاجیت، پژوهشنامه اقتصادی، ویژه‌نامه بانک(۶)،۶۱ـ۸۹.

تقوی، م.، لطفی، ع. ا.، سهرابی، ع. (۱۳۸۷). مدل ریسک اعتباری و رتبه‌بندی مشتریان حقوقی بانک کشاورزی. *پژوهشنامهٔ اقتصادی*، ۴، ۹۹ـ۱۲۸.

تقوی‌فرد، م. ت.، نادعلی، ا. (۱۳۹۱). طبقه‌بندی متقاضیان تسهیلات اعتباری بانکی با استفاده از داده‌کاوی و منطق فازی. *مطالعات مدیریت صنعتی*، ۲۵، ۸۵ـ۱۰۷.

تهرانی، ر.، و فلاح شمس، م. (۱۳۸۴). طراحی و تبیین الگوی ریسک اعتباری در نظام بانکی کشور. *مجلهٔ علوم اجتماعی و انسانی*، ۲۲(۲)، ۶۰ـ۴۵.

تهرانی، ر.، محمدی، م.، و رحیمی، ا. م. (۱۳۸۸). نظام سنجش اعتبار و جایگاه آن در بهبود نظام تأمین مالی. اولین کنفرانس بین‌المللی توسعهٔ نظام تأمین مالی در ایران، مرکز مطالعات تکنولوژی دانشگاه صنعتی شریف.

حیدرپور، ف.، و کارذبحی، م. (۱۳۸۸). طراحی الگویی جهت اعتبارسنجی مشتریان حقوقی بانک با استفاده از معیار c5. *مجلهٔ مطالعات مالی*، ۲(۲)، ۱۳۵ـ۱۵۴.

خالقی‌فر، ح. (۱۳۹۲). ارائهٔ مدل عوامل مؤثر بر ریسک اعتباری با استفاده از رگرسیون لاجیت در بانک اقتصاد نوین شعبهٔ جزیرهٔ کیش. نخستین کنفرانس ملی توسعهٔ مدیریت پولی و بانکی.

رشیدیان، س.، سلطان‌پناه، ه.، و فارقی، ه. (۱۳۹۰). طبقه‌بندی مشتریان شبکهٔ بانکی بر اساس ریسک اعتباری با استفاده از مدل‌های پیش‌بینی و تصمیم‌گیری چندمعیاره (مطالعهٔ موردی: بانک کارآفرین) (پایان‌نامهٔ کارشناسی ارشد). دانشکدهٔ علوم‌انسانی، دانشگاه آزاد اسلامی، تهران.

زمانی، ش.، و ودیعی، م. ح. (۱۳۹۱)، پیش‌بینی ریسک اعتباری مشتریان بانکی با استفاده از مدل شبکهٔ عصبی. کنفرانس ملی حسابداری، مدیریت مالی و سرمایه‌گذاری.

سفیدگران، س.، و حاله، ح. (۱۳۹۵). شناسایی و رتبه‌بندی مشتریان حقوقی متقاضی تسهیلات بانکی از منظر ریسک اعتباری با استفاده از روش ترکیبی ANP-PROMETHEE. *مدیریت توسعه و تحول، ویژه‌نامه*، ۱۹۰ـ۱۷۹.

سهرابی، ب.، رئیسی وانانی، ا.، و زارع مبارک‌آباد، ف. (۱۳۹۵). طراحی سیستم توصیه‌گر به‌منظور بهینه‌سازی و مدیریت تسهیلات بانکی بر مبنای الگوریتم‌های خوشه‌بندی و طبقه‌بندی تسهیلات. *پژوهش‌های نوین در تصمیم‌گیری*، ۱(۲)، ۵۳ـ۷۶.

شریعت‌پناهی، م.، و هاشمی بردکاهی، س. (۱۳۸۸). ارائهٔ مدلی برای اعتبارسنجی مشتریان در بانک صنعت و معدن. *فصلنامهٔ مطالعات حسابداری*، ۶(۲۱)، ۶۱ـ۸۲.

شورورزی، م. ر.، مسیح‌آبادی، ا.، و غیاثی شهرکی، م. (۱۳۹۱). ارزیابی ریسک اعتباری مشتریان حقوقی بانک صادرات بر مبنای دو مدل شبکه‌های عصبی مصنوعی و لوجیت و مقایسهٔ آن‌ها. کنفرانس ملی حسابداری، مدیریت مالی و سرمایه‌گذاری.

صفری، س.، ابراهیمی شقاقی، م.، و شیخ، م. ج. (۱۳۸۹). مدیریت ریسک اعتباری مشتریان حقوقی در بانک‌های تجاری با رویکرد تحلیل پوششی داده‌ها (رتبه‌بندی اعتباری). *پژوهش‌های مدیریت در ایران*، ۱۴(۴)، ۱۶۴ـ۱۳۷.

