پژوهش‌های نوین در تصمیم‌گیری
دوره ۱۰، شماره ۱، بهار ۱٤۰٤، صص ۳۱-٦٥
نوع مقاله: پژوهشی

اعتبارسنجی مشتریان بانک همراه با تعیین بهینه پارامترهای تسهیلات با استفاده از مدل شبیه‌سازی-بهینه‌یابی

امیر خرمی [۱]، محمود دهقان نیری [*۲]، علی رجب زاده [۳]

۱. دانشجوی دکتری مدیریت صنعتی، دانشکده مدیریت و اقتصاد، دانشگاه تربیت مدرس، تهران، ایران
۲. دانشیار، گروه مدیریت صنعتی، دانشکده مدیریت و اقتصاد، دانشگاه تربیت مدرس، تهران، ایران
۳. استاد، گروه مدیریت صنعتی، دانشکده مدیریت و اقتصاد، دانشگاه تربیت مدرس، تهران، ایران.

تاریخ دریافت: ۱٤۰۳/۰۹/۲۰ تاریخ پذیرش: ۱٤۰۳/۱۲/۱٥

چکیده

در این مقاله، روشی جدید برای اعتبارسنجی و تعیین پارامترهای بهینه تسهیلات بانک‌ها توسط رویکرد شبیه‌سازی-بهینه‌یابی ارائه شده است. روش پیشنهادی شامل سه مرحله آماده‌سازی داده‌ها، مدل اعتبارسنجی و مدل شبیه‌سازی-بهینه‌یابی می‌باشد. در آماده‌سازی داده‌ها، اطلاعات تسهیلات بانکی و صورت‌های مالی شرکت‌ها گردآوری شده و ویژگی‌های مورد نیاز محاسبه می‌شوند. انتخاب ویژگی‌های مهم توسط الگوریتم حداقل افزونگی حداکثر ارتباط (MRMR) انجام می‌گیرد. سپس برای حل مسئله اعتبارسنجی، از پنج روش طبقه‌بندی شامل رگرسیون لجستیک (LR)، ماشین بردار پشتیبان (SVM)، شبکه عصبی مصنوعی (ANN)، تقویت گرادیان شدید (XGB) و جنگل تصادفی (RF) استفاده می‌شود. عملکرد این مدل‌ها توسط معیارهایی چون دقت، نمره F۱ و سطح زیر منحنی (AUC) ارزیابی شده و بهترین مدل برای مرحله بعد انتخاب می‌شود. در مرحله شبیه‌سازی-بهینه‌یابی، مشخصات بهینه تسهیلات اعطایی به مشتریان با هدف حداقل‌سازی نرخ نکول تسهیلات انجام می‌گیرد. برای این منظور، اندازه تسهیلات، نرخ بهره و مدت زمان بازپرداخت تسهیلات به عنوان متغیرهای مسئله بهینه‌سازی در نظر گرفته می‌شوند. حل مسئله بهینه‌سازی توسط الگوریتم ممتیک (MA) در چهار حالت صورت می‌گیرد. در الگوریتم ممتیک، برای تخمین احتمال نکول مشتریان، از مدل اعتبارسنجی از پیش آموزش دیده استفاده می‌شود. مطالعه موردی بر روی داده‌های ۱۰۰۰ مشتری حقوقی یک بانک تجاری در ایران صورت گرفته است. از بین ۳۰ ویژگی تعریف شده، ۱۱ ویژگی برای استفاده در اعتبارسنجی انتخاب شدند. روش جنگل تصادفی (RF) بهترین عملکرد را در بین مدل‌های اعتبارسنجی داشته است. رویکرد شبیه‌سازی-بهینه‌یابی موفق شده با کاستن از اندازه تسهیلات و نرخ بهره و افزایش مدت تسهیلات، نرخ نکول را از ۳۸٪ به ۲۰٪ کاهش دهد. این نتایج نشان‌گر کارایی روش پیشنهادی این تحقیق در کاهش ریسک اعتباری بانک‌ها می‌باشد.

کلیدواژه‌ها: ریسک اعتباری، اعتبارسنجی، طبقه‌بندی، الگوریتم ممتیک، مدل شبیه‌سازی-بهینه‌یابی.

---

* نویسنده مسئول مقاله E-mail: Mdnayeri@modares.ac.ir

۱- مقدمه و بیان مسئله

یک سیستم بانکی کارآمد، نقشی حیاتی در رشد اقتصادی کشورها ایفا می‌کند و از طریق تسهیل جریان پس‌انداز و سرمایه‌گذاری، به تخصیص بهینه منابع و پویایی اقتصاد کمک می‌کند [۱]. در میان فعالیت‌های متنوع بانکداری، اعطای تسهیلات[۱] به عنوان یکی از محوری‌ترین آن‌ها، با وجود مزایای فراوان برای توسعه اقتصادی، همواره با ریسک اعتباری[۲] همراه است [۲]. ریسک اعتباری، احتمال عدم بازپرداخت اصل و فرع تسهیلات توسط وام‌گیرنده است و در صورت عدم مدیریت صحیح، می‌تواند منجر به مشکلات جدی در سیستم مالی، از جمله بحران‌های مالی و ورشکستگی بانک‌ها گردد [۳] [٤].

اگرچه حذف کامل ریسک اعتباری امری ناممکن است، مدیریت و کاهش آن از طریق سازوکارهای مؤثر، ضرورتی اجتناب‌ناپذیر برای بانک‌ها و مؤسسات مالی است. اعتبارسنجی[۳] که فرآیندی برای ارزیابی مشتریان و تعیین میزان اعتبار آن‌ها است، به عنوان یکی از کارآمدترین راهکارها در این زمینه شناخته می‌شود [٥] [٦]. هدف اصلی اعتبارسنجی، پیش‌بینی احتمال نکول[٤] تسهیلات اعطایی است که به دلیل ناتوانی یا عدم تمایل مشتریان در بازپرداخت اقساط تسهیلات بر اساس زمانبندی تعیین‌شده رخ می‌دهد [٤].

ایده اصلی اعتبارسنجی، مقایسه ویژگی‌ها و خصوصیات مشتری درخواست‌کننده تسهیلات با مشتریانی است که پیش‌تر تسهیلات دریافت کرده‌اند [۷].از منظر تاریخی، روش‌های سنجش اعتبار از رویکرد قضاوتی[٥] (ارزیابی کیفی) به مدل‌های کمّی اعتبارسنجی تکامل یافته‌اند [۸]. در رویکرد قضاوتی، کارشناس تحلیل‌گر اعتباری بر اساس معیارهایی چون ٥C درخواست‌های وام را بررسی کرده و در مورد آن‌ها تصمیم‌گیری می‌کند [۹]. این رویکرد با چالش‌هایی همچون تأثیر شدید نگرش‌های ذهنی تصمیم‌گیرندگان، احتمال بروز خطاهای سهوی و زمان‌بر بودن فرآیند ارزیابی مواجه بوده است. برای حل این چالش‌ها، مدل‌های اعتبارسنجی توسعه یافته‌اند [۸] [۱۰]. در مدل‌های اعتبارسنجی، فرآیند درخواست تسهیلات به‌صورت خودکار و

---

[۱] Facility
[۲] Credit risk
[۳] Credit scoring
[٤] Default
[٥] Judgemental

سریع انجام می‌شود. این مدل‌ها با بهره‌گیری از پایگاه داده مشتریان و بکارگیری الگوریتم‌های ریاضی، وضعیت اعتباری مشتریان را پیش‌بینی می‌کنند [٤] [۸].

از دیدگاه محاسباتی، مدل‌های اعتبارسنجی از روش‌های طبقه‌بندی[۱] برای پیش‌بینی وضعیت مشتریان (خوش‌حساب/ بدحساب) بهره می‌گیرند. در طول سه دهه‌ی گذشته، این مدل‌ها شاهد تحولات چشمگیری بوده‌اند. در مراحل اولیه، روش‌های ساده‌تری همچون تفکیک خطی[۲] (LDA) و رگرسیون لجستیک[۳] (LR) به عنوان رویکردهای غالب در حوزه‌ی اعتبارسنجی مطرح بودند [۱۰]. پیشرفت در فناوری‌های محاسباتی همراه با عواملی مانند توافقنامه بین‌المللی بازل[٤] و رخداد بحران مالی جهانی در دهه ۱۹۹۰، موجب افزایش توجه پژوهشگران به اعتبارسنجی گردید [۱۱]. در نتیجه روش‌های مدرن یادگیری ماشین[٥] همچون درخت تصمیم[٦] (DT)، شبکه‌های عصبی مصنوعی[۷] (ANN) و ماشین بردار پشتیبان[۸] (SVM) به تدریج در اعتبارسنجی مورد استفاده قرار گرفتند [۹] [۱۰]. در سالیان اخیر، الگوریتم‌های یادگیری جمعی[۹] نظیر جنگل تصادفی[۱۰] (RF)، تقویت حداکثر گرادیان[۱۱] (XGBoost) و بوستینگ تطبیقی[۱۲] (Adaboost) با اقبال زیادی در حوزه اعتبارسنجی مواجه شده‌اند [۱۰] [۱۱].

بررسی پیشینه تحقیق در بخش دوم انجام گرفته و خلاصه‌ای از تحقیقات اخیر در جداول ۱ تا ۳ ارائه شده است. بررسی تحقیقات پیشین نشان‌گر این است که پژوهش‌های قبلی هر یک ویژگی‌های خاصی را برای مدلسازی مشتریان و پیش‌بینی وضعیت اعتباری آن‌ها مورد استفاده قرار داده‌اند. همانطور که در جدول ۱ نمایش داده شده است، برای مشتریان حقیقی، مشخصات فردی (مانند سن، جنسیت و وضعیت تاهل) در کنار ویژگی‌های مالی (از جمله وضعیت مسکن، میانگین موجودی حساب و درآمد) مد نظر بوده‌اند. همچنین بر اساس جدول ۲ برای مشتریان حقوقی، مشخصات شرکت (چون سابقه، حوزه فعالیت و وضعیت تملیکی) و ویژگی‌های استخراج‌شده از صورت‌های مالی (به‌ویژه نسبت‌های مالی) مورد توجه بوده‌اند. علاوه بر این ویژگی‌ها، مشخصات مربوط به تسهیلات اعطایی نیز برای هر دو گروه

---

[۱] Classification methods
[۲] Linear discriminant analysis
[۳] Logistic regression
[٤] Basel Accords
[٥] Machine learning
[٦] Decision tree
[۷] Artificial neural network
[۸] Support vector machine
[۹] Ensemble learning
[۱۰] Random forest
[۱۱] Extreme gradient boosting
[۱۲] Adaptive Boosting

مشتریان در تحقیقات متعددی همچون [۱٤]، [۱۷]، [۱۸]، [۲۱]، [۲٤]، [۲٦]، [۲۷]، [۲۹]، و [۳۱] در نظر گرفته شده‌اند. مهمترین این مشخصات، مبلغ تسهیلات، نرخ سود و مدت زمان بازپرداخت اقساط تسهیلات می‌باشند.

اغلب ویژگی‌های به‌کاررفته در مدل‌های اعتبارسنجی جزو پارامترهای ورودی بوده و ثابت هستند. به طور نمونه، ویژگی‌های مشتری حقیقی (سن، جنسیت، شغل و میزان درآمد) یا مشتری حقوقی (حوزه فعالیت، وضعیت ملکی و صورت‌های مالی شرکت) همگی در مقطع زمانی درخواست تسهیلات، معیّن[۱] بوده و بانک قابلیت تغییر آن را ندارد. با این وجود، پارامترهای تسهیلات اعطایی (همچون مبلغ، سود و مدت زمان تسهیلات) که برای هر دو دسته مشتریان حقیقی و حقوقی کاربرد دارند، از منظر بانک‌ها متغیر[۲] بوده و قابل تغییر و تنظیم می‌باشند. بنابراین بانک‌ها می‌توانند این پارامترها را در محدوده مشخصی تنظیم کنند و از این طریق، بر احتمال نکول مشتریان تأثیرگذار باشند.

بررسی ادبیات تحقیق نشان‌گر این است که تأثیر تعدیل پارامترهای تسهیلات (مانند مبلغ، نرخ بهره و مدت بازپرداخت) بر احتمال نکول مشتریان در تحقیقات حوزه اعتبارسنجی مغفول مانده است. همچنین در مطالعات حوزه وام‌دهی بهینه[۳] که در آن‌ها مشخصات تسهیلات به عنوان متغیر در نظر گرفته می‌شوند، وضعیت اعتباری مشتریان همواره ثابت فرض می‌شود. به منظور پر کردن شکاف پژوهشی در این حوزه، پژوهش حاضر به بررسی تأثیر تنظیم پارامترهای تسهیلات بر احتمال نکول مشتریان می‌پردازد. هدف اصلی این پژوهش، ارائه یک رویکرد نوین اعتبارسنجی با در نظر گرفتن پارامترهای تسهیلات به عنوان متغیرهای قابل کنترل است تا از این طریق، احتمال نکول تسهیلات کاهش یافته و مدیریت ریسک اعتباری بانک‌ها ارتقا یابد.

