



<https://amf.ui.ac.ir>

Journal of Asset Management and Financing

E-ISSN: 2383-1189

Vol. 13, Issue 4, No. 51, 2026, p 77-96

Received: 12/10/2024

Accepted: 14/04/2025

Research Paper

Predicting the Likelihood of Operational Risk Occurrence in the Banking Industry Using Machine Learning Algorithms

Hamed Naderi

Ph.D. Candidate, Department of Industrial Engineering, Faculty of Industrial & Systems Engineering, Tarbiat Modares University, Tehran, Iran
h.naderi@modares.ac.ir

Mohamad Ali Rastegar Sorkhe * 

Assistant Professor, Department of Systems and Productivity Management, Faculty of Industrial & Systems Engineering, Tarbiat Modares University, Tehran, Iran
ma_rastegar@modares.ac.ir

Bakhtiar Ostadi

Associate Professor, Department of Systems and Productivity Management, Faculty of Industrial & Systems Engineering, Tarbiat Modares University, Tehran, Iran
bostadi@modares.ac.ir

Mehrdad Kargari

Associate Professor, Department of Information Technology Engineering, Faculty of Industrial & Systems Engineering, Tarbiat Modares University, Tehran, Iran
m_kargari@modares.ac.ir

Abstract

This study investigates and predicts the likelihood of operational risk occurrence in the banking industry using machine learning algorithms. The primary objective is to analyze operational risk data and evaluate the performance of various machine learning models to develop effective tools for enhancing risk management and minimizing financial losses in banks and financial institutions. Operational risk data were collected, pre-processed, and then used for predictions with machine learning models, including Random Forest (RF), Decision Tree (DT), Support Vector Machine (SVM), Logistic Regression (LR), Naïve Bayes (NB), and k-Nearest Neighbors (KNN). Model performance was assessed using evaluation metrics such as accuracy, precision, recall, F1-score, and the Area Under the Curve (AUC) to determine the most effective model for risk prediction. The findings indicate that the RF and SVM algorithms outperform other models in predicting operational risk across all scenarios. Furthermore, the results demonstrate the strong predictive capability of machine learning algorithms in assessing operational risk, highlighting their potential as valuable decision-making tools for risk management in the banking sector.

Keywords: Risk Prediction, Operational Risk, Risk Management, Machine Learning

Introduction

Operational risk is defined as the risk arising from external factors or failures in internal controls or information systems, which may lead to both anticipated and unexpected losses (Crouchy et al., 1998). Lopez (2002) characterizes it as any unquantifiable risk that a bank may encounter. According to the Basel II Agreement, operational risk refers to the probability of loss resulting from deficiencies, breakdowns, or inefficiencies in human resources, processes, technologies, infrastructure, or internal and external events (Pena et al., 2018).

To estimate the capital required to cover operational risk, the Basel framework introduces three approaches: the Basic Indicator Approach (BIA), the Standardized Approach (SA), and the Advanced Measurement Approach (AMA) (Mora Valencia, 2010; Mora Valencia et al., 2017). The BIA and SA estimate capital requirements based on annual gross income, with the key distinction being that the SA categorizes a bank's activities into eight business lines. Under the BIA, an alpha coefficient (α) of 15% is applied, whereas in the SA, each business line has a specific beta coefficient (β) ranging between 12% and 18%. The AMA employs both quantitative

*Corresponding author

Naderi, H., Rastegar Sorkhe, M. A., Ostadi, B. and Kargari, M. (2026). Predicting the likelihood of occurrence of operational risk in the banking industry using machine learning algorithms. *Journal of Asset Management and Financing*, 13(4), 77-96.



and qualitative methods for operational risk modeling, leveraging databases to collect statistical data and utilizing the loss distribution approach (LDA) to model frequency and severity distributions. Capital coverage is then determined based on the cumulative distribution of these variables. Since the LDA is data-driven, the Basel framework (BCBS, 2004) emphasizes the necessity of a robust database for collecting operational risk data. Four key databases are required: internal loss event data, external loss event data, scenario-based analysis data, and a database of business environment and internal control factors.

Compared to other banking risks, such as credit and market risks, measuring, monitoring, and managing operational risk is considerably more complex. This risk has gained increasing attention in recent years, as large operational losses have led to the liquidation of financial institutions (Abdymomunov et al., 2020; Afonso et al., 2019). Crisanto and Perino (2017) identify cyber threats and cyber fraud as critical factors influencing operational risk capital estimation. These risks have intensified with the growth of electronic banking services and include illegal access, system disruptions, and the misuse or theft of digital assets for financial gain (BCBS, 2016; Drew & Farrell, 2018). To quantify potential losses in electronic banking transactions, Bouveret (2018) proposed a Bayesian Network (BN) model to estimate operational risk capital requirements in financial institutions.

Machine learning has emerged as one of the most promising yet challenging approaches in modern finance (Tsai & Wu, 2008). These methods have transformed the financial industry, with deep learning (DL) being extensively studied and applied due to its adaptability and predictive capabilities (Ivanov, 2019). Pena et al. (2021) employed a fuzzy convolutional deep learning model to estimate the maximum operational risk value at a 99.9% confidence level. Similarly, Zhou et al. (2020) utilized semi-supervised machine learning algorithms to classify operational risks based on financial news, analyzing 5,843 documents from financial articles and newspapers in the Asia-Pacific region between February and March 2019. Their model demonstrated the capability to predict various types of risks in the banking industry. In another study, Akbari and Yazdani (2023) applied machine learning algorithms to determine optimal thresholds for operational loss severity data, classifying the data and estimating the capital required to cover operational risk by integrating severity and frequency distribution functions with Monte Carlo simulation.

Method and Data

In this study, operational risk data were collected, pre-processed, and then used for predictions with machine learning models, including RF, DT, SVM, LR, NB, and KNN. The models' performance was assessed using evaluation metrics such as accuracy, precision, recall, F1-score, and AUC to identify the most effective model for predicting the likelihood of risk occurrence.

Findings

The results indicate that the RF and SVM algorithms exhibit strong performance in predicting operational risk across all scenarios. Specifically, the RF algorithm achieved an accuracy of 0.9690, while the SVM algorithm attained an accuracy of 0.9587 in State 1, making them the most effective models in this setting. Both algorithms demonstrated comparable performance across other modes.

Conclusion and Discussion

This study analyzes and predicts operational risk occurrence in the banking industry using machine learning algorithms. The findings indicate that various algorithms, particularly RF and SVM, demonstrate strong predictive performance. These results have the potential to transform operational risk management in banks, leading to significant reductions in associated costs and losses.

A key insight from this study is that leveraging large and diverse datasets can substantially enhance prediction accuracy. Machine learning models can process complex datasets, identify hidden patterns, and facilitate early risk detection, enabling banks to implement preventive measures before risks materialize. Moreover, integrating machine learning into risk management enhances decision-making by providing precise, data-driven predictions, allowing for more effective strategies and efficient resource allocation.

Future research could incorporate additional data, such as historical records, economic indicators, and internal process information, to further improve prediction accuracy. With advancements in technology, more sophisticated techniques—such as reinforcement learning methods (e.g., DQN, Q-Learning, DDPG, and Meta-Learning)—could enhance the accuracy and efficiency of operational risk prediction models.

مقاله پژوهشی

پیش‌بینی احتمال وقوع ریسک عملیاتی در صنعت بانکداری با استفاده از الگوریتم‌های یادگیری ماشین

حامد نادری

دانشجوی دکتری، مهندسی صنایع، گروه مهندسی صنایع، دانشکده مهندسی صنایع و سیستم‌ها، دانشگاه تربیت مدرس، تهران، ایران

h.naderi@modares.ac.ir

محمدعلی رستگار سرخه* 

استادیار، گروه مدیریت سیستم و بهره‌وری، دانشکده مهندسی صنایع و سیستم‌ها، دانشگاه تربیت مدرس، تهران، ایران

ma_rastegar@modares.ac.ir

بختیار استادی

دانشیار، گروه مدیریت سیستم و بهره‌وری، دانشکده مهندسی صنایع و سیستم‌ها، دانشگاه تربیت مدرس، تهران، ایران

bostadi@modares.ac.ir

مهرداد کارگری

دانشیار، گروه مهندسی فناوری اطلاعات، دانشکده مهندسی صنایع و سیستم‌ها، دانشگاه تربیت مدرس، تهران، ایران

m_kargari@modares.ac.ir

چکیده

این پژوهش با هدف پیش‌بینی احتمال وقوع ریسک عملیاتی در صنعت بانکداری با استفاده از الگوریتم‌های یادگیری ماشین انجام شده است. پژوهش حاضر با تحلیل داده‌های ریسک عملیاتی و ارزیابی عملکرد الگوریتم‌های یادگیری ماشین به منظور ارائه الگوریتم‌هایی مؤثر برای پیش‌بینی دقیق‌تر احتمال وقوع ریسک عملیاتی و مدیریت بهتر آن در صنعت بانکداری صورت گرفته است. در این پژوهش، داده‌های مرتبط با ریسک عملیاتی از سال ۱۳۹۵ تا ۱۴۰۲ جمع‌آوری و پیش‌پردازش شد و سپس با استفاده از مدل‌های یادگیری ماشین مانند RF، DT، SVM، LR، NB و KNN پیش‌بینی انجام شد. عملکرد مدل‌ها با معیارهایی همچون دقت، صحت، بازخوانی، F1-score و AUC ارزیابی شد تا بهترین مدل برای پیش‌بینی احتمال وقوع ریسک انتخاب شود. نتایج نشان می‌دهند که الگوریتم‌های RF و SVM در پیش‌بینی احتمال وقوع ریسک عملیاتی در تمامی حالت‌ها عملکرد بسیار خوبی دارند؛ به علاوه که الگوریتم‌های یادگیری ماشین توانایی بالایی در پیش‌بینی وقوع ریسک عملیاتی دارند و می‌توانند ابزار مؤثری برای تصمیم‌گیری‌های مدیریتی در صنعت بانکداری فراهم کنند.

کلیدواژه‌ها: پیش‌بینی ریسک، ریسک عملیاتی، مدیریت ریسک، یادگیری ماشین

طبقه‌بندی Jel: G2، G21، G32

* نویسنده مسئول

نادری، حامد، رستگار سرخه، محمدعلی، استادی، بختیار و کارگری، مهرداد. (۱۴۰۴). پیش‌بینی احتمال وقوع ریسک عملیاتی در صنعت بانکداری با استفاده از الگوریتم‌های یادگیری ماشین، مدیریت دارایی و تأمین مالی، ۱۳ (۴)، ۷۷-۹۶.



مقدمه

امروزه ریسک عملیاتی به عنوان یکی از ریسک‌های کلیدی در صنعت بانکداری شناخته می‌شود. توجه به این نوع ریسک پس از وقوع بحران‌های مالی جهانی افزایش یافته است. در صنعت بانکداری، اقدامات متعددی به منظور مدیریت ریسک عملیاتی انجام شده است (González-Carrasco et al., 2019; Hoffman, 2002). طبق توافق‌نامهٔ بال ۲، ریسک عملیاتی به معنای احتمال متحمل شدن زیان ناشی از وقایع داخلی مانند نقص‌ها و نارسایی‌ها در فرایندها، سیستم‌ها، افراد و وقایع خارجی است (Bank for International Settlements, 2016). به منظور کاهش پیامدهای منفی ریسک عملیاتی، مؤسسات مالی باید به مدیریت این ریسک توجه ویژه‌ای داشته باشند. کمیتهٔ بال ریسک عملیاتی را به هفت دسته تقسیم کرده است که شامل تقلب داخلی، تقلب خارجی، روابط کار، مشتریان، آسیب به دارایی‌های ثابت، خرابی‌های فناوری و نارسایی در اجرای فرایندها و مدیریت آنها است (Elarif & Hinti, 2014).