فیضی، ع.، و هاشمی، م. ع. (۱۳۹۲). ارائهٔ مدلی جهت ارزیابی ریسک اعتباری بانک‌ها (مطالعهٔ موردی: بانک‌های پذیرفته‌شده در بورس اوراق بهادار تهران). نخستین کنفرانس ملی توسعهٔ مدیریت پولی و بانکی.

عرب مازار، ع.، و روئین‌تن، پ. (۱۳۸۵)، عوامل مؤثر بر ریسک اعتباری مشتریان بانکی (مطالعهٔ موردی: بانک کشاورزی). *دوفصلنامهٔ جستارهای اقتصادی*، ۳(۶)، ۸۲ـ۴۶.

غلامیان، م. ر.، سادات رسول، م.، و حاجی محمدی، ز. (۱۳۹۱). استخراج قواعد انجمنی رتبه‌بندی اعتبار با استفاده از الگوریتم ژنتیک: موردکاوی بانک ایرانی. سومین همایش ملی مهندسی صنایع و سیستم.

کاظمی، ا.، قاسمی، ج. و زندیه، و.(۱۳۹۰). رتبه بندی اعتباری مشتریان حقیقی بانک‌ها با استفاده از مدل‌های مختلف شبکه‌های عصبی: مطالعه موردی یکی از بانک‌های خصوصی ایران، مطالعات مدیریت صنعتی، ۹(۲۳)، ۱۳۱ـ ۱۶۱.

کیمیاگری، م. ع.، امینی، م. ج.، طباطبایی آرانی، ح.، حسینی، ل. (۱۳۹۱). مدل ریسک اعتباری بازپرداخت تسهیلات مشتریان اعتباری بانک (مطالعهٔ موردی: حوزهٔ شهرضا بانک ملی ایران)، نهمین کنفرانس مهندسی صنایع.

محمدخان، م.، اسماعیلی، م. ا.، و یاراحمدی، م. (۱۳۸۷). طراحی مدل ارزیابی ریسک اعتباری مشتریان بانک با استفاده از مدل رگرسیون لجستیک. ششمین کنفرانس بین‌المللی مهندسی صنایع.

محمدی، ت.، شاکری، ع.، اسکندری، ف.، و کریمی، د. (۱۳۹۶). عوامل مؤثر بر شکل‌گیری مطالبات غیرجاری در نظام بانکی کشور: مطالعهٔ موردی. *مجلس و راهبرد*، ۲۴(۸۹)، ۲۶۹ـ ۳۰۰.

مروج، م. (۱۳۸۳). افزودن قابلیت داده‌کاوی فازی به بانک‌های اطلاعاتی رابطه‌ای (پایان‌نامهٔ کارشناسی ارشد). دانشگاه صنعتی امیرکبیر، تهران.

مقدم، خ.، ارشدی، ع.، و رضایی، ک. (۱۳۹۲). شناسایی و تعیین عوامل مؤثر بر رفتار اعتباری متقاضیان حقیقی تسهیلات خرد در شهر تهران با نگرش بر کاهش نرخ رشد مطالبات در یکی از بانک‌های خصوصی. *پژوهش‌های پولی و بانکی*، ۶(۱۶)،۲۷ـ ۵۸.

مکیان، ن.، المدرسی، م. ت.، و کریمی تکلو، س. (۱۳۸۹). مقایسهٔ مدل شبکه‌های عصبی مصنوعی با روش‌های رگرسیون لجستیک و تحلیل ممیزی در پیش‌بینی ورشکستگی شرکت‌ها. *فصلنامهٔ پژوهش‌های اقتصادی*، ۱۰(۲)، ۱۶۱ـ ۱۴۱.

میرزایی، ح.، نظریان، ر.، و باقری، ر. (۱۳۹۰). بررسی عوامل مؤثر بر ریسک اعتباری اشخاص حقوقی بانک‌ها (مطالعهٔ موردی شعب بانک ملی ایران، شهر تهران). *فصلنامهٔ روند پژوهش‌های اقتصادی*، ۱۹(۵۸)، ۹۸ـ ۶۷.

میرعمادی، ع. ا. (۱۳۹۲). برآورد ریسک اعتباری مشتریان مؤسسات اعتباری با استفاده از متغیرهای حسابداری. نخستین کنفرانس ملی توسعهٔ مدیریت پولی و بانکی.

مهرآرا، م.، موسایی، م.، تصوری، م.، و حسن‌زاده، آ. (۱۳۸۸). رتبه‌بندی اعتباری مشتریان حقوقی بانک پارسیان. *مدل‌سازی اقتصادی*، ۳، پیاپی ۹، ۱۵۰ـ ۱۲۱.