در این راستا، پژوهش حاضر یک چارچوب ترکیبی شبیه‌سازی-بهینه‌سازی[٤] را پیشنهاد می‌دهد. در این چارچوب، ابتدا از روش‌های طبقه‌بندی برای ارزیابی اعتبار مشتریان استفاده می‌شود. سپس، با بهره‌گیری از شبیه‌سازی، تأثیر تغییر پارامترهای تسهیلات بر احتمال نکول مشتریان تحلیل شده و در نهایت، با استفاده از الگوریتم بهینه‌سازی ممتیک[۵]، مقادیر بهینه این

---

[۱] Given
[۲] Variable
[۳] Loan optimization / Optimal lending
[٤] Simulation-Optimization model
[۵] Memetic optimization algorithm

پارامترها تعیین می‌گردد. مطالعه موردی این پژوهش بر روی داده‌های واقعی مشتریان حقوقی یک بانک تجاری در ایران انجام شده است. مدل شبیه‌سازی-بهینه‌سازی ارائه شده، نخستین تلاش برای بهره‌گیری از قابلیت تنظیم پارامترهای تسهیلات در راستای کاهش ریسک اعتباری بانک‌ها است.

بخش‌های بعدی این مقاله شامل مروری بر پیشینه تحقیق، تشریح روش‌شناسی، تحلیل یافته‌ها و در نهایت، بحث و نتیجه‌گیری خواهد بود. این پژوهش امیدوار است تا با ارائه یک رویکرد نوین، گامی مؤثر در جهت بهبود فرآیند اعتبارسنجی و کاهش ریسک‌های مالی بانک‌ها بردارد.

۲- پیشینه تحقیق

در

جدول ۳ خلاصه‌ای از تحقیقات اخیر در زمینه اعتبارسنجی ارائه شده و ویژگی‌های کلیدی هر تحقیق بررسی شده است. از بین ۲۵ تحقیق بررسی شده، سه تحقیق مشتریان حقوقی را در نظر گرفته، یک تحقیق به هر دو دسته مشتریان حقیقی و حقوقی پرداخته و مابقی تحقیقات، بر مشتریان حقیقی تمرکز کرده‌اند. بنابراین مشاهده می‌شود که تحقیق پیشین، کمتر به اعتبارسنجی مشتریان حقوقی پرداخته‌اند.

مدل‌های مورد استفاده در اعتبارسنجی در این تحقیقات، محدوده وسیعی از مدل‌ها را در بر می‌گیرد. در اغلب این تحقیقات، دو یا گاه چندین مدل متفاوت مورد استفاده قرار گرفته و نتایج آنها با هم مقایسه شده‌اند. در بین مدل‌های مورد استفاده، برخی مدل‌ها همچون رگرسیون لجستیک (LR)، انواع کلاسیک درخت تصمیم (DT) و روش k-نزدیک‌ترین همسایه (KNN) نقش مدل پایه را داشته‌اند و نتایج مدل‌های پیچیده‌تر و جدیدتر (همچون شبکه‌های عصبی و SVM) با آنها مقایسه شده است. در بین مدل‌های مدرن‌تر، روش‌های جنگل تصادفی (RF)، بوستینگ تطبیقی (AdaBoost) و تقویت گرادیان شدید (XGBoost) پرکاربردتر از سایرین بوده‌اند.

در بین تحقیقات معرفی شده در

جدول ۳، تعداد ۷ تحقیق مطالعه موردی را بر روی مشتریان واقعی یک بانک انجام داده‌اند، درحالیکه ۱۸ تحقیق از دیتاست‌های آماده برای این منظور استفاده کرده‌اند. تعداد و نوع ویژگی‌های مورد استفاده در مدل‌های اعتبارسنجی در این تحقیقات، به اطلاعات در دسترس

وابسته بوده است. برای دیتاست‌های آماده، ویژگی‌ها از پیش مشخص هستند و محققین تنها می‌توانند با استفاده از روش‌های انتخاب ویژگی، زیرمجموعه‌ای از آنها را برای استفاده در مدل‌های اعتبارسنجی انتخاب کنند. در تحقیقاتی که از داده‌های بانک‌های واقعی استفاده می‌شود، ویژگی‌های مورد استفاده محدود به اطلاعاتی است که محقق گردآوری کرده است. به دلیل تفاوت ماهوی بین مشتریان حقیقی و حقوقی، اطلاعات با اهمیت این مشتریان به منظور استفاده در مدل‌های اعتبارسنجی نیز با هم متفاوت بوده است. مهمترین و پرکاربردترین ویژگی‌هایی که در تحقیقات فوق برای مشتریان حقیقی و حقوقی مورد استفاده قرار گرفته است، به ترتیب در جدول ۱ و
جدول ۲ ارائه شده‌اند.

جدول ۱: پارامترهای استفاده شده در تحقیقات پیشین (مشتریان حقیقی)

اطلاعات مربوط به متقاضی تسهیلات	اطلاعات مربوط به تسهیلات اعطایی
سن	مبلغ تسهیلات
جنسیت	مدت تسهیلات
تحصیلات	نرخ سود تسهیلات
شغل	تعداد اقساط تسهیلات
سابقه کار	تعداد تسهیلات دریافتی پیشین
درآمد	نوع/ ارزش وثیقه
میانگین موجودی حساب سه ماهه اخیر	
وضعیت تاهل	

جدول ۲: پارامترهای استفاده شده در تحقیقات پیشین (مشتریان حقوقی)

اطلاعات اساسی شرکت‌ها	اطلاعات صورت‌های مالی شرکت‌ها	اطلاعات مربوط به تسهیلات اعطایی
حوزه فعالیت شرکت	نسبت آنی	مبلغ تسهیلات
سال تاسیس شرکت	نسبت جاری	مدت تسهیلات
وضعیت مالکیت محل	نسبت بدهی	نرخ سود تسهیلات
سابقه همکاری با بانک	نسبت مالکانه	تعداد اقساط تسهیلات
سرمایه اولیه شرکت	نسبت حقوق صاحبان سهام	تعداد تسهیلات دریافتی پیشین
حسن شهرت	نسبت بدهی‌ها به دارایی‌ها	نوع/ ارزش وثیقه
معدل حساب	نسبت بدهی کل به دارایی کل	
	نسبت دارایی جاری به بدهی جاری	
	نسبت فروش به دارایی	
	نسبت سود خالص به فروش خالص	

	نسبت دارایی جاری به دارایی کل	
	سرمایه در گردش	

جدول ۳: روش‌های اعتبارسنجی مورد استفاده در مقالات سالیان اخیر

شماره تحقیق	سال انتشار	نوع مشتریان	مدل‌های اعتبارسنجی	ویژگی‌های مورد استفاده در مدل	داده‌های مورد استفاده در تحلیل
[۱۳]	۲۰۱۸	حقوقی	SVM Logistic Regression,	۱۰ ویژگی	۱۴۱۹۱ مشتری بانک‌های لهستانی
[۱٤]	۲۰۱۸	حقیقی	Logistic Decision Tree, Regression, Neural Network, Neural Network with Bagging	بین ۱۵ تا ۲۲ ویژگی	Bosnian dataset, German dataset, Australian dataset
[۱٥]	۲۰۱۹	حقیقی	Adaptive Network-Based Fuzzy Inference System (ANFIS)	۷ ویژگی	۹۰۰۰ مشتری یک بانک ایرانی
[۱٦]	۲۰۱۹	حقیقی	Logistic Regression, SVM, Random Forest, XGBoost, Neural Network, Deep Learning	۳٦۱ ویژگی	SCF dataset
[۱۷]	۲۰۲۰	حقیقی	Decision Tree, Naïve Bayes, Generalized Linear Model, SVM	۱٤ ویژگی	٦۲٥٦ مشتری یک بانک
[۱۸]	۲۰۲۰	حقیقی و حقوقی	Random Forest, KNN, Naïve Bayes, Logistic Regression, SVM, Neural Network, Bagging, Adaptive Boosting	بین ۷ تا ٦٥ ویژگی	German dataset, Australian dataset, Japanese dataset, Taiwan dataset, Polish dataset, Indian dataset
[۱۹]	۲۰۲۰	حقیقی	SVM, KNN, Fuzzy System, DGHNL	۲۰ ویژگی	German dataset
[۲۰]	۲۰۲۱	حقیقی	Decision Tree, Random Forest, XGBoost, Adaboost, KNN, Neural Network	۷ ویژگی	٤٤٥۰ مشتری یک بانک در غنا
[۲۱]	۲۰۲۱	حقیقی	XGBoost, Random Forest, Logistic Regression	۳٤ ویژگی	٥ میلیون مشتری بانک‌های لهستانی
[۲۲]	۲۰۲۱	حقیقی	Logistic Regression, SVM, KNN, Random Forest	۲٤ ویژگی	Taiwan dataset
[۲۳]	۲۰۲۱	حقیقی	Naïve Bayes, Decision Tree, Random Forest, Neural Network, SVM, KNN	بین ٦ تا ۲٤ ویژگی	Australian dataset, Japanese dataset, German dataset, Bankruptcy dataset, Taiwan dataset
[۲٤]	۲۰۲۲	حقیقی	Random Forest, Decision Tree, Naïve Bayes, Neural Networks, Logistic Regression KNN,	۲۰ ویژگی	German dataset
[۲٥]	۲۰۲۲	حقیقی	Logistic Regression, KNN, NaiveBayes, SVM, XGboost, Random Forest	بین ۱۵ تا ۲۱ ویژگی	German dataset, Australian dataset, Japanese dataset
[۲٦]	۲۰۲۲	حقیقی	Logistic regression, SVM, Neural Networks	۱٥ ویژگی	۳۱٥۳۲ مشتری یک بانک در

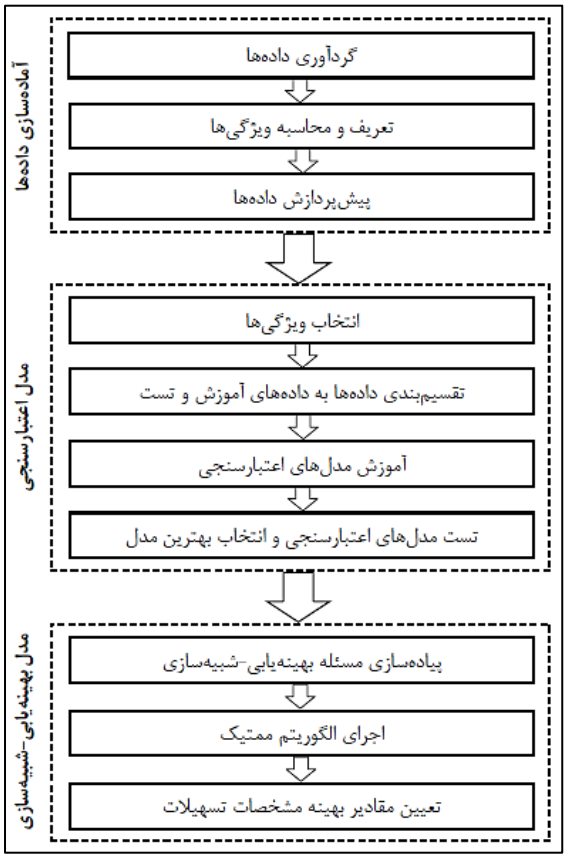
شماره تحقیق	سال انتشار	نوع مشتریان	مدل‌های اعتبارسنجی	ویژگی‌های مورد استفاده در مدل	داده‌های مورد استفاده در تحلیل
					مالزی
[۲۷]	۲۰۲۲	حقوقی	SVM based on fuzzy integral	۲۶ ویژگی	۱٤٦ مشتری یک بانک چینی
[۲۸]	۲۰۲۳	حقوقی	AdaBoost, Logistic Regression, Random Forest, XGBoost	۱۰ ویژگی	Tinubu Square dataset
[۲۹]	۲۰۲۳	حقیقی	Bagging Supervised Autoencoder Classifier (BSAC)		Taiwan dataset, Lending Club dataset
[۳۰]	۲۰۲۳	حقیقی	Logistic Regression, Decision Tree, KNN, Random Forest, SVM, Adaboost, XGboost	بین ۱۰ تا ۳۰ ویژگی	Australian dataset, German dataset, Chinese personal loan dataset, Give me some credit dataset, Credit card fraud detection dataset
[۳۱]	۲۰۲۳	حقیقی	AdaBoost, Decision Tree, KNN, Logistic Regression, Random Forest	بین ۱٤ تا ۲۱ ویژگی	Australian dataset, German dataset
[۳۲]	۲۰۲٤	حقیقی	Model-Agnostic Explanations (LIME), SHapley Additive exPlanations (SHAP)	۱٦ ویژگی	European Datawarehouse dataset
[۳۳]	۲۰۲٤	حقیقی	XGBoost	۷۸ ویژگی	Japanese dataset, Australian dataset, German dataset, Give me some credit dataset
[۳٤]	۲۰۲٤	حقیقی	SVM, Decision Tree, Random Forest, XGBoost	۲۰ ویژگی	German dataset
[۳۵]	۲۰۲٤	حقیقی	Deep Forest (DF)	۱۱۰ ویژگی	Lending Club dataset
[۳٦]	۲۰۲٤	حقیقی	XGBoost, Neural Network	۲۳ ویژگی	Credit Approval dataset
[۳۷]	۲۰۲٤	حقیقی	Deep Learning	۲۵ ویژگی	Default of Credit Card Clients dataset

۳- روش‌شناسی پژوهش

روش‌شناسی پیشنهادی این پژوهش برای طراحی مدل ترکیبی شبیه‌سازی و بهینه‌سازی اعتبارسنجی مبتنی بر الگوریتم ممتیک در شکل ۱ ترسیم شده است. همانگونه که در این فلوچارت مشهود است، چارچوب روش‌شناختی از سه مرحله کلیدی آماده‌سازی داده‌ها[۱]، مدل اعتبارسنجی[۲] و مدل شبیه‌سازی- بهینه‌یابی[۳] تشکیل شده است. در ادامه این بخش، هر یک از این فرآیندها معرفی و تشریح می‌شوند.

