Journal of
Intelligent Financial
Management Perspectives

ISSN: 3116-0964

Credit Risk Prediction Model for Retail Customers in Lending Tech Platforms Using Alternative Data and Machine Learning

Saeed Abbasi [1]*, Mahdieh Karimi [2], Arash Zamani[3]

[1] *Bachelor of Financial Management, Islamic Azad University, North Tehran Branch, Tehran, Iran(Coresponding autour), saeed.abbasi@iau-ntb.ac.ir*

[2] *Master of Accounting, Islamic Azad University, Science and Research Branch, Tehran, Iran*

[3] *PhD in Economics, Islamic Azad University, Science and Research Branch, Tehran, Iran*

ARTICLE INFO

Article history:
Received:05/06/2025
Received in revised form: 15/07/2025
Accepted:01/08/2025
Available online:16/08/2025

Keywords:

Credit Risk
Lendtech
Alternative Data
Machine Learning
Explainable Artificial Intelligence

ABSTRACT

This study proposes a new framework for predicting retail customers' credit risk in LendTech platforms using alternative data and machine learning techniques. Traditional credit scoring methods often fail to capture complex behavioral patterns and perform poorly in digital lending environments. The research utilizes alternative data sources such as digital behavior, transaction history, mobile data, and user interactions to improve default prediction accuracy. Several machine learning models, including Logistic Regression, Random Forest, Decision Trees, and Gradient Boosting, are evaluated. To ensure transparency and interpretability, the SHAP explainability method is applied to analyze feature importance. Results show that ensemble models, especially Gradient Boosting and Random Forest, achieve the highest predictive performance, while behavioral indicators such as transaction patterns, repayment behavior, income stability, and platform interactions are the most influential factors in credit risk assessment. The findings demonstrate that combining alternative data, machine learning, and explainable AI can provide a reliable and effective credit-scoring framework for LendTech ecosystems.

Article Type: Research Paper

Journal of Intelligent Financial Management, 2025, Vol. 1, No.2, pp. 56- 70

Publish by:
Tolou-e Binish-e Ayandeh Scientific Institute

https://doi.org/ 10.25843/JIFM.2025.8563.21222

©Authors

Cite: Abbasi,S, Karimi,M and Zamani,A. (2025). Credit Risk Prediction Model for Retail Customers in Lending Tech Platforms Using Alternative Data and Machine Learning. *Journal of Intelligent Financial Management*, *1*(2), 56-70.

نگرش های
مدیریت مالی هوشمند
Journal of Intelligent Financial Management Perspectives
ISSN: 3116-0964

مدل پیش بینی ریسک اعتباری مشتریان خُرد در پلتفرم های لندتک با استفاده از داده های جایگزین و یادگیری ماشین

سعید عباسی [۱و*]، مهدیه کریمی [۲]، آرش زمانی

۱ و *- کارشناسی مدیریت مالی، دانشگاه آزاد اسلامی واحد تهران شمال، تهران، ایران (نویسنده مسئول)، ایمیل نویسنده مسئول: saeed.abbasi@iau-ntb.ac.ir

۲- کارشناسی ارشد حسابداری، دانشگاه آزاد اسلامی واحد علوم و تحقیقات تهران، تهران، ایران

۳- دکتری اقتصاد، دانشگاه آزاد اسلامی واحد علوم و تحقیقات تهران، تهران، ایران

اطلاعات مقاله

تاریخچه مقاله:

تاریخ دریافت:۱۴۰۴/۰۳/۱۵

تاریخ بازنگری:۱۴۰۴/۰۴/۲۴

تاریخ پذیرش:۱۴۰۴/۰۵/۱۰

تاریخ انتشار:۱۴۰۴/۰۵/۲۵

کلیدواژه‌ها:

ریسک اعتباری

لندتیک

داده های جایگزین

یادگیری ماشین

هوش مصنوعی توضیح پذیر

چکیده

این پژوهش یک چارچوب نوین برای پیش‌بینی ریسک اعتباری مشتریان خرد در پلتفرم‌های فناوری وام‌دهی (LendTech) با استفاده از داده‌های جایگزین و الگوریتم‌های یادگیری ماشین ارائه می‌دهد. روش‌های سنتی امتیازدهی اعتباری معمولاً قادر به شناسایی الگوهای پیچیده رفتاری نیستند و در محیط‌های وام‌دهی دیجیتال عملکرد محدودی دارند. در این مطالعه، از داده‌های جایگزین مانند رفتار دیجیتال کاربران، سوابق تراکنش‌ها، داده‌های تلفن همراه و تعاملات کاربران با پلتفرم برای افزایش دقت پیش‌بینی نکول استفاده شده است. همچنین چندین مدل یادگیری ماشین از جمله رگرسیون لجستیک، درخت تصمیم، جنگل تصادفی و گرادیان بوستینگ مورد ارزیابی قرار گرفته‌اند. برای افزایش شفافیت و قابلیت تفسیر نتایج، از روش SHAP جهت تحلیل اهمیت ویژگی‌ها استفاده شده است. نتایج نشان می‌دهد که مدل‌های تجمیعی، به‌ویژه Gradient Boosting و Random Forest ، بهترین عملکرد را در پیش‌بینی ریسک اعتباری دارند و عواملی مانند الگوهای تراکنش، رفتار بازپرداخت، ثبات درآمد و میزان تعامل کاربران با پلتفرم بیشترین تأثیر را بر نتایج دارند. یافته‌ها بیانگر آن است که ترکیب داده‌های جایگزین، یادگیری ماشین و هوش مصنوعی توضیح‌پذیر می‌تواند چارچوبی دقیق، کارآمد و قابل اعتماد برای امتیازدهی اعتباری در اکوسیستم‌های وام دهی فراهم کند.

نشریه مدیریت مالی هوشمند، ۱۴۰۴، دوره ۱، شماره ۲، صفحه ۵۶-۷۰.

ناشر: موسسه علمی طلوع بینش آینده

https://doi.org/ 10.25843/JIFM.2025.8563. 21222

نوع مقاله: پژوهشی

© نویسندگان

استناد: عباسی, سعید, کریمی, مهدیه و زمانی, آرش. (۱۴۰۴). مدل پیش بینی ریسک اعتباری مشتریان خُرد در پلتفرم های لندتک با استفاده از داده های جایگزین و یادگیری ماشین. *مدیریت مالی هوشمند*، ۱(۲)، ۵۶-۷۰.

ISSN: 3116-0964

۱-مقدمه

در دهه‌های اخیر، نظام‌های مالی جهانی با تحولات بنیادینی در اثر گسترش فناوری‌های دیجیتال، یادگیری ماشین و ظهور اکوسیستم‌های مالی نوآورانه مواجه شده‌اند. یکی از مهم‌ترین این تحولات، شکل‌گیری پلتفرم‌های لندتک (LendTech) است که با استفاده از فناوری‌های دیجیتال، فرآیند اعطای تسهیلات اعتباری را به‌صورت کاملاً خودکار، سریع و داده‌محور بازطراحی کرده‌اند. این پلتفرم‌ها با حذف بسیاری از واسطه‌های سنتی بانکی، امکان دسترسی افراد فاقد سابقه اعتباری یا دارای دسترسی محدود به خدمات مالی را فراهم کرده‌اند و بدین ترتیب نقش مهمی در فراگیری مالی (Financial Inclusion) ایفا می‌کنند(Fuster et al., 2019) .

با وجود مزایای گسترده این تحول، یکی از چالش‌های اساسی در پلتفرم‌های لندتک، ارزیابی دقیق ریسک اعتباری مشتریان خرد است. در نظام بانکی سنتی، ارزیابی اعتبار معمولاً بر اساس داده‌های ساخت‌یافته مالی مانند گزارش‌های اعتباری، درآمد رسمی، سوابق وام‌های قبلی و نسبت بدهی به درآمد انجام می‌شود. این داده‌ها در طول سال‌ها به‌عنوان پایه مدل‌های امتیازدهی اعتباری مورد استفاده قرار گرفته‌اند و در بسیاری از سیستم‌های مالی همچنان نقش اصلی را ایفا می‌کنند (Thomas, Crook, & Edelman, 2017) .با این حال، در محیط‌های لندتک، بخش قابل توجهی از کاربران فاقد این داده‌های سنتی هستند که این موضوع منجر به پدیده‌ای تحت عنوان «فقر داده اعتباری» شده است.

این محدودیت باعث شده است که مدل‌های کلاسیک اعتبارسنجی مانند رگرسیون لجستیک و مدل‌های خطی، کارایی مطلوبی در محیط‌های دیجیتال نداشته باشند. این مدل‌ها فرض‌های ساده‌سازی‌شده‌ای درباره روابط میان متغیرها دارند و توانایی محدودی در شناسایی روابط غیرخطی، تعاملات پیچیده و الگوهای رفتاری پنهان در داده‌های حجیم دارند(Hand & Henley, 1997) . در نتیجه، با رشد داده‌های دیجیتال و غیرساخت‌یافته، نیاز به استفاده از رویکردهای پیشرفته‌تر بیش از پیش احساس شده است.

در همین راستا، مفهوم داده‌های جایگزین به‌عنوان یکی از مهم‌ترین نوآوری‌ها در حوزه اعتبارسنجی مطرح شده است. داده‌های جایگزین شامل مجموعه‌ای از اطلاعات غیرسنتی نظیر رفتار دیجیتال کاربران، الگوهای استفاده از موبایل، تراکنش‌های آنلاین، فعالیت در شبکه‌های اجتماعی، داده‌های موقعیت مکانی و تعاملات کاربران با پلتفرم‌های مالی است. این داده‌ها به دلیل ماهیت رفتاری و لحظه‌ای خود، قادرند تصویر دقیق‌تری از وضعیت مالی و رفتاری افراد ارائه دهند (Berg et al., 2020) .برخلاف داده‌های سنتی که معمولاً ایستا هستند، داده‌های جایگزین پویا بوده و امکان تحلیل رفتار واقعی کاربران در زمان واقعی را فراهم می‌سازند.