هزار، ر.، اسکندری، ح. ر.، و اشرفی، م. (۱۳۸۸). رتبه‌بندی اعتباری مشتریان با به‌کارگیری شبکهٔ عصبی در صنعت بانکداری (مطالعهٔ موردی بانک اقتصاد نوین). سومین کنفرانس جهانی بانکداری الکترونیکی.

Basel Committee on Banking Supervision. (2000). *Credit rating and complementary sources of credit quality information* (Working Paper No.3).

Chapman, et al. (1999) CRISP-DM 1.0: Step-by-step data mining guide, retrieved from: http://www.crisp-dm.org/CRISPwP-0800.pdf.

Doumpos, M., & Zopounidis, C. (2014). Evaluation and decision models with multiple criteria. *International Handbooks on Information Systems*, 563-587.

Edward, F. R., & Mishkin, F. S. (1995). The decline of traditional banking: implication for financial stability and regulatory policy. *Federal Reserve Bank of New York policy Review*, 27-45.

Fisher, R. A. (1936). The use-of multiple measurement in Taxonomic problem, *Annals of Eugenics*.

Chi, Q., & Li, W. (2017). Economic policy uncertainty, credit risks and banks' lending decisions: Evidence from Chinese commercial banks. *China Journal of Accounting Research*, 10, 33-50.

Chye, K. H., Tan, W., & Goh, C. (2006). A Two-step Method to Construct Credit Scoring Models with Data Mining Techniques, International Journal of Business and Information, Volume 1 Number 1, 96-118.

Kotu, V., & Deshpande, B. (2014). Predictive Analytics and Data Mining Concepts and Practice with RapidMiner, MK publications, www.mkp.com.

Mongi, L. (2018). Comparative study on Credit Risk in Islamic Banking, Quarterly Review of Economics and Finance.

Nazari, M., & Alidadi, M. (2013). Measuring Credit Risk of BankCustomers Using Artificial Neural Network, Journal of Management Research, Vol5, No 2, PP17-27.

Olivas, R. (2007). Decision trees: A primer for decision-making professionals. 1-20.

Singh, S., Murthi, B. P. S., & Steffes, E. (2013). Developing a measure of risk adjusted revenue (RAR) in credit cards market: Implications forcustomer relationship management. European Journal of Operational Research, 224(2), pp 425-434

Sinkey F. Joseph (1992). Commercial Bank Financial Management in the Financial-services Industry, 4th Edition; Macmillan, p.519.

Thomas, L. C., Eldman, D. B., & Crook, J. N. (2002). Credit Scoringand Its Applications. Philadelphia: Siam.

Yu, L., Wang, S., Lai, K. (2008). Credit risk assessment with a multistage neural network ensemble learning approach. Expert Systems with Applications.No 34, pp 1434– 1444.

پیوست

جدول پیوست ۱

منابع گزینش شاخصه‌های اساسی شکل‌دهندهٔ رفتار مشتریان

نام متغیر	منابع
اسناد دریافتنی	عرب مازار و روئین‌تن (۱۳۸۵)؛ میرزایی، نظریان، و باقری (۱۳۹۰)؛ خالقی‌فر (۱۳۹۲)
بخش اقتصادی (هدف اعتبار)	تهرانی، محمدی، و رحیمی (۱۳۸۸)؛ تقوی، لطفی، و سهرابی (۱۳۸۷)؛ میرزایی، نظریان، و باقری (۱۳۹۰)؛ البرزی و همکاران (۱۳۹۲)؛ تهرانی و فلاح شمش (۱۳۸۴)؛ غلامیان، سادات رسول، و حاجی‌محمدی (۱۳۹۱)، سهرابی و همکاران (۱۳۹۵)، چی و لی (۲۰۱۷)
بازده فروش	شورورزی، مسیح آبادی، و غیاثی شهرکی (۱۳۹۱)؛ زمانی و ودیعی (۱۳۹۱)
(سود خالص به دارایی‌ها) بازده کل دارایی‌ها	زمانی و ودیعی (۱۳۹۱)؛ شورورزی و همکاران (۱۳۹۱)؛ عرب مازار و روئین‌تن (۱۳۸۵)؛ محمدخان و همکاران (۱۳۸۷)؛ فیضی و هاشمی (۱۳۹۲)، مونگی (۲۰۱۸)
بدهکاران	عرب مازار و روئین‌تن (۱۳۸۵)؛ میرزایی و همکاران (۱۳۹۰)؛ خالقی‌فر (۱۳۹۲)
بدهی به بانک‌ها	میرزایی و همکاران (۱۳۹۰)؛ خالقی‌فر (۱۳۹۲)، سهرابی و همکاران (۱۳۹۵)
بدهی جاری	میرزایی و همکاران (۱۳۹۰)؛ خالقی‌فر (۱۳۹۲)؛ غلامیان و همکاران (۱۳۹۱)، تقوی فرد و نادعلی(۱۳۹۱)
بستانکاران (حساب‌های پرداختنی)	میرزایی، نظریان، و باقری (۱۳۹۰)؛ خالقی‌فر (۱۳۹۲)
تعداد اقساط	تقوی، لطفی، و سهرابی (۱۳۸۷)؛ میرزایی و همکاران (۱۳۹۰)؛ البرزی و همکاران (۱۳۹۲) (بخشی از شاخص‌ها)؛ خالقی‌فر (۱۳۹۲)؛ کیمیاگری و همکاران (۱۳۹۱)
تعداد چک برگشتی	میرزایی و همکاران (۱۳۹۰)؛ کیمیاگری و همکاران (۱۳۹۱)؛ خالقی‌فر (۱۳۹۲)
داشتن ضامن معتبر	میرزایی و همکاران (۱۳۹۰)؛ کیمیاگری و همکاران (۱۳۹۱)
گزارش حسابرسی	میرزایی و همکاران (۱۳۹۰)؛ شورورزی و همکاران (۱۳۹۱)؛ حیدرپور و کارذبحی (۱۳۸۸)
دارایی ثابت	میرزایی و همکاران (۱۳۹۰)؛ خالقی‌فر (۱۳۹۲)؛ عرب مازار و روئین‌تن (۱۳۸۵)؛ غلامیان و همکاران (۱۳۹۱)

