ارزیابی ریسک اعتباری در سیستم‌های بانکی با استفاده از تکنیک‌های داده‌کاوی

نیما همتا[۱،*]، محمد احسانی‌فر[۲]، بهاره محمدی[۳]

استادیار، گروه مهندسی مکانیک (ساخت و تولید)، دانشگاه صنعتی اراک، اراک، ایران(عهده دار مکاتبات)

استادیار، گروه مهندسی صنایع، واحد اراک، دانشگاه آزاد اسلامی، اراک، ایران

دانشجوی کارشناسی ارشد، گروه مهندسی صنایع، واحد اراک، دانشگاه آزاد اسلامی، اراک، ایران

تاریخ دریافت: بهمن ماه ۱۳۹۶، اصلاحیه: فروردین ماه ۱۳۹۷، پذیرش: اردیبهشت ماه ۱۳۹۷

چکیده

این مقاله با هدف شناسایی عوامل مؤثر بر ریسک اعتباری و ارائه مدلی جهت پیش‌بینی ریسک اعتباری و رتبه‌بندی اعتباری مشتریان حقوقی متقاضی تسهیلات بانک سپه شهرستان دزفول، با استفاده از روش‌های خوشه‌بندی، شبکه عصبی و ماشین بردار پشتیبان انجام گرفته است. در این مقاله، ۲۷ متغیر توضیح‌دهنده شامل متغیرهای مالی و غیرمالی مورد بررسی قرار گرفت که از بین این متغیرها، ۸ متغیر تأثیرگذار بر ریسک اعتباری انتخاب گردید که به وسیله روش خوشه‌بندی مجموعه داده‌ها به خوشه‌ها دسته‌بندی شدند. هم‌چنین متغیرهای انتخابی به عنوان بردار ورودی شبکه عصبی پرسپترون ۳ لایه وارد مدل شد و در نهایت با استفاده از ماشین بردار پشتیبان، به منظور پیش‌بینی عملکرد مالی مشتریان حقوقی بانک ارائه گردید. نتایج حاصل از مدل شبکه عصبی و ماشین بردار پشتیبان حاکی از آن است که مدل شبکه عصبی در پیش‌بینی ریسک اعتباری مشتریان حقوقی و رتبه‌بندی اعتباری از کارایی بیشتری برخودار است.

کلمات کلیدی: رتبه بندی اعتباری، خوشه‌بندی، شبکه عصبی، ماشین بردار پشتیبان، داده‌کاوی.

۱- مقدمه

ارتباط صحیح بین نظام‌های مالی و تولیدی در هر کشوری از مهم‌ترین عوامل رشد و توسعه اقتصادی محسوب خواهد شد. بانک‌ها به عنوان بخش اصلی نظام مالی نقش مهمی را در تأمین مالی بخش‌های تولیدی، تجاری و مصرفی و حتی دولتی به عهده خواهد داشت. در ایران نیز با توجه به ساختار اقتصادی کشور و به دلایلی هم چون عدم توسعه بازار-های سرمایه و سایر شبکه‌های غیر بانکی و قراردادی تأمین مالی بخش-های واقعی اقتصاد بر عهده شبکه بانکی کشور است [۱]. بنابراین بانک‌ها به منظورآگاهی از نیازمندی‌های مشتریان خود، در اعطای تسهیلات اعتباری باید به شناسایی ویژگی‌های آنها بپردازند. این امر از طریق اعتبارسنجی، منجر به کاهش ریسک‌های بانکی از جمله ریسک اعتباری می‌شود. اعتبارسنجی به عملی اطلاق می‌شود که در آن اعتبار مشتریان حقیقی و حقوقی مؤسسات مالی اعتباری و بانک‌ها با توجه به اطلاعات دریافتی از آنها اندازه‌گیری شده و امکان شناخت بیشتر را نسبت به وضعیت و توان مالی افراد جهت بازپرداخت تسهیلات دریافتی و دریافت خدمات بیشتر فراهم می‌کند[۹]. به کارگیری سیستم رتبه‌بندی، شاخص و معیاری برای تنظیم نرخ سود و میزان وثایق فراهم می‌کند تا مشتریان با وضعیت اعتباری مناسب از آنها بهره‌مند شوند. هم‌چنین کاهش مطالبات معوق، توسعه ضریب نفوذ اعتبار، افزایش کارایی و سرعت، بهبود سودآوری بانک و گردآوری اطلاعات آماری، ارتباط بالقوه با سیستم‌های مکانیزه امتیازدهی و فراهم نمودن اطلاعات مشابه، حذف بخش عمده‌ای از تقلب‌ها و حداقل کردن ریسک اعتباری، امکان دسترسی سریع به اطلاعات به‌روز و دسترسی به زیر ساخت‌های اطلاعاتی و مدیریت آسان داده‌ها برای صنایع مرتبط با مدیریت وصول مطالبات و بهبود نسبت بدهی‌های ناوصول از اثرات نظام سنجش و اعتبار است. بنابراین وجود یک سیستم اطلاعاتی متمرکز و مبتنی بر تکنولوژی نوین و با معماری ویژه و ساختار داده استاندارد که دخالت عوامل انسانی را به حداقل می‌رساند،

*nima.hamta@gmail.com

زمینه تنبیه متخلفان و تشویق افراد خوش‌حساب را فراهم خواهد کرد. در واقع مدل‌های رتبه‌بندی اعتباری یکی از مهم‌ترین و اساسی‌ترین سیستم-های این تصمیم‌گیری‌ها هستند که بخش عمده‌ای از اطلاعات مورد نیاز مؤسسه اعتباردهنده در مدیریت اعتبار را فراهم می‌کنند. بانک‌ها و مؤسسات مالی با استفاده از رتبه‌بندی اعتباری با توجه به سوابق متقاضی تسهیلات، احتمال عدم بازپرداخت توسط وی را ارزیابی کرده و امتیاز-بندی می‌کند. به عبارت دیگر رتبه بندی اعتباری مجموعه‌ای از مدل‌های تصمیم‌گیری و روش های مرتبط با آنها است که به اعتباردهندگان در اعطای اعتبار به مشتریان کمک می‌کند. با توجه به مطالعات انجام شده و نتایج به دست آمده در زمینه رتبه‌بندی اعتباری مشتریان، می‌توان گفت که در حال حاضر روش‌های آماری و مبتنی بر محاسبات عددی، به دلیل حجم محاسبات ریاضی بالا و پیچیده، دقت پیش‌بینی پایین و هم-چنین وقت‌گیر بودن آنها، کمتر مورد استفاده قرار می‌گیرند. این در حالی است که روش‌های مبتنی بر هوش مصنوعی و دانش داده‌کاوی از جمله شبکه عصبی، خوشه‌بندی و ماشین بردار پشتیبان به خوبی نقاط ضعف اشاره شده را پوشش می‌دهند و قابلیت انعطاف‌پذیری بسیار بالایی در برخورد با مسأله پیش‌رو است. تنظیم دقت و نامحدود بودن در حجم ورودی‌های مسأله از جمله مزایای برتری آنها نسبت به سایر روش‌ها است [۵].