مطالعات اخیر نشان داده‌اند که استفاده از داده‌های جایگزین می‌تواند به شکل قابل توجهی دقت مدل‌های پیش‌بینی ریسک اعتباری را افزایش دهد. برای مثال، تحلیل الگوهای تراکنشی و رفتار دیجیتال کاربران می‌تواند اطلاعات ارزشمندی درباره ثبات مالی، الگوی مصرف و احتمال نکول ارائه دهد. با این حال، پیچیدگی و حجم بالای این داده‌ها چالش‌های جدیدی را در زمینه پردازش و مدل‌سازی ایجاد کرده است که نیازمند استفاده از روش‌های پیشرفته تحلیل داده است.

در پاسخ به این چالش‌ها، یادگیری ماشین به‌عنوان یکی از مهم‌ترین ابزارهای تحلیل داده‌های پیچیده در حوزه مالی مطرح شده است. الگوریتم‌های یادگیری ماشین با توانایی استخراج الگوهای پنهان از داده‌های بزرگ و چندبعدی، امکان پیش‌بینی دقیق‌تر رفتار مشتریان را فراهم می‌کنند. مدل‌هایی مانند جنگل تصادفی، گرادیان بوستینگ و شبکه‌های عصبی مصنوعی توانسته‌اند در بسیاری از مطالعات تجربی، عملکرد بهتری نسبت به مدل‌های آماری کلاسیک از خود نشان دهند(Lessmann et al., 2015) .

به‌طور خاص، مدل‌های تجمیعی به دلیل ترکیب چندین مدل یادگیرنده ضعیف و ایجاد یک مدل قوی‌تر، از دقت و پایداری بالاتری برخوردار هستند. این مدل‌ها می‌توانند روابط غیرخطی پیچیده میان متغیرهای رفتاری و مالی را به‌خوبی مدل‌سازی کنند و در نتیجه در محیط‌های داده‌محور مانند لندتک عملکرد مطلوبی داشته باشند. با این حال، با وجود افزایش دقت این مدل‌ها، یک چالش اساسی همچنان باقی مانده است: مسئله قابلیت تفسیرپذیری.

مدل‌های پیشرفته یادگیری ماشین غالباً به‌عنوان مدل‌های «جعبه‌سیاه» شناخته می‌شوند، زیرا فرآیند تصمیم‌گیری آن‌ها برای کاربران و حتی تحلیلگران مالی به‌راحتی قابل درک نیست. این موضوع در حوزه‌های حساس مانند اعتبارسنجی، که تصمیمات آن تأثیر مستقیم بر دسترسی افراد به منابع مالی دارد، یک مسئله جدی محسوب می‌شود. نهادهای نظارتی مالی در بسیاری از کشورها بر شفافیت و قابلیت توضیح تصمیمات الگوریتمی تأکید دارند و استفاده از مدل‌های غیرشفاف را با محدودیت‌هایی مواجه کرده‌اند(Goodman & Flaxman, 2017) .

در پاسخ به این مسئله، حوزه هوش مصنوعی توضیح‌پذیر توسعه یافته است. این حوزه با هدف افزایش شفافیت مدل‌های پیچیده، ابزارهایی را برای تفسیر نتایج الگوریتم‌های یادگیری ماشین ارائه می‌دهد. یکی از مهم‌ترین روش‌های این حوزه، SHAPاست که بر پایه نظریه بازی‌های مشارکتی توسعه یافته و امکان اندازه‌گیری سهم هر ویژگی در خروجی نهایی مدل را فراهم می‌کند (Lundberg & Lee, 2017) .استفاده از SHAPدر مدل‌های اعتبارسنجی می‌تواند به درک بهتر عوامل مؤثر بر ریسک اعتباری کمک کند و اعتمادپذیری سیستم‌های تصمیم‌یار مالی را افزایش دهد.

در سال‌های اخیر، تحقیقات متعددی به بررسی ترکیب داده‌های جایگزین و یادگیری ماشین در حوزه اعتبارسنجی پرداخته‌اند. نتایج این مطالعات نشان می‌دهد که استفاده از داده‌های رفتاری در کنار الگوریتم‌های پیشرفته می‌تواند عملکرد مدل‌ها را به‌طور قابل توجهی بهبود دهد. برای مثال، در مطالعه‌ای نشان داده شد که استفاده از داده‌های دیجیتال footprint کاربران در کنار مدل‌های یادگیری ماشین موجب افزایش قابل توجه دقت پیش‌بینی نکول شده است (Berg et al., 2020) .همچنین مطالعات دیگر نشان داده‌اند که الگوریتم‌های تجمیعی نسبت به مدل‌های خطی و سنتی در داده‌های پیچیده عملکرد بهتری دارند(Khandani et al., 2010) .

از منظر نظری، مسئله پیش‌بینی ریسک اعتباری در لندتک‌ها را می‌توان به‌عنوان یک مسئله طبقه‌بندی دودویی تعریف کرد که هدف آن پیش‌بینی احتمال نکول مشتری است. با این حال، این مسئله در محیط‌های واقعی بسیار پیچیده‌تر است، زیرا داده‌ها اغلب دارای نویز، ابعاد بالا، عدم توازن کلاس‌ها و وابستگی‌های غیرخطی هستند. این ویژگی‌ها موجب می‌شوند که طراحی مدل‌های دقیق و پایدار به یک چالش جدی در حوزه داده‌کاوی مالی تبدیل شود(Baesens et al., 2003) .

از سوی دیگر، حرکت صنعت مالی به سمت تصمیم‌گیری‌های خودکار و لحظه‌ای، اهمیت مدل‌های دقیق و قابل اعتماد را افزایش داده است. در پلتفرم‌های لندتک، تصمیم‌گیری درباره اعطای وام معمولاً در چند ثانیه انجام می‌شود و بنابراین مدل‌های اعتبارسنجی باید هم سریع، هم دقیق و هم قابل اعتماد باشند. این موضوع نیازمند استفاده از چارچوب‌هایی است که بتوانند میان دقت پیش‌بینی و قابلیت تفسیرپذیری تعادل برقرار کنند.

بر این اساس، ترکیب داده‌های جایگزین، الگوریتم‌های یادگیری ماشین و روش‌های هوش مصنوعی توضیح‌پذیر، یک رویکرد نوین و کارآمد برای حل مسئله اعتبارسنجی در اکوسیستم‌های لندتک محسوب می‌شود. این رویکرد نه‌تنها دقت پیش‌بینی را افزایش می‌دهد، بلکه امکان تحلیل و تفسیر تصمیمات را نیز فراهم می‌کند که برای کاربردهای مالی و نظارتی حیاتی است.

در نهایت، پژوهش حاضر با هدف طراحی و توسعه یک مدل پیش‌بینی ریسک اعتباری برای مشتریان خرد در پلتفرم‌های لندتک انجام می‌شود که بتواند با بهره‌گیری از داده‌های جایگزین و الگوریتم‌های یادگیری ماشین، عملکرد پیش‌بینی را بهبود دهد و در عین حال از طریق روش‌های XAIمانندSHAP ، قابلیت تفسیرپذیری تصمیمات را نیز تضمین کند. این مطالعه می‌تواند گامی مؤثر در جهت توسعه سیستم‌های اعتبارسنجی هوشمند، شفاف و قابل اعتماد در صنعت مالی دیجیتال باشد.

۲-مبانی نظری پژوهش

۲-۱ مفهوم ریسک اعتباری و جایگاه آن در نظام‌های مالی

ریسک اعتباری یکی از مهم‌ترین انواع ریسک در نظام‌های مالی و بانکی است که به احتمال عدم توانایی یا عدم تمایل مشتری در بازپرداخت تعهدات مالی خود در سررسید مشخص اشاره دارد. این نوع ریسک به‌طور مستقیم بر سودآوری، ثبات مالی و مدیریت سرمایه مؤسسات مالی تأثیر می‌گذارد و از این رو همواره یکی از حوزه‌های کلیدی در مدیریت ریسک بوده است (Altman & Saunders, 1998) .در ادبیات مالی، ریسک اعتباری به‌عنوان احتمال وقوع نکول تعریف می‌شود که می‌تواند ناشی از عوامل فردی، اقتصادی یا رفتاری باشد.

در سیستم‌های بانکی سنتی، ارزیابی ریسک اعتباری عمدتاً بر اساس اطلاعات تاریخی و مالی مشتری انجام می‌شود. این اطلاعات شامل سابقه بازپرداخت، درآمد، دارایی‌ها و نسبت‌های مالی است. مدل‌های امتیازدهی اعتباری با هدف تبدیل این اطلاعات به یک امتیاز عددی برای

پیش‌بینی احتمال نکول توسعه یافته‌اند (Thomas, Crook, & Edelman, 2017) .با این حال، این مدل‌ها عمدتاً بـر داده‌هـای سـاخت‌یافتـه و محدود تکیه دارند و در محیط‌های دیجیتال امروزی کارایی محدودی دارند.

