Publisher: Shahid Beheshti University
Journal of Financial Management Perspective
2023, Vol. 13, No.42
PP: 9-34
DOI: 10.48308/jfmp.2023.103948
P-ISSN: 2645-4637
E-ISSN: 2645-4645
Copyright: © 2023 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by /4.0/)

Machine learning in estimating operational risk coverage capital of banks with a loss distribution Approach

Mahdi Akbari [*], **Ahmadreza Yazdanian**[**]

Research Paper

Abstract

Operational risk is one of the most important risks of financial institutions. After the approval of the Basel committee, it has started to be noticed all over the world. The increasing increase in operating losses in various business lines has caused the attention of financial institution managers to be directed to the area of operational risk. In this research, a method for estimating the appropriate threshold for operational loss data and also a method for classifying operational loss data is presented, and the capital required to cover operational risk by combining the intensity distribution function and frequency of operating loss data and Monte Carlo simulation are obtained. Also, the dependence between matrix cells of business lines and loss making events has also been investigated. For this research, operational loss data related to a banking group including several Asian, European and American banks have been used. The research results indicate that the loss distribution approach with the combination of critical value theory and machine learning algorithms (clustering), as well as the loss distribution approach with the combination of machine learning algorithms (classification), is more efficient than other methods.

Keywords: Operational Risk; Loss Distribution Approach; Critical Value Theory; Classification Algorithms; Clustering Algorithms.

Received: 2022. November. 27, Accepted: 2023. April. 03.
* M.A. Student in Financial Mathematics, University of Kharazmi, Tehran, Iran.
E-Mail: mehdi.akbari.financial.sciences@gmail.com
** Assistant Prof., Department of Financial Mathematics, Kharazmi University, Tehran, Iran. (Corresponding Author).
E-Mail: yazdanian@khu.ac.ir

ناشر: دانشگاه شهیدبهشتی
نشریه چشم‌انداز مدیریت مالی
۱۴۰۲، دوره ۱۳، شماره ۴۲
ص‌ص. ۹- ۳۴

شاپای چاپی: ۴۶۳۷- ۲۶۴۵
شاپای الکترونیکی: ۴۶۴۵- ۲۶۴۵

Copyright: © 2023 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/)

یادگیری ماشین در تخمین سرمایه پوششی ریسک‌عملیاتی بانک‌ها با رویکرد توزیع زیان

مهدی اکبری *، احمدرضا یزدانیان **

مقاله پژوهشی

چکیده

ریسک عملیاتی یکی از مهم‌ترین ریسک‌های مؤسسات مالی است. توجه به آن پس از مصوبات کمیته بال، در سراسر دنیا آغاز شده‌است. افزایش روزافزون زیان‌های عملیاتی در خطوط کسب‌وکار مختلف سبب شده‌است تا توجّه مدیران مؤسسات مالی معطوف به حوزه ریسک عملیاتی شود. در این پژوهش، روشی جهت تخمین آستانه مناسب برای داده‌های شدت زیان عملیاتی و همچنین روشی جهت طبقه‌بندی داده‌های شدت زیان عملیاتی ارائه شده‌است و سرمایه مورد نیاز برای پوشش ریسک عملیاتی با تجمیع تابع توزیع شدت و فرکانس داده‌های زیان عملیاتی و شبیه‌سازی مونتکارلو به‌دست آمده‌است. همچنین وابستگی بین سلول‌های ماتریس خطوط کسب‌وکار و حوادث ضررساز نیز مورد بررسی قرار گرفته‌است. برای این پژوهش داده‌های زیان عملیاتی مربوط به‌یک مجموعه بانکداری شامل چند بانک آسیایی، اروپایی و آمریکایی به‌کارگرفته شده‌است. نتایج پژوهش حاکی از آن است که رویکرد توزیع زیان با ترکیب تئوری مقدار بحرانی و الگوریتم‌های‌یادگیری ماشین (خوشه‌بندی)، همچنین رویکرد توزیع زیان با ترکیب الگوریتم‌های‌یادگیری ماشین (طبقه‌بندی)، نسبت به سایر روش‌ها کارآمدتر است.

کلیدواژه‌ها: ریسک عملیاتی؛ رویکرد توزیع زیان؛ نظریه مقدار بحرانی؛ الگوریتم‌های طبقه‌بندی؛ الگوریتم‌های خوشه‌بندی.

تاریخ دریافت مقاله: ۱۴۰۱/۰۹/۰۶، تاریخ پذیرش مقاله: ۱۴۰۲/۰۱/۱۴.

