فصل‌نامه مطالعات مالی و بانکداری اسلامی
مؤسسه عالی آموزش بانکداری ایران
دوره ۴، شماره ۹
تابستان ۱۳۹۷
صفحه ۷۱-۳۳

تصمیمات تسهیلات‌دهی بانک با استفاده از روش الگوریتم ژنتیک (مطالعه موردی: مشتریان حقیقی بانک سپه)

رضا حبیبی[۱]
حسن کوهی[۲]
حسین بعیدی مفردنیا[۳]

چکیده

با توجه به ماهیت فعالیت‌های صنعت بانکداری که عمدتاً مبتنی بر تجهیز و تخصیص منابع است، این صنعت به‌طور گسترده با ریسک‌های اعتباری مواجه است. بنابراین شناخت منشأ ریسک اعتباری و تخمین آن همواره یک مسئله اساسی برای صنعت بانکداری است. در همین خصوص، تحقیقی از نوع داده‌کاوی با هدف شناسایی ویژگی‌های مؤثر بر ریسک اعتباری مشتریان حقیقی بانک سپه و همچنین طراحی مدلی برای پیش‌بینی احتمال نکول تسهیلات، از طریق مدل‌های الگوریتم ژنتیک و رگرسیون پروبیت انجام گرفته است. داده‌های این تحقیق مربوط به تسهیلات اعتباری پرداخت شده به اشخاص حقیقی در سال ۱۳۹۵ است. از میان کلیه تسهیلات اعتباری پرداخت شده به اشخاص حقیقی در سال ۱۳۹۵ دو نمونه ۳۶۰۰ تایی (به‌منظور برازش مدل) و ۴۰۰ تایی (به‌منظور راستی‌آزمایی مدل به‌وسیله منحنی ROC) به‌صورت تصادفی انتخاب شدند. همچنین به‌منظور تجزیه و تحلیل داده‌ها از نرم‌افزار متلب استفاده شده است.

نتایج تحقیق نشان می‌دهد که روش الگوریتم ژنتیک در تعیین متغیرها در سه سطح متفاوت براساس درجه اهمیت، توانایی بالاتری در پیش‌بینی احتمال نکول تسهیلات نسبت به روش رگرسیون پروبیت دارد. نتایج راستی‌آزمایی نشان می‌دهد که سطح زیر منحنی ROC در روش الگوریتم ژنتیک برابر ۰/۹۲، اما در روش رگرسیون پروبیت برابر ۰/۷۲ است و همچنین نتایج در ماتریس ROC نشان داد که روش الگوریتم ژنتیک ۹۱/۸ درصد پیش‌بینی صحیح نموده و روش رگرسیون پروبیت ۹۰ درصد پیش‌بینی صحیح کرده است.

واژه‌های کلیدی: رتبه‌بندی اعتباری، الگوریتم ژنتیک، رگرسیون پروبیت، منحنی ROC.
طبقه‌بندی JEL: G24, G32, C35, C45.

---

[۱] عضو هیأت علمی، مؤسسه عالی آموزش بانکداری ایران (نویسنده مسئول)؛ R_habibi@ibi.ac.ir
[۲] عضو هیأت علمی، مؤسسه عالی آموزش بانکداری ایران؛ Kouhi@ibi.ac.ir
[۳] کارشناس ارشد بانکداری اسلامی، مؤسسه عالی آموزش بانکداری ایران؛ Hossein.baidi@yahoo.com

تاریخ دریافت: ۹۶/۱۲/۱۱ **تاریخ پذیرش: ۹۷/۰۸/۲۳**

مقدمه

بانک‌ها به‌عنوان یکی از مهم‌ترین بنگاه‌های اقتصادی در کسب و کار بایستی توانایی شناسایی ویژگی‌های مشتریان خود را داشته باشند. شناخت دقیق مشتریان سبب آگاهی از نیازها و رفتارهای مشتریان و ارائه خدمات مالی بهتر به آن‌ها می‌شود. یکی از مهم‌ترین خدمات مالی بانک اعطای تسهیلات به مشتریان است. امروزه در هر کسب و کاری، رضایتمندی مشتریان از مباحث اصلی و با اهمیت برای کسب درآمد و رقابت مؤثر در عرصه بازار تلقی می‌شود. لذا، شناخت از مشتریان در نهایت به رضایتمندی مشتریان و وفاداری آن‌ها به بانک منجر خواهد شد. از طرف دیگر، هر کسب و کاری با مجموعه‌ای از ریسک‌ها در فرایندهای خود مواجه است. یکی از مهم‌ترین ریسک‌های بانکی، ریسک اعتباری است که بانک‌ها بایستی بتوانند این ریسک را مدیریت کنند (البرزی و همکاران، ۱۳۸۹).

از جمله دلایل اهمیت سنجش و مدیریت ریسک اعتباری می‌توان موارد زیر را نام برد:

۱- هم‌اکنون مهم‌ترین عامل ورشکستگی بانک‌ها ریسک اعتباری است. اگر مشتری به‌موقع تعهدات خود را بازپرداخت نکند، این تسهیلات به‌صورت مطالبات معوق بانکی درمی‌آید و قدرت تسهیلات‌دهی بانک به افراد جدید کاهش می‌یابد و در نتیجه سودآوری بانک نیز کاهش خواهد یافت که به اختلال در اقتصاد کشور می‌انجامد.

۲- اندازه‌گیری ریسک اعتباری با پیش‌بینی زیان‌های عدم بازپرداخت اعتبارات و ایجاد رابطه منطقی بین ریسک و بازده، امکان بهینه‌سازی ترکیب پرتفوی اعتباری، قیمت‌گذاری دارایی‌ها و تعیین سرمایه اقتصادی بانک‌ها را به‌منظور کاهش هزینه‌های سرمایه‌ای و حفظ توان رقابتی فراهم و نوعی مزیت نسبی برای بانک‌ها و مؤسسات مالی و اعتباری ایجاد می‌کند.

۳- در ایران از یک‌طرف فعالیت بانک‌ها براساس قانون بانکداری بدون ربا و مبتنی بر عقود اسلامی است، بنابراین نمی‌توان بین بازار پول و سرمایه مرزی قائل شد. از طرف دیگر، با توجه به ساختار اقتصادی کشور، عملیات بازار سرمایه (بازار اوراق بهادار و سهام) و سایر شبکه‌های غیربانکی، پیشرفت فراوانی نداشته و از این‌رو سهم قابل‌توجهی از سرمایه‌گذاری از طریق بازار بانکی انجام می‌گیرد، بنابراین موفقیت بانک‌ها در انجام این امور اهمیت خاصی دارد (کشاورز حداد و آیتی گازار، ۱۳۸۶).

اعتبارسنجی مشتریان به‌عنوان یک ابزار مهم می‌تواند به بانک در شناسایی مشتریان خوش‌حساب و بدحساب و اعطای بهتر تسهیلات اعتباری کمک کند. اعتبارسنجی بهینه نیازمند اجرای مدل‌های آماری و اقتصادسنجی مناسب است. مدل‌های رتبه‌بندی اعتباری یکی از مهم‌ترین و اساسی‌ترین سیستم‌های تصمیم‌گیری هستند که بخش عمده‌ای از اطلاعات مورد نیاز مؤسسات اعتباردهنده در

مدیریت مؤثر اعتبارات را فراهم می‌کنند. این مدل‌ها به بانک‌ها و مؤسسات مالی و اعتباری کمک می‌کنند تا با صرف هزینه کمتر و دقت بیشتر، درخواست مشتریان اعتباری خود را بررسی کنند.

اگر چه تحقیقات نظری متعددی با استفاده از مدل‌های رتبه‌بندی مشابه خارجی در مورد بانک‌ها و مؤسسات مالی و اعتباری ایران انجام شده است، اما با مقایسه اطلاعات کسب شده از بانک‌های سایر کشورها با بانک‌های کشور، می‌توان پی برد که در بانک‌ها و مؤسسات مالی و اعتباری کشور برخلاف سایر کشورها، مدلی جامع و کاربردی برای ارزیابی ریسک وجود ندارد. بالا بودن ذخایر مطالبات مشکوک‌الوصول بانک‌ها و تسهیلات اعطایی سوخت شده یا معوق شده بانک‌ها نیز گویای نبود مدل‌های مناسب اندازه‌گیری ریسک اعتباری و سیستم‌های مدیریت ریسک در شبکه بانکی کشور است.

با توجه به اهمیت این موضوع، استفاده از مدل مناسب رتبه‌بندی اعتباری داخلی برای پیش‌بینی وضعیت عدم بازپرداخت تسهیلات توسط مشتریان، می‌تواند برای بانک‌ها و مؤسسات مالی و اعتباری تبدیل به یک مزیت رقابتی شود (رستم کلایی، ۱۳۸۶). در این تحقیق تلاش شده است که مدلی برای رتبه‌بندی اعتباری مشتریان حقیقی بانک سپه ارائه شود تا به کمک این مدل، مدیران و متقاضی مسئولان بخش اعتباری بانک‌ها و مؤسسات مالی (به‌ویژه بانک سپه)، با صرف کمترین هزینه و کوتاه‌ترین زمان به تصمیمات تسهیلات‌دهی مرتبط با مشتریان حقیقی اقدام کنند.

مبانی نظری و پیشینه تحقیق

امتیازدهی اعتباری، نظامی است که از طریق آن بانک‌ها و مؤسسات اعتباری با استفاده از اطلاعات حال و گذشته متقاضی، احتمال عدم بازپرداخت تسهیلات توسط وی را ارزیابی می‌کنند و به ایشان امتیاز می‌دهند. به‌عبارت دیگر، امتیازدهی به‌معنی کمی نمودن احتمال نکول در آینده است. امتیازدهی اعتباری ابزاری عینی برای مدیریت ریسک است که مشتریان اعتباری را با بی‌طرفی و براساس آمار و اطلاعات رتبه‌بندی می‌کند؛ در حالی‌که روش‌های قدیمی برای ارزیابی مشتریان، عمدتاً ذهنی و متکی بر دیدگاه مسئولان پرداخت تسهیلات است. تعاریف گوناگونی برای رتبه‌بندی اعتباری ارائه شده است. از نظر مؤسسه استاندارد اند پورز[1]، سیستم رتبه‌بندی اظهارنظر در مورد ارزش اعتباری یک بدهکار براساس عوامل ریسک است. همچنین از نظر مؤسسه مودیز[2]، رتبه‌بندی عبارت است از اظهارنظر در مورد توانایی آتی بدهکار و تعهد حقوقی منتشرکننده اوراق، برای انجام پرداخت‌های به‌موقع اصل

[1] Standard and Poor's
[2] Moody's

و بهره، روی اوراق بهادار با درآمد ثابت و مشخص. از جمله اهداف امتیازدهی اعتباری که با گسترش بانکداری الکترونیک و به‌دنبال آن فراگیرشدن استفاده از کارت‌های اعتباری ضرورت می‌یابد این است که لزوم بازنگری در فرایندهای تسهیلات بانکی و امتیازدهی و به هنگام کردن این فرایند با استفاده از فناوری روز در زمینه‌های امتیازسنجی، امری اجتناب‌ناپذیر است.

در زیر چهار هدف از اهداف مدل‌های امتیازدهی ارائه می‌شود:

۱- **بازاریابی:** بازاریابی باعث شناسایی مشتریان بالقوه برای تسهیلات‌گیری می‌شود که خود سبب کاهش هزینه‌های تبلیغاتی است. همچنین احتمال از دست‌دادن مشتریان با ارزش باید تعیین شود، زیرا سازمان را به تنظیم راهبرد حفظ مشتری قادر خواهد ساخت.

۲- **ارزیابی درخواست‌ها:** در مورد تمدید اعتبار و میزان آن تصمیم‌گیری می‌شود و رفتار آینده متقاضی جدید تسهیلات با پیش‌بینی احتمال نکول در آینده تعیین می‌شود.

۳- **کنترل مشتریان:** رفتار آینده مشتریان فعلی که تسهیلات دریافت کرده‌اند را با توجه به شرایط اقتصادی باید پیش‌بینی کرد. بنابراین می‌توان توجه بیشتری به مشتریانی داشت که دچار مشکل شده‌اند و بدین‌صورت می‌توان احتمال نکول را کاهش داد.

۴- **مدیریت تسهیلات بد:** بر این اساس می‌توان خط‌مشی مشابه برای به حداقل رساندن هزینه مجموعه با افزایش مقدار بازیافت شده از حساب‌های معوق شده اتخاذ کرد.

ارائه مدل‌های اعتبارسنجی همواره با نتایج خوبی در بانک روبه‌رو بوده است. در حال حاضر مدل‌های مختلفی برای اعتبارسنجی مشتریان بانک‌ها و مؤسسات مالی به‌کار می‌رود. از طرفی تحقیقات متنوعی روی کارایی این مدل‌ها در داخل کشور و کاربرد آن در اعتبارسنجی صورت گرفته است. در سایر کشورها نیز با توسعه روش‌های اعتبارسنجی در دهه‌های ۱۹۵۰ و ۱۹۶۰، انواع مدل‌های آماری و داده کاوی برای ارزیابی اعتباری مشتریان ارائه شد. امروزه اغلب بانک‌ها و مؤسسه‌های اعتباری از اعتبارسنجی مشتریان به‌منظور کنترل ریسک اعتباری خود استفاده می‌کنند. روش‌های متنوعی برای ارزیابی اعتباری مشتریان وجود دارد و مورد استفاده قرار می‌گیرد که معرفی و مرور برخی از آن‌ها به‌شرح زیر ارائه شده است.

برخی مطالعات نشان داده‌اند که روش‌های هوش مصنوعی و محاسبات تکاملی می‌توانند پیش‌بینی‌های دقیق‌تری را در اختیار استفاده‌کنندگان قرار دهند. در مقابل برخی دیگر، هنوز بر این باورند که روش‌های قدیمی بهتر و مؤثرتر بوده یا استفاده هم‌زمان از هر دو دسته روش‌ها، نتیجه خوبی را به‌همراه خواهد

داشت (دسای و همکاران[1]، ۱۹۹۷). برای نمونه در تحقیقاتی که انجام شده، با بررسی انواع روش‌های اعتبارسنجی، عملکرد این روش‌ها را روی مجموعه‌هایی از داده‌های بانک‌های فرانسوی و آمریکایی می‌توان دید که روش‌های رگرسیون لجستیک و شبکه‌های عصبی از لحاظ دقت عملکرد مشابهی دارند، اما از لحاظ سرعت انجام تحلیل‌ها، مدل لاجیت از بقیه بهتر است (چن و هوانگ[2] ، ۲۰۰۳). همچنین در ادبیات اعتبارسنجی مشتریان، برنامه‌ریزی ژنتیکی برای ساخت مدل‌های اعتبارسنجی پیشنهاد شده است که نتایج به‌دست آمده در مقایسه با مدل‌های متعارف اعتبارسنجی شامل شبکه‌های عصبی، درخت تصمیم، مجموعه‌های سخت و رگرسیون لجستیک، نشان می‌دهد که برنامه‌ریزی ژنتیکی عملکرد بهتری دارد (اونگ و همکاران[3]، ۲۰۰۵).