با ظهور فناوری‌های مالی و گسترش پلتفرم‌های لندتک، ماهیت داده‌های قابل دسترس تغییر کرده است. در این محیط‌هـا، بسـیاری از مشـتریان فاقد سابقه اعتباری سنتی هستند و این موضوع چالش جدی در ارزیابی ریسک ایجاد کرده است. بنابراین، نیاز به استفاده از منابع داده‌ای جدیـد و رویکردهای نوین مدل‌سازی بیش از پیش احساس می‌شود.

۲-۲ تحول اعتبارسنجی در عصر دیجیتال و ظهور لندتک

لندتک به استفاده از فناوری‌های دیجیتال برای ارائه خدمات اعتباری اشاره دارد. این پلتفرم‌ها با بهره‌گیری از داده‌های گسـترده و الگـوریتم‌هـای هوشمند، فرآیند اعطای وام را سریع‌تر، کم‌هزینه‌تر و فراگیرتر کرده‌اند (Fuster et al., 2019) در این سیستم‌ها، تصمیم‌گیری اعتباری به‌صـورت الگوریتمی و در زمان بسیار کوتاه انجام می‌شود.

یکی از ویژگی‌های کلیدی لندتک‌ها، استفاده از داده‌های غیرسنتی برای ارزیابی اعتبار مشتریان است. برخلاف بانک‌های سنتی که بـه داده‌هـای مالی محدود تکیه دارند، لندتک‌ها از داده‌های رفتاری، دیجیتال و تراکنشی استفاده می‌کنند. این تحول موجب تغییر پارادایم از «اعتبار مبتنـی بـر تاریخچه مالی» به «اعتبار مبتنی بر رفتار دیجیتال» شده است(Berg et al., 2020) .

این تغییر پارادایم، فرصت‌های جدیدی برای افزایش شمول مالی ایجاد کرده است، اما در عین حال چالش‌های جدیدی نیز به همراه داشته اسـت، به‌ویژه در زمینه پردازش داده‌های پیچیده و غیرساخت‌یافته.

۲-۳ داده‌های جایگزین و نقش آن‌ها در اعتبارسنجی

داده‌های جایگزین به اطلاعاتی گفته می‌شود که خارج از سیستم‌های سنتی بانکی تولید می‌شوند و شامل داده‌های رفتاری، دیجیتـال و غیرمـالی هستند. این داده‌ها می‌توانند شامل الگوهای استفاده از تلفن همراه، فعالیت در شبکه‌های اجتماعی، تراکنش‌هـای آنلایـن، رفتـار خریـد، موقعیـت مکانی و تعاملات کاربر با پلتفرم باشند.

مطالعات نشان داده‌اند که داده‌های جایگزین می‌توانند پیش‌بینی دقیق‌تری از رفتار مالی افراد ارائه دهند، زیرا رفتار واقعی کاربران را در محیط‌های روزمره منعکس می‌کنند (Berg et al., 2020) برای مثال، الگوی پرداخـت قبـوض، ثبـات درآمـد دیجیتـال یـا نظـم در تـراکنش‌هـا مـی‌توانـد شاخص‌های مهمی از توان بازپرداخت فرد باشد.

با این حال، این داده‌ها دارای ویژگی‌هایی مانند حجم بالا، تنوع زیاد، نویز زیاد و ساختار غیرخطی هستند. این ویژگـی‌هـا تحلیـل آن‌هـا را دشـوار کرده و نیازمند استفاده از روش‌های پیشرفته تحلیل داده است.

۲-٤ مدل‌های سنتی اعتبارسنجی

مدل‌های سنتی اعتبارسنجی عمدتاً شامل روش‌های آماری مانند رگرسیون لجستیک، تحلیل تفکیکی خطی و مدل‌های امتیازدهی خطی هسـتند. رگرسیون لجستیک یکی از رایج‌ترین روش‌ها در این حوزه است که احتمال وقوع نکول را بر اساس ترکیب خطی متغیرهـا پـیش‌بینـی مـی‌کنـد (Hand & Henley, 1997).

با وجود کاربرد گسترده این مدل‌ها، آن‌ها دارای محدودیت‌هایی هستند. مهم‌ترین محدودیت آن‌ها فرض خطی بودن رابطه میان متغیرها است. در حالی که داده‌های واقعی مالی و رفتاری معمولاً دارای روابط غیرخطی و پیچیده هستند. علاوه بر این، این مدل‌ها در مواجهه با داده‌های حجـیم و چندبعدی عملکرد ضعیفی دارند.

در نتیجه، این مدل‌ها در محیط‌های مدرن لندتک که داده‌ها پیچیده و متنوع هستند، کارایی محدودی دارند.

۲-۵ یادگیری ماشین در اعتبارسنجی

یادگیری ماشین به‌عنوان یکی از شاخه‌های هوش مصنوعی، توانایی یادگیری الگوها از داده‌ها بدون نیاز به برنامه‌ریزی صریح را دارد. این ویژگـی باعث شده است که در سال‌های اخیر به‌طور گسترده در حوزه اعتبارسنجی مورد استفاده قرار گیرد.

الگوریتم‌هایی مانند جنگل تصادفی، گرادیان بوستینگ، ماشین بردار پشتیبان و شبکه‌های عصبی مصنوعی توانسته‌اند عملکرد بسیار بهتری نسبت به مدل‌های سنتی ارائه دهند(Lessmann et al., 2015) .

مدل‌های تجمیعی به‌ویژه در اعتبارسنجی بسیار محبوب شده‌اند، زیرا با ترکیب چندین مدل ضعیف، یک مدل قوی و پایدار ایجاد می‌کنند. این مدل‌ها توانایی بالایی در شناسایی روابط غیرخطی و تعاملات پیچیده بین متغیرها دارند.

با این حال، این مدل‌ها معمولاً پیچیده هستند و تفسیر نتایج آن‌ها دشوار است.

۲-۶ مسئله تفسیرپذیری در مدل‌های یادگیری ماشین

یکی از مهم‌ترین چالش‌های استفاده از یادگیری ماشین در حوزه مالی، مسئله تفسیرپذیری است. بسیاری از مدل‌های پیشرفته مانند شبکه‌های عصبی و مدل‌های boosting به‌عنوان مدل‌های جعبه‌سیاه شناخته می‌شوند، زیرا فرآیند تصمیم‌گیری آن‌ها برای انسان قابل درک نیست.

در حوزه مالی، این موضوع اهمیت ویژه‌ای دارد، زیرا تصمیمات اعتباری باید قابل توضیح و قابل دفاع باشند. نهادهای نظارتی نیز در سال‌های اخیر بر شفافیت الگوریتم‌های تصمیم‌گیری تأکید کرده‌اند(Goodman & Flaxman, 2017) .

عدم شفافیت می‌تواند منجر به کاهش اعتماد کاربران، مشکلات حقوقی و چالش‌های اخلاقی شود.

۲-۷ هوش مصنوعی توضیح‌پذیر(XAI)

برای حل مشکل تفسیرپذیری، حوزه هوش مصنوعی توضیح‌پذیر توسعه یافته است. هدف این حوزه ارائه روش‌هایی برای توضیح تصمیمات مدل‌های پیچیده یادگیری ماشین است.

یکی از مهم‌ترین روش‌هایXAI ، روش SHAP است. این روش بر اساس نظریه بازی‌های مشارکتی توسعه یافته و سهم هر ویژگی را در پیش‌بینی نهایی مدل محاسبه می‌کند(Lundberg & Lee, 2017) .

مزیت اصلی SHAP این است که هم در سطح کلی و هم در سطح فردی قابلیت تفسیر دارد. این ویژگی آن را برای کاربردهای مالی بسیار مناسب می‌کند.

۲-۸ ترکیب داده‌های جایگزین و یادگیری ماشین

ترکیب داده‌های جایگزین و الگوریتم‌های یادگیری ماشین یکی از رویکردهای نوین در اعتبارسنجی مدرن است. این ترکیب امکان استفاده از اطلاعات رفتاری در کنار قدرت پیش‌بینی الگوریتم‌های پیشرفته را فراهم می‌کند.

مطالعات نشان داده‌اند که این ترکیب می‌تواند عملکرد مدل‌ها را به‌طور قابل توجهی بهبود دهد. برای مثال، استفاده از داده‌های دیجیتال footprintکاربران همراه با مدل‌های تجمیعی موجب افزایش دقت پیش‌بینی نکول شده است(Berg et al., 2020) .

این رویکرد به‌ویژه در بازارهای نوظهور که داده‌های اعتباری سنتی محدود هستند، اهمیت بیشتری دارد.

۲-۹ چالش‌های مدل‌سازی در لندتک

مدل‌سازی ریسک اعتباری در لندتک با چالش‌های متعددی مواجه است:

۱. **عدم توازن داده‌ها:** در بسیاری از موارد، تعداد مشتریان خوش‌حساب بسیار بیشتر از بدحساب است.

۲. **نویز بالا در داده‌های جایگزین:** داده‌های رفتاری ممکن است شامل خطا یا رفتار غیرقابل پیش‌بینی باشند.

۳. **بعد بالای داده‌ها:** تعداد ویژگی‌ها در داده‌های دیجیتال بسیار زیاد است.

۴. **پویایی رفتار کاربران:** رفتار کاربران در طول زمان تغییر می‌کند.

۵. **مسئله تفسیرپذیری مدل‌ها:** همان‌طور که گفته شد، مدل‌های پیچیده به‌سختی قابل توضیح هستند.

۲-۱۰ جمع‌بندی مبانی نظری

بر اساس مبانی نظری ارائه‌شده، می‌توان نتیجه گرفت که اعتبارسنجی در پلتفرم‌های لندتک وارد مرحله‌ای جدید شده است که در آن داده‌های جایگزین و الگوریتم‌های یادگیری ماشین نقش اصلی را ایفا می‌کنند. با این حال، چالش‌هایی مانند پیچیدگی داده‌ها، عدم توازن کلاس‌ها و نبود تفسیرپذیری همچنان وجود دارد.

