

چارچوب تبیین پذیر برای پیش بینی ریسک اعتباری بانکی

زینب طالبی جویباری

دانشجوی کارشناسی ارشد آمار زیستی دانشگاه علوم پزشکی پیامبر اعظم مازندران. بانک ملی شعبه انبارتوشه راه آهن. تهران، ایران.
آدرس پست الکترونیک نویسنده : talebizainab3@gmail.com

جمشید یزدانی چراتی

هیئت علمی و مدیر آموزشی دانشگاه علوم پزشکی پیامبر اعظم مازندران. ایران.
آدرس پست الکترونیک نویسنده : jamshid.charati@gmail.com

فرزانه امینی

دانشجو دکتری آمار زیستی، دانشگاه علوم پزشکی پیامبر اعظم مازندران. ایران.
آدرس پست الکترونیک نویسنده : farzane19694@gmail.com

چکیده

مدیریت ریسک نکول اعتباری یکی از چالش های اساسی نظام بانکی است که مستقیماً بر ثبات مالی و کارایی تخصیص منابع اثر می گذارد. هدف این پژوهش، توسعه و مقایسه مدل های پیش بینی ریسک نکول و تبیین پذیری نتایج آن ها با تأکید بر شناسایی محرک های اصلی ریسک اعتباری است. داده های مورد استفاده شامل اطلاعات مشتریان اعتباری برگرفته از مجموعه داده German Credit بوده که پس از پیش پردازش، استخراج متغیرهای مشتق شده و شاخص های نوین فشار مالی، به مجموعه ای غنی از عوامل اقتصادی و رفتاری تبدیل شده است. در این مطالعه، سه رویکرد مدل سازی شامل رگرسیون لجستیک با جریمه LASSO، جنگل تصادفی و XGBoost به کار گرفته شد. ارزیابی عملکرد مدل ها با استفاده از معیار AUC نشان داد که مدل LASSO با مقدار AUC برابر با ۰.۷۷۶، جنگل تصادفی با ۰.۷۷۳ و XGBoost با ۰.۷۴۱ عملکرد مناسبی در تفکیک مشتریان خوش حساب و بد حساب دارند. اگرچه اختلاف عملکرد مدل ها محدود است، مدل های غیرخطی عملکرد بهتری در استخراج الگوهای پیچیده و غیرخطی از داده ها از خود نشان دادند. به منظور افزایش شفافیت و تفسیرپذیری، از روش SHAP برای تحلیل مدل جنگل تصادفی استفاده شد. نتایج SHAP نشان داد که شاخص هایی نظیر فشار بدهی، شدت تنش مالی، مبلغ اعتبار و مدت بازپرداخت بیشترین نقش را در پیش بینی نکول ایفا می کنند. همچنین جهت اثر متغیرها حاکی از آن است که افزایش فشار بازپرداخت و بار بدهی احتمال نکول را افزایش داده، در حالی که شاخص های نقدینگی و ثبات درآمد نقش محافظتی دارند.

واژگان کلیدی: ریسک اعتباری، یادگیری ماشین، جنگل تصادفی، XGBoost، تبیین پذیری مدل.

در سال های اخیر، با رشد سریع فناوری های اطلاعاتی و افزایش پیچیدگی بازارهای مالی، پیش بینی دقیق ریسک اعتباری به یکی از چالش های اساسی بانک ها و مؤسسات مالی تبدیل شده است. تصمیمات نادرست در زمینه اعطای اعتبار می تواند منجر به افزایش مطالبات غیرجاری، کاهش سودآوری و حتی بروز ریسک های سیستمیک در نظام مالی شود. از این رو، مؤسسات مالی نیازمند ابزارهایی هستند که بتوانند با تحلیل داده های متنوع و پیچیده، ریسک ناشی از وام دهی و سرمایه گذاری را به طور دقیق تری برآورد کنند. در این راستا، مدل های آماری پیشرفته و الگوریتم های یادگیری ماشین به عنوان ابزارهایی کارآمد برای شناسایی الگوهای پنهان در داده های مالی و ارائه پیش بینی هایی با دقت بالا مورد توجه قرار گرفته اند.

مطالعات متعدد نشان می دهند که هوش مصنوعی و فناوری های نوین حسابداری می توانند نقش مهمی در بهبود کیفیت گزارش های مالی و ارتقای تصمیم گیری سرمایه گذاران ایفا کنند. برای مثال، ارائه یک چارچوب برای پذیرش هوش مصنوعی در گزارش دهی مالی نشان می دهد که درک مفید بودن و سهولت استفاده از سیستم های مبتنی بر هوش مصنوعی، تأثیر معناداری بر بهبود کیفیت اطلاعات مالی دارد (حیدریوسف و همکاران ۱۴۰۴). همچنین مطالعات دیگر بیان می کنند که به کارگیری فناوری های نوین حسابداری در بخش دولتی منجر به افزایش دقت و کاهش خطاهای گزارش های مالی می شود (Ahmad, et al., 2024).

افزون بر این، استفاده از الگوریتم های هوش مصنوعی در گزارشگری پایداری موجب افزایش معنادار دقت و کارآمدی تحلیل داده های مرتبط با شاخص های زیست محیطی و اجتماعی می شود (Ahmad et al., 2023). علاوه بر این، استفاده از ابزارهای مبتنی بر هوش مصنوعی در فرآیندهای گزارش دهی مالی، نقش معناداری در ارتقای کیفیت اطلاعات و تسهیل تصمیم گیری آگاهانه سرمایه گذاران ایفا می کند (Anantharaman et al., 2023). افزون بر این، به کارگیری تکنیک های هوش مصنوعی می تواند ضمن افزایش دقت اطلاعات مالی، از طریق کاهش عدم تقارن اطلاعاتی، ریسک تصمیم گیری سرمایه گذاران را نیز کاهش دهد (Hamidi et al., 2023).

هم زمان با این تحولات، ادغام داده های سنتی با داده های نوظهور، ظرفیت های جدیدی را در زمینه پیش بینی ریسک اعتباری فراهم کرده است. داده های نوظهور، از جمله اطلاعات دیجیتال، تراکنش های آنلاین و الگوهای رفتاری مشتریان در بسترهای الکترونیکی، امکان ارائه تصویری جامع تر و پویاتر از وضعیت اعتباری و رفتار مالی مشتریان را فراهم می سازند. شواهد تجربی حاکی از آن است که به کارگیری توأمان این داده ها در کنار مدل های آماری و الگوریتم های یادگیری ماشین، موجب بهبود معنادار دقت پیش بینی ریسک اعتباری و افزایش توان شناسایی مشتریان پرریسک می شود (Agarwal et al., 2020; Berg et al., 2020). با این حال، یکی از چالش های اساسی در استفاده از مدل های پیچیده یادگیری ماشین، ماهیت «جعبه سیاه» آن هاست که می تواند پذیرش نتایج مدل را در محیط های نظارتی و تصمیم گیری مالی با محدودیت مواجه سازد.