همچنین در برخی از تحقیقات به مقایسه روش‌های گوناگون در اعتبارسنجی پرداخته‌اند؛ به‌عنوان مثال عملکرد دو روش داده‌کاوی، شامل کارت[4] و مارس[5] در اعتبارسنجی مشتریان بررسی شده است. برای سنجش کارایی این دو روش، از مجموعه داده‌های کارت اعتباری یک بانک آمریکایی استفاده شده است. نتایج نشان داد عملکرد دو روش یاد شده نسبت به رویکردهای سنتی تحلیل ممیزی، رگرسیون لجستیک، شبکه‌های عصبی و ماشین بردار پشتیبان بهتر است (لی و همکاران[6]، ۲۰۰۶).

سایر مدل‌های به‌کار رفته در اعتبارسنجی شامل از مدل‌های تحلیل ممیزی گام به گام و رگرسیون لجستیک با استفاده از نسبت مالی به‌عنوان متغیرهای مؤثر بر رفتار اعتباری و تفکیک مشتریان از نظر ریسک اعتباری است. عموماً متغیر نسبت جاری بیشترین قدرت تفکیک مشتریان خوش‌حساب و بدحساب را در میان سایر متغیرها دارد. از سوی دیگر، روش‌های تحلیل ممیزی و رگرسیون لجستیک در دسته‌بندی مشتریان از نظر ریسک اعتباری، نتایج مشابهی ارائه می‌دهند (ذکاوت، ۱۳۸۲). همچنین، مدل‌های پروبیت، لاجیت و درختان تصمیم در طبقه‌بندی و رگرسیون هم در بین مدل‌های رتبه‌بندی مشاهده می‌شوند (ملا ابراهیم لو، ۱۳۸۴؛ کشاورز حداد و آیتی گازار، ۱۳۸۶).

[1] Desai et al.
[2] Chen & Huang
[3] Ong et al.
[4] Classification and Regression Trees (CART)
[5] Multivariate Adaptive Regression Splines (MARS)
[6] Lee et al.

پیشینه تحقیق

هوانگ و همکاران[1] (۲۰۰۷) یک روش هیبریدی برای اعتبارسنجی ارائه دادند. بدین‌صورت که ترکیب الگوریتم ژنتیک و ماشین بردار پشتیبان هم برای انتخاب خصیصه‌ها و هم برای بهینه‌سازی پارامترهای مدل اعتبارسنجی به‌کار گرفته شده است. مقایسه‌هایی نیز بین روش ماشین بردار پشتیبان و شبکه‌های عصبی، برنامه‌ریزی ژنتیکی و درخت تصمیم صورت گرفته که نتایج کمابیش یکسانی داشته است. ولی نتایج تجربی نشان می‌دهد که ماشین بردار پشتیبان نسبت به سایر روش‌های داده‌کاوی نتایج بهتری ارائه می‌دهد.

ابدوو[2] (۲۰۰۸) در تحقیق خود از دو روش مدل رگرسیون پروبیت و برنامه‌ریزی ژنتیک برای مدل‌سازی ارزیابی اعتباری در بین بانک‌های بخش دولتی کشور مصر استفاده کرده است. در نخستین روش با استفاده از مدل رگرسیون پروبیت درصد تأثیر هر یک از متغیرها بر متغیر وابسته (نکول تسهیلات) و در روش دوم با استفاده از برنامه‌ریزی ژنتیک متغیرهای تأثیرگذار بر متغیر وابسته مشخص شدند. در نهایت با استفاده از منحنی ROC هر دو روش بررسی شدند تا مشخص شود بین دو روش پروبیت و ژنتیک کدام‌یک توانایی بیشتری در تعیین ویژگی‌های تأثیرگذار مشتری دارد و همچنین بهتر می‌تواند با این ویژگی‌ها احتمال نکول تسهیلات را پیش‌بینی کند؛ سپس با استفاده از یک نمونه به راستی‌آزمایی و تشخیص صحت هر دو روش پرداخته شد. در نهایت نشان داده شد که بین روش رگرسیون پروبیت و برنامه‌ریزی ژنتیک، روش برنامه‌ریزی ژنتیک بهتر می‌تواند متغیرهای تأثیرگذار را مشخص کند.

پالیوال و کومار[3] (۲۰۰۹) تحقیق خود را بر پایه ترکیب طبقه‌بندی با ناظر و بدون ناظر استوار کرده‌اند. در گام نخست با استفاده از الگوریتم خوشه‌بندی[4]، مشتریان به گروه‌هایی با ویژگی‌های مشابه دسته‌بندی شدند و در گام بعدی برای هر گروه از مشتریان درخت تصمیم ساخته شد. روش پیشنهاد شده آن‌ها به بانک اجازه می‌دهد تا از قوانین متفاوت برای مشتریان مختلف که ریسک‌های بالایی دارند، استفاده کند.

---

[1] Huang et al.
[2] Abdou
[3] Paliwal & Kumar
[4] Clustering Algorithm

کنستانگیورا[1] (۲۰۱۱) با استفاده از روش بگینگ[2] احتمال نکول را بر مبنای رگرسیون لاجیت بهبود بخشید. وی نشان داد که هزینه‌های نظارت هزینه‌های سرباری بر بانک هستند، اما استفاده از یک سیستم نظارتی مناسب حتی با هزینه بالا هم در نهایت به کاهش ریسک اعتباری بانک و حتی در مواردی هم سودآوری می‌انجامد. تحقیق وی با استفاده از داده‌های واقعی مؤسسه بانکداری بوداپست انجام شد و در آن سه دسته از متغیرها بررسی شدند. نخستین سری متغیرها، مشخصه‌های جمعیت‌شناختی مشتریان بود. دومین سری وضعیت مالی گیرنده تسهیلات و سومین دسته از متغیرها شامل تاریخچه تسهیلات و بازپرداخت تسهیلات قبلی بود.

لوند[3] (۲۰۱۵) رگرسیون لجستیک دو مرحله‌ای را برای بهبود رتبه‌بندی اعتباری در سطح مشتریان حقیقی بانک‌های هلند بین دوره زمانی ۲۰۱۴-۲۰۱۲ به‌کار برد و به این نتیجه رسید که ضریب جینی به‌عنوان معیاری از نیکویی برازش مدل به میزان یک درصد افزایش یافته است. او ابتدا داده‌ها را به دو قسمت برازش و ارزیابی تقسیم کرد؛ سپس با استفاده از داده‌های بخش اول مدل لجستیک برازش و با استفاده از داده‌های سری دوم، عملکرد مدل ارزیابی شد.

جو و سون[4] (۲۰۱۷) مجموعه روش‌های کمی و کیفی را برای اعتبارسنجی مشتریان ارائه دادند. روش‌های آنان شامل تحلیل همبستگی کانونی[5]، رتبه‌بندی و تحلیل ممیزی کیفی و تحلیل شناسایی الگو[6] بود که آن‌ها را در بین مشتریان حقیقی بانک‌های چین به‌کار بردند.

متیو[7] (۲۰۱۷) با استفاده از رگرسیون لجستیک مرحله‌ای به رتبه‌بندی ۱۵۰ هزار مشتری بانک‌های آمریکا با ۱۱ ویژگی ثبت شده مشتریان پرداخت. وی نشان داد که این روش از روش‌های تجزیه مقدار منفرد[8] عادی و وزنی به میزان ۱۰ و ۱۵ درصد دقت بیشتری دارد. همچنین، روشی کاربردی برای انتخاب ویژگی که یکی از مهم‌ترین مسائل در اعتبارسنجی مشتریان است، ارائه داد.

ابدوو (۲۰۱۸) امتیازدهی اعتباری را با استفاده از برنامه‌ریزی ژنتیک در بانک‌های مصر طی دوره زمانی ۲۰۱۴-۲۰۱۳ انجام داد و به این نتیجه رسید که این روش از لحاظ شاخص‌های وزن شهود و

---

[1] Constangioara
[2] Bagging
[3] Lund
[4] Ju & Sohn
[5] Canonical Correlation Analysis
[6] Pattern Recognition
[7] Mathew
[8] Singular Value Decomposition (SVD)

متوسط طبقه‌بندی صحیح و هزینه طبقه‌بندی اشتباه، روش بهتری نسبت به روش‌های رگرسیون پروبیت و برخی روش‌های داده‌کاوی است. در این تحقیق، ابدوو به‌دنبال یافتن بهترین تابع به‌عنوان تابع ممیزی بین مشتریان بانک‌های مصر در دوره مطالعه است و هر تابع به‌صورت یک نمودار درختی ترسیم شده و بهترین تابع با استفاده برنامه‌ریزی ژنتیک حاصل می‌شود.

رستم‌کلایی (۱۳۸۶) با استفاده از ۲۲ متغیر و ۲۳۵ قانون در چارچوب یک سیستم خبره فازی، الگویی برای ارزیابی اعتباری مشتریان تسهیلات اعتباری بانکی طراحی کرده است. استخراج دانش و ارزیابی متغیرها در این سیستم به‌گونه‌ای طراحی شده که به فرایند ذهنی ارزیابان نزدیک باشد. مدل استفاده شده، سیستم استدلال میدانی است و خروجی سیستم، مشتریان را در چهار گروه مختلف براساس میزان اعتبارشان طبقه‌بندی می‌کند.

نیلی و سبزواری (۱۳۸۷) برای ارزیابی ۸۶ مورد از مشتریان حقوقی بانک پارسیان، با استفاده از اطلاعات مالی آن‌ها، مدل‌های امتیازدهی لاجیت و روش غیرپارامتری AHP را بررسی و این دو روش را از نظر میزان دقت در پیش‌بینی مشتریان خوب و بد با یکدیگر مقایسه کردند. بر این اساس، روش AHP که یک روش کارشناس محور است، توانست با دقت بهتری ارزش اعتباری و در نهایت خوب و بد بودن مشتریان را تعیین کند.

البرزی و همکاران (۱۳۸۹) نیز در تحقیقی برای اعتبارسنجی مشتریان بانک‌ها از الگوریتم ژنتیک در بهینه‌سازی درختان تصمیم‌گیری استفاده کردند که نتایج آنان نشان داد، دقت طبقه‌بندی مدل پیشنهادی به‌طور تقریبی از تمام مدل‌های درخت تصمیم‌گیری مقایسه شده در این تحقیق بیشتر است.

فلاح شمس و مهدوی‌راد (۱۳۹۱) به طراحی مدل اعتبارسنجی و پیش‌بینی ریسک اعتباری مشتریان تسهیلات لیزینگ ایران خودرو پرداختند. آن‌ها مشتریان حقیقی تسهیلات لیزینگ شرکت ایران خودرو را از سال ۱۳۸۱ تا ۱۳۸۴، با به‌کارگیری آزمون تفاضل میانگین و ضریب تعیین برای پنج متغیر خالص درآمد ماهیانه متقاضی، مدت تسهیلات، میزان تسهیلات، خالص درآمد ماهیانه ضامن و سابقه کار بررسی کردند و نشان دادند که کارایی مدل لاجیت ۹۷/۳۹ درصد و مدل پروبیت ۹۸/۴۹ درصد است.

مهرگان و همکاران (۱۳۹۵) اعتبارسنجی مشتریان حقیقی بانک‌های گوناگون را از طریق روش‌های داده‌کاوی K-Means و Kohonen بررسی کردند. آن‌ها با بررسی حدود ۱۰۰ مدل گوناگون، مشتریان بانک را براساس معیارهای مبلغ تسهیلات، مبلغ ضمانت، زمان بازپرداخت تسهیلات و متغیرهای جمعیت‌شناختی، به پنج خوشه با مراکز متفاوت با فاصله معنی‌دار تقسیم کردند. همچنین آن‌ها نشان دادند که در بین متغیرهای مورد بررسی متغیر مبلغ تسهیلات بیشترین تأثیر را داراست.

اسکندری و روحی (۱۳۹۶) به مدیریت ریسک مشتریان بانکی با استفاده از روش ماشین بردار تصمیم بهبود یافته با الگوریتم ژنتیک با رویکرد داده‌کاوی پرداختند. آن‌ها از آمارهای رسمی بانک مرکزی در مورد مطالبات معوق بانک‌ها استفاده نمودند و نشان دادند که الگوی پیشنهادی دقت بیشتری در مورد پیش‌بینی درصد وصول مطالبات قراردادهای با احتمال ریسک مطالباتی بالا قبل از اعطای تسهیلات نسبت به مدل‌های استانداردی نظیر رگرسیون لجستیک داراست.

اقبالی و همکاران (۱۳۹۶) به ارزیابی مقایسه‌ای عملکرد توابع شایستگی گوناگون الگوریتم ژنتیک در رتبه‌بندی مشتریان پرداختند. آن‌ها نشان دادند که الگوریتم ژنتیک با انتخاب بهترین تابع شایستگی بهتر از روش‌های متعارفی مانند رگرسیون لجستیک و تحلیل پوششی داده‌ها است.

روش‌شناسی تحقیق

این تحقیق از نظر هدف، کاربردی و از نظر تجزیه و تحلیل اطلاعات، توصیفی است. بخشی از اطلاعات و داده‌های این تحقیق از طریق مطالعات کتابخانه‌ای جمع‌آوری شده که این مطالعات شامل بررسی مقالات، کتب، تحقیقات و تحلیل‌های انجام شده مرتبط با موضوع تحقیق است.

هدف اصلی این تحقیق شناخت ویژگی‌های مهم رتبه‌بندی مشتریان حقیقی بانک سپه، با استفاده از مدل‌های رگرسیون پروبیت و الگوریتم ژنتیک و همچنین ارائه یک روش مناسب برای اتخاذ تصمیمات تسهیلات‌دهی صحیح به مشتریان حقیقی بانک سپه از میان دو مدل بیان شده است. اهداف کاربردی شامل، شناسایی ویژگی‌های اعتباری مشتریان حقیقی بانک سپه و طبقه‌بندی بهینه مشتریان جدید براساس آن ویژگی‌ها برای ارائه خدمات بانکی، تصمیم‌گیری بانک در اعطای تسهیلات به کمک یک الگوی عملی، تصمیم‌گیری در مورد اعطای تسهیلات و خدمات اعتباری به مشتریان و کاهش ریسک ارائه خدمات و تسهیلات اعتباری است.

پرسش‌های اصلی تحقیق به‌شرح ذیل است:

۱- کدام‌یک از ویژگی‌های مشتریان حقیقی بانک سپه بیشترین تأثیر را بر احتمال نکول تسهیلات دارند؟

۲- کدام‌یک از مدل‌های رگرسیون پروبیت و الگوریتم ژنتیک عملکرد بهتری در پیش‌بینی رتبه‌بندی اعتباری مشتریان حقیقی بانک سپه دارند؟

جامعه آماری در این تحقیق کلیه اشخاص حقیقی است که در سال ۱۳۹۵ از تسهیلات اعتباری بانک سپه استفاده کرده‌اند. در ابتدا تعداد تسهیلات نکول نشده (تسهیلات فاقد اقساط معوق) به کل تسهیلات در جامعه آماری (۳۲۱۹۲۵ فقره است) و همچنین نسبت تسهیلات نکول شده (تسهیلات

با حداقل ۳ قسط معوق) به کل تسهیلات در جامعه آماری مشخص شد. پس از تعیین این نسبت‌ها از بین جامعه آماری دو نمونه ۳۶۰۰ تایی و ۴۰۰ تایی به‌طور تصادفی انتخاب شد که نمونه ۳۶۰۰ تایی به‌منظور برآورد تأثیر متغیرها در مدل پروبیت و مدل الگوریتم ژنتیک و همچنین نمونه ۴۰۰ تایی به‌منظور راستی‌آزمایی دو مدل و مقایسه آن‌ها با هم درنظر گرفته شد.