نام متغیر	منابع
دارایی جاری	میرزایی، نظریان، و باقری (۱۳۹۰)؛ خالقی‌فر (۱۳۹۲)؛ عرب مازار و روئین‌تن (۱۳۸۵)؛ غلامیان و همکاران (۱۳۹۱)
دورهٔ زمانی بازپرداخت تسهیلات	تقوی، لطفی، و سهرابی (۱۳۸۷)؛ میرزایی، نظریان، و باقری (۱۳۹۰)؛ تهرانی و فلاح شمش (۱۳۸۴)؛ تهرانی و همکاران (۱۳۸۸)؛ هزار، اسکندری، و اشرفی (۱۳۸۸)
سابقهٔ چک برگشتی	زمانی و ودیعی (۱۳۹۱)؛ شورورزی و همکاران (۱۳۹۱)؛ تقوی، لطفی، و سهرابی (۱۳۸۷)؛ صابری؛ اکرامی، و راهنما اسکی (۱۳۸۸)
سابقهٔ فعالیت شرکت	تقوی و همکاران (۱۳۸۷)؛ میرزایی و همکاران (۱۳۹۰)؛ کیمیاگری و همکاران (۱۳۹۱)؛ خالقی‌فر (۱۳۹۲)؛ غلامیان و همکاران (۱۳۹۱)؛ شورورزی و همکاران (۱۳۹۱)؛ تهرانی و همکاران (۱۳۸۸)؛ یو و همکاران (۲۰۰۸)
سابقهٔ همکاری با بانک	حیدرپور و کارذبحی (۱۳۸۸)؛ تقوی و همکاران (۱۳۸۷)؛ میرزایی و همکاران (۱۳۹۰)؛ خالقی‌فر (۱۳۹۲)؛ کیمیاگری و همکاران (۱۳۹۱)؛ غلامیان و همکاران (۱۳۹۱)
سرمایه	صابری و خالقی‌فر (۱۳۹۲)؛ هزار، اسکندری، و اشرفی (۱۳۸۸)؛ غلامیان و همکاران (۱۳۹۱)، مونگی (۲۰۱۸)
سرمایهٔ در گردش به کل دارایی‌ها	شریعت پناهی و همکاران (۱۳۸۸)؛ تهرانی و فلاح شمش (۱۳۸۴)؛ فیضی و هاشمی (۱۳۹۲)؛ پرویزیان، ذکاوت، و محمدیان (۱۳۸۸)
گردش بستانکار شش‌ماههٔ حساب‌های وام‌گیرنده	کیمیاگری و همکاران (۱۳۹۱)؛ میرزایی و همکاران (۱۳۹۰)؛ عرب مازار و روئین‌تن (۱۳۸۵)، چی و لی (۲۰۱۷)
سود انباشته به کل دارایی	تهرانی و فلاح شمش (۱۳۸۴)، مونگی (۲۰۱۸)
نرخ تسهیلات	هزار و همکاران (۱۳۸۸)؛ تقوی و همکاران (۱۳۸۷)؛ تهرانی و فلاح شمش (۱۳۸۴)؛ کیمیاگری و همکاران (۱۳۹۱)
متوسط گردش موجودی کالا (فروش به موجودی کالا)	عرب مازار و روئین‌تن (۱۳۸۵)؛ میرزایی و همکاران (۱۳۹۰)
مدت‌زمان تنفس	میرزایی و همکاران (۱۳۹۰)؛ کیمیاگری و همکاران (۱۳۹۱)؛
موجودی کالا	عرب مازار و روئین‌تن (۱۳۸۵)؛ خالقی‌فر (۱۳۹۲)؛ غلامیان و همکاران (۱۳۹۱)
میانگین ماندهٔ شش‌ماههٔ حساب‌های وام‌گیرنده	میرزایی و همکاران (۱۳۹۰)؛ کیمیاگری و همکاران (۱۳۹۱)؛ شورورزی و همکاران (۱۳۹۱)؛ تقوی، لطفی، و سهرابی (۱۳۸۷)
نسبت ارزش وثیقهٔ ارائه‌شده توسط متقاضی به ارزش تسهیلات	کیمیاگری و همکاران (۱۳۹۱)؛ خالقی‌فر (۱۳۹۲)