* کارشناسی ارشد ریاضیات مالی، دانشگاه خوارزمی، تهران، ایران.
E-Mail: mehdi.akbari.financial.sciences@gmail.com
** استادیار، گروه ریاضیات مالی، دانشگاه خوارزمی، تهران، ایران (نویسنده مسئول).
E-Mail: yazdanian@khu.ac.ir

۱. مقدمه

زمانی که ناظران بانکی آلمان در۲۶ ژوئن۱۹۷۴ بانک هرستات[1] را منحل کردند و مدتی پس از از آن کمیته نظارت بر امور بانکی بازل[2] را تشکیل دادند، مدیریت ریسک عملیاتی جایگاه خاصی یافت [۱۰]. براساس تعریف کمیته بازل ریسک عملیاتی عبارت است از زیان ناشی از فرآیندهای داخلی، افراد و سیستم‌های معیوب یا ریسک زیان ناشی از حوادث خارجی. با توجه به تعریف ریسک عملیاتی این شاخه از ریسک به حدی می‌تواند گسترده باشد که چه‌بسا به ورشکستگی مؤسسه مالی بیانجامد [۹]. برای داشتن مقایسه‌ای بین ریسک‌های عملیاتی، بازار و اعتباری می توان گفت ریسک عملیاتی با عناصری مانند افراد، فرآیندها، سیستم‌ها، ریسک بازار با عناصری مانند اوراق‌بهادار و ریسک اعتباری با عناصری مانند وام‌ها سازوکار دارند. اگر عناصر هر دسته از ریسک‌ها را به عنوان یک مجموعه در نظر بگیریم، حداکثر زیان ناشی از دسته ریسک‌های بازار و اعتباری و نقدینگی محدود و مشخص است. امّا با توجه به افزایش گروه‌های حوادث در ریسک عملیاتی حداکثر زیان ناشی از این دسته ریسک نامشخص و نامحدود است [۹].

شکل۱. ریسک عملیاتی در بازل ۲

از شکل (۱) مشاهده می‌گردد که طبق بیانیه‌های کمیته‌بال، بانک‌ها می‌توانند الزامات کفایت سرمایه خود را با توجّه به ریسک‌های عملیاتی، بازار، نقدینگی و اعتباری تعیین کنند. این کمیته برای محاسبه ریسک عملیاتی روش‌های شاخص پایه[3]، استاندارد[4]و پیشرفته[1] را پیشنهاد می‌دهد

[1] Herstatt
[2] Basel
[3] Basic Indicator Approach
[4] Standardized Approach

[۹]. رویکرد توزیع زیان در سال ۲۰۰۱ توسط کمیته بازل ارائه شد که در آن با استفاده از ماتریس ریسک عملیاتی[۲]، توزیع‌های آماری فراوانی و شدت زیان سالانه به‌طور جداگانه تخمین‌زده می‌شود، سپس با ترکیب این دو تابع توزیع، تابع توزیع زیان تجمیعی به‌دست می‌آید. با کم کردن ارزش در معرض خطر ۹۹/۹ درصد از میانگین تابع توزیع زیان تجمیعی، ذخیره سرمایه برای هر سلول به‌دست می‌آید، سپس با درنظر گرفتن همبستگی بین سلول‌های ماتریس می‌توان ذخیره سرمایه کل بانک را به‌دست‌آورد [۹]. در این تحقیق دو رویکرد براساس الگوریتم‌های یادگیری ماشین با محوریت رویکرد توزیع زیان ارائه شده‌است که ضمن در نظر گرفتن همبستگی بین سلول‌های ماتریس ریسک عملیاتی، کاهش پیچیدگی محاسباتی و زمان محاسبه رویکرد توزیع زیان را به دنبال دارد.

نوآوری‌های این پژوهش در مدل‌سازی داده‌های شدت زیان عملیاتی است. در مدل‌سازی داده‌های شدت زیان عملیاتی، یک راه استفاده از نظریه مقدار بحرانی[۳] است. برای استفاده از این روش ملزم به انتخاب یک آستانه مناسب هستیم ولی روش یکتا و منحصر به فردی برای انتخاب این آستانه وجود ندارد. در واقع در بحث تئوری مقدار بحرانی داده‌های شدت، روشی برای تخمین آستانه مناسب با استفاده از الگوریتم‌های خوشه‌بندی ارائه می‌دهیم. درمدل‌سازی داده‌های شدت زیان عملیاتی، یک راه استفاده از برازش پارامتری است. در برازش پارامتری داده‌های شدت، روشی مناسب برای طبقه‌بندی داده‌های شدت با استفاده از الگوریتم‌های طبقه بندی ارائه می‌دهیم.

در نظام بانکداری، حرکت به سوی فعالیت‌های الکترونیکی سبب افزایش تعداد مشتریان، افزایش و پیچیده‌تر شدن خدمات قابل‌ارائه به مشتریان می‌شود. تمامی این امور سبب به وجود آمدن ریسک عملیاتی است بنابراین باید با حرکت به‌سوی نظام بانکداری الکترونیک تدابیری برای مقابله با زیان‌های بالقوه آن اندیشید. داشتن برنامه مناسب برای مدیریت ریسک عملیاتی سبب می‌شود تا مدیران مؤسسات مالی سرمایه لازم را برای مقابله با ریسک عملیاتی در نظر بگیرند. ادغام و قبضه مالکیت سبب اختلاط صنایع گوناگون با یکدیگر شده و منجر به افزایش ریسک عملیاتی می‌شود. بنابراین به طورخلاصه می‌توان ضرورت و اهمیت بررسی ریسک عملیاتی را «جهت‌گیری فعالیت‌های بانکی به‌سمت فعالیت‌های بانکی الکترونیک»، «دستورات مدیریت ریسک» و «ادغام و قبضه مالکیت» دانست [۸].