نکته بسیار مهم این است که باید در زمان نمونه‌گیری تصادفی، در حجم نمونه نسبت تسهیلات معوق نشده (تسهیلات فاقد اقساط معوق) به کل تسهیلات در جامعه آماری (۳۲۱۹۲۵ فقره است) و همچنین نسبت تسهیلات معوق شده (تسهیلات با حداقل ۳ قسط معوق) به کل تسهیلات رعایت شود. در همین خصوص نیاز است که نمونه‌گیری از تسهیلات معوق نشده و تسهیلات معوق شده، به‌صورت جداگانه انجام شود. پس از بررسی جامعه آماری، تعداد تسهیلات معوق نشده ۲۶۳۸۳۵ فقره و تعداد تسهیلات معوق شده ۵۸۰۹۰ فقره است، یعنی تسهیلات معوق نشده ۰/۸۲ و تسهیلات معوق شده ۰/۱۸ است. بنابراین در حجم نمونه ۳۶۰۰ تایی باید تعداد ۲۹۵۲ فقره تسهیلات معوق نشده و تعداد ۶۴۸ فقره تسهیلات معوق شده وجود داشته باشد و همچنین در حجم نمونه ۴۰۰ تایی، تعداد ۳۲۸ فقره تسهیلات معوق نشده و تعداد ۷۲ فقره تسهیلات معوق شده وجود داشته باشد. نمونه‌گیری تصادفی با نسبت درست تسهیلات معوق نشده و تسهیلات معوق شده به جامعه آماری انجام و هر دو حجم نمونه برای اجرای مدل الگوریتم ژنتیک و رگرسیون پروبیت و همچنین راستی‌آزمایی هر دو مدل آماده شد.

مدل و متغیرهای تحقیق

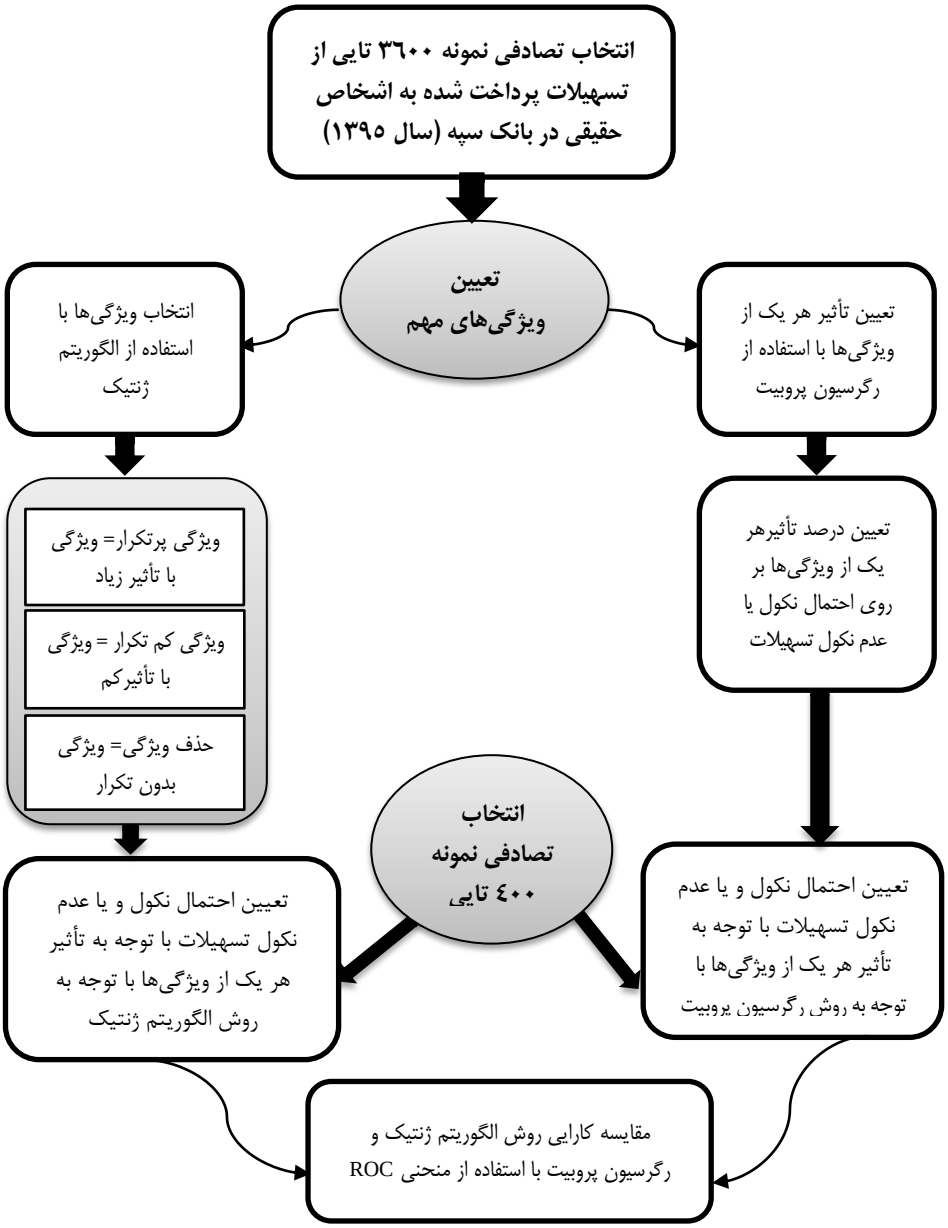
یکی از ویژگی‌ها و مزایای این تحقیق نسبت به سایر تحقیقات مشابه، وجود یک پایگاه داده با حجم بالا (تعداد مشتریان زیاد) و متغیرهای مستقل زیاد و مرتبط است. مرتبط بودن متغیرها از روی مرور ادبیات و مبانی نظری و پیشینه تحقیق استدلال می‌شود. روش کار تحقیق همان‌طورکه در نمودار (۱) دیده می‌شود، هر دو روش الگوریتم ژنتیک و رگرسیون پروبیت است که به انتخاب ویژگی (متغیر) خواهند پرداخت و در نهایت، عملکردشان باهم مقایسه می‌شود. در ذیل متغیرهای مستقل و وابسته آورده شده‌اند.

متغیرهای مستقل: نوع قرارداد (قرض‌الحسنه)، (مشارکت مدنی)، (مضاربه)، (سلف)، (فروش اقساطی مواد اولیه)، (فروش اقساطی ماشین‌آلات)، (فروش اقساطی مسکن)، (اجاره به‌شرط تملیک)، (جعاله)، (خرید دین)، (مرابحه)، رشته فعالیت (صنعت و معدن)، (کشاورزی)، (بازرگانی)، (مسکن و

ساختمان)، (خدمات)، مبلغ پرداختی، مبلغ بازگشتی، سود سهم مشتری، نرخ سود سهم مشتری، سود سهم دولت، نرخ سود سهم دولت، تعداد اقساط، مبلغ وثیقه، کد وثیقه (غیرمنقول)، (لازم الاجراء)، (چک)، (سفته)، (سپرده بانکی)، مبلغ پیش‌پرداخت، عادی، تکلیفی، روش تقسیط (پلکانی)، (عادی)، (تک قسطی)، (تنفسی)، (قرض‌الحسنه)، (تدریجی)، (سایر)، نرخ جریمه، شهر مورد مصرف (کلان‌شهر)، شهر مورد مصرف (غیرکلان‌شهر)، جنسیت (مرد)، (زن)، شغل (صاحبان مشاغل آزاد کلان)، (صاحبان مشاغل آزاد خرد)، (کارمند خصوصی)، (کارمند دولتی)، (نیروهای مسلح)، (خانه‌دار)، (بازنشسته)، (سایر)، اصل مبلغ پرداختی تسهیلات تا ۹۶/۴/۳۰ مانده بدهی مشتری تا ۹۶/۴/۳۰، سن، وضعیت مسکن (ملکی)، (استیجاری)، (سایر)، درآمد، میانگین موجودی حساب سه ماه آخر.

متغیر وابسته: معوقات (در این تحقیق زمانی یک تسهیلات را معوق تلقی می‌کنیم که حداقل سه قسط پرداخت نشده داشته باشد).

نمودار ۱: مدل تحقیق تعیین احتمال بازپرداخت یا عدم بازپرداخت با توجه به تأثیر ویژگی‌ها در الگوریتم ژنتیک



پس از انتخاب متغیرهای تحقیق، با استفاده از مدل الگوریتم ژنتیک به تعیین ویژگی‌های تأثیرگذار از بین متغیرهای مستقل اقدام می‌شود، که در سه سطح بدون تأثیر، کم تأثیر و با تأثیر زیاد تقسیم می‌شود. این مرحله با استفاده از نمونه ۳۶۰۰ تایی انجام می‌شود. در مرحله بعد نوبت به اجرای مدل رگرسیون پروبیت برای تعیین درصد تأثیر هر یک از متغیرهای مستقل بر متغیر وابسته می‌رسد که به‌وسیله همان نمونه ۳۶۰۰ تایی انجام می‌شود. در پایان درصد تأثیر هر یک از متغیرهای مستقل بر متغیر وابسته مشخص می‌شود. در مرحله آخر به‌وسیله نمونه ۴۰۰ تایی و به کمک سطح زیر منحنی ROC، توانایی مدل الگوریتم ژنتیک در مقابل مدل رگرسیون پروبیت و راستی‌آزمایی آن‌ها بررسی می‌شود. شمای کلی تحقیق و مراحل انجام آن، به‌ترتیب در نمودارهای (۱)، (۲) و (۳) آورده شده است.

الف- مدل الگوریتم ژنتیک: الگوریتم ژنتیک یک روش جستجوی کلی است که برای یافتن پاسخ‌های نزدیک به بهینه برای مسائل بهینه‌سازی غیرخطی به‌کار می‌رود. الگوریتم ژنتیک، الهامی از علم ژنتیک و نظریه تکامل داروین است و براساس بقای برترین‌ها یا انتخاب طبیعی استوار است. یک کاربرد متداول الگوریتم ژنتیک، استفاده از آن به‌عنوان تابع بهینه‌کننده است. الگوریتم ژنتیک ابزاری سودمند در بازشناسی الگو، انتخاب ویژگی، درک تصویر و یادگیری ماشینی است. به‌عبارت دیگر، الگوریتم‌های ژنتیک، الگوریتم‌های جستجو با هدف کلی هستند که از اصولی که از جمعیت ژنتیک طبیعی الهام گرفته شده برای تولید راه‌حل‌هایی برای مسائل جستجوی پیچیده استفاده می‌کنند.

در یک الگوریتم ژنتیک، جمعیتی از افراد طبق مطلوبیت آن‌ها در محیط بقا می‌یابند و افرادی با قابلیت‌های برتر، شانس ازدواج و تولید مثل بیشتری را خواهند یافت. بنابراین پس از چند نسل فرزندانی با کارایی بهتر به‌وجود می‌آیند. در الگوریتم ژنتیک هر فرد از جمعیت به‌صورت یک کروموزوم[1] معرفی می‌شود. کروموزوم‌ها طی چندین نسل کامل‌تر می‌شوند. در هر نسل کروموزوم‌ها ارزیابی می‌شوند و متناسب با ارزش خود امکان بقا و تکثیر می‌یابند. تولید نسل در بحث الگوریتم ژنتیک با عملگرهای همبری[2] و جهش[3] صورت می‌گیرد. والدین برتر براساس یک تابع برازندگی انتخاب می‌شوند.

در هر مرحله از اجرای الگوریتم ژنتیک، یک دسته از نقاط فضای جستجو مورد پردازش‌های تصادفی قرار می‌گیرند. به این صورت که به هر نقطه دنباله‌ای از کاراکترها نسبت داده می‌شود و بر روی این دنباله‌ها عملگرهای ژنتیکی اعمال می‌شود. سپس دنباله‌های به‌دست آمده رمزگشایی می‌شود تا نقاط جدیدی در فضای جستجو به‌دست آید. در نهایت بر اساس این که تابع هدف در هر

[1] Chromosome
[2] Cross Over
[3] Mutation

یک از نقاط چه مقدار باشد، احتمال شرکت نمودن آن‌ها در مرحله بعد تعیین می‌شود. الگوریتم‌های ژنتیک را می‌توان یک روش بهینه‌سازی تصادفی جهت‌دار دانست که به تدریج به سمت نقطه بهینه حرکت می‌کنند. در مورد ویژگی‌های الگوریتم ژنتیک در مقایسه با دیگر روش‌های بهینه‌سازی می‌توان گفت الگوریتمی است که بدون داشتن هیچ‌گونه اطلاعی از مسئله و هیچ‌گونه محدودیتی بر نوع متغیرهای آن، برای هر مسئله‌ای قابل اعمال است و کارایی اثبات شده‌ای در یافتن بهینه کلی[۱] دارد. توانایی این روش در حل مسائل پیچیده بهینه‌سازی است که روش‌های کلاسیک یا قابل اعمال نیستند و یا بهینه کلی قابل اطمینان نیست.

به‌طورکلی، الگوریتم‌های ژنتیک از اجزای زیر تشکیل می‌شوند:

۱- **کروموزوم:** در الگوریتم‌های ژنتیک، هر کروموزوم نشان‌دهنده یک نقطه در فضای جستجو و یک راه حل ممکن برای مسئله مورد نظر است. خود کروموزوم‌ها (راه‌حل‌ها) از تعداد ثابتی ژن[۲] (متغیر) تشکیل می‌شوند. برای نمایش کروموزوم‌ها معمولاً از کدگذاری‌های دودویی (رشته‌های بیتی) استفاده می‌شود.

۲- **جمعیت[۳]:** مجموعه‌ای از کروموزوم‌ها یک جمعیت را تشکیل می‌دهند. با تأثیر عملگرهای ژنتیکی بر هر جمعیت، جمعیت جدیدی با همان تعداد کروموزوم تشکیل می‌شود.

۳- **تابع برازندگی[٤]:** به‌منظور حل هر مسئله با استفاده از الگوریتم‌های ژنتیکی، ابتدا باید یک تابع برازندگی برای آن مسئله ابداع شود. برای هر کروموزوم، این تابع عددی غیرمنفی را برمی‌گرداند که نشان‌دهنده شایستگی یا توانایی فردی آن کروموزوم است.