نام متغیر	منابع
نسبت آنی	عرب مازار و روئین‌تن (۱۳۸۵)؛ میرزایی و همکاران (۱۳۹۰)؛ هزار و همکاران (۱۳۸۸)؛ خالقی‌فر (۱۳۹۲)؛ شورورزی و همکاران (۱۳۹۱)
نسبت بدهی	حیدرپور و کارذبحی (۱۳۸۸)، تقوی فرد و نادعلی(۱۳۹۱)
وجه نقد	عرب مازار و روئین‌تن (۱۳۸۵)؛ میرزایی و همکاران (۱۳۹۰)؛ خالقی‌فر (۱۳۹۲)
نسبت بدهی جاری به ارزش دفتری سرمایه	میرعمادی (۱۳۹۲)؛ هزار و همکاران (۱۳۸۸)؛ یو و همکاران (۲۰۰۸)
نسبت بدهی جاری به ارزش ویژه	عرب مازار و روئین‌تن (۱۳۸۵)؛ یو و همکاران (۲۰۰۸)
نسبت بدهی جاری به کل دارایی	شریعت پناهی و همکاران (۱۳۸۸)؛ میرعمادی (۱۳۹۲)؛ یو و همکاران (۲۰۰۸)
نسبت جاری	عرب مازار و روئین‌تن (۱۳۸۵)؛ میرزایی، نظریان، و باقری (۱۳۹۰)؛ تهرانی و فلاح شمش (۱۳۸۴)؛ حیدرپور و کارذبحی (۱۳۸۸)؛ مکیان، المدرسی، و کریمی تکلو (۱۳۸۹)؛ زمانی و ودیعی (۱۳۹۱)؛ میرعمادی (۱۳۹۲)؛ خالقی‌فر (۱۳۹۲)؛ رشیدیان، سلطان‌پناه، و فارقی (۱۳۹۰)؛ شورورزی، مسیح‌آبادی، غیاثی شهرکی (۱۳۹۱)؛ یو و همکاران (۲۰۰۸)
نوع شرکت	تقوی، لطفی، و سهرابی (۱۳۸۷)؛ هزار، اسکندری، و اشرفی (۱۳۸۸)؛ غلامیان (۱۳۹۱)
نوع عقد	تقوی، لطفی، و سهرابی (۱۳۸۷)؛ البرزی، محمدپورزرندی، و خان‌بابایی (۱۳۹۲) (بخشی از شاخص‌ها)؛ کیمیاگری و همکاران (۱۳۹۱)؛ خالقی‌فر (۱۳۹۲)؛ میرزایی، نظریان، و باقری (۱۳۹۰)
نوع فعالیت شرکت	خالقی‌فر (۱۳۹۲)؛ شورورزی، مسیح‌آبادی، و غیاثی شهرکی (۱۳۹۱)؛ کیمیاگری و همکاران (۱۳۹۱)؛ عرب مازار و روئین‌تن (۱۳۸۵)
نوع وثایق	هزار، اسکندری، و اشرفی (۱۳۸۷)؛ تهرانی و همکاران (۱۳۸۸)؛ البرزی، محمدپورزرندی، و خان‌بابایی (۱۳۹۲) (بخشی از شاخص‌ها)؛ اکرامی و راهنما اسکی (۱۳۸۸)؛ شورورزی، مسیح‌آبادی، و غیاثی شهرکی (۱۳۹۱)؛ زمانی و ودیعی (۱۳۹۱)؛ کیمیاگری و همکاران (۱۳۹۱)
نسبت حاشیهٔ سود (سود خالص به فروش خالص)	میرزایی، نظریان، و باقری (۱۳۹۰)؛ خالقی‌فر (۱۳۹۲)؛
نسبت دارایی جاری (دارایی جاری به کل دارایی)	عرب مازار و روئین‌تن (۱۳۸۵)؛ میرزایی و همکاران (۱۳۹۰)؛ هزار و همکاران (۱۳۹۱)؛ شورورزی، مسیح‌آبادی، و غیاثی شهرکی (۱۳۹۱)؛ خالقی‌فر (۱۳۹۲)؛ یو و همکاران (۲۰۰۸)