---

[۱] Data preparation
[۲] Credit scoring model
[۳] Simulation-Optimization model



شکل ۱: فلوچارت چارچوب روش پیشنهادی این مقاله

۳-۱- آماده‌سازی داده‌ها

داده‌های مورد استفاده در این پژوهش، شامل اطلاعات ۱۰۰۰ شرکت حقوقی است که در بازه زمانی ۱۳۹٦ تا ۱٤۰۰ از یک بانک تجاری ایرانی تسهیلات دریافت کرده‌اند. این داده‌ها به صورت کاملاً ناشناس و با همکاری بانک مذکور برای اهداف تحقیقاتی در اختیار پژوهشگران قرار گرفته است. برای هر شرکت، علاوه بر مشخصات تسهیلات اعطاشده نظیر نوع تسهیلات، مبلغ اصلی تسهیلات ، نرخ سود و دوره بازپرداخت، گزارش‌های مالی شامل ترازنامه و صورت سود و زیان این شرکت‌ها در سال ۱٤۰۰ نیز جمع‌آوری شده است. این ترکیب داده‌ای، امکان

تحلیل جامعی از رابطه بین پارامترهای تسهیلات و رفتار بازپرداخت مشتریان حقوقی را فراهم می‌کند.

۳-۱-۱- تعریف و محاسبه ویژگی‌ها

بر مبنای داده‌های جمع‌آوری‌شده، ویژگی‌های تحلیلی برای کاربرد در مدل‌های ارزیابی اعتبار تعریف و محاسبه گردید .برچسب‌دهی داده‌ها (وضعیت نکول) برای هر شرکت به صورت باینری (۰ برای عدم نکول و ۱ برای نکول) انجام شد. بر اساس معیار پذیرفته‌شده در مطالعات پیشین، تسهیلاتی که بیش از ۹۰ روز تأخیر در بازپرداخت اقساط داشته‌اند، به عنوان نکول شده طبقه‌بندی و برچسب ۱ دریافت کردند [۲۳]. بررسی داده‌ها نشان می‌دهد که از بین ۱۰۰۰ شرکت، ٤١٥ شرکت (٤١/٥٪) دچار نکول شده‌اند. مطابق با تحقیقات پیشین، لگاریتم مبلغ اصل قرارداد (لگاریتم در پایه ۱۰) محاسبه شده و به جای مبلغ اصل قرارداد از آن استفاده شد. علاوه بر این، سود سالیانه تسهیلات (برحسب درصد) و مدت زمان بازپرداخت تسهیلات (برحسب سال) برای استفاده در مدل‌های اعتبارسنجی در نظر گرفته شدند. بنابراین در مجموع سه ویژگی از بین اطلاعات تسهیلات بانکی در نظر گرفته می‌شود که عبارتند از:

- لگاریتم مبلغ اصل قرارداد (اندازه وام)
- سود سالیانه تسهیلات (نرخ بهره)
- مدت زمان بازپرداخت تسهیلات

به منظور ایجاد امکان مقایسه وضعیت شرکت‌ها با یکدیگر، به طور معمول از نسبت‌های مالی[۱] [۳۸] استفاده می‌شود. بر اساس اطلاعات گردآوری شده صورت‌های مالی، تعداد ۲۷ نسبت مالی محاسبه گردید که شامل نسبت‌های نقدینگی[۲]، نسبت‌های فعالیت[۳] (کارایی)، نسبت‌های اهرمی[٤] (بدهی) و نسبت‌های سودآوری[٥] می‌باشند. این نسبت‌های مالی عبارتند از [۸]، [۹]، [۱۰]، [۱۷]، [۱۸]، [۲۲]، [۲٤]، [۲٥]:

- نسبت‌های نقدینگی: نسبت جاری، نسبت سریع (آنی)، نسبت نقدینگی (نسبت وجه نقد)، نسبت دارایی جاری

---

[۱] Financial ratios
[۲] Liquidity ratios
[۳] Activity/Efficiency ratios
[٤] Leveraging/Debt ratios
[٥] Profitability ratios

- نسبت‌های فعالیت: دوره گردش موجودی کالا، نسبت کالا به سرمایه در گردش، گردش سرمایه جاری، گردش سرمایه ثابت، نسبت گردش دارایی
- نسبت‌های اهرمی: نسبت بدهی، نسبت سرمایه، نسبت بدهی-سرمایه، نسبت کل بدهی به ارزش ویژه (حقوق صاحبان سهام)، نسبت بدهی جاری به ارزش ویژه، نسبت بدهی بلندمدت (غیرجاری) به ارزش ویژه، نسبت دارایی ثابت به ارزش ویژه، نسبت مالکانه، نسبت حقوق صاحبان سهام به کل بدهی‌ها، نسبت حقوق صاحبان سهام به دارایی‌های ثابت، نسبت پوشش بدهی
- نسبت‌های سودآوری: حاشیه (نسبت) سود خالص، حاشیه‌سود عملیاتی، حاشیه سود ناخالص، بازده حقوق صاحبان سهام، بازده دارایی‌ها، بازده سرمایه در گردش، بازده دارائی ثابت

۳-۱-۲- پیش‌پردازش داده‌ها

پیش از بکارگیری ویژگی‌ها در مدل‌های اعتبارسنجی، به منظور اطمینان از صحت و کیفیت داده‌ها، پیش‌پردازش داده‌ها[۱] به صورت زیر انجام گرفت:

- جایگزینی داده‌های ناقص[۲]: برای جایگزینی داده‌های ناقص، مشابه تحقیقاتی چون [۱۹] و [۳۹]، از روش انتساب میانگین[۳] استفاده گردید.
- شناسایی و ترمیم داده‌های پرت[٤]: به منظور شناسایی و ترمیم داده‌های پرت، مشابه تحقیقاتی چون [٤۰] و [٤۱]، از روش فاصله بین چارکی[۵] استفاده شد.
- نرمال‌سازی داده‌ها[٦]: برای رفع اثرات منفی ناشی از تفاوت محدوده تغییرات ویژگی‌های مختلف، از روش نرمال‌سازی حداقل-حداکثر[۷] برای نگاشت[۸] داده‌ها به محدوده صفر تا یک استفاده گردید [۳۹] و [٤۲].

۳-۲- مدل‌های اعتبارسنجی

۳-۲-۱- انتخاب ویژگی‌ها

---

[۱] Data pre-processing
[۲] Missing data
[۳] Mean imputation
[٤] Outlier detection and treatment
[۵] Interquartile range
[٦] Data normalization
[۷] Minimum–maximum normalization method
[۸] Mapping

بر اساس توضیحات ارائه شده، اطلاعات ۳۰ ویژگی (۳ ویژگی تسهیلات بانکی و ۲۷ نسبت مالی) برای استفاده در مدل‌های اعتبارسنجی آماده شد. این ویژگی‌ها به عنوان متغیرهای مستقل[۱] مسئله در نظر گرفته شدند. برچسب داده‌ها (وضعیت نکول)، متغیر وابسته[۲] بوده و مقدار آن توسط روش‌های طبقه‌بندی، پیش‌بینی خواهد شد. زیاد بودن تعداد ویژگی‌ها گاه موجب کاهش کارایی و افزایش زمان محاسبات می‌گردد. بنابراین نیاز است که توسط روش‌های انتخاب ویژگی، زیرمجموعه‌ای از ویژگی‌ها که توانایی بالاتری در تخمین احتمال نکول دارند، انتخاب شوند. برای انتخاب ویژگی‌ها از بین ۲۷ نسبت مالی، از الگوریتم حداقل افزونگی حداکثر ارتباط[۳] (MRMR) [٤٣] استفاده شد. در روش MRMR، برای هر یک از ویژگی‌ها، یک امتیاز محاسبه می‌شود که نشان‌گر میزان اهمیت ویژگی مورد نظر است. برای کاستن از تاثیر منفی همبستگی ویژگی‌ها با یکدیگر، امتیازدهی ویژگی‌ها بگونه‌ای صورت می‌گیرد که کمترین افزونگی[٤] بین ویژگی‌ها وجود داشته باشد. ویژگی‌هایی که دارای امتیازی بالاتر از ۰.۰۱ هستند برای استفاده در مدل‌های اعتبارسنجی انتخاب شدند.

۳-۲-۲- روش‌های طبقه‌بندی

در این پژوهش، از پنج الگوریتم طبقه‌بندی جهت اعتبارسنجی مشتریان و تفکیک آنها به دو گروه خوش‌حساب و بد حساب استفاده گردید. در ادامه، مروری مختصر بر این روش‌ها ارائه می‌شود:

رگرسیون لجیتسک[۵] (LR): این روش به عنوان یک مدل خطی تعمیم‌یافته[٦] (GLM) شناخته می‌شود که برای پیش‌بینی متغیرهای هدف باینری (مانند نکول/عدم نکول) طراحی شده است. در رگرسیون لجستیک، از تابع لاجیت[۷] برای بیان رابطه بین متغیرهای مستقل (ویژگی‌ها) و متغیر وابسته (برچسب داده‌ها) استفاده می‌شود. احتمال آنکه یک داده در کلاس یک (نکول) قرار گیرد، توسط تابع لاجیت به صورت عددی بین صفر و یک محاسبه می‌شود. برای دسته‌بندی مشتریان در دو دسته خوش‌حساب و بدحساب، از یک حد آستانه[۸] استفاده می‌شود

[۱] Independent variables
[۲] Dependent variable
[۳] Minimum redundancy maximum relevance (MRMR) algorithm
[٤] Redundancy
[۵] Logistic regression
[٦] Generalized linear model
[۷] Logit function
[۸] Threshold

که به طور پیش‌فرض، برابر با ۰/۵ است. برای آموزش مدل رگرسیونی به منظور یافتن ضرایب ویژگی‌ها (مقادیر β)، از روش حداکثر احتمال[۱] استفاده می‌شود [۱۷] و [٤٥].

ماشین بردار پشتیبان[۲] (SVM): الگوریتم SVM یک روش قدرتمند در یادگیری ماشین برای طبقه‌بندی داده‌ها است. هدف اصلی این الگوریتم، یافتن یک ابرصفحه بهینه است که بیشترین فاصله (حاشیه) را بین دو یا چند کلاس داده فراهم کند. مشخصات این ابرصفحه با استفاده از بردارهای پشتیبان، یعنی نزدیک‌ترین نقاط داده به ابرصفحه، تعیین می‌شود. در صورتی که داده‌ها به صورت خطی قابل جداسازی نباشند، SVM از توابع هسته برای تبدیل داده‌ها به فضاهای ابعاد بالاتر استفاده می‌کند تا بدین طریق، جداسازی خطی امکان‌پذیر شود. این روش به دلیل عملکرد خوب در فضاهای با ابعاد بالا و توانایی حل مسائل غیرخطی، در زمینه‌های مختلفی مانند تشخیص تصویر، طبقه‌بندی متن و تشخیص پزشکی و نیز مسئل اعتبارسنجی کاربرد دارد [۱۷] و [۱۸].

شبکه عصبی مصنوعی[۳] (ANN): این روش که الهام گرفته از ساختار عصبی مغز انسان است، از قابلیت انطباق‌پذیری قابل توجهی در حل مسائل پیچیده برخوردار است. ساختار پایه‌ای آن شامل سه لایه اصلی است: لایه ورودی (دریافت ویژگی‌های داده)، لایه پنهان (انجام محاسبات غیرخطی برای استخراج الگوها) و لایه خروجی (تولید پیش‌بینی نهایی). با استفاده از فرایند پس انتشار خطا، وزن اتصالات بین نورون‌ها به گونه‌ای تنظیم می‌شود که اختلاف بین پیش‌بینی‌ها و مقادیر واقعی کمینه گردد. شبکه‌های عمیق با لایه‌های پنهان متعدد می‌توانند روابط پیچیده‌تری را مدل‌سازی کنند [۹] و [۱۸].