زیان‌های درخور توجهی که مؤسسات مالی و غیرمالی از ناحیهٔ فرایندها و عوامل غیراعتباری و غیربازاری متحمل شده‌اند، سبب جلب توجه مدیران و تصمیم‌گیران به حوزهٔ ریسک عملیاتی شده است. مدیریت ریسک عملیاتی یکی از مباحث برجسته در صنعت بانکداری است که تأثیر چشمگیری بر عملکرد بانک‌ها و مؤسسات مالی دارد. بررسی‌های انجام‌شده نشان می‌دهد که پژوهشگران به زوایای مختلف این ریسک پرداخته‌اند. برخی بر تعریف ریسک، برخی بر طبقه‌بندی رویدادهای مرتبط، برخی بر اندازه‌گیری و ویژگی‌های مدیریت ریسک و گروهی دیگر بر تحلیل‌های مقایسه‌ای تأکید دارند (Barakat & Hussainey, 2013). ریسک عملیاتی زمانی مطرح می‌شود که یک سازمان نتواند عملیات خود را به درستی انجام دهد. ادامهٔ این وضعیت می‌تواند به افت عملکرد سازمان و کاهش نرخ بازده سرمایه‌گذاری منجر شود (Basel Committee on Banking Supervision, 2006). به طور کلی، ریسک عملیاتی به دلیل مجموعه‌ای از احتمالات بروز خطا و نقص در عملیات خاص بنگاه تجاری یا مالی به وجود می‌آید (Wang & Hsu, 2013).

مدیریت ریسک عملیاتی یکی از چالش‌های مدیریت ریسک است که صنعت بانکداری با آن روبه‌رو است؛ بنابراین، پژوهش حاضر با هدف رفع چالش مدیریت ریسک عملیاتی انجام شده است. سؤال اصلی پژوهش این است که چگونه می‌توان با استفاده از داده‌های ریسک عملیاتی بانک و الگوریتم‌های یادگیری ماشین، احتمال وقوع ریسک عملیاتی را پیش‌بینی کرد و بهترین الگوریتم برای این پیش‌بینی کدام است؛ بنابراین، این سؤال شامل دو بخش پیش‌بینی احتمال وقوع ریسک عملیاتی براساس داده‌های ریسک عملیاتی و ارزیابی و انتخاب بهترین الگوریتم یادگیری ماشین برای پیش‌بینی احتمال وقوع ریسک عملیاتی است. پژوهش حاضر رویکردی جدید برای مدیریت ریسک عملیاتی ارائه می‌دهد و این پژوهش به دنبال طراحی مدلی است که بتواند احتمال وقوع ریسک عملیاتی را پیش‌بینی کند.

مبانی نظری

ریسک عملیاتی به عنوان ریسکی تعریف می‌شود که از عوامل خارجی یا نقص در کنترل‌های داخلی یا سیستم‌های اطلاعاتی ناشی می‌شود و می‌تواند به زیان منجر شود؛ خواه این زیان پیش‌بینی‌شده باشد یا غیرمنتظره (Crouchy et al., 1998). توافق‌نامهٔ بال ۲، ریسک عملیاتی را به عنوان احتمال زیان ناشی از کمبودها، خرابی‌ها یا نارسایی‌ها در منابع انسانی، فرایندها، فناوری‌ها، زیرساخت‌ها یا رویدادهای داخلی و خارجی تعریف می‌کند (Pena et al., 2018). پسالچک (Posavljak, 2024) تعریف ریسک عملیاتی را در قالب چالش‌های زیرساختی و پشتیبانی از تصمیم‌گیری در مدیریت دارایی‌ها بیان می‌کند. ریسک عملیاتی در بانکداری به عنوان تهدیدات مرتبط با فرایندهای روزمره شامل امنیت زیرساخت‌ها، دقت داده‌ها و سیاست‌های

مدیریتی تعریف می‌شود. ریسک عملیاتی شامل اختلالات ناشی از دیجیتالی‌شدن در فرایندهای بانکی، اتوماسیون و چالش‌های نظارتی مرتبط با فناوری مالی تعریف شده است (Rahman et al., 2024).

براساس توافق‌نامه^۱ بال، برای تخمین سرمایه لازم برای پوشش ریسک‌های عملیاتی، سه رویکرد شاخص پایه، استاندارد و اندازه‌گیری پیشرفته معرفی شده است (Mora-Valencia, 2010; Mora-Valencia & Zapata-Jaramillo, 2017). رویکرد شاخص پایه و استاندارد براساس درآمد ناخالص سالیانه سرمایه پوششی را تخمین می‌زنند، با این تفاوت که در رویکرد استاندارد، فعالیت‌های بانک به ۸ خط کسب‌وکار^۲ تقسیم می‌شود. در رویکرد شاخص پایه از ضریب آلفا (α) ۱۵ درصد استفاده می‌شود؛ اما در رویکرد استاندارد هر خط کسب‌وکار ضریب بتا (β) مربوط به خود را دارد که بین ۱۲ تا ۱۸ درصد است. رویکرد اندازه‌گیری پیشرفته شامل روش‌های کمی و کیفی برای مدل‌سازی ریسک عملیاتی است. این روش شامل استفاده از پایگاه داده‌هایی برای جمع‌آوری داده‌های آماری لازم است و از روش توزیع زیان برای برازش توزیع‌های فراوانی و شدت استفاده می‌کند. سرمایه پوششی براساس توزیع تجمعی فراوانی و شدت محاسبه می‌شود. روش توزیع زیان روشی داده‌محور است و برای محاسبه ریسک عملیاتی نیاز به داده دارد؛ بنابراین، طبق توافق‌نامه^۳ بال ۲ نیاز به پایگاه داده برای جمع‌آوری داده لازم وجود دارد. برای جمع‌آوری داده‌های ریسک عملیاتی، چهار پایگاه داده لازم است: پایگاه داده داخلی^۴، پایگاه داده خارجی^۵، پایگاه داده تجزیه و تحلیل مبتنی بر سناریو^۴ و پایگاه داده عوامل محیط کسب‌وکار و کنترل داخلی^۵ (Bank for International Settlements, 2004).

برای مدیریت اثربخش ریسک‌های عملیاتی، درک دقیق ابعاد و مؤلفه‌های آن امری ضروری است. (Naderi & Rastegar, 2022) در پژوهشی که با استفاده از روش فراترکیب انجام شده است، چارچوبی جامع برای اجزای مدیریت ریسک عملیاتی در صنعت بانکداری ارائه داده‌اند. در این پژوهش با تحلیل و ترکیب یافته‌های مطالعات پیشین، جنبه‌های مختلف این نوع ریسک شناسایی و طبقه‌بندی شده است. این چارچوب به مدیران بانکی کمک می‌کند تا دیدگاهی ساختاریافته به مدیریت ریسک عملیاتی داشته باشند و اقدامات مناسب‌تری برای کاهش اثرات آن طراحی کنند. ریسک‌های عملیاتی در نظام بانکی اسلامی از موضوعات پرکاربرد در پژوهش‌های مالی است؛ در این راستا، مطالعات مختلفی انجام شده است که این نوع ریسک‌ها را تفکیک و تحلیل کرده‌اند. (Talebi et al., 2011) در مطالعه‌ای توصیفی-تحلیلی، ریسک‌های عملیاتی در بانکداری اسلامی را به دو گروه عمده شامل ریسک‌های مشترک با نظام بانکی متعارف و ریسک‌های ویژه بانکداری اسلامی دسته‌بندی کردند. یافته‌های آنها بر اهمیت تفکیک این دو نوع ریسک در مدیریت ریسک عملیاتی تأکید دارد. (Khoshsima & Shehikdash, 2013) نیز با بررسی رابطه میان کارایی بانک‌ها و ریسک‌های اعتباری، عملیاتی و نقدینگی از روش‌های پارامتریک و ناپارامتریک استفاده کردند. نتایج نشان داد که ریسک‌های مذکور اثر معناداری بر کارایی بانک‌های ایران دارند و نیازمند تحلیل دقیق‌تر و مدیریت جامع‌تری هستند. (Nosrati & Pakizeh, 2014) با بهره‌گیری از رویکرد اندازه‌گیری پیشرفته، سرمایه لازم را برای ریسک عملیاتی برآورد کردند. این پژوهش بر مبنای روش توزیع زیان از مدل‌های کلاسیک و توزیع‌های دنباله پهن مانند توزیع آلفای پایدار استفاده کرد. آنها موفق شدند چارچوبی کاربردی برای محاسبه ذخایر سرمایه در بانک‌های ایرانی ارائه دهند. (Ostadi et al., 2018) در مطالعه‌ای، ارزیابی ریسک عملیاتی را با استفاده از روش‌های مبتنی بر استنتاج بیزی بررسی کردند. تمرکز اصلی پژوهش آنها بر ترکیب داده‌های داخلی و نظرات کارشناسی برای تخمین پارامترهای توزیع زیان بود. نتایج نشان داد که با افزایش دوره‌های پیش‌بینی، مقادیر پارامترهای زیان کاهش یافته و نشان‌دهنده کاهش ریسک با گذشت

¹ Business line

² Internal Data

³ External Data

⁴ Scenario-based analysis data

⁵ Business Environment and Internal Control Factors

زمان است. (Mostafaei et al., 2019) با استفاده از تکنیک نگاشت شناختی فازی، ریسک‌های عملیاتی را شناسایی و تحلیل کردند. آنها با ارائه یک چارچوب نگاشت مبتنی بر فرایند، راهکاری برای شناسایی و مدیریت ساختاریافته ریسک‌های عملیاتی پیشنهاد کردند. این مطالعه به‌طور خاص بر قابلیت نگاشت شناختی در درک و تحلیل ریسک‌های عملیاتی تأکید دارد.