نام متغیر	منابع
نسبت سرمایهٔ در گردش به کل دارایی‌ها	مکیان، المدرسی، و کریمی تکلو (۱۳۸۹)؛ میرعمادی (۱۳۹۲)
نوع تسهیلات برحسب دورهٔ زمانی بازپرداخت	کیمیاگری و همکاران (۱۳۹۱)؛ غلامیان و همکاران (۱۳۹۱)
نوع تضمین	عرب مازار و روئین‌تن (۱۳۸۵)؛ میرزایی و همکاران (۱۳۹۰)؛ خالقی‌فر (۱۳۹۲)؛ تقوی، لطفی، و سهرابی
نسبت سود قبل از کسر بهره و مالیات به کل دارایی‌ها	آلتمن؛ میرعمادی (۱۳۹۲)؛ مکیان، المدرسی، و کریمی تکلو (۱۳۸۹)، مونگی (۲۰۱۸)
نسبت سودآوری (سود به سرمایه)	هزار و همکاران (۱۳۸۷)؛ تهرانی و فلاح شمش (۱۳۸۴)؛ میرعمادی (۱۳۹۲)، مونگی (۲۰۱۸)
نسبت سود انباشته به کل دارایی‌ها	میرعمادی (۱۳۹۲)؛ فیضی و هاشمی (۱۳۹۲)
نسبت سود پیش از کسر مالیات و بهره به خالص فروش	میرعمادی (۱۳۹۲)؛ مکیان و همکاران (۱۳۸۹)؛ یو و همکاران (۲۰۰۸)
نسبت کل بدهی	زمانی و ودیعی (۱۳۹۱)؛ شورورزی وهمکاران (۱۳۹۱)
نسبت گردش سرمایهٔ جاری (فروش به سرمایه در گردش)	عرب مازار و روئین‌تن (۱۳۸۵)؛ میرزایی و همکاران (۱۳۹۰)؛ خالقی‌فر (۱۳۹۲)؛ شورورزی و همکاران (۱۳۹۱)؛ یو و همکاران (۲۰۰۸)
نسبت گردش کل دارایی‌ها (فروش به کل دارایی‌ها)	عرب مازار و روئین‌تن (۱۳۸۵)؛ میرزایی و همکاران (۱۳۹۰)؛ خالقی‌فر (۱۳۹۲)؛ آلتمن؛ پرویزیان و همکاران (۱۳۸۸)
نسبت مالکانه (ارزش ویژه به کل دارایی)	عرب مازار و روئین‌تن (۱۳۸۵)؛ میرزایی، نظریان، و باقری (۱۳۹۰) شورورزی و همکاران (۱۳۹۱)؛ خالقی‌فر (۱۳۹۲)؛ زمانی و ودیعی (۱۳۹۱)
نسبت نقدی (وجه نقد به دارایی)	عرب مازار و روئین‌تن (۱۳۸۵)؛ میرزایی و همکاران (۱۳۹۰)
تعداد حساب‌های بانک	میرزایی و همکاران (۱۳۹۰)؛ خالقی‌فر (۱۳۹۲)
نرخ بازده سرمایه	شورورزی و همکاران (۱۳۹۱)
سرمایه به دارایی	میرعمادی (۱۳۹۲)؛ فیضی و هاشمی (۱۳۹۲)
تاریخ اولین سررسید	البرزی و همکاران (۱۳۹۲) (بخشی از شاخص‌ها)
تاریخ آخرین سررسید	البرزی و همکاران (۱۳۹۲) (بخشی از شاخص‌ها)
تاریخ پرداخت تسهیلات	البرزی و همکاران (۱۳۹۲) (بخشی از شاخص‌ها)
اندازهٔ بنگاه	پرویزیان و همکاران (۱۳۸۸)
سهم بانک	مونگی (۲۰۱۸)، چی و لی (۲۰۱۷)
کل بدهی‌ها	غلامیان و همکاران (۱۳۹۱)