تقویت گرادیان شدید[٤] (XGB): الگوریتم XGBoost که یکی از انواع روش‌های یادگیری جمعی است، یک روش کارآمد در یادگیری ماشین است که بر پایه اصل تقویت گرادیان عمل می‌کند. این الگوریتم با ساخت درخت‌های تصمیم‌گیری متوالی، به بهبود عملکرد مدل می‌پردازد. هر درخت جدید تلاش می‌کند خطاهای درخت‌های قبلی را اصلاح کند. XGBoost با استفاده از تکنیک‌های پیشرفته‌ای مانند تنظیم‌سازی[٥] و موازی‌سازی[٦]، سرعت و دقت بالایی را در

---

[۱] Maximum likelihood method
[۲] Support vector machine
[۳] Artificial neural network
[٤] Extreme gradient boosting
[٥] Regularization
[٦] Parallelization

پردازش داده‌های حجیم و پیچیده فراهم می‌کند. این الگوریتم عملکرد خوبی در طبقه‌بندی انواع الگوهای داده‌ها دارد و از عوامل منفی همچون داده‌های ناقص، داده‌های پرت و الگوهای غیرخطی، تاثیرپذیری اندکی دارد [۹] و [۱۵].

جنگل تصادفی[۱] (**RF**): روش جنگل تصادفی که از روش‌های یادگیری جمعی است، شامل مجموعه‌ای از درخت‌های تصمیم است که با یکدیگر تجمیع شده و پاسخ نهایی را تشکیل می‌دهند. به هر یک از درخت‌ها، یک زیرمجموعه تصادفی از مجموعه ویژگی‌ها تخصیص داده شده و آموزش درخت‌ها توسط داده‌های متفاوتی صورت می‌گیرد. برای تخصیص تصادفی داده‌ها به درخت‌ها، از روش‌های بازنمونه‌گیری[۲] همچون بوت‌استرپینگ[۳] استفاده می‌شود. با توجه به این تفاوت‌ها، هر درخت به شیوه متفاوتی آموزش دیده و پاسخ‌های متفاوتی تولید می‌کند. پاسخ‌های ایجاد شده توسط درخت‌ها توسط روش‌هایی چون رای‌گیری اکثریت با یکدیگر تجمیع می‌شوند [۱۷] و [۱۸].

۳-۲-۳- معیارهای ارزیابی

به منظور ارزیابی عملکرد مدل‌های اعتبارسنجی، مجموعه داده‌ها به صورت تصادفی، به دو بخش داده‌های آموزش (۷۰٪ داده‌ها) و داده‌های تست (۳۰٪ داده‌ها) تقسیم‌بندی می‌شوند. پس از آموزش مدل‌ها توسط داده‌های آموزش، عملکرد آنها بر روی داده‌های تست ارزیابی می‌گردد. ارزیابی نتایج هر یک از مدل‌ها با محاسبه ماتریس درهم‌ریختگی[۴] آغاز می‌شود. در این ماتریس، هر چهار حالت پیش‌بینی برچسب (کلاس) داده‌ها در نظر گرفته می‌شود: مثبت صادق[۵] (TP)، مثبت کاذب[۶] (FP)، منفی صادق[۷] (TN) و منفی کاذب[۸] (FN). بر اساس مقادیر ماتریس درهم‌ریختگی، شاخص‌های زیر محاسبه شده و در ارزیابی مدل‌های اعتبارسنجی استفاده می‌شود [۱۰] و [۱۸]:

---

۱ Random forest
۲ Resampling
۳ Bootstrapping
۴ Confusion matrix
۵ True Positive
۶ False Positive
۷ True Negative
۸ False Negative

- شاخص صحت[۱]: بیان‌گر این است که از بین داده‌هایی که مدل به عنوان مثبت (نکول) شناسایی کرده، چه بخشی واقعا دارای برچسب مثبت بوده‌اند (نکول داشته‌اند). محاسبه شاخص صحت بر اساس رابطه زیر صورت می‌گیرد:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \qquad (۱)$$

- شاخص بازیابی[۲]: بیان‌گر این است که از بین داده‌هایی که واقعا دارای برچسب مثبت بوده‌اند (نکول داشته‌اند)، چه بخشی را مدل توانسته به عنوان مثبت (نکول) شناسایی کند. محاسبه شاخص بازیابی بر اساس رابطه زیر صورت می‌گیرد:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \qquad (۲)$$

- نمره F۱ [۳]: ترکیبی از شاخص‌های صحت و بازیابی بوده و به طور کلی بیان‌گر توانایی کلی مدل در تفکیک مشتریان دارای نکول از مشتریان بدون نکول می‌باشد. محاسبه این شاخص بر اساس رابطه زیر صورت می‌گیرد:

$$\text{F1\_Score} = \frac{۲ \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \qquad (۳)$$

- شاخص دقت[٤]: بیان‌گر این است که مدل تا چه حد توانایی این را داشته که به طور همزمان، داده‌های با برچسب مثبت را، مثبت و داده‌های با برچسب منفی را، منفی شناسایی کند. محاسبه شاخص دقت بر اساس رابطه زیر صورت می‌گیرد:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{P} + \text{N}} \qquad (٤)$$

- سطح زیر منحنی[٥] (AUC): بیان‌گر عملکرد مدل‌های طبقه‌بندی بر اساس منحنی مشخصه عملکرد سیستم[٦] بوده و نشان می‌دهد که در شناسایی مشتریان دارای نکول، تا چه میزان

---

۱ Precision
۲ Recall
۳ F۱-Score
٤ Accuracy
٥ Area under curve
٦ Receiver operating characteristic (ROC) curve

نرخ مثبت صادق[۱] (TPR) بالاتر از نرخ مثبت کاذب[۲] (FPR) بوده است. محاسبه AUC بر اساس سطح زیر منحنی (انتگرال) نمودار صحت (Precision) بر حسب بازیابی (Recall) به صورت زیر صورت می‌پذیرد:

$$\mathrm{AUC} = \int_0^1 \mathrm{Precision}(t).\, d(\mathrm{Recall}(t)) \qquad (۵)$$

۳-۳- مدل شبیه‌سازی-بهینه‌یابی

پس از ارزیابی عملکرد مدل‌های اعتبارسنجی بر پایه شاخص‌های استاندارد، بهترین مدل انتخاب شده و در چارچوب ترکیبی شبیه‌سازی و بهینه‌سازی به کار گرفته می‌شود. فرآیند بهینه‌سازی با استفاده از الگوریتم ممتیک انجام می‌پذیرد. در ادامه، مراحل کلیدی این رویکرد به اختصار تشریح می‌شود. رویکرد شبیه‌سازی-بهینه‌یابی، با تنظیم بهینه مشخصات تسهیلات (شامل مبلغ، نرخ سود و دوره بازپرداخت)، به دنبال کاهش نرخ نکول و در نتیجه کاهش ریسک اعتباری بانک‌ها و موسسات مالی و اعتباری است. اجزای کلیدی این رویکرد به شرح زیر است:

بهینه‌یابی پارامترهای تسهیلات:

- متغیرهای تصمیم: سه پارامتر اصلی تسهیلات شامل مبلغ تسهیلات، نرخ سود و مدت بازپرداخت به عنوان عوامل قابل کنترل در نظر گرفته می‌شوند.
- تابع هدف: کمینه‌سازی تعداد موارد نکول با یافتن ترکیب بهینه پارامترهایی که احتمال بازپرداخت موفق را حداکثر می‌کند.
- الگوریتم ممتیک: این الگوریتم با ترکیب جستجوی سراسری (الگوریتم ژنتیک) و بهینه‌سازی محلی، با دقت و سرعت بالا به پاسخ‌های بهینه دست می‌یابد.

شبیه‌سازی:

- چالش واقع‌گرایی: در شرایط عملیاتی، آزمون مستقیم تأثیر تغییر پارامترهای تسهیلات بر رفتار مشتریان به دلیل محدودیت زمانی و ریسک مالی غیرممکن است.
- جایگزین تحلیلی: با استفاده از مدل اعتبارسنجی آموزش‌دیده، اثرات تغییر پارامترها بر احتمال نکول به صورت مجازی شبیه‌سازی می‌شود. به این ترتیب، برای هر تنظیم جدید

---

[۱] True positive rate

[۲] False positive rate

پارامترها، ویژگی‌های به‌روزرسانی‌شده مشتری به مدل وارد شده و پیش‌بینی تجدیدنظر شده از ریسک اعتباری مشتریان به دست می‌آید.

دو جزء فوق (بهینه‌یابی و شبیه‌سازی)، به صورت حلقه بازخوردی با یکدیگر تعامل دارند:

۱. الگوریتم ممتیک مقادیر جدیدی برای پارامترهای تسهیلات پیشنهاد می‌دهد.
۲. شبیه‌ساز با استفاده از مدل اعتبارسنجی، ریسک نکول متناظر با این مقادیر را محاسبه می‌کند.
۳. تابع برازش، کیفیت راهکار پیشنهادی را ارزیابی می‌کند.
٤. فرایند فوق تا دستیابی به پارامترهای بهینه، در یک حلقه تکرار می‌شود.

متغیرهای مسئله بهینه‌سازی به صورت زیر تعریف می‌شوند:

- متغیر ۱: اندازه تسهیلات[۱] (اصل مبلغ تسهیلات) که متغیری پیوسته است: همانند مدل اعتبارسنجی، بازنمایی این متغیر به صورت لگاریتمی انجام می‌شود. بر اساس اطلاعات گردآوری شده شرکت‌ها، مقدار حداقل و حداکثر اندازه تسهیلات، به ترتیب برابر با ۹ و ۱۳ در نظر گرفته می‌شود.
- متغیر ۲: نرخ بهره[۲] (نرخ سود تسهیلات) که متغیری گسسته است: بر اساس دستورالعمل‌های بانک مورد استفاده، نرخ بهره تسهیلات بسته به نوع تسهیلات، یکی از مقادیر ٤٪ یا ۱۸٪ را اختیار می‌کند.
- متغیر ۳: مدت وام[۳] (زمان بازپرداخت تسهیلات) که متغیری گسسته است: بر اساس دستورالعمل‌های بانک مورد استفاده، بر اساس نوع تسهیلات، مدت تسهیلات یکی از مقادیر ۱، ۲، ۳، ۵ یا ۱۰ سال را اختیار می‌کند.

به منظور بهینه‌یابی مشخصات تسهیلات اعطایی به مشتریان بانک، از الگوریتم ممتیک[٤] (MA) استفاده می‌شود. الگوریتم ممتیک به طور کلی، نسخه‌ای بهبودیافته از الگورتیم ژنتیک[٥] (GA) [٤٤] محسوب می‌شود [٤٥]. ایده الگوریتم ممتیک، از مفهوم میم‌ها[٦] سرچشمه گرفته که برخلاف ژن‌ها[۷]، قادر به تطبیق خود بر اساس وضعیت محیط می‌باشند. نخستین بار موسکاتو[۸]

---

[۱] Facility size
[۲] Interest rate
[۳] Facility age
[٤] Memetic algorithm
[٥] Genetic algorithm
[٦] Meme
[۷] Genes
[۸] Pablo Moscato

[٤٦] از ایده میم استفاده کرده و الگوریتم ممتیک را به عنوان الگوریتمی مشابه با ژنتیک، اما با امکان یادگیری فردی به منظور دست‌یابی به بهبودهای محلی، معرفی نمود. در الگوریتم ممتیک از عملگر جست‌وجوی محلی[١] در کنار عملگرهای متداول ژنتیک استفاده می‌شود و همین موجب افزایش کارایی این الگوریتم و کاهش احتمال همگرایی زودرس[٢] (پیش از موعد) می‌گردد [٤٧] و [٤٨].

مدل شبیه‌سازی-بهینه‌یابی تنها بر روی داده‌های تست (شامل ٣٠٠ مشتری)، پیاده‌سازی می‌شود تا امکان ارزیابی منصفانه نتایج آن فراهم گردد. بازنمایی متغیرهای مسئله توسط الگوریتم ممتیک، به صورت زیر انجام می‌گیرد. هر کروموزوم از ٣٠٠ بخش تشکیل شده که هر بخش، متناظر با یکی از مشتریان است. اندیس i، شمارش‌گر مشتری‌ها بوده و n تعداد کل مشتری‌ها است. در هر بخش، سه ژن وجود دارد که معادل با سه متغیر برای آن مشتری است. بنابراین کروموزوم در مجموع دارای ٩٠٠ ژن می‌باشد:

$$[LS_1, IR_1, LA_1, \dots, LS_i, IR_i, LA_i, \dots, LS_n, IR_n, LA_n] \qquad (٦)$$

ژن LS_i بیان‌گر اندازه تسهیلات مشتری شماره i بوده و IR_i و LA_i به ترتیب نرخ بهره و مدت تسهیلات را برای مشتری شماره i نشان می‌دهند. به طور نمونه بخش [١٠,٤,٥] برای یک مشتری نشان‌گر یک تسهیلات با اندازه لگاریتمی ١٠ (١٠ میلیارد ریال)، نرخ بهره ٤٪ و مدت زمان بازپرداخت ٥ سال می‌باشد.