٤- **عملگرهای ژنتیکی:** در الگوریتم‌های ژنتیک، در طی مرحله تولید مثل[۵] از عملگرهای ژنتیک استفاده می‌شود. با تأثیر این عملگرها بر روی یک جمعیت، نسل[۶] بعدی آن جمعیت تولید می‌شود. عملگرهای انتخاب[۷]، آمیزش[۸] و جهش معمولاً بیشترین کاربرد را در الگوریتم‌های ژنتیکی دارند. دارند.

---

[1] Global Optimum
[2] Gene
[3] Population
[4] Fitness Function
[5] Reproduction
[6] Generation
[7] Selection
[8] Crossover

عملگرهای ژنتیکی را می‌توان به‌صورت زیر شرح داد:

۱- **عملگر انتخاب:** این عملگر از بین کروموزوم‌های موجود در یک جمعیت، تعدادی کروموزوم را برای تولید مثل انتخاب می‌کند. کروموزوم‌های برازنده‌تر شانس بیشتری دارند تا برای تولید مثل انتخاب شوند.

۲- **عملگر آمیزش:** عملگر آمیزش بر روی یک زوج کروموزوم از نسل مولد عمل کرده و یک زوج کروموزوم جدید تولید می‌کند. عملگرهای آمیزش متعددی از قبیل، آمیزش تک نقطه‌ای[1] و آمیزش دو نقطه‌ای وجود دارد. در آمیزش تک نقطه‌ای، یک موقعیت تصادفی بین دو ژن در نظر گرفته می‌شود. سپس تمامی ژن‌های طرف راست یا طرف چپ این موقعیت در کروموزوم‌های والد با یکدیگر جابه‌جا می‌شوند تا کروموزوم‌های جدید به‌دست آیند. در آمیزش دو نقطه‌ای، دو موقعیت به‌صورت تصادفی انتخاب می‌شود و تمامی ژن‌ها بین این دو موقعیت در کروموزوم‌های والد با یکدیگر جابه‌جا می‌شوند. شایان ذکر است که آمیزش معمولاً بر روی همه زوج کروموزوم‌های انتخاب شده برای جفت‌گیری به‌کار برده نمی‌شود. معمولاً احتمال آمیزش برای هر زوج کروموزوم بین ۰/۶ تا ۰/۹۵ در نظر گرفته می‌شود که به این عدد نرخ آمیزش[2] یا احتمال آمیزش آمیزش گفته می‌شود و با Pc نمایش داده می‌شود. در صورتی‌که روی یک زوج کروموزوم عمل آمیزش صورت نگیرد، فرزندان با تکرار نمودن والدین تولید می‌شوند.

۳- **عملگر جهش:** پس از اتمام عمل آمیزش, عملگر جهش روی کروموزوم‌ها اثر داده می‌شود. این عملگر یک ژن از یک کروموزوم را به‌طور تصادفی انتخاب می‌کند و سپس محتوای آن ژن را تغییر می‌دهد. اگر ژن از جنس اعداد دودویی باشد، آن را به وارونش تبدیل می‌کند و چنانچه متعلق به یک مجموعه باشد، مقدار یا عنصر دیگری از آن مجموعه را به‌جای آن ژن قرار می‌دهد.

در این تحقیق پس از انتخاب متغیرهای تأثیرگذار، با استفاده از مدل الگوریتم ژنتیک اقدام به تعیین ویژگی‌های تأثیرگذار از بین متغیرهای مستقل می‌شود که در سه سطح بدون تأثیر، کم تأثیر و با تأثیر زیاد تقسیم می‌شود. تابع بهینه شده با الگوریتم ژنتیک در این تحقیق، تابع اشتباه در رده‌بندی[3] است که توسط شاخص‌های معروفی نظیر ضریب جینی[4] و یا آنتروپی[5] تعریف می‌شود. این مرحله با

[1] One-point Crossover
[2] Crossover Rate
[3] Misclassification
[4] Gini
[5] Entropy

با استفاده از نمونه ۳۶۰۰ تایی انجام می‌شود. مراحل انجام و چگونگی انتخاب ویژگی‌های تأثیرگذار با استفاده از الگوریتم ژنتیک در نمودار (۲) نشان داده شده است.

ب- مدل رگرسیون پروبیت: در مدل‌های آماری، تحلیل رگرسیون یک فرایند آماری برای تخمین روابط بین متغیرها است. این فرایند وقتی‌که تمرکز روی روابط بین متغیر وابسته و یک یا چند متغیر مستقل باشد، شامل روش‌های زیادی برای مدل‌سازی و تحلیل متغیرهای خاص و منحصر به‌فرد است. یکی از ویژگی‌های مورد نیاز هنگام اجرای مدل‌های رگرسیون، پیوسته بودن متغیر وابسته است. با وجود این، در برخی از موارد کاربردی برحسب نیاز، متغیرهای وابسته مورد نظر گسسته هستند. در ساده‌ترین حالت فرض کنیم که متغیر وابسته y_i $(i=1,2,...,n)$ دو مقدار $\{0,1\}$ را اخذ می‌کند و می‌خواهیم $P(y_i=0|\mathbf{x}_i)=1-P\left(y_i=1\middle|\mathbf{x}_i\right)$ را محاسبه کنیم که بردار $\mathbf{x}_i$ مجموعه‌ای از متغیرهای تأثیرگذار در مسئله است. حال فرض کنیم متغیری مانند z_i وجود دارد که $z_i=-\beta'\mathbf{x}_i+\varepsilon_i$ و در آن علامت $'$ نشان‌دهنده ترانهاده یک بردار است و همچنین $y_i=\begin{cases}1 & z_i\le 0\\ 0 & z_i>0\end{cases}$ می‌شود که ε_iها مستقل و هم‌توزیع هستند و F_{ε_i} نیز تابع توزیع ε_i است. در این صورت،

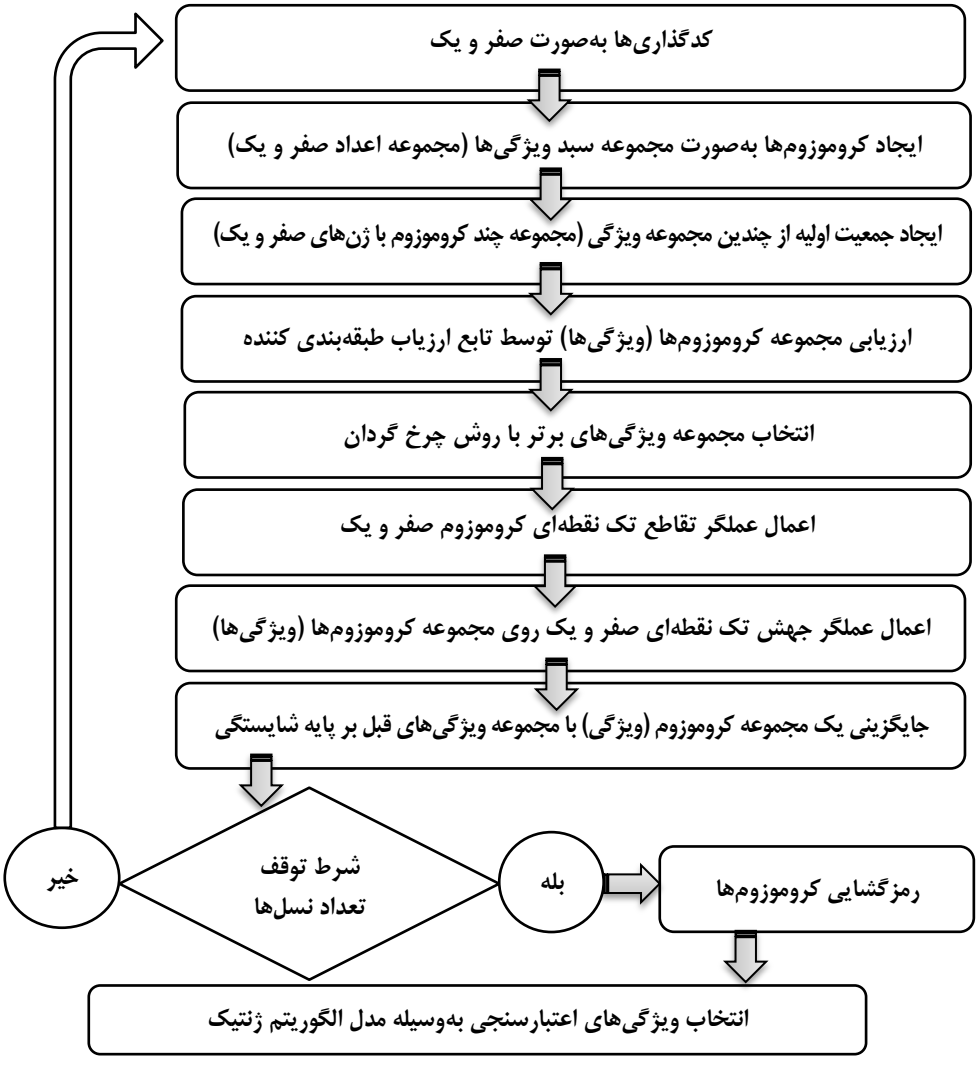
$$P(y_i=1|\mathbf{x}_i)=P(z_i\le 0|\mathbf{x}_i)=P(\varepsilon_i\le\beta'\mathbf{x}_i)=F_{\varepsilon_i}(\beta'\mathbf{x}_i)$$

با فرض $p_i=F_{\varepsilon_i}(\beta'\mathbf{x}_i)$، مشاهده می‌شود که $\beta'\mathbf{x}_i=F_{\varepsilon_i}^{-1}(p_i)$. یک رهیافت متعارف در این حالت استفاده از رگرسیون‌های لاجیت و پروبیت است. در این دو حالت، به‌ترتیب ε_iها را دارای توزیع لجستیک و نرمال استاندارد فرض می‌کنیم. در حالت توزیع لجستیک، $F_{\varepsilon_i}^{-1}(p_i)=\ln(\frac{p_i}{1-p_i})$ رگرسیون لجستیک تولید و در حالت توزیع نرمال، $F_{\varepsilon_i}^{-1}(p_i)=\Phi^{-1}(p_i)$ رگرسیون پروبیت ایجاد می‌شود.

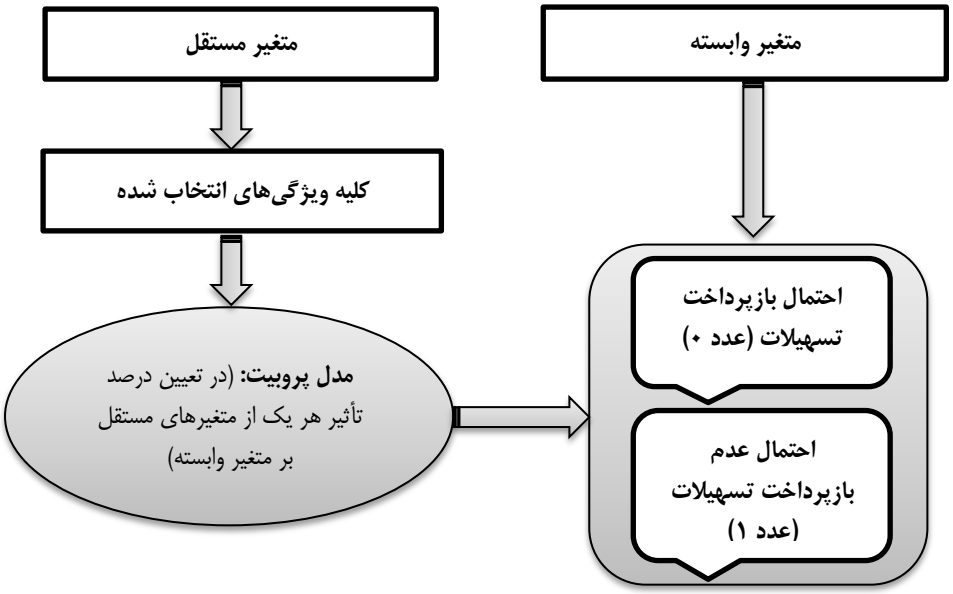
در حالتی‌که توزیع تجمعی پاسخ‌ها به‌شرط حضور متغیرهای توضیح‌گر از توزیع نرمال تبعیت کند، پیشنهاد می‌شود به‌جای رگرسیون لاجیت از رگرسیون پروبیت استفاده شود و در صورتی‌که توزیع تجمعی داده‌های متغیر وابسته از توزیع نرمال تبعیت نکند، اگر تعداد متغیرهای مستقل بسیار زیاد باشد تحلیل لاجیت ترجیح داده می‌شود. با این استدلال و اینکه تعداد متغیرهای به‌کار رفته در مقایسه با متغیرهای به‌کار رفته در مقالات لاتین مشابه متوسط به‌نظر می‌رسد، رگرسیون پروبیت ترجیح داده

شد. در نمودار (۳) جریان چگونگی تعیین مقدار تأثیر ویژگی‌ها بر متغیر وابسته با استفاده از مدل پروبیت به تفصیل بیان شده است.

نمودار ۲: نمودار جریان چگونگی انتخاب ویژگی‌ها توسط الگوریتم ژنتیک



نمودار ۳: نمودار جریان چگونگی تعیین مقدار تأثیر ویژگی‌ها بر متغیر وابسته با استفاده از مدل رگرسیون پروبیت



ج- منحنی ROC: منحنی مشخصه کارکرد گیرنده، یعنی منحنی ROC، یک نمودار گرافیکی است که توانایی تشخیص یک سیستم طبقه‌بندی دودویی را نشان می‌دهد. منحنی ROC نمودار پراکندگی حساسیت برای یک سیستم صفر و یک است و فضای زیر این منحنی برای ارزیابی دقت پیش‌بینی مدل‌های طبقه‌بندی به‌کار می‌رود. برای تشکیل نمودارهای ROC، به دو طبقه‌بندی واقعی و پیش‌بینی شده نیاز است. برای ترسیم نمودار ROC باید محورهای افقی و عمودی تعریف شوند. محور افقی، درصد مثبت کاذب[1] (FP) و محور عمودی، درصد مثبت واقعی[2] (TP) در نظر گرفته می‌شود. در واقع این منحنی با رسم TP در برابر FP به‌ازای برش‌های گوناگون ایجاد می‌شود. میزان TP نیز به‌عنوان حساسیت یا احتمال تشخیص در روش‌های یادگیری ماشین شناخته می‌شود. در واقع TP نسبت متغیرهای مثبتی که صحیح طبقه‌بندی شده‌اند به کل متغیرهای مثبت است. معیار FP نیز درصد مشاهدات منفی است که مثبت پیش‌بینی شده‌اند (همانند خطای نوع دوم در آزمون

[1] False Positive
[2] True Positive

فرض). همچنین می‌توان از FP-۱ به‌عنوان یک توان تشخیص (توان آزمون در آزمون فرض) در مسائل تصمیم‌گیری استفاده کرد.

هر چه نقاط به سمت بالا و چپ نزدیک‌تر باشد مدل پیش‌بینی مناسب‌تر و به حالت ایده آل خود نزدیک‌تر است. مختصات نقطه‌ی (۱ و ۰) حالت ایده آل است. این نقطه نشان می‌دهد که آنچه مدل پیش‌بینی ارائه می‌دهد، انطباق کامل بر مدل واقعی دارد. نقطه مقابل آن دارای مختصات (۰ و ۱) است و به این معنی است که هر چه مدل پیش‌بینی ارائه داده است، عکس مدل واقعی است. در صورتی‌که مدل تصادفی عمل کند، نقاط در اطراف خط Y=X قرار می‌گیرند.