اندازه‌گیری، نظارت و مدیریت این ریسک در مقایسه با سایر ریسک‌های بانکی مانند ریسک‌های اعتباری و بازار بسیار دشوار است. این ریسک به‌خصوص برای سازمان‌های بانکی مهم است و در سال‌های اخیر به آن توجه شده است؛ زیرا زیان‌های عملیاتی بزرگ می‌تواند منجر به انحلال مؤسسات مالی شود (Abdymomunov et al., 2020; Afonso et al., 2019). کریسانتو و پرنیو (Crisanto & Prenio, 2017) منابعی را شناسایی کرده‌اند که تهدیدات سایبری و کلاهبرداری سایبری می‌تواند بر فرایند تخمین سرمایه ریسک عملیاتی تأثیر بگذارد. این جرائم که با معرفی خدمات بانکداری الکترونیک بیشتر شده است، شامل دسترسی غیرقانونی، اختلال سیستم و سوءاستفاده یا دزدی از دستگاه‌ها برای دستیابی به مزیت مالی است (Bank for International Settlements, 2016; Drew & Farrell, 2018). برای تعیین کمیت زیان‌های احتمالی در معاملات بانکداری الکترونیکی، مدل شبکه بیزی^۱ (BN) برای تخمین سرمایه پوششی ریسک عملیاتی در شرکت‌های مالی طراحی شده است (Bouveret, 2018). یادگیری ماشین یکی از امیدوارکننده‌ترین و چالش‌برانگیزترین رویکردهای کلیدی در امور مالی مدرن است (Tsai & Wu, 2008). این روش‌ها نحوه کار صنعت مالی را تغییر داده است و یادگیری عمیق^۲ (DL) به دلیل تطبیق‌پذیری و قابلیت‌های پیش‌بینی، بیشترین پژوهش و کاربرد را دارد (Ivanov et al., 2019). پنا و همکاران (Pena et al., 2021) با استفاده از مدل یادگیری عمیق کانولوشنال فازی، به دنبال محاسبه حداکثر ارزش در معرض خطر ریسک عملیاتی با سطح اطمینان ۹۹/۹ درصد بوده‌اند. (Zhou et al., 2021) با استفاده از الگوریتم‌های یادگیری ماشین نیمه‌نظارت‌شده، سعی در طبقه‌بندی ریسک‌های عملیاتی با توجه به اخبار مالی داشته‌اند. مدل ارائه‌شده قادر به پیش‌بینی انواع ریسک‌های صنعت بانکداری است. اکبری و یزدانیان (Akbari & Yazdani, 2023) از الگوریتم‌های یادگیری ماشین برای تخمین آستانه مناسب برای داده‌های شدت زیان عملیاتی و طبقه‌بندی این داده‌ها استفاده کرده و سرمایه لازم برای پوشش ریسک عملیاتی با تجمیع تابع توزیع شدت و فراوانی داده‌ها و شبیه‌سازی مونت کارلو را تخمین زده‌اند.

با بررسی پژوهش‌های گذشته، مدیریت ریسک عملیاتی یکی از چالش‌های صنعت بانکداری است. برای مدیریت بهتر این ریسک نیاز به بهبود فرایندهای مدیریت ریسک عملیاتی وجود دارد. پژوهش‌های گذشته بیشتر بر تخمین سرمایه لازم برای پوشش ریسک عملیاتی تمرکز داشته‌اند؛ اما پژوهشی مبنی بر پیش‌بینی احتمال وقوع ریسک عملیاتی انجام نشده است. پیش‌بینی احتمال وقوع ریسک عملیاتی نیز می‌تواند فرایند مدیریت ریسک را بهبود بخشد. در این پژوهش، با استفاده از الگوریتم‌های یادگیری ماشین مدیریت ریسک عملیاتی بهبود داده شده است.

روش پژوهش

در این پژوهش، داده‌ها از بانک پارسیان و از طریق سامانه مدیریت ریسک سمر جمع‌آوری شده‌اند. مجموعه داده مربوط به سال‌های ۱۳۹۵ تا ۱۴۰۲ است و شامل ۴۲۱۳ رکورد و ۱۰ ویژگی است. این داده‌ها براساس چارچوب کمیته بال در نظارت بانکی^۳ (BCBS) جمع‌آوری شده است (Pereira et al., 2018; Bank for International Settlements, 2004). نمره احتمال وقوع ریسک براساس چارچوب سند کوزو (COSO, 2017) داده شده است. ابتدا پیش‌پردازش بر روی مجموعه داده اصلی انجام شد و مجموعه داده تمیز شد. در جدول ۱ اطلاعات مربوط به مجموعه داده بیان شده است.

¹ Bayesian Network

² Deep Learning

³ Basel Committee on Banking Supervision

جدول (۱) مجموعه داده

Table (1) Dataset

ردیف	ویژگی	نوع ویژگی	توضیحات
۱	خط کسب‌وکار	کیفی	نوع خط کسب‌وکار بانک مطالعه‌شده
۲	عنوان اصلی	کیفی	عنوان اصلی ریسک در هر خط کسب‌وکار
۳	عنوان فرعی	کیفی	عنوان فرعی ریسک در هر خط کسب‌وکار
۴	فرایند اصلی	کیفی	فرایندهای اصلی که در بانک مطالعه‌شده انجام می‌شود.
۵	فرایند فرعی	کیفی	فرایندهای فرعی که در بانک مطالعه‌شده انجام می‌شود.
۶	عامل وقوع رویداد	کیفی	عواملی که باعث وقوع ریسک شده است.
۷	نوع رویداد	کیفی	رویدادهای که رخ داده است.
۸	مالک ریسک	کیفی	مالک ریسک که در واحد مربوط رخ داده است.
۹	محل وقوع ریسک	کیفی	محل وقوع ریسک در آن رخ داده است.
۱۰	احتمال وقوع ریسک	کمی	نمره احتمال وقوع ریسک

طبق استاندارد کوزو ۲۰۱۷ احتمال وقوع ریسک عملیاتی نمره‌دهی شده است. در جدول ۲ نمره احتمال وقوع ریسک مشخص شده است.

جدول (۲) احتمال وقوع ریسک

Table (2) Risk occurrence likelihood

نمره	احتمال
۱	نادر
۲	بعید
۳	ممکن
۴	محتمل
۵	تقریباً قطعی

از ۶ الگوریتم پرکاربرد یادگیری ماشین استفاده شده است. (۱) k -نزدیکترین همسایه^۱ (KNN) یک الگوریتم یادگیری ماشین است که به‌طور گسترده برای مسائل طبقه‌بندی و رگرسیون استفاده می‌شود. فرض نزدیکی نقاط مشابه اساس KNN است. از این‌رو، ابتدا مقدار k تعریف می‌شود که تعداد همسایگان به دنبال اندازه‌گیری فاصله اقلیدسی در بین آنها است (Poongodi et al., 2023؛ ۲) بیز ساده^۲ (NB) یک الگوریتم طبقه‌بندی ساده، اما مؤثر و متداول یادگیری ماشین است که در دسته یادگیری با ناظر^۳ جای می‌گیرد. بیز ساده الگوریتمی احتمالی است که براساس نظریه بیز برای طبقه‌بندی استفاده می‌شود. مفهوم اصلی حداقل مربعات زیربنای یک طبقه‌بندی باینری، یعنی دو طبقه و چند طبقه است. این تکنیک برای طبقه‌بندی باینری استفاده می‌شود (Liu et al., 2017؛ ۳) جنگل تصادفی^۴ (RF) مدل طبقه‌بندی و رگرسیون مبتنی بر مجموعه است. از جنگل تصادفی می‌توان برای انتخاب ویژگی نیز استفاده کرد. یک الگوریتم یادگیری ماشین است که براساس درخت تصمیم است. از روش دسته‌بندی برای ایجاد زیرمجموعه‌های متعدد داده با انتخاب تصادفی مشاهدات از مجموعه داده‌های اصلی با

¹ K-Nearest Neighbor² Naive Bayes³ Supervised Learning⁴ Random Forest

جایگزینی استفاده می‌شود و با استفاده از این زیرمجموعه‌ها درخت‌های کوچک‌تری می‌سازد. سپس، تمام این درختان را به صورت موازی اجرا می‌کند و پیش‌بینی نهایی با میانگین‌گیری پیش‌بینی از همه درختان تصمیم محاسبه می‌شود. برای بهبود دقت پیش‌بینی در جنگل تصادفی، سه پارامتر حیاتی هستند: تعداد درختان، حداکثر تعداد ویژگی‌هایی که برای تقسیم یک گره در نظر گرفته می‌شوند و حداقل تعداد برگ‌های لازم برای تقسیم یک گره داخلی است (Simaiya et al., 2021b؛ ۴) رگرسیون لجستیک^۱ (LR) روش دیگری است که یادگیری ماشینی از حوزه آمار به دست آورده است. عمدتاً برای مسائل طبقه‌بندی براساس ویژگی‌های دو کلاسه ترجیح داده می‌شود. این یک تکنیک آماری است که داده‌های دریافتی حاصل از یافته‌های قبلی از مجموعه داده‌های داده‌شده را پیش‌بینی می‌کند. روش رگرسیون لجستیک با مطالعه ارتباط بین یک یا چند عامل پیش‌بینی از پیش موجود، پارامترهای داده‌های متعدد را پیش‌بینی می‌کند (Simaiya et al., 2021a؛ ۵) درخت تصمیم^۲ (DT) مدل یادگیری ماشین است که برای طبقه‌بندی داده‌ها استفاده می‌شود. این روش با استفاده از یک سری قوانین مدلی غیرخطی بر روی داده‌های آموزشی ایجاد می‌کند. درخت تصمیم یکی از رایج‌ترین انواع مدل‌ها است. این فقط یک ساختار فلوچارت‌مانند با شبکه‌ای است که آزمایشی را روی یک تابع نشان می‌دهد. طبقه‌بندی درخت تصمیم، همان‌طور که قبلاً گفته شد، طبقه‌بندی‌ها را به زیرمجموعه‌ها تقسیم می‌کند و یکی از پرکاربردترین مدل‌های یادگیری ماشین است (Jabbar et al., 2016؛ ۶) ماشین بردار پشتیبان^۳ (SVM) مبتنی بر طبقه‌بندی نظارت‌شده است که می‌تواند برای یافتن راه‌حلی برای وظایف مختلف رگرسیون و طبقه‌بندی استفاده شود و بیشتر برای غلبه بر وظایف رگرسیون و طبقه‌بندی استفاده می‌شود. این روش به عنوان یک مرحله گذار در یک همسایگی n بعدی تحلیل می‌شود و n نشان‌دهنده تعداد ویژگی‌ها است (Garg et al., 2021).

در رویکردهای یادگیری ماشین نیاز است تا ارتباط بین متغیرهای مسئله مشخص شود و سعی شده است با استفاده از ماتریس همبستگی^۴ این ارتباط مشخص شود. در ماتریس همبستگی متغیرها همان ویژگی‌های^۵ مجموعه داده هستند. این ماتریس ۱۰ سطر و ۱۰ ستون دارد و مقارن است. روش‌های مختلفی برای اندازه‌گیری ضریب همبستگی^۶ وجود دارد. روش اول که یک روش پارامتری است، ضریب همبستگی پیرسون^۷ و روش دوم نیز که روشی ناپارامتری است، ضریب همبستگی اسپیرمن^۸ نامیده می‌شوند. در شکل ۱ ماتریس همبستگی اسپیرمن رسم شده است. باتوجه به نتایج ضریب همبستگی با روش اسپیرمن مشخص شد که هیچ همبستگی مثبت و منفی معناداری بین متغیرها وجود ندارد. در تحلیل همبستگی، واحدهای اندازه‌گیری مربوط به ریسک‌های ثبت‌شده در بانک پارسیان طی این سال‌ها هستند؛ بنابراین، همبستگی بین متغیرها در سطح رکوردهای ثبت‌شده از ریسک‌ها انجام شده است. از آنجاکه داده‌ها مربوط به چندین سال هستند، همبستگی بین رخدادهای مختلف ریسک در بازه‌های زمانی ۱۳۹۵ تا ۱۴۰۲ بررسی شده است.