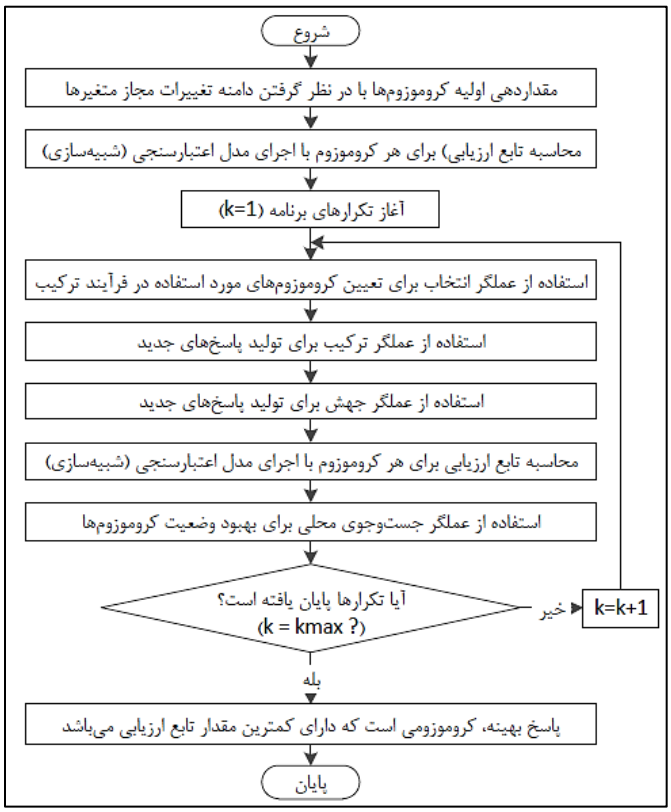
در الگوریتم ممتیک، عملگرهای ترکیب و جهش[٣] به شیوه تک‌نقطه‌ای[٤] پیاده‌سازی می‌شوند. برای انتخاب کروموزوم‌ها، از روش چرخ رولت[٥] استفاده می‌گردد که به کروموزوم‌های با برازش بالاتر شانس بیشتری می‌دهد. عملگر جست‌وجوی محلی[٦] مشابه با عملگر جهش کار می‌کند، اما پس از تولید پاسخ جدید، تابع هدف مسئله ارزیابی شده و در صورت عدم بهبود، فرآیند تا دستیابی به پاسخ بهتر یا رسیدن به حداکثر تکرار مجاز ادامه می‌یابد. تابع هدف مسئله با فراخوانی بهترین مدل اعتبارسنجی (از بین پنج مدل آموزش‌دیده) و شبیه‌سازی تأثیر

---

[١] Local search operators
[٢] Premature convergence
[٣] Mutation operator
[٤] Single-point
[٥] Roulette wheel selection
[٦] Local search operator

پارامترهای وام بر احتمال نکول مشتریان و در نهایت محاسبه تعداد تسهیلاتی که نکول شده‌اند، ارزیابی می‌شود.

از آنجا که پنج مدل اعتبارسنجی در این پژوهش استفاده شده، مدلی با بالاترین دقت به عنوان پایه شبیه‌سازی انتخاب می‌گردد. فرآیند بهینه‌سازی با تولید جمعیت اولیه از کروموزوم‌ها آغاز شده و در هر تکرار، با اعمال عملگرهای ترکیب، جهش و جستجوی محلی، نسل جدیدی از کروموزوم‌ها ایجاد می‌شود. هر کروموزوم با استفاده از مدل اعتبارسنجی برتر ارزیابی شده و تا رسیدن به معیارهای توقف ادامه می‌یابد. برای درک بهتر تعامل بین این مراحل، در شکل ۲ فلوچارت کامل فرآیند شبیه‌سازی-بهینه‌یابی ارائه شده است. این نمودار شامل ورودی‌ها (داده‌های مشتریان و محدودیت‌ها) و خروجی‌ها (پارامترهای بهینه وام) است.



شکل ۲: مراحل اجرای رویکرد بهینه‌یابی-شبیه‌سازی

به منظور بررسی تاثیر هر یک از سه مشخصه تسهیلات بر وضعیت نکول مشتریان و ریسک اعتباری بانک، رویکرد شبیه‌سازی-بهینه‌یابی در چهار حالت مختلف اجرا شده و نتایج آن با یکدیگر مقایسه می‌گردد:

- حالت ۱: مدت تسهیلات و نرخ بهره، ثابت بوده و اندازه تسهیلات، متغیر می‌باشد.
- حالت ۲: اندازه تسهیلات و مدت تسهیلات، ثابت بوده و نرخ بهره، متغیر می‌باشد.
- حالت ۳: اندازه تسهیلات و نرخ بهره، ثابت بوده و مدت تسهیلات، متغیر می‌باشد.
- حالت ٤: اندازه، نرخ بهره و مدت تسهیلات، هر سه متغیرهای مسئله می‌باشند.

۴- نتایج تحقیق

مطالعه موردی بر روی اطلاعات گردآوری شده ۱۰۰۰ مشتری حقوقی یک بانک تجاری ایرانی انجام گرفته است. برای این منظور، اطلاعات تسهیلات بانکی و اطلاعات صورت‌های مالی گردآوری و تحلیل شده است. پس از تعریف و محاسبه ویژگی‌ها و پیش‌پردازش داده‌ها، در نهایت تعداد ۳۰ ویژگی برای استفاده در مدل‌های اعتبارسنجی در نظر گرفته شد.

به منظور اعتبارسنجی مشتریان، از پنج مدل مختلف طبقه‌بندی استفاده شد. پیاده‌سازی این مدل‌ها با استفاده از جعبه‌ابزار یادگیری آماری و ماشینی[۱] در نرم‌افزار متلب[۲] صورت گرفت. در روش SVM، تابع هسته از نوع شعاعی پایه (RBF) با گامای ۰.۱ و پارامتر تنظیم‌سازی C برابر با ۱۰ در نظر گرفته شد. در روش ANN، از یک شبکه چند لایه پرسپترون با سه لایه پنهان که به ترتیب دارای ۳۰، ۱۵ و ۱۰ نورون می‌باشند، استفاده شد. آموزش شبکه توسط الگوریتم بازانتشار خطای LM[۳] انجام گرفت. در روش XGB، تعداد درخت‌های تصمیم‌گیری برابر با ۱۰۰، حداکثر عمق هر درخت برابر با ۳ و نرخ یادگیری برابر با ۰.۱ در نظر گرفته شد. در روش RF، حداکثر تعداد تقسیم تصمیم برابر با ۱۰، حداقل مشاهدات در هر برگ برابر با ۷ و تعداد دوره‌های آموزش برابر با ۳۰ قرار داده شد. در الگوریتم ممتیک، تعداد کروموزوم‌ها برابر با ۲۰، حداکثر تعداد تکرارهای الگوریتم برابر با ۱۵۰، درصد عملیات ترکیب برابر با ۸۰٪، درصد عملیات جهش، برابر با ۲۰٪ و حداکثر تعداد تلاش‌ها در جست‌وجوی محلی برابر با ۳۰ در نظر گرفته شد.

با اجرای فرآیند تخصیص تصادفی، مجموعه داده‌ها به ۷۰۰ داده آموزش (۷۰٪ داده‌ها) و ۳۰۰ داده تست (۳۰٪ داده‌ها) تقسیم‌بندی شد. تمامی مراحل آماده‌سازی داده‌ها، اعتبارسنجی و شبیه‌سازی-بهینه‌یابی، در نرم‌افزار متلب (نسخه ۲۰۲۰b) پیاده‌سازی و اجرا گردید.

---

[۱] Statistics and Machine Learning Toolbox

[۲] Matlab

[۳] Levenberg-Marquardt error backpropagation algorithm

۴-۱- انتخاب ویژگی‌ها

به منظور انتخاب ویژگی‌ها از بین ۲۷ نسبت مالی، از الگوریتم حداقل افزونگی حداکثر ارتباط (MRMR) استفاده شد. بر اساس امتیاز محاسبه شده برای هر ویژگی، تعداد ۸ ویژگی با امتیازی بالاتر از ۰.۰۱ انتخاب شدند که عبارتند از:

- نسبت نقدینگی
- دوره گردش موجودی کالا
- نسبت کالا به سرمایه در گردش
- نسبت بدهی
- نسبت بدهی بلندمدت به حقوق صاحبان سهام
- نسبت حقوق صاحبان سهام به کل بدهی‌ها
- نسبت حقوق صاحبان سهام به دارایی‌های ثابت
- نسبت پوشش بدهی

نسبت نقدینگی جزء نسبت‌های نقدینگی بوده و دوره گردش موجودی کالا و نسبت کالا به سرمایه در گردش، جزء نسبت‌های فعالیت (کارایی) می‌باشند. سایر نسبت‌های مالی انتخاب شده، جزء نسبت‌های اهرمی (بدهی) در نظر گرفته می‌شوند. هشت نسبت مالی فوق به همراه سه ویژگی تسهیلات (اندازه، سود و مدت تسهیلات)، که مجموعا ۱۱ ویژگی را شامل می‌شوند، برای استفاده در مدل اعتبارسنجی انتخاب شدند.

۴-۲- نتایج مدل‌های اعتبارسنجی

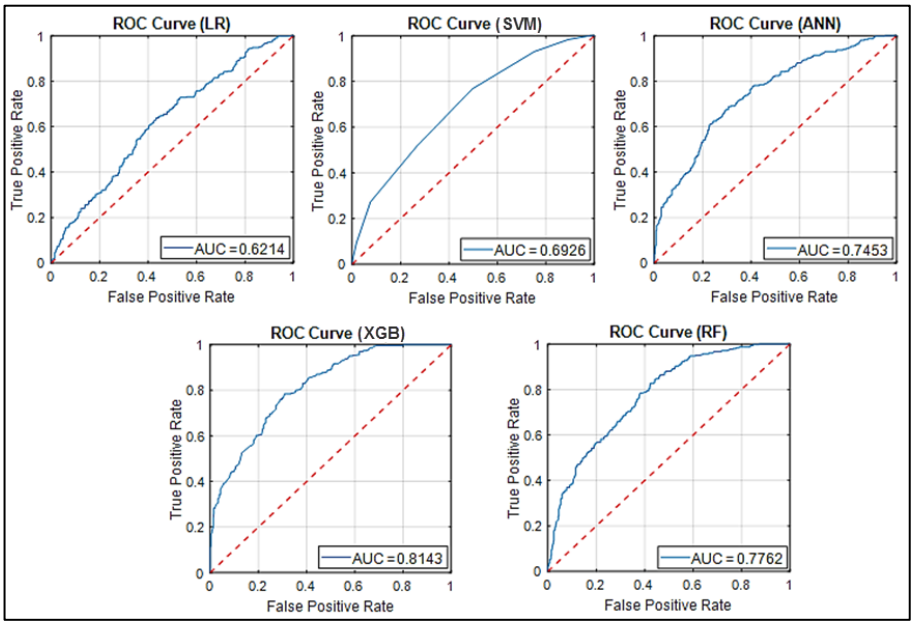
۴-۲-۱- نتایج مدل‌ها بر روی داده‌های آموزش

برای هر یک از مدل‌های اعتبارسنجی، منحنی مشخصه عملکرد سیستم (ROC) برای داده‌های آموزش در شکل ۳ نمایش داده شده و مقدار سطح زیر منحنی (AUC) برای هر مدل مشخص شده است. در این منحنی‌ها، محور عمودی بیانگر نرخ مثبت صادق (TPR) و محور افقی بیانگر نرخ مثبت کاذب (FPR) است. بالاتر بودن مقدار AUC از مقدار آستانه ۰/۵۰ که بیانگر بالاتر بودن TPR از FPR است، یکی از معیارهای اصلی ارزیابی عملکرد مدل‌های طبقه‌بندی می‌باشد. بررسی شکل ۳ نشان می‌دهد که تمامی پنج مدل اعتبارسنجی، از منظر AUC عملکرد خوبی نشان داده‌اند و مقادیر AUC برای تمامی مدل‌ها بالاتر از ۰/۶۰ بوده است.

معیارهای ارزیابی مدل‌های اعتبارسنجی در داده‌های آموزش، در جدول ٤ ارائه شده است. برای هر یک از معیارها، بهترین عملکرد، برجسته[۱] شده است. همانطور که در این جدول مشاهده می‌شود، روش‌های XGB و ANN به ترتیب بهترین عملکرد را از منظر معیارهای صحت و بازیابی داشته‌اند. با ترکیب این دو معیار، نمره F۱ محاسبه شده که مقدار آن برای تمامی روش‌ها بجز رگرسیون لجستیک (LR)، بالاتر از ۰/۵۰ بوده است. بهترین عملکرد را از منظر نمره F۱، روش ANN با مقدار ۰.٦۲۰۸ داشته است. دقت تمامی مدل‌ها نیز در حدود ۰/٦۰ یا بالاتر از آن بوده و بهترین عملکرد مربوط به روش XGB با دقت ۰/۷۲ بوده است. همچنین سطح زیر منحنی (AUC) برای تمامی مدل‌ها بالاتر از ۰/٦۰ بوده که بهترین عملکرد را روش XGB با مقدار ۰/۸۱ داشته است. بررسی این نتایج نشان‌گر این است که تمامی مدل‌ها از منظر معیارهای ارائه شده، عملکرد قابل قبولی داشته‌اند و مسئله کم‌برازش[۲] (برازش ضعیف مدل) در آنها رخ نداده است.