---

[۱] Advanced Measurement Approach

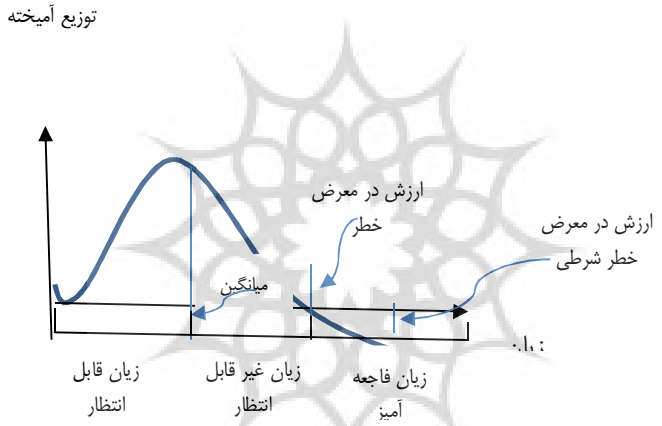
[۲] BLETM

[۳] Critical value theory

دو روش شاخص پایه و استاندارد نسبت به رویکرد توزیع زیان ساده ترند ولی دقت کافی را ندارند و سبب می‌شوند مقدار سرمایه مورد نیاز به‌دقت تخمین زده نشود. اکثر تحقیقات قبلی از روش‌های شاخص پایه و استاندارد برای اندازه‌گیری ریسک عملیاتی استفاده کرده‌اند که درنهایت ذخیره سرمایه را به‌طور دقیق تخمین نزده‌اند. در این پژوهش درصدد پیاده‌سازی رویکرد توزیع زیان برمبنای الگوریتم‌های یادگیری‌ماشین هستیم.

۲. مبانی نظری و پیشینه پژوهش

اندازه‌گیری ریسک عملیاتی براساس معیار ارزش در معرض خطر انجام می‌شود. ارزش در معرض خطر همان چندک یا کوانتایل ۹/۹۹ درصد تابع توزیع زیان تجمیعی در یک سطح اطمینان مشخص است. مطابق استانداردهای بال این دوره زمانی‌یک سال و سطح اطمینان برابر ۹۹/۹ درصد تعریف شده‌است. مطابق شکل (۲) زیان قابل انتظار همان میانگین توزیع و زیان غیرقابل انتظار تفاضل ارزش در معرض خطر از میانگین توزیع است. زیان فاجعه‌آمیز، عبارت است از هر زیانی بیشتر از زیان غیرقابل انتظار [۲،۳،۴،۵،۲۹].



شکل ۲ . ارزش در معرض خطر و ارزش در معرض خطر شرطی

شناسایی مصادیق ریسک عملیاتی کار آسانی نیست و به فرآیندها داخلی سازمان بستگی دارد. جدول (۱) گروه حوادث زیانباری که ممکن است در یک مؤسسه مالی رخ دهد را تشریح می‌کند. طبق نظر کمیته بال طبقه‌بندی زیر به شناسایی مصادیق ریسک عملیاتی کمک می‌کند [۸].

جدول ۱. طبقه‌بندی حوادث زیان‌بار عملیاتی

گروه حوادث	تعریف
سرقت داخلی	زیان ناشی از اعمال متقلبانه، اختلاص دارایی‌ها یا دور زدن مقررات، اصول و سیاست‌های سازمان که حداقل به یک‌طرف داخل سازمان مربوط می‌شود.
سرقت خارجی	زیان ناشی از اعمال متقلبانه، اختلاص دارایی‌ها یا دور زدن مقررات توسط یک نفر خارج از سازمان.
شیوه استخدام و ایمنی محیط کار	زیان‌های ناشی از اعمال خلاف قوانین یا مقررات استخدام، سلامت و ایمنی کارکنان، یا زیان ناشی از پرداخت بابت دعوای خسارت شخصی یا زیان ناشی از مسائل تبعیض‌آمیز میان کارکنان.
مشتریان، محصولات و شیوه کسب‌وکار	زیان‌های ناشی از خطاهای سهوی یا بی‌دقتی در جهت برآوردن تعهدات نسبت به مشتریان خاص یا زیان ناشی از ساختار یک محصول.
اختلال در کسب‌وکار و خرابی سیستم‌ها	زیان‌های ناشی از اختلال در کسب‌وکار یا از کار افتادگی سیستم‌ها.
خسارت‌های وارده بر دارایی‌های فیزیکی	زیان‌های مربوطه به دارایی‌های فیزیکی که از بلاهای طبیعی یا دیگر حوادث ناشی می‌شوند.
اجرا، تحویل و مدیریت فرایندها	زیان‌های ناشی از کاستی در مدیریت فرآیندها و یا زیان ناشی از روابط میان رقبا و فروشندگان.