¹ Logistic Regression

² Decision Tree

³ Support Vector Machine

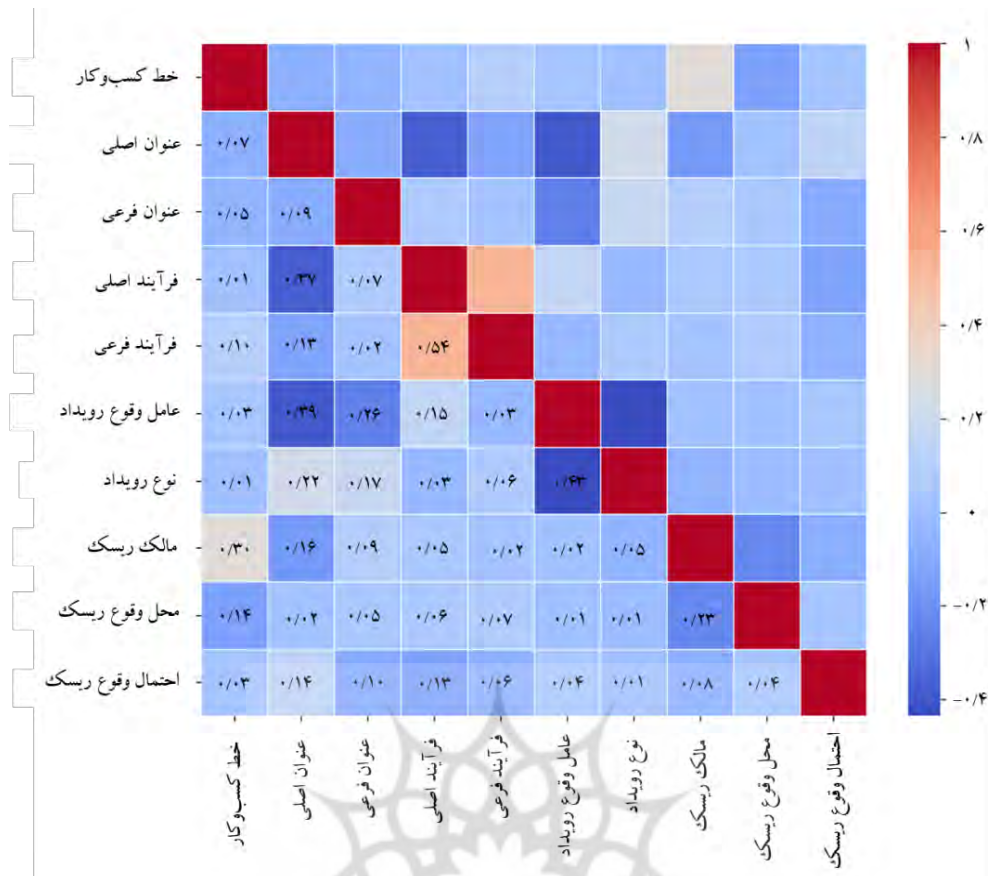
⁴ Correlation Matrix

⁵ Features

⁶ Correlation Coefficient

⁷ Pearson Correlation Coefficient

⁸ Spearman Correlation Coefficient



شکل (۱) ماتریس همبستگی اسپیرمن

Figure (1) Spearman's correlation matrix

پس از ساخت مدل لازم است تا صحت مدل سنجیده شود. فرض کنید مدلی برای پیش‌بینی احتمال وقوع ریسک عملیاتی ساخته شده است. یکی از سؤالات مهمی که در این مرحله مطرح می‌شود، چگونگی ارزیابی عملکرد مدل به دست آمده است. در این مرحله مطمئناً نیاز به برآوردی از صحت پیش‌بینی مدل است؛ بدین معنی که مدل چقدر می‌تواند احتمال وقوع ریسک براساس رویدادهایی که مدل برای آنها، ساخته نشده است درست پیش‌بینی کند. ممکن است مدل‌های مختلفی ساخته شود؛ بنابراین، یکی از سؤالاتی که مطرح می‌شود این است که کدام یک از مدل‌ها عملکرد بهتری در پیش‌بینی احتمال وقوع ریسک خواهد داشت. معمولاً بر روی مجموعه داده آموزشی مدل ساخته می‌شود و با استفاده از داده‌های آزمون مدل ارزیابی می‌شود. معیارهای ارزیابی به این سؤال پاسخ می‌دهند که مدل تا چه حدی قادر است پیش‌بینی صحیحی انجام دهد. معیارهای ارزیابی مدل شامل دقت^۱، صحت^۲، بازخوانی^۳، F1-score و AUC است که دقت و صحت مهم‌ترین این معیارها محسوب می‌شوند. اطلاعات لازم برای محاسبه معیارهای مذکور در یک ماتریس با عنوان ماتریس درهم‌ریختگی^۴ در جدول ۳ نشان داده شده است.

¹ Accuracy² Precision³ Recall⁴ Confusion Matrix

جدول (۳) ماتریس درهم ریختگی

Table (3) Confusion matrix

برچسب پیش بینی			
	مثبت (۱)	منفی (۰)	
FN		TN	منفی (۰)
TP		FP	مثبت (۱)

در اینجا: (۱) مثبت درست (TP)^۱: تعداد نمونه های مثبتی است که به طور درست توسط مدل برچسب گذاری شده است؛ (۲) منفی درست (TN)^۲: تعداد نمونه های منفی است که توسط مدل درست برچسب گذاری شده است؛ (۳) مثبت غلط (FP)^۳: تعداد نمونه های منفی است که توسط مدل به اشتباه به عنوان مثبت برچسب گذاری شده است؛ (۴) منفی غلط (FN)^۴: تعداد نمونه های مثبتی است که توسط مدل به اشتباه برچسب منفی گرفته است. عملکرد مدل ها با معیارهای مختلفی اندازه گیری شده است. معیار دقت^۵ به نمونه هایی اشاره دارد که به درستی از همه نمونه ها شناسایی شده است (Guo et al., 2020). نحوه محاسبه این معیار در رابطه ۱ آمده است:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

صحت و یادآوری توانایی مدل را برای تشخیص دقیق وقوع یک نمونه کلاس مثبت اندازه گیری می کند (Cantarella et al., 2012; Afroz et al., 2023). نحوه محاسبه این دو معیار در رابطه ۲ و ۳ آمده است:

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

F1-Score میانگین هارمونیک یادآوری و صحت یک مدل پیش بینی را تعیین می کند. F1-Score برای مسائل طبقه بندی ای خوب است که برچسب های هدف نامتعادل هستند (Goutte & Gaussier, 2005). نحوه محاسبه این معیار در رابطه ۴ آمده است:

$$F1_Score = \frac{2 * precision * Recall}{precision + Recall} \quad (4)$$

حساسیت^۶ با عنوان نرخ مثبت درست نامیده می شود. این معیار بیانگر نمونه های مثبتی است که درست پیش بینی شده است. نحوه محاسبه این معیار در رابطه ۵ آمده است:

$$Sensitivity = \frac{TP}{TP + FN} \quad (5)$$

خصوصیت^۷ نیز نشان دهنده نمونه های منفی است که درست پیش بینی شده است. این معیار نرخ منفی درست نامیده می شود. نحوه محاسبه این معیار در رابطه ۶ آمده است:

¹ True Positive

² True Negative

³ False Positive

⁴ False Negative

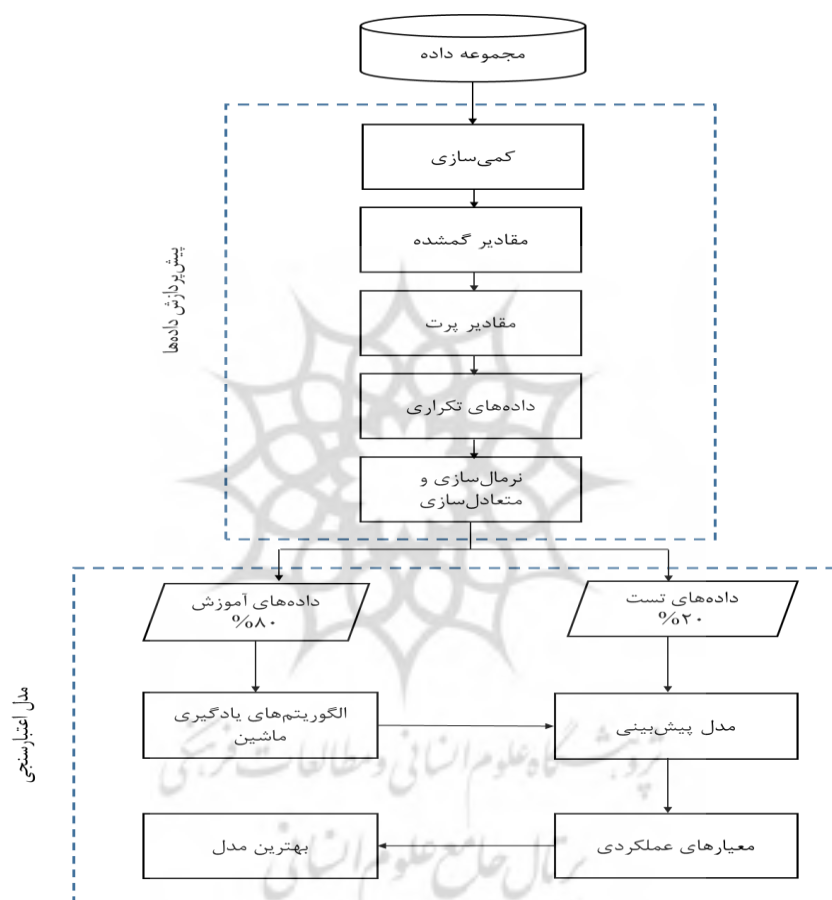
⁵ Accuracy

⁶ Sensitivity

⁷ Specificity

$$Specificity = \frac{TN}{FP + TN} \quad (۶)$$

در شکل ۲ فلوچارت مراحل این پژوهش رسم شده است که ابتدا پیش‌پردازش داده‌ها صورت می‌گیرد که شامل کمی‌سازی داده‌ها، حذف داده‌های تکراری، پرت و خالی است و در ادامه نرمال‌سازی داده‌ها و بالانس کردن کلاس‌ها انجام شده است. در مرحله بعد مجموعه‌های آموزش و آزمون از هم تفکیک شده است و با استفاده از الگوریتم‌های یادگیری ماشین پیش‌بینی احتمال وقوع ریسک صورت می‌گیرد. از معیارهای ارزیابی الگوریتم‌های طبقه‌بندی‌کننده برای ارزیابی مدل استفاده می‌شود و بهترین الگوریتم‌ها انتخاب می‌شوند.