[۱] Bold

[۲] Underfitting



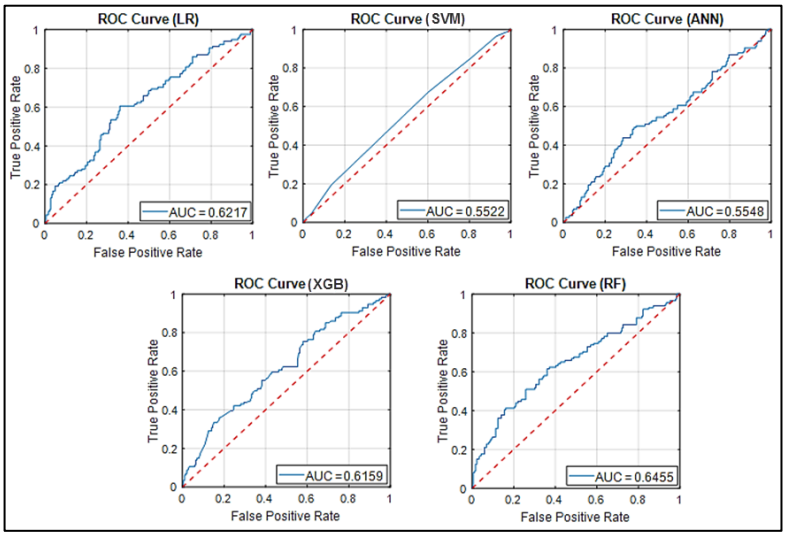
شکل ۳: منحنی مشخصه عملکرد سیستم (ROC) برای هر یک از مدل‌های اعتبارسنجی (داده‌های آموزش)

جدول ۴: معیارهای ارزیابی مدل‌های اعتبارسنجی (داده‌های آموزش)

Credit Scoring Model	Precision	Recall	F۱-Score	Accuracy	AUC
Logistic Regression (LR)	۰,۵۲۱۳	۰,۳۲۵٦	۰,٤۰۰۸	۰,۵۸۱٤	۰,٦۲۱٤
Support Vector Machine (SVM)	۰,۵۹۳۹	۰,۵۱۵۰	۰,۵۵۱٦	۰,٦٤۰۰	۰,٦۹۲٦
Artificial Neural Network (ANN)	۰,٦٦۱۷	**۰,۵۸٤۷**	۰,٦۲۰۸	۰,٦۹۲۹	۰,۷٤۵۳
Extreme Gradient Boosting (XGB)	**۰,۷۵۱۲**	۰,۵۲۱٦	۰,٦۱۵۷	۰,۷۲۰۰	**۰,۸۱٤۳**
Random Forest (RF)	۰,۷٤۰۱	۰,٤۳۵۲	۰,۵٤۸۱	۰,٦۹۱٤	۰,۷۷٦۲

۴-۲-۲- نتایج مدل‌ها بر روی داده‌های تست

برای هر یک از مدل‌های اعتبارسنجی، منحنی مشخصه عملکرد سیستم (ROC) برای داده‌های تست در شکل ٤ نمایش داده شده است. همانطور که در این شکل مشاهده می‌شود، در تمامی مدل‌های اعتبارسنجی، مقدار AUC بالاتر از ۰/۵۵ بوده است. سه مدل LR، XGB و RF بهترین عملکرد را داشته‌اند.



شکل ۴: منحنی مشخصه عملکرد سیستم (ROC) برای هر یک از مدل‌های اعتبارسنجی (داده‌های تست)

نتایج به دست آمده برای معیارهای ارزیابی مدل‌های اعتبارسنجی در داده‌های تست، در جدول ٥ ارائه شده و بهترین عملکرد در هر معیار، به صورت برجسته مشخص شده است. همانطور که در این جدول مشاهده می‌شود، از منظر معیارهای صحت و بازیابی، بهترین عملکرد را به ترتیب روش‌های RF و ANN داشته‌اند. ترکیب این دو شاخص در نمره F۱ نشان‌گر این است که روش RF با مقدار ۰/٤٦ بهترین عملکرد را داشته است. از منظر معیارهای دقت و AUC نیز اگرچه روش‌های LR و XGB عملکرد خوبی داشته‌اند، اما بهترین عملکرد مربوط به روش RF با دقت ۰/٦٨ و AUC برابر با ۰/٦٤ بوده است.

جدول ٥: معیارهای ارزیابی مدل‌های اعتبارسنجی (داده‌های تست)

Credit Scoring Model	Precision	Recall	F۱-Score	Accuracy	AUC
Logistic Regression (LR)	۰,٤٦٣٤	۰,۳۳۳۳	۰,۳۸۷۸	۰,٦۰۰۰	۰,٦۲۱۷
Support Vector Machine (SVM)	۰,٤۲۲۰	۰,٤۰۳٥	۰,٤۱۲٦	۰,٥٦۳۳	۰,٥٥۲۲
Artificial Neural Network (ANN)	۰,٤۸۰۸	۰,٤۳۸٦	۰,٤٥۸۷	۰,٦۰٦۷	۰,٥٥٤۸
Extreme Gradient Boosting (XGB)	۰,٥٥٤۱	۰,۳٥۹٦	۰,٤۳٦۲	۰,٦٤٦۷	۰,٦۱٥۹
Random Forest (RF)	۰,٦٤۰٦	۰,۳٥۹٦	۰,٤٦۰۷	۰,٦۸۰۰	۰,٦٤٥٥

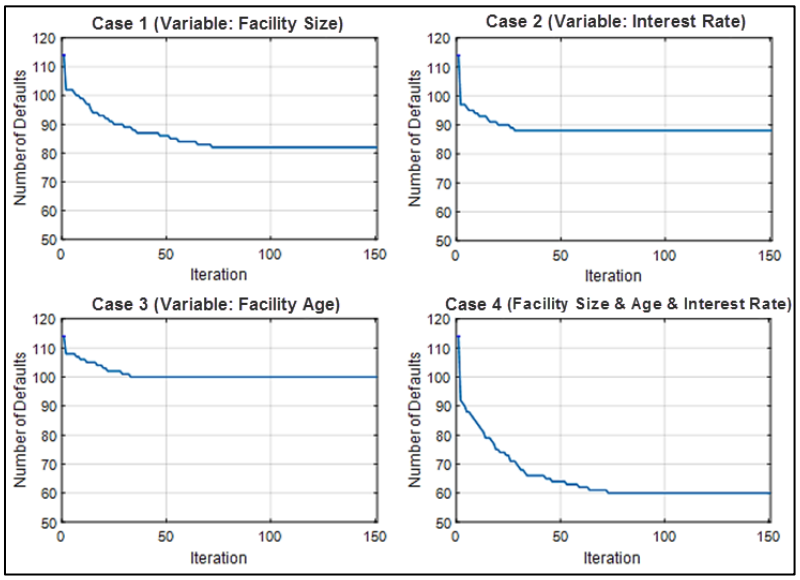
بررسی نتایج ارائه شده در جدول ٥ نشان‌گر این است که عملکرد کلی مدل‌های اعتبارسنجی مورد استفاده در این مقاله، قابل قبول بوده است. میانگین نمره F۱ مدل‌ها برابر با ۰/٤۳، میانگین دقت برابر با ۰/٦۲ و میانگین AUC برابر با ۰/٥۹ بوده است. با در نظر گرفتن سه معیار نمره F۱، دقت و AUC، می‌توان نتیجه گرفت که روش جنگل تصادفی (RF) در داده‌های تست مورد استفاده در این تحقیق، بهترین عملکرد را در بین مدل‌های مورد استفاده داشته است. بدین ترتیب، از مدل آموزش دیده RF برای ایفای نقش شبیه‌سازی در رویکرد شبیه‌سازی-بهینه‌یابی استفاده خواهد شد.

۴-۳- نتایج مدل شبیه‌سازی-بهینه‌یابی

در رویکرد شبیه‌سازی-بهینه‌یابی تلاش می‌شود با تغییر مشخصات تسهیلات اعطایی به مشتریان (اندازه، نرخ بهره و مدت تسهیلات)، احتمال نکول مشتریان و به تبع آن، ریسک اعتباری بانک کاهش داده شود. برای شبیه‌سازی فرآیند تاثیر تغییر در مشخصات تسهیلات بر احتمال نکول مشتریان، از مدل جنگل تصادفی (RF) که بهترین عملکرد را در بین مدل‌های اعتبارسنجی مورد استفاده داشته، استفاده شده است. تابع هدف مسئله بهینه‌سازی، حداقل‌سازی تعداد مشتریانی است که دچار نکول می‌شوند. در داده‌های تست که شامل ۳۰۰ مشتری است، تعداد ۱۱٤ مشتری (۳۸٪) در واقعیت دچار نکول شده‌اند. با هدف بررسی تاثیر هر یک از سه مشخصه تسهیلات بر وضعیت نکول مشتریان، رویکرد شبیه‌سازی-بهینه‌یابی در چهار حالت مختلف اجرا شده و نتایج آن با یکدیگر مقایسه می‌گردد.

منحنی همگرایی الگوریتم ممتیک در حالات مختلف بهینه‌سازی در شکل ٥ نمایش داده شده است. در این نمودارها، محور عمودی بیان‌گر تابع هدف مسئله (تعداد نکول تسهیلات) بوده و محور افقی، تکرارهای الگوریتم ممتیک را نشان می‌دهد. همانطور که مشاهده می‌شود، در تمامی حالات بهینه‌سازی، مقدار تابع هدف در ابتدا برابر با ۱۱٤ بوده، اما به تدریج توسط الگوریتم ممتیک کاهش یافته و به یک مقدار بهینه همگرا شده است. زمان حل مسئله بهینه‌سازی در هر حالت، حدودا ۳۰ دقیقه بوده است. بهترین عملکرد متعلق به حالت ٤ شبیه‌سازی بوده که در آن، هر سه مشخصه تسهیلات (اندازه، نرخ بهره و مدت تسهیلات) به عنوان متغیر در نظر گرفته شده‌اند.

جزئیات نتایج حالات مختلف بهینه‌سازی در جدول ٦ ارائه شده است. در این جدول به منظور بررسی تاثیر اجرای بهینه‌سازی، نتایج حالت پیش از بهینه‌سازی نیز ارائه شده است. برای هر یک از حالات بهینه‌سازی، تعداد و درصد نکول مشتریان به همراه مقادیر میانگین متغیرهای مسئله (اندازه، نرخ بهره و مدت تسهیلات) نمایش داده شده است.



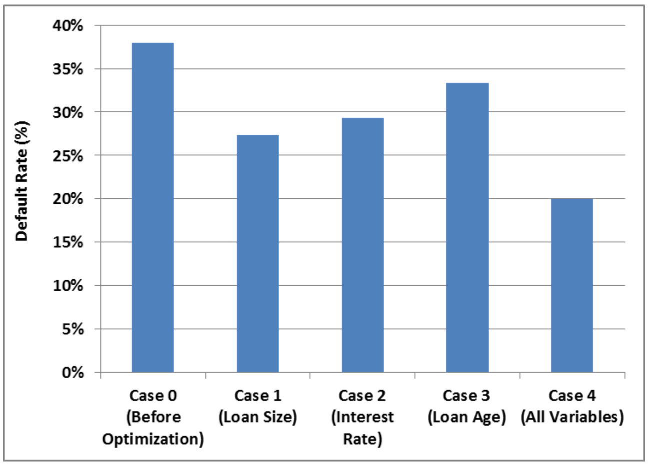
شکل ۵: منحنی همگرایی الگوریتم ممتیک در حالات مختلف بهینه‌سازی

جدول ۶: نتایج حالات مختلف بهینه‌سازی توسط الگوریتم ممتیک

Case	No. of Defaults	Default Rate (%)	Average Facility Size (Log)	Average Interest Rate (%)	Average Facility Age (year)
Case ۰ (Before Optimization)	۱۱٤	۳۸,۰۰٪	۱۰,۲۸	۱۵,۵۳	۳,۱۲
Case ۱ (Facility Size)	۸۲	۲۷,۳۳٪	۹,۸٦	۱۵,۵۳	۳,۱۲
Case ۲ (Interest Rate)	۸۸	۲۹,۳۳٪	۱۰,۲۸	۱۰,۲۱	۳,۱۲
Case ۳ (Facility Age)	۱۰۰	۳۳,۳۳٪	۱۰,۲۸	۱۵,۵۳	٤,۰٦
Case ٤ (All Variables)	٦۰	۲۰,۰۰٪	۹,۹۸	۱۰,۱٦	٤,۲۵

در حالت ۱ بهینه‌سازی، اندازه تسهیلات متغیر مسئله بهینه‌سازی بوده است. بر اساس نتایج ارائه شده در جدول ٦، تغییرات ایجاد شده در اندازه تسهیلات مشتریان موجب کاهش تعداد موارد نکول از ۱۱٤ مورد (۳۸٪) به ۸۲ مورد (۲۷/۳۳٪) شده است که نسبت به حالت پیش از بهینه‌سازی، ۱۰/٦۷٪ کاهش در نرخ نکول را نشان می‌دهد. در این حالت، متوسط لگاریتم اندازه تسهیلات اعطایی به مشتریان، از ۱۰/۲۸ (بیش از ۱۰ میلیارد ریال) به ۹/۸٦ (کمتر از ۱۰ میلیارد ریال) کاهش یافته است. در حالت ۲ بهینه‌سازی، نرخ بهره (سود تسهیلات) به عنوان متغیر مسئله در نظر گرفته شده است. تغییرات ایجاد شده در نرخ بهره مشتریان موجب کاهش تعداد موارد نکول از ۱۱٤ مورد (۳۸٪) به ۸۸ مورد (۲۹/۳۳٪) شده که معادل با ۸/٦۷٪ کاهش است. در این حالت، متوسط نرخ بهره تسهیلات مشتریان از ۱۵/۵۳٪ به ۱۰/۲۱٪ کاهش یافته است. در حالت ۳ بهینه‌سازی، مدت تسهیلات متغیر مسئله بوده است. تغییرات ایجاد شده در مدت تسهیلات موجب کاهش تعداد موارد نکول از ۱۱٤ مورد (۳۸٪) به ۱۰۰ مورد (۳۳/۳۳٪) شده که ٤/٦۷٪ کاهش را نشان می‌دهد. متوسط مدت تسهیلات مشتریان در این حالت، از ۳/۱۲ سال به ٤/۰٦ سال افزایش یافته است.