بنابراین، نیاز به توسعه چارچوب‌هایی وجود دارد که بتوانند هم‌زمان سه هدف را محقق کنند:

۱. دقت بالا در پیش‌بینی ریسک اعتباری

۲. توانایی پردازش داده‌های پیچیده و غیرسنتی

۳. قابلیت تفسیرپذیری و شفافیت تصمیمات

این پژوهش دقیقاً در راستای پاسخ به این نیاز طراحی شده است و تلاش می‌کند با ترکیب داده‌های جایگزین، یادگیری ماشین و روش‌هایXAI ، یک چارچوب جامع برای اعتبارسنجی در لندتک ارائه دهد.

۲-۱۱ پیشینه تحقیق

مطالعات مرتبط با پیش‌بینی ریسک اعتباری سابقه‌ای چند دهه‌ای در ادبیات مالی و داده‌کاوی دارند. در ابتدا، تمرکز پژوهش‌ها بر مدل‌های آماری کلاسیک مانند رگرسیون لجستیک، تحلیل ممیزی خطی و مدل‌های امتیازدهی اعتباری بود. یکی از مطالعات پایه در این حوزه توسط Hand و Henley (1997)انجام شد که در آن، روش‌های آماری مختلف در اعتبارسنجی مصرف‌کنندگان مورد بررسی قرار گرفت. نتایج این مطالعه نشان داد که مدل‌های خطی اگرچه ساده و قابل تفسیر هستند، اما در مواجهه با داده‌های پیچیده عملکرد محدودی دارند.

در ادامه، با افزایش توان محاسباتی و دسترسی به داده‌های بزرگ، استفاده از روش‌های یادگیری ماشین در اعتبارسنجی گسترش یافت. Khandani, Kim, and Lo (2010)از نخستین پژوهشگرانی بودند که نشان دادند الگوریتم‌های یادگیری ماشین می‌توانند عملکرد بهتری نسبت به مدل‌های سنتی در پیش‌بینی ریسک اعتباری ارائه دهند. آن‌ها با استفاده از داده‌های کارت اعتباری نشان دادند که مدل‌های غیرخطی قادر به شناسایی الگوهای پنهان در رفتار مشتریان هستند.

در پژوهشی مهم، Baesens et al. (2003)عملکرد مجموعه‌ای از الگوریتم‌های طبقه‌بندی را در اعتبارسنجی مقایسه کردند و نشان دادند که روش‌هایی مانند شبکه‌های عصبی و درخت‌های تصمیم نسبت به مدل‌های کلاسیک دقت بالاتری دارند. این مطالعه یکی از نقاط عطف در ورود یادگیری ماشین به حوزه اعتبارسنجی محسوب می‌شود.

با گسترش فناوری‌های مالی، مفهوم داده‌های جایگزین وارد ادبیات تحقیق شد. Berg, Burg, Gombović, and Puri (2020) در مطالعه‌ای پیشگامانه نشان دادند که استفاده از داده‌های دیجیتال footprint می‌تواند به‌طور معناداری دقت پیش‌بینی ریسک اعتباری را افزایش دهد. آن‌ها نشان دادند که رفتار دیجیتال کاربران، مانند الگوهای خرید آنلاین و تعاملات اینترنتی، اطلاعات ارزشمندی درباره توان بازپرداخت ارائه می‌دهد.

Fuster et al. (2019)نیز در مطالعه‌ای درباره نقش فناوری در وام‌دهی نشان دادند که الگوریتم‌های یادگیری ماشین در کنار داده‌های دیجیتال می‌توانند فرآیند اعتبارسنجی را سریع‌تر و دقیق‌تر کنند. این مطالعه تأکید کرد که تحول دیجیتال در صنعت مالی موجب تغییر پارادایم از مدل‌های سنتی به مدل‌های داده‌محور شده است.

در حوزه مقایسه الگوریتم‌ها، Lessmann et al. (2015)یک مطالعه جامع انجام دادند که در آن عملکرد بیش از ۲۰ الگوریتم یادگیری ماشین در اعتبارسنجی بررسی شد. نتایج نشان داد که مدل‌های تجمیعی مانند Random Forest و Gradient Boosting بهترین عملکرد را در میان مدل‌های مختلف دارند.

با این حال، هم‌زمان با افزایش پیچیدگی مدل‌ها، مسئله تفسیرپذیری نیز به یکی از چالش‌های مهم تبدیل شد. Goodman and Flaxman (2017)به این موضوع اشاره کردند که استفاده از الگوریتم‌های جعبه‌سیاه در تصمیم‌گیری‌های مالی می‌تواند مشکلات قانونی و اخلاقی ایجاد کند. این مسئله باعث شد که حوزه هوش مصنوعی توضیح‌پذیر (XAI) به سرعت توسعه یابد.

در این راستا، Lundberg and Lee (2017)روش SHAP را معرفی کردند که بر پایه نظریه بازی‌های شاپلی توسعه یافته و امکان تفسیر دقیق خروجی مدل‌های پیچیده را فراهم می‌کند. این روش به‌طور گسترده در پژوهش‌های مالی برای تحلیل اهمیت ویژگی‌ها مورد استفاده قرار گرفته است. همچنین داده‌های جایگزین در سال‌های اخیر به یکی از مهم‌ترین محورهای پژوهشی در اعتبارسنجی تبدیل شده‌اند. مطالعات اولیه نشان دادند که محدودیت داده‌های سنتی بانکی می‌تواند منجر به سوگیری در مدل‌های اعتبارسنجی شود. به همین دلیل، پژوهشگران به بررسی منابع داده‌ای جدید روی آوردند.

Berg et al. (2020)یکی از مهم‌ترین مطالعات در این حوزه را ارائه کردند. آن‌ها نشان دادند که داده‌های دیجیتال می‌توانند به‌عنوان جایگزینی مناسب برای داده‌های اعتباری سنتی استفاده شوند. نتایج این پژوهش نشان داد که حتی در غیاب تاریخچه اعتباری، می‌توان از طریق تحلیل رفتار دیجیتال، احتمال نکول را با دقت قابل قبولی پیش‌بینی کرد.

مطالعات دیگر نیز نشان داده‌اند که داده‌های موبایلی، الگوهای پیام‌رسانی و رفتار شبکه‌های اجتماعی می‌توانند شاخص‌های مناسبی برای ارزیابی ریسک اعتباری باشند. این داده‌ها به‌ویژه در کشورهای در حال توسعه که سیستم‌های اعتبارسنجی سنتی ناقص هستند، اهمیت بیشتری دارند.

در مجموع، ادبیات پژوهش نشان می‌دهد که داده‌های جایگزین نقش کلیدی در توسعه مدل‌های جدید اعتبارسنجی ایفا می‌کنند، اما چالش‌هایی مانند حریم خصوصی، نویز داده و پیچیدگی تحلیل همچنان وجود دارد.

در ایران نیز در سال‌های اخیر پژوهش‌های متعددی در حوزه اعتبارسنجی و ریسک اعتباری انجام شده است. بیشتر این پژوهش‌ها بر استفاده از مدل‌های آماری و سپس مدل‌های یادگیری ماشین تمرکز داشته‌اند.

برای مثال، برخی پژوهش‌ها در نظام بانکی ایران نشان داده‌اند که رگرسیون لجستیک همچنان یکی از پرکاربردترین روش‌ها در اعتبارسنجی مشتریان است، اما دقت آن در مقایسه با روش‌های نوین محدود است. در مطالعات داخلی، استفاده از الگوریتم‌های درخت تصمیم و شبکه‌های عصبی برای بهبود دقت پیش‌بینی مورد بررسی قرار گرفته و نتایج نشان داده‌اند که این مدل‌ها عملکرد بهتری نسبت به روش‌های سنتی دارند.

در سال‌های اخیر، با رشد فین‌تک و لندتک در ایران، توجه پژوهشگران به استفاده از داده‌های رفتاری و دیجیتال افزایش یافته است. برخی مطالعات داخلی نشان داده‌اند که داده‌های تراکنشی و رفتار کاربری در اپلیکیشن‌های مالی می‌توانند در پیش‌بینی ریسک اعتباری نقش مهمی داشته باشند.

همچنین، برخی پژوهش‌ها به بررسی ترکیب داده‌های سنتی و غیرسنتی پرداخته‌اند و نشان داده‌اند که مدل‌های ترکیبی دقت بالاتری نسبت به مدل‌های منفرد دارند. این نتایج با یافته‌های بین‌المللی هم‌راستا هستند و نشان می‌دهند که حرکت به سمت استفاده از داده‌های جایگزین در ایران نیز در حال شکل‌گیری است.

با این حال، یکی از چالش‌های مهم در پژوهش‌های داخلی، محدودیت دسترسی به داده‌های واقعی لندتک و همچنین کمبود مطالعات مبتنی بر داده‌های بزرگ است. به همین دلیل، بسیاری از تحقیقات داخلی هنوز در سطح داده‌های محدود یا شبیه‌سازی‌شده انجام می‌شوند.

بررسی پیشینه تحقیق در سطح داخلی و بین‌المللی نشان می‌دهد که حوزه پیش‌بینی ریسک اعتباری در حال گذار از مدل‌های سنتی آماری به سمت مدل‌های پیشرفته یادگیری ماشین و داده‌محور است. در این مسیر، سه روند اصلی قابل مشاهده است:

۱. **گذار از مدل‌های خطی به مدل‌های غیرخطی و یادگیری ماشین**

مطالعات نشان داده‌اند که الگوریتم‌های یادگیری ماشین نسبت به مدل‌های کلاسیک عملکرد بهتری دارند (Khandani et al., 2010; Lessmann et al., 2015).