نام متغیر	منابع
کل دارایی‌ها	غلامیان و همکاران (۱۳۹۱)، مونگی (۲۰۱۸)
جمع گردش بدهکار حساب‌های شش‌ماههٔ شرکت	عرب مازار و روئین‌تن (۱۳۸۵)
حساب‌های وام‌گیرنده میانگین ماندهٔ ۶شش‌ماهه	چی و لی (۲۰۱۷)
قدمت شرکت	هزار و همکاران (۱۳۸۷)
ماندهٔ حساب به مبلغ تسهیلات	اکرامی و راهنما اسکی (۱۳۸۸)، چی و لی (۲۰۱۷)
مبلغ آخرین تسهیلات دریافتی	خان‌بابایی و همکاران (۱۳۸۹) (بخشی از شاخص‌ها)؛ عرب مازار و روئین‌تن (۱۳۸۵)
مبلغ گردش حساب‌های شش‌ماههٔ شرکت	عرب مازار و روئین‌تن (۱۳۸۵)
مبلغ وثیقه	تهرانی و همکاران (۱۳۸۸)
معدل موجودی	اکرامی و راهنما اسکی (۱۳۸۸)
تسهیلات نزد بانک‌های دیگر	چی و لی (۲۰۱۷)
سابقهٔ اعتباری (جاری یا غیرجاری) در بانک یا نظام بانکی	تهرانی و همکاران؛ هزار، اسکندری و همکاران (۱۳۸۵)؛ مونگی (۲۰۱۸) اکرامی و راهنما اسکی (۱۳۸۸)، چی و لی (۲۰۱۷)
سال اخذ آخرین تسهیلات	تقوی و همکاران (۱۳۸۷)
زمینهٔ تولیدی متقاضی تسهیلات	اکرامی و راهنما اسکی (۱۳۸۸)
سوابق مدیران ارشد	غلامیان و همکاران (۱۳۹۱)
سرمایه به بدهی	مکیان و همکاران (۱۳۸۹) مونگی (۲۰۱۸)
بدهی به دارایی	پرویزیان و همکاران (۱۳۸۸)؛ تهرانی و فلاح شمش (۱۳۸۴)، مونگی (۲۰۱۸)
فروش خالص به سرمایهٔ در گردش و برعکس	پرویزیان و همکاران (۱۳۸۸)
سرمایه به بدهی بلندمدت و برعکس	میرعمادی (۱۳۹۲)؛ حیدرپور و کارذبحی (۱۳۸۸)، مونگی (۲۰۱۸)
سرمایه به دارایی و برعکس	پرویزیان و همکاران (۱۳۸۸)
بدهی به سرمایه	شریعت پناهی و هاشمی بردکاهی (۱۳۸۸)
موجودی نقدی به دارایی	پرویزیان و همکاران (۱۳۸۸)
موجودی نقدی به بدهی جاری	پرویزیان و همکاران (۱۳۸۸)؛ فیضی و هاشمی (۱۳۹۲)
وضعیت حساب پس‌انداز و جاری	تهرانی و همکاران (۱۳۸۸)، سهرابی و همکاران (۱۳۹۵)

جدول پیوست ۲

شاخص‌های منتخب و میزان دسترسی به داده‌های آن در ۵۲۱ پرونده اعتباری

شاخص	تعداد داده‌های در دسترس	شاخص	تعداد داده‌های در دسترس
فروش به موجودی کالا	۲۰۵	دارای طرح توجیهی است؟	۵۱۹
سود به فروش	۲۱۰	کد نوع عقد فنّاوری	۵۲۰
سود خالص به فروش	۲۲۶	نرخ سود	۵۲۰
دارایی جاری به فروش	۲۳۱	دورهٔ تنفس	۵۲۰
هزینه به درآمد	۲۴۳	مدت قرارداد	۵۲۰
سرمایه به بدهی بلندمدت	۲۵۵	فاصلهٔ اقساط	۵۲۰
شغل ضامن ۳	۳۲۲	تعداد اقساط	۵۲۰
سود خالص به سرمایه	۳۵۰	اسناد تجاری (سفته)	۵۲۰
فروش به دارایی ثابت	۳۵۸	محل اجرای طرح	۵۲۰
سود خالص به دارایی ثابت	۳۶۳	وثیقهٔ ملکی خارج از طرح	۵۲۰
درآمد به دارایی	۳۷۱	سایر	۵۲۰
سود عملیاتی به دارایی	۳۷۶	جمع تضمینات	۵۲۰
فروش به دارایی	۳۷۹	تعداد جاری	۵۲۰
موجودی نقد به بدهی	۳۸۱	تعداد کوتاه‌مدت	۵۲۰
درآمد کل	۳۸۳	تعداد بلندمدت	۵۲۰
سود خالص به دارایی	۳۸۴	غیرجاری	۵۲۰
شغل ضامن ۲	۳۸۷	نکول بر مبنای مبلغ وام	۵۲۰
سود عملیاتی	۳۸۸	کد نکول تفصیلی	۵۲۰
فروش	۳۹۱	تعداد امهال	۵۲۰
سرمایه به بدهی	۳۹۴	مبلغ امهال	۵۲۰
بدهی جاری به سرمایه	۴۰۳	ماندهٔ سپردهٔ جاری زمان دریافت تسهیلات	۵۲۰
شغل ضامن ۱	۴۰۹	تعداد پروندهٔ مانده‌دار در نظام بانکی	۵۲۰
سرمایه به دارایی ثابت	۴۱۸	تعداد پرونده در بانک توسعه	۵۲۰
بدهی جاری به دارایی	۴۴۰	جمع مبالغ تسهیلات دریافتی در بانک توسعه	۵۲۰
سرمایه به دارایی	۴۴۵	تعداد کل چک‌های تحویل‌شده در بانک توسعه	۵۲۰
میانگین قرض‌الحسنهٔ سه ماه قبل	۴۴۶	تعداد کل چک‌های برگشتی قبل از دریافت وام در بانک توسعه	۵۲۰
میانگین قرض‌الحسنهٔ شش ماه قبل	۴۴۶	مبلغ کل چک های برگشتی قبل از دریافت وام در بانک توسعه	۵۲۰
مبلغ چک اخذشده از شرکت	۴۶۱	تاریخ شروع کار	۵۲۰
موجودی نقد به دارایی	۴۶۴	اصل و سود محاسباتی	۵۲۰
دارایی جاری به دارایی	۴۶۸	مبلغ وام	۵۲۰