۲- روش شناسی پژوهش

با توجه به ماهیت تحقیق که استفاده از داده‌کاوی جهت طبقه‌بندی مشتریان بانک سپه می‌باشد، این تحقیق از نوع داده‌محور[1] می‌باشد. پایه اصلی تحقیق حاضر بر کشف دانش از پایگاه داده‌های بانک سپه است. از این‌رو استاندارد جهانی CRISP-DM[2] جهت انجام فرآیند تحقیق مورد استفاده قرار گرفته است، که این مراحل شامل این است که از کار با داده‌ها شروع شده و سعی بر آن است تا مواردی را که قبلاً آگاهی نسبت به آنها وجود نداشته است، کشف کرده و برای آنها قوانینی ساخته شود. بنابراین از روش کشف دانش استفاده شده است. در این تحقیق پس از جمع‌آوری داده‌های مشتریان سابق بانک از پایگاه داده مربوطه و پس از آن، پالایش داده‌ها، به شناسایی متغیرهای تأثیرگذار در رتبه‌بندی مشتریان پرداخته می‌شودکه این کار از طریق مصاحبه با کارشناسان امر و مستندات علمی، انجام گرفته است و بعد از این مرحله برای تمامی مشتریان نمونه نهایی، با توجه به تعریفی که از خوش‌حساب و بدحساب بودن مشتریان وجود دارد، یک برچسب طبقه یا همان تعریف در نظر گرفته می‌شود. در مرحله بعد با استفاده از تکنیک‌های خوشه‌بندی، ماشین بردار پشتیبان و شبکه عصبی و با کمک نرم‌افزارهای مربوطه، مشتریان بر اساس ویژگی‌هایشان طبقه‌بندی می‌گردند و رفتار آنها پیش-بینی می‌شود. در انتها، قوانین و الگوهای موجود در داده‌های کل مشتریان را بر اساس مشتریان تعریف شده پیدا نموده و به عنوان چارچوبی جهت پیش‌بینی اعتبار متقاضیان جدید ارائه می‌گردد تا با قرار گرفتن در طبقات تعریف شده، پیش‌بینی شوند.

۲-۱- جامعه آماری

جامعه آماری این تحقیق، اطلاعات تراکنشی و جمعیت شناختی ۲۰۰ نفر از مشتریان است که به منظور رتبه‌بندی اعتباری مشتریان حقوقی بانک سپه، از اطلاعات مربوط به پرونده‌های این مشتریان که صورت کاملاً تصادفی انتخاب شده‌اند، استفاده می‌گردد. به منظور رتبه‌بندی اعتباری، از تحلیل اطلاعات مربوط به مشتریان حقوقی بانک سپه شهرستان دزفول استفاده می‌شود. این حجم نمونه از مشتریان دارای حساب منتهی به سال ۱۳۹۳تا ۱۳۹۴ انتخاب شده‌اند.

۲-۲- نمونه و نمونه‌گیری

در این پژوهش، مشتریان حقوقی که از بانک سپه شهرستان دزفول در یک بازه زمانی یک ساله تسهیلات اعتباری دریافت نموده و آنها را به بانک عودت داده یا نداده‌اند به عنوان جامعه آماری تعریف می‌گردند. با توجه به این موضوع برای نمونه‌گیری، کلیه مشتریان اعتباری حقیقی و حقوقی شعبه شهری این بانک در بازه زمانی مورد نظر مورد بررسی قرار گرفته و از میان ۴۰۰ مشتری اعتباری، با بکارگیری نمونه‌گیری تصادفی ساده، ۲۰۰ مشتری که در خلال این بازه زمانی وام دریافت کرده بودند، انتخاب شدند. بر این اساس فرمول نمونه‌گیری از جامعه محدود بدون جای‌گذاری برابر است با:

(۱)

$$n = \frac{NZ_{\frac{\alpha}{2}}^{2}\sigma^2 x}{\varepsilon^2(N-1)+Z_{\frac{\alpha}{2}}^{2}\sigma^2 x}$$

N=حجم جامعه ، $\sigma^2 x$=انحراف معیار جامعه ، n=حداقل حجم نمونه، ε=درصدخطای مجاز، $Z_{\frac{\alpha}{2}}^{2}$=سطح اطمینان ۹۵٪

۲-۳- ابزار های پژوهش

[1] Data Oriented
[2] Cross Industry Standard Process for Data Mining

۲-۳-۱- خوشه‌بندی

خوشه‌بندی[3] از مفیدترین کارکردهای داده‌کاوی برای کشف گروه‌ها و تعیین توزیع‌های مورد علاقه و الگوها در داده‌ها است. مسأله خوشه بندی در مورد جداسازی یک مجموعه داده به گروه‌ها (خوشه) به نحوی است که داده‌های موجود در یک خوشه، نسبت به نقاط موجود در خوشه‌های دیگر شباهت بیشتری به یکدیگر داشته باشند. در روش‌های خوشه‌بندی، تحلیل خوشه گسسته را بر پایه فاصله اقلیدسی انجام می‌دهند. این فاصله‌های اقلیدسی از یک یا تعدادی متغیر هسته محاسبه شده است که توسط الگوریتم، تولید و به‌روز می‌شوند. می‌توان معیار خوشه‌بندی را تعیین کرد که برای اندازه‌گیری فاصله بین مشاهدات و هسته‌ها به کار می‌رود. مشاهدات به خوشه‌هایی تقسیم می‌شوند به گونه‌ای که هر مشاهده حداکثر به یک خوشه تعلق داشته باشد. هم‌چنین از مطالعات خوشه‌بندی به یادگیری یا بخش‌بندی کنترل‌نشده تعبیر می‌شود. یادگیری کنترل‌نشده، فرآیند خوشه‌بندی با هدف نامشخص است. هدف تقسیم کردن موارد به کلاس‌های گسسته است که نسبت به ورودی همگن باشند.

روش خوشه‌بندی K-means

مجموعه داده‌هایی با N داده n بعدی x^n را در نظر می‌گیریم، هدف تعیین تقسیم‌بندی طبیعی مجموعه داده‌ای به K خوشه است. می‌دانیم که K خوشه ناپیوسته شامل N_j نقطه داده‌ای با بردار نشان‌گر μ_i وجود دارد که $j = 1,...,K$ الگوریتم K-means تلاش می‌کند تا مجموع مربعات تابع خوشه‌بندی را از رابطه (۲) به حداقل برساند.

$$J = \sum_{j=1}^{k} \sum_{n \in S_j} \left\| x^n - \mu_j \right\|^2 \qquad (۲)$$

که در این رابطه μ_j میانگین نقاط داده‌ای در خوشه S_j است و از رابطه (۳) به دست می‌آید.

$$\mu_j = \frac{1}{N_j} \sum_{n \in S_j} x^n \qquad (۳)$$

این دنباله با تخصیص تصادفی نقاط به K خوشه انجام می‌شود. سپس بردارهای میانگین μ_j از N_j نقطه را در هر خوشه محاسبه می‌کند. برای هر نقطه مجدداً خوشه جدیدی تعیین می‌شود که بر اساس آن بردار نزدیک‌ترین میانگین به‌دست می‌آید. سپس بردارهای میانگین مجدداً محاسبه می‌شوند. خوشه‌بندی K-means به شرح زیر پیش می‌رود:

- تعداد خوشه‌ها را با K نشان داده می‌شود، مشخص می‌کنیم؛

- K دسته اولیه را انتخاب می‌کنیم؛

- مواردی را تعیین می‌کنیم که به عضو j ازدسته j که $j = 1,...,K$ نزدیک‌ترند؛

- میانگین نمونه‌ها در هر خوشه را محاسبه می‌کنیم و مرکز خوشه‌های K را به میانگین خوشه‌هایشان نزدیک می‌کنیم؛

- نزدیک‌ترین موارد به مرکز خوشه جدید j متعلق به خوشه j را مجدداً تخصیص می‌دهیم؛

- میانگین نمونه‌ها را در هر خوشه به عنوان یک مرکز خوشه جدید در نظر می‌گیریم؛ این روش را آنقدر تکرار می‌شود تا در خوشه‌بندی تغییر بیشتری دیده نشود.

خوشه‌بندی K-means یک روش طبقه‌بندی بدون نظارت است. از نظر محاسباتی روشی کارآمد است که دانه‌های خوشه اولیه را فراهم می‌کند که به طرز هوشمندانه‌ای قرار دارند. و روش‌های خوشه‌بندی به معیار فاصله یا تشابه بین نقاط بستگی دارد. استانداردهای فاصله‌ای متفاوتی در خوشه‌بندی به روش K-means استفاده می‌شود که می‌تواند خوشه‌های متفاوتی به وجود بیاورد [۷].

۲-۳-۲- شبکه عصبی

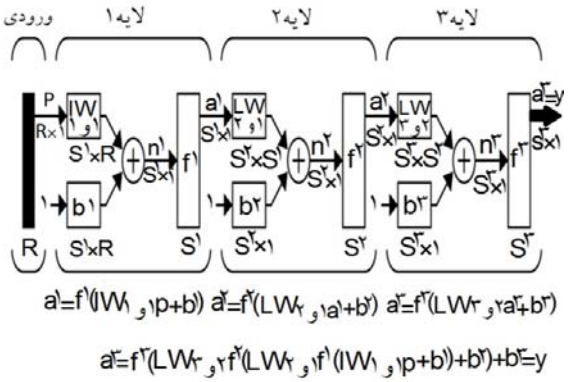
یک سیستم عصبی مصنوعی در حقیقت، فرآیند یادگیری انسانی را شبیه‌سازی می‌کند. سیستم شبکه عصبی مصنوعی[4] با تقلید از سیستم مغزی و عصبی انسان می‌کوشد که ارتباط بین داده‌ها و ستاده‌ها را از راه تکرار نمونه‌برداری از مجموعه داده‌های گذشته یاد گیرد. شبکه عصبی دارای یک برتری اساسی نسبت به سیستم خبره بوده و آن این است که هنگامی که داده‌ها کامل نبوده و یا دارای پارازیت باشند، از راه آموخته‌های گذشته حدس منطقی از داده‌ها می‌سازد. شبکه عصبی بر اساس سه ویژگی داده‌های ورودی، وزن‌ها و لایه‌های پنهان مشخص می‌شود. مدل شبکه عصبی استفاده شده در این تحقیق از نوع پیش‌خورمی‌باشد. هر نرون شبکه پیش‌خور[5] در هر لایه با تمامی نرون‌های لایه پیشین خود در

[3] Clustering

[4] Artificial Neural Network

[5] Feed Forward Neural Network

ارتباط می‌باشد و آن بدین علت است که بردار خروجی در هر لایه، بردار ورودی لایه بعد خود می‌باشد. از مهم‌ترین و پرکاربردترین انواع شبکه‌های پیش‌خور، شبکه‌های عصبی پرسپترون[6] چند لایه است که در این تحقیق بکار گرفته شده است. مدل استفاده شده در این تحقیق یک شبکه عصبی سه لایه است که از بردار ورودی با ۸ متغیر ورودی که عبارتند از: میزان سرمایه ثبتی، مدت وام، نرخ بهره، نوع وثایق، سابقه همکاری، نسبت بدهی، مبلغ مصوب وام استفاده می‌کند. شمای ساختاری شبکه پرسپترون سه لایه مورد استفاده در این مقاله در شکل (۱) نشان داده شده است.



شکل (۱): شمای ساختاری شبکه پرسپترون

P بردار ورودی و $IW_{1,1} IW^{1,1}$ ماتریس وزن در لایه پنهان اول می‌باشد. $n^i n^i$ بردار نرون‌ها، $a^i a^i$ بردار خروجی، $f^i f^i$ تابع انتقال، $b^i b^i$ بردار بایاس در لایه iام است و $LW_{i,j}$ ماتریس وزن در لایه پنهان iام می‌باشد. $a^3 = y$ بردار خروجی است که از رابطه (۴) به‌دست می‌آید:

$$\begin{aligned} &a = p \\ &a^i = f^i(w^t . a^{i-1} + b^i) \\ &i = 1,2,3 \\ &where w^1 = IW_{1,1}, W^2 = LW_{2,1}, W^3 = LW_{3,2} \end{aligned} \qquad (۴)$$

تعداد نرون‌ها در لایه پنهان اول برابر با تعداد متغیرها فرض شده است که این تعداد بهینه‌ترین تعداد نرون در این لایه می‌باشد. تابع انتقال استفاده شده در لایه سوم تابع لگاریتمی سیگوئید $g \succ 0 \ \frac{1}{1+e^{-net.}} where$ است که $\frac{1}{1+e^{-net.}} whereg \succ 0$ در شبکه‌های پس انتشار بسیار مورد استفاده قرار می‌گیرد. این تابع انتقال موجب می‌شود خروجی شبکه عددی بین ۰ تا ۱ باشد که نتیجه مطلوب در این تحقیق است [۱]. قواعد یادگیری، رویه‌ای برای اصلاح وزن‌ها و بایس‌ها[7] است و در راستای آموزش شبکه استفاده قرار می‌گیرد. در ورودی‌ها به صورت مجموعه $\{[p_1,t_1][p_1,t_2][p_q,t_q]\}$ که در آن p_q p_q ها ورودی شبکه و t_q t_q هدف متناظر با هر ورودی است، تعریف می‌گردند. با ورود بردار ورودی به شبکه، خروجی آن با مقدار هدف مقایسه می‌شود. قواعد یادگیری با تنظیم وزن‌ها و بایس‌ها، خروجی شبکه را به مقدار هدف نزدیک می‌کند. از انواع الگوریتم‌های مؤثر جهت آموزش یک شبکه پرسپترون چند لایه، الگوریتم پس انتشار خطا[8] است. الگوریتم پس انتشار بر مبنای تخمین بیشترین کاهش گرادیان عمل می‌کند که شاخص کارایی در این الگوریتم، میانگین مربعات خطا است. شبکه‌های پرسپترون چند لایه که توسط الگوریتم پس انتشار آموزش داده شوند پرکاربردترین نوع از شبکه‌های عصبی می‌باشند. از انواع الگوریتم‌های پس انتشار، الگوریتم لونبرگ- مارکوت[9] که بر گرفته از روش نیوتن بوده و روش بهینه‌ای در آموزش شبکه‌های عصبی می‌باشد، جهت آموزش در این تحقیق مورد استفاده قرار گرفته است. اجرایی کردن مدل شامل خلق، آموزش و آزمایش داده‌ها توسط نرم‌افزار R پیاده‌سازی شده است. الگوریتم، زمانی همگرا است که نرم گرادیان کمتر از مقدار از پیش تعریف شده باشد و یا اینکه خطای مجموع مربعات از حد معینی کوچکتر گردد. در مدل شبکه عصبی مورد نظر در این تحقیق نتیجه، ریسک اعتباری مشتری است که با یک عدد اعشاری بین ۰ تا ۱ مشخص می‌شود. عدد نزدیک به صفر ریسک بالای مشتری را در پرداخت اصل و اقساط تسهیلات و عدد نزدیک به یک کمترین ریسک اعتباری مشتری در پرداخت اقساط را مشخص می‌کند. مهم‌ترین مسأله در استفاده از شبکه‌های عصبی، انتخاب یک الگوی آموزشی مناسب است. لذا لازم است ابتدا داده‌های ورودی به دو مجموعه آموزش و آزمایش تقسیم شوند. داده‌های این تحقیق شامل ۲۰۰ پرونده می‌باشد که ۱۸۰ پرونده جهت آموزش و ۲۰ پرونده جهت آزمایش شبکه مورد نظر قرار گرفته شده است. مجموعه آزمایش برای تنظیم اتصالات بین نرون‌ها و در جهت مینیمم کردن خطا مورد استفاده قرار می‌گیرد. در روند آموزش شبکه و با تکرار

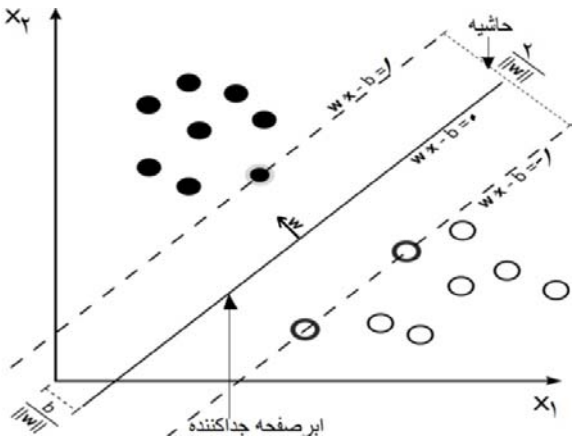
[6] Perceptron Network

[7] Bias
[8] Back Propagation
[9] Levenberg – Marquardt

فرآیند یادگیری، وزن‌ها و بایس‌ها (۲) به گونه‌ای تنظیم می‌شوند که خطای برازش داده‌های آزمایش همگرا به صفر باشد [۱].

۲-۳-۳- ماشین بردار پشتیبان[10]

ماشین بردار پشتیبان در واقع یک طبقه‌بندی‌کننده دودویی است که دو کلاس را با استفاده از یک مرز خطی از هم جدا می‌کند. در تقسیم خطی داده‌ها هدف دست‌یابی به تابعی است که تعیین‌کننده ابرصفحه‌ای با بیشترین حاشیه می‌باشد. با حداکثر شدن این ابرصفحه[11]، تفکیک بین طبقات حداکثر می‌گردد. فرض کنید که $S=\{x_i,y_i\}$ یک نمونه آموزشی است که از دو کلاس $y_i=\pm1$ و هر کلاس از $i=1,...,m$ ویژگی تشکیل شده است. همان‌گونه که در شکل ۲ نشان داده شده است، خط $(\prec w\ .x_i\succ+b)=0$ داده‌های موجود را در دو کلاس ±1 طبقه‌بندی می‌کند. به این خط ابرصفحه جداکننده گفته می‌شود. دو خط $(\prec w\ .x_i\succ+b)=+1$ و $(\prec w\ .x_i\succ+b)=-1$ به ترتیب بیان‌گر مرز ناحیه دسته‌های $y=+1$ و $y=-1$ می‌باشند. به نزدیک‌ترین داده‌های آموزشی به ابرصفحه‌های جداکننده، بردار پشتیبان[12] نامیده می‌شوند [۱۳].



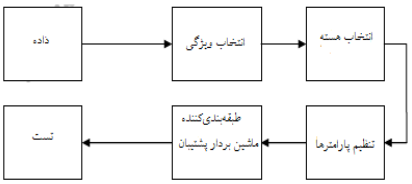
شکل (۲): ابرصفحه‌ای جداکننده دو گروه۱- و۱+

در ماشین بردار پشتیبان به دو طریق خطی و غیرخطی می‌توان مجموعه نقاط را از یکدیگر جدا نمود. از این‌رو در این مقاله از تابع غیرخطی آر بی اف استفاده شده است. که تابع آر بی اف جزء توابع کرنلی[13] است که نسبت که سایر توابع از عملکرد مطلوب‌تری در مسائل طبقه‌بندی اعتباری از خود نشان داده است. از این‌رو مسأله بهینه سازی به صورت زیر تعریف می‌شود:

(۵)

$$Max\sum\alpha_i-\frac{1}{2}\sum_i\sum_j\alpha_i\alpha_j y_i y_j k(x_i x_j)$$
$$s.t\sum_i\alpha_i y_i=0,\alpha_i\geq0$$
$$0\leq\alpha_i\leq C,\forall i$$

اعتبارسنجی در واقع شامل طبقه‌بندی و رتبه‌های (خوب و بد) تخصیص یافته به متقاضیان و شرایط متقاضی از قبیل مقدار وام، موقعیت کاری، اطلاعات شخصی، سن و... به عنوان متغیرهای طبقه‌بندی است. رویه اعتبارسنجی مبتنی بر ماشین بردار پشتیبان شامل جمع‌آوری داده و پیش‌پردازش آن، انتخاب ویژگی‌ها، انتخاب هسته و پارامترهای مربوطه، آموزش، استفاده از طبقه‌بندی کننده ماشین بردار پشتیبان، آزمایش و در نهایت اعتبارسنجی برای نمونه‌های جدید می‌باشد. همان‌طور که می‌دانیم تنظیم درست پارامترهای مدل، می‌توانند صحت مدل طبقه‌بندی کننده بهینه ماشین بردار پشتیبان را بالا ببرند. پارامترهایی که بایستی در صورت استفاده از هسته RBF بهینه شوند. عبارتند از: گاما و C. در این تحقیق برای پیدا کردن پارامترها از Grid Search استفاده شده است. شکل (۳) رویه مدل مبتنی بر ماشین بردار پشتیبان را نشان می‌دهد [۸].



شکل (۳): رویه مبتنی بر ماشین بردار پشتیبان

۳- مبانی نظری و مروری بر پیشینه پژوهش

در بازاری که حاشیه سود بانک‌ها به دلیل تشدید رقابت همواره در حال کاهش بوده و همواره فشاری برای کاهش بیشتر هزینه‌ها احساس می‌شود. مدل‌های ریسک اعتباری با پیش‌بینی زیان‌های عدم بازپرداخت وام‌ها، نوعی برتری نسبی برای انک‌ها و نهادهای اعتباری ایجاد خواهد کرد. مدل‌های ریسک اعتباری با اندازه‌گیری ریسک می‌توانند با ایجاد ارتباط خردمندانه‌ای بین ریسک و بازده، امکان قیمت‌گذاری دارایی‌ها را فراهم سازد. هم‌چنین مدل‌های ریسک اعتباری امکان بهینه‌سازی ترکیب

[10] Support Vector Machine
[11] Hyper plane
[12] Support Vector
[13] Kernal methods

پرتفوی اعتباری و تعیین سرمایه اقتصادی بانک‌ها برای کاهش هزینه‌های سرمایه‌ای را فراهم خواهد ساخت [۱].

در تحقیقی تحت عنوان رتبه‌بندی اعتباری مشتریان بانک با استفاده

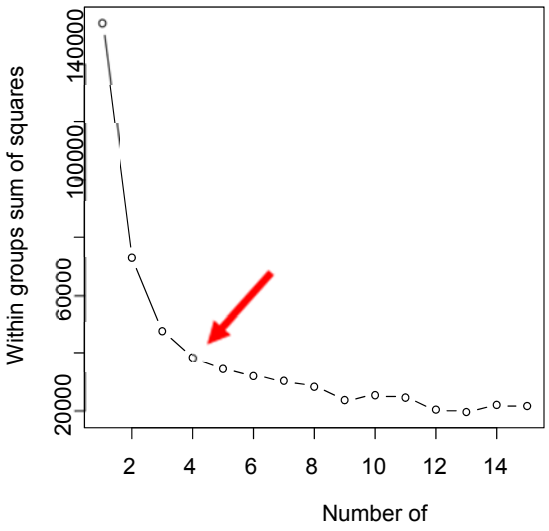
در تحقیقی تحت عنوان کاربرد استخراج داده‌ها در بانکداری به این نتیجه رسیدند که، استخراج داده‌ها یک ابزار به کار رفته برای استخراج اطلاعات مهم داده‌های موجود است و تصمیم‌گیری بهتر را در سرتاسر صنایع بانکداری و خرده‌فروشی توانا می‌سازد. روش‌های استخراج داده‌ها می‌تواند بر حسب بانک‌ها برای هدف‌گیری بهتر و به دست آوردن مشتریان جدید، کشف کلاهبرداری در کوتاه‌ترین زمان ارائه بخش‌های مبنی بر فرآورده‌ها، تجزیه و تحلیل، الگوهای خرید مشتری در طی زمان برای حفظ و رابطه بهتر به کار رفته باشند [۱۴].

در تحقیقی تحت عنوان مدل‌های مجموع ماشین‌های بردار پشتیبان حداقل مربعات برای امتیازبندی اعتبار، چندین مدل ماشین بردار پشتیبان برای ارزیابی ریسک اعتباری مطرح می‌شوند، اگر چه مدل‌های دسته جمعی مبنی بر قابلیت اعتماد نمی‌توانند برای کلیه مجموعه داده‌ها به بهترین عملکرد دسترسی پیدا کنند، آنها می‌توانند به خوبی بهترین نحوه با سطح ۵ درصد اعتماد دسترسی پیدا کنند. آزمایشات نشان دادند که مدل‌های دسته جمعی خوب و قوی هستند و می‌توانند راه‌حل‌های امیدوارکننده‌ای برای تجزیه و تحلیل ریسک اعتباری در همان زمان ارائه دهند و آنها علل بالقوه زیادی برای حل دیگر طبقه‌بندی، طبقه‌بندی دوتایی (مضاعف) دارند. هم‌چنین نتایج نشان دادند که این استراتژی‌های دسته جمعی می‌توانند به بهبود عملکرد تا یک میزان کمک کنند و برای ساخت مدل‌های امتیازبندی اعتبار مفید ومؤثر هستند [۱۶].

درتحقیقی تحت عنوان سیستم ایجاد ریسک اعتباری- مدیریت ریسک اعتباری در شرایط اقتصادی متغیر در بانک‌های مقدونی به این نتیجه رسیدند که، مدیریت موفقیت‌آمیز ریسک اعتباری بر اساس حفظ کیفیت بالای سهام (موجودی اوراق بهادار) وام با یک بخش مناسب و پراکندگی جغرافیایی است. این ریسک اعتباری به طور ساده به عنوان یک عامل بالقوه تعریف شده است که وام‌گیرنده بانک یا شریک برای جبران تعهدات آن بر طبق شرایط پذیرفته شده با شکست مواجه شده. برای اغلب بانک‌ها، این وام‌ها بزرگترین و واضح‌ترین منبع ریسک اعتباری است و لیکن دیگر منابع ریسک اعتباری در سرتاسر فعالیت‌های بانک وجود دارد. بانک‌ها به طور فزاینده با این ریسک بر حسب ابزار مالی متفاوت غیر از دیگر وام‌ها شامل: پذیرش حیطه معاملات بانکی، مالیات، تجارت، معاملات برواتی خارجی، ابزار مالی، معاوضه، دارایی‌های خاص و ضمانت‌ها و تسویه معاملات مواجه می‌شوند [۱۱].

۴- یافته‌های پژوهش

در این بخش فرضیه ارتباط رتبه‌بندی مشتریان در قالب مشتریان خوش-حساب و بدحساب با کاهش ریسک و سودآوری بانک و جذب مشتریان خوب با روش خوشه‌بندی که یکی از روش‌های داده‌کاوی می‌باشد، مورد بررسی و تحلیل قرار می‌گیرد. شناسایی مشتریان مستلزم تحلیل مشتریان هدف و دسته‌بندی مشتریان است، که منجر به یافتن گروه‌هایی از مشتریان خوش‌حساب براساس ویژگی‌های آنها می‌شود. به منظور آزمون این فرضیه ابتدا با استفاده از پیشینه پژوهشی و با استفاده از ایده-های محقق متغیرهای کیفی و کمی و نسبت‌های مالی انتخاب گردیدند. سپس با استفاده از نرم‌افزار R بخش‌بندی مشتریان با استفاده از روش خوشه‌بندی k-means انجام گرفت و ویژگی‌های مشتریان در هریک از بخش‌ها مشخص شد.قبل از انجام روش‌های داده‌کاوی، به منظور افزایش میزان دقت نتایج به دست آمده لازم است تعداد داده‌های نامناسب حذف شوند و هم‌چنین کلیه مقادیر به صورت عددی و نرمال سازی شده برای تحلیل به کار برده شوند. بنابراین در این مقاله برای نرمال‌سازی دادگان از روش نرمال‌سازی z استفاده شده است. در این تحقیق از شاخص مجموع مربعات درون گروهی برای تعیین تعداد خوشه اولیه (k) استفاده شد. به این مفهوم که با کمتر شدن پراکندگی بین گروه‌ها می‌توان خوشه مناسب را به دست آورد. با توجه به نمودار ۴ مشاهده می‌شود که بعد از ۴ خوشه مجموع مربعات درون گروهی تغییرات زیادی ندارند بنابراین تعداد خوشه بهینه را برابر ۴ در نظر گرفته می‌شود.



شکل (۴): تعداد خوشه‌ها

به منظور پیاده‌سازی این مدل از نرم‌افزار R به عنوان یکی از تخصصی‌ترین نرم‌افزارهای موجود در زمینه داده‌کاوی استفاده شد نتایج به دست‌آمده از پیاده‌سازی مدل در جداول زیر نشان داده شده است. جدول (۱) ویژگی هر یک از متغیرها در خوشه‌های به دست‌آمده از الگوریتم K-Means را نشان می‌دهد و جدول (۲) نسبت مشتریان بدحساب در هر خوشه و درنتیجه ارزیابی و برچسب‌گذاری هر یک از خوشه‌ها نشان داده شده است. همان‌گونه که نشان داده شد، ویژگی‌های هر یک از خوشه‌ها پس از خوشه‌بندی مشخص شد. با توجه به جدول (۲) با قرار گرفتن ۰/۰۳ از مشتریان بدحساب در خوشه یک و ۰/۰۲ مشتریان بدحساب در خوشه ۳، می‌توان بیان کرد که مشتریانی که ویژگی‌های خوشه یک و سه را دارا هستند توانایی بیشتری در بازپرداخت وام خود دارند. در مقابل مشتریانی که در خوشه ۲ و ۴ قرار می‌گیرند ریسک بالاتری در پرداخت وام خواهند داشت.

جدول (۱): میانگین هر یک از متغیرها در خوشه های به دست آمده با استفاده از الگوریتم K-Means

متغیرها	۱	۲	۳	۴
میزان سرمایه ثبتی	۴۵۹۱/۶۵۴۸	۶۵۴۶/۶۵۱۵	۷۹۵۱/۸۵۴۱	۷۵۸۴/۸۴۶
مدت وام	۳۱/۹۵۱۴۶	۲۷/۳۵۱۴۵	۴۸/۶۵۶۴۱	۳۲/۶۵۴۶۱
نرخ بهره	۲۷/۱۶۳۵ ۱	۲۵/۸۵۶۵۶۴	۲۶/۰۴۶۳۵	۲۶/۳۱۲۳۵
نوع وثایق	۱/۵۴۳۹۵۶	۱/۷۶۹۲	۱/۰۲۵۱	۱/۹۵۴۴
سابقه همکاری	۹/۶۵۴۱	۸/۳۵۱۶	۱۲/۳۵۴۶	۷/۵۵۶۱۴
نسبت جاری	۱/۳۱۲۳	۱/۰۱۶۱۲	۱/۵۳۵۴	۰/۹۵۶۵۴
نسبت بدهی	۰/۸۷۴۶۸	۱/۲۴۴۱۰۹	۰/۹۶۱۱۲۶	۱/۴۷۴۲۹
مبلغ مصوب وام	۳۱۶۹/۲۵۴۸	۴۸۱۱/۶۲	۱۱۴۹/۲۵۴۸	۶۵۱۱/۶۲

مجموع مربعات بین خوشه‌ها برابر ۷۷/۶ درصد می‌باشد. این عدد هرچه بیشتر باشد نشان‌دهنده تمایز بین خوشه‌ها می‌باشد، بنابراین تأییدکننده تعداد خوشه‌های مناسب و بهینه می‌باشد.

جدول (۲) : نتیجه ارزیابی خوشه های تشکیل شده در الگوریتم K-Means

شماره خوشه	نسبت افراد بدحساب درخوشه	نتیجه ارزیابی
۱	۰،۰۳	ریسک پایین
۲	۰،۲۱	ریسک نسبتاً بالا
۳	۰،۰۲	ریسک پایین
۴	۰،۷۲	ریسک بالا

همانطور که قبلاً بیان شد سعی بر آن است که مشتریان تسهیلات بانکی بانک سپه با کمک شبکه‌های عصبی رتبه‌بندی شود و سپس ارتباط آن با کاهش ریسک و سودآوری بانک و همچنین جذب مشتریان خوب سنجیده شود. با کمک این رتبه‌بندی مشتریان در دسته‌های مختلفی از نظر اعتبار رتبه‌بندی می‌شوند تا مشخص شود که کدام مشتریان و با چه مشخصاتی شرایط دریافت تسهیلات را دارند و از طرفی چه کسانی شایستگی دریافت اعتبار را ندارند. قابل ذکر می‌باشد که در این روش نرمال‌سازی، داده‌ها در بازه ۱ و ۱- قرار می‌گیرند. پس از نرمال‌سازی، به صورت تصادفی ۹۰ درصد از داده‌ها (۱۸۰مشتری) به عنوان داده های آموزش و ۲۵ درصد (۲۰ مشتری) برای داده‌های تست انتخاب شده‌اند. نتیجه حاصل از صحت طبقه‌بندی این الگوریتم و اعتبار مدل در جدول (۳) ارائه شده است:

جدول (۳): دقت مدل در دو گروه آموزش و تست

دقت مدل	گروه آموزش	گروه آزمایش
درست	٪۹۴	٪۹۰
نادرست	٪۶	٪۱۰
تعداد نمونه	۱۸۰	۲۰

میزان صحت مدل شبکه عصبی در مجموعه داده‌های آموزش ۹۴ درصد می‌باشد. درصدهای مربوط به مجموعه داده‌های آزمایش که معیاری برای ارزیابی اعتبار و صحت مدل می‌باشد، ۹۰ درصد می‌باشد. این نتایج در گروه آموزش با ۱۸۰ مشتری و در گروه آزمایش با ۲۰ مشتری به دست‌آمده است. پیش‌بینی‌های مورد انتظار برای ۲۰ مشتری در مجموعه داده‌های تست برای مدل شبکه عصبی در جدول (۴) نشان داده شده است.

جدول (۴): عملکرد پیش بینی شبکه عصبی در مرحله آزمایش

اعتبار مشتریان / پیش بینی مدل شبکه عصبی	بدحساب	خوش حساب
بدحساب	۲	۱
خوش حساب	۱	۱۶

همان‌طور که در جدول (۴) مشخص است از مجموع ۱۷ مشتری خوش‌حساب، تعداد ۱۶ مشتری به عنوان خوش‌حساب و ۱ مشتری (به اشتباه) به عنوان بدحساب تشخیص داده شده‌اند، به طوری که دقت الگو در این خصوص ۹۴ درصد محاسبه شده است. به همین ترتیب از ۳ مشتری