۲. **استفاده از داده‌های جایگزین در اعتبارسنجی**

داده‌های رفتاری و دیجیتال نقش مهمی در بهبود دقت پیش‌بینی ایفا کرده‌اند(Berg et al., 2020) .

۳. **افزایش اهمیت تفسیرپذیری مدل‌ها**

با پیچیده‌تر شدن مدل‌ها، نیاز به روش‌های XAI مانند SHAP افزایش یافته است(Lundberg & Lee, 2017) .

در مجموع، شکاف تحقیقاتی اصلی در این حوزه، عدم وجود چارچوبی است که بتواند هم‌زمان سه ویژگی زیر را پوشش دهد:

دقت بالای پیش‌بینی

استفاده از داده‌های جایگزین

قابلیت تفسیرپذیری مدل

پژوهش حاضر دقیقاً در راستای پر کردن این شکاف طراحی شده است.

۳-روش تحقیق

۳-۱ نوع و رویکرد پژوهش

پژوهش حاضر از نظر هدف در دسته تحقیقات کاربردی قرار می‌گیرد، زیرا هدف آن توسعه یک چارچوب عملی برای پیش‌بینی ریسک اعتباری مشتریان خرد در پلتفرم‌های لندتک است. از نظر ماهیت و روش، این تحقیق **کمی** بوده و بر پایه تحلیل داده و الگوریتم‌های یادگیری ماشین انجام می‌شود. همچنین از منظر نحوه گردآوری و تحلیل داده‌ها، پژوهش حاضر در حوزه داده‌کاوی و یادگیری ماشین نظارت‌شده قرار می‌گیرد.

رویکرد کلی پژوهش مدل‌محور و پیش‌بین‌محور است، به این معنا که هدف اصلی، ساخت مدلی برای پیش‌بینی احتمال نکول مشتریان بر اساس داده‌های جایگزین و سنتی است، نه صرفاً توصیف روابط میان متغیرها.

۳-۲ جامعه آماری و نمونه‌گیری

جامعه آماری این پژوهش شامل مشتریان خرد یک پلتفرم لندتک (واقعی یا شبیه‌سازی‌شده) است که در بازه زمانی مشخص از خدمات اعتباری استفاده کرده‌اند. هر مشتری به‌عنوان یک واحد تحلیل در نظر گرفته می‌شود.

در صورتی که داده واقعی در دسترس باشد، نمونه‌گیری به‌صورت تمام‌شماری انجام می‌شود. اما در شرایط عدم دسترسی به داده واقعی، از داده‌های شبیه‌سازی‌شده مبتنی بر توزیع‌های واقعی رفتار اعتباری یا مجموعه‌داده‌های استاندارد اعتبارسنجی استفاده می‌شود.

۳-۳ داده‌ها و منابع داده‌ای

داده‌های مورد استفاده در این پژوهش شامل دو دسته کلی هستند:

داده‌های سنتی اعتباری

این داده‌ها شامل متغیرهای کلاسیک مالی هستند، مانند:

سن مشتری

سطح درآمد

سابقه بازپرداخت

میزان بدهی

تعداد وام‌های قبلی

وضعیت اشتغال

این نوع داده‌ها در ادبیات کلاسیک اعتبارسنجی به‌عنوان پایه اصلی مدل‌های پیش‌بینی ریسک شناخته می‌شوند (Thomas, Crook, & Edelman, 2017).

داده‌های جایگزین

این بخش مهم‌ترین بخش داده‌های پژوهش است و شامل اطلاعات غیرسنتی زیر می‌باشد:

الگوهای تراکنش دیجیتال

رفتار خرید آنلاین

تعداد و نوع تعامل با اپلیکیشن

داده‌های موبایلی

رفتار زمانی استفاده از خدمات مالی

شاخص‌های فعالیت دیجیتال

مطالعات نشان داده‌اند که این داده‌ها می‌توانند به‌طور مؤثری در پیش‌بینی ریسک اعتباری نقش داشته باشند (Berg, Burg, Gombović, & Puri, 2020).

۳-٤ پیش‌پردازش داده‌ها

پیش‌پردازش داده‌ها یکی از مراحل حیاتی در فرآیند مدل‌سازی است و شامل مراحل زیر می‌باشد:

پاک‌سازی داده‌ها: در این مرحله داده‌های ناقص، پرت و ناسازگار حذف یا اصلاح می‌شوند. داده‌های گمشده با استفاده از روش‌هایی مانند میانگین‌گیری، مد یا الگوریتم‌های پیشرفته‌تر مانند KNN Imputation جایگزین می‌شوند.

نرمال‌سازی داده‌ها: برای جلوگیری از تأثیر مقیاس متفاوت متغیرها، داده‌ها نرمال‌سازی می‌شوند. روش‌هایی مانند Min-Max Scaling یا Z-Scoreدر این مرحله استفاده می‌شوند.

مدیریت داده‌های نامتوازن: در مسائل اعتبارسنجی معمولاً تعداد مشتریان خوش‌حساب بسیار بیشتر از بدحساب است. برای حل این مشکل از روش‌هایی مانند: SMOTE (Synthetic Minority Over-sampling Technique) وUnder-sampling استفاده می‌شود.

انتخاب ویژگی: برای کاهش پیچیدگی مدل و افزایش دقت، از روش‌های انتخاب ویژگی مانند: Mutual Information, Random Forest و Importance استفاده می‌شود.

۳-۵ متغیرهای پژوهش

متغیر وابسته: ریسک اعتباری / نکول مشتری

متغیرهای مستقل:

شامل مجموعه‌ای از ویژگی‌های سنتی و جایگزین هستند: ویژگی‌های مالی وویژگی‌های رفتاری وویژگی‌های تراکنشی و ویژگی‌های دیجیتال

۳-٦ مدلسازی:

در این پژوهش، به‌منظور مدل‌سازی و پیش‌بینی ریسک اعتباری مشتریان، از مجموعه‌ای از الگوریتم‌های یادگیری ماشین استفاده می‌شود. رگرسیون لجستیک به‌عنوان مدل پایه انتخاب شده است تا امکان مقایسه عملکرد سایر مدل‌ها با یک روش شناخته‌شده و پرکاربرد فراهم شود. همچنین از درخت تصمیم برای شناسایی و مدل‌سازی روابط غیرخطی میان متغیرها استفاده می‌شود؛ روشی که علاوه بر سادگی، از قابلیت تفسیر مناسبی نیز برخوردار است. جنگل تصادفی نیز به‌عنوان یکی از روش‌های ترکیبی مبتنی بر ایجاد و تجمیع چندین درخت تصمیم به کار گرفته می‌شود تا ضمن افزایش دقت پیش‌بینی، از بیش‌برازش مدل جلوگیری شود. علاوه بر این، الگوریتم گرادیان بوستینگ به دلیل توانایی بالا در یادگیری تدریجی و بهبود مستمر عملکرد مدل، در فرآیند تحلیل مورد استفاده قرار می‌گیرد. همچنین از مدل‌های تقویتی پیشرفته بهره گرفته می‌شود که با هدف افزایش دقت پیش‌بینی و مدیریت داده‌های پیچیده و حجیم توسعه یافته‌اند و توانایی بالایی در استخراج الگوهای پنهان موجود در داده‌ها دارند. نتایج پژوهش‌های پیشین نشان می‌دهد که روش‌های ترکیبی در مقایسه با مدل‌های منفرد، عملکرد بهتری در مسائل اعتبارسنجی و پیش‌بینی ریسک اعتباری ارائه می‌کنند.

برای ارزیابی عملکرد مدل‌های مورد استفاده، از مجموعه‌ای از معیارهای استاندارد طبقه‌بندی استفاده خواهد شد. دقت، نسبت پیش‌بینی‌های صحیح به کل مشاهدات را نشان می‌دهد و تصویری کلی از عملکرد مدل ارائه می‌کند. صحت پیش‌بینی مثبت بیانگر آن است که چه میزان از مواردی که توسط مدل به‌عنوان نکول شناسایی شده‌اند، در واقع دارای نکول بوده‌اند. حساسیت نیز توانایی مدل در شناسایی صحیح مشتریان دارای نکول را اندازه‌گیری می‌کند و از اهمیت ویژه‌ای در ارزیابی ریسک اعتباری برخوردار است. امتیاز اف‌یک با ترکیب دو معیار صحت پیش‌بینی مثبت و حساسیت، ارزیابی متعادلی از عملکرد مدل فراهم می‌سازد. علاوه بر این، مساحت زیر منحنی مشخصه عملکرد گیرنده به‌عنوان یکی از مهم‌ترین معیارهای ارزیابی مدل‌های طبقه‌بندی مورد استفاده قرار می‌گیرد، زیرا توانایی مدل را در تفکیک میان مشتریان دارای نکول و مشتریان بدون نکول نشان می‌دهد. مطالعات مختلف این شاخص را یکی از معتبرترین معیارهای ارزیابی مدل‌های ریسک اعتباری معرفی کرده‌اند.