جدول ۱: حالت‌های گوناگون در طبقه‌بندی داده‌ها در مدل‌های پیش‌بینی (ماتریس ارزیابی عملکرد)

Predicted Negative Class **(منفی پیش‌بینی‌شده)**	Predicted Positive Class **(مثبت پیش‌بینی‌شده)**	
FN (False Negative) منفی کاذب	TP (True positive) مثبت واقعی	Actual Positive Class **(مثبت واقعی)**
TN (True Negative) منفی واقعی	FP(False Positive) مثبت کاذب	Actual Negative Class **(منفی واقعی)**

جدول (۱) که همان ماتریس درهم‌ریختگی یا ماتریس ارزیابی عملکرد است، مفاهیم TP و FP را شرح می‌دهد. به‌عنوان مثال در بخش‌بندی مشتریان به خوش‌حساب (مثبت) و بدحساب (منفی) (نکول‌کننده و نکول‌نکننده تسهیلات)، TP نشان‌دهنده درصد افراد خوش‌حسابی است که در عمل هم توسط بانک خوش‌حساب در نظر گرفته شده‌اند و FP بیانگر درصد افراد بدحسابی است که توسط بانک خوش‌حساب تلقی شده‌اند. بقیه سلول‌های این ماتریس مشابه تعریف می‌شود. حال فرض کنیم بخواهیم منحنی ROC رگرسیون لجستیک را رسم کنیم. پس از برازش رگرسیون لجستیک و محاسبه احتمال نکول به‌ازای همه افراد، به ازای برش‌های گوناگون (آستانه‌های گوناگون برای احتمال نکول که اگر احتمال نکول فردی از آن آستانه (برش) بالاتر شد آن فرد را بدحساب (نکول‌کننده تسهیلات) تلقی کنیم)، مقادیر TP و FP را محاسبه و در مقابل هم رسم می‌کنیم و بدین ترتیب منحنی ROC رگرسیون لجستیک به‌دست می‌آید. برای الگوریتم ژنتیک هم با یافتن احتمال نکول به طریقی مشابه عمل می‌کنیم.

یافته‌های تحقیق و تجزیه و تحلیل آن‌ها

آمار توصیفی

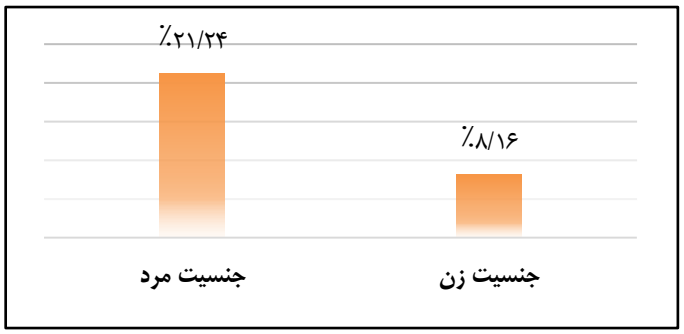
برای تشریح علت انتخاب تعدادی از متغیرهای این تحقیق در جدول (۲) درصد نکول تسهیلات در متغیر مورد نظر برای کل جامعه آورده شده که با توضیحات مختصری علت انتخاب متغیر مورد نظر بیان می‌شود. در همین خصوص شهرهای ایران به دو گروه کلان‌شهر و غیرکلان‌شهر تقسیم شد. تعداد تسهیلات پرداخت شده در کلان‌شهرها در سال ۱۳۹۵ برابر ۱۱۴۱۱۰ فقره که معادل ۳۵/۴۴۴ درصد و تعداد تسهیلات پرداخت‌شده در غیرکلان‌شهرها در سال ۱۳۹۵ برابر ۲۰۷۸۳۱ فقره که معادل ۶۴/۵۵۵ درصد است. آمار توصیفی مربوط به کلان‌شهرها و غیرکلان‌شهرها در جدول (۲) آورده شده است.

جدول ۲: آمار توصیفی در ارتباط با کلان‌شهرها و غیرکلان‌شهرها

شرح	شهر مورد نظر	کل تعداد تسهیلات	تعداد تسهیلات نکول شده (معوق)	درصد
کل تسهیلات پرداختی در سال ۱۳۹۵	کلان‌شهرها	۱۱۴۱۱۰	۱۸۴۲۲	٪۱۶/۱۴۵
کل تسهیلات پرداختی در سال ۱۳۹۵	غیرکلان شهرها	۲۰۷۸۳۱	۳۹۵۲۷	٪۱۹/۰۱۳
کل تسهیلات پرداختی به مبلغ ۲۰۰ میلیون ریال در سال ۱۳۹۵	کلان‌شهرها	۱۲۶۹۸	۱۳۴۱	٪۱۰/۵۶۱
کل تسهیلات پرداختی به مبلغ ۲۰۰ میلیون ریال در سال ۱۳۹۵	غیرکلان شهرها	۲۲۶۲۵	۴۸۲۹	٪۲۱/۳۴۲
کل تسهیلات پرداختی به مبلغ ۱۰۰ میلیون ریال در سال ۱۳۹۵	کلان‌شهرها	۲۸۲۵۹	۴۵۶۱	٪۱۶/۱۳۸
کل تسهیلات پرداختی به مبلغ ۱۰۰ میلیون ریال در سال ۱۳۹۵	غیرکلان شهرها	۴۷۸۴۳	۶۲۶۰	٪۱۳/۰۸۰
کل تسهیلات پرداختی به مبلغ ۵۰ میلیون ریال در سال ۱۳۹۵	کلان‌شهرها	۸۹۴	۸۹	٪۱۰/۰۰۰
کل تسهیلات پرداختی به مبلغ ۵۰ میلیون ریال در سال ۱۳۹۵	غیرکلان شهرها	۷۲۴۴	۷۱۵	٪۹/۸۷۹

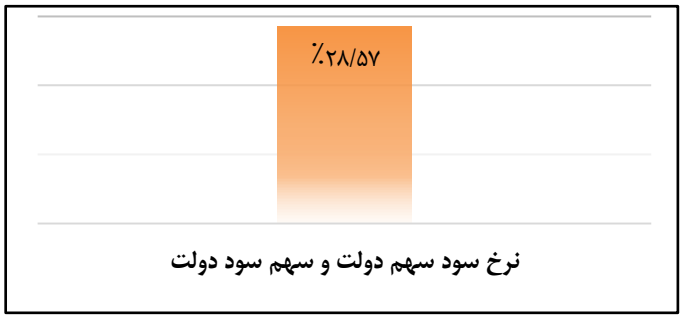
آمار توصیفی مربوط به متغیر نوع جنسیت (زن و مرد) و درصد نکول تسهیلات برای این دو متغیر در نمودار (۴) آورده شده است. تعداد تسهیلات پرداخت شده به جنسیت مرد در سال ۱۳۹۵ برابر ۲۷۹۶۸۱ فقره که معادل ۸۶/۸۷۴ درصد و تعداد تسهیلات پرداخت شده به جنسیت زن در سال ۱۳۹۵ برابر ۴۲۲۶۰ فقره که معادل ۱۳/۱۲۶ درصد است.

نمودار ٤: درصد نکول تسهیلات در جنسیت



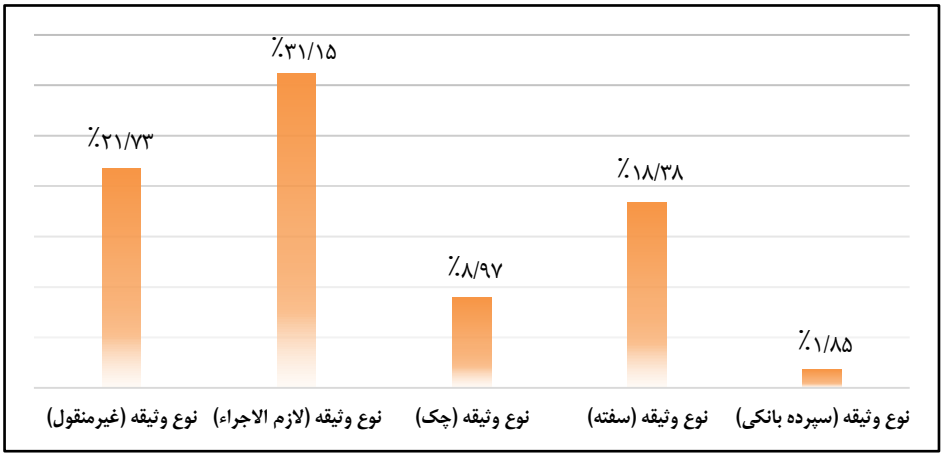
تعداد تسهیلات پرداخت شده که دارای نرخ سود سهم دولت و سهم سود دولت است، در سال ۱۳۹۵ برابر ۲۹۰۴۸ فقره که معادل ۹/۰۲۲ درصد است. آمار توصیفی مربوط به نرخ سود سهم دولت و سهم سود دولت در نمودار (۵) بیان شده است. اما نکته مهم مربوط به نرخ سود سهم دولت است، بدین صورت‌که هرچه این نرخ بیشتر باشد، درصد نکول تسهیلات کمتر است و هر چه این نرخ کمتر باشد، درصد نکول تسهیلات بیشتر می‌شود.

نمودار ٥: درصد نکول تسهیلات در نرخ سود سهم دولت و سهم سود دولت



همچنین آمار توصیفی مربوط به متغیر نوع وثایق که ۵ فقره است در نمودار (۶) آورده شده است. تعداد تسهیلات پرداخت شده به نوع وثیقه (غیرمنقول) در سال ۱۳۹۵ برابر ۸۷۸۱۴ فقره که معادل ۲۷/۲۷۶ درصد، تعداد تسهیلات پرداخت شده به نوع وثیقه (لازم الاجراء) در سال ۱۳۹۵ برابر ۱۳۰۷۱۲ فقره است که معادل ۴۰/۶۰۱ درصد، تعداد تسهیلات پرداخت شده به نوع وثیقه (چک) در سال ۱۳۹۵ برابر ۴۰۱۰۲ فقره که معادل ۱۲/۴۵۶ درصد، تعداد تسهیلات پرداخت شده به نوع وثیقه (سفته) در سال ۱۳۹۵ برابر ۱۴۳۹۱ فقره که معادل ۴/۴۷۰ درصد و تعداد تسهیلات پرداخت شده به نوع وثیقه (سپرده بانکی) در سال ۱۳۹۵ برابر ۴۸۹۲۲ فقره می‌باشد که معادل ۱۵/۱۹۵درصد است.

نمودار ٦: درصد نکول تسهیلات در وثایق مختلف



مدل الگوریتم ژنتیک

پارامترهای الگوریتم ژنتیک در این تحقیق در جدول (۳) ارائه شده است.

جدول ۳: پارامترهای الگوریتم ژنتیک

پارامتر	مقدار
تعداد نسل و جمعیت اولیه	۶۰
تابع ارزیاب	5.4 C
نرخ تقاطع	۰/۹
نرخ جهش	۰/۰۱
عدد تصادفی	۱

برای تشریح این پارامترها، می‌توان گفت که تعداد نسل اولیه تعداد ویژگی‌هایی است که می‌خواهیم از بین آن‌ها انتخاب کنیم. تابع ارزیاب تابعی است که در هر مرحله از تکامل، ژن‌های قوی و ضعیف را از هم جدا می‌کند که در اینجا الگوریتم مبتنی بر درخت تصمیم c۵.۴ از مبحث داده کاوی انتخاب می‌شود. تقاطع برای جابه‌جایی ژن‌ها بین کروموزوم‌ها در استخر جفت‌گیری استفاده می‌شود. هدف این فرایند جایگزینی ژن‌های ضعیف‌تر کروموزوم‌ها با قوی‌ترهای سایر کروموزوم‌هاست تا روی جواب انتخاب شده تغییراتی را اعمال کند و آن را به‌سوی بهینگی سوق دهد. نرخ تقاطع نسبت تعداد نوزادان تولید شده در هر نسل به اندازه جمعیت اصلی تعریف می‌شود. این نرخ، تعداد مورد انتظار کروموزوم‌هایی را که توسط عملگر تقاطع دچار تغییر می‌شوند، تعیین می‌کند. نرخ جهش برای عدم گیرافتادن در جواب‌های بهینه و محلی و هدایت درست به‌سوی بهترین جواب با حداکثر پوشش فضای جواب است. در واقع نرخ جهش عبارت است از درصدی از کل تعداد ژن‌های موجود که دچار تغییر می‌شوند. اگر نرخ جهش خیلی کوچک باشد، تعداد زیادی از ژن‌ها که می‌توانستند مفید باشند تست نمی‌شوند. اما اگر نرخ جهش خیلی بزرگ باشد، اختلالی تصادفی به‌وجود آمده و نوزادان و والدین شباهتشان را از دست خواهند داد. در روش چرخ گردان که اساسی‌ترین روش برای تعیین احتمال بقا برای هر کروموزوم است، در هر مرحله یک یا چند عدد تصادفی برای اجرای الگوریتم انتخاب می‌شود. پارامتر عدد تصادفی تعیین می‌کند که در هر مرحله از تکامل الگوریتم چند عدد تصادفی بایستی انتخاب شود.

با استفاده از روش انتخاب ویژگی‌ها مبتنی بر الگوریتم ژنتیک، تعداد ۲۴ ویژگی از بین ۶۰ ویژگی مورد استفاده در رتبه‌بندی اعتباری مشتریان حذف شد. همچنین تعداد ۲۰ ویژگی از بین ۶۰ ویژگی مورد استفاده مهم در نظر گرفته شده‌اند و در پایان مراحل الگوریتم ژنتیک باقی ماندند. ضمناً باقی ویژگی‌ها که ۱۶ فقره هستند جزء متغیرهای کم تکرار در مدل الگوریتم ژنتیک تشخیص داده شدند. این ویژگی‌ها در جدول (۴) آورده شده است.