دارایی کل	۴۸۲	جنسیت مدیرعامل	۵۲۰
مدرک تحصیلی مدیرعامل	۴۹۵	مدیرعامل سابقهٔ کار مرتبط دارد یا خیر	۵۲۰
آیا صورت مالی شرکت حسابرسی‌شده است؟	۵۱۸	نوع عقد	۵۲۰
سن مدیرعامل	۵۱۹	محل تصویب پرونده	۵۲۰
آیا شرکت خانوادگی است؟	۵۱۹	علت پرداخت تسهیلات	۵۲۰
وضعیت تأهل مدیر عامل	۵۱۹	نوع اقساط	۵۲۰
آیا مدیرعامل شرکت از بانک توسعهٔ تعاون تسهیلات اخذ کرده است؟	۵۱۹	نوع شرکت	۵۲۰
آیا مدیرعامل شرکت در بانک توسعهٔ تعاون ماندهٔ بدهی غیرجاری بلی؟	۵۱۹	بخش اقتصادی	۵۲۰

جدول پیوست ۳

تعداد داده‌های پرت

شاخص	دادهٔ پرت	شاخص	دادهٔ پرت
دورهٔ تنفس	۱۷	تعداد اقساط	۳
نرخ سود	۱۴	تعداد کوتاه‌مدت	۳
تعداد امهال	۱۳	میانگین قرض‌الحسنهٔ سه ماه قبل	۲
تعداد پرونده در بانک توسعه	۱۲	اوراق بهادار	۲
فاصلهٔ اقساط	۱۱	بدهی جاری به سرمایه	۲
ماندهٔ سپردهٔ جاری زمان دریافت تسهیلات	۱۰	سود خالص به فروش	۲
جمع مبالغ تسهیلات دریافتی در بانک توسعه	۱۰	اصل و سود محاسباتی	۲
تعداد کل چک‌های تحویل‌شده در بانک توسعه	۱۰	مدت قرارداد	۱
سایر تضمینات	۹	تعداد پروندهٔ مانده‌دار در نظام بانکی	۱
غیرجاری	۸	فروش	۱
مبلغ سهم آوردهٔ متقاضی به مبلغ وام	۸	درآمد کل	۱
تعداد ضامنین	۸	موجودی نقد به دارایی	۱
محل اجرای طرح	۷	دارایی جاری به دارایی	۱
وثیقهٔ ملکی خارج از طرح	۷	دارایی جاری به فروش	۱
تعداد چک برگشتی در نظام بانکی	۷	بدهی جاری به دارایی	۱
تعداد حساب جاری	۶	فروش به موجودی کالا	۱
مبلغ امهال	۵	هزینه به درآمد	۱
مبلغ چک برگشتی در نظام بانکی	۵	سود عملیاتی به دارایی	۱
تعداد کل چک‌های برگشتی قبل از دریافت وام	۵	سود خالص به دارایی ثابت	۱
میانگین قرض‌الحسنهٔ شش ماه قبل	۴	مبلغ وام	۱
مبلغ سهم آوردهٔ متقاضی	۴	مبلغ چک اخذشده	۱