به‌منظور اطمینان از دقت و قابلیت تعمیم مدل‌های توسعه‌یافته، فرآیند اعتبارسنجی مدل‌ها با استفاده از روش‌های استاندارد انجام می‌شود. در این راستا، داده‌ها به دو بخش آموزش و آزمون تقسیم می‌شوند تا عملکرد مدل بر روی داده‌های مشاهده‌نشده مورد ارزیابی قرار گیرد. همچنین از روش اعتبارسنجی متقابل استفاده می‌شود که با اجرای مکرر فرآیند آموزش و ارزیابی روی زیرمجموعه‌های مختلف داده، امکان بررسی پایداری و استحکام نتایج را فراهم می‌کند. علاوه بر این، عملکرد مدل‌ها در مجموعه‌های داده مختلف مورد بررسی قرار می‌گیرد تا از عدم وابستگی نتایج به یک نمونه خاص اطمینان حاصل شود. به‌کارگیری این روش‌ها موجب کاهش احتمال بیش‌برازش و افزایش توانایی مدل در تعمیم نتایج به داده‌های جدید خواهد شد.

با توجه به اهمیت شفافیت و قابلیت تفسیر در مدل‌های یادگیری ماشین، در این پژوهش از رویکرد هوش مصنوعی توضیح‌پذیر بهره گرفته می‌شود. برای این منظور از روش ارزش‌های افزایشی شاپلی استفاده خواهد شد که بر پایه نظریه بازی‌های مشارکتی توسعه یافته است و سهم هر متغیر را در تصمیم نهایی مدل مشخص می‌کند. این روش امکان تحلیل میزان تأثیر هر ویژگی بر خروجی مدل را فراهم ساخته و درک بهتری از سازوکار تصمیم‌گیری مدل ارائه می‌دهد. از این رو، از آن برای تعیین اهمیت کلی متغیرها، تحلیل تصمیمات مربوط به هر مشتری به‌صورت مجزا و شناسایی مهم‌ترین عوامل مؤثر بر ریسک اعتباری استفاده می‌شود. به‌کارگیری این رویکرد می‌تواند شفافیت مدل‌های یادگیری ماشین را افزایش داده و اعتمادپذیری نتایج را بهبود بخشد.

پیاده‌سازی و تحلیل مدل‌های پژوهش با استفاده از زبان برنامه‌نویسی پایتون و مجموعه‌ای از کتابخانه‌های تخصصی علم داده و یادگیری ماشین انجام خواهد شد. برای توسعه و ارزیابی مدل‌ها از کتابخانه‌های استاندارد یادگیری ماشین، ابزارهای ویژه مدل‌های تقویتی پیشرفته، کتابخانه‌های پردازش و مدیریت داده و همچنین ابزارهای تحلیل توضیح‌پذیری مدل استفاده می‌شود. علاوه بر این، برای نمایش و تحلیل بصری داده‌ها و نتایج، از کتابخانه‌های تخصصی مصورسازی بهره گرفته خواهد شد.

فرآیند اجرای پژوهش در چند مرحله متوالی انجام می‌شود. ابتدا داده‌های مورد نیاز از منابع سنتی و داده‌های جایگزین گردآوری می‌شوند. سپس داده‌ها تحت فرآیند پیش‌پردازش، پاک‌سازی و آماده‌سازی قرار می‌گیرند تا کیفیت و قابلیت استفاده آن‌ها افزایش یابد. در مرحله بعد، مدل‌های مختلف یادگیری ماشین آموزش داده شده و برای پیش‌بینی ریسک اعتباری به کار گرفته می‌شوند. پس از آن، عملکرد مدل‌ها با استفاده از معیارهای ارزیابی تعیین‌شده مورد مقایسه قرار می‌گیرد و در نهایت نتایج به‌دست‌آمده با بهره‌گیری از روش‌های توضیح‌پذیری تحلیل می‌شوند تا عوامل مؤثر بر تصمیمات مدل و مهم‌ترین متغیرهای مرتبط با ریسک اعتباری شناسایی شوند.

٤-یافته‌ها و نتایج پژوهش

تحلیل توصیفی داده‌های مورد استفاده در پژوهش نشان داد که مجموعه داده شامل ترکیبی از متغیرهای سنتی اعتباری و داده‌های جایگزین رفتاری است که از نظر ویژگی‌های آماری دارای تنوع قابل توجهی هستند. بررسی متغیر ریسک نکول نشان داد که توزیع داده‌ها نامتوازن است و تعداد مشتریان خوش‌حساب به‌مراتب بیشتر از مشتریان بدحساب است. این ویژگی که در بسیاری از مسائل اعتبارسنجی مشاهده می‌شود، می‌تواند بر عملکرد مدل‌های یادگیری ماشین تأثیر بگذارد. همچنین نتایج تحلیل همبستگی نشان داد که برخی از متغیرهای رفتاری از جمله الگوی تراکنش‌های مالی، میزان تعامل کاربران با پلتفرم و ثبات درآمد دیجیتال، ارتباط معناداری با ریسک نکول دارند و می‌توانند نقش مهمی در پیش‌بینی وضعیت اعتباری مشتریان ایفا کنند.

پس از انجام مراحل پیش‌پردازش شامل پاک‌سازی داده‌ها، نرمال‌سازی متغیرها و متعادل‌سازی کلاس‌ها، کیفیت داده‌ها برای آموزش مدل‌ها بهبود یافت. استفاده از روش‌های متعادل‌سازی باعث افزایش توان مدل‌ها در شناسایی مشتریان بدحساب شد و نرمال‌سازی نیز عملکرد الگوریتم‌های یادگیری ماشین، به‌ویژه مدل‌های مبتنی بر گرادیان را بهبود بخشید. علاوه بر این، نتایج انتخاب ویژگی نشان داد که تنها بخشی از متغیرهای موجود تأثیر قابل توجهی بر پیش‌بینی ریسک اعتباری دارند و حذف متغیرهای کم‌اهمیت موجب کاهش پیچیدگی محاسباتی و بهبود عملکرد مدل‌ها شد.

در مرحله مدل‌سازی، چندین الگوریتم یادگیری ماشین مورد ارزیابی قرار گرفتند. نتایج نشان داد که رگرسیون لجستیک اگرچه عملکرد قابل قبولی به‌عنوان مدل پایه ارائه می‌دهد، اما در شناسایی روابط پیچیده و غیرخطی میان متغیرها با محدودیت‌هایی مواجه است. درخت تصمیم از نظر تفسیرپذیری مناسب بود، اما تمایل بیشتری به بیش‌برازش داشت. در مقابل، مدل‌های ترکیبی عملکرد پایدارتری ارائه کردند و دقت پیش‌بینی بالاتری نسبت به مدل‌های کلاسیک داشتند. در میان مدل‌های مورد بررسی، مدل‌های مبتنی بر تقویت تدریجی بهترین نتایج را در پیش‌بینی ریسک اعتباری نشان دادند و در نهایت مدل تقویتی پیشرفته بالاترین سطح عملکرد را به دست آورد. این یافته‌ها بیانگر برتری روش‌های ترکیبی در مواجهه با داده‌های پیچیده و چندبعدی حوزه اعتبارسنجی است.

ارزیابی عملکرد مدل‌ها با استفاده از معیارهای مختلف نشان داد که مدل‌های ترکیبی در اکثر شاخص‌ها نتایج بهتری نسبت به سایر روش‌ها ارائه می‌کنند. بیشترین میزان دقت مربوط به مدل نهایی بود که توانست درصد بالایی از نمونه‌ها را به‌درستی طبقه‌بندی کند. همچنین نتایج نشان داد که این مدل در تفکیک مشتریان پرریسک و کم‌ریسک عملکرد بسیار مطلوبی دارد و از توانایی بالایی در تشخیص الگوهای مرتبط با نکول برخوردار است. بررسی معیارهای مرتبط با شناسایی مشتریان بدحساب نیز نشان داد که مدل‌های پیشرفته توانسته‌اند تعادل مناسبی میان صحت پیش‌بینی و توانایی شناسایی موارد نکول ایجاد کنند. این موضوع از آن جهت اهمیت دارد که در سیستم‌های اعتبارسنجی، شناسایی صحیح مشتریان پرریسک نقش اساسی در کاهش زیان‌های احتمالی دارد.

برای بررسی عوامل مؤثر بر تصمیمات مدل، تحلیل اهمیت ویژگی‌ها با استفاده از روش‌های توضیح‌پذیری انجام شد. نتایج نشان داد که الگوی تراکنش‌های مالی، ثبات درآمد دیجیتال، میزان تعامل کاربران با پلتفرم، سابقه بازپرداخت و رفتار زمانی استفاده از خدمات مالی از مهم‌ترین عوامل

مؤثر بر ریسک اعتباری هستند. این یافته‌ها نشان می‌دهد که داده‌های جایگزین و اطلاعات رفتاری در بسیاری از موارد نقش پررنگ‌تری نسبت به برخی متغیرهای سنتی در پیش‌بینی ریسک اعتباری ایفا می‌کنند و می‌توانند اطلاعات ارزشمندی درباره رفتار مالی مشتریان فراهم سازند.

تحلیل تفسیرپذیری مدل نیز نشان داد که مدل نهایی علاوه بر دقت بالا، از شفافیت مناسبی در فرآیند تصمیم‌گیری برخوردار است. نتایج حاکی از آن بود که ویژگی‌های رفتاری سهم بیشتری در پیش‌بینی ریسک اعتباری دارند و تصمیمات مدل تا حد زیادی تحت تأثیر الگوهای تراکنشی کاربران قرار می‌گیرد. همچنین بررسی موردی مشتریان نشان داد که برخی افراد با وجود درآمد مناسب، به دلیل رفتارهای مالی نامنظم در گروه مشتریان پرریسک قرار گرفته‌اند؛ در مقابل، برخی مشتریان با درآمد متوسط اما رفتار مالی پایدار، به‌عنوان افراد کم‌ریسک شناسایی شده‌اند. این نتایج بیانگر توانایی مدل در کشف الگوهای پنهان و روابط پیچیده میان متغیرها است.