در حالت ٤ بهینه‌سازی، هر سه مشخصه تسهیلات (اندازه، نرخ بهره و مدت) به عنوان متغیرهای مسئله در نظر گرفته شده‌اند. تغییرات ایجاد شده در مشخصات تسهیلات موجب کاهش تعداد موارد نکول از ۱۱٤ مورد (۳۸٪) به ٦۰ مورد (۲۰٪) شده که نسبت به حالت پیش از بهینه‌سازی، ۱۸٪ کاهش در نرخ نکول را نشان می‌دهد. در این حالت نیز در مقایسه با حالت پیش از بهینه‌سازی، متوسط اندازه و نرخ بهره تسهیلات دچار کاهش شده و متوسط مدت تسهیلات، افزایش یافته است. همانطور که در شکل ٦ نیز نمایش داده شده است، با وجود اینکه در هر چهار حالت بهینه‌سازی، نرخ نکول نسبت به حالت پیش از بهینه‌سازی کاهش یافته، کمترین نرخ نکول مربوط به حالت ٤ بوده است که در آن هر سه مشخصه تسهیلات (اندازه، نرخ بهره و مدت) متغیر بوده‌اند. در بین سه حالت نخست بهینه‌سازی نیز حالات ۱ و ۳ به ترتیب بیشترین و کمترین کاهش در نرخ نکول را موجب شده‌اند. بنابراین اندازه تسهیلات دارای بیشترین تاثیر و مدت تسهیلات دارای کمترین تاثیر بر نرخ نکول بوده و نرخ بهره از نظر تاثیرگذاری، بین این دو جای گرفته است.



شکل ۶: درصد نکول در حالات مختلف بهینه‌سازی

۵- بحث و نتیجه‌گیری

در این مقاله، طراحی مدل شبیه‌سازی-بهینه‌یابی به منظور اعتبارسنجی و تعیین مشخصات بهینه تسهیلات مشتریان انجام شده و مطالعه موردی بر روی بنگاه‌های مشتری یک بانک تجاری صورت گرفت. روش مورد استفاده در این مقاله شامل سه مرحله است. در مرحله اول، آماده‌سازی داده‌ها انجام گرفت. بر اساس اطلاعات صورت‌های مالی و تسهیلات بانکی، ۳۰ ویژگی تعریف و محاسبه شده و پیش‌پردازش داده‌ها صورت گرفت. در مرحله دوم، با بکارگیری الگوریتم حداقل افزونگی حداکثر ارتباط (MRMR)، ۱۱ ویژگی برای استفاده در اعتبارسنجی انتخاب گردید. سپس پنج مدل اعتبارسنجی مختلف آموزش داده شده و بهترین مدل بر اساس عملکرد روی داده‌های تست انتخاب شد. در مرحله سوم، با بکارگیری بهترین مدل اعتبارسنجی در فرآیند شبیه‌سازی-بهینه‌یابی، مشخصات بهینه تسهیلات اعطایی به مشتریان توسط الگوریتم ممتیک در چهار حالت مختلف تعیین گردید.

تعریف و محاسبه ویژگی‌های مورد استفاده در مدل‌های اعتبارسنجی، بر اساس داده‌های گردآوری شده از مشتریان حقوقی یک بانک تجاری در ایران صورت گرفت. از بین ۲۷ نسبت مالی محاسبه شده، تعداد ۸ نسبت مالی که شامل نسبت‌های نقدینگی، فعالیت (کارایی) و اهرمی (بدهی) می‌باشند، دارای تاثیر قابل توجه بر متغیر پاسخ (وضعیت نکول) بودند. به دلیل عدم

دسترسی به برخی اطلاعات، محاسبه برخی از نسبت‌های مالی (از جمله نسبت‌های ارزش بازار[1])، ممکن نبود. با این وجود، باید توجه داشت که با در نظر گرفتن مشکلات دسترسی به داده‌های تسهیلات بانکی و صورت‌های مالی شرکت‌های متناظر با تسهیلات، گردآوری داده‌های مورد استفاده در این مقاله از منظر حجم نمونه و نوع اطلاعات، کاری دشوار بوده است. امید است تحقیقات آینده بتوانند با افزایش حجم نمونه و افزایش اطلاعات گردآوری شده، رویکرد شبیه‌سازی-بهینه‌یابی این تحقیق را با داده‌های بیشتری آزمایش کنند.

در این مقاله از پنج مدل مختلف طبقه‌بندی به منظور اعتبارسنجی مشتریان استفاده شد. روش LR از نوع روش‌های کلاسیک یادگیری ماشین بوده و نقش مدل پایه را ایفا می‌کنند. روش‌های XGB و RF که هر دو در دسته مدل‌های یادگیری جمعی جای می‌گیرند، از جمله جدیدترین روش‌های پرکاربرد در حل مسئله اعتبارسنجی می‌باشند. روش‌های SVM و ANN نیز بین مدل‌های کلاسیک و مدل‌های پیش‌رو جای می‌گیرند. بررسی معیارهای ارزیابی مدل‌ها نشان داد که عملکرد تمامی مدل‌ها، حداقل‌های لازم را دارا بوده است. با آزمایش مدل‌ها بر روی داده‌های تست، روش جنگل تصادفی (RF) بهترین عملکرد را از منظر معیارهای نمره F۱، دقت و سطح زیر منحنی (AUC) داشت. بنابراین در مجموع روش RF که جزء روش‌های یادگیری جمعی است، به عنوان بهترین روش برگزیده شد. این نتیجه هم راستا با نقش مؤثر روش‌های یادگیری جمعی در تحقیقات اخیر در حوزه اعتبارسنجی است [۱۰] و [۱۱].

نوآوری اصلی این مقاله، معرفی و پیاده‌سازی رویکرد شبیه‌سازی-بهینه‌یابی است. در این رویکرد، پس از آموزش مدل‌های اعتبارسنجی، مشخصات تسهیلات اعطایی به مشتریان به عنوان متغیر در نظر گرفته شد و این امکان در نظر گرفته شد که بانک‌ها در یک محدوده مشخص، مشخصات تسهیلات اعطایی به مشتریان خود را تغییر دهند. برای ایجاد امکان بررسی تاثیر مشخصات تسهیلات بر احتمال نکول مشتریان، از ایده شبیه‌سازی استفاده شد. در شبیه‌سازی، تمامی ویژگی‌های مشتریان بجز مشخصات تسهیلات، ثابت بوده و در نتیجه امکان بررسی تاثیر تغییرات در این مشخصات بر احتمال نکول مشتریان فراهم می‌شود. با این مفروضات، فرآیند شبیه‌سازی-بهینه‌یابی به صورت یک مسئله بهینه‌سازی مدلسازی شد که در آن، مشخصات تسهیلات (اندازه، نرخ بهره و مدت) متغیرهای مسئله می‌باشند. حل مسئله

[1] Market ratios

بهینه‌سازی توسط الگوریتم ممتیک صورت گرفت. نتایج حاصل از الگوریتم ممتیک نشان‌گر کارایی این الگوریتم در حل مسئله بهینه‌سازی در زمانی نسبتا کوتاه (حدود ۳۰ دقیقه) می‌باشد. با هدف مقایسه حالات مختلف، حل مسئله بهینه‌سازی در چهار حالت صورت گرفت. در حالات اول تا سوم، به ترتیب تنها اندازه تسهیلات، نرخ بهره و مدت زمان بازپرداخت به عنوان متغیرهای مسئله بهینه‌سازی در نظر گرفته شدند. در حالت چهارم، هر سه مشخصه فوق به طور همزمان به عنوان متغیر لحاظ شدند. پیش از حل مسئله بهینه‌سازی، نرخ نکول برابر با ٪۳۸ بوده است. نرخ نکول در حالات اول تا چهارم بهینه‌سازی به ترتیب به ۲۷/۳۳٪، ۲۹/۳۳٪، ۳۳٪/۳۳ و ۲۰٪ رسیده است. در نتیجه رویکرد شبیه‌سازی-بهینه‌یابی موفق شده با تغییر در مشخصات تسهیلات اعطایی به مشتریان، احتمال نکول آنها را تا حد قابل توجهی کاهش داده و از این طریق، از ریسک اعتباری بانک بکاهد. بررسی نتایج نشان می‌دهد که اندازه تسهیلات بیشترین تاثیر و مدت تسهیلات کمترین تاثیر را بر کاهش نرخ نکول داشته است. بر اساس نتایج حاصله، هر تغییری در تسهیلات که موجب کاهش در مبلغ اقساط پرداختی مشتریان شود (چه از طریق کاهش اندازه و سود تسهیلات و چه از طریق افزایش زمان بازپرداخت)، احتمال نکول آنها را کاهش می‌دهد.

ایده شبیه‌سازی-بهینه‌یابی، می‌تواند افق جدیدی در فرآیند اعطای تسهیلات بانک‌ها و مؤسسات مالی ایجاد کند. در فرآیند رایج، ابتدا اعتبارسنجی مشتریان انجام شده و سپس در صورتی که مشتری در دسته خوش‌حساب جای بگیرد، مشخصات تسهیلات مشتری تعیین می‌گردد. در نتیجه روش مرسوم، تاثیر مشخصات تسهیلات بر احتمال نکول مشتریان را در فرآیند اعتبارسنجی نادیده می‌گیرد. ایده شبیه‌سازی-بهینه‌یابی تعیین مشخصات بهینه تسهیلات را با فرآیند اعتبارسنجی ترکیب می‌کند تا از این طریق، تاثیر مشخصات تسهیلات بر احتمال نکول مشتریان را لحاظ کرده و از آن برای کاهش ریسک اعتباری بانک‌ها بهره ببرد.

تحلیل نتایج حاصل از مطالعه موردی بر روی مشتریان یک بانک تجاری در ایران نشان داد که تاثیر مدل بهینه‌یابی-شبیه‌سازی بر نرخ نکول مشتریان، از طریق کاهش اندازه (مبلغ وام) و سود (نرخ بهره) و افزایش مدت زمان بازپرداخت وام صورت پذیرفته است. بنابراین می‌توان نتیجه گرفت که بر اساس یافته‌های مطالعه موردی، هر تغییری در تسهیلات که موجب کاهش در مبلغ اقساط پرداختی مشتریان شود، احتمال نکول آنها را کاهش می‌دهد. این یافته به طور

بالقوه می‌تواند پیشنهاداتی را به بانک‌ها و موسسات دولتی جهت مدیریت ریسک اعتباری خود از طریق تعیین بهینه پارامترهای تسهیلات اعطایی به مشتریان ارائه دهد. بانک‌های کشور می‌توانند در هنگام عقد قرارداد تسهیلات با مشتریان حقوقی و حقیقی، به برآورد توانایی آنها در بازپرداخت تسهیلات (بر اساس درآمدها و هزینه‌های ماهیانه آنها) بپردازند و در صورتی که اقساط تسهیلات نسبت به وضعیت مالی آنها سنگین است، اقدام به تصحیح پارامترهای وام نمایند. این تصحیح می‌تواند از طریق کاهش اندازه وام و سود آن و یا افزایش مدت زمان بازپرداخت وام باشد. بدین ترتیب توانایی و تمایل مشتریان به بازپرداخت اقساط خود افزایش یافته و از این طریق، نرخ نکول آنها کاهش خواهد یافت.

روش پیشنهادی این تحقیق، علاوه بر کاربردی که برای بانک‌های تجاری دارد، همچنین می‌تواند توسط بانک مرکزی و سایر سازمان‌های سیاست‌گذار و قانون‌گذار در حوزه پولی و بانکی مورد استفاده قرار گیرد. بانک مرکزی و سایر ارگان‌های حاکمیتی می‌توانند با گردآوری اطلاعات مشتریان حقیقی و حقوقی تمامی بانک‌های کشور، مدل بهینه‌یابی-شبیه‌سازی را بر روی آنها اجرا کنند و بر اساس یافته‌های آن، به سیاست‌گذاری و قانون‌گذاری در مورد نحوه تعیین پارامترهای وام (از جمله اندازه، سود و مدت وام و همچنین نوع و ارزش وثیقه) بپردازند. با توجه با اینکه سیستم بانکی کشور با معضل افزایش حجم بالای نکول تسهیلات و در نتیجه افزایش معوقات بانکی مواجه است، استفاده از ایده‌هایی چون مدل بهینه‌یابی-شبیه‌سازی می‌تواند راهکاری برای کاستن از نرخ نکول تسهیلات و کاهش ریسک اعتباری بانک‌ها باشد.