جدول ٤: ترکیب ویژگی‌ها تأثیرگذار با استفاده از مدل الگوریتم ژنتیک

ویژگی‌های باقی‌مانده (پرتکرار)	ویژگی‌های کم تکرار	ویژگی‌های حذف شده (بدون تکرار)
نوع قرارداد (قرض‌الحسنه)	نوع قرارداد (مشارکت مدنی)	نوع قرارداد (سلف)
نوع قرارداد (مضاربه)	نوع قرارداد (فروش اقساطی مواد اولیه)	نوع قرارداد (فروش اقساطی مسکن)
نوع قرارداد (فروش اقساطی ماشین‌آلات)	نوع قرارداد (جعاله)	نوع قرارداد (اجاره به‌شرط تملیک)
نوع قرارداد (مرابحه)	رشته فعالیت (صنعت و معدن)	نوع قرارداد (خرید دین)
رشته فعالیت (خدمات)	تعداد اقساط	رشته فعالیت (کشاورزی)
مبلغ پرداختی	کد وثیقه (غیرمنقول)	رشته فعالیت (بازرگانی)
مبلغ بازگشتی	روش تقسیط (عادی)	رشته فعالیت (مسکن و ساختمان)
سود سهم مشتری	جنسیت (مرد)	سود سهم دولت
نرخ سود سهم مشتری	جنسیت (زن)	نرخ سود سهم دولت
کد وثیقه (لازم الاجراء)	شغل (صاحبان مشاغل آزاد خرد)	کد وثیقه (چک)
کد وثیقه (سپرده بانکی)	شغل (کارمند خصوصی)	کد وثیقه (سفته)
مبلغ وثیقه	شغل (نیروهای مسلح)	مبلغ پیش‌پرداخت
نرخ جریمه	سن	عادی یا تکلیفی (عادی)
شهر مورد مصرف (کلان‌شهر)	وضعیت مسکن (سایر)	عادی یا تکلیفی (تکلیفی)
شهر مورد مصرف (غیر کلان‌شهر)	اصل مبلغ پرداختی تسهیلات	روش تقسیط (پلکانی)
شغل (صاحبان مشاغل آزاد کلان)	مانده بدهی مشتری	روش تقسیط (تک قسطی)
شغل (کارمند دولتی)		روش تقسیط (تنفسی)
درآمد		روش تقسیط (قرض‌الحسنه)
میانگین موجودی سه ماه آخر		روش تقسیط (تدریجی)
وضعیت مسکن (ملکی)		روش تقسیط (سایر)
		شغل (خانه‌دار)
		شغل (بازنشسته)
		شغل (سایر)
		وضعیت مسکن (استیجاری)

در جدول (۴) پس از اجرای مدل الگوریتم ژنتیک مشخص شد که کدام یک از ویژگی‌ها تأثیر بیشتری بر احتمال نکول تسهیلات دارد، کدام متغیر تاثیر کمتری دارد و کدام یک از متغیرها تاثیری بر احتمال نکول تسهیلات ندارد. در واقع، در الگوریتم ژنتیک برخی از ژن‌ها (ویژگی‌ها) غالب هستند (پرتکرارند) و برخی دیگر مغلوب که این نتایج در جدول (۴) آمده است. در این مرحله و پس از تعیین ویژگی‌ها اقدام به طراحی مدل اعتبارسنجی با استفاده از الگوریتم ژنتیک در نمونه ۳۶۰۰ تایی می‌کنیم. مدل طراحی شده با استفاده از الگوریتم ژنتیک در نرم‌افزار متلب قادر است به‌وسیله داده‌های همان ۶۰ ویژگی، پیش‌بینی کند که شخص حقیقی پس از دریافت تسهیلات، آیا این تسهیلات را نکول می‌نماید یا خیر؟

برای آزمون و راستی‌آزمایی مدل طراحی شده با استفاده از الگوریتم ژنتیک، مدل فوق را با نمونه ۴۰۰ تایی راستی‌آزمایی کرده و نتایج ذیل حاصل شد.

- از بین ۳۲۸ فقره تسهیلاتی که معوق نشده بودند، تعداد ۳۱۶ فقره را صحیح تشخیص داد. این قسمت در ماتریس درهم ریختگی همان قسمت TP یعنی بخش مثبت واقعی است.
- از بین ۳۲۸ فقره تسهیلاتی که معوق نشده بودند، تعداد ۱۲ فقره را صحیح تشخیص نداد. این قسمت در ماتریس درهم ریختگی همان قسمت FP یعنی بخش مثبت کاذب است.
- از بین ۷۲ فقره تسهیلاتی که معوق شده بودند، تعداد ۵۱ فقره را صحیح تشخیص داد. این قسمت در ماتریس درهم ریختگی همان قسمت TN یعنی بخش منفی واقعی است.
- از بین ۷۲ فقره تسهیلاتی که معوق شده بودند، تعداد ۲۱ فقره را صحیح تشخیص نداد. این قسمت در ماتریس درهم ریختگی همان قسمت FP یعنی بخش منفی کاذب است.

نکته قابل توجه در این قسمت روش تشخیص احتمال نکول و احتمال عدم نکول تسهیلات در نمونه ۴۰۰تایی به روش الگوریتم ژنتیک است. روش تشخیص احتمال نکول یا عدم نکول تسهیلات به این صورت است که در روش الگوریتم ژنتیک، متغیر وابسته که همان احتمال نکول و عدم نکول تسهیلات است را در بازه صفر تا یک در نظر می‌گیریم. این عدد بیان‌کننده احتمال نکول و عدم نکول تسهیلات است، بدین‌ترتیب که اگر این عدد از ۰/۵ کوچک‌تر باشد، بیان‌کننده وجود احتمال نکول تسهیلات است. همچنین اگر این عدد از ۰/۵ بزرگ‌تر باشد، احتمال عدم نکول تسهیلات را بیان می‌کند. اگر مقدار متغیر وابسته در مدل الگوریتم ژنتیک برابر با ۰/۵ باشد، در این حالت احتمال نکول تسهیلات و احتمال عدم نکول تسهیلات در این قرارداد تسهیلات با هم برابر است. ضمناً بیان این نکته بسیار ضروری است که هر چه این مقدار از عدد ۰/۵ فاصله گرفته و به عدد صفر نزدیک

شود، احتمال نکول تسهیلات بیشتر و بیشتر می‌شود و بالعکس. به‌طور مثال اگر مقدار متغیر وابسته برای دو فقره از تسهیلات برابر با ۰/۴ و ۰/۱ شود، بیان‌کننده این مطلب است که تسهیلاتی که مقدار عددی متغیر وابسته آن برابر با ۰/۱ شود، از تسهیلاتی که مقدار عددی متغیر وابسته آن برابر با ۰/۴ شود، احتمال نکول بسیار بیشتری دارد و همچنین تسهیلاتی که مقدار عددی متغیر وابسته آن ۰/۱ شود، از تسهیلاتی که مقدار عددی متغیر وابسته آن برابر با ۰/۴ شود، برای بانک ریسک اعتباری بسیار بالاتری دارد.

پس از انجام راستی‌آزمایی با استفاده از نمونه ۴۰۰ تایی به‌وسیله مدل طراحی‌شده در الگوریتم ژنتیک، برای امکان مقایسه با نتایج حاصل از مدل رگرسیون پروبیت، نتایج به‌دست آمده با استفاده از مدل ROC در نرم‌افزار متلب اجرا و ماتریس درهم ریختگی و نمودار ROC حاصل به‌شرح جدول (۵) به‌دست آمد.

جدول ۵: ماتریس درهم‌ریختگی مربوط به مدل الگوریتم ژنتیک

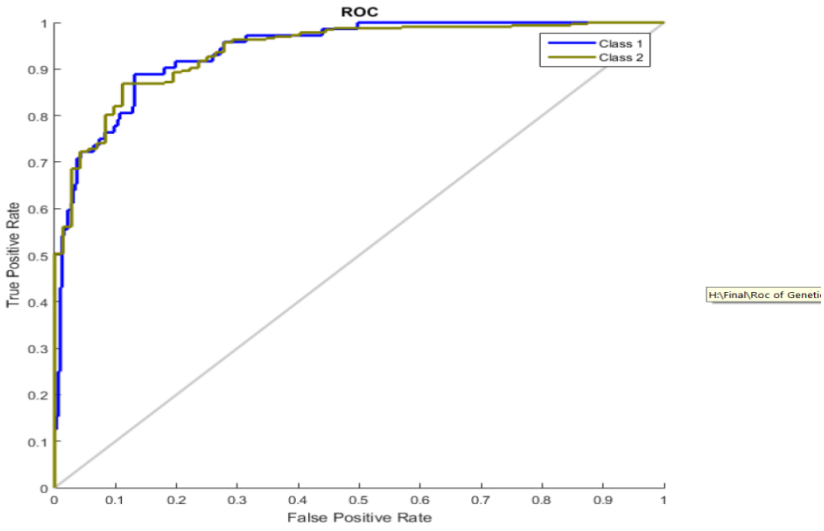
Class۱	۵۱ ٪۱۲/۸	۱۲ ٪۳	٪۸۱ ٪۱۹
Class۲	۲۱ ٪۵/۳	۳۱۶ ٪۷۹	٪۹۳/۸ ٪۶/۲
	٪۷۰/۸ ٪۲۹/۲	٪۹۶/۳ ٪۳/۷	٪۹۱/۸ ٪۸/۳
	Class۱	Class۲	

جدول (۵) نشان می‌دهد که (ماتریس درهم ریختگی) در دو کلاس طبقه‌بندی شده است. کلاس ۱ نشان‌دهنده تسهیلات معوق شده در نمونه ۴۰۰ تایی است که تعداد این تسهیلات ۷۲ فقره است. از بین این ۷۲ فقره تسهیلات معوق شده، تعداد ۵۱ فقره را صحیح (TN) (منفی واقعی) و تعداد ۲۱ فقره را اشتباه (FN) (منفی کاذب) پیش‌بینی کرده است. به بیان دیگر در کلاس ۱ مدل طراحی شده توانسته است ۷۰/۸٪ صحیح و ۲۹/۲٪ اشتباه پیش‌بینی کند. در این ماتریس، کلاس ۲ نشان‌دهنده تسهیلات معوق نشده در نمونه ۴۰۰ تایی است که تعداد این تسهیلات ۳۲۸ فقره است. از بین این ۳۲۸ فقره تسهیلات معوق نشده، تعداد ۳۱۶ فقره را صحیح (TP) (مثبت واقعی) و تعداد ۱۲ فقره را اشتباه (FP)

(مثبت کاذب) پیش‌بینی کرده است. به بیان دیگر در کلاس ۲ مدل طراحی شده توانسته است ۷۹٪ صحیح و ۲۱٪ اشتباه پیش‌بینی کند. ضمناً مشاهده می‌شود که هرچه منحنی به سمت بالا و چپ کشیدگی داشته باشد، مدل بیشتر توانایی پیش‌بینی داشته است، یعنی مدل طراحی شده به‌وسیله الگوریتم ژنتیک در نمونه ۴۰۰ تایی توانسته است که در مجموع ۹۱/۸٪ صحیح و ۸/۳٪ اشتباه پیش‌بینی کند.

نکته نهایی در این مرحله مساحت زیر منحنی ROC (نمودار (۷)) است که در روش الگوریتم ژنتیک در کلاس ۱ مساحت زیر منحنی ROC برابر ۰/۹۳ و در کلاس ۲ مساحت زیر منحنی ROC برابر ۰/۹۱ است. یکی از روش‌های تشخیص توانایی یک مدل، مساحت سطح زیر منحنی ROC است که هر چه این سطح بزرگ‌تر باشد، نشان می‌دهد که مدل توانایی بهتری در پیش‌بینی دارد.

نمودار ۷: منحنی ROC مربوط به مدل الگوریتم ژنتیک



مدل رگرسیون پروبیت

در این مرحله با رگرسیون پروبیت و نمونه ۳۶۰۰ تایی در نرم‌افزار EViews اقدام به طراحی مدل اعتبارسنجی می‌نماییم و همچنین تعیین می‌کنیم که وزن هر یک از این ۶۰ فقره ویژگی در نکول شدن و نکول نشدن تسهیلات به اشخاص حقیقی در بانک سپه چه مقدار است. قبل از اجرای

مدل رگرسیون پروبیت نیاز است که متغیرهای کیفی و متغیرهای کمی را جداگانه آماده کنیم. در ابتدا کلیه متغیرهای کمی همانند مدل الگوریتم ژنتیک به‌شرح جدول (۶) نام‌گذاری و آماده می‌شوند.

جدول ٦: متغیرهای کمی تحقیق برای اجرای مدل رگرسیون پروبیت

متغیر	شرح ویژگی
۱X	مبلغ پرداختی
۲X	مبلغ بازگشتی
۳X	سود سهم مشتری
۴X	سود سهم دولت
۵X	نرخ سود سهم مشتری
۶X	نرخ سود سهم دولت
۷X	تعداد اقساط
۸X	مبلغ وثیقه
۹X	مبلغ پیش‌پرداخت
۱۰X	نرخ جریمه
۱۱X	اصل مبلغ پرداختی تسهیلات توسط مشتری
۱۲X	مانده بدهی مشتری
۱۳X	سن
۱۴X	درآمد
۱۵X	میانگین موجودی حساب سه ماه آخر

سپس نوبت به متغیرهای کیفی می‌رسد. در این دسته از متغیرها با n سطح متمایز، برای اجرای رگرسیون‌ها و برای جلوگیری از هم‌خطی، باید به تعداد (n−۱) متغیر مجازی به‌ازای هر سطح متغیر کیفی به‌جز یکی از سطوح ایجاد شود. به‌عنوان مثال، متغیر کیفی وضعیت مسکن که دارای ۳ سطح مجزا است در این بخش به دو متغیر مجازی ۳۶R و ۳۷R تقسیم می‌شود و متغیر ۳۸R که مربوط به سطح سوم وضعیت مسکن است برای جلوگیری از هم‌خطی در این قسمت حذف می‌شود. شرح کامل متغیرهای اسمی در جدول (۷) آورده شده است.

جدول ۷: متغیرهای مجازی تحقیق برای اجرای مدل رگرسیون پروبیت

نام کامل ویژگی	متغیر	نوع ویژگی
نوع قرارداد (قرض‌الحسنه)	۱R	نوع قرارداد
نوع قرارداد (مشارکت مدنی)	۲R	نوع قرارداد
نوع قرارداد (مضاربه)	۳R	نوع قرارداد
نوع قرارداد (فروش اقساطی مواد اولیه)	۴R	نوع قرارداد
نوع قرارداد (فروش اقساطی ماشین‌آلات)	۵R	نوع قرارداد
نوع قرارداد (فروش اقساطی مسکن)	۶R	نوع قرارداد
نوع قرارداد (اجاره به‌شرط تملیک)	۷R	نوع قرارداد
نوع قرارداد (جعاله)	۸R	نوع قرارداد
نوع قرارداد (خرید دین)	۹R	نوع قرارداد
رشته فعالیت (صنعت و معدن)	۱۱R	رشته فعالیت
رشته فعالیت (کشاورزی)	۱۲R	رشته فعالیت
رشته فعالیت (مسکن و ساختمان)	۱۴R	رشته فعالیت
رشته فعالیت (خدمات)	۱۵R	رشته فعالیت
عادی یا تکلیفی (عادی)	۱۶R	عادی یا تکلیفی
روش تقسیط (عادی)	۱۸R	روش تقسیط
روش تقسیط (تک قسطی)	۱۹R	روش تقسیط
روش تقسیط (تنفسی)	۲۰R	روش تقسیط
روش تقسیط (قرض‌الحسنه)	۲۱R	روش تقسیط
روش تقسیط (تدریجی)	۲۲R	روش تقسیط
شهر مورد مصرف (کلان‌شهر)	۲۴R	کد شهر مورد مصرف
جنسیت (مرد)	۲۶R	جنسیت
شغل (صاحبان مشاغل آزاد کلان)	۲۸R	شغل
شغل (صاحبان مشاغل آزاد خرد)	۲۹R	شغل
شغل (کارمند خصوصی)	۳۰R	شغل
شغل (کارمند دولتی)	۳۱R	شغل
شغل (نیروهای مسلح)	۳۲R	شغل
شغل (خانه‌دار)	۳۳R	شغل
شغل (بازنشسته)	۳۴R	شغل
وضعیت مسکن (ملکی)	۳۶R	وضعیت مسکن
وضعیت مسکن (استیجاری)	۳۷R	وضعیت مسکن
کد وثیقه (غیرمنقول)	۳۹R	نوع وثیقه
کد وثیقه (لازم‌الاجراء)	۴۰R	نوع وثیقه
کد وثیقه (چک)	۴۱R	نوع وثیقه
کد وثیقه (سفته)	۴۲R	نوع وثیقه

جدول ۸: متغیرهای مجازی حذف‌شده در مدل پروبیت

نام کامل ویژگی حذف‌شده	متغیر	نوع ویژگی
نوع قرارداد (مرابحه)	۱۰R	نوع قرارداد
رشته فعالیت (بازرگانی)	۱۳R	رشته فعالیت
عادی یا تکلیفی (تکلیفی)	۱۷R	عادی یا تکلیفی
روش تقسیط (سایر)	۲۳R	روش تقسیط
شهر مورد مصرف (غیرکلان‌شهر)	۲۵R	کد شهر مورد مصرف
جنسیت (زن)	۲۷R	جنسیت
شغل (سایر)	۳۵R	شغل
وضعیت مسکن (سایر)	۳۸R	وضعیت مسکن
کد وثیقه (سپرده بانکی)	۴۳R	نوع وثیقه

پس از آماده‌سازی متغیرها نوبت به اجرای مدل رگرسیون پروبیت با استفاده از نمونه ۳۶۰۰ تایی در نرم‌افزار EViews است که نتایج اجرای مدل به شرح جدول (۹) است.