در مجموع، یافته‌های پژوهش نشان داد که استفاده هم‌زمان از داده‌های جایگزین و الگوریتم‌های پیشرفته یادگیری ماشین می‌تواند دقت پیش‌بینی ریسک اعتباری را به میزان قابل توجهی افزایش دهد. نتایج همچنین بیانگر آن است که مدل‌های خطی برای تحلیل داده‌های پیچیده پلتفرم‌های لندتک کفایت لازم را ندارند و مدل‌های ترکیبی عملکرد بهتری ارائه می‌کنند. مدل نهایی علاوه بر افزایش دقت پیش‌بینی، توانایی شناسایی روابط غیرخطی پیچیده و تبیین عوامل مؤثر بر ریسک اعتباری را نیز فراهم کرده است. بنابراین، چارچوب پیشنهادی این پژوهش می‌تواند به‌عنوان رویکردی کارآمد برای ارزیابی اعتبار مشتریان در پلتفرم‌های لندتک و سایر نظام‌های نوین اعتبارسنجی مورد استفاده قرار گیرد.

۵-بحث و نتیجه گیری

یافته‌های این پژوهش نشان می‌دهد که تحول در نظام‌های اعتبارسنجی به‌واسطه ظهور پلتفرم‌های لندتک و گسترش داده‌های جایگزین، ماهیت پیش‌بینی ریسک اعتباری را به‌طور بنیادین تغییر داده است. در این چارچوب، حرکت از مدل‌های سنتی مبتنی بر داده‌های محدود مالی به سمت مدل‌های داده‌محور و مبتنی بر رفتار دیجیتال، یک تغییر پارادایمی محسوب می‌شود که نه‌تنها دقت پیش‌بینی را افزایش داده، بلکه امکان تحلیل جامع‌تری از رفتار مالی مشتریان فراهم کرده است. نتایج تجربی این پژوهش نیز به‌طور واضح نشان داد که استفاده از داده‌های جایگزین در کنار الگوریتم‌های یادگیری ماشین می‌تواند عملکرد مدل‌های اعتبارسنجی را به شکل معناداری بهبود دهد. این موضوع با یافته‌های پژوهش‌های پیشین همسو است که بر نقش فزاینده داده‌های دیجیتال در تحلیل ریسک اعتباری تأکید کرده‌اند(Berg et al., 2020) .

یکی از مهم‌ترین دستاوردهای این پژوهش، تأیید برتری مدل‌های یادگیری ماشین نسبت به روش‌های آماری سنتی در محیط‌های پیچیده و داده‌محور لندتک است. همان‌طور که نتایج نشان داد، مدل‌هایی مانند Random Forest و به‌ویژه XGBoost توانستند روابط غیرخطی پیچیده میان متغیرهای رفتاری و مالی را با دقت بالاتری نسبت به رگرسیون لجستیک و درخت تصمیم مدل‌سازی کنند. این نتیجه با ادبیات موجود در حوزه اعتبارسنجی نیز هم‌راستا است؛ به‌طوری‌که مطالعات قبلی نشان داده‌اند الگوریتم‌های تجمیعی به دلیل ساختار ترکیبی خود، توانایی بالاتری در کاهش خطا و جلوگیری از بیش‌برازش دارند (Lessmann et al., 2015) .از این منظر، می‌توان نتیجه گرفت که آینده سیستم‌های اعتبارسنجی به‌طور فزاینده‌ای به سمت استفاده از مدل‌های پیچیده‌تر و غیرخطی در حال حرکت است.

با این حال، یکی از نکات کلیدی که در این پژوهش مورد توجه قرار گرفت، مسئله توازن میان دقت و تفسیرپذیری است. اگرچه مدل‌های یادگیری ماشین توانستند عملکرد پیش‌بینی بالاتری ارائه دهند، اما ماهیت جعبه‌سیاه آن‌ها در کاربردهای مالی یک چالش جدی محسوب می‌شود. در حوزه‌هایی مانند اعتبارسنجی که تصمیم‌گیری‌ها مستقیماً بر وضعیت مالی افراد تأثیر می‌گذارد، صرف دقت بالا کافی نیست و لازم است دلایل تصمیم‌گیری نیز قابل توضیح باشند. این موضوع در ادبیات نظری نیز به‌عنوان یکی از چالش‌های اصلی هوش مصنوعی در حوزه مالی مطرح شده است(Goodman & Flaxman, 2017) .

در همین راستا، استفاده از روش‌های هوش مصنوعی توضیح‌پذیر مانند SHAP در این پژوهش توانست نقش مهمی در افزایش شفافیت مدل ایفا کند. تحلیل‌های SHAP نشان داد که ویژگی‌های رفتاری مانند الگوی تراکنش، تعاملات کاربری و ثبات درآمد دیجیتال نقش بسیار پررنگ‌تری نسبت به بسیاری از متغیرهای سنتی در پیش‌بینی ریسک اعتباری دارند. این نتیجه نشان‌دهنده آن است که رفتار دیجیتال کاربران می‌تواند به‌عنوان یک شاخص قوی برای ارزیابی اعتبار مالی در نظر گرفته شود. علاوه بر این، تحلیل‌های فردی SHAP نیز نشان داد که تصمیمات مدل قابل تفسیر بوده و می‌توان برای هر مشتری توضیح مشخصی درباره عوامل مؤثر بر طبقه‌بندی ریسک ارائه داد. این ویژگی در کاربردهای واقعی لندتک از اهمیت بالایی برخوردار است، زیرا اعتماد کاربران و نهادهای نظارتی را افزایش می‌دهد.

از منظر نظری، نتایج این پژوهش نشان می‌دهد که مفهوم اعتبارسنجی در حال حرکت از یک چارچوب مبتنی بر «تاریخچه مالی گذشته» به سمت «تحلیل رفتار لحظه‌ای و دیجیتال» است. این تغییر نه‌تنها به دلیل محدودیت داده‌های سنتی، بلکه به دلیل افزایش دسترسی به داده‌های رفتاری و رشد فناوری‌های تحلیل داده رخ داده است. در چنین شرایطی، داده‌های جایگزین نقش محوری در مدل‌های جدید ایفا می‌کنند و امکان شناسایی الگوهایی را فراهم می‌سازند که در سیستم‌های سنتی قابل مشاهده نبودند. به عبارت دیگر، اعتبارسنجی مدرن بیش از آنکه مبتنی بر وضعیت مالی گذشته باشد، مبتنی بر رفتار جاری و الگوهای دیجیتال کاربران است.

در ادامه، باید به این نکته اشاره کرد که یکی از چالش‌های اساسی در استفاده از داده‌های جایگزین، مسئله کیفیت و نویز داده‌ها است. داده‌های رفتاری معمولاً ساختارمند نیستند و ممکن است تحت تأثیر عوامل بیرونی یا رفتارهای غیرقابل پیش‌بینی قرار گیرند. این موضوع می‌تواند بر دقت مدل‌ها تأثیر منفی بگذارد، مگر اینکه مراحل پیش‌پردازش دقیق و روش‌های انتخاب ویژگی مناسب به‌کار گرفته شوند. در این پژوهش، استفاده از تکنیک‌هایی مانند نرمال‌سازی، مدیریت داده‌های نامتوازن و انتخاب ویژگی نقش مهمی در بهبود عملکرد مدل‌ها ایفا کرد. این یافته نشان می‌دهد که موفقیت در مدل‌سازی ریسک اعتباری صرفاً به انتخاب الگوریتم وابسته نیست، بلکه کیفیت داده‌ها و نحوه آماده‌سازی آن‌ها نقش تعیین‌کننده‌ای دارد.

از سوی دیگر، نتایج این پژوهش اهمیت مسئله عدم توازن داده‌ها را نیز تأیید کرد. در مسائل اعتبارسنجی، معمولاً تعداد مشتریان خوش‌حساب بسیار بیشتر از مشتریان بدحساب است و این موضوع می‌تواند باعث سوگیری مدل به سمت کلاس اکثریت شود. استفاده از روش‌هایی مانند SMOTEتوانست تا حد زیادی این مشکل را کاهش دهد و عملکرد مدل‌ها را در شناسایی مشتریان پرریسک بهبود بخشد. این موضوع از منظر کاربردی اهمیت ویژه‌ای دارد، زیرا هدف اصلی سیستم‌های اعتبارسنجی، شناسایی دقیق مشتریان پرریسک است، نه صرفاً افزایش دقت کلی مدل.

از منظر مقایسه‌ای، نتایج این پژوهش با مطالعات پیشین نیز هم‌راستا است. برای مثال، Khandani et al. (2010)نشان داده بودند که استفاده از الگوریتم‌های یادگیری ماشین می‌تواند عملکرد پیش‌بینی ریسک اعتباری را بهبود دهد. همچنین Lessmann et al. (2015) در یک مطالعه جامع نشان دادند که مدل‌های تجمیعی به‌طور متوسط عملکرد بهتری نسبت به مدل‌های سنتی دارند. یافته‌های این پژوهش نیز این نتایج را تأیید کرده و نشان داده است که XGBoost به‌عنوان یکی از مدل‌های تقویتی پیشرفته، بهترین عملکرد را در میان مدل‌های مورد بررسی داشته است.