شکل (۲) مدل پیشنهادی

Figure (2) Proposed model

یافته‌ها

برای پیش‌بینی احتمال وقوع ریسک عملیاتی، از داده‌های ریسک عملیاتی و الگوریتم‌های یادگیری ماشین استفاده شده است. متغیرهای ورودی‌های مدل در این پژوهش شامل خط کسب‌وکار، عنوان اصلی ریسک، عنوان فرعی ریسک، فرایند اصلی ریسک، فرایند فرعی ریسک، عامل وقوع رویداد، نوع رویداد، مالک ریسک و محل وقوع ریسک هستند. این ویژگی‌ها به عنوان ورودی‌ها برای پیش‌بینی احتمال وقوع ریسک به مدل یادگیری ماشین داده می‌شوند. هدف پیش‌بینی احتمال وقوع ریسک عملیاتی به عنوان خروجی مدل است که دارای ۵ حالت مختلف است. به منظور ساده‌سازی مسئله و بهبود عملکرد مدل، کلاس‌های خروجی به دو کلاس تبدیل شده است. روش کار به شرح جدول ۴ است:

جدول (۴) خروجی‌های مدل

Table (4) Model outputs

حالت	کلاس ۱	کلاس ۰
حالت ۱	نمره ۱ (نادر)	نمره ۲،۳،۴،۵
حالت ۲	نمره ۲ (بعید)	نمره ۱،۳،۴،۵
حالت ۳	نمره ۳ (ممکن)	نمره ۱،۲،۴،۵
حالت ۴	نمره ۴ (محتمل)	نمره ۱،۲،۳،۵
حالت ۵	نمره ۵ (تقریباً قطعی)	نمره ۱،۲،۳،۴

در این پژوهش، احتمال وقوع ریسک در ۵ سطح قرار دارد که طبق جدول ۲ دارای ۵ نمره است؛ با این حال، استفاده از یک مدل یادگیری ماشین برای پیش‌بینی این ۵ سطح به صورت هم‌زمان چالش‌هایی به همراه دارد؛ از جمله کاهش دقت مدل در تشخیص کلاس‌های مختلف. برای حل این مشکل، روش تبدیل مسئله به چندین طبقه‌بندی دوکلاسه اتخاذ شده است. در این روش، ۵ حالت مختلف تعریف شده است که در هر حالت یکی از ۵ سطح احتمال وقوع ریسک به عنوان کلاس ۱ و سایر سطوح به عنوان کلاس ۰ در نظر گرفته شده‌اند. مدل برای هر یک از این ۵ حالت آموزش داده می‌شود؛ برای مثال:

حالت ۱: کلاس ۱ شامل نمره ۱ (نادر) و کلاس ۰ شامل نمرات ۲، ۳، ۴، ۵ (بعید، ممکن، محتمل، تقریباً قطعی) است. اگر مدل به درستی پیش‌بینی کند، به این معناست که احتمال وقوع ریسک نادر (۱) است؛ در غیر این صورت، ریسک در یکی از سطوح دیگر (۲، ۳، ۴، ۵) قرار دارد.

حالت ۲: کلاس ۱ شامل نمره ۲ (بعید) و کلاس ۰ شامل نمرات ۱، ۳، ۴، ۵ است. در این حالت اگر مدل پیش‌بینی درستی داشته باشد، احتمال وقوع ریسک بعید (۲) است، در غیر این صورت، در یکی از سطوح دیگر (۱، ۳، ۴، ۵) قرار دارد. این روش به مدل اجازه می‌دهد تا برای هر سطح ریسک به طور جداگانه آموزش ببیند و دقت کلی پیش‌بینی افزایش یابد.

به این دلایل، کاهش کلاس‌ها صورت گرفته است: (۱) ساده‌سازی مسئله: با تبدیل مسئله به چندین مسئله دوکلاسه^۱ به جای یک مسئله پنج‌کلاسه پیچیدگی محاسباتی کاهش می‌یابد و مدل‌ها ساده‌تر می‌شوند. این کار به مدل اجازه می‌دهد تا بهتر روی داده‌ها آموزش ببیند و نتایج فهم‌پذیرتری تولید کند؛ (۲) افزایش دقت مدل‌ها: مدل‌های دوکلاسه معمولاً دقت بیشتری در مقایسه با مدل‌های چندکلاسه دارند. با تبدیل مسئله به مجموعه‌ای از مدل‌های دوکلاسه می‌توان به هر کلاس توجه بیشتری کرد و دقت پیش‌بینی‌ها را افزایش داد؛ (۳) قابلیت تعمیم‌دهی بهتر: مدل‌های دوکلاسه معمولاً توانایی تعمیم‌دهی بهتری از مدل‌های چندکلاسه دارند. این امر به ویژه زمانی بسیار مهم است که داده‌ها نامتوازن هستند.

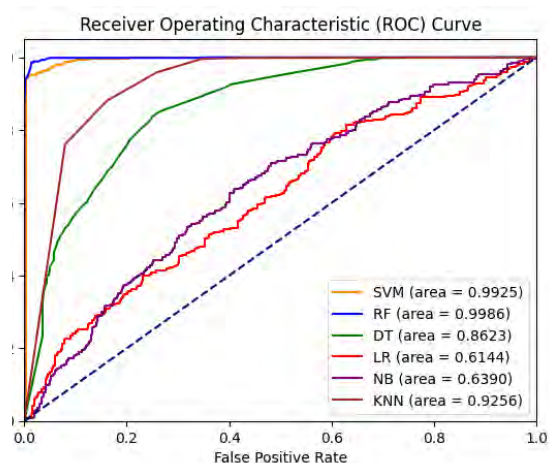
با این رویکرد، مدل می‌تواند هر کلاس را به صورت جداگانه و دقیق‌تر پیش‌بینی کند؛ علاوه بر این، با ترکیب نتایج مدل‌های دوکلاسه می‌توان به سیستم پیش‌بینی جامع‌تر و دقیق‌تری دست یافت. در جدول ۵ معیارهای ارزیابی مدل از جمله دقت، صحت، بازیابی، F1-score و AUC آورده شده است. از مهم‌ترین معیارهای ارزیابی مدل می‌توان به دقت و صحت اشاره کرد. در حالت ۱ الگوریتم RF با مقدار دقت و صحت ۰/۹۶۹۰ و ۰/۹۴۲۶ به عنوان بهترین الگوریتم شناسایی شده است و الگوریتم SVM با اختلاف کمی در مقایسه با RF با مقدار دقت و صحت ۰/۹۵۸۷ و ۰/۹۴۰۳ به عنوان الگوریتم دوم شناسایی شده است. در حالت دوم و سوم و پنجم به ترتیب الگوریتم‌های SVM و RF نتایج قابل قبولی داشته‌اند و در حالت ۴ الگوریتم‌های RF و SVM به ترتیب توانسته‌اند با دقت و صحت بالایی پیش‌بینی را انجام دهند. همان‌طور که مشاهده شد، الگوریتم‌های RF و SVM الگوریتم‌هایی هستند که برای این مورد مطالعه نتایج قابل قبولی دارند؛ بنابراین، می‌توان نتیجه گرفت که می‌توان از این الگوریتم‌ها برای پیش‌بینی احتمال وقوع ریسک عملیاتی در صنعت بانکداری استفاده کرد.

¹ Binary Classification

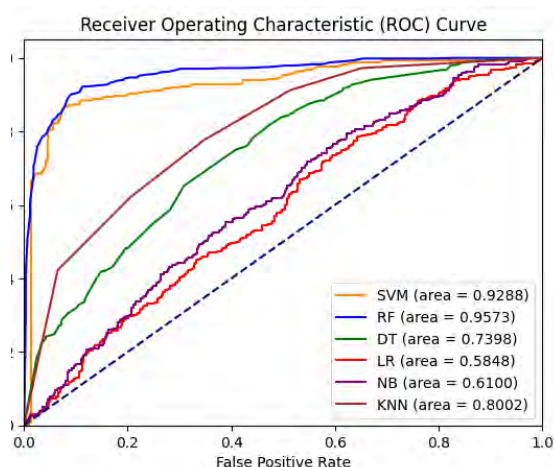
جدول (۵) نتایج معیارهای ارزیابی مدل
Table (5) Results of model evaluation criteria

AUC	F1-score	بازخوانی	صحت	دقت	الگوریتم	
۰/۹۹۲۵	۰/۹۵۹۴	۰/۹۷۹۳	۰/۹۴۰۳	۰/۹۵۸۷	SVM	الف ۱
۰/۹۹۸۶	۰/۹۶۹۸	۰/۹۹۸۶	۰/۹۴۲۶	۰/۹۶۹۰	RF	
۰/۸۷۵۶	۰/۷۹۰۹	۰/۷۸۸۷	۰/۷۹۳۱	۰/۷۹۲۰	DT	
۰/۶۱۴۴	۰/۵۶۶۱	۰/۵۶۴۹	۰/۵۶۷۳	۰/۵۶۸۲	LR	
۰/۶۳۹۰	۰/۵۸۰۱	۰/۵۴۲۸	۰/۶۲۲۸	۰/۶۰۸۱	NB	
۰/۹۲۵۶	۰/۸۶۴۳	۰/۹۵۸۶	۰/۷۸۶۸	۰/۸۴۹۹	KNN	
۰/۹۲۸۸	۰/۸۸۳۳	۰/۸۸۱۰	۰/۸۸۵۶	۰/۸۸۵۶	SVM	الف ۲
۰/۹۵۷۳	۰/۸۸۳۸	۰/۹۳۷۹	۰/۸۳۵۶	۰/۸۷۸۸	RF	
۰/۷۳۹۸	۰/۶۸۶۹	۰/۷۶۰۳	۰/۶۳۰۹	۰/۶۶۳۶	DT	
۰/۵۸۴۸	۰/۵۳۲۲	۰/۵۳۴۵	۰/۵۲۹۹	۰/۵۳۸۱	LR	
۰/۶۱۰۰	۰/۴۸۸۸	۰/۴۱۲۱	۰/۶۰۰۵	۰/۵۷۶۳	NB	
۰/۸۰۰۲	۰/۷۲۷۰	۰/۷۷۵۹	۰/۶۸۳۹	۰/۷۱۳۶	KNN	
۰/۹۲۸۳	۰/۸۷۳۹	۰/۸۵۲۵	۰/۸۹۶۶	۰/۸۸۰۷	SVM	الف ۳
۰/۹۴۴۸	۰/۸۷۵۹	۰/۹۱۲۶	۰/۸۴۲۰	۰/۸۷۴۶	RF	
۰/۷۴۳۱	۰/۷۰۳۷	۰/۷۹۶۰	۰/۶۳۰۶	۰/۶۷۴۹	DT	
۰/۵۶۲۳	۰/۵۶۷۴	۰/۶۲۴۸	۰/۵۱۹۷	۰/۵۳۸۰	LR	
۰/۵۸۷۹	۰/۶۰۹۹	۰/۷۱۰۴	۰/۵۳۴۲	۰/۵۵۹۲	NB	
۰/۷۹۸۶	۰/۷۲۸۲	۰/۷۹۰۵	۰/۶۷۵۰	۰/۷۱۳۸	KNN	
۰/۹۷۶۱	۰/۹۳۵۲	۰/۹۴۹۲	۰/۹۲۱۶	۰/۹۳۶۴	SVM	الف ۴
۰/۹۸۵۳	۰/۹۴۸۵	۰/۹۷۹۰	۰/۹۱۹۹	۰/۹۴۸۷	RF	
۰/۷۹۴۷	۰/۶۶۴۰	۰/۶۱۸۸	۰/۷۱۶۳	۰/۶۹۷۳	DT	
۰/۶۰۲۷	۰/۵۹۹۲	۰/۶۶۳۷	۰/۵۴۶۱	۰/۵۷۰۸	LR	
۰/۶۲۷۰	۰/۶۴۲۴	۰/۸۰۵۷	۰/۵۳۴۲	۰/۵۶۶۵	NB	
۰/۹۰۳۷	۰/۸۳۲۷	۰/۹۳۷۲	۰/۷۴۹۱	۰/۸۱۷۹	KNN	
۰/۹۹۳۴	۰/۹۷۶۴	۰/۹۸۸۷	۰/۹۶۴۴	۰/۹۷۶۲	SVM	الف ۵
۰/۹۹۴۰	۰/۹۵۲۴	۰/۹۹۶۲	۰/۹۱۲۴	۰/۹۵۰۴	RF	
۰/۹۱۱۰	۰/۸۲۶۴	۰/۸۵۰۳	۰/۸۰۳۸	۰/۸۲۱۹	DT	
۰/۶۲۹۳	۰/۶۲۰۷	۰/۶۸۴۳	۰/۵۶۷۸	۰/۵۸۳۱	LR	
۰/۶۰۶۱	۰/۵۹۶۹	۰/۶۹۵۶	۰/۵۲۲۷	۰/۵۳۱۷	NB	
۰/۹۶۰۹	۰/۹۴۰۶	۰/۹۹۶۲	۰/۸۹۰۸	۰/۹۳۷۳	KNN	

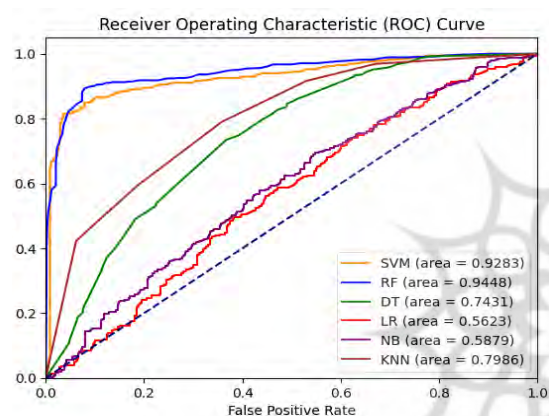
برای ارزیابی دقت پیش‌بینی الگوریتم‌های استفاده‌شده از ناحیه زیرمنحنی ROC(AUC) استفاده شده است. منحنی ROC به صورت بصری عملکرد مدل طبقه‌بندی را درآستانه‌های مختلف نشان می‌دهد که تقابل بین نرخ مثبت واقعی و نرخ مثبت کاذب را نشان می‌دهد و توانایی یک طبقه‌بندی‌کننده را برای تمایز بین کلاس‌ها نشان می‌دهد. مقدار AUC بالاتر عملکرد خوب مدل را نشان می‌دهد که می‌تواند کلاس را به طور مؤثر از هم جدا کند. شکل ۳ منحنی‌های ROC را برای الگوریتم‌های استفاده‌شده در ۵ حالت بیان شده رسم شده است. همان‌طور که مشاهده می‌شود، الگوریتم‌های RF و SVM نتایج قابل قبولی دارند.



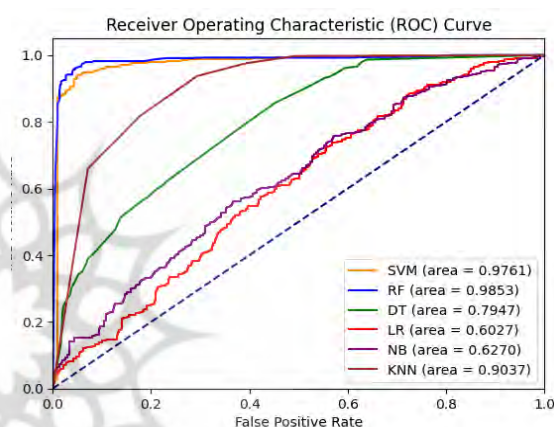
ROC of State 1



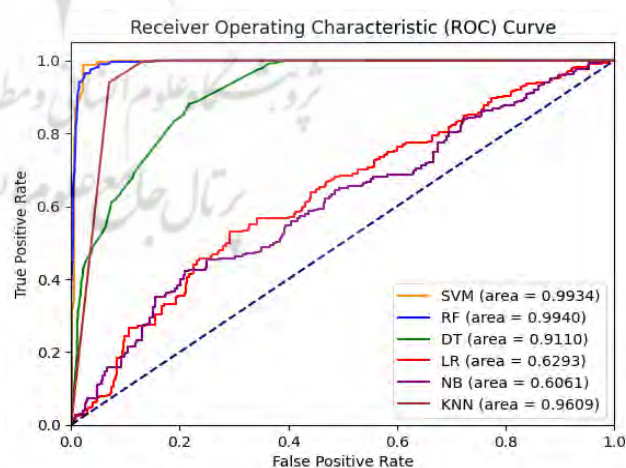
ROC of State 2



ROC of State 3



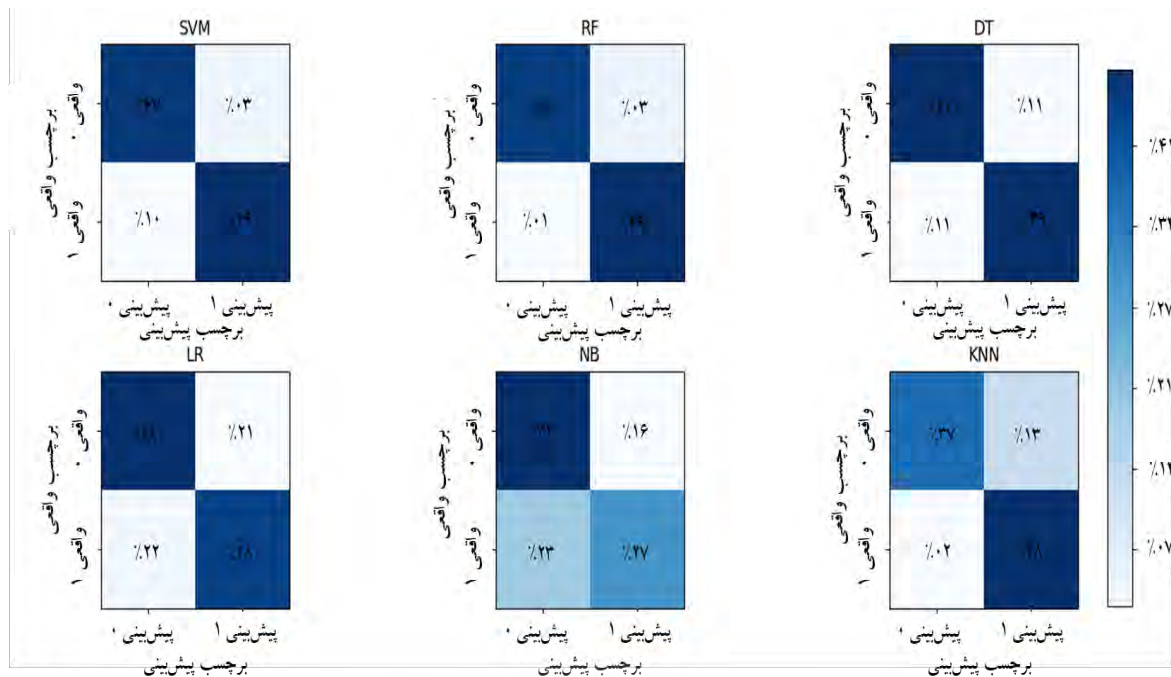
ROC of State 4



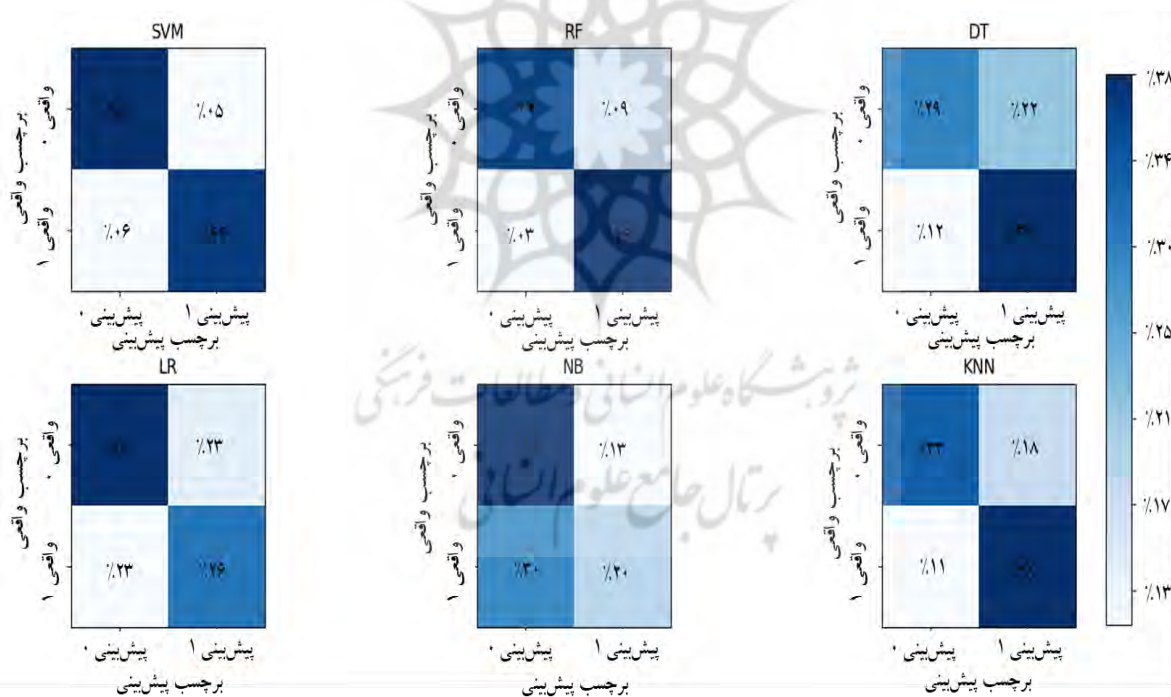
ROC of State 5

شکل (۳) نمودار ROC برای ۵ حالت
Figure (3) ROC diagram for 5th state

ماتریس درهم‌ریختگی برای مدل مد نظر در ۵ حالت بیان‌شده، رسم شد. عناصر پررنگ، کلاس‌هایی هستند که به‌درستی پیش‌بینی شده‌اند. شکل‌های ۴ تا ۸ ماتریس درهم‌ریختگی برای ۵ حالت بیان‌شده را نشان می‌دهند.

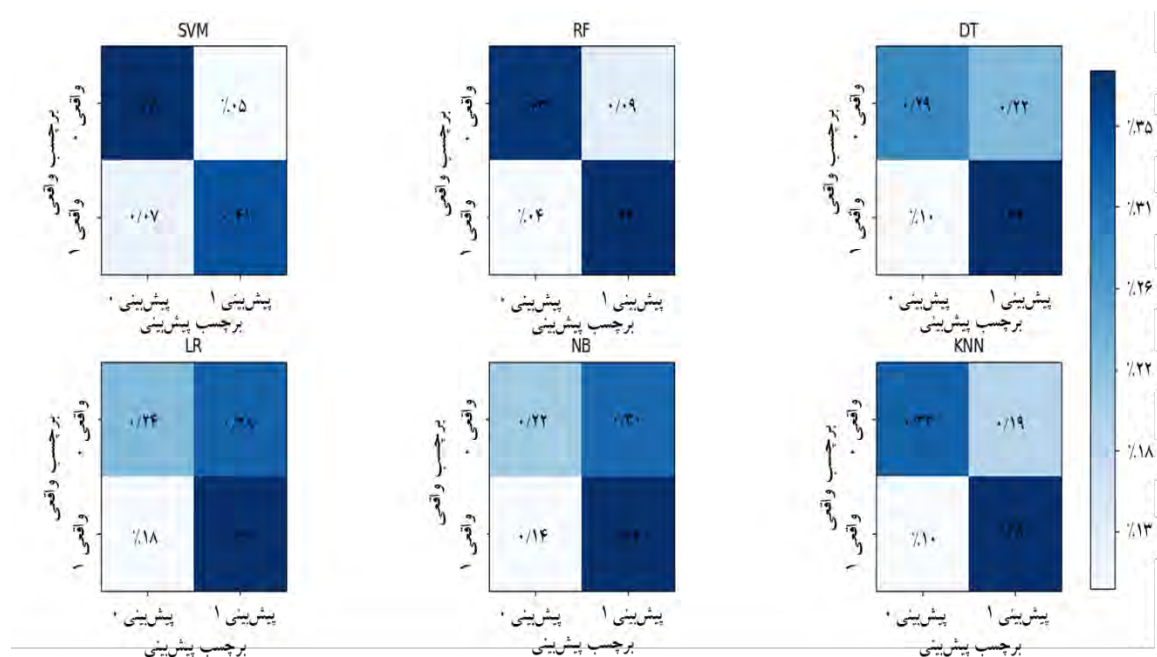


شکل (۴) ماتریس درهم‌ریختگی برای حالت ۱

Figure (4) Confusion matrix for 1st state

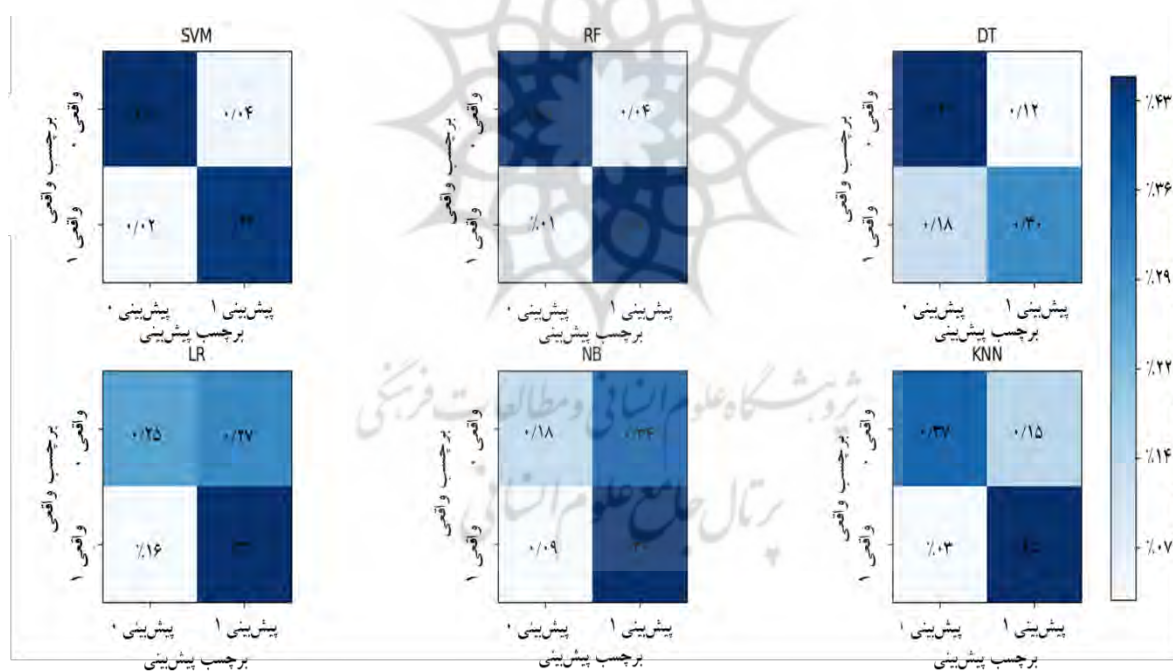
شکل (۵) ماتریس درهم‌ریختگی برای حالت ۲

Figure (5) Confusion matrix for 2nd state



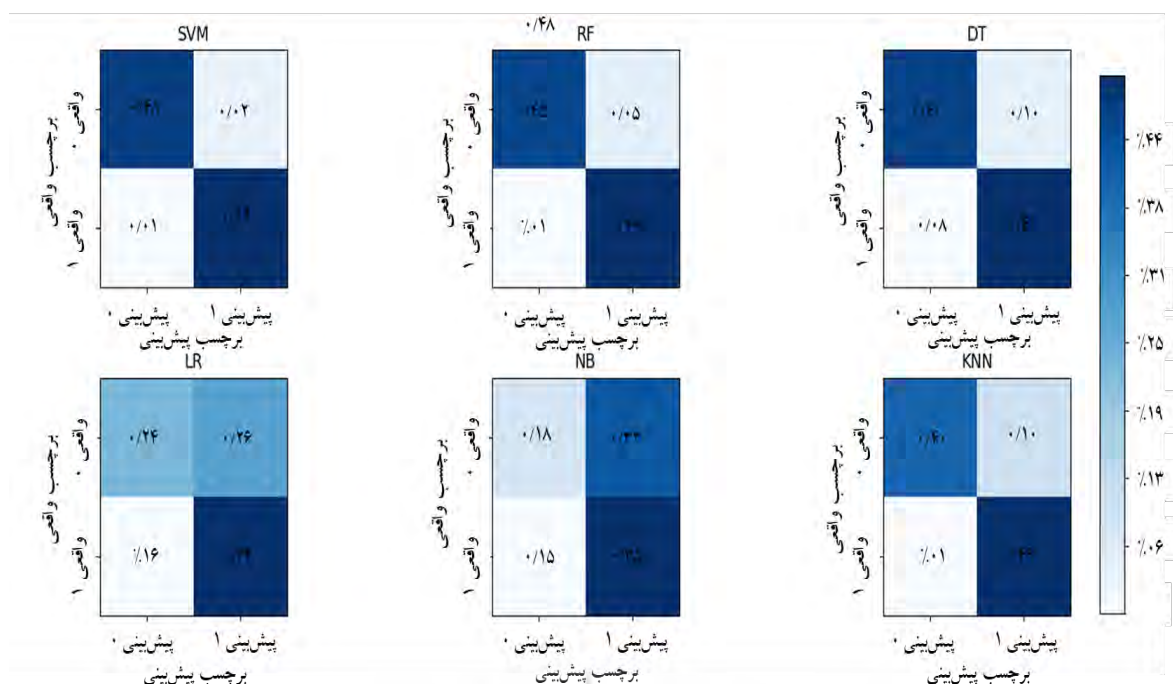
شکل (۶): ماتریس درهم‌ریختگی برای حالت ۳

Figure (6) Confusion matrix for 3rd state



شکل (۷): ماتریس درهم‌ریختگی برای حالت ۴

Figure (7) Confusion matrix for 4th state



شکل (۸) ماتریس درهم‌ریختگی برای حالت ۵

Figure (8) Confusion matrix for 5th state

نتایج و پیشنهادها

در این پژوهش تلاش شد تا با استفاده از الگوریتم‌های یادگیری ماشین، وقوع ریسک‌های عملیاتی در صنعت بانکداری بررسی و پیش‌بینی شود. نتایج نشان می‌دهند که الگوریتم‌های مختلف، به‌ویژه جنگل تصادفی و ماشین بردار پشتیبان عملکرد قابل‌قبولی در پیش‌بینی وقوع ریسک‌های عملیاتی دارند. این نتایج می‌تواند تحولی در شیوه مدیریت ریسک‌های عملیاتی در بانک‌ها ایجاد کند و باعث کاهش چشمگیری در هزینه‌ها و خسارات ناشی از این ریسک‌ها شود. یکی از مهم‌ترین یافته‌ها این است که استفاده از داده‌های بزرگ و متنوع می‌تواند به‌طور چشمگیری دقت پیش‌بینی‌ها را افزایش دهد. مدل‌های یادگیری ماشین قادرند با تحلیل داده‌های حجیم و پیچیده، الگوهای پنهان را کشف کنند و به‌این‌ترتیب به شناسایی ریسک‌های بالقوه کمک کنند. این امر می‌تواند به بانک‌ها این امکان را بدهد که پیش از وقوع ریسک‌ها اقدامات پیشگیرانه‌ای را اتخاذ کنند و بدین ترتیب از وقوع خسارات مالی و اعتباری جلوگیری کنند. استفاده از مدل‌های یادگیری ماشین در مدیریت ریسک عملیاتی، می‌تواند به بهبود فرایند تصمیم‌گیری‌های مدیریتی کمک کند. این مدل‌ها با ارائه تحلیل‌های دقیق و پیش‌بینی‌های مبتنی بر داده، به مدیران بانک‌ها این امکان را می‌دهند که تصمیم‌های بهتری بگیرند و راهکارهای مؤثرتری برای مدیریت ریسک‌ها اجرا کنند؛ به علاوه، این روش‌ها می‌توانند زمان و هزینه‌های مرتبط با مدیریت ریسک را به‌طور چشمگیری کاهش دهند و به بانک‌ها این امکان را بدهند که منابع خود را به شکلی بهینه‌تر مدیریت کنند. برای بهبود دقت پیش‌بینی‌ها می‌توان داده‌های بیشتری را جمع‌آوری و متغیرهای جدیدی را به مدل‌ها اضافه کرد که شامل داده‌های تاریخی، داده‌های اقتصادی و اطلاعات مرتبط با فرایندهای داخلی بانک‌ها می‌شود. با پیشرفت تکنولوژی و ظهور الگوریتم‌های جدید می‌توان از تکنیک‌های پیشرفته‌تری مانند یادگیری تقویتی مانند الگوریتم‌های DQN، Q-Learning، DDPG و Meta-Learning استفاده کرد. این تکنیک‌ها می‌توانند دقت و کارایی مدل‌ها را بهبود بخشند.

منابع

- استادی، بختیار، خزایی، سجاد، و حسین‌زاده کاشان، علی (۱۳۹۷). ارزیابی ریسک عملیاتی با استفاده از روش استنتاج بیزی و با در نظر گرفتن ترکیب منابع داده‌ای و فرض وابستگی بین نظرات کارشناسان و داده‌های زیان داخلی. *راهبرد مدیریت مالی*، ۶(۱)، ۵۳-۷۲. <https://doi.org/10.22051/jfm.2018.17227.1486>
- خوش‌سیما، رضا، و شهیکی تاش، محمدنبی (۱۳۹۱). تأثیر ریسک‌های اعتباری، عملیاتی و نقدینگی بر کارایی نظام بانکی ایران. *فصلنامه پژوهشنامه اقتصاد و برنامه ریزی*، ۱۷(۴)، ۶۹-۹۵.
- طالبی، محمد، کاوندف، مجتبی، و حسین‌پور، محمد (۱۳۹۰). تحلیل و رتبه‌بندی ریسک‌های عملیاتی در بانکداری اسلامی؛ مطالعه موردی: بانکداری بدون ربا در ایران. *اقتصاد اسلامی*، ۱۱(۴۴)، ۱۵۷-۱۸۴.
- مصطفائی دولت‌آباد، خدیجه، آذر، عادل، و مقبل باعرض، عباس (۱۳۹۷). شناسایی و تحلیل ریسک‌های عملیاتی با استفاده از نگاشت شناختی فازی. *مدیریت دارایی و تأمین مالی*، ۶(۴)، ۱-۱۸.
- <https://doi.org/10.22108/amf.2018.103404.1087>
- نادری، حامد، و رستگار، محمدعلی (۱۴۰۱). به‌کارگیری روش فراترکیب در روش‌شناسی مدیریت ریسک عملیاتی بانکی. *مدیریت دارایی و تأمین مالی*، ۱۰(۴)، ۱۱۵-۱۳۲. <https://doi.org/10.22108/amf.2023.135765.1767>
- نصرتی، هاشم، و پاکیزه، کامران (۱۳۹۳). تخمین ذخیره سرمایه ریسک عملیاتی در صنعت بانکداری با استفاده از رویکرد توزیع زیان (LDA). *مهندسی مالی و مدیریت اوراق بهادار (مدیریت پرتفوی)*، ۵(۲۰)، ۱-۲۶.

References

- Afonso, G., Curti, F., & Mihov, A. (2019). *Coming to terms with operational risk* (No. 20190107). Federal Reserve Bank of New York.
- Abdymomunov, A., Curti, F., & Mihov, A. (2020). US banking sector operational losses and the macroeconomic environment. *Journal of Money, Credit and Banking*, 52(1), 115-144. <https://doi.org/10.1111/jmcb.12661>.
- Afroz, S., Brennan, M., & Greenstadt, R. (2012). Detecting hoaxes, frauds, and deception in writing style online. In *2012 IEEE symposium on security and privacy*. San Francisco, CA, USA. <https://doi.org/10.1109/SP.2012.34>
- Akbari, M., & Yazdani, A. (2023). Machine learning in estimating operational risk coverage capital of banks with a loss distribution Approach. *Financial Management Perspective*, 13(42), 9-34. <https://doi.org/10.48308/jfmp.2023.103948>
- Bank for International Settlements. (2004). *Basel II: International Convergence of Capital Measurement and Capital Standards: A Revised Framework*. Basel Committee on Banking Supervision, CH-4002 Basel, Switzerland.
- Bank for International Settlements. (2016). *Standardised Measurement Approach for Operational Risk*. Consultative Document.
- Barakat, A., & Hussainey, K. (2013). Bank governance, regulation, supervision, and risk reporting: Evidence from operational risk disclosures in European banks. *International Review of Financial Analysis*, 30, 254-273. <https://doi.org/10.1016/j.irfa.2013.07.002>
- Basel Committee on Banking Supervision. (2006). *International convergence of capital measurement and capital standards: A Revised Framework - Comprehensive Version*. Bank of International Sett.
- Bouveret, A. (2018). *Cyber Risk for the Financial Sector: A Framework for Quantitative Assessment*. International Monetary Fund.
- Committee of Sponsoring Organizations of the Treadway Commission (COSO). (2017). *Enterprise Risk Management: Integrating with Strategy and Performance*. COSO.
- Crisanto, J. C., & Prenio, J. (2017). *Regulatory Approaches to Enhance Banks' Cyber-Security Frameworks*. Bank for International Settlements. Financial Stability Institute. Retrieved from.
- Crouchy, M., Galai, D., & Mark, R. (1998). Key steps in building consistent operational risk measurement and management. *Operational Risk and Financial Institutions*, 17(3), 45-62.
- Cantarella, M., Fraccaroli, N., & Volpe, R. (2023). Does fake news affect voting behaviour? *Research Policy*, 52(1), 104628. <https://doi.org/10.1016/j.respol.2022.104628>
- Drew, J. M., & Farrell, L. (2018). Online victimization risk and self-protective strategies: Developing police-led

- cyber fraud prevention programs. *Police Practice and Research*, 19(6), 537-549. <https://doi.org/10.1080/15614263.2018.1507890>
- Elarif, F. Z., & Hinti, S. (2014). Methods of quantifying operational risk in Banks: Theoretical approaches. *American Journal of Engineering Research*, 3(03), 283-244.
- Garg, A., Lilhore, U. K., Ghosh, P., Prasad, D., & Simaiya, S. (2021, August). Machine learning-based model for prediction of student's performance in higher education. In *8th International Conference on Signal Processing and Integrated Networks (SPIN)* (pp. 162-168). Noida, India. <https://doi.org/10.1109/SPIN52536.2021.9565999>
- Guo, B., Ding, Y., Yao, L., Liang, Y., & Yu, Z. (2020). The future of false information detection on social media: New perspectives and trends. *ACM Computing Surveys (CSUR)*, 53(4), 1-36. <https://doi.org/10.1145/3393880>
- González-Carrasco, I., Jiménez-Márquez, J. L., López-Cuadrado, J. L., & Ruiz-Mezcua, B. (2019). Automatic detection of relationships between banking operations using machine learning. *Information Sciences*, 485, 319-346. <https://doi.org/10.1016/j.ins.2019.02.030>
- Goutte, C., & Gaussier, E. (2005, March). A probabilistic interpretation of precision, recall and F-score, with implication for evaluation. In D.E. Losada, & J.M. Fernández-Luna (Eds.), *Advances in Information Retrieval. ECIR 2005* (pp. 345-359). Lecture Notes in Computer Science, vol 3408. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-31865-1_25
- Hoffman, D. G. (2002). *Managing Operational Risk: 20 Firmwide Best Practice Strategies*. John Wiley & Sons. Inc.
- Ivanov, D., Dolgui, A., & Sokolov, B. (2019). The impact of digital technology and Industry 4.0 on the ripple effect and supply chain risk analytics. *International Journal of Production Research*, 57(3), 829-846. <https://doi.org/10.1080/00207543.2018.1488086>
- Jabbar, M. A., Deekshatulu, B. L., & Chandra, P. (2016). Intelligent heart disease prediction system using random forest and evolutionary approach. *Journal of Network and Innovative Computing*, 4, 10.
- Khoshsima, R., & Shahikitash, M. N. (2013). The impact of credit, operational and liquidity risks on the efficiency of Iran's banking system. *Planning and Budgeting*, 17(4), 69-95. [In Persian].
- Liu, X., Wang, X., Su, Q., Zhang, M., Zhu, Y., Wang, Q., & Wang, Q. (2017). A hybrid classification system for heart disease diagnosis based on the RFRS method. *Computational and Mathematical Methods in Medicine*, (1), 8272091. <https://doi.org/10.1155/2017/8272091>
- Mor -Valencia, A., & Zapata-Jaramillo, W. (2017). Quantifying operational risk using the loss distribution approach (lda) model. In *Proceedings of the Seventh European Academic Research Conference on Global Business, Economics, Finance and Banking (EAR17Swiss Conference)* (pp. 0-10).
- Mora-Valencia, A. (2010). Cuantificación del riesgo operativo en entidades financieras en Colombia. *Cuadernos de Administración*, 23(41), 185-211.
- Mostafaei, K., Azar, A., & Moghbel, A. (2019). Identification and analysis of operational risks: A fuzzy cognitive map approach. *Journal of Asset Management and Financing*, 6(4), 1-18. <https://doi.org/10.22108/amf.2018.103404.1087> [In Persian].
- Naderi, H., & Rastegar, M. A. (2022). Applying the meta-synthesis method in banking operational risk management methodology. *Journal of Asset Management and Financing*, 10(4), 115-132. <https://doi.org/10.22108/amf.2023.135765.1767> [In Persian].
- Nosrati, H., & Pakizeh, K. (2014). Estimation of operating capital reserves in the banking industry. *Financial Engineering and Portfolio Management*, 5(20), 1-26. [In Persian].
- Ostadi, B., Khazayi, S., & Husseinzadeh, A. K. (2018). Operational risk Assessment using Bayesian inference with regard to the composition of data sources and the assumption of dependence between experts and internal loss data. *Financial Management Strategy*, 6(1), 53-72. <https://doi.org/10.22051/jfm.2018.17227.1486> [In Persian].
- Poongodi, M., Nguyen, T. N., Hamdi, M., & Cengiz, K. (2023). A measurement approach using smart-IoT based architecture for detecting the COVID-19. *Neural Processing Letters*, 55, 877. <https://doi.org/10.1007/s11063-021-10602-x>
- Pena, A., Patino, A., Chiclana, F., Caraffini, F., Gongora, M., Gonzalez-Ruiz, J. D., & Duque-Grisales, E. (2021). Fuzzy convolutional deep-learning model to estimate the operational risk capital using multi-source risk events. *Applied Soft Computing*, 107, 107381. <https://doi.org/10.1016/j.asoc.2021.107381>
- Pereira, P., Silva, E. S., & Pereira, A. (2018). Operational risk management: The basel II. *The International Journal of Business Management and Technology*, 2.
- Peña, A., Bonet, I., Lochmuller, C., Chiclana, F., & Góngora, M. (2018). An integrated inverse adaptive neural fuzzy system with Monte-Carlo sampling method for operational risk management. *Expert Systems with Applications*, 98, 11-26. <https://doi.org/10.1016/j.eswa.2018.01.001>
- Posavljak, M. (2024). *Development of Tools for Infrastructure Asset Management Cross-Asset Trade-off Analysis and Universal Performance Measure for Public Agencies* [Doctoral dissertation, University of Waterloo].

- Rahman, M. M., Akhi, S. S., Hossain, S., Ayub, M. I., Siddique, M. T., Nath, A., Nath, P. C., & Hassan, M. M. (2024). Evaluating machine learning models for optimal customer segmentation in banking: A comparative study. *The American Journal of Engineering and Technology*, 6(12), 68-83. <https://doi.org/10.37547/tajet/Volume06Issue12-08>
- Simaiya, S., Lilhore, U. K., Prasad, D., & Verma, D. K. (2021a). MRI brain tumour detection & image segmentation by hybrid hierarchical K-means clustering with FCM based machine learning model. *Annals of the Romanian Society for Cell Biology*, 25(1), 88-94.
- Simaiya, S., Gautam, V., Lilhore, U. K., Garg, A., Ghosh, P., Trivedi, N. K., & Anand, A. (2021b, October). EEPsA: Energy efficiency priority scheduling algorithm for cloud computing. In *2nd International Conference on Smart Electronics and Communication (ICOSEC)* (pp. 1064-1069). IEEE. <https://doi.org/10.1109/ICOSEC51865.2021.9591967>
- Tsai, C. F., & Wu, J. W. (2008). Using neural network ensembles for bankruptcy prediction and credit scoring. *Expert Systems with Applications*, 34(4), 2639-2649. <https://doi.org/10.1016/j.eswa.2007.05.019>
- Talebi, M., Kavand, M. & Hosseinpour, M. (2011). Analysis and ranking of operational risks in Islamic banking; Case study: Interest-free banking in Iran. *Islamic Economics*, 11(44), 157-184. [In Persian].
- Wang, T., & Hsu, C. (2013). Board composition and operational risk events of financial institutions. *Journal of Banking & Finance*, 37(6), 2042-2051. <https://doi.org/10.1016/j.jbankfin.2013.01.027>
- Zhou, F., Qi, X., Xiao, C., & Wang, J. (2021). MetaRisk: Semi-supervised few-shot operational risk classification in banking industry. *Information Sciences*, 552, 1-16. <https://doi.org/10.1016/j.ins.2020.11.027>