در این راستا، تبیین پذیری مدل ها و رویکردهای هوش مصنوعی قابل توضیح به عنوان یکی از الزامات اساسی در کاربردهای مالی، به طور فزاینده ای مورد توجه قرار گرفته است. روش هایی مانند SHAP با تخصیص سهم هر متغیر در پیش بینی نهایی مدل، امکان بررسی هم زمان شدت و جهت اثر متغیرها بر خروجی مدل را فراهم می کنند و بدین ترتیب شفافیت تصمیمات مدل را به طور معناداری افزایش می دهند. بهره گیری از چنین چارچوب های تبیین پذیری، افزون بر تقویت اعتماد نهادهای نظارتی و مدیران مالی، زمینه استفاده عملی از نتایج مدل ها را در فرآیندهای تصمیم گیری اعتباری فراهم می سازد (Lundberg et al. 2017).

بر این اساس، پژوهش حاضر با تمرکز بر ادغام مدل های آماری و الگوریتم های یادگیری ماشین، بهره گیری از داده های نوظهور و به کارگیری چارچوب های تبیین پذیر، تلاش می کند بخشی از خلأ های موجود در ادبیات ریسک اعتباری را پوشش دهد. نتایج این پژوهش می تواند به توسعه مدل های پیش بینی دقیق تر، شفاف تر و کاربردی تر در محیط های مالی کمک کرده و از طریق بهبود تصمیم گیری های اعتباری، کاهش ریسک سیستمیک و افزایش اعتماد سرمایه گذاران، نقش مؤثری در پیشرفت مطالعات حوزه فین تک و مدیریت ریسک ایفا کند.

۱-۱. ریسک اعتباری و مدل های یادگیری ماشین

ریسک اعتباری^۱ به احتمال عدم ایفای تعهدات مالی توسط وام گیرنده اطلاق می شود و ارزیابی دقیق آن نقش اساسی در کاهش زیان ها و بهبود تصمیم گیری اعتباری دارد (Altman 1968). یکی از مهم ترین شاخص های کمی در این حوزه، احتمال نکول^۲ است که احتمال وقوع نکول در یک افق زمانی مشخص را نشان می دهد و در چارچوب های نظارتی بازل و رویه های مبتنی بر رتبه بندی داخلی (IRB) کاربرد گسترده ای دارد (Crouhy et al. 2006).

مدل های سنتی اعتبارسنجی عمدتاً بر داده های رسمی مالی متکی هستند؛ با این حال، محدودیت این داده ها به ویژه برای مشتریان دارای سابقه اعتباری کوتاه، موجب توجه به داده های نوظهور شده است (Vidal and Barbon 2019). هم زمان، پژوهش های اولیه نشان دادند که الگوریتم های یادگیری ماشین، از جمله شبکه های عصبی و روش های تکاملی، توانایی بالاتری در شناسایی روابط غیرخطی و پیچیده میان متغیرهای مالی و PD دارند (West 2000; Yobas et al. 2000). در سال های اخیر، روش های تجمیعی مانند Random Forest، AdaBoost و Gradient Boosting به طور گسترده برای افزایش دقت پیش بینی ریسک اعتباری استفاده شده اند (Moscato et al. 2021; Tripathi et al. 2022). مطالعه مرجع Lessmann با ارائه یک چارچوب مقایسه ای جامع نشان داد که بسیاری از الگوریتم های یادگیری ماشین عملکردی بهتر از مدل های سنتی نظیر رگرسیون لجستیک دارند و مسیر تحقیقات بعدی را مشخص کرد (Lessmann, Baesens et al. 2015). همچنین، مدل های ترکیبی و چند مرحله ای به عنوان راهکاری برای کاهش محدودیت های الگوریتم های منفرد توسعه یافته اند (Lee and Chen 2005; Dushimimana et al. 2020).

۱-۲. داده های نوظهور و فین تک

ظهور فناوری های مالی^۳ و پلتفرم های وام دهی دیجیتال، استفاده از داده های نوظهور مانند ردپاهای دیجیتال، داده های رفتاری موبایل و تعاملات آنلاین را امکان پذیر کرده است. این داده ها می توانند تصویری غنی تر از رفتار اعتباری مشتریان ارائه دهند و دقت پیش بینی ریسک اعتباری را به طور قابل توجهی افزایش دهند (Berg et al. 2020).

مطالعات نشان می دهد که ترکیب داده های نوظهور با الگوریتم های یادگیری ماشین، به ویژه در محیط های وام دهی همتا به همتا (Serrano-Cinca and Gutiérrez-Nieto)، منجر به بهبود معنادار عملکرد مدل ها می شود (بنی هاشمی، علی مندگاری et al. 1397). همچنین، شواهد تجربی حاکی از آن است که داده های رفتاری دیجیتال می توانند کمبود داده های مالی سنتی را جبران کرده و

^۱ Credit Risk
^۲ Probability of Default – PD
^۳ FinTech

شمول مالی را تقویت کنند (Serrano-Cinca and Gutiérrez-Nieto 2016, Bastani, Asgari et al. 2019). این تحولات، چارچوب های جدیدی برای ارزیابی ریسک اعتباری در اکوسیستم فین تک فراهم کرده اند (Vallee and Zeng 2019).

۱-۳. تبیین پذیری و XAI در مدل های هوش مصنوعی

با افزایش استفاده از مدل های پیچیده یادگیری ماشین در حوزه های حساس مالی، مسئله تبیین پذیری و شفافیت تصمیمات مدل اهمیت ویژه ای یافته است. بسیاری از این مدل ها به عنوان «جعبه سیاه» شناخته می شوند و درک سازوکار تصمیم گیری آن ها برای نهادهای نظارتی و تصمیم گیران دشوار است. در این راستا، هوش مصنوعی تبیین پذیر^۴ به عنوان پاسخی به این چالش مطرح شده است (Adadi and Berrada 2018).

روش های مدل-ناوابسته مانند LIME و به ویژه SHAP، امکان تبیین سهم متغیرها در پیش بینی مدل را فراهم می کنند (Ribeiro, Singh et al. 2016). SHAP با اتکا بر نظریه بازی ها، هم تبیین محلی و هم تبیین جهانی ارائه می دهد و به یکی از پرکاربردترین ابزارهای XAI در مطالعات مالی تبدیل شده است (Lundberg and Lee 2017, Giudici and Raffinetti 2021). پژوهش ها نشان می دهند که استفاده از XAI می تواند ریسک مدل را کاهش داده، قابلیت ممیزی را افزایش دهد و اعتماد نهادهای نظارتی و مشتریان را تقویت کند (Bussmann, Giudici et al. 2021, Bücker, Szepannek et al. 2022).

۱-۴. ریسک سیستمیک و شبکه های مالی

ریسک سیستمیک به خطر سرایت شوک ها در سراسر سیستم مالی اشاره دارد و در دهه های اخیر به یکی از محورهای اصلی تحقیقات مالی تبدیل شده است. شاخص هایی مانند CoVaR برای سنجش اثر متقابل مؤسسات مالی توسعه یافته اند (Acharya, Engle et al. 2012) (Tobias and Brunnermeier 2016). مطالعات نشان می دهد که وابستگی های شبکه ای و غیرخطی نقش مهمی در انتقال ریسک دارند (Adams, Füss et al. 2014, Abbassi, Brownlees et al. 2017).

پژوهش های جدیدتر با بهره گیری از مدل های شبکه ای و الگوریتم های یادگیری ماشین، امکان شناسایی الگوهای پنهان، خوشه های ریسک و نقاط شکننده سیستم مالی را فراهم کرده اند. این رویکردها نشان می دهند که ترکیب یادگیری ماشین و تحلیل شبکه ای می تواند دیدی جامع تر و پیش بینی پذیرتر از ریسک سیستمیک ارائه دهند (Giudici, Hadji-Misheva et al. 2020, Kumar, Rahman et al. 2024).

۲- روش تحقیق

در این پژوهش، چارچوبی ترکیبی از آمار کلاسیک و یادگیری ماشین برای پیش بینی ریسک اعتباری و شناسایی ریسک های نوظهور بانکی ارائه شده است. این چارچوب شامل پنج لایه تحلیلی مکمل است که کل چرخه داده، مدل سازی، ارزیابی، تبیین و تصمیم گیری را پوشش می دهد:

۱. لایه تولید و آماده سازی داده ها^۵
۲. لایه انتخاب ویژگی ها^۶
۳. لایه مدل سازی آماری و یادگیری ماشین^۷
۴. لایه ارزیابی آگاه از ریسک^۸
۵. لایه تبیین پذیری و تحلیل ریسک^۹

هدف اصلی این ساختار، ارائه پیش بینی دقیق احتمال نکول (Habibi tabar) همراه با شفافیت و قابلیت تبیین برای مدیران ریسک و نهادهای نظارتی است. استفاده از روش های تبیین پذیر مانند مقادیر SHAP امکان شناسایی محرک های اصلی ریسک و تحلیل منطق تصمیم گیری مدل را فراهم می سازد.

۱-۲. داده ها و انتخاب ویژگی ها

داده های مورد استفاده شامل مجموعه German Credit Data منتشر شده توسط مخزن UCI است که شامل ۱۰۰۰ مشاهده، ۲۰ متغیر توضیحی و یک متغیر هدف (نکول/عدم نکول) ۱۰ می باشد. این داده ها شامل اطلاعات اعتباری، مالی و فردی وام گیرندگان بوده و با چارچوب های استاندارد ارزیابی احتمال نکول در رویکردهای مبتنی بر رتبه بندی داخلی (IRB) هم راستا هستند.

انتخاب ویژگی ها در دو سطح انجام شد: نخست، متغیرهای اصلی اعتباری شامل مشخصات وام، سابقه اعتباری، توان بازپرداخت ۱۱ و ویژگی های فردی؛ که انتخاب آن ها مبتنی بر ادبیات کلاسیک ریسک اعتباری و مطالعاتی نظیر (Kleinbaum, Thomas et al., Klein et al.) و (Ahmad, Shamsuddin et al.) (Lessmann et al.) است.

جدول ۱. متغیرهای ورودی مدل ریسک اعتباری

نقش در ریسک	تعریف عملیاتی	متغیر	دسته
شدت تعهد مالی	کل مبلغ تسهیلات	مبلغ وام	مشخصات وام
افق ریسک	تعداد ماه	مدت بازپرداخت	مشخصات وام
ریسک رفتاری	رفتار گذشته	سابقه اعتباری	سابقه اعتباری
توان بازپرداخت	فشار نقدینگی	نسبت قسط به درآمد	توان بازپرداخت
ضربه گیر شوک	سطح ذخایر مالی	پس انداز	وضعیت مالی
ناهمگنی ریسک	چرخه عمر	سن	ویژگی فردی
متغیر وابسته	نکول/عدم نکول	نکول	برچسب

Data Layer^۵
Feature Selection Layer^۶
Modeling Layer^۷
Risk-aware Evaluation Layer^۸
Explainability & Risk Analysis Layer^۹
Modeling Layer Default (Target)^{۱۰}
Installment Rate^{۱۱}

جدول ۱ متغیرهای اصلی ورودی و نیز شاخص های مشتق شده منتخب را نشان می دهد.

در سطح دوم، مجموعه ای از ویژگی های مشتق شده به منظور شناسایی ریسک های نوظهور طراحی شد؛ از جمله شاخص فشار بدهی^{۱۲}، شاخص شکنندگی مالی^{۱۳}، تراکم اعتباری^{۱۴} و شاخص تنش کوتاه مدت^{۱۵}. این شاخص ها با هدف استخراج سیگنال های زودهنگام فشار مالی و الگوهای پریسک ساخته شده اند.

جدول ۱ متغیرهای اصلی ورودی و جدول ۲ شاخص های مشتق شده منتخب را نشان می دهد.

در سطح دوم، مجموعه ای از ویژگی های مشتق شده به منظور شناسایی ریسک های نوظهور طراحی شد؛ از جمله شاخص فشار بدهی^{۱۶}، شاخص شکنندگی مالی^{۱۷}، تراکم اعتباری^{۱۸} و شاخص تنش کوتاه مدت^{۱۹}. این شاخص ها با هدف استخراج سیگنال های زودهنگام فشار مالی و الگوهای پریسک ساخته شده اند.

جدول ۲. شاخص های مشتق شده منتخب برای شناسایی ریسک های نوظهور

نوع ریسک	منطق محاسبه	شاخص
فشار بدهی	مبلغ وام / مدت	شاخص فشار بدهی
ریسک زودهنگام	قسط × مبلغ	شاخص شکنندگی مالی
بیش اعتباری	اعتبارات فعال / سابقه شغلی	شدت اعتباری
ریسک نقدینگی	وضعیت پس انداز + حساب	شاخص نقدینگی
هشدار زودهنگام	قسط / مدت	شاخص تنش کوتاه مدت

^{۱۲} Debt Burden Index

^{۱۳} Financial Stress Indicator

^{۱۴} Credit Intensity

^{۱۵} Short-Term Stress Index

^{۱۶} Debt Burden Index

^{۱۷} Financial Stress Indicator

^{۱۸} Credit Intensity

^{۱۹} Short-Term Stress Index

۲-۲. لایه مدل سازی آماری و یادگیری ماشین

در این لایه، مدل های مختلف با هدف پیش بینی PD و مقایسه عملکرد توسعه داده شدند. از یک سو، مدل های آماری کلاسیک شامل رگرسیون لجستیک و لاسو^{۲۰} به عنوان مدل های پایه و قابل تفسیر مورد استفاده قرار گرفتند. بطوریکه لاسو با منظم سازی، مسئله هم خطی و تفکیک کامل^{۲۱} داده ها را کنترل می کند. از سوی دیگر، الگوریتم های یادگیری ماشین شامل درخت تصمیم، جنگل تصادفی و XGBoost برای مدل سازی روابط غیرخطی و تعاملات پیچیده میان متغیرها به کار گرفته شدند.

فرآیند مدل سازی شامل تقسیم داده ها به مجموعه های آموزش و آزمون، پیش پردازش داده ها، آموزش مدل ها و تنظیم پارامترها است. ارزیابی عملکرد مدل ها بر اساس معیارهایی نظیر AUC-ROC، F1-Score، Accuracy و آماره KS انجام شد و مدل نهایی با توجه به دقت، پایداری و قابلیت تبیین انتخاب گردید.

۲-۳. لایه ارزیابی آگاه از ریسک:

ارزیابی مدل ها صرفاً به معیارهای متداول یادگیری ماشین محدود نشد، بلکه شاخص های ریسک محور نیز مدنظر قرار گرفت. این شاخص ها شامل توزیع PD، تحلیل زیان مورد انتظار و بررسی رفتار مدل در نواحی پرریسک است. نخست، توزیع احتمال نکول (Habibi tabar) برای بررسی نحوه تفکیک وام گیرندگان در سطوح مختلف ریسک تحلیل شد؛ به گونه ای که توان مدل در شناسایی مشتریان پرریسک و خوشه بندی سطوح ریسک به صورت دقیق ارزیابی گردد. سپس، زیان مورد انتظار به عنوان یکی از شاخص های کلیدی مدیریت ریسک اعتباری، با ترکیب مؤلفه های PD، نرخ زیان در صورت نکول و میزان قرارگیری در معرض ریسک مورد توجه قرار گرفت تا پیامدهای مالی پیش بینی های مدل به صورت کمی قابل ارزیابی باشد.

در ادامه، رفتار مدل در نواحی پرریسک توزیع با استفاده از شاخص های Tail Risk شامل Value at Risk (Bastani, Asgari et al.) و Conditional Value at Risk (CVaR) بررسی شد. این تحلیل امکان ارزیابی پایداری مدل در شرایط بحرانی و سناریوهای بدبینانه را فراهم می سازد؛ شرایطی که در آن ها خطای پیش بینی می تواند منجر به زیان های قابل توجه بانکی شود. در نهایت، آزمون های تنش رای شبیه سازی شوک های اعتباری و ارزیابی واکنش مدل نسبت به تغییرات شدید متغیرهای کلیدی انجام شد. این رویکرد ارزیابی چندبعدی، امکان مقایسه مدل ها را نه تنها از منظر دقت آماری، بلکه از دیدگاه کارایی عملی، پایداری ریسک و قابلیت اتکای تصمیم گیری در محیط های واقعی بانکی فراهم می کند.

۲-۴. لایه تبیین پذیری و تحلیل ریسک

به منظور افزایش شفافیت تصمیمات مدل های غیرخطی و اطمینان پذیری کاربردهای بانکی از روش های تبیین پذیر شامل مقادیر SHAP و نمودارهای وابستگی جزئی برای تحلیل اثر واقعی متغیرها و تعاملات آن ها بر احتمال نکول استفاده شد. این رویکرد

LASSO^{۲۰}
Complete Separation^{۲۱}

امکان شناسایی متغیرهای کلیدی، جهت و شدت اثر آنها بر ریسک اعتباری، خوشه بندی وام گیرندگان بر اساس پروفایل ریسک و شناسایی زود هنگام ریسک های نوظهور را فراهم می سازد. بدین ترتیب، قدرت پیش بینی مدل های یادگیری ماشین با قابلیت تبیین و الزامات نظارتی ترکیب شده و تصمیم گیری اعتباری به سطحی شفاف و قابل اتکا می یابد.

۳. یافته ها:

۳-۱. تحلیل توصیفی داده ها:

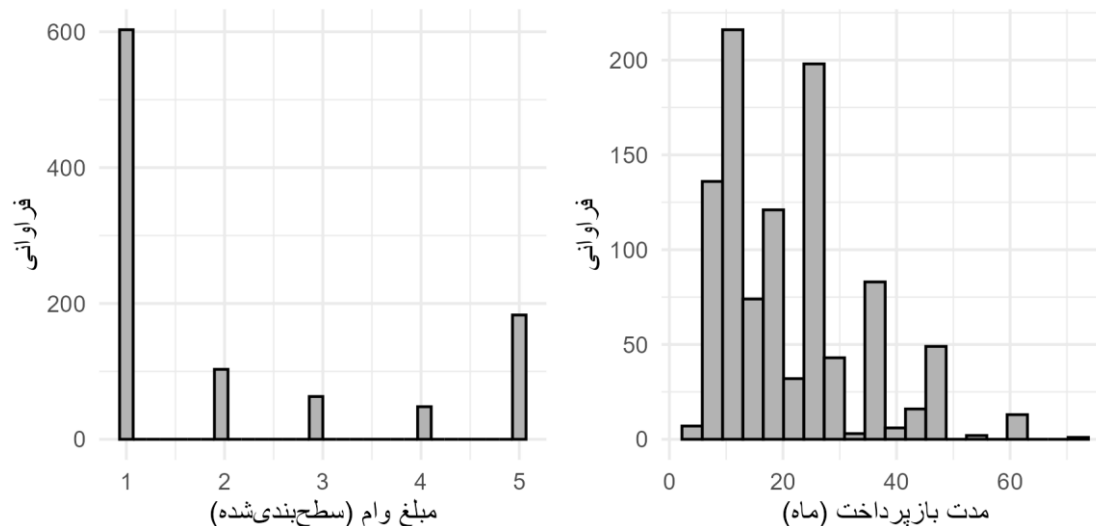
در این بخش، آمار توصیفی مجموعه ای از متغیرهای کلیدی مورد استفاده در مدل های پیش بینی احتمال نکول ارائه می شود تا تصویری کلی از ساختار داده ها و ناهمگنی ریسک در نمونه پژوهش فراهم گردد. تمرکز بر متغیرهایی است که در ادبیات ریسک اعتباری به عنوان محرک های اصلی نکول شناخته می شوند. (Thomas et al., 2002; Brown & Mues, 2012)

با توجه به ماهیت کدگذاری شده و دسته ای بخش قابل توجهی از متغیرهای مجموعه داده German Credit، آمار توصیفی صرفاً برای متغیرهای پیوسته و ترتیبی با تفسیر اقتصادی مستقیم شامل مدت بازپرداخت، مبلغ وام، نرخ تعهد اقساط و سن وام گیرنده گزارش شده است.

جدول ۲. آمار توصیفی متغیرهای منتخب

متغیر	تعداد	میانگین	انحراف معیار	مینیمم	ماکسیمم
مبلغ وام (کدگذاری شده)	۱۰۰۰	۲۰۱۰	۱۰۵۸	۱	۵
مدت بازپرداخت (ماه)	۱۰۰۰	۲۰۹۰	۱۲۰۶	۴	۷۲
نرخ تعهد اقساط	۱۰۰۰	۲۸۵	۱۰۱۰	۱	۴
سن (کدگذاری شده)	۱۰۰۰	۱۰۱۶	۰۳۶	۱	۲

جدول ۲ نشان می دهد که متغیر مدت بازپرداخت از پراکندگی قابل توجهی برخوردار است که بیانگر ناهمگنی افق های زمانی بازپرداخت در میان وام گیرندگان می باشد. همچنین، توزیع متغیرهای ترتیبی مرتبط با مبلغ وام، نرخ اقساط و سن، وجود سطوح متفاوتی از فشار مالی، توان بازپرداخت و ریسک رفتاری را در نمونه مورد بررسی منعکس می کند. این ناهمگنی، ضرورت استفاده از مدل های انعطاف پذیر و غیرخطی برای تحلیل ریسک اعتباری و ریسک های نوظهور را تقویت می کند.



شکل ۱. توزیع مبلغ وام و مدت بازپرداخت در داده های اعتباری

شکل ۱ توزیع متغیرهای کلیدی مبلغ وام و مدت بازپرداخت را نشان می دهد. پراکندگی نسبتاً بالای متغیر مدت بازپرداخت بیانگر ناهمگنی قابل توجه در افق های زمانی بازپرداخت وام گیرندگان است که می تواند منجر به تفاوت در تجمع ریسک نکول در دوره های مختلف شود. از سوی دیگر، متغیر مبلغ وام که به صورت ترتیبی کدگذاری شده است، وجود سطوح متفاوتی از ریسک اعتباری را در میان مشتریان منعکس می کند و نشان دهنده تمایز میان وام های با فشار بدهی پایین تا بالا می باشد. این الگوها مؤید ضرورت استفاده از مدل های انعطاف پذیر و غیرخطی برای شناسایی ناهمگنی ریسک و تحلیل ریسک های نوظهور بانکی هستند.

۳-۲. مقایسه عملکرد مدل های پیش بینی ریسک اعتباری

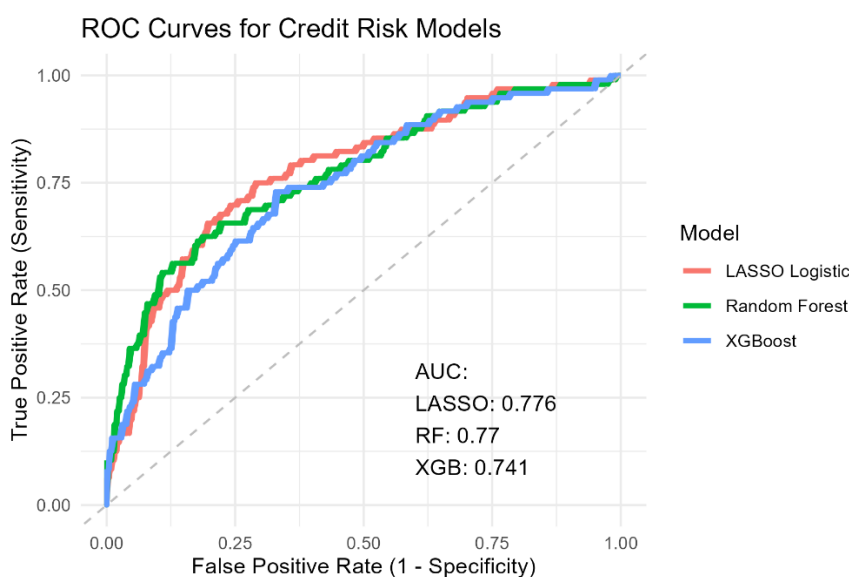
در این بخش، عملکرد مدل های آماری کلاسیک و الگوریتم های یادگیری ماشین در پیش بینی احتمال نکول مورد مقایسه قرار می گیرد. هدف اصلی، ارزیابی توان تمایز مدل ها، پایداری پیش بینی ها و قابلیت کاربرد آن ها در چارچوب مدیریت ریسک اعتباری است. منحنی گیرنده-عملیاتی (Breiman) بیانگر رابطه بین نرخ کشف صحیح نکول^{۲۲} و نرخ خطای هشدار کاذب^{۲۳} در سطوح مختلف آستانه تصمیم گیری است. مقدار سطح زیر منحنی ROC به عنوان یک معیار مستقل از آستانه، توان مدل را در رتبه بندی وام گیرندگان بر اساس ریسک نکول نشان می دهد؛ به گونه ای که مقادیر بالاتر AUC بیانگر قدرت تفکیک بهتر بین مشتریان نکول کننده و غیر نکول کننده هستند. در ادبیات مدیریت ریسک، مقادیر AUC بالاتر از ۰.۷۵ به عنوان عملکرد مناسب برای کاربردهای عملی نظیر رتبه بندی اعتباری، تخصیص سرمایه و سیستم های هشدار زودهنگام تلقی می شوند.

^{۲۲} True Positive Rate
^{۲۳} False Positive Rate

جدول ۳. عملکرد مدل ها بر اساس معیار AUC-ROC

مساحت زیر منحنی ROC	مدل
۰,۷۷۶	رگرسیون لجستیک لاسو
۰,۷۷۳	جنگل تصادفی
۰,۷۴۱	XGBoost

جدول ۳ عملکرد سه مدل اصلی شامل رگرسیون لجستیک لاسو، جنگل تصادفی و XGBoost را بر اساس معیار AUC-ROC گزارش می کند. نتایج بیانگر آن است که مدل رگرسیون لجستیک لاسو با مقدار AUC برابر با ۰,۷۷۶، بالاترین توان تفکیک بین وام گیرندگان نکول کننده و غیر نکول کننده را ارائه می دهد. این یافته نشان می دهد که در حضور ویژگی های مشتق شده و کنترل هم خطی متغیرها، مدل رگرسیون لاسو همچنان می تواند عملکردی رقابتی و قابل اتکا داشته باشد؛ موضوعی که با نتایج مطالعات پیشین هم راستا است. (Lessmann et al., 2015)



شکل ۲. منحنی ROC برای رتبه بندی وام گیرندگان

شکل ۲ منحنی های ROC مربوط به سه مدل رگرسیون لجستیک لاسو، جنگل تصادفی و XGBoost را نشان می دهد. مقادیر AUC بیانگر توان تمایز مدل ها در تفکیک نکول کنندگان از غیر نکول کنندگان هستند و نشان می دهند که هر سه مدل از قدرت پیش بینی قابل قبولی برخوردارند. نمودار ROC امکان ارزیابی عملکرد مدل ها را از سه منظر فراهم می کند: توان تفکیک کلی از طریق AUC، رفتار مدل ها در آستانه های مختلف تصمیم گیری، و تعیین نقطه بهینه تصمیم گیری عملی با استفاده از شاخص Youden که موازنه مناسبی میان حساسیت و ویژگی ایجاد می کند.

هرچند ROC صرفاً عملکرد کلی مدل ها را نشان می دهد، تفاوت میان مدل های آماری و یادگیری ماشین را می توان با توجه به ویژگی های ساختاری الگوریتم ها تبیین کرد. به طور خاص، جنگل تصادفی و XGBoost به دلیل توانایی در مدل سازی روابط

غیرخطی و تعاملات پیچیده میان متغیرها، ظرفیت بالاتری برای شناسایی الگوهای رفتاری و ریسک های نوظهور دارند (Breiman, 2001; Chen, 2016).

۳-۳. تبیین پذیری مدل و تحلیل ریسک های نوظهور

۳-۳-۱. تحلیل تبیینی مبتنی بر SHAP

به منظور تبیین تصمیمات مدل جنگل تصادفی و شناسایی عوامل کلیدی مؤثر بر احتمال نکول، از چارچوب SHAP استفاده شد. تحلیل SHAP در این پژوهش به صورت تبیین پذیری کلی و مبتنی بر نمونه گیری از کل داده ها انجام شده است؛ به طوری که میانگین قدر مطلق اثر SHAP²⁴ به عنوان شاخص شدت اثر متغیرها و میانگین SHAP به عنوان شاخص جهت اثر آنها بر ریسک نکول محاسبه شد. (Lundberg, Lee, 2017)

جدول ۴. مهم ترین متغیرهای اثرگذار بر خروجی مدل جنگل تصادفی بر اساس مقادیر SHAP

متغیر	(جهت اثر) میانگین SHAP	(شدت اثر) قدر مطلق میانگین SHAP
فشار بدهی بلند مدت	-۰.۰۵۳	۰.۰۵۳
شاخص فشار مالی	-۰.۰۳۶	۰.۰۳۶
دوره باز پرداخت وام	-۰.۰۳۶	۰.۰۳۶
شاخص استرس کوتاه مدت	-۰.۰۳۰	۰.۰۳۰
بدهکاران کوتاه مدت	۰.۰۳۰	۰.۰۳۰
سابقه اعتباری	-۰.۰۲۶	۰.۰۲۶
وضعیت شخصی/جنسیت	-۰.۰۲۳	۰.۰۲۳
وضعیت حساب جاری	۰.۰۲۳	۰.۰۲۳
مبلغ اعطا وام	-۰.۰۲۲	۰.۰۲۲
وضعیت حساب جاری (عددی)	۰.۰۱۹	۰.۰۱۹

نتایج جدول ۴ نشان می دهد که متغیرهای وضعیت حساب جاری، مدت بازپرداخت و سابقه اعتباری، بالاترین میانگین قدر مطلق اثر SHAP را داشته و به عنوان محرک های اصلی ریسک نکول شناسایی می شوند. همچنین، علامت میانگین جهت اثر SHAP بیانگر نقش افزایشی یا کاهششی این متغیرها بر احتمال نکول در سطح کل داده ها است.

²⁴ SHAP (Mean |SHAP|)

از منظر تفسیر بانکی، وضعیت حساب جاری به عنوان شاخص نقدینگی کوتاه مدت و توان بازپرداخت، بیشترین سهم را در تغییرات پیش بینی احتمال نکول دارند. همچنین، دوره های بازپرداخت طولانی و سابقه اعتباری ضعیف، اثر معناداری بر افزایش احتمال نکول دارند؛ یافته ای که با ادبیات کلاسیک ریسک اعتباری همخوانی دارد. (Altman, 1968; Thomas et al., 2002)

۵. بحث و تفسیر نتایج

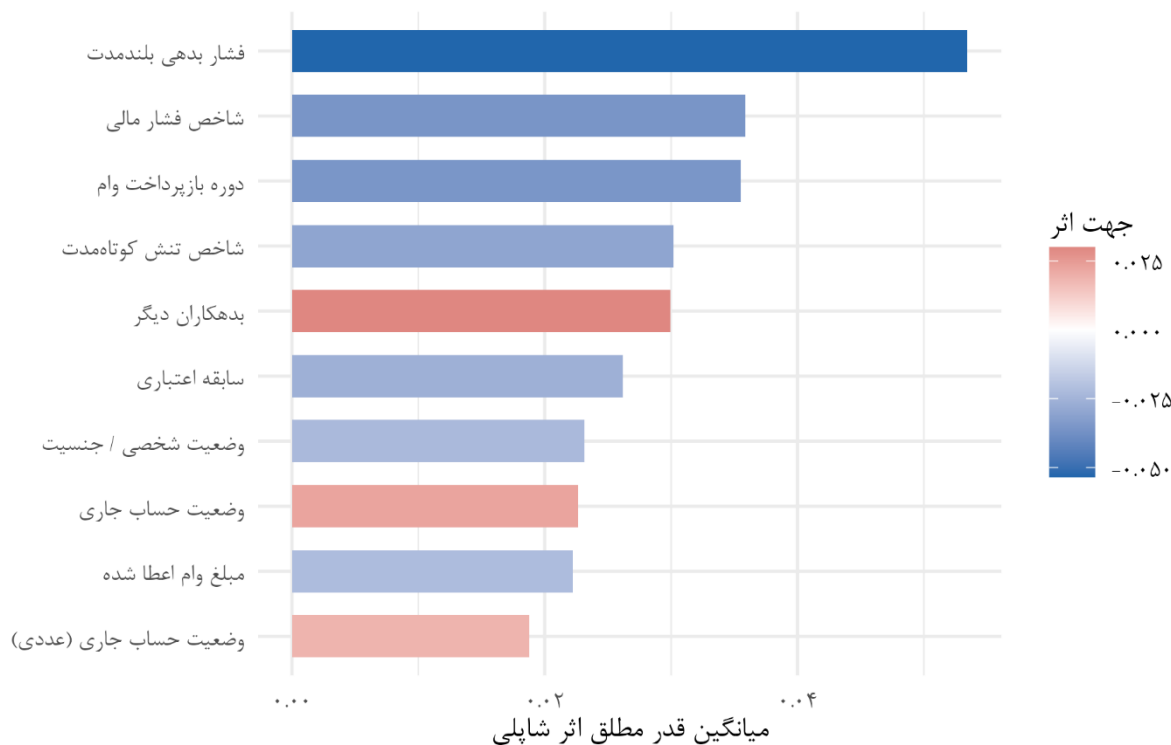
۵-۱. چرایی تمرکز تحلیل SHAP بر مدل جنگل تصادفی

اگرچه در این پژوهش سه مدل شامل رگرسیون لجستیک لاسو، جنگل تصادفی و XGBoost توسعه داده شد، تحلیل تبیین پذیری مبتنی بر SHAP صرفاً برای مدل جنگل تصادفی ارائه گردید. دلیل این انتخاب، ماهیت تعادلی جنگل تصادفی میان دقت پیش بینی و پایداری تبیین پذیری است. برخلاف مدل لاسو که ساختاری خطی و ذاتاً قابل تفسیر دارد، مدل های درختی غیرخطی نیازمند ابزارهای XAI برای استخراج منطق تصمیم گیری هستند. از سوی دیگر، اگرچه XGBoost نیز مدلی قدرتمند است، اما ساختار تقویتی و حساسیت بالاتر آن به تنظیمات پارامترها، موجب نوسان بیشتر مقادیر SHAP و کاهش پایداری تفسیر در سطح کل داده ها می شود. مطالعات پیشین نشان داده اند که جنگل تصادفی در مقایسه با مدل های Boosting، تفسیر SHAP پایدارتر و قابل اعتمادتری در کاربردهای بانکی و نظارتی ارائه می دهد. (Lundberg et al. 2017; Bussmann et al., 2021; Bücker et al., 2022) بنابراین، تمرکز بر جنگل تصادفی امکان ارائه تبیینی شفاف، قابل اتکا و سازگار با الزامات نظارتی را فراهم می سازد.

۵-۲. محرک های اصلی ریسک نکول (تفسیر اقتصادی متغیرهای کلیدی)

نتایج تحلیل SHAP نشان می دهد که شاخص های مشتق شده فشار مالی مهم ترین نقش را در پیش بینی ریسک نکول ایفا می کنند. بر اساس میانگین قدر مطلق مقادیر SHAP، شاخص فشار بدهی بلندمدت قوی ترین محرک نکول شناسایی شد و پس از آن، شاخص فشار کوتاه مدت، مدت بازپرداخت وام و شاخص تنش مالی قرار دارند. این یافته نشان می دهد که ابعاد پنهان فشار مالی و شکنندگی بازپرداخت، نقش تعیین کننده تری نسبت به متغیرهای کلاسیک در رفتار نکول دارند.

شکل ۲ نمودار توزیع اثر این متغیرها را بر خروجی مدل نشان می دهد. تمرکز جرم نقاط و شدت رنگ ها بیانگر آن است که با افزایش فشار بدهی و تنش مالی، احتمال نکول به طور غیرخطی افزایش می یابد. طول میله ها نشان دهنده اهمیت متغیرها و رنگ آن ها بیانگر جهت اثر متغیر بر ریسک نکول است (قرمز: اثر افزایشی، آبی: اثر کاهش). این حساسیت بالا نسبت به متغیرهای مشتق شده نشان می دهد که الگوریتم جنگل تصادفی، در مقایسه با مدل های آماری کلاسیک، ظرفیت تحلیلی بالاتری برای کشف ریسک های نوظهور و رفتارهای اعتباری پیچیده دارد. (Giudici, Raffinetti, 2021)



شکل ۳. نمودار توزیع اثر برای ۱۰ متغیر برتر در مدل جنگل تصادفی

در کنار این متغیرها، مدت بازپرداخت وام به عنوان شاخص فشار مالی بلندمدت، و وضعیت حساب جاری به عنوان نمایانگر نقدینگی کوتاهمدت، نقش مکملی در پیش بینی نکول دارند. دوره های بازپرداخت طولانی تر و وضعیت نامطلوب حساب جاری، انعطاف پذیری مالی مشتری را کاهش داده و احتمال نکول را افزایش می دهند.

سابقه اعتباری و مبلغ وام اگرچه اثر معناداری دارند، اما شدت تأثیر آن ها نسبت به شاخص های مشتق شده فشار مالی کمتر است. این نتیجه نشان می دهد که مدل جنگل تصادفی، فراتر از شاخص های سنتی، قادر به شناسایی الگوهای پیچیده و غیرخطی ریسک نکول و ریسک های نوظهور اعتباری است؛ یافته ای که با ادبیات نوین ریسک اعتباری هم خوانی دارد (Altman, 1968; Thomas, 2017).

از منظر اقتصادی و بانکی، وضعیت حساب جاری نمایانگر نقدینگی کوتاهمدت و جریان وجوه در دسترس وام گیرنده است. مشتریانی با وضعیت نامطلوب حساب جاری معمولاً توان کمتری در مدیریت شوک های نقدینگی دارند و در مواجهه با نوسانات درآمدی، سریع تر به نکول سوق داده می شوند. مدت بازپرداخت وام به عنوان شاخصی از فشار مالی بلندمدت عمل می کند؛ دوره های بازپرداخت طولانی تر می توانند منجر به انباشت بدهی، افزایش حساسیت به تغییرات اقتصادی و کاهش انعطاف پذیری مالی شوند. سابقه اعتباری نیز پروکسی مستقیم رفتار مالی گذشته وام گیرنده است و نتایج این پژوهش هم راستا با ادبیات کلاسیک

نشان می دهد که سابقه ضعیف بازپرداخت، احتمال نکول آینده را به طور معناداری افزایش می دهد (Altman, 1968, Thomas, 2017).

۶. نتیجه گیری

نتایج حاصل از تحلیل مدل های پیش بینی ریسک اعتباری نشان می دهد که همه مدل ها، شامل رگرسیون لجستیک لاسو، جنگل تصادفی و XGBoost، توانایی قابل قبولی در تفکیک وام گیرندگان نکول کننده و غیر نکول کننده دارند، اما تفاوت های ساختاری و عملکردی آن ها قابل توجه است. مدل رگرسیون لجستیک لاسو با AUC برابر با ۰.۷۷۶ بالاترین توان تفکیک را ارائه داد، که نشان دهنده مزیت استفاده از مدل های آماری منظم شده در حضور ویژگی های مشتق شده و کنترل هم خطی متغیرها است. این یافته با نتایج مطالعات پیشین در زمینه رتبه بندی اعتباری و پیش بینی نکول هم راستا است (Lessmann et al., 2015).

از سوی دیگر، مدل های درختی مانند جنگل تصادفی و XGBoost به دلیل توانایی مدل سازی روابط غیر خطی و تعاملات پیچیده میان متغیرها، ظرفیت بالاتری برای شناسایی الگوهای رفتاری نوظهور و ریسک های مخفی دارند. تحلیل های مبتنی بر SHAP نشان داد که جنگل تصادفی به طور خاص توانایی بالایی در شناسایی محرک های اصلی ریسک و الگوهای نوظهور دارد و نسبت به XGBoost پایداری تبیین پذیری بالاتری ارائه می دهد. حساسیت بالای XGBoost به تنظیمات پارامترها و ساختار تقویتی آن موجب نوسان بیشتر مقادیر SHAP و کاهش پایداری تفسیر در سطح کل داده ها می شود، که محدودیت مهمی در کاربرد عملی و گزارش به مدیران ریسک به شمار می رود.

از منظر اقتصادی و عملیاتی، نتایج تحلیل SHAP نشان می دهد که متغیرهایی مانند وضعیت حساب جاری، مدت بازپرداخت و سابقه اعتباری بیشترین نقش را در پیش بینی ریسک نکول دارند. این متغیرها نمایانگر نقدینگی کوتاه مدت، فشار مالی بلندمدت و رفتار مالی گذشته وام گیرندگان هستند و تأثیر معناداری بر احتمال نکول دارند. شاخص های مشتق شده مانند فشار بدهی و تنش مالی نیز به شناسایی زود هنگام ریسک های نوظهور کمک می کنند و نشان می دهند که مدل های یادگیری ماشین قادر به کشف روابط غیر خطی و پیچیده در داده های اعتباری هستند که مدل های آماری کلاسیک کمتر قادر به استخراج آن ها هستند.

با توجه به یافته ها، ترکیب مدل های آماری تبیین پذیر با الگوریتم های یادگیری ماشین، تعادلی مؤثر میان دقت پیش بینی، شفافیت و قابلیت کاربرد عملی در محیط بانکی ایجاد می کند. این ترکیب به مدیران ریسک امکان می دهد که نه تنها تصمیمات مبتنی بر داده اتخاذ کنند، بلکه منطق پشت پیش بینی های مدل را نیز درک کنند و ریسک های نوظهور را به موقع شناسایی نمایند.

در نهایت این پژوهش نشان داد که جنگل تصادفی تعادل مطلوبی میان دقت پیش بینی و تبیین پذیری ارائه می دهد و توانایی شناسایی محرک های اصلی و ریسک های نوظهور را دارد. متغیرهایی مانند وضعیت حساب جاری، مدت بازپرداخت و سابقه اعتباری بیشترین نقش را در پیش بینی نکول دارند و شاخص های مشتق شده مانند فشار بدهی و تنش مالی به شناسایی زود هنگام ریسک ها کمک می کنند.

کاربردهای عملی شامل رتبه بندی اعتباری مشتریان، طراحی سیستم های هشدار زود هنگام و تخصیص بهینه سرمایه است.

پیشنهادهای برای تحقیقات آینده:

۱. استفاده از داده های گسترده تر و متنوع شامل تراکنش های مشتریان،
۲. توسعه مدل های ترکیبی برای افزایش دقت و پایداری،
۳. بررسی شاخص ها و ویژگی های مشتق شده جدید برای شناسایی ریسک های نوظهور،
۴. تحلیل عملکرد مدل ها در شرایط بحرانی یا شوک های اقتصادی.

منابع

- Abbassi, P., C. Brownlees, C. Hans and N. Podlich (2017). "Credit risk interconnectedness: What does the market really know?" *Journal of Financial Stability* 29: 1-12
- Acharya, V., R. Engle and M. Richardson (2012). "Capital shortfall: A new approach to ranking and regulating systemic risks." *American Economic Review* 102(3): 59-64.
- Adadi, A. and M. Berrada (2018). "Peeking inside the black-box: a survey on explainable artificial intelligence (XAI)." *IEEE access* 6: 52138-52160.
- Adams, Z., R. Füss and R. Gropp (2014). "Spillover effects among financial institutions: A state-dependent sensitivity value-at-risk approach." *Journal of Financial and Quantitative Analysis* 49(3): 575-598.
- enAgarwal, S., S. Alok, P. Ghosh and S. Gupta (2020). "Financial inclusion and alternate credit scoring for the millennials: role of big data and machine learning in fintech." *Business School, National University of Singapore Working Paper, SSRN 3507827*.
- Aghaei, A. M., J. Gholami, A. Sangchooli, Y. Rostam-Abadi, S. Olamazadeh, M. Ardeshtir, S. Baheshmat, B. Shadloo, M. Taj, K. Saeed and A. Rahimi-Movaghar (2023). "Prevalence of injecting drug use and HIV, hepatitis B, and hepatitis C in people who inject drugs in the Eastern Mediterranean region: a systematic review and meta-analysis." *Lancet Glob Health* 11(8): e1225-e1237.
- Ahmad, A., H. Abusaimeh, A. Rababah, M. Alqsass, N. Al-Olima and M. Hamdan (2024). "Assessment of effects in advances of accounting technologies on quality financial reports in Jordanian public sector." *Uncertain Supply Chain Management* 12(Ratkowsky, Lowry, McMeekin, Stokes, & Chandler): 133-142.
- Ahmad, S., N. R. Shamsuddin and H. Farizad (2015). "Risks of divorce: comparison between Cox and Parametric Models".
- Altman, E. I. (1968). "Financial ratios, discriminant analysis and the prediction of corporate bankruptcy." *The journal of finance* 23(Ratkowsky et al.): 589-609.
- Amin, M. S. M. (2002). *Synthesis and Characterization of New Chelating Resins Containing Thiourea as the Functional Group*, Universiti Putra Malaysia.
- Anantharaman, D., Rozario, A., & Parker, C. A. (2023). *Artificial intelligence and financial reporting quality. Available at SSRN 4625279*.
- Bastani, K., E. Asgari and H. Namavari (2019). "Wide and deep learning for peer-to-peer lending." *Expert Systems with Applications* 134: 209-224.
- Berg, T., V. Burg, A. Gombović and M. Puri (2020). "On the rise of fintechs: Credit scoring using digital footprints." *The Review of Financial Studies* 33(Ratkowsky et al.): 2845-289.
- Bracke, P., A. Datta, C. Jung and S. Sen (2019). "Machine learning explainability in finance: an application to default risk analysis".
- Breiman, L. (2001). "Random Forests." *Machine Learning* 45(Ratkowsky et al.): 5-32.
- Bücker, M., G. Szepannek, A. Gosiewska and P. Biecek (2022). "Transparency, auditability, and explainability of machine learning models in credit scoring." *Journal of the Operational Research Society* 73(Ratkowsky et al.): 70-90.
- Bussmann, N., P. Giudici, D. Marinelli and J. Papenbrock (2021). "Explainable machine learning in credit risk management." *Computational Economics* 57(Ratkowsky et al.): 203-216.
- Crouhy, M., D. Galai and R. Mark (2006). *The essentials of risk management*, McGraw-Hill New York.
- Dushimimana, B., Y. Wambui, T. Lubega and P. E. McSharry (2020). "Use of machine learning techniques to create

- a credit score model for airtime loans." Journal of Risk and Financial Management 13(8): 180.
- Giudici, P., B. Hadji-Misheva and A. Spelta (2020). "Network based credit risk models." Quality Engineering 32(Ratkowsky et al.): 199-211.
- Giudici, P. and E. Raffinetti (2021). "Shapley-Lorenz eXplainable artificial intelligence." Expert systems with applications 167: 114104.
- Hamidi, K. S., Al-Saadi, M. N., & Al-Shibli, G. I. (2023). Using artificial intelligence techniques to improve financial reporting and its impact on investor decisions. *Tikrit Journal of Administrative and Economic Sciences*, 19(Ratkowsky et al.), 58-75.
- Habibi tabar, J. (2009). "Divorce and the Family Support Bill: With an emphasis on the prevention of divorce." Women's Strategic Studies 11(43): 35-64.
- Khajeh-Kazemi, R., B. Golestan, K. Mohammad, M. Mahmoudi, S. Nedjat and M. Pakravan (2011). "Comparison of generalized estimating equations and quadratic inference functions in superior versus inferior Ahmed glaucoma valve implantation." Journal of research in medical sciences: the official journal of Isfahan University of Medical Sciences 16(3): 235.
- Kleinbaum, D. G., M. Klein, D. G. Kleinbaum and M. Klein (2012). "Competing risks survival analysis." Survival Analysis: A self-learning text: 425-495.
- Kumar, G., M. R. Rahman, A. Rajverma and A. K. Misra (2024). "Predicting systemic risk of banks: a machine learning approach." Journal of Modelling in Management 19(Ratkowsky et al.): 441-469.
- Lee, T.-S. and I.-F. (Chen, 2016) (2005). "A two-stage hybrid credit scoring model using artificial neural networks and multivariate adaptive regression splines." Expert Systems with applications 28(Ratkowsky et al.): 743-752.
- Lessmann, S., B. Baesens, H.-V. Seow and L. C. Thomas (2015). "Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research." European journal of operational research 247(Ratkowsky et al.): 124-136.
- Lundberg, S. M. and S.-I. Lee (2017). "A unified approach to interpreting model predictions." Advances in neural information processing systems 30.
- Moscato, V., A. Picariello and G. Sperli (2021). "A benchmark of machine learning approaches for credit score prediction." Expert Systems with Applications 165: 113986.
- Ribeiro, M. T., S. Singh and C. Guestrin (2016). "Model-agnostic interpretability of machine learning." arXiv preprint arXiv:1606.05386.
- Ribeiro, M. T., S. Singh and C. Guestrin (2016). "Why should i trust you?" Explaining the predictions of any classifier. Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining.
- Serrano-Cinca, C. and B. Gutiérrez-Nieto (2016). "The use of profit scoring as an alternative to credit scoring systems in peer-to-peer (P2P) lending." Decision Support Systems 89: 113-122.
- Tobias, A. and M. K. Brunnermeier (2016). "CoVaR." The American Economic Review 106(Ratkowsky et al.): 1705.
- Tripathi, D., A. K. Shukla, B. R. Reddy, G. S. Bopche and D. Chandramohan (2022). "Credit scoring models using ensemble learning and classification approaches: a comprehensive survey." Wireless Personal Communications 123(Ratkowsky et al.): 785-812.
- Vallee, B. and Y. Zeng (2019). "Marketplace lending: A new banking paradigm?" The Review of Financial Studies 32(Ratkowsky et al.): 1939-1982.
- Vidal, M. F. and F. Barbon (2019). "Credit scoring in financial inclusion." Washington: CGAP.
- West, D. (2000). "Neural network credit scoring models." Computers & operations research 27(11-12): 1131-1152.
- Yobas, M. B., J. N. Crook and P. Ross (2000). "Credit scoring using neural and evolutionary techniques." IMA Journal of Management Mathematics 11(Ratkowsky et al.): 111-125.
- حیدریوسف، ه.، پ. پیری و پ. چالاکي (۱۴۰۴). "مدل آینده‌نگری به‌کارگیری هوش مصنوعی در گزارشگری مالی با تأکید بر سودمندی درک‌شده و سهولت استفاده‌شده." تحقیقات مالی ۷۹(۲۷): ۶۶۲-۶۸۴.

Chen, T. (2016). XGBoost: A Scalable Tree Boosting System. *Cornell University*.

Ratkowsky, D., Lowry, R., McMeekin, T., Stokes, A., & Chandler, R. (1983). Model for bacterial culture growth rate throughout the entire biokinetic temperature range. *Journal of bacteriology*, 154(3), 1222-1226.