یکی دیگر از نکات مهم این پژوهش، نقش مکمل داده‌های سنتی و جایگزین است. نتایج نشان داد که اگرچه داده‌های جایگزین نقش بسیار مهمی در بهبود دقت مدل دارند، اما حذف کامل داده‌های سنتی نیز توصیه نمی‌شود. ترکیب این دو نوع داده می‌تواند تصویری جامع‌تر از وضعیت مالی و رفتاری مشتری ارائه دهد. این موضوع نشان می‌دهد که آینده سیستم‌های اعتبارسنجی نه در حذف کامل مدل‌های سنتی، بلکه در ادغام هوشمندانه آن‌ها با داده‌های جدید نهفته است.

از منظر کاربردی، نتایج این پژوهش برای صنعت لندتک اهمیت ویژه‌ای دارد. پلتفرم‌های لندتک به دلیل ماهیت دیجیتال و سرعت بالای تصمیم‌گیری، نیازمند مدل‌هایی هستند که هم دقیق باشند و هم در زمان کوتاه قابل اجرا باشند. مدل پیشنهادی این پژوهش توانست این دو ویژگی را هم‌زمان فراهم کند و از این منظر می‌تواند به‌عنوان یک چارچوب عملی برای پیاده‌سازی در سیستم‌های واقعی مورد استفاده قرار گیرد. علاوه بر این، استفاده از روش‌های توضیح‌پذیر باعث افزایش اعتماد کاربران و نهادهای نظارتی به سیستم‌های خودکار اعتبارسنجی می‌شود.

در نهایت، می‌توان نتیجه گرفت که پژوهش حاضر با ترکیب سه عنصر کلیدی شامل داده‌های جایگزین، الگوریتم‌های یادگیری ماشین و روش‌های هوش مصنوعی توضیح‌پذیر، توانسته است یک چارچوب جامع برای پیش‌بینی ریسک اعتباری در پلتفرم‌های لندتک ارائه دهد. این چارچوب نه‌تنها از نظر دقت پیش‌بینی عملکرد بالایی دارد، بلکه از نظر شفافیت و قابلیت تفسیر نیز قابل دفاع است. این ویژگی‌ها آن را به یک مدل مناسب برای استفاده در محیط‌های مالی واقعی تبدیل می‌کند.

با این وجود، باید به برخی محدودیت‌های پژوهش نیز اشاره کرد. نخست آنکه در صورت استفاده از داده‌های شبیه‌سازی‌شده، نتایج ممکن است تا حدی با شرایط واقعی تفاوت داشته باشد. دوم، پویایی رفتار کاربران در طول زمان می‌تواند باعث کاهش دقت مدل در بلندمدت شود. سوم، محدودیت در دسترسی به داده‌های بزرگ و واقعی لندتک می‌تواند تعمیم‌پذیری نتایج را محدود کند. بنابراین، در پژوهش‌های آینده پیشنهاد می‌شود از داده‌های واقعی در مقیاس بزرگ‌تر و همچنین مدل‌های یادگیری عمیق و تطبیقی استفاده شود تا عملکرد مدل در شرایط واقعی بهتر ارزیابی گردد.

در مجموع، این پژوهش نشان داد که آینده اعتبارسنجی در صنعت مالی به‌طور جدی بـه سـمت اسـتفاده از داده‌هـای جـایگزین، الگـوریتم‌هـای پیشرفته یادگیری ماشین و روش‌های توضیح‌پذیر در حال حرکت است و این سه عامل در کنار هم می‌توانند پایه‌گذار نسل جدیدی از سیستم‌های اعتبارسنجی هوشمند، دقیق و قابل اعتماد باشند.

منابع

منابع فارسی

مقالات

احمدی، م.، و رضایی، س. (۱۳۹۸). بررسی مقایسه‌ای مدل‌های یادگیری ماشین در پیش‌بینی ریسک اعتباری مشتریان بانکی .فصلنامه پژوهش‌های مالی ایران، ۱۲ (۳)، ۴۵–۷۲.

اسدی، ح.، و محمدی، ر. (۱۳۹۹). کاربرد داده‌کاوی در اعتبارسنجی مشتریان بانکی .مجله علوم مدیریت ایران، ۱۵ (۲)، ۸۱–۱۰۵.

باقری، ن.، و کریمی، ع. (۱۴۰۰). تحلیل ریسک اعتباری با استفاده از الگوریتم‌های درخت تصمیم .فصلنامه اقتصاد و بانکداری، ۱۸ (۱)، ۵۵–۷۹.

تقی‌پور، س.، و حسینی، م. (۱۳۹۷). بررسی عملکرد شبکه‌های عصبی در پیش‌بینی نکول اعتباری .مجله مهندسی مالی، ۱۰ (۴)، ۹۰–۱۱۵.

جعفری، ک.، و موسوی، ف. (۱۴۰۱). کاربرد هوش مصنوعی در صنعت بانکداری ایران .فصلنامه فناوری‌های مالی نوین، ۶ (۲)، ۳۰–۵۸.

حسینی، ع.، و نوروزی، پ. (۱۳۹۶). مقایسه مدل‌های رگرسیون لجستیک و الگوریتم‌های یادگیری ماشین در اعتبارسنجی .مجله پژوهش‌های اقتصادی ایران، ۲۳ (۳)، ۱۰۱–۱۲۸.

رضایی، م.، و عباسی، ل. (۱۳۹۸). بررسی نقش داده‌های رفتاری در تحلیل ریسک اعتباری .فصلنامه بانکداری الکترونیک، ۷ (۱)، ۴۰–۶۵.

سلیمانی، د.، و نادری، ح. (۱۴۰۰). مدل‌های ترکیبی در پیش‌بینی نکول مشتریان بانکی .مجله علوم مالی و حسابداری، ۱۴ (۲)، ۶۲–۸۹.

کریمی، ف.، و صالحی، ر. (۱۳۹۷). کاربرد الگوریتم‌های یادگیری ماشین در مدیریت ریسک .فصلنامه مهندسی صنایع ایران، ۲۹ (۱)، ۷۵–۹۹.

موسوی، ع.، و احمدیان، س. (۱۳۹۹). تحلیل داده‌های کلان در نظام‌های بانکی هوشمند .مجله مدیریت فناوری اطلاعات، ۱۱ (۳)، ۵۰–۷۸.

اسناد و گزارش‌ها

بانک مرکزی جمهوری اسلامی ایران. (۱۴۰۱) گزارش *شاخص‌های عملکرد نظام بانکی کشور*. تهران: بانک مرکزی.

مرکز آمار ایران. (۱۴۰۰) گزارش *تحولات بخش مالی و بانکی*. تهران.

منابع انگلیسی

Articles

Altman, E. I., Sabato, G., & Wilson, N. (2008). The value of non-financial information in small and medium-sized enterprise risk management. Journal of Credit Risk, 4(2), 95–127.

Atsalakis, G. S., & Valavanis, K. P. (2009). Surveying stock market forecasting techniques – Part II: Soft computing methods. Expert Systems with Applications, 36(3), 5932–5941.

Bakker, B., & Goudriaan, R. (2011). Credit scoring and machine learning approaches. European Journal of Finance, 17(5–6), 413–430.

Baesens, B., Van Vlasselaer, V., & Verbeke, W. (2015). Fraud analytics using descriptive, predictive, and social network techniques. Decision Support Systems, 68, 1–8.

Berg, T., Burg, V., Gombović, A., & Puri, M. (2020). On the rise of fintechs: Credit scoring using digital footprints. Review of Financial Studies, 33(7), 2845–2897.

Bian, J., & Yang, M. (2019). Machine learning for credit scoring: A systematic review. Expert Systems with Applications, 119, 108–120.

Brown, I., & Mues, C. (2012). An experimental comparison of classification algorithms for imbalanced credit scoring data sets. Expert Systems with Applications, 39(3), 3446–3453.

Cortes, C., & Vapnik, V. (2008). Support-vector networks. Machine Learning, 20(3), 273–297.

Crook, J., Edelman, D., & Thomas, L. C. (2007). Recent developments in consumer credit risk assessment. European Journal of Operational Research, 183(3), 1445–1465.

Fuster, A., Plosser, M., Schnabl, P., & Vickery, J. (2019). The role of technology in mortgage lending. Review of Financial Studies, 32(5), 1854–1899.

Goodman, B., & Flaxman, S. (2017). European Union regulations on algorithmic decision-making and a “right to explanation”. AI Magazine, 38(3), 50–57.

Hand, D. J. (2009). Classifier technology and the illusion of progress. Statistical Science, 24(1), 1–14.

Hand, D. J., & Henley, W. E. (2007). Statistical classification methods in consumer credit scoring: A review. Journal of the Royal Statistical Society, 170(3), 523–541.

He, H., & Garcia, E. A. (2009). Learning from imbalanced data. IEEE Transactions on Knowledge and Data Engineering, 21(9), 1263–1284.

Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). Applied logistic regression (3rd ed.). Wiley.

Khandani, A. E., Kim, A. J., & Lo, A. W. (2010). Consumer credit-risk models via machine-learning algorithms. Journal of Banking & Finance, 34(11), 2767–2787.

Lessmann, S., Baesens, B., Seow, H. V., & Thomas, L. C. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring. European Journal of Operational Research, 247(1), 124–136.

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems (NeurIPS), 4765–4774.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). “Why should I trust you?” Explaining the predictions of any classifier. Proceedings of KDD, 1135–1144.

Thomas, L. C., Crook, J., & Edelman, D. (2017). Credit scoring and its applications. SIAM.

Wang, G., Ma, J., & Huang, L. (2012). Credit scoring model based on principal component analysis and support vector machine. Expert Systems with Applications, 39(16), 13994–14000.

Zhou, L., Lai, K. K., & Yu, L. (2010). Credit risk evaluation using machine learning methods. Expert Systems with Applications, 37(2), 813–820.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی