فصلنامهٔ پژوهش‌های پولی_بانکی
سال هفدهم، شمارهٔ ۶۱، پاییز ۱۴۰۳
صفحات ۵۳۵_۵۶۴

مدیریت ریسک اعتباری جهت بهبود اعطای تسهیلات به مشتریان بانکی

فاطمه محمدی* **محسن مهرآرا†**
علی جدیدزاده‡

تاریخ دریافت: ۱۴۰۳/۰۹/۱۹ تاریخ پذیرش: ۱۴۰۳/۱۱/۱۶

چکیده

نظام بانکی به‌عنوان یکی از ارکان مالی کشور نقشی اساسی در توسعهٔ اقتصادی دارد. یکی از چالش‌های مهم بانک‌ها در ارائهٔ تسهیلات، شناخت دقیق مشتریان و تمایز میان گروه‌های مختلف آن‌هاست. این پژوهش با هدف شناسایی و دسته‌بندی مشتریان بانکی بر اساس ریسک اعتباری و پیش‌بینی احتمال نکول آن‌ها با استفاده از مدل‌های یادگیری ماشین انجام می‌شود. نتایج دسته‌بندی با الگوریتم میانگین کا (K-means) نشان داد که مشتریان به چهار گروه اصلی تقسیم می‌شوند: خوشهٔ چهارم شامل مشتریان پرریسکی است که بدهی‌های معوقهٔ بالا و تسهیلات نکول‌شده دارند و نیازمند نظارت دقیق‌اند. خوشهٔ سوم مشتریان بحران‌سازی را شامل می‌شود که با بدهی‌های معوقهٔ بالا در معرض انتقال به گروه پرریسک قرار دارند. خوشهٔ دوم شامل مشتریان با ریسک متوسط است که هرچند پایدارند، به نظارت نیاز دارند. خوشهٔ اول نیز مشتریان کم‌ریسکی را شامل می‌شود که بدهی‌های معوقهٔ کمی دارند و بهترین گزینه برای اعطای تسهیلات جدیدند. در این پژوهش، از مدل‌های ایکس‌جی‌بوست و لاجیت برای پیش‌بینی نکول استفاده شده است. مدل ترکیبی ارائه‌شده به بانک‌ها کمک می‌کند تا راهبردهای مؤثرتری برای مدیریت ریسک و تخصیص بهینهٔ منابع مالی تدوین و جایگاه رقابتی خود را در بازار تقویت کنند.

واژه‌های کلیدی: ریسک اعتباری، احتمال نکول، دسته‌بندی مشتریان، تسهیلات بانکی، پیش‌بینی وضعیت اعتباری.
طبقه‌بندی JEL: C38، C53، G21

* دانشجوی مقطع ارشد اقتصاد انرژی دانشگاه تهران (نویسندهٔ مسئول)؛ fatemehmohamadi78@ut.ac.ir
† استاد دانشکدهٔ اقتصاد دانشگاه تهران؛ mmehrara@ut.ac.ir
‡ استادیار دانشکدهٔ اقتصاد دانشگاه تهران؛ jadidzadeh@ut.ac.ir

۱ مقدمه

ارزیابی دقیق و مؤثر اعتبار متقاضیان تسهیلات بانکی یکی از عوامل کلیدی در مدیریت بهینهٔ منابع مالی و کاهش ریسک‌های مرتبط با نکول[1] تسهیلات است. بانک‌ها به‌عنوان نهادهای مالی بزرگ و تأثیرگذار در اقتصاد کشور مسئولیت سنگینی در تأمین سرمایه برای پروژه‌های تولیدی، سرمایه‌گذاری‌های کلان، و تسهیلات خرد دارند. این نقش، به‌ویژه در شرایط پیچیده و ناپایدار اقتصادی اهمیت بیشتری می‌یابد. در واقع، در یک نظام مالی پویا و مقاوم بانک‌ها با ارزیابی دقیق و به‌موقع ریسک‌های اعتباری می‌توانند ضمن کاهش احتمال نکول تسهیلات، ثبات مالی خود را حفظ کرده و در توسعهٔ پایدار اقتصادی کشور نیز مشارکت مؤثری داشته باشند. ازاین‌رو، فرایند اعتبارسنجی[2] به یکی از حساس‌ترین و راهبردی‌ترین فرایندهای بانک‌ها و مؤسسات مالی تبدیل شده است. در سال‌های اخیر، افزایش حجم و تنوع تقاضاهای تسهیلاتی و پیچیدگی‌های ناشی از نوسانات اقتصادی موجب شده است که روش‌های سنتی ارزیابی اعتبار کارایی لازم را نداشته باشند. این روش‌ها که بیشتر بر قضاوت‌های فردی، تجربه، و اطلاعات محدود تکیه دارند، به‌درستی تمام جنبه‌های ریسک اعتباری[3] مشتریان را پوشش نمی‌دهند. در این مطالعه، از الگوریتم میانگین کا (K-means) برای دسته‌بندی مشتریان استفاده می‌شود که نشان‌دهندهٔ تقسیم آن‌ها به چهار گروه اصلی است. مشتریان پرریسک، که در خوشهٔ چهارم قرار دارند، دارای بالاترین بدهی‌های معوقه و تسهیلات نکول شده‌اند. بانک‌ها باید از اعطای تسهیلات به این گروه تا حد امکان جلوگیری کنند. خوشهٔ سوم شامل مشتریان بحران‌سازی است که در معرض ریسک بالای نکول قرار دارند، درحالی‌که خوشهٔ دوم مشتریان با ریسک متوسط را دربرمی‌گیرد. خوشهٔ اول نیز شامل مشتریان معتبر و کم‌ریسک است که بهترین گزینه‌ها برای اعطای تسهیلات جدید به‌شمار می‌روند.

در دههٔ اخیر، بانک‌ها به‌منظور بهبود عملکرد مالی خود و کاهش ریسک معوقات، به‌سمت استفاده از مدل‌های ارزیابی پیشرفته‌تر و داده‌محور حرکت کرده‌اند. این مدل‌ها که بر پایهٔ فنّاوری‌های نوین؛ همچون یادگیری ماشین و تحلیل داده‌های با حجم بالا[4] توسعه یافته‌اند، کمک می‌کنند تا بانک‌ها بتوانند با دقت بیشتری متقاضیان تسهیلات را ارزیابی کرده و ریسک

---

[1] default

[2] credit scoring

[3] credit risk

[4] big data

اعتباری را به حداقل برسانند. پژوهش حاضر با هدف ارائهٔ یک مدل ترکیبی و جامع برای ارزیابی ریسک اعتباری متقاضیان تسهیلات بانکی انجام می‌شود. این مدل بر تحلیل داده‌های بانکی و استفاده از الگوریتم‌های یادگیری ماشین استوار است و تلاش می‌کند دقت پیش‌بینی ریسک نکول را افزایش دهد. یکی از نوآوری‌های این پژوهش ترکیب چندین روش مختلف یادگیری ماشین با داده‌های متنوع از جنبه‌های مالی، اعتباری، و رفتاری مشتریان است. به‌این‌ترتیب، مدل پیشنهادی نه‌تنها می‌تواند به شناسایی متقاضیان کم‌ریسک کمک کند، بلکه به بانک‌ها ابزاری کارآمد در بهبود مدیریت تسهیلات و افزایش بهره‌وری مالی ارائه می‌دهد.

در بخش‌های بعدی این پژوهش، ابتدا به مرور ادبیات موجود درزمینهٔ ارزیابی ریسک اعتباری و کاربردهای یادگیری ماشین پرداخته می‌شود؛ سپس مدل پیشنهادی و مراحل مختلف پیاده‌سازی آن معرفی می‌شود؛ درنهایت، نتایج حاصل از پیاده‌سازی این مدل در یک مجموعه دادهٔ واقعی از مشتریان بانکی موردارزیابی قرار می‌گیرد و مزایا و محدودیت‌های آن به‌تفصیل بررسی می‌شود.

۲ پیشینهٔ تحقیق

مدیریت اعطای تسهیلات به مشتریان یکی از مهم‌ترین وظایف نظام بانکی به‌شمار می‌رود و تأثیر بسزایی بر تأمین مالی، رونق اقتصادی، و کاهش ریسک اعتباری دارد. بااین‌حال، اعطای تسهیلات به متقاضیانی که توانایی بازپرداخت ندارند، برای بانک‌ها ریسک‌هایی جدی به همراه دارد. ازاین‌رو، بانک‌ها از روش‌ها و مدل‌های متنوعی برای بهینه‌سازی فرایند اعطای تسهیلات و کاهش ریسک اعتباری بهره می‌برند. در همین راستا، پژوهش‌های متعددی در زمینه‌های مختلفی همچون بخش‌بندی مشتریان، تحلیل ارزش چرخهٔ عمر مشتری، رتبه‌بندی اعتباری، و پیش‌بینی نکول انجام شده است.

۲/۱ پژوهش‌های داخلی

از مؤثرترین روش‌های مدیریت تسهیلات بانکی، بخش‌بندی مشتریان بر اساس ویژگی‌های آن‌ها و تحلیل ارزش چرخهٔ عمر[1] آن‌هاست. این روش به بانک‌ها کمک می‌کند تا مشتریان خود را به گروه‌های مختلفی مانند مشتریان باارزش بالا و پایین دسته‌بندی و راهبردهای متناسب با هر دسته را تدوین کنند.

[1] Customer Lifetime Value (CLV)

مرادی و همکاران (۱۴۰۲) در پژوهشی به بررسی رتبه‌بندی گروه‌های مشتریان و تعیین بخش‌های برتر در صنعت بانکداری پرداخته و از مدل آراف‌ام و شبکهٔ عصبی خودسازمان‌دهنده استفاده کردند. در این تحقیق، پس از پیش‌پردازش داده‌ها با مدل آراف‌ام و به‌کارگیری الگوریتم خوشه‌بندی، مشتریان به ۱۰ خوشه تقسیم شدند. با رتبه‌بندی این خوشه‌ها، سه خوشهٔ برتر شناسایی و به‌عنوان مشتریان هدف برای اعطای تسهیلات پیشنهاد شدند. این نتایج نشان‌دهندهٔ تأثیر مؤثر روش‌های پیشرفته در بهبود شناسایی مشتریان کلیدی در صنعت بانکداری است. تقوی‌فرد (۱۳۹۶) در تحقیق خود با عنوان «دسته‌بندی مشتریان حقوقی و پیش‌بینی توانایی سوددهی آن‌ها با استفاده از ارزش طول عمر مشتری و رویکرد زنجیرهٔ مارکوف[1]» به بررسی ارزش طول عمر مشتری و اهمیت دسته‌بندی مشتریان پرداخت. او با استفاده از مدل آراف‌ام، تکنیک سلسله‌مراتبی،[2] و نیز نظرات کارشناسان بانکی، متغیرهای مختلف را وزن‌دهی و مشتریان را گروه‌بندی کرد. نتایج این پژوهش شامل استخراج ماتریس احتمال برای پیش‌بینی جابه‌جایی مشتریان و تعیین ضریب برای رتبه‌بندی دسته‌های مختلف مشتریان بود. این دو مطالعه به‌خوبی نشان می‌دهند که بهره‌گیری از روش‌های پیشرفتهٔ داده‌کاوی و تحلیل آماری می‌تواند در شناسایی و دسته‌بندی مشتریان کلیدی به بانک‌ها کمک کرده و بهبود راهبردهای بازاریابی و ارائهٔ خدمات هدفمندتر را به‌دنبال داشته باشد.

در پژوهش دیگری، تقوی فرد و خواجوند (۱۳۹۱) با استفاده از تحلیل تابع رادیال[3] و الگوریتم دو بخشی[4] به بخش‌بندی مشتریان بانک صادرات ایران پرداختند. آن‌ها مشتریان را به چهار دستهٔ «طلایی، ارزشمند، وفادار، و کم‌ارزش» تقسیم کردند. این دسته‌بندی به بانک کمک کرد تا باتوجه‌به ویژگی‌های هر گروه، راهبردهای متناسبی برای مدیریت ارتباط با مشتریان تدوین کند. آجرلو و همکاران (۱۳۹۰) با استفاده از اطلاعات تراکنشی پنج هزار مشتری حقیقی و حقوقی بانک ملت، مدلی برای سنجش و اندازه‌گیری ارزش طول عمر مشتریان ارائه کردند. این مدل به بانک‌ها کمک کرد تا مشتریان را بر اساس ارزش اقتصادی‌شان اولویت‌بندی کنند.

---

[1] Markov Chain

[2] Analytic Hierarchy Process (AHP)

[3] Radial Basis Function (RBF)

[4] Bipartite Algorithm

۲/۲ پژوهش‌های خارجی

در سطح بین‌المللی، لی و همکاران (۲۰۲۱)[1] در مقاله‌ای رویکردی جامع برای شناسایی، بهینه‌سازی، و تفسیر خوشه‌های داده‌های مالی ارائه دادند که نشان می‌دهد خوشه‌بندی مالی ابزاری کارآمد برای تحلیل رفتار مشتریان و بهینه‌سازی تصمیمات مالی است. عماد و همکاران (۲۰۱۹)[2] در مقالهٔ خود نقش الگوریتم‌های یادگیری ماشین در تحلیل و بهبود پروفایل‌های مشتریان بانکی را بررسی کردند. نتایج این پژوهش نشان داد که استفاده از این الگوریتم‌ها فرایندهای تصمیم‌گیری بانک‌ها را بهبود بخشیده و به شناسایی دقیق‌تر رفتارهای مشتریان کمک می‌کند. علاوه‌براین، سکیتوفسکی و شارلیا (۲۰۱۴)[3] در پژوهش خود با استفاده از تحلیل خوشه‌ای به بخش‌بندی مشتریان و بهبود سیستم‌های اعتبارسنجی پرداخته‌اند و نشان داده‌اند که این روش می‌تواند به پیش‌بینی بهتر امتیازات اعتباری و کاهش ریسک نکول کمک کند. مطالعات بین‌المللی بررسی‌شده به‌طور قابل‌توجهی با اهداف مدنظر برای این پژوهش هم‌راستا بودند، چراکه این پژوهش به‌دنبال استفاده از ترکیب مدل‌های یادگیری ماشین و آماری برای بهبود ارزیابی ریسک اعتباری است. درحالی‌که مطالعات قبلی بیشتر بر بخش‌بندی و بهبود تصمیم‌گیری تمرکز داشته‌اند؛ ادغام داده‌های چندگانه و مدل‌های ترکیبی پیشرفته در این پژوهش گامی فراتر برداشته و رویکردی جامع‌تر و پیش‌بینانه‌تر ارائه می‌دهد.

[1] LI et al.
[2] Emad et al.
[3] Scitovski & Šarlija

جدول ۱

خلاصهٔ مطالعات انجام‌شده در حوزهٔ دسته‌بندی و پیش‌بینی مشتریان متقاضی تسهیلات

ردیف	سال	نویسنده(گان)	عنوان مقاله	یافته‌ها	کاستی‌ها
۱	۱۴۰۲	مرادی و همکاران	مطالعات کمّی در مدیریت صنعت بانکداری برای افزایش رضایتمندی و سودآوری مشتریان (مطالعهٔ موردی: بانک ملت).	در این پژوهش، از مدل RFM و شبکهٔ عصبی SOM برای خوشه‌بندی مشتریان استفاده شده است. مشتریان در قالب ۱۰ خوشه طبقه‌بندی شدند که خوشه‌های ۵، ۱، و ۷ به‌عنوان بهترین خوشه‌ها شناخته شدند.	این مطالعه تنها به رتبه‌بندی اعتباری مشتریان پرداخته و پیش‌بینی ریسک نکول موردبررسی قرار نگرفته است.
۲	۱۳۹۶	تقوی‌فرد	دسته‌بندی مشتریان حقوقی و پیش‌بینی توانایی سوددهی آنان با استفاده از ارزش طول عمر مشتری و رویکرد زنجیرهٔ مارکوف.	استفاده از مدل RFM و تحلیل سلسله مراتبی برای گروه‌بندی و پیش‌بینی جابه‌جایی مشتریان.	پیچیدگی مدل و نیاز به داده‌های دقیق و کامل.
۳	۱۳۹۱	تقوی‌فرد و خواجوند	بخش‌بندی مشتریان بانک صادرات ایران با استفاده از داده‌کاوی.	تقسیم مشتریان به چهار دسته (طلایی، ارزشمند، وفادار، کم‌ارزش) و تدوین راهبردهای متناسب هر دسته.	تمرکز محدود بر مشتریان بانک صادرات، عدم‌استفاده از مدل پیش‌بینی نرخ نکول برای هر خوشه، و استفاده از متغیرهای محدودی برای خوشه‌بندی.
۴	۱۳۹۰	آجرلو و همکاران	الگویی برای تعیین ارزش چرخهٔ عمر مشتریان CLV در صنعت بانکداری.	مشتریان با بالاترین CLV به‌دست‌آمده از روش RFM، از نظر ارزش مالی، مهم‌ترین گروه برای بانک محسوب می‌شوند. مشتریانی با وفاداری پایین و مانده‌حساب متوسط به راهبردهای خاصی برای افزایش تعامل نیاز دارند.	استفاده از روش RFM برای تعیین ارزش مشتریان با تأکید بر طول عمر مشتری. نبود مدل‌سازی برای پیش‌بینی نکول. تمرکز بیشتر بر راهبردهای تعامل و تخصیص منابع به مشتریان با CLV بالا و عدم پوشش نکول، محدود به مشتریان بانک ملت.
۵	۲۰۲۱	لی و همکاران	رویکردی یکپارچه برای تشخیص، بهینه‌سازی، و تفسیر خوشه‌های مالی	محققان یک روش تشخیص خوشه‌ای معرفی کرده‌اند که شامل الگوریتم‌های بهینه‌سازی برای شناسایی ساختارهای پنهان در داده‌های مالی است. این رویکرد به تجزیه‌وتحلیل بهتر و تفسیری دقیق‌تر از داده‌های مالی کمک می‌کند.	تمرکز بر خوشه‌بندی و عدم‌تمرکز بر مشتریان حقیقی بانکی.

۶	۲۰۱۹	عماد و همکاران	بهبود پروفایل رفتاری مشتریان بانک با استفاده از یادگیری ماشین.	محققان این پژوهش تأکید دارند که ترکیب داده‌های Transaction و اطلاعات جمعیت‌شناسی می‌تواند به بهبود دقت پروفایل‌سازی کمک کند. این تحقیق چهار تکنیک یادگیری ماشین K-means، K-means بهبودیافته، C-means فازی، و شبکه‌های عصبی را روی یک مجموعه دادهٔ برچسب‌گذاری‌شده از یک بانک تایوانی اعمال کرد.	نیاز به حجم بزرگ‌داده و تنظیمات پیچیدهٔ مدل‌های یادگیری ماشین.
۷	۲۰۱۴	سکیتوفسکی و شارلیا	تحلیل خوشه‌ای در دسته‌بندی خرده‌فروشان برای انجام اعتبارسنجی	بهبود فرایندهای اعتبارسنجی و کاهش ریسک نکول از طریق بخش‌بندی مشتریان خرده‌فروشی.	تمرکز محدود بر بخش‌بندی خرده‌فروشی، نیاز به بررسی دقیق‌تر متغیرهای دیگر.
۸	۲۰۱۴	خبزی و همکاران	کاربرد جدید خوشه‌بندی RFM برای دسته‌بندی مشتریان و استخراج الگوی استفاده از خدمات پرداخت الکترونیک بانک‌ها	شناسایی الگوهای رفتاری مشتریان و بهبود راهبردهای بازاریابی بانک‌ها با استفاده از خوشه‌بندی RFM.	تأکید بر روش RFM و ویژگی‌های مالی معاملات.

مطالعهٔ حاضر با استفاده از الگوریتم میانگین کا (K-means) برای دسته‌بندی مشتریان متقاضی تسهیلات بر داده‌های تاریخی و رفتار اعتباری مشتریان تمرکز دارد و به‌منظور پیش‌بینی و مدیریت ریسک‌های اعتباری، چهارچوبی دقیق‌تر و کاربردی‌تر برای بانک‌ها فراهم می‌آورد. این رویکرد می‌تواند به بهبود فرایند اعتبارسنجی و راهبردهای اعطای تسهیلات در بانک‌ها کمک کند و باتوجه‌به نتایج مطالعات پیشین، بهینه‌سازی قابل‌توجهی در فرایندهای تصمیم‌گیری بانک‌ها به ارمغان آورد.

۳ مبانی نظری

۳/۱ خوشه‌بندی

خوشه‌بندی یک روش دسته‌بندی جمعیت‌های ناهمگن به خوشه‌های همگن است (ژیو و همکاران، ۲۰۰۹)[1]. در آغاز فرایند خوشه‌بندی، اطلاعاتی درمورد تعداد، شکل، یا ویژگی‌های خوشه‌ها در دست نیست و به‌دلیل نبود دانش قبلی دربارهٔ خوشه‌ها، این روش به‌عنوان یک تکنیک بدون ناظر[2] شناخته می‌شود. الگوریتم میانگین کا (K-means) یکی از پرکاربردترین و ساده‌ترین روش‌های دسته‌بندی‌است که برای تقسیم یک مجموعه از نقاط داده به k بخش استفاده می‌شود. هدف اصلی این الگوریتم به‌حداقل‌رساندن مجموع مربعات فاصلهٔ نقاط داده از مراکز خوشه‌هاست. مراحل اصلی این الگوریتم عبارت‌اند از:

۱) انتخاب تصادفی k مرکز اولیه: در ابتدا، k مرکز به‌صورت تصادفی از میان نقاط داده انتخاب می‌شود.
۲) اختصاص نقاط به نزدیک‌ترین مرکز: هر نقطه داده به مرکز خوشه‌ای که کمترین فاصله را با آن دارد، اختصاص داده می‌شود. معمولاً از فاصلهٔ اقلیدسی برای محاسبهٔ فاصله استفاده می‌شود.
۳) محاسبهٔ مراکز جدید: مراکز خوشه‌ها با میانگین نقاط اختصاص‌داده‌شده به هر خوشه محاسبه می‌شود.
٤) تکرار مراحل ۲ و ۳: این مراحل تا زمانی تکرار می‌شوند که مراکز خوشه‌ها تغییر نکنند یا تغییرات بسیار کم باشد. فرمول اصلی این روش که در رابطهٔ ۳ دیده می‌شود، بر اساس کمینه‌سازی مجموع مربعات فاصله‌ها بین نقاط و مراکز خوشه تخصیص‌یافته می‌باشد:

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} \left\| x_i^{(j)} - c_j \right\|^2 \qquad (۱)$$

در رابطهٔ (۱):

- k نشان‌دهندهٔ تعداد خوشه‌هاست.
- n تعداد نمونه‌ها یا داده‌های موجود در هر خوشه را نشان می‌دهد.
- $x_i^{(j)}$ به هر نمونه در خوشهٔ j اشاره دارد.
- c_j مرکز خوشهٔ j است.

[1] Xu et al.
[2] Unsupervised Learning

- $\left\| x_i^{(j)} - c_j \right\|^2$ فاصلهٔ اقلیدسی بین نمونهٔ $x_i^{(j)}$ و مرکز خوشهٔ c_j است.

هدف این تابع یافتن مراکز خوشه‌ها به‌گونه‌ای است که مجموع فاصله‌ها بین هر نمونه و مرکز خوشهٔ خود حداقل شود. با نگاشت داده‌ها به این محورهای جدید، داده‌ها با کمترین تعداد ممکن از مؤلفه‌ها نمایش داده می‌شوند و تحلیل‌های بعدی با دقت و سرعت بیشتری انجام می‌شوند.

۳/۲ معیار ارزیابی مدل‌های خوشه‌بندی

شاخص دیویس بولدین یکی از معیارهای ارزیابی مدل‌های خوشه‌بندی است که دیویس و بولدین در ۱۹۷۹ آن را معرفی کردند. این شاخص براساس فاصلهٔ بین خوشه‌ها و میزان پراکندگی درون‌خوشه‌ای تعریف می‌شود و هدف آن، کمینه‌سازی میزان شباهت بین خوشه‌ها و افزایش تمایز آن‌هاست. شاخص دیویس بولدین بر این اصل استوار است که خوشه‌های بهینه باید دارای واریانس درونی کم و فاصلهٔ بین‌خوشه‌ای زیادی باشند. مقدار کمتر این شاخص نشان‌دهندهٔ مدل‌های بهتر و خوشه‌بندی مطلوب است.

$$DB = \frac{1}{n_c}\sum_{i=1}^{n_c} R_i \qquad (۲)$$
$$R_i = \max_{j=1,\dots,nc, j\neq i}\left(R_{ij}\right),\ i = 1, \dots, n$$

در رابطهٔ (۲)؛

- R_i برای هر خوشهٔ i نشان‌دهندهٔ بیشترین فاصله بین خوشه و سایر خوشه‌هاست. این فاصله نمایانگر نزدیک‌ترین فاصله بین خوشه و نزدیک‌ترین خوشهٔ دیگر است که در فرمول به‌صورت زیر تعریف می‌شود:

$$R_i = \max_{j=1,\dots,nc, j\neq i}\left(R_{ij}\right),\ i = 1, \dots, n$$

در آن R_{ij} نشان‌دهندهٔ فاصلهٔ بین مرکز خوشهٔ i و مرکز خوشهٔ j است.

- n نشان‌دهندهٔ تعداد خوشه‌هایی است که در مجموعه داده موردبررسی قرارگرفته‌اند.
- $\frac{1}{n_c}\sum_{i=1}^{n_c} R_i$ میانگین بیشترین فاصلهٔ بین خوشه‌هاست.

۳/۳ پیش‌بینی

پیش‌بینی در داده‌کاوی به‌معنای استفاده از مدل‌های آماری و یادگیری ماشین برای پیش‌بینی خروجی‌های ناشناخته بر اساس داده‌های گذشته است. در زمینهٔ تحلیل مالی و اعتباری،

پیش‌بینی نکول تسهیلات ازجمله کاربردهای رایج است که از مدل‌های مختلف برای این منظور استفاده می‌شود. هدف اصلی پیش‌بینی در این حوزه، تشخیص مشتریانی است که احتمال بیشتری برای نکول تسهیلات دارند و این امر به بانک‌ها و مؤسسات مالی کمک می‌کند تا ریسک‌های اعتباری خود را مدیریت کنند. مدل‌های پیش‌بینی از دهۀ ۱۹۵۰ با معرفی مدل‌های رگرسیون لاجیت به‌صورت گسترده مورداستفاده قرار گرفتند.

۳/۳/۱ مدل لاجیت

رگرسیون لاجیت یکی از اولین و معروف‌ترین مدل‌های پیش‌بینی در این زمینه است که توسط دیوید کاکس در ۱۹۵۸ معرفی شد. مدل لاجیت برای پیش‌بینی احتمال وقوع یک رخداد (نظیر نکول تسهیلات) بر اساس یک یا چند متغیر مستقل استفاده می‌شود. فرمول کلی مدل لاجیت به‌صورت زیر است:

$$ln\left(\frac{p}{1-p}\right) = \beta 0 + \beta 1X1 + \beta 2X2 + \cdots + \beta nXn \qquad (۳)$$

در این مدل، $ln\left(\frac{p}{1-p}\right)$ نسبت احتمال وقوع بر عدم‌وقوع یک رویداد (نکول یا عدم‌نکول) را نشان می‌دهد و x متغیرهای توضیحی (ویژگی‌های مشتریان) است. مزیت این مدل در سادگی و شفافیت تفسیر نتایج آن است؛ اما یکی از محدودیت‌های آن ناتوانی در مدیریت داده‌های بزرگ و پیچیده است.

۳/۳/۲ مدل XGBoost

با گسترش حجم داده‌ها و پیچیدگی آن‌ها، روش‌های نوین‌تری مانند XGBoost معرفی شدند. این مدل را چن در ۲۰۱۶ معرفی کرد و به‌سرعت به یکی از محبوب‌ترین مدل‌های پیش‌بینی تبدیل شد. این روش بر پایۀ تکنیک بوستینگ[1] بنا شده است و از ترکیب چندین مدل ضعیف‌تر (مانند درخت‌های تصمیم) برای ایجاد یک مدل قوی‌تر استفاده می‌کند. مدل فوق قادر به مدیریت داده‌های بزرگ و پیچیده است و به‌دلیل استفاده از تکنیک‌های بهینه‌سازی؛ مانند کاهش درجۀ دوم، دقت بالایی در پیش‌بینی دارد. در پیش‌بینی نکول تسهیلات در هر خوشه، استفاده از XGBoost و رگرسیون لاجیت به‌دلیل ترکیب مزایای هر دو روش می‌تواند رویکرد مؤثری باشد. این ترکیب به تحلیلگران امکان می‌دهد تا با دقت بیشتری ریسک نکول را در هر خوشه شناسایی کنند و از نتایج آن برای مدیریت بهینۀ ریسک

[1] Boosting

و تدوین راهبردهای مناسب استفاده کنند. تابع هدف کلی XGBoost شامل دو قسمت است:

- زیان پیش‌بینی‌ها[1]: تفاوت بین مقدار واقعی و مقدار پیش‌بینی‌شده که به کمک یک تابع زیان محاسبه می‌شود.

$$\Omega(f_t) + \sum_{i=1}^{n}\left[g_i f_t(X_i) + \frac{1}{2}h_i f_t^2(X_i)\right] \approx Obj^{(t)} \quad (۴)$$

در آن:

g_i گرادیان تابع زیان به پیش‌بینی در مرحلهٔ قبلی است.

h_i، مشتق دوم تابع زیان به پیش‌بینی در مرحلهٔ قبلی است.

- جریمهٔ پیچیدگی مدل[2]: برای جلوگیری از بیش‌برازش[3]، پیچیدگی مدل با استفاده از یک تابع منظم‌سازی محاسبه و مدل جریمه می‌شود.

$$\Omega(f_t) = \lambda_j^2 \sum_{j=1}^{T} w\frac{1}{2} + \gamma T \quad (۵)$$

۱) γ: پارامتر جریمه برای تعداد برگ‌های درخت[4]

۲) λ: پارامتر منظم‌سازی برای وزن‌های برگ‌های درخت

۳) w_i: وزن‌های برگ‌های درخت

۴ داده‌ها و متدولوژی

هدف اصلی این پژوهش ارائهٔ مدلی جامع برای ارزیابی عملکرد تسهیلات بانکی از منظر احتمال نکول و پیش‌بینی وضعیت اعتباری متقاضیان قبل از اعطای تسهیلات است. در این راستا، اطلاعات مربوط به ۲۰،۰۰۰ نفر از مشتریان حقیقی تمام بانک‌ها از قبیل بانک‌های دولتی، خصوصی، و قرض‌الحسنه مورداستفاده قرار گرفته است.[5] روش انجام این تحقیق شامل تحلیل خوشه‌بندی به‌وسیلهٔ الگوریتم‌های میانگین کا (K-means) و خوشه‌بندی

---

[1] training loss

[2] regularization term

[3] overfitting

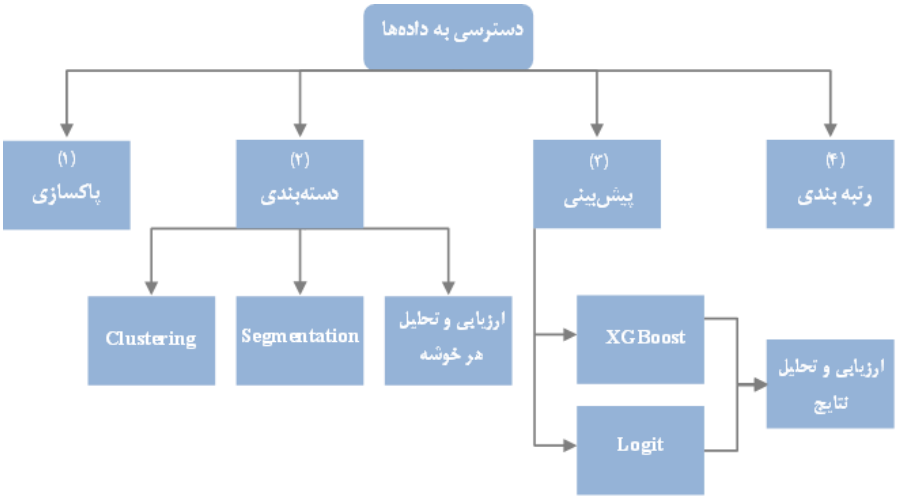
[4] complexity penalty

[5] داده‌های فوق با همکاری شرکت **مشاورهٔ رتبه‌بندی اعتباری ایران** تهیه شده است که به‌عنوان تنها شرکت اعتبارسنجی تحت نظارت بانک مرکزی جمهوری اسلامی شناخته می‌شود.

هیبریدی با ترکیب روش‌های K-means و ژنتیک برای دسته‌بندی مشتریان براساس ویژگی‌های اعتباری و مالی است. سپس، برای پیش‌بینی احتمال نکول در هر خوشه از مدل‌های یادگیری ماشینی مانند XGBoost و مدل لاجیت استفاده شده است. درنهایت، خوشه‌ها با استفاده از روش‌های چندمعیاره برای رتبه‌بندی و ارزیابی عملکرد تسهیلات مورد بررسی قرار گرفته‌اند.

۴/۱ آماده‌سازی و پیش‌پردازش داده‌ها

جامعهٔ آماری این پژوهش به‌طور تصادفی از میان متقاضیانی انتخاب شده است که در سال‌های اخیر (۱۳۹۵ـ۱۴۰۰) از بانک تسهیلات دریافت کرده‌اند. اطلاعات تراکنشی و جمعیت‌شناختی این مشتریان برای تحلیل داده‌ها مورداستفاده قرار گرفته است. روند انجام پژوهش در شکل ۱ شرح داده شده است:



شکل ۱. گام‌های پژوهشی فرایند مدل‌سازی

در فرایند مدل‌سازی و تحلیل داده‌ها، یکی از مراحل حیاتی پاک‌سازی و پیش‌پردازش داده‌هاست. این مرحله برای افزایش دقت نتایج و بهبود کارایی مدل‌های مورداستفاده انجام می‌گیرد. در این پژوهش، پس از جمع‌آوری داده‌های اولیه، اقدامات زیر به‌صورت سیستماتیک در راستای آماده‌سازی داده‌ها برای مراحل بعدی صورت پذیرفت.

۴/۲ کنترل و تعدیل مقادیر پرت

وجود مقادیر پرت می‌تواند اثرات زیادی بر نتایج مدل‌سازی داشته باشد. در این پژوهش برای مواجهه با این چالش، از رویکرد چارک‌بندی بهره گرفته شد. در این روش، مقادیر خارج از محدودهٔ چارک‌های اول و سوم شناسایی و به تناسب تعدیل شدند. با اتخاذ این رویکرد، تعادل بین مقادیر مختلف داده‌ها حفظ شد و اثرات منفی مقادیر ناهنجار کاهش یافت. پس از اجرای رویکرد چارک‌بندی و حذف برخی مقادیر پرت تعداد مشاهدات از ۲۰,۰۰۰ به ۱۹,۷۰۰ مشتری رسید.

۴/۳ نرمال‌سازی و مقیاس‌گذاری داده‌ها

باتوجه به دامنه‌های متفاوت مقادیر در متغیرهای مختلف، برای همگن‌سازی و جلوگیری از بروز مشکلات ناشی از تفاوت در مقیاس داده‌ها فرایند نرمال‌سازی روی داده‌ها اعمال شد. در این فرایند، تمام مقادیر به بازه‌ای مشخص بین ۰ و ۱ تبدیل شدند. این اقدام نه‌تنها به همسان‌سازی مقادیر منجر شد، بلکه زمینه را برای اجرای دقیق‌تر و کارآمدتر الگوریتم‌های دسته‌بندی و پیش‌بینی فراهم کرد. بدین‌ترتیب، داده‌های آماده‌سازی‌شده با حفظ دقت و یکپارچگی وارد مراحل بعدی تحلیل شدند.

۴/۴ تعریف متغیر هدف

برای ساخت مدل ارتقای کارایی اعطای تسهیلات، در ابتدا، لازم است تعریف مناسبی از متغیر هدف (وقوع یا عدم‌وقوع نکول) تعیین کنیم. در این پژوهش، برای انتخاب قرارداد برای تعیین وضعیت نکول شخص، پارامترهای مختلفی بررسی شده است که در جدول ۲ به توضیح هریک می‌پردازیم.

جدول ۲

پارامترهای موردبررسی و قیود مشخص برای استخراج متغیر هدف

ردیف	پارامتر	نوع محدودیت
۱	نقش متقاضی در قرارداد	متقاضی اصلی، شریک
۲	شروع بازه زمانی دریافت اعتبار	۱۴۰۰/۹/۱
۳	پایان بازه زمانی بازپرداخت اعتبار	۱۴۰۱/۱۱/۱۱
۴	تعداد ماه برای محاسبهٔ ارزیابی وضعیت قراردادها	۳۶ ماه
۵	حداقل تعداد ثبت وضعیت (snapshot) برای هر قرارداد در بازه عملکرد	حداقل ۹ اسنپشات در هر قرارداد

بعد از تعیین پارامترها در فرایند مدل‌سازی، طبق قانون بازل، تعریف نکول مشخص خواهد بود.

۴/۵ انتخاب ویژگی‌ها پس از کاهش ابعاد

مرحلهٔ بسیار مهم دیگر پس از تعریف مناسب متغیر هدف که در فرایند مدل‌سازی انجام شد، انتخاب متغیرهای مستقل بود که برای ارتقای کارایی مدل در دسته‌بندی مشتریان و پیش‌بینی نکول و تخصیص بهینهٔ تسهیلات نقش کلیدی دارد. برای شناسایی و استخراج این متغیرها، شاخص‌هایی از حوزه‌های گوناگون ازجمله تسهیلات بانکی، چک‌های برگشتی، و صدک درآمدی تعیین شد. فرایند انتخاب این شاخص‌ها با برگزاری جلسات کارشناسی و کارگروه‌هایی با حضور متخصصان حوزهٔ اعتبارسنجی و نخبگان علمی به انجام رسید. درنتیجهٔ این بررسی‌ها، ۳۳ شاخص نهایی به‌عنوان متغیرهای اصلی مدل انتخاب شدند. پس از آماده‌سازی داده‌ها، برای کاهش ابعاد و افزایش کارایی مدل از تکنیک تحلیل مؤلفه‌های اصلی (PCA) استفاده شد. این روش به کاهش تعداد ابعاد داده کمک کرده و درعین‌حال اطلاعات اصلی داده‌ها را حفظ کرد. پس از اعمال PCA، تعداد ابعاد به ۶ بعد اصلی کاهش یافت. در ادامه، جدول ۳ برای معرفی متغیرهای استفاده‌شده در مدل ارائه شده است:

جدول ۳

معرفی ویژگی‌های مورداستفاده در مدل‌سازی

دسته بندی	ویژگی‌ها
داده‌های تسهیلاتی	مبلغ کل تسهیلات در ۳۶ ماه گذشته
	مبلغ تسهیلات دارای وضعیت منفی در ۳۶ ماه گذشته

	مبلغ تسهیلات در جریان شخص در ۳۶ ماه گذشته
داده‌های مربوط به وضعیت قراردادها	تعداد قراردادهای شخص که دارای وضعیت منفی هستند.
	تعداد قراردادهای شخص که فاقد وضعیت منفی هستند.
	تعداد ماه‌هایی که قرارداد شخص در وضعیت منفی قرار دارد.
	تعداد ماه‌هایی که قرارداد شخص فاقد وضعیت منفی است.
	تعداد بانک‌هایی که شخص در آن دارای قرارداد با وضعیت منفی است.
	تعداد بانک‌هایی که شخص در آن دارای قرارداد بدون وضعیت منفی است.
داده‌های مربوط به چک‌های برگشتی	مبلغ چک‌های برگشتی شخص در ۳۶ ماه گذشته
	تعداد چک‌های برگشتی در ۳۶ ماه گذشته
	تعداد بانک‌های دارای چک برگشتی
شاخص‌های مالی و اعتباری	تعداد روزهای گذشته از آخرین اسنپ‌شات با مبلغ سررسیدشدهٔ پرداخت‌نشده
	تعداد ماه‌های بدون تأخیر در پرداخت در ۳۶ ماه گذشته
	تعداد قراردادهایی که در ۳۶ ماه گذشته دارای بدهی معوقه بودند.
	نسبت میانگین بدهی‌های آتی ۶ ماه اخیر به میانگین بدهی آتی ۷ الی ۱۲ ماه اخیر
داده‌های جمعیت‌شناختی	سن اشخاص
	وضعیت شغلی شخص (شاغل، بازنشسته، نامعلوم)
	صدک درآمدی
	جنسیت

۵ مدل و نتایج

در بخش بعدی مدل‌سازی جهت ارتقای کارایی اعطای تسهیلات، برای شناسایی الگوهای رفتاری و تحلیل مشتریان از روش‌های مختلف دسته‌بندی استفاده شده است.

۵/۱ خوشه‌بندی با روش K-means

یکی از رایج‌ترین روش‌های خوشه‌بندی مشتریان که در این پژوهش نیز مورداستفاده قرار گرفته الگوریتم میانگین کا (K-means) است. برای تعیین تعداد بهینهٔ خوشه‌ها (k)، از روش آرنج[1] بهره گرفته شد. پس از انجام تحلیل‌های لازم و بررسی‌های مکرر، تعداد بهینهٔ خوشه‌ها ۴ تعیین شد. در این الگوریتم، هر مشتری به خوشه‌ای اختصاص داده می‌شود که فاصلهٔ کمتری از مرکز آن خوشه دارد. فرایند تنظیم و بهینه‌سازی پارامتر k چندین‌بار انجام شد تا بهترین تقسیم‌بندی حاصل شود. نتایج این خوشه‌بندی نشان داد که مشتریان به‌طور

[1] Elbow Method

طبیعی به ۴ گروه مجزا تقسیم می‌شوند که هر گروه شامل مشتریانی با ویژگی‌های رفتاری و اعتباری مشابه است. این گروه‌ها به شکلی معنادار رفتار اعتباری مشترک میان اعضای خوشه‌ها را نمایش می‌دهند و این امر امکان تحلیل دقیق‌تر هر خوشه را فراهم می‌کند. درنهایت، استفاده از روش آرنج و ارزیابی مستمر به انتخاب تعداد خوشهٔ بهینه و ارائهٔ تقسیم‌بندی منجر شد که با درنظرگرفتن شباهت‌های میان مشتریان، آن‌ها را به گروه‌هایی با ویژگی‌های یکسان و رفتار اعتباری مشابه دسته‌بندی کرد. هر خوشه بر اساس ویژگی‌های رفتاری و اعتباری مشتریان تقسیم‌بندی شده است. جزئیات این خوشه‌بندی در جدول ۴ آورده شده است.

جدول ۴

توصیف آماری ویژگی‌های مالی در هر خوشهٔ به‌دست‌آمده از روش K-means

ردیف	ویژگی‌های مالی	میانگین در خوشهٔ ۱	میانگین در خوشهٔ ۲	میانگین در خوشهٔ ۳	میانگین در خوشهٔ ۴
۱	مبلغ کل تسهیلات در ۳۶ ماه گذشته (میلیون ریال)	۴۶۰,۴۶۴,۱۴۸	۱۴۱,۳۸۰,۲۸۶	۳۲۳,۸۴۲,۲۰۵	۲۳۵,۷۶۴,۰۰۶
۲	مبلغ تسهیلات با وضعیت منفی در ۳۶ ماه گذشته (میلیون ریال)	۱۲,۹۵۴,۴۰۳	۴,۵۰۵,۰۳۰	۲۰,۳۷۶,۵۷۲	۱۴۰,۹۲۳,۹۱۶
۳	تعداد قراردادهای دارای وضعیت منفی شخص	۰/۰۹۷	۰/۰۸۲	۰/۱۹۳	۰/۹۳۱
٤	تعداد قراردادهای فاقد وضعیت منفی شخص	۲.۶۷۱	۲/۱۴۰	۲/۱۱۴	۲/۹۲۲
٥	تعداد ماه‌های با قرارداد وضعیت منفی شخص	۰/۱۹۱	۰/۲۱۸	۰.۴۵۲	۲.۵۷۶
٦	تعداد ماه‌های با قرارداد فاقد وضعیت منفی شخص	۹/۸۶۴	۸/۹۳۸	۹/۲۵۲	۹/۳۴۴
۷	تعداد بانک‌های دارای قرارداد با وضعیت منفی	۰/۰۸۸	۰/۰۷۶	۰/۱۷۰	۰/۸۳۳
۸	تعداد بانک‌های فاقد قرارداد با وضعیت منفی	۱/۸۹۷	۱/۵۲۲	۱/۵۵۰	۲/۰۹۷
۹	مبلغ چک‌های برگشتی شخص در ۳۶ ماه گذشته (میلیون ریال)	۲۰۳,۹۱۱,۲۰۱	۵۵,۷۵۲,۳۳۳	۲۴,۰۶۲,۵۵۵	۲۰۱,۰۷۷,۸۳۹
۱۰	تعداد چک‌های برگشتی در ۳۶ ماه گذشته	۱/۰۷۶	۰.۴۰۵	۰/۲۰۴	۱/۵۴۸
۱۱	تعداد اعضای هر خوشه	۴۴۸۵	۴۴۸۷	۶۳۷۲	۳۸۳۶

باتوجه‌به داده‌های جدول ۴، خوشهٔ ۱ با ۴۴۸۵ عضو از نظر مالی پایدارتر به‌نظر می‌رسد. اعضای این خوشه بیشترین مبلغ کل تسهیلات درخواستی در ۳۶ ماه گذشته را با میانگین ۴۶۰,۴۶۴,۱۴۸ میلیون ریال دارند که نشان می‌دهد این گروه از نظر درخواست تسهیلات فعالیت بیشتری دارد. همچنین، اعضای خوشهٔ ۱ کمترین تعداد ماه‌های با قرارداد وضعیت منفی میانگین ۰/۱۹۱ ماه و بالاترین تعداد ماه‌های با قرارداد فاقد وضعیت منفی حدود ۱۰ ماه را دارند. این موضوع نشان می‌دهد که اعضای خوشهٔ ۱ رفتار مالی باثبات‌تری دارند و کمتر با مشکلات اعتباری مواجه می‌شوند. خوشهٔ ۲ با ۴۴۸۷ عضو کمترین میانگین مبلغ کل تسهیلات درخواستی؛ یعنی ۱۴۱,۳۸۰,۲۸۶ میلیون ریال را در بین خوشه‌ها دارد که شاید به‌دلیل ریسک پایین‌تر یا نیاز کمتر اعضای این خوشه به تسهیلات بالا باشد. باوجوداین، تعداد ماه‌های با قرارداد وضعیت منفی در این خوشه با میانگین ۰/۲۱۸ ماه بالاتر از خوشهٔ ۱ است؛ اما همچنان از خوشه‌های دیگر پایین‌تر است. این خوشه کمترین تعداد چک‌های برگشتی (میانگین ۰/۴۰۵ چک) را بین خوشه‌ها دارد که نشان‌دهندهٔ مدیریت مالی بهتر و ریسک اعتباری نسبتاً پایین‌تر است.

خوشهٔ ۳ با ۶۳۷۲ عضو، از نظر مبلغ تسهیلات با وضعیت منفی (میانگین ۲۰,۳۷۶,۵۷۲ میلیون ریال) در میانه قرار دارد و نشان می‌دهد که بخشی از اعضای این خوشه با مشکلات اعتباری مواجه‌اند .تعداد قراردادهای دارای وضعیت منفی در این خوشه با میانگین ۰/۱۹۳ قرارداد نسبتاً بالاست و تعداد ماه‌های با قرارداد وضعیت منفی نیز (میانگین ۰/۴۵۲ ماه) بیشتر از خوشه‌های ۱ و ۲ است. خوشهٔ ۴ با ۳۸۲۶ عضو، بالاترین ریسک اعتباری را بین خوشه‌ها دارد. این گروه دارای بیشترین مبلغ تسهیلات با وضعیت منفی با میانگین ۱۴۰,۹۲۳,۹۱۶ میلیون ریال و بالاترین تعداد قراردادهای دارای وضعیت منفی (میانگین ۰/۹۳۱ قرارداد) است که نشان می‌دهد اعضای این خوشه بیشترین احتمال ناتوانی در پرداخت تعهدات خود را دارند. تعداد چک‌های برگشتی نیز در این خوشه بیشترین مقدار (میانگین ۱/۵۴۸ چک) را دارد که نشانهٔ دیگری از وضعیت مالی نامطلوب و عدم‌پایبندی اعضا به تعهدات مالی‌شان است. به‌طور خلاصه، خوشهٔ ۱ نمایانگر گروهی با رفتار مالی پایدار و کمترین ریسک اعتباری است؛ درحالی‌که خوشهٔ ۴ بیشترین ریسک اعتباری را دارد و اعضای آن با مشکلات متعددی در پرداخت تعهدات مالی مواجه‌اند. خوشه‌های ۲ و ۳ در میانهٔ این دو قرار دارند و رفتارهای مالی متفاوتی از خود نشان می‌دهند.

جدول ۵
وضعیت متغیر هدف در هرخوشه به‌دست‌آمده از روش K-means

ردیف	تعداد خوشه	تعداد مشاهدات	وضعیت نکول
۱	خوشهٔ اول	۴۴۸۵	۰/۱۹۲
۲	خوشهٔ دوم	۴۴۸۷	۰/۱۹۴
۳	خوشهٔ سوم	۶۳۷۲	۰/۲۷۱
۴	خوشهٔ چهارم	۳۸۲۶	۰/۵۱۵

جدول ۵ نشان‌دهندهٔ وضعیت متغیر هدف (نرخ نکول) در هر خوشه است که تفاوت‌های مهمی میان خوشه‌ها از نظر ریسک اعتباری افراد مشخص می‌کند. خوشهٔ چهارم با نرخ نکول ۰/۵۱۵ بالاترین میزان نکول را میان خوشه‌ها دارد که نشان‌دهندهٔ بیشترین ریسک اعتباری و وضعیت مالی نامطلوب افراد این گروه است. این خوشه به‌طور واضح پرریسک‌ترین گروه بوده و رفتار اعتباری نامناسب‌تری در مقایسه با سایر خوشه‌ها دارد. در مقابل، خوشهٔ اول با نرخ نکول ۰/۱۹۲ کمترین میزان نکول را دارد که نشان‌دهندهٔ رفتار اعتباری پایدارتر و ریسک کمتر افراد این گروه است. افراد این خوشه از نظر مالی عملکرد بهتری دارند و از ریسک پایین‌تری نسبت به دیگر خوشه‌ها برخوردارند. خوشهٔ دوم با نرخ نکول ۰/۱۹۴ عملکردی مشابه خوشهٔ اول دارد و جزو خوشه‌های با ریسک پایین‌تر قرار می‌گیرد. خوشهٔ سوم نیز با نرخ نکول ۰/۲۷۱ به نسبت خوشه‌های اول و دوم ریسک بالاتری دارد؛ اما همچنان در مقایسه با خوشهٔ چهارم وضعیت بهتری از نظر ریسک اعتباری دارد. این تحلیل نشان می‌دهد که خوشه‌های اول و دوم به‌عنوان خوشه‌های کم‌ریسک‌تر و با رفتار مالی باثبات‌تر شناسایی می‌شوند، درحالی‌که خوشهٔ چهارم بیشترین ریسک اعتباری را داراست و به توجه ویژه‌ای در فرایندهای اعتبارسنجی نیاز دارد.

۵/۲ خوشه‌بندی با روش ترکیبی

علاوه بر روش K-means، در این پژوهش از یک روش هیبریدی نیز استفاده شده است که ترکیبی از الگوریتم ژنتیک و K-means است. این روش به‌دلیل توانایی‌های منحصربه‌فرد هریک از این دو الگوریتم انتخاب شده است. در این روش هیبریدی، ابتدا الگوریتم ژنتیک برای پیداکردن بهترین نقاط شروع خوشه‌بندی به‌کار گرفته شده است. این الگوریتم با استفاده از فرایندهای انتخاب طبیعی و جهش، به‌تدریج بهترین موقعیت‌های اولیه برای الگوریتم میانگین کا (K-means) را شناسایی می‌کند. فرایند به این‌صورت انجام شده است

که ابتدا تابع هدفی برای ارزیابی کیفیت خوشه‌بندی تعریف شده است. این تابع هدف بر اساس معیار مجموع مربعات درون خوشه‌ها طراحی شد که معیاری برای سنجش تراکم خوشه‌ها محسوب می‌شود. هدف از این تابع، کمینه‌کردن مجموع مربعات وزنی مشاهدات درون خوشه برای به‌دست‌آوردن خوشه‌هایی با بیشترین شباهت درونی بود. سپس الگوریتم ژنتیک برای جست‌وجوی تعداد بهینهٔ خوشه‌ها در بازه مشخص ۲۰ خوشه اجرا شد. الگوریتم با استفاده از جمعیت اولیه‌ای از تعداد خوشه‌ها شروع به‌کار کرد و به‌طور مکرر این تعداد را براساس عملکرد تابع هدف به‌روز کرد. درنهایت، پس از ۱۰۰ تکرار از فرایند جست‌وجو، بهترین تعداد خوشه انتخاب شد که به کمترین مقدار مجموع مربعات وزنی بین خوشه‌ای منجر شد. نتیجهٔ بهینهٔ این جست‌وجو، ۱۰ خوشه ۱۰ بود که بر اساس تابع هدف تعیین شد.

۵/۳ تحلیل و ارزیابی نتایج دسته‌بندی مشتریان

برای ارزیابی و مقایسهٔ عملکرد مدل‌های خوشه‌بندی از شاخص دیویس_بولدین استفاده شده است. این شاخص یکی از معیارهای استاندارد برای سنجش کیفیت خوشه‌بندی است که مقدار کمتر آن نشان‌دهندهٔ تمایز بهتر خوشه‌ها و همگنی بیشتر نقاط درون هر خوشه است.

جدول ۶

مقایسهٔ ارزیابی نتایج دسته‌بندی مشتریان

ردیف	الگوریتم	شاخص DB
۱	K-means	۱/۳۵۱
۲	K-means +GA	۱/۴۳۹

همان‌طور که در جدول ۶ دیده می‌شود، دو روش خوشه‌بندی موردارزیابی قرارگرفتند:

۱) K-means: مقدار شاخص دیویس_بولدین برای این روش برابر با ۱/۳۵۱ بود.

۲) الگوریتم هیبریدی (ترکیب K-means + GA): مقدار شاخص دیویس_بولدین در این روش برابر با ۱/۴۳۹محاسبه شد.

باتوجه‌به نتایج، روش الگوریتم میانگین کا (K-means) عملکرد بهتری به روش هیبریدی از نظر این شاخص دارد، چراکه مقدارکمتری برای شاخص دیویس_بولدین کسب کرده است. بر همین اساس، خوشه‌بندی نهایی در این پژوهش بر مبنای نتایج الگوریتم میانگین کا (K-means) انجام شد تا دقت و کیفیت بیشتری در تفکیک داده‌ها حاصل شود.

۵/۴ پیش‌بینی نکول در هر خوشه

در ادامهٔ فرایند مدل‌سازی پس از دسته‌بندی مشتریان بانکی به‌واسطهٔ ویژگی‌های تسهیلات، به بررسی پیش‌بینی ریسک نکول در هر خوشه با استفاده از دو مدل XG-Boost و رگرسیون لاجیت پرداخته شده است. عملکرد هر مدل بر اساس معیارهایی مانند سطح زیر منحنی[1] و اختصاصی‌بودن[2] در چهار خوشهٔ مشتری به‌دست‌آمده با الگوریتم میانگین کا (K-means) ارزیابی شد. در ادامه به‌تفصیل، نتایج هر مدل در خوشه‌های مختلف توضیح داده می‌شود.

۵/۴/۱ پیش‌بینی با مدل XG-Boost

در این پژوهش XG-Boost، به‌عنوان یک الگوریتم گرادیان بوستینگ قدرتمند، برای پیش‌بینی ریسک نکول در چهار دسته به‌کار رفته است. معیارهای عملکرد این مدل نشان‌دهندهٔ توانایی آن در شناسایی روابط پیچیده در داده‌هاست که معمولاً از مدل‌های ساده‌تری مانند رگرسیون لاجیت بهتر عمل می‌کند. نتایج به‌صورت زیر بودند:

جدول ۷

ارزیابی نتایج مدل XG-Boost بر تمام مشتریان

ردیف	مدل پیش‌بینی	تعداد مشتریان	معیار AUC	معیار Specificity
۱	XG-Boost	۱۹۱۷۰	۰/۸۳۶	۰/۸۹۳

این نتایج نشان‌دهندهٔ عملکرد کلی مدل در پیش‌بینی نکول بر اساس ویژگی‌های موجود است. در مرحلهٔ بعد، پس از دسته‌بندی داده‌ها، مدل به‌صورت جداگانه روی هر دسته از مشتریان اجرا شد. این کمک می‌کند تا میزان مؤثربودن دسته‌بندی مشتریان بر بهبود دقت مدل پیش‌بینی بررسی شود. نتایج به‌دست‌آمده از هر دسته به شرح زیر است:

[1] Area under the curve (AUC)

[2] specificity

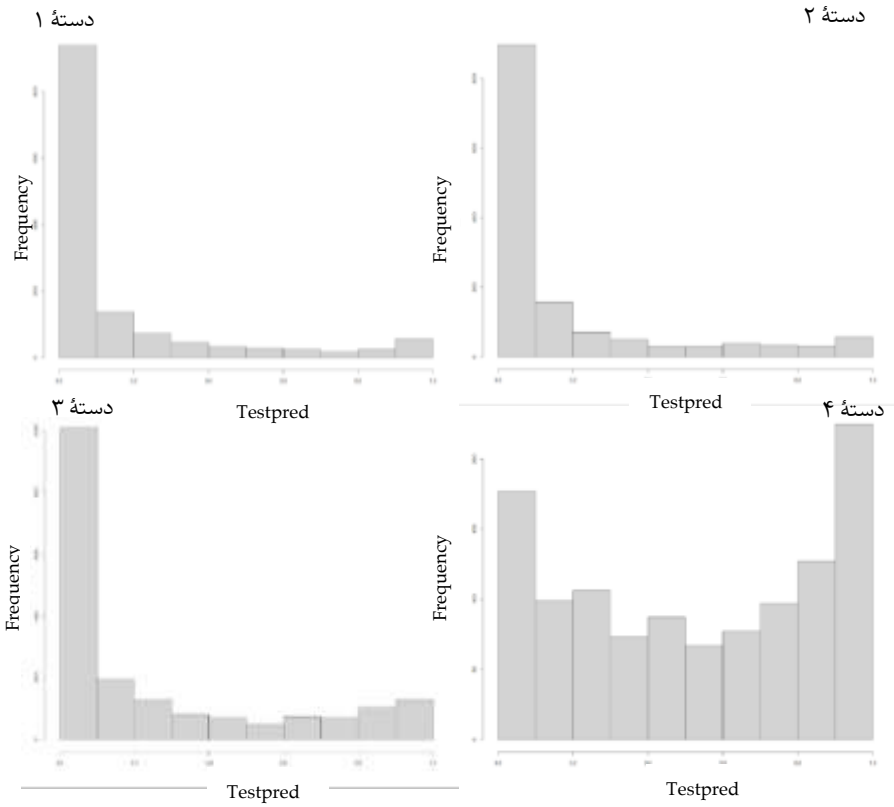
جدول ۸

ارزیابی نتایج مدل XG-Boost روی هر دسته از مشتریان

ردیف	دسته مشتریان	تعداد مشتریان	معیار AUC	معیار Specificity
۱	دستهٔ اول	۴۴۸۵	۰/۷۶۰	۰/۹۴۳
۲	دستهٔ دوم	۴۴۸۷	۰/۸۰۷	۰/۹۳۴
۳	دستهٔ سوم	۶۳۷۲	۰/۸۲۵	۰/۸۹۸
٤	دستهٔ چهارم	۳۸۲۶	۰/۷۴۲	۰/۶۷۰

باتوجه‌به نتایج به‌دست‌آمده از مدل که در جدول ۸ ذکر شده است، می‌توان گفت که دسته‌بندی به بهبود دقت پیش‌بینی نکول کمک کرده و مدل توانسته است عملکرد بهتری روی هر دسته نسبت به کل داده‌ها ارائه دهد. AUC بالاتر از ۰/۷ در تمام دسته‌ها نشان‌دهندهٔ عملکرد مناسب مدل در تفکیک نکول‌کنندگان از سایر مشتریان است.

شکل ۲، توزیع نکول پیش‌بینی‌شده به روش XG-Boost را در چهار دستهٔ مختلف مشتریان به تصویر می‌کشد. در دستهٔ اول (بالا سمت چپ)، توزیع به‌سمت چپ متمایل است، به این معنی که اکثر مشتریان در این دسته دارای ریسک نکول پایین‌اند و تنها تعداد محدودی از آن‌ها با ریسک نکول بالا مواجه‌اند. این نشان می‌دهد که این دسته به‌طور عمده شامل مشتریانی با اعتبار قوی است. در دستهٔ دوم (بالا سمت راست)، توزیع مشابهی مشاهده می‌شود که نشان‌دهندهٔ تمرکز بیشتر ریسک نکول در مقادیر پایین‌تر و تعداد اندکی مشتریان با ریسک نکول بالاست. دستهٔ سوم (پایین سمت چپ) توزیعی پراکنده‌تر دارد، به‌طوری که مشتریان این دسته به‌طور گسترده‌تری از نظر ریسک نکول توزیع شده‌اند. این الگو نشان می‌دهد که ریسک نکول در این گروه متنوع‌تر است. درنهایت، در دستهٔ چهارم (پایین سمت راست)، توزیع به‌سمت راست چولگی دارد که نشان‌دهندهٔ وجود تعداد زیادی مشتری با ریسک متوسط یا بالای نکول است. این الگو بیانگر افزایش احتمال نکول در این دسته از مشتریان است.



شکل ۲. توزیع نکول پیش‌بینی‌شده با روش XG-Boost در هر دسته از مشتریان

۵/۴/۲ پیش‌بینی با مدل لاجیت

علاوه بر پیش‌بینی با مدل XG-Boost، برای پیش‌بینی ریسک نکول در این پژوهش از مدل رگرسیون لاجیت نیز استفاده شده است. مدل لاجیت یکی از مدل‌های پایه در تحلیل‌های پیش‌بینی با فرض وجود روابط خطی میان متغیرهای مستقل و احتمال وقوع رویداد (در اینجا نکول)، ارائه می‌شود. این مدل به‌دلیل سادگی در تفسیر نتایج، برای تحلیل‌گران جذاب است؛ اما به‌دلیل ناتوانی در شناسایی روابط غیرخطی و پیچیده، عملکرد آن به‌طورکلی، ضعیف‌تر از مدل‌هایی مانند XG-Boost است.

در این پژوهش، متغیر وابسته به‌عنوان وضعیت نکول مشتریان و متغیرهای مستقل شامل تمامی ویژگی‌های مرتبط با مشتریان برای مدل‌سازی انتخاب شده‌اند. نتایج ارزیابی کلی مدل لاجیت بر تمام مشتریان در جدول ۹ ارائه شده است:

جدول ۹

ارزیابی نتایج مدل لاجیت برکلیهٔ مشتریان

ردیف	مدل پیش‌بینی	تعداد مشتریان	معیار AUC	معیار Specificity
۱	Logit	۱۹۱۷۰	۰/۸۰۶	۰/۹۲۲

باتوجه‌به جدول ۹، مدل لاجیت توانسته است به معیار AUC معادل ۰/۸۰۶ دست یابد که نشان‌دهندهٔ عملکرد مناسب این مدل در تفکیک مشتریان نکول‌کننده و غیرنکول‌کننده است. همچنین، معیار اختصاصیت مدل برابر با ۰/۹۲۲ است که نشان‌دهندهٔ توانایی مدل در شناسایی مشتریان غیرنکول‌کننده است. این نتایج بیانگر این موضوع است که باوجود ساختار سادهٔ مدل لاجیت، این مدل همچنان قادر به پیش‌بینی ریسک نکول با دقتی قابل‌قبول است. در مرحلهٔ بعد، مشتریان به چهار دستهٔ مختلف تقسیم شدند و مدل لاجیت به‌طور جداگانه روی هر دسته اعمال شد. نتایج ارزیابی این مدل روی دسته‌های مختلف مشتریان در جدول ۱۰ آمده است:

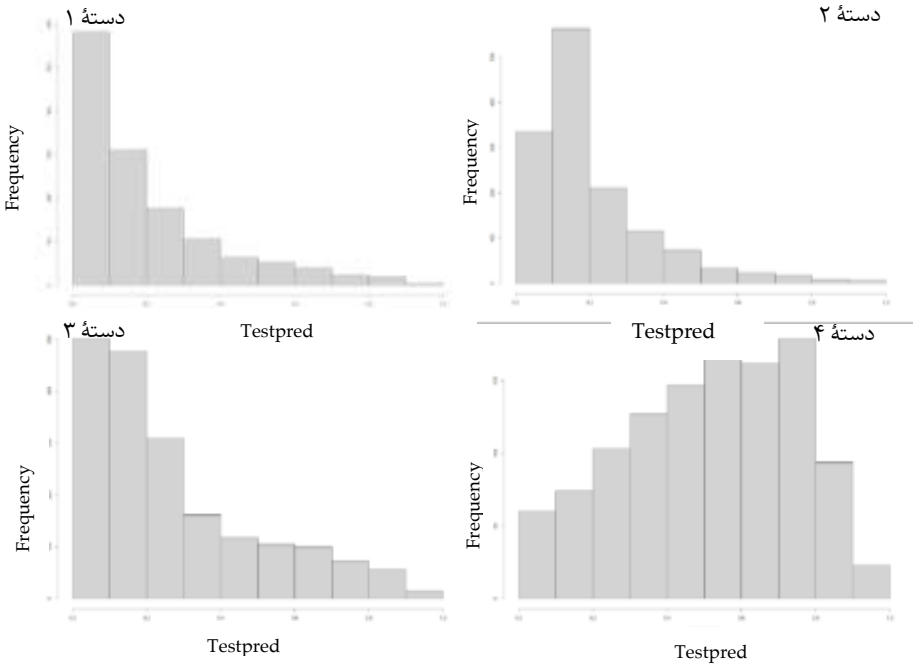
جدول ۱۰

ارزیابی نتایج مدل لاجیت روی هر دسته از مشتریان

ردیف	دسته مشتریان	تعداد مشتریان	معیار AUC	معیار Specificity
۱	دستهٔ اول	۴۴۸۵	۰/۷۲۲	۰/۹۲۱
۲	دستهٔ دوم	۴۴۸۷	۰/۷۶۰	۰/۹۷۰
۳	دستهٔ سوم	۶۳۷۲	۰/۸۱۴	۰/۹۳۶
٤	دستهٔ چهارم	۳۸۲۶	۰/۷۴۳	۰/۶۷۰

در تحلیل بر اساس دسته‌بندی مشتریان، مشاهده می‌شود که مدل لاجیت در هر چهار دسته عملکرد مناسبی دارد؛ به‌طوری‌که در تمام دسته‌ها معیار AUC بالاتر از ۰/۷ بوده و نشان‌دهندهٔ توانایی نسبی مدل در تفکیک مشتریان نکول‌کننده و غیرنکول‌کننده است. در دستهٔ سوم که شامل ۶۳۷۲ مشتری است، بالاترین AUC با مقدار ۰/۸۱۴ به‌دست آمده است

که نشان‌دهندهٔ عملکرد بهتر مدل لاجیت در این دسته است. علاوه‌براین، معیار اختصاصیت در دسته‌های اول، دوم، و سوم بالاتر از ۰/۹ بوده که نشان‌دهندهٔ دقت بالای مدل در شناسایی مشتریان غیرنکول‌کننده در این دسته‌هاست. بااین‌حال، در دستهٔ چهارم اختصاصیت به ۰/۶۷۰ کاهش یافته است که به‌معنی عملکرد ضعیف‌تر مدل در شناسایی مشتریان غیرنکول‌کننده در این دسته است.



شکل۳. توزیع نکول پیش‌بینی‌شده با روش لاجیت در هر دسته از مشتریان

شکل ۳ توزیع نکول پیش‌بینی‌شده به روش لاجیت را در چهار دسته از مشتریان نشان می‌دهد. در دستهٔ اول (بالا سمت چپ)، توزیع به‌سمت چپ متمایل است که یعنی بیشتر مشتریان در این دسته دارای ریسک نکول پایین‌اند و احتمال نکول آن‌ها نزدیک به صفر است. این توزیع بیانگر این است که اکثر مشتریان در این دسته نکول نخواهند کرد. در دستهٔ دوم (بالا سمت راست)، توزیع به‌سمت چپ چولگی دارد، به این معنی که مشتریان این دسته نیز احتمال نکول پایینی دارند؛ اما در مقایسه با دستهٔ اول تعداد بیشتری از مشتریان در معرض

ریسک متوسط نکول قرار دارند. دستهٔ سوم (پایین سمت چپ)، دارای توزیعی مشابه دسته‌های اول و دوم است، با این تفاوت که چولگی کمتر بوده و برخی مشتریان احتمال نکول بیشتری دارند. در دستهٔ چهارم (پایین سمت راست)، توزیع به‌سمت راست متمایل است، به این معنا که مشتریان این دسته بیشتر در معرض ریسک بالای نکول قرار دارند و احتمال نکول برای بسیاری از آن‌ها بیشتر از دسته‌های دیگر است.

۶ رتبه‌بندی دسته‌های مشتریان

باتوجه‌به نتایج به‌دست‌آمده از مدل، مشتریان براساس معیارهایی ازجمله میانگین مبلغ تسهیلات دارای وضعیت منفی در ۳۶ ماه گذشته، نسبت میانگین بدهی‌های آتی به میانگین بدهی‌های اخیر و میانگین تعداد قراردادهای دارای بدهی معوقه در ۳۶ ماه گذشته انجام شده است. همچنین، یک نمرهٔ ترکیبی بر اساس همین معیارها محاسبه شده است که نشان‌دهندهٔ ارزش هر خوشه است.

جدول ۱۱

رتبه‌بندی دسته‌های مشتریان

ردیف	نام دسته	مبلغ تسهیلات دارای وضعیت منفی در ۳۶ ماه گذشته	نسبت میانگین بدهی‌های آتی به میانگین بدهی‌های اخیر	تعداد قراردادهای دارای بدهی معوقه در ۳۶ ماه گذشته	وزن هر خوشه
۱	دسته دوم	۰/۰۱۷	۰/۰۰۸۹۷	۰/۰۳۴	ـ۰/۰۶۲
۲	دسته اول	۰/۰۵۱	۰/۰۰۸۹۸	۰/۰۳۷	ـ۰/۰۹۶
۳	دسته سوم	۰/۰۸۰	۰/۰۰۸۶۹	۰/۰۳۳	ـ۰/۱۱۸
٤	دسته چهارم	۰/۵۵۵	۰/۰۰۸۶۴	۰/۰۷۳	ـ۰/۶۳۷

تحلیل خوشه‌ها با توجه به جدول ۱۱ شرح زیر است:

- **مشتریان پرریسک**[1](خوشهٔ چهارم): این دسته از مشتریان بالاترین ریسک را دارند و وزن آن‌ها برابر با ۰/۶۳۷ـ است که نشان‌دهندهٔ بیشترین سطح ریسک اعتباری است. مبلغ تسهیلات دارای وضعیت منفی در این خوشه ۰/۵۵۵ بوده و تعداد قراردادهای دارای بدهی معوقه ۰/۰۷۳ است. این خوشه شامل مشتریانی است که رفتار اعتباری آن‌ها با

[1] high-risk clients

بدحسابی همراه بوده است. بانک‌ها باید توجه داشته باشند که اعطای تسهیلات به این خوشه ریسک بالایی دارد و از اعطای تسهیلات کلان به این گروه جلوگیری شود.

- **مشتریان بحران‌ساز**[1] (خوشهٔ سوم): مشتریان این خوشه با وزن ۰/۱۱۸ــ در دستهٔ مشتریان با ریسک بالاتر از متوسط قرار دارند. مبلغ تسهیلات دارای وضعیت منفی ۰/۰۸۰ و تعداد قراردادهای دارای بدهی معوقه برابر با ۰/۰۳۳ است. این گروه نشان‌دهندهٔ مشتریانی است که با بدهی‌های معوقهٔ بالا مواجه‌اند و احتمالاً در آیندهٔ نزدیک به دستهٔ مشتریان پرریسک منتقل خواهند شد. این دسته به مدیریت ویژه‌ای نیاز دارند تا از ورود به وضعیت نکول جلوگیری شود؛ برای مثال، طرح‌های تشویقی با اعطای تسهیلات دارای محدودیت سقف معین برای این گروه در نظرگرفته شود تا رفتار اعتباری این دسته از مشتریان بهبود یابد.
- **مشتریان پایدار با ریسک متوسط**[2] (خوشهٔ دوم): این خوشه با وزن ۰/۰۹۶ــ در ردهٔ دوم کم‌ریسک‌ترین خوشه‌ها قرار دارد. مبلغ تسهیلات دارای وضعیت منفی در این دسته ۰/۰۵۱ و تعداد قراردادهای دارای بدهی معوقه ۰/۰۳۷ است. مشتریان این گروه در وضعیت اعتباری نسبتاً پایداری قرار دارند؛ اما برخی نشانه‌های ریسک مثل بدهی‌های معوقه وجود دارد که بانک‌ها باید به آن‌ها دقت داشته باشند. بانک‌ها می‌توانند با ایجاد شرایط نظارتی، همانند دریافت وثیقه یا ضامن‌های بیشتر، با اطمینان بیشتری به این خوشه از مشتریان تسهیلات اعطا کنند.
- **مشتریان معتبر و کم‌ریسک**[3] (خوشهٔ اول): این دسته کمترین ریسک را دارد و با وزن ۰/۰۶۲ــ در رتبهٔ کم‌ریسک‌ترین خوشه‌ها قرار دارد. مبلغ تسهیلات دارای وضعیت منفی برابر با ۰/۰۱۷ و تعداد قراردادهای دارای بدهی معوقه ۰/۰۳۴ است. مشتریان این خوشه دارای سابقهٔ اعتباری بسیار خوب و کمترین میزان ریسک از منظر بانکی‌اند. این دسته شامل مشتریانی است که بدهی‌های معوقه و تسهیلات منفی کمی دارند و وضعیت اعتباری پایدار و خوبی دارند. اعطای تسهیلات به این گروه از مشتریان که از خوش‌حساب‌ترین مشتریان برای بانک‌ها محسوب می‌شوند، ریسک اعتباری بسیار کمی به‌همراه دارد.

---

[1] critical risk clients

[2] stable clients with moderate risk

[3] low-risk trusted clients

۷ نتیجه‌گیری

این پژوهش نشان‌دهندهٔ اهمیت استفاده از مدل‌های ترکیبی و تکنیک‌های پیشرفتهٔ یادگیری ماشین در بهبود فرایندهای شناسایی، دسته‌بندی مشتریان، و پیش‌بینی نکول در نظام بانکی است. با به‌کارگیری الگوریتم میانگین کا (K-means) برای دسته‌بندی مشتریان و مدل‌های XGBoost و لاجیت برای پیش‌بینی نکول، مدلی طراحی شد که به‌طور مؤثری رفتار اعتباری مشتریان را تحلیل کرده و احتمال ریسک نکول آن‌ها را تخمین می‌زند. نتایج به‌دست‌آمده نشان می‌دهد مشتریان به چهار گروه اصلی تقسیم می‌شوند: پرریسک (بدهی معوقهٔ زیاد و نیاز به مدیریت ویژه)، بحران‌ساز (رفتار اعتباری ناپایدار و نیازمند مدیریت پیشگیرانه)، ریسک متوسط (بدهی کمتر؛ اما نیازمند نظارت)، و کم‌ریسک (بهترین گزینه برای تسهیلات جدید).

مزیت اصلی مدل پیشنهادی توانایی آن در پیش‌بینی دقیق‌تر ریسک نکول و ارائهٔ راهکارهای بهینه برای مدیریت ریسک است. این مدل به مدیران بانکی کمک می‌کند تا رویکردی هوشمندانه‌تر و کارآمدتر در تخصیص منابع مالی اتخاذ کنند. بانک‌ها می‌توانند تسهیلات بیشتری به مشتریان کم‌ریسک اختصاص دهند و برای مشتریان با ریسک بالا یا متوسط راهبردهای حمایتی و نظارتی دقیق‌تری اجرا کنند. درنهایت، این پژوهش پیشنهاد می‌کند بانک‌ها از مدل‌های ترکیبی یادگیری ماشین در کنار روش‌های سنتی اعتبارسنجی استفاده کنند تا ضمن کاهش ریسک نکول و افزایش بازده تسهیلات، کیفیت خدمات خود را نیز بهبود بخشند. این رویکرد به افزایش رضایت مشتریان و تقویت جایگاه رقابتی بانک‌ها در بازار مالی منجر خواهد شد. بااین‌حال، انجام مطالعات بیشتر در این زمینه می‌تواند به بهبود عملکرد مدل کمک کند. به‌کارگیری الگوریتم‌های پیشرفته‌تر؛ مانند شبکه‌های عصبی عمیق[1] و مقایسهٔ آن‌ها با مدل‌های فعلی می‌تواند دقت پیش‌بینی‌ها را افزایش دهد. همچنین، اضافه‌کردن متغیرهای جدید نظیر بررسی متغیرهای جدید (متغیرهایی مانند سطح تحصیلات مشتری، تعداد حساب‌های فعال شخص، میزان گردش حساب در دوره‌های زمانی مشخص و...) می‌تواند تحلیل رفتار اعتباری مشتریان را بهبود بخشد. علاوه‌براین، گسترش کاربرد مدل‌های پیشنهادی در سایر حوزه‌ها نظیر صنعت بیمه و سرمایه‌گذاری می‌تواند به بهبود مدیریت ریسک در این بخش‌ها کمک کند. این پیشنهادها می‌توانند به بانک‌ها و مؤسسات مالی کمک کنند تا ضمن کاهش ریسک نکول و افزایش بازدهی تسهیلات، خدمات خود را به‌صورت هدفمندتر و مؤثرتر ارائه دهند.

---

[1] Deep Learning

فهرست منابع

مرادی، م، حری، م و نوری، ا. (۱۴۰۲). مطالعات کمی در مدیریت صنعت بانکداری به منظور افزایش رضایتمندی و سودآوری مشتریان (مطالعه موردی: بانک ملت). *مطالعات کمّی در مدیریت*، ۱۴(۵۳)، ۸۴ـ۵۲.

تقوی فرد، م. (۱۳۹۶). دسته‌بندی مشتریان حقوقی و پیش‌بینی توانایی سوددهی آنان با استفاده از ارزش طول عمر مشتری و رویکرد زنجیرهٔ مارکوف (مورد مطالعه: مشتریان یک بانک خصوصی). *مطالعات مدیریت صنعتی*، ۴۵. ۴۵ـ۷۴.

مهرآرا، م.، و مهران‌فر، م. (۱۳۹۲). عملکرد بانکی و عوامل کلان اقتصادی در مدیریت ریسک. *مدلسازی اقتصادی*، ۷(۱ (پیاپی ۲۱))، ۲۱ـ۳۷

خواجوند، س.، تقوی فرد، م.، و نجفی، ا. (۱۳۹۱). بخش‌بندی مشتریان بانک صادرات ایران با استفاده از داده کاوی. *مطالعات مدیریت بهبود و تحول*، ۲۱(۶۷)، ۲۰۰ـ۱۷۹

سهرابی، ب.، خانلری، ا.، و آجرلو، ن. (۱۳۹۰). الگویی برای تعیین ارزش چرخهٔ عمر مشتریان (CLV) در صنعت بانکداری. *پژوهش های مدیریت در ایران*، ۱۵(۱ (پیاپی ۷۰))، ۲۲۳ـ۲۳۹

Li, T., Kou, G., Peng, Y., & Philip, S. Y. (2021). An integrated cluster detection, optimization, and interpretation approach for financial data. *IEEE transactions on cybernetics*, 52(12), 13848-13861.

Dawood, E. A. E., Elfakhrany, E., & Maghraby, F. A. (2019). Improve profiling bank customer's behavior using machine learning. *Ieee Access,* 7, 109320-109327.

Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. *In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785-794).

Scitovski, S., & Šarlija, N. (2014). Cluster analysis in retail segmentation for credit scoring. *Croatian Operational Research Review*, 235-245.

Khobzi, H., Akhondzadeh-Noughabi, E., & Minaei-Bidgoli, B. (2014). A new application of RFM clustering for guild segmentation to mine the pattern of using banks'e-payment services. *Journal of Global Marketing*, 27(3), 178-190.

Haenlein, M., Kaplan, A. M., & Beeser, A. J. (2007). A model to determine customer lifetime value in a retail banking context. *European Management Journal,* 25(3), 221-234.

Heffernan, S. (2003). 14. The causes of bank failures. *Handbook of International Banking*, 366.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of educational psychology, 24*(6), 417.

Pearson, K. (1901). LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin philosophical magazine and journal of science*, 2(11), 559-572.