جدول ۹: نتایج حاصل از مدل رگرسیون پروبیت

متغیر	ضریب	انحراف استاندارد	z	مقدار احتمال
R1	−۱/۰۴۷۳۳۸	۰/۰۵۷۷۱۶	−۱۸/۱۴۶۴۶	۰/۰۰۰۰
R2	−۱/۵۱۶۹۵۷	۰/۰۸۵۰۲۹	−۱۷/۸۴۰۵۳	۰/۰۰۰۰
R3	−۱/۵۴۶۶۳۵	۰/۱۵۵۰۸۱	−۹/۹۷۳۰۶۵	۰/۰۰۰۰
R4	−۰/۹۴۶۷۸۲	۰/۱۸۵۲۸۸	−۵/۱۰۹۷۸	۰/۰۰۰۰
R5	−۰/۶۱۹۳۰۷	۰/۱۰۳۸۶۷	−۵/۹۶۲۰۰۸	۰/۰۰۰۰
R6	−۰/۷۰۶۳۰۳	۰/۱۹۴۵۳	−۳/۶۳۰۸۰۶	۰/۰۰۰۳
R7	−۰/۴۳۰۷۲۷	۰/۵۲۹۹۵۴	−۰/۸۱۲۷۶۴	۰/۴۱۶۴
R8	−۰/۳۹۷۶۲۲	۰/۱۷۴۱۶	−۲/۲۸۳۰۷۹	۰/۰۲۲۴
R9	−۱/۰۶۷۵۷۱	۰/۵۸۶۸۷۶	−۱/۸۱۹۰۷۴	۰/۰۶۸۹
R11	−۰/۲۹۶۴۹۳	۰/۱۱۱۸۹۵	−۲/۶۴۹۷۳۶	۰/۰۰۸۱
R12	−۰/۵۲۱۰۴۷	۰/۰۹۸۵۸	−۵/۲۸۵۵۴۲	۰/۰۰۰۰
R14	−۰/۵۳۲۱۳۸	۰/۱۳۲۶۳۳	−۴/۰۱۲۱۲	۰/۰۰۰۱
R15	−۰/۰۱۴۹۱۵	۰/۰۵۷۱۱۷	−۰/۲۶۱۱۳۲	۰/۷۹۴۰
R18	۰/۰۸۳۳۰۳	۰/۲۲۷۸۴۱	۰/۳۶۵۶۱۷	۰/۷۱۴۷
R19	−۰/۷۲۲۳۰۷	۰/۲۳۹۶۵۹	−۳/۰۱۳۸۹۸	۰/۰۰۲۶
R21	−۰/۲۰۵۷۱۷	۰/۲۳۳۱۴۶	−۰/۸۸۲۳۵۱	۰/۳۷۷۶

ادامه جدول ۹: نتایج حاصل از مدل رگرسیون پروبیت

متغیر	ضریب	انحراف استاندارد	Z	مقدار احتمال
R22	۰/۲۷۵۶۷۲	۰/۵۵۰۷۹۶	۰/۵۰۰۴۹۸	۰/۶۱۶۷
R28	−۱/۱۳۴۳۸۷	۰/۲۸۵۵۱۶	−۳/۹۷۳۱۱۲	۰/۰۰۰۱
R29	−۱/۱۵۳۲۸۶	۰/۱۵۳۲۵	−۷/۵۲۵۵۴	۰/۰۰۰۰
R30	−۰/۶۰۰۵۶۳	۰/۱۴۸۹۲۸	−۴/۰۳۲۵۶۴	۰/۰۰۰۱
R31	−۰/۷۸۲۶۷۹	۰/۱۵۳۲۵۷	−۵/۱۰۶۹۷۱	۰/۰۰۰۰
R32	−۱/۱۸۲۴۴۴	۰/۲۰۸۱۲۶	−۵/۶۸۱۳۹۱	۰/۰۰۰۰
R33	−۱/۳۸۱۸۶۵	۰/۲۳۸۴۱۴	−۵/۷۹۶۰۷۷	۰/۰۰۰۰
R34	−۱/۷۲۷۲۴۸	۰/۳۴۴۹۴۹	−۵/۰۰۷۲۵۶	۰/۰۰۰۰
R39	۱/۶۴۵۵۹۵	۰/۳۶۶۰۱۳	۴/۴۹۶۰۰۸	۰/۰۰۰۰
R40	۱/۶۵۴۹۹۹	۰/۳۴۸۲۴۳	۴/۷۵۲۴۲۷	۰/۰۰۰۰
R41	۲/۶۱۲۳۵۵	۰/۳۷۹۶۰۱	۶/۸۸۱۸۶۴	۰/۰۰۰۰
R42	۱/۵۱۹۷۷۴	۰/۳۹۹۶۷	۳/۸۰۲۵۷۲	۰/۰۰۰۱
C	۳/۱۲۱۱۹۹	۰/۵۶۶۸۹۳	۵/۵۰۵۷۹۶	۰/۰۰۰۰
R16	۰/۲۵۶۵۵۲	۰/۳۴۰۵	۰/۷۵۳۴۵۷	۰/۴۵۱۲
R24	−۰/۱۳۲۸۰۸	۰/۰۹۷۷۵۴	−۱/۳۵۸۵۹۳	۰/۱۷۴۳
R36	−۰/۸۱۶۱۸۷	۰/۱۴۵۳۳۵	−۵/۶۱۵۸۹۹	۰/۰۰۰۰
R37	۰/۴۸۶۸۱۶	۰/۱۴۵۹۷۱	۳/۳۳۵۰۱۴	۰/۰۰۰۹
X1	-۶/۸۸E-۰۹	۶/۷۶E-۰۹	−۱/۰۱۷۲۶	۰/۳۰۹۰
X2	۱/۳۵E-۰۸	۷/۶۴E-۰۹	۱/۷۶۱۶۶۴	۰/۰۷۸۱
X3	۲/۴E-۰۹	۷/۴۳E-۰۹	۰/۳۲۲۵۰۹	۰/۷۴۷۱
X4	-۲/۸۴E-۰۹	۱/۲۹E-۰۹	−۲/۲۱۲۸۷۷	۰/۰۲۶۹
X5	−۰/۰۸۵۵۸	۰/۰۴۰۹۱۱	−۲/۰۹۱۸۵۹	۰/۰۳۶۵
X6	−۰/۰۶۱۳۰۳	۰/۰۳۵۶۲۸	−۱/۷۲۰۶۲۸	۰/۰۸۵۳
X7	۰/۰۱۷۶۵۵	۰/۰۰۴۲۶۲	۴/۱۴۱۹۹۵	۰/۰۰۰۰
X8	-۲/۶۶E-۱۰	۴/۳۸E-۱۰	−۰/۶۰۸۵۱	۰/۵۴۲۸
X9	۶/۵۶E-۰۹	۳/۵۹E-۰۹	۱/۸۲۹۳	۰/۰۶۷۴
X10	۰/۰۹۱۵۹۸	۰/۰۴۲۲۷۱	۲/۱۶۶۸۹۸	۰/۰۳۰۲
X11	-۸/۱۸E-۰۹	۲/۲۳E-۰۹	−۳/۶۶۶۷۱۳	۰/۰۰۰۲
X12	-۱/۰۸E-۰۸	۱/۳۸E-۰۹	−۷/۸۰۲۱۶	۰/۰۰۰۰
X13	−۰/۰۷۵۹۹۶	۰/۰۰۵۸۷۳	−۱۲/۹۴۰۶۲	۰/۰۰۰۰
X14	-۵/۸۴E-۰۸	۴/۲۶E-۰۹	−۱۳/۷۰۶۱۱	۰/۰۰۰۰
X15	-۷/۲۶E-۰۹	۹/۹۵E-۱۰	−۷/۲۹۶۲۲۷	۰/۰۰۰۰
Log likelihood: −۱۹۳۹/۱۱۷				

پس از تعیین وزن هر یک از ویژگی‌ها در احتمال نکول و عدم نکول تسهیلات به‌وسیله رگرسیون پروبیت به کمک نمونه ۳۶۰۰ تایی، مدل اعتبارسنجی آماده شد. در این مرحله مدل اعتبارسنجی آماده شده به‌وسیله منحنی ROC و به کمک نمونه ۴۰۰ تایی مورد راستی‌آزمایی قرار می‌گیرد.

برای انجام راستی‌آزمایی مدل اعتبارسنجی آماده شده به‌وسیله رگرسیون پروبیت، نیاز است که نمونه ۴۰۰ تایی را در مدل طراحی شده اجرا کرد. با اجرای اعتبارسنجی در نمونه نتایج زیر به‌دست آمد:

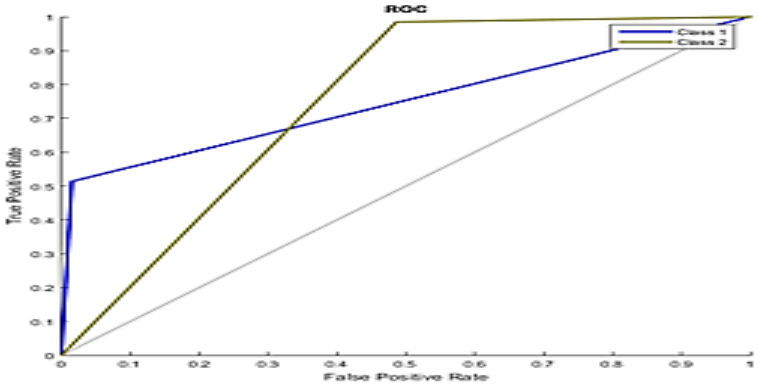
- از بین ۳۲۸ فقره تسهیلاتی که معوق نشده بودند، تعداد ۳۲۳ فقره را صحیح تشخیص داد. این قسمت در ماتریس درهم‌ریختگی همان قسمت TP یعنی بخش مثبت واقعی است.
- از بین ۳۲۸ فقره تسهیلاتی که معوق نشده بودند، تعداد ۵ فقره را صحیح تشخیص نداد. این قسمت در ماتریس درهم‌ریختگی همان قسمت FP یعنی بخش مثبت کاذب است.
- از بین ۷۲ فقره تسهیلاتی که معوق شده بودند، تعداد ۳۷ فقره را صحیح تشخیص داد. این قسمت در ماتریس درهم‌ریختگی همان قسمت TN یعنی بخش منفی واقعی است.
- از بین ۷۲ فقره تسهیلاتی که معوق شده بودند، تعداد ۳۵ فقره را صحیح تشخیص نداد. این قسمت در ماتریس درهم‌ریختگی همان قسمت FP یعنی بخش منفی کاذب است.

در این بخش روش تشخیص احتمال نکول یا عدم نکول تسهیلات در مدل رگرسیون پروبیت در نمونه ۴۰۰ تایی، همانند تشخیص در مدل الگوریتم ژنتیک است. پس از انجام راستی‌آزمایی به‌وسیله مدل طراحی شده در رگرسیون پروبیت، برای امکان مقایسه این نتایج با نتایج حاصل از مدل الگوریتم ژنتیک، ماتریس درهم‌ریختگی و نمودار ROC حاصل به شرح جدول (۱۰) آورده شده است.

جدول ۱۰: ماتریس درهم‌ریختگی مربوط به مدل رگرسیون پروبیت

Class۱	۳۷ ٪۹/۳	۵ ٪۱/۳	٪۸۸/۱ ٪۱۱/۹
Class۲	۳۵ ٪۸/۸	۳۲۳ ٪۸۰/۸	٪۹۰/۲ ٪۹/۸
	٪۵۱/۴ ٪۴۸/۶	٪۹۸/۵ ٪۱/۵	٪۹۰ ٪۱۰
	Class۱	Class۲	

نمودار ۸: منحنی ROC مربوط به مدل رگرسیون پروبیت



جدول (۱۰) در دو کلاس طبقه‌بندی شده است. کلاس ۱ که نشان‌دهنده تسهیلات معوق شده در نمونه ۴۰۰ تایی است و تعداد این تسهیلات ۷۲ فقره است. از بین این ۷۲ فقره تسهیلات معوق شده، تعداد ۳۷ فقره صحیح و تعداد ۳۵ فقره اشتباه پیش‌بینی شده است. به بیان دیگر در کلاس ۱ مدل طراحی شده توانسته است ۵۱/۴٪ صحیح و همچنین ۴۸/۶٪ اشتباه پیش‌بینی نماید. در این ماتریس کلاس ۲ نشان‌دهنده تسهیلات معوق نشده در نمونه ۴۰۰ تایی است و تعداد این تسهیلات ۳۲۸ فقره است. از بین این ۳۲۸ فقره تسهیلات معوق نشده، تعداد ۳۲۳ فقره صحیح و تعداد ۵ فقره اشتباه پیش‌بینی شده است. به بیان دیگر در کلاس ۲ مدل طراحی شده توانسته است ۹۸/۵٪ صحیح و همچنین ۱/۵٪ اشتباه پیش‌بینی نماید.

نکته نهایی در این مرحله مساحت سطح زیر منحنی ROC است که درروش رگرسیون پروبیت در کلاس ۱ مساحت زیر منحنی ROC (نمودار (۸)) برابر ۰/۷۱ و در کلاس ۲ مساحت زیر منحنی ROC برابر ۰/۷۳ است. یکی از روش‌های تشخیص توانایی یک مدل مساحت سطح زیر منحنی ROC است که هر چه این سطح بزرگ‌تر باشد، نشان می‌دهد که مدل توانایی بهتری در پیش‌بینی دارد، یعنی مدل طراحی شده در رگرسیون پروبیت در نمونه ۴۰۰ تایی توانسته است در مجموع ۹۰٪ صحیح پیش‌بینی نماید و ۱۰٪ اشتباه پیش‌بینی نماید.

تفسیر نتایج نمودار ROC

قبل از بررسی ماتریس‌ها و نمودارهای ROC برای تفسیر نتایج، توجه به متغیرهای تاثیرگذار و مهم از بین کلیه متغیرهای مستقل تحقیق که به‌وسیله مدل الگوریتم ژنتیک و رگرسیون پروبیت انتخاب شده‌اند، بسیار ضروری است.

جدول ۱۱: جدول متغیرهای تأثیرگذار در دو مدل الگوریتم ژنتیک و رگرسیون پروبیت

متغیرهای پرتکرار در مدل الگوریتم ژنتیک	متغیرهای با ضریب معنی‌داری بالا در مدل پروبیت
نوع قرارداد (قرض‌الحسنه)	نوع قرارداد (مضاربه)
نوع قرارداد (مضاربه)	نوع قرارداد (مشارکت مدنی)
نوع قرارداد (فروش اقساطی ماشین‌آلات)	رشته فعالیت (بازرگانی)
نوع قرارداد (مرابحه)	رشته فعالیت (مسکن و ساختمان)
نوع قرارداد (خدمات)	شغل (بازنشسته)
مبلغ پرداختی	شغل (خانه‌دار)
مبلغ بازگشتی	شغل (نیروهای مسلح)
سود سهم مشتری	کد وثیقه (سپرده بانکی)
نرخ سود سهم مشتری	وضعیت مسکن (ملکی)
کد وثیقه (لازم الاجراء)	وضعیت مسکن (سایر)
کد وثیقه (سپرده بانکی)	سود سهم دولت
مبلغ وثیقه	تعداد اقساط
نرخ جریمه	نرخ سود سهم مشتری
شهر مورد مصرف (کلان‌شهر)	نرخ جریمه
شهر مورد مصرف (غیرکلان‌شهر)	سن
شغل (صاحبان مشاغل آزاد کلان)	درآمد
شغل (کارمند دولتی)	میانگین موجودی سه ماه آخر
درآمد	
میانگین موجودی سه ماه آخر	
وضعیت مسکن (ملکی)	

در جدول (۱۱) به‌طور واضح بیان شده که در دو مدل الگوریتم ژنتیک و مدل رگرسیون پروبیت کدام‌یک از متغیرها و ویژگی‌ها تأثیر بیشتری بر متغیر وابسته یعنی احتمال نکول تسهیلات دارند. این

تفاوت در متغیرها از تفاوت در روش کار این دو مدل سرچشمه می‌گیرد. نکته قابل تامل در جدول (۱۱) این است که مسئولان اعتبارات شعب در زمان پرداخت تسهیلات به مشتریان حقیقی باید این ویژگی‌ها را بسیار زیاد مورد توجه قرار دهند و در این ویژگی‌ها تامل کنند.

با توجه به داده‌های جداول (۵) و (۱۰) (ماتریس درهم‌ریختگی)، مشاهده می‌شود که مدل طراحی شده به‌وسیله رگرسیون پروبیت در کلاس ۲، از مدل طراحی شده به‌وسیله الگوریتم ژنتیک توانایی بهتری در پیش‌بینی داشته است. کلاس ۲ مربوط به تسهیلات معوق نشده در نمونه ۴۰۰ تایی است که تعداد این تسهیلات ۳۲۸ فقره است. مدل الگوریتم ژنتیک در کلاس ۲، ۳۱۶ فقره را درست و ۱۲ فقره را اشتباه پیش‌بینی کرده است، اما مدل رگرسیون پروبیت در کلاس ۲، ۳۲۳ فقره را درست و تنها ۵ فقره را اشتباه پیش‌بینی کرده است.

همچنین با توجه به داده‌های جداول (۵) و (۱۰) (ماتریس درهم‌ریختگی)، مشاهده می‌شود که مدل طراحی شده به‌وسیله الگوریتم ژنتیک در کلاس ۱، از مدل طراحی شده به‌وسیله رگرسیون پروبیت توانایی بهتری در پیش‌بینی داشته است. کلاس ۱ مربوط به تسهیلات معوق شده در نمونه ۴۰۰ تایی و تعداد این تسهیلات ۷۲ فقره است. مدل الگوریتم ژنتیک در کلاس ۱، ۵۱ فقره را درست و ۲۱ فقره را اشتباه پیش‌بینی کرده است، اما مدل رگرسیون پروبیت در کلاس ۱، تنها ۳۷ فقره را درست و ۳۵ فقره را اشتباه پیش‌بینی کرده است. بنابراین در تجزیه و تحلیل کلاس ۱ به این نتیجه می‌رسیم که مدل طراحی شده به‌وسیله الگوریتم ژنتیک بهتر از مدل طراحی شده به‌وسیله رگرسیون پروبیت است.

بحث و نتیجه‌گیری

با تجزیه و تحلیل ماتریس درهم‌ریختگی این نتیجه حاصل می‌شود که مدل طراحی شده به‌وسیله الگوریتم ژنتیک از بین ۴۰۰ فقره تسهیلات، تعداد ۳۶۷ فقره برابر با ۹۱/۸٪ را درست و تعداد ۳۳ فقره برابر با ۸/۳٪ را اشتباه پیش‌بینی کرده است، اما مدل طراحی شده به‌وسیله رگرسیون پروبیت از بین ۴۰۰ فقره تسهیلات، تعداد ۳۶۰ فقره برابر با ۹۰٪ را درست و تعداد ۴۰ فقره برابر با ۱۰٪ را اشتباه پیش‌بینی کرده است.

با توجه به توضیحات نمودار ROC، مشاهده شد که هر دو منحنی ROC مربوط به کلاس ۱ و ۲ در مدل طراحی شده توسط الگوریتم ژنتیک به نقطه (۱ و۰) نزدیک‌تر است و همچنین مساحت زیر منحنی ROC در روش الگوریتم ژنتیک برابر ۰/۹۲، اما در روش رگرسیون پروبیت برابر ۰/۷۲ است که بیان‌کننده پیش‌بینی بهتر مدل الگوریتم ژنتیک نسبت به مدل رگرسیون پروبیت است، اما انطباق

کامل بر مدل واقعی ندارد؛ زیرا منحنی ROC در هر دو نمودار به نقطه (۱ و۰) نمی‌رسد، اما منحنی ROC مدل طراحی شده به‌وسیله الگوریتم ژنتیک نسبت به نمودار ROC رگرسیون پروبیت به نقطه (۱ و۰) بسیار نزدیک‌تر است که مساحت زیر منحنی نیز این مطلب را تأیید می‌کند؛ زیرا مساحت زیر منحنی در روش الگوریتم ژنتیک بیشتر از مساحت زیر منحنی در روش رگرسیون پروبیت است.

در مورد الگوریتم ژنتیک، متغیرهای نوع قرارداد (قرض‌الحسنه)، نوع قرارداد (مضاربه)، نوع قرارداد (فروش اقساطی ماشین‌آلات)، نوع قرارداد (مرابحه)، رشته فعالیت (خدمات)، مبلغ پرداختی، مبلغ بازگشتی، سود سهم مشتری، نرخ سود سهم مشتری، کد وثیقه (لازم‌الاجراء)، کد وثیقه (سپرده بانکی)، مبلغ وثیقه، نرخ جریمه، شهر مورد مصرف (کلان‌شهر)، شغل (صاحبان مشاغل آزاد کلان)، شغل (کارمند دولتی) و درآمد مهم‌ترین متغیرهایی هستند که در پیش‌بینی به‌کار می‌روند.

همچنین دیده می‌شود که در رگرسیون پروبیت به‌جز متغیرهای روش تقسیط (عادی)، روش تقسیط (تدریجی)، کد وثیقه (غیرمنقول)، کد وثیقه (لازم‌الاجراء)، کد وثیقه (چک)، کد وثیقه (سفته)، وضعیت مسکن (استیجاری)، مبلغ پرداختی، مبلغ بازگشتی، تعداد اقساط، مبلغ پیش‌پرداخت و نرخ جریمه که تاثیر منفی بر احتمال بازپرداخت تسهیلات دارند، مابقی متغیرها اثر مثبت دارند.

با توجه به این مطلب که مدل‌های رتبه‌بندی فعلی معمولاً از روش‌های رگرسیونی استفاده می‌کنند و نیاز به تعدادی متغیر مهم در مدل‌هایشان دارند که عدم وجود این متغیرها باعث نتایج خطادار می‌شود و همچنین تعدادی از متغیرها در حال حاضر در بانک سپه به‌دلیل سیستم نرم‌افزاری قدیمی، وجود ندارد، استفاده از روش الگوریتم ژنتیک می‌تواند بسیار مهم باشد و این مشکل را حل کند. همچنین پیش‌بینی بهتر این روش در مقایسه با یکی از مدل‌های رگرسیونی، دلیل دیگری بر مزیت این روش است.

پیشنهادات کاربردی

۱- به‌کارگیری روش‌های هوش مصنوعی الگوریتم ژنتیک در رتبه‌بندی اعتباری مشتریان حقیقی بانک،

۲- توسعه مدل این تحقیق در رتبه‌بندی اعتباری مشتریان بانک با نرم‌افزار کاربردی رتبه‌بندی اعتباری،

۳- ایجاد یک بانک اطلاعاتی جامع به‌منظور دریافت و ثبت مناسب داده‌های رتبه‌بندی اعتباری مشتریان بانک با استفاده از الگوریتم‌های جدید همچون الگوریتم مورچگان،

۴- طراحی و ساخت یک سیستم پشتیبانی تصمیم‌گیری برای رتبه‌بندی اعتباری مشتریان بانک،

۵- با توجه به توصیه کمیته بال در این زمینه بهتر است بانک‌ها برای رتبه‌بندی اعتباری مشتریان خود به یک روش بسنده نکنند و روش‌ها و مدل‌های مختلفی را برای این امر به‌کار برند.

منابع و مآخذ

الف. منابع فارسی

اسکندری جعفری، میثم و روحی، میلاد. (۱۳۹۶): ***مدیریت ریسک اعتباری مشتریان بانکی با استفاده از روش ماشین بردار تصمیم بهبودیافته با الگوریتم ژنتیک با رویکرد داده کاوی***. فصلنامه مدیریت دارایی و تامین مالی ۱۹، ۱۷-۳۲.

اقبالی، علی، رضوی حاجی آقا، سیدحسن و عموزاد مهدیرجی، حنان. (۱۳۹۶): ***ارزیابی مقایسه‌ای عملکرد توابع شایستگی الگوریتم ژنتیک در رتبه‌بندی مشتریان***. نشریه مدیریت صنعتی ۹، ۲۴۵-۲۶۴.

البرزی، خان‌بابایی و محمدپورزرندی (۱۳۸۹): ***رتبه‌بندی اعتباری مشتریان حقیقی بانک سپه (مطالعه موردی شعب شهر تهران)***. پایان‌نامه کارشناسی ارشد. دانشگاه آزاد اسلامی واحد تهران مرکز.

ذکاوت، س، م، (۱۳۸۱): ***رتبه‌بندی داخلی مشتریان بانک‌ها با استفاده از مدل‌های رگرسیون لاجیت***. پایان‌نامه کارشناسی ارشد. دانشگاه تهران.

رستم‌کلایی (۱۳۸۶): ***مدل‌سازی ریسک اعتباری در بانک کشاورزی***. رویکرد مدل‌های لاجیت، پروبیت و شبکه‌های عصبی. رساله دکتری. دانشگاه علامه طباطبائی.

فلاح شمس، میرفیض و مهدوی راد، حمید. (۱۳۹۱): ***طراحی مدل اعتبار سنجی و پیش‌بینی ریسک اعتباری مشتریان تسهیلات لیزینگ (مورد مطالعه: شرکت لیزینگ ایران خودرو)***. فصلنامه پژوهشنامه اقتصادی (رویکرد اسلامی- ایرانی) ۴۴، ۲۱۳-۲۳۴.

کشاورز حداد و آیتی گازار (۱۳۸۶): ***ارزیابی عملکرد و رتبه‌بندی واحد اعتبارات و تسهیلات مؤسسه اعتباری توسعه***. پایان‌نامه کارشناسی ارشد. دانشکده صنایع دانشگاه تربیت مدرس.

ملا ابراهیم لو (۱۳۸۴): ***بررسی عوامل مؤثر بر ریسک اعتباری مشتریان حقوقی بانک (بررسی موردی بانک کشاورزی).*** پایان‌نامه کارشناسی ارشد. دانشگاه شهید بهشتی.

مهرگان، محمدرضا، تهرانی، رضا و نظری، عبدالله. (۱۳۹۵): ***اعتبارسنجی مشتریان بانک با استفاده از خوشه بندی به روش*** **K-Means**. مطالعات کمی در مدیریت ۲۵، ۱۲۳-۱۴۱.

ب. منابع فارسی

Abdou, H. A. (2018): ***Genetic programming for credit scoring: The case of Egyptian public sector banks***. Expert Systems with Applications 39, 11402-11417.

Abdou, H. A. Pointon, J., and Elmasry, A. (2008): ***Neural network versus conventional techniques in credit scoring in Egyptian bank***. Expert Systems with Applications 35, 1275-1292.

Chen, M. & Huang, S. (2003): ***Credit scoring and rejected instances reassigning through evolutionary computation techniques***. Expert Systems with Applications 24, 433–441.

Constangioara, A. (2011): ***Consumer credit scoring***. Romanian Journal of Economic Forecasting 3, 162-177.

Desai, V. S., Crook, J. N., & Overstreet, G. A. (1997): ***A comparison of neural networks and linear scoring models in the credit union environment***. European Journal of Operational Research 95, 24–37.

Huang, C., Chen, M., & Wang, C. (2007): ***Credit scoring with a data mining approach based on support vector machines***. Expert Systems with Applications 33, 847–856.

Lee, T., Chiu, C., Lu, C., & Chen, I. (2006): ***Credit scoring using the hybrid neural discriminant technique***. Expert Systems with Applications 23, 245–254.

Lund, A. (2015): ***Two-stage logistic regression models for improved credit scoring***. Master's Thesis in Computer Science. Royal Institute of Technology, Stockholm. Sweden.

Mathew, A. (2017): ***Credit scoring using logistic regression***. Master's Theses and Graduate Research. San Jose State University. USA.

Ong, C., Huang, J., & Tzeng, G. (2005): ***Building credit scoring models using genetic programming***. Expert Systems with Applications 29, 41–47.

Paliwal, M., & Kumar, U. A. (2009): ***Neural networks and statistical techniques: A review of applications***. Expert Systems with Applications 36, 2–17.

Ju, Y., and Sohn, S. Y. (2017): ***Technology credit scoring based on a quantification method***. Sustainability 9, 1-16.