۶- منابع

[۱] Soui, M., Gasmi, I., Smiti, S., & Ghédira, K. (۲۰۱۹). Rule-based credit risk assessment model using multi-objective evolutionary algorithms. *Expert systems with applications*, *126*, ۱٤٤-۱٥۷.

[۲] Minaei, B., Afshar, A., & Houshdar Mahjoub, R. (۲۰۱٤). Customer credit clustering for Present appropriate facilities. *Management Research in Iran*, *7*(٤), ۱-۲٤.

[۳] Safari, S., Ebrahimi, M., & Sheik, M. J. (۲۰۱۱). Managing the Credit Risk of the Bank's Clients in Commercial Banks DEA Approach (Credit Rating). *Management Research in Iran*, *14*(٤), ۱۳۷-۱٦٤.

[٤] Ahmed, M. I., & Rajaleximi, P. R. (۲۰۱۹). An empirical study on credit scoring and credit scorecard for financial institutions. *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, *8*(۷), ۲۲۷۸-۱۳۲۳.

[٥] Bluhm, C., Overbeck, L., & Wagner, C. (۲۰۱٦). *Introduction to credit risk modeling*. Crc Press.

[٦] Ghasemnia Arabi, N., & Safaei Ghadikolaei, A. (۲۰۱٦). A New Approach of applying multi criteria decision making models for classifying the credit customers of bank. *Modern Research in Decision Making*, *1*(۳), ٤۳-٦٤.

[۷] Sohrabi, B., Raeesi Vanani, I., & Zareh Mirkabad, F. (۲۰۱٦). Designing a Recommender System for Optimizing and Managing Bank Facilities through the Utilization of Clustering and Classification Algorithms. *Modern Research in Decision Making*, *1*(۲), ٥۳-۷٦.

[۸] Abdou, H. A., & Pointon, J. (۲۰۱۱). Credit scoring, statistical techniques and evaluation criteria: a review of the literature. *Intelligent systems in accounting, finance and management*, *18*(۲-۳), ٥۹-۸۸.

[۹] Koutanaei, F. N., Sajedi, H., & Khanbabaei, M. (۲۰۱٥). A hybrid data mining model of feature selection algorithms and ensemble learning classifiers for credit scoring. *Journal of Retailing and Consumer Services*, *27*, ۱۱-۲۳.

[۱۰] Dastile, X., Celik, T., & Potsane, M. (۲۰۲۰). Statistical and machine learning models in credit scoring: A systematic literature survey. *Applied Soft Computing*, *91*, ۱۰٦۲٦۳.

[۱۱] Markov, A., Seleznyova, Z., & Lapshin, V. (۲۰۲۲). Credit scoring methods: Latest trends and points to consider. *The Journal of Finance and Data Science.*

[۱۲] Dželihodžić, A., Đonko, D., & Kevrić, J. (۲۰۱۸). Improved credit scoring model based on bagging neural network. *International Journal of Information Technology & Decision Making*, *17*(۰٦), ۱۷۲٥-۱۷٤۱.

[۱۳] Nehrebecka, N. (۲۰۱۸). Predicting the default risk of companies. Comparison of credit scoring models: LOGIT vs Support Vector Machines. Econometrics. Ekonometria. Advances in Applied Data Analytics, ۲۲(۲), ٥٤-۷۳.

[۱٤] Dželihodžić, A., Đonko, D., & Kevrić, J. (۲۰۱۸). Improved credit scoring model based on bagging neural network. International Journal of Information Technology & Decision Making, ۱۷(۰٦), ۱۷۲٥-۱۷٤۱.

[۱٥] Moradi, S., & Mokhatab Rafiei, F. (۲۰۱۹). A dynamic credit risk assessment model with data mining techniques: evidence from Iranian banks. Financial Innovation, ٥(۱), ۱-۲۷.

[۱٦] Munkhdalai, L., Munkhdalai, T., Namsrai, O. E., Lee, J. Y., & Ryu, K. H. (۲۰۱۹). An empirical comparison of machine-learning methods on bank client credit assessments. Sustainability, ۱۱(۳), ٦۹۹.

[۱۷] Nalić, J., & Martinovic, G. (۲۰۲۰). Building a credit scoring model based on data mining approaches. International Journal of Software Engineering and Knowledge Engineering, ۳۰(۰۲), ۱٤۷-۱٦۹.

[۱۸] Boughaci, D., & Alkhawaldeh, A. A. (۲۰۲۰). Appropriate machine learning techniques for credit scoring and bankruptcy prediction in banking and finance: A comparative study. Risk and Decision Analysis, ۸(۱-۲), ۱٥-۲٤.

[۱۹] Pławiak, P., Abdar, M., Pławiak, J., Makarenkov, V., & Acharya, U. R. (۲۰۲۰). DGHNL: A new deep genetic hierarchical network of learners for prediction of credit scoring. Information Sciences, ٥۱٦, ٤۰۱-٤۱۸.

[۲۰] Ampountolas, A., Nyarko Nde, T., & Constantinescu, C. (۲۰۲۱). A machine learning approach for micro-credit scoring. Risks, ۹(۳), ٥۰.

[۲۱] Biecek, P., Chlebus, M., Gajda, J., Gosiewska, A., Kozak, A., Ogonowski, D., ... & Wojewnik, P. (۲۰۲۱). Enabling machine learning algorithms for credit scoring--explainable artificial intelligence (XAI) methods for clear understanding complex predictive models. arXiv preprint arXiv:۲۱۰٤,۰٦۷۳٥.

[۲۲] Laborda, J., & Ryoo, S. (۲۰۲۱). Feature selection in a credit scoring model. Mathematics, ۹(۷), ۷٤٦.
[۲۳] Tripathi, D., Edla, D. R., Bablani, A., Shukla, A. K., & Reddy, B. R. (۲۰۲۱). Experimental analysis of machine learning methods for credit score classification. Progress in Artificial Intelligence, ۱۰, ۲۱۷-۲٤۳.
[۲٤] Hussin Adam Khatir, A. A., & Bee, M. (۲۰۲۲). Machine learning models and data-balancing techniques for credit scoring: What is the best combination?. Risks, ۱۰(۹), ۱٦۹.
[۲٥] Lenka, S. R., Bisoy, S. K., Priyadarshini, R., & Sain, M. (۲۰۲۲). Empirical analysis of ensemble learning for imbalanced credit scoring datasets: A systematic review. Wireless Communications and Mobile Computing, ۲۰۲۲.
[۲٦] Sum, R. M., Ismail, W., Abdullah, Z. H., & Shah, N. F. M. N. (۲۰۲۲). A New Efficient Credit Scoring Model For Personal Loan Using Data Mining Technique For Sustainability Management. *Journal of Sustainability Science and Management*, *12*(۲), ۱۳٥٤٦۷.
[۲۷] Zhou, M. (۲۰۲۲). Credit Risk Assessment Modeling Method Based on Fuzzy Integral and SVM. Mobile Information Systems, ۲۰۲۲.
[۲۸] El Qadi, A., Trocan, M., Diaz-Rodriguez, N., & Frossard, T. (۲۰۲۳). Feature contribution alignment with expert knowledge for artificial intelligence credit scoring. Signal, Image and Video Processing, ۱۷(۲), ٤۲۷-٤۳٤.
[۲۹] Abdoli, M., Akbari, M., & Shahrabi, J. (۲۰۲۳). Bagging Supervised Autoencoder Classifier for credit scoring. Expert Systems with Applications, ۲۱۳, ۱۱۸۹۹۱.
[۳۰] Runchi, Z., Liguo, X., & Qin, W. (۲۰۲۳). An ensemble credit scoring model based on logistic regression with heterogeneous balancing and weighting effects. Expert Systems with Applications, ۲۱۲, ۱۱۸۷۳۲.
[۳۱] Aljadani, A., Alharthi, B., Farsi, M. A., Balaha, H. M., Badawy, M., & Elhosseini, M. A. (۲۰۲۳). Mathematical Modeling and Analysis of Credit Scoring Using the LIME Explainer: A Comprehensive Approach. Mathematics, ۱۱(۱۹), ٤۰٥٥.
[۳۲] Chen, Y., Calabrese, R., & Martin-Barragan, B. (۲۰۲٤). Interpretable machine learning for imbalanced credit scoring datasets. European Journal of Operational Research, ۳۱۲(۱), ۳٥۷-۳۷۲.
[۳۳] Mushava, J., & Murray, M. (۲۰۲٤). Flexible loss functions for binary classification in gradient-boosted decision trees: An application to credit scoring. Expert Systems with Applications, ۲۳۸, ۱۲۱۸۷٦.
[۳٤] Rofik, R., Aulia, R., Musaadah, K., Ardyani, S. S. F., & Hakim, A. A. (۲۰۲٤). The Optimization of Credit Scoring Model Using Stacking Ensemble Learning and Oversampling Techniques. Journal of Information System Exploration and Research, ۲(۱), ۱۱-۲۰.
[۳٥] Xiao, J., Zhong, Y., Jia, Y., Wang, Y., Li, R., Jiang, X., & Wang, S. (۲۰۲٤). A novel deep ensemble model for imbalanced credit scoring in internet finance. International Journal of Forecasting, ٤۰(۱), ۳٤۸-۳۷۲.
[۳٦] Krishna, S. J. S., Aarif, M., Bhasin, N. K., Kadyan, S., & Bala, B. K. (۲۰۲٤, July). Predictive Analytics in Credit Scoring: Integrating XG Boost and Neural Networks for Enhanced Financial Decision Making. In ۲۰۲٤ International Conference on Data Science and Network Security (ICDSNS) (pp. ۱-٦). IEEE.

[۳۷] Talaat, F. M., Aljadani, A., Badawy, M., & Elhosseini, M. (۲۰۲٤). Toward interpretable credit scoring: integrating explainable artificial intelligence with deep learning for credit card default prediction. Neural Computing and Applications, ۳٦(۹), ٤۸٤۷-٤۸٦٥.

[۳۸] Groppelli, A. A., & Nikbakht, E. (۲۰۰۰). Finance, Barron's Educational Series. Inc., ISBN, ۷٦٤۱۱۲۷٥۷.

[۳۹] Yang, F., Qiao, Y., Huang, C., Wang, S., & Wang, X. (۲۰۲۱). An automatic credit scoring strategy (ACSS) using memetic evolutionary algorithm and neural architecture search. Applied Soft Computing, ۱۱۳, ۱۰۷۸۷۱.

[٤۰] Acuna, E., & Rodriguez, C. (۲۰۰٤). A meta analysis study of outlier detection methods in classification. Technical paper, Department of Mathematics, University of Puerto Rico at Mayaguez, ۱, ۲٥.

[٤۱] Dawson, R. (۲۰۱۱). How significant is a boxplot outlier?. Journal of Statistics Education, ۱۹(۲).

[٤۲] Deng, X., Li, M., Deng, S., & Wang, L. (۲۰۲۲). Hybrid gene selection approach using XGBoost and multi-objective genetic algorithm for cancer classification. Medical & Biological Engineering & Computing, ٦۰(۳), ٦٦۳-٦۸۱.

[٤۳] Xie, S., Zhang, Y., Lv, D., Chen, X., Lu, J., & Liu, J. (۲۰۲۳). A new improved maximal relevance and minimal redundancy method based on feature subset. The Journal of Supercomputing, ۷۹(۳), ۳۱٥۷-۳۱۸۰.

[٤٤] Louzada, F., Ara, A., & Fernandes, G. B. (۲۰۱٦). Classification methods applied to credit scoring: Systematic review and overall comparison. Surveys in Operations Research and Management Science, ۲۱(۲), ۱۱۷-۱۳٤.

[٤٥] Sörensen, K., & Sevaux, M. (۲۰۰٦). MA| PM: memetic algorithms with population management. Computers & Operations Research, ۳۳(٥), ۱۲۱٤-۱۲۲٥.

[٤٦] Moscato, P. (۱۹۸۹). On evolution, search, optimization, genetic algorithms and martial arts: Towards memetic algorithms. Caltech concurrent computation program, C۳P Report, ۸۲٦(۱۹۸۹), ۳۷.

[٤۷] Neri, F., & Cotta, C. (۲۰۱۲). Memetic algorithms and memetic computing optimization: A literature review. Swarm and Evolutionary Computation, ۲, ۱-۱٤.

[٤۸] Kumar, R., & Memoria, M. (۲۰۲۰). A review of memetic algorithm and its application in traveling salesman problem. Int. J. Emerg. Technol., ۱۱(۲), ۱۱۱۰-۱۱۱٥.

[٤۹] Holland, J. H. (۱۹۷۳). Genetic algorithms and the optimal allocation of trials. SIAM journal on computing, ۲(۲), ۸۸-۱۰٥.