بدحساب، ۲ مشتری به عنوان بدحساب و ۱ مشتری (به اشتباه) خوش-حساب تشخیص داده شده‌اند و دقت الگو در این حالت حدود ۶۷ درصد می‌باشد. بنابراین در کل دقت پیش‌بینی در این مدل ۹۰ درصد است. برای آزمون این فرضیه از نرم‌افزار Weka۳,۷,۱۳ استفاده شده است. بدین صورت که ابتدا داده‌های خام جمع‌آوری و طبقه‌بندی و سپس در قالب CSV وارد نرم‌افزار کردیم. برای آماده‌سازی و پیش‌پردازش داده‌ها در نرم‌افزار Weka، ابزار فیلتر به کار برده می‌شود که شامل بخش‌های نرمال‌سازی، نمونه‌گیری، انتخاب ویژگی، تبدیل و نظایر آن می‌باشد. بنابراین برای تنظیم درست و بهینه‌ای از پارامترهای مدل از روش Grid Search با 10-fold اعتبار متقابل استفاده شده است. همچنین گروه-های آموزش ۷۵ درصد و تست ۲۵ درصد به صورت تصادفی در نظرگرفته شده است. پارامترهای با کمترین مقدار خطا و کمترین پراکندگی به عنوان پارامترهای بهینه به شمار می‌آیند. با در نظر گرفتن پارامترها، الگوریتم ماشین‌بردارپشتیبان اجرا شد و مدل ماشین‌بردار براساس آنها ساخته شد. نتیجه حاصل از صحت طبقه‌بندی این الگوریتم و اعتبار مدل در جدول (۵) ارائه شده است.

جدول (۵): دقت مدل ماشین بردار پشتیبان

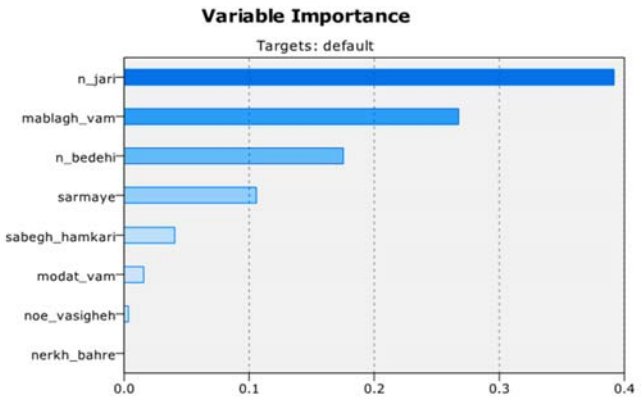
دقت مدل	تعدادپیش بینی	درصد پیش بینی
درست	۱۵۳	٪۷۶
نادرست	۴۷	٪۲۳

میزان صحت مدل ماشین‌بردارپشتیبان (SVM) در مجموعه داده‌ها ٪۷۶ می‌باشد. این مقدار مربوط به صحت طبقه‌بندی و معیاری برای ارزیابی اعتبار و صحت مدل می‌باشد. همچنین تعداد داده‌هایی که درست پیش-بینی شده‌اند ۱۵۳ و تعداد داده‌هایی که نادرست پیش‌بینی شده‌اند ۴۷ می‌باشند. پیش‌بینی‌های مورد انتظار برای ۲۰۰ مشتری در مجموعه کل داده‌ها، برای مدل ماشین‌بردارپشتیبان در جدول (۶) مشاهده می‌شود. همان‌طور که در جدول (۶) مشخص است، از مجموع ۱۵۱ مشتری خوش‌حساب، تعداد ۱۲۸ مشتری به عنوان خوش‌حساب و ۲۳ مشتری

جدول (۶): عملکرد پیش بینی مدل ماشین بردار پشتیبان

اعتبار مشتریان / پیش‌بینی مدل شبکه عصبی	بدحساب	خوش حساب
بدحساب	۲۵	۲۳
خوش حساب	۲۴	۱۲۸

(به اشتباه) به عنوان بدحساب تشخیص داده شده‌اند، به طوری که دقت الگو در این خصوص ۸۵/۷۶ درصد محاسبه شده است. به همین ترتیب از ۴۹ مشتری بدحساب، ۲۵ مشتری به عنوان بدحساب و ۲۴ مشتری (به اشتباه) خوش‌حساب تشخیص داده شده‌اند و دقت الگو در این حالت حدود ۵۱/۰۱ درصد می‌باشد. بنابراین میانگین دقت پیش‌بینی در این مدل ۶۸/۵۰ درصد است. ضعف این مدل در پیش‌بینی مشتریان بدحساب می‌باشد که میزان دقت پیش‌بینی در این حالت ۵۱/۰۱ درصد می‌باشد که در مقایسه با سایر مدل‌ها از قدرت پایین‌تری برخوردار می‌باشد. همچنین بر اساس این مدل متغیرهای با اهمیت در شکل (۵) نمایش داده است.



شکل (۵): متغیرهای با اهمیت

براساس شکل (۵) متغیرهای با اهمیت عبارت است از: نسبت جاری، مبلغ مصوب وام، نسبت بدهی، سرمایه ثبتی شرکت، سابقه همکاری، مدت وام، نوع وثیقه و نرخ بهره.

۵- نتیجه‌گیری

اعطای تسهیلات اعتباری به مشتریان از جمله مهمترین وظایف بانک‌ها به شمار می‌رود. بانک‌ها در هر کشور پس از جمع‌آوری منابع مالی، این منابع را به بخش‌های مختلف اقتصادی تخصیص می‌دهند. در حقیقت این اقدام بانک‌ها، بخش‌های مختلف اقتصادی را در هر کشور در انجام بهتر وظایفشان تقویت و نهایتاً زمینه لازم را برای رشد و توسعه اقتصادی کشور فراهم می‌آورند. در این پژوهش با توجه به بررسی داده‌ها به روش خوشه‌بندی، شبکه عصبی و ماشین بردار پشتیبان به این نتیجه رسیده‌ایم که این اعداد نشان می‌دهند که کارایی مدل شبکه عصبی بهتر و مؤثرتر از روش ماشین بردار پشتیبان در هر دو حالت خوش‌حساب و بدحساب می‌باشد و هم‌چنین دقت پایین مدل ماشین بردار پشتیبان در پیش‌بینی مشتریان بدحساب در این روش می‌باشد. هم‌چنین به این نتیجه رسیده‌ایم که شاخص‌های مؤثر در ریسک اعتباری مشتریان بانک سپه شهرستان دزفول در روش خوشه‌بندی به ترتیب: نسبت جاری، نسبت بدهی، مبلغ مصوب وام و سابقه همکاری می‌باشد که اهمیت این شاخص‌ها در روش ماشین بردار پشتیبان به ترتیب: نسبت جاری، مبلغ مصوب وام، نسبت بدهی، سرمایه ثبتی، سابقه همکاری، مدت وام، نوع وثایق و نرخ بهره می‌باشد که نشان می‌دهد که در هر دو روش نوع وثایق و نرخ بهره تأثیر زیادی بر ریسک اعتباری مشتریان بانک مورد مطالعه نداشته‌اند. هم‌چنین کارایی بالای مدل شبکه عصبی در پیش‌بینی مشتریان خوش‌حساب و بدحساب به این نتیجه رسیدیم که بانک مورد مطالعه توانایی بالایی در شناسایی مشتریان خوش‌حساب دارد، که این شناسایی باعث می‌شود که بانک سپه قبل از اعطای تسهیلات، مشتریان خوش‌حساب خود را شناسایی کرده و تسهیلات اعتباری را به مشتریان واجد شرایط اعطا نماید. همین امر باعث کاهش ریسک و سودآوری بانک و جذب مشتریان خوش‌حساب می‌شود. با توجه به نتایج تحقیق می‌توان پیشنهادات زیر را ارائه داد:

- افزایش اطلاعات مدیران از تکنولوژی‌های داده‌کاوی و مزایای آن.

- با توجه به این که برای ایجاد، توسعه و بهبود مدل‌های ریسک اعتباری داشتن اطلاعات مالی و اقتصادی روزآمد نقش مهم و اساسی ایفا می‌نماید. لذا برخورداری از بانک اطلاعاتی مشتریان و بنابراین طراحی و تنظیم سیستم کارآمد اخذ اطلاعات اعتباری به عنوان یکی از ابزارهای پایه مطرح است. بر این اساس پیشنهاد می‌شود که، زیرساخت‌های اطلاعاتی یکپارچه و پایگاه‌های اطلاعاتی نرمال در بانک‌ها ایجاد گردد.

- با توجه به روند فزاینده سهم مطالبات معوق از کل مطالبات بخش غیر دولتی، استقرار سیستم رتبه‌بندی اعتباری مشتریان، بانک‌ها را در تخصیص بهینه منابع و تسهیلات مالی خود یاری می‌دهد. تحت این سیستم تسهیلات به مشتریان مطلوب تخصیص می‌یابد. مشتری مطلوب، مشتری است که ضمن هزینه نمودن تسهیلات دریافتی در بخش‌های مختلف اقتصادی آن را به سیستم بانکی بازگردانده و بدین ترتیب باعث افزایش منابع مالی در اختیار بانک‌ها و به تبع آن افزایش قدرت وام‌دهی، درآمد ملی و تولید ناخالص ملی گردد. هم‌چنین با استقرار سیستم رتبه‌بندی اعتباری مشتریان، این امکان برای سیستم بانکی کشور فراهم می‌آید تا میزان و نوع وثیقه دریافتی از هر مشتری را بر اساس ریسک هر یک تعیین نماید و با توجه به میزان ریسک اعتباری هر مشتری نوع وثیقه تعیین شده به گونه‌ای باشد که قادر به پوشش ریسک اعتباری آنها گردد.

- براساس نتایج به دست آمده ایجاد سیستم نظارتی و کنترلی قوی جهت دریافت به موقع صورت‌های مالی به منظور بررسی نسبت‌ها و شاخص‌های مالی مورد نیاز جهت استفاده در سیستم رتبه‌بندی اعتباری، به مبحثی قابل تأمل تبدیل می‌گردد.

- ثبت کامل تمامی اطلاعات در فیلدهای بانک اطلاعاتی.

- افزایش آگاهی مدیران نسبت به سودآوری حاصل از جذب مشتریان خوش‌حساب در کاهش ریسک اعتباری.

منابع و ماخذ

[۱] ابراهیمی، مرضیه، دریابر، عبدالله، (۱۳۹۱)، **مدیریت ریسک اعتباری در نظام بانکی- رویکرد تحلیل پوششی داده‌ها و رگرسیون لجستیک و شبکه عصبی**، فصلنامه دانش سرمایه‌گذاری، ش ۲، ص ۶۲- ۳۵.

[۲] بهرامی‌زنوز، مریم، (۱۳۸۹)، **داده‌کاوی: کشف پنهان داده‌ها**، اداره تحقیقات و کنترل ریسک بانک سپه، ص ۱۲- ۲.

[۳] تارخ، محمدجعفر، شریفیان، کبری، (۱۳۸۹)، **کاربرد داده‌کاوی در بهبود مدیریت ارتباط با مشتری**، فصلنامه علمی-پژوهشی مطالعات مدیریت صنعتی، ش ۱۷، ص ۱۸۱- ۱۵۳.

[۴] دستگیر، محسن، شفیعی‌سردشت، مرتضی، (۱۳۸۸)، **فناوری داده‌کاوی؛ رویکردی نوین در حوزه مالی**، فصلنامه حسابرسی. دوره جدید، ش ۵، ص ۲۲- ۶.

[۵] دادمحمدی، دانیال، احمدی، عباس، (۱۳۹۳)، **رتبه‌بندی اعتباری مشتریان بانک با استفاده از شبکه عصبی با اتصالات جانبی**، فصلنامه توسعه مدیریت پولی و بانکی، ش ۳، ص ۲۸-۱.

[۶] رادفر، رضا، نظافتی، نوید، یوسفی‌اصلی، سعید، (۱۳۹۲)، **طبقه‌بندی مشتریان اینترنت بانک با کمک الگوریتم‌های داده‌کاوی**، مدیریت فناوری اطلاعات. دوره ۶، ش ۱، ص ۹۰-۷۱.

[۷] قره‌خانی، محسن، ابوالقاسمی، مریم، (۱۳۸۹)، **کاربردهای داده‌کاوی در صنعت بیمه، تازه‌های جهان بیمه**، ش ۲۰، ص ۱۵۸-۵.

[۸] طلوعی‌اشلقی، عباس، نیکومرام، هاشم، مقدوری شربیانی، فرناز، (۱۳۸۹)، **طبقه‌بندی متقاضیان تسهیلات اعتباری بانک‌ها با استفاده از تکنیک ماشین بردار پشتیبان**، مجله پژوهش‌های مدیریت، ش ۸، ص ۱۸-۲.

[۹] مهرآرا، محسن، موسایی، میثم، تصوری، مهسا. حسن‌زاده آیت، (۱۳۹۰)، **رتبه‌بندی اعتباری مشتریان حقوقی بانک پارسیان**. فصلنامه مدل‌سازی اقتصادی، ش ۳، ص ۱۵۰-۱۲۱.

[10] Chen, W., Xiang, G., Liu, Y., Wang, K., (2012), **Credit Risk Evaluation by Hybrid Data Mining Technique**, System Engineering Procedia 3, 194-200.

[11] Dicevska, S., (2012), **Credit Risk- Creating System of Credit Risk Management in Changing Economic Conditions in Macedonian Banks**, Procedia- Social and Behavioral Sciences 44, 460-469.

[12] Deshpande, S. P., Thakare, V., M., (2010), **Data Mining System and Application: A Review**, International Journal of Distributed and Parallel system 1, 22-40.

[13] Harris, T., (2015), **Credit Scoring Using the Clustered Support Vector Machine**, Expert system with Applications 42, 741-750.

[14] Moin, K. I., Ahmed, D. Q. B., (2012), **Use of Data Mining in Banking,** International Journal of Engineering Research and Applications 2, 738-742.

[15] Oreski, S., Oreski, G., Oreski, D.S., (2012). **Hybrid system with Genetic Algorithm and Artificial Neural Networks and Its Application to Retail Credit Risk Assessment**, Expert System with Applications 39, 12605-12617.

[16] Zhou, L., Lia, K. K., Yu, L. (2010). **Least Squares Support Vector Machines Ensemble Models for Credit Scoring**, Expert System with Applications 37, 127-133.