پس از جمع‌آوری داده‌های زیان عملیاتی، برای هر سلول ماتریس ریسک عملیاتی، باید یک توزیع پیوسته برای داده‌های شدت زیان ویک توزیع گسسته برای داده‌های فرکانس زیان مدل‌سازی کرد. برای مدل‌سازی داده‌های شدت معمولاً دو راه پیشنهادی توسط کمیته‌بال وجود دارد [۹]:

۱. برازش تابع توزیع‌های پیوسته به‌کل داده‌های شدت زیان عملیاتی.
۲. استفاده از نظریه مقدار بحرانی[1].

برای تخمین پارامترهای توزیع نیز روش‌های زیر وجود دارد:

۱. روش گشتاورها[2].
۲. روش حداکثر درست نمایی[3].

در انتخاب توزیع‌های پارامتری باید مناسب‌ترین گزینه را انتخاب کنیم. معمولاً دو راه برای انتخاب توزیع‌های پارامتری مناسب وجود دارد:

۱. استفاده از آزمون‌های فرض نظیر نکویی برازش.
۲. استفاده از نمودار چارک-چارک (Q-Q)[4].

[1] Critical value theory
[2] Moments method
[3] Maximum likelihood method
[4] Quartile-Quartile

گاهی‌اوقات مدیران ریسک برای مدل‌سازی داده‌های شدت از نظریه مقدار بحرانی استفاده می‌کنند. نظریه مقدار بحرانی، همانطور که از نام آن پیدا است، شاخه‌ای از علم آمار می‌باشد که به بررسی و مطالعه پدیده‌های بحرانی می‌پردازد. در مبحث ریسک عملیاتی، فرض این نظریه این است که می‌توان توزیع مربوط به داده‌های شدت زیان عملیاتی را به دو دسته بدنه و دنباله تقسیم‌بندی کرد. در واقع پدیده‌های بحرانی، حوادث نادری هستند که در دنباله توزیع قرار می‌گیرند. عقیده و تمرکز اصلی این نظریه بر این است که داده‌های شدت مربوط به حوادث زیان بار عملیاتی که حیات‌یک مؤسسه مالی را تهدید می‌کنند، در دنباله توزیع قرار دارند. برای جداسازی داده‌های شدت زیان عملیاتی به دو دسته بدنه و دنباله توزیع معمولاً دو راه وجود دارد [۹]:

۱. **روش پیشنهادی کمیته بال:** در این روش بازه زمانی (معمولاً یک‌سال) به زیر بازه‌های مجزا تقسیم می‌شود و برای هر زیر بازه بزرگ‌ترین مقدار به‌عنوان مقدار بحرانی در نظر گرفته‌می‌شود.
۲. **روش آستانه:** یک مقدار به‌عنوان آستانه یا مرز انتخاب می‌شود و مقادیر بیشتر از آستانه به توزیع پارتو عمومی برازش می‌شوند. در حقیقت مقادیر کمتر از آستانه به‌عنوان بدنه توزیع و مقادیر بیشتر از آستانه به‌عنوان دنباله توزیع هستند.

از نکات حائز اهمیت انتخاب آستانه مناسب است که در تخمین سرمایه پوششی مورد نیاز برای ریسک عملیاتی نقش به‌سزایی دارد. معمولاً این آستانه به‌صورت تجربی انتخاب می‌شود. مفروضات زیر را در مدل‌سازی داده‌های فرکانس و شدت در نظر می‌گیریم:

۱. داده‌های فرکانس از توزیع‌های آماری گسسته نظیر پوآسن‌یا دو جمله‌ای منفی پیروی می‌کنند.
۲. داده‌های شدت از توزیع‌های آماری پیوسته نظیر نمایی، لاگ نرمال، لجستیک و... پیروی می‌کنند.
۳. بین سلول‌های ماتریس خطوط کسب‌وکار–رویدادهای زیان‌بار همبستگی وجود دارد.
۴. توزیع‌های شدت و زیان از یکدیگر مستقل هستند.

پس از مدل‌سازی داده‌های فرکانس و شدت، باید تابع توزیع زیان تجمیعی را به‌دست‌آوریم. این تابع از درهم آمیختن تابع توزیع‌های شدت و فرکانس به‌دست می‌آید. مناسب‌ترین فرم برای رسیدن به تابع توزیع زیان تجمیعی استفاده از شبیه‌سازی مونت‌کارلو است [۱۷]. اگر برای مدل‌سازی داده‌های شدت زیان از برازش توزیع‌های پیوسته به‌کل داده‌های شدت زیان استفاده

کرده باشیم در این صورت برای تعیین تابع زیان تجمیعی کل به روشی که در ادامه توضیح داده خواهد شدعمل می‌کنیم.

ابتدا فرکانس X را از توزیع‌های گسسته نظیر پوآسن‌یا دوجمله‌ای منفی به‌دست می‌آوریم. سپس به تعداد X شدت زیان از تابع توزیع پیوسته برازش شده انتخاب می‌کنیم. سپس با استفاده از رویکرد ارزش در معرض خطر، داده‌های تولیدی را به‌صورت غیر نزولی مرتبط می‌کنیم و زیان مرتبط با سطح اطمینان ۹۹/۹ درصد را به‌دست می‌آوریم.

اگر برای مدل‌سازی داده‌های شدت زیان از نظریه مقدار بحرانی استفاده کرده باشیم در این صورت برای تعیین تابع زیان تجمیعی کل به روشی که در ادامه توضیح داده خواهد شد، عمل می‌کنیم.

ابتدا آستانه مناسب را با روش‌های تجربی به‌دست می‌آوریم. سپس توزیع پارتو عمومی (GPD)[1] را با استفاده از روش‌های تخمین پارامتر نظیر حداکثر درست نمایی به مقادیر دم توزیع یا مقادیر بزرگتر از مقدار آستانه برازش می‌کنیم. فرکانس X را از توزیع‌های گسسته نظیر پوآسن‌یا دوجمله‌ای منفی به‌دست می‌آوریم. درصدی از X نشان‌دهنده بخشی از داده‌ها است که در بدنه توزیع قرار دارند ($X_{\text{بدنه}}$) . درصدی از X نشان‌دهنده بخشی از داده‌ها است در دنباله توزیع هستند ($X_{\text{دنباله}}$). با جمع تمامی $X_{\text{بدنه}}$ها و $X_{\text{دنباله}}$ زیان کل سالانه به‌دست می‌آید. سرانجام کل ذخیره سرمایه مورد نیاز برای ماتریس ریسک عملیاتی با جمع سلول‌های ماتریس عملیاتی در ضمن در نظر گرفتن همبستگی بین عناصر محاسبه می‌شود [۹].

با توجه به مطالعات قبلی در روش اول، یعنی؛ برازش تابع توزیع‌های پیوسته به‌کل داده‌های شدت زیان عملیاتی، چون حجم داده‌های موجود در بدنه بسیار بیشتر از حجم داده‌های موجود در دنباله است، سبب می‌شود برازش تابع توزیع به‌خوبی انجام نشود و تخمین سرمایه لازم برای پوشش ریسک عملیاتی به‌خوبی تخمین‌زده نشود [۱۷]. در روش دوم، یعنی استفاده از نظریه مقدار بحرانی، روش خاصی برای انتخاب آستانه مناسب وجود ندارد، لذا همین امر سبب تخمین بیش از حد یا کمتر سرمایه مورد نیاز می‌شود. نکته حائز اهمیت در مباحث شبیه‌سازی رسیدن به یک جواب مسأله با کمترین واریانس برآوردگر می‌باشد [۱۵]. در این دو روش واریانس برآوردگر زیاد می‌باشد.

یکی از ایرادات نظریه مقدار بحرانی عدم شناسایی یک آستانه مشخص برای داده‌های زیان عملیاتی است. اکنون در این پژوهش مدلی مبنی بر ابزارهای‌یادگیری ماشین ارائه می‌دهیم تا

[1] Generalized pareto distribution

داده‌های شدت زیان عملیاتی را با استفاده از الگوریتم‌های خوشه‌بندی به دو دسته بدنه و دنباله جداسازی کند و سپس با توزیع‌های مناسب این ذخیره سرمایه را در ضمن در نظر گرفتن همبستگی، محاسبه کند. همچنین مدلی مبنی بر ابزارهای‌یادگیری ماشین ارائه می‌دهیم تا داده‌های شدت زیان عملیاتی را با استفاده از الگوریتم‌های طبقه‌بندی در طبقه های مجزا قرار دهد و سپس با توزیع‌های مناسب این ذخیره سرمایه را در ضمن در نظر گرفتن همبستگی، محاسبه کند.

با افزایش تخلفات در بانک‌ها، ریسک عملیاتی اهمیّت شایان ذکری یافته است. در مجموع ریسک عملیاتی از دسته ریسک‌های غیرمالی است که می‌تواند ادامه حیات بانک‌ها را با خطر مواجه کند. در سال‌های ۱۹۸۰-۱۹۷۰ میلادی بسیاری از وام‌گیرندگان از بانک‌ها نسبت به بازپرداخت قصور خود کوتاهی کردند و این امر مشکلات جدی برای بانک‌ها به وجود آورد. در همین حوالی بانک‌ها پی بردند که ذخیره سرمایه مناسبی برای مواجه با چنین بحران‌هایی در نظر بگیرند. یک شاخص ارائه‌شده توسط بال برای تعیین میزان ذخیره لازم، نسبت سرمایه به دارایی بود که امروزه با نسبت کفایت سرمایه شناخته می‌شود. نسبت کفایت سرمایه، شاخصی است که میزان مقاومت بانک را در شرایط بحرانی در مقابل دارایی‌های بانک مورد بررسی قرار می‌دهد. مهم‌ترین محدودیت این معیار این است این نسبت به ریسک دارایی‌های بانک بی‌توجّه است[۱][۸]. بنابراین کمیته بین‌المللی بال در سپتامبر ۲۰۰۱ تغییرات الزامات کفایت سرمایه را در بیانیه مهمی اعلام نمود[۲]. به‌طور کلی می‌توان گفت دو دسته تحقیق در خصوص ریسک عملیاتی تاکنون انجام‌شده‌است:

۱. دسته اول مقالات و استانداردهای کمیته بال
۲. دسته دوم مقالات علمی

مقاله فارچوت و همکاران[۳](۲۰۰۱) از دسته اولین مقالاتی است که به بررسی رویکرد توزیع زیان می‌پردازد. در این تحقیق داده‌های شدت زیان را با توزیع لاگ نرمال و فراوانی آن‌ها را با

---

[۱] این معیار نسبت به اندازه و حجم فعالیت های بانک بی اعتنا است. اگر چه نسبت بزرگتر سرمایه به دارایی به معنی پوشش بالاتر دارایی ها توسط سرمایه است امّا استفاده از این معیار محدودیت هایی دارد. برای مثال دو صنعت بانکداری را ملاحظه بفرمایید که صنعت اول منابع خود را در دارایی های نقدی سرمایه گذاری کرده است و صنعت دوم منابع خود را به وام اختصاص داده است. اگر فرض شود نسبت داریی به سرمایه دو صنعت برابر باشد، در این صورت مخاطره های احتمالی برای دارایی های صنعت دوم بیشتر است در واقع کفایت سرمایه اولی در مقابله با مخاطرات، بیشتر از دومی است.

[۲] کمیته بین المللی بازل (بال) کمیته ای است که به تعیین مقررات و نظارت بر فعالیت های بانکی می پردازد. وظیفه اصلی این کمیته تدوین استانداردها نظیر نگهداری سرمایه مناسب در بانک‌ها است.

[۳] Frachot et al

توزیع پوآسن مدل‌سازی کرده و در نهایت مقدار یکسانی را برای ذخیره سرمایه در هر بخش در نظر گرفتند [۱].

فونتنوول و همکاران[1](۲۰۰۳) با استفاده از داده‌های جمع‌آوری‌شده اقدام به محاسبه ذخیره سرمایه چند بانک بزرگ بین‌المللی کردند. آن‌ها از خانواده توزیع‌های پارتو برای مدل کردن داده‌های شدت و از توزیع پوآسن برای مدل کردن فرکانس اتفاقات استفاده کردند. درنهایت نشان دادند که مقدار ذخیره سرمایه مورد نیاز برای مقابله با ریسک عملیاتی بیشتر از ذخیره سرمایه مورد نیاز برای ریسک بازار است [۲۲].

چاوز و همکاران[2](۲۰۰۵) به بررسی مسأله مقدار بحرانی پرداختند و در سطح اطمینان ۹۹/۹ درصد ذخیره سرمایه را اندازه‌گیری کردند. آن‌ها نشان دادند به‌دلیل عدم دسترسی به فرم بسته تابع توزیع تجمیعی این ذخیره سرمایه کمتر برآورد می‌شود [۱۲].

جامع‌ترین و کامل‌ترین پژوهش با محوریت دم توزیع با مطالعه‌ی دوتا و پری[3](۲۰۰۷) بر روی داده‌های جمع‌آوری‌شده انجام شد. تمرکز اصلی آن‌ها مدل کردن داده‌های فرین در دم توزیع‌ها بود، چرا که اعتقاد داشتند مسأله اصلی در مدل‌سازی داده‌های شدت، ساختار مدل‌سازی دم داده ها است. از این رو به بررسی و مقایسه رده‌ای از توزیع‌های دنباله پهن نظیر نمایی و پارتو تعمیم‌یافته پرداختند. آن‌ها با استفاده از چولگی و کشیدگی رهنمون به‌سمت خانواده توزیع‌های «جی-اچ» [4]شدند [۱۷].

کاپل و همکاران[5](۲۰۰۸)روشی را معرفی کردند که چگونه می‌توان از ترکیب داده‌های زیان داخلی و خارجی برای تخمین سرمایه مورد نیاز برای ریسک عملیاتی استفاده‌کرد. در همین سال واله و همکاران[6] با درنظر گرفتن کمبود داده‌های زیان عملیاتی برای محاسبه ذخیره سرمایه، با زنجیره‌های مارکف مونت‌کارلو سعی کردند این مشکل را برطرف سازند و به مدل‌سازی این ریسک پرداختند [۱۳].

ژویانگ[7](۲۰۱۱) با در نظر گرفتن همبستگی موجود از دو توزیع برای مدل کردن داده‌های شدت استفاده‌کرد (یکی برای داده‌ها با شدت معمولی و دیگری شدت بیش‌ازحد). در همین سال لی و همکاران[8]به بررسی عوامل بیمه در دارایی و منابع بانک در اندازه‌گیری ریسک عملیاتی

[1] Fontnouvelle et al
[2] Chavez et al
[3] Dutta & Perry
[4] g-h distribution
[5] Chapelle et al
[6] Valla et al
[7] Zhuang
[8] Lee et al

پرداختند و با استفاده از تئوری مقدار بحرانی و ارزش در معرض خطر ذخیره سرمایه برای مقابله با ریسک عملیاتی را مدل‌سازی کردند [۲۷].

دمتیس[1](۲۰۱۸) در حال ایجاد یک سیستم‌یادگیری ماشینی برای اسکن تراکنش‌های کوچک و مشتریان تجاری بزرگ هستند تا از این طریق سرمایه لازم برای ریسک عملیاتی را کاهش دهند [۱۶].

در ایران نیز تحقیقاتی در این زمینه انجام‌گرفته است. در سال ۱۳۸۵ بانک اقتصاد نوین با همکاری متخصصان سوئیسی اقداماتی در خصوص مدیریت ریسک انجام داد و نتایج کار را در قالب چند کتاب ارائه داد. یکی از مهم‌ترین این نتایج ایجاد و توسعه نرم‌افزار مدیریت ریسک عملیاتی بود که پس از بهره‌برداری از آن، نتایج را در قالب کتاب‌های مذکور ارائه نمودند [۲۱].

در سال ۱۳۸۵ امیر عرفانیان و همکاران تحقیقاتی در مورد بانک صنعت و معدن انجام دادند و روش‌های شاخص پایه و استاندارد را برای محاسبه ریسک عملیاتی این بانک ارائه نمودند. آن‌ها معتقد بودند که دو روش مذکور برای بانک‌ها با فعالیت ساده مناسب است ولی دقت کافی را ندارند و برای بانک‌ها با فعالیت‌های پیچیده تجاری لازم است از روش‌های پیشرفته استفاده‌کرد [۱۹].

در سال ۱۳۹۲ کامران پاکیزه و همکاران به بررسی رویکرد توزیع زیان در بانک مفروض پرداختند و با استفاده از توزیع‌های آلفا پایدار و دنباله پهن به مدل‌سازی داده‌های شدت پرداختند و نتایج را در نرم‌افزارR با در نظر گرفتن ضرایب همبستگی اعلام نمودند. طبق نتایج این پژوهش در نظر گرفتن ضریب همبستگی بین گروه حوادث، ذخیره سرمایه را بین۲۰درصد تا ۳۰ درصد کاهش می‌دهد [۲۵].

۳. روش‌شناسی پژوهش

داده‌های پژوهش

داده‌های مورد استفاده در این پژوهش، مربوط به چند بانک استرالیایی، اروپایی، آمریکای شمالی، برزیلی و هندی به‌عنوان یک سبد مربوط به سال‌های ۲۰۰۲ تا ۲۰۰۸ جمع‌آوری شده‌است [۱۱]. در ادامه ویژگی آمار توصیفی داده‌ها و خصوصیات آن‌ها بیان می‌شود.

پرتال جامع علوم انسانی

[1] Demetis

جدول ۲. بانک‌های شرکت‌کننده در پرتفوی

بانک‌های شرکت‌کننده	تعداد	تعداد بانک‌هایی که از روش‌های پیشرفته استفاده کرده‌اند	تعداد بانک‌هایی که از سایر روش‌ها استفاده کرده‌اند
استرالیایی	۱۱	۵	۶
اروپایی	۶۰	۲۰	۴۰
ژاپن	۱۸	۷	۱۱
آمریکای شمالی	۲۳	۱۰	۱۳
برزیل / هند	۹	۰	۹
جمع	۱۲۱	۴۲	۷۹

درجدول۲ نام بانک‌های شرکت‌کننده به تفکیک استفاده از روش‌های اندازه‌گیری ریسک عملیاتی آورده شده‌است.

جدول۳. داده‌های ضرر همراه با سال وقوع آن‌ها

تمام اعضا	قبل ۲۰۰۲	۲۰۰۲	۲۰۰۳	۲۰۰۴	۲۰۰۵	۲۰۰۶	۲۰۰۷	۲۰۰۸
تعداد ضررها								
مبلغ کل ضرر (میلیون‌یورو)	۱۴۰۱۷	۱۰۲۱۶	۱۳۶۹۱	۲۲۱۵۲	۳۳۲۱۶	۳۶۳۸۶	۳۶۶۲۲	۵۵۸۲
تعداد بانک‌های گزارشگر	۳۸۲۹	۱۲۰۶۹	۴۵۶۲	۷۲۱۲	۹۷۴۰	۷۴۴۶	۷۸۷۵	۹۷۱
تعداد زیان در هر بانک	۲۴	۳۵	۵۵	۶۸	۱۰۸	۱۱۵	۱۱۷	۸۴
مبلغ کل ضرر برای هر بانک	۵۸۴	۲۹۲	۲۴۹	۳۲۶	۳۰۸	۳۱۶	۳۱۳	۶۶
(میلیون‌یورو)	۱۶۰	۳۴۵	۸۳	۱۰۶	۹۰	۶۵	۶۷	۱۲
میانگین ضرر در هر بانک	۰/۲۷	۱/۱۸	۰/۳۳	/۳۳	۰/۲۹	۰/۲۰	۰/۲۲	۰/۱۷
(میلیون‌یورو)								
استفاده از روش‌های پیشرفته	قبل ۲۰۰۲	۲۰۰۲	۲۰۰۳	۲۰۰۴	۲۰۰۵	۲۰۰۶	۲۰۰۷	۲۰۰۸
تعداد ضررها								
مبلغ کل ضرر (میلیون‌یورو)	۹۷۲۱	۸۸۳۵	۱۱۳۰۱	۱۷۰۷۰	۲۲۴۷۶	۲۴۷۵۵	۲۴۱۸۲	۳۵۶۸
تعداد بانک‌های گزارشگر	۳۵۱۲	۱۱۹۶۷	۴۲۷۱	۶۶۱۳	۷۵۱۴	۵۹۱۰	۵۹۹۷	۷۲۰
تعداد زیان در هر بانک	۱۶	۲۶	۳۱	۳۶	۴۰	۴۱	۴۱	۳۲
مبلغ کل ضرر برای هر بانک	۶۰۸	۳۴۰	۳۶۵	۴۷۴	۵۶۲	۶۰۴	۵۹۰	۱۱۲
(میلیون‌یورو)	۲۱۹	۴۶۰	۱۳۸	۱۸۴	۱۸۸	۱۴۴	۱۴۶	۲۳
میانگین ضرر در هر بانک	۰/۳۶	۱/۳۵	۰/۳۸	۰/۳۹	۰/۳۸	۰/۲۴	۰/۲۵	۰/۲۰
(میلیون‌یورو)								
استفاده از سایر روش‌ها	قبل ۲۰۰۲	۲۰۰۲	۲۰۰۳	۲۰۰۴	۲۰۰۵	۲۰۰۶	۲۰۰۷	۲۰۰۸
تعداد ضررها								
مبلغ کل ضرر (میلیون‌یورو)	۴۲۹۶	۱۳۸۱	۲۳۹۰	۵۰۸۲	۱۰۷۴۰	۱۱۶۳۱	۱۲۴۴۰	۲۰۱۴
تعداد بانک‌های گزارشگر	۳۱۷	۱۰۱	۲۹۲	۵۹۹	۲۲۲۶	۱۵۳۶	۱۸۷۸	۲۵۱
تعداد زیان در هر بانک	۸	۹	۲۴	۳۲	۶۸	۷۴	۷۶	۵۲
مبلغ کل ضرر برای هر بانک	۵۳۷	۱۵۳	۱۰۰	۱۵۹	۱۵۸	۱۵۷	۱۶۴	۳۹
(میلیون‌یورو)	۴۰	۱۱	۱۲	۱۹	۳۳	۲۱	۲۵	۵
میانگین ضرر در هر بانک	۰/۰۷	۰/۰۷	۰/۱۲	۰/۱۲	۰/۲۱	۰/۱۳	۰/۱۵	۰/۱۲
(میلیون‌یورو)								

از جدول۳ اطلاعات مربوط به تمامی بانک‌های شرکت‌کننده از قبیل تعداد ضرر و مبلغ ضرر براساس روش های اندازه‌گیری ریسک عملیاتی قابل مشاهده است.

جدول ۴. ویژگی آمار توصیفی داده‌های نوسان

گروه حوادث	مجموع	مینیمم	ماکزیمم	میانگین	واریانس	چولگی	کشیدگی
سرقت داخلی	۱۳۷۴/۳	۳/۵	۹۷۹/۴	۱۵۲/۶۹	۹۹۶۳۶/۶۹	۲/۳۲	۳/۶۵
سرقت خارجی	۸۵۶۴/۹	۱۱/۵	۷۳۱۱/۹	۹۵۱/۶۷	۵۷۳۵۴۵۰/۵	۲/۴۴	۴/۰۱
استخدام و ایمنی محیط کار	۵۷۱۴/۵	۱۲/۸	۳۲۰۳/۴	۶۳۴/۹۴	۱۳۷۱۶۲۸/۷۷	۱/۵۴	۰/۶۴
مشتریان، محصولات و شیوه کسب‌وکار	۵۹۱۴/۹	۴۴/۹	۲۳۸۱	۶۵۷/۲۲	۹۰۹۵۵۷/۸۴	۱/۲۵	-۰/۳۲
آسیب به دارایی‌های فیزیکی	۳۸۲/۹	۱/۹	۲۴۵/۴	۴۲/۵۶	۶۱۵۲/۳۹	۲/۲۲	۳/۳۳
اختلال در کسب‌وکار و خرابی سیستم‌ها	۶۴۲/۳	۴/۶	۲۹۳/۸	۷۱/۳۷	۹۱۰۱/۸۲	۱/۶۵	۱/۳۴
تحویل، اجرا و مدیریت فرآیندها	۹۹۷۰/۵	۶۹/۱	۳۷۴۳/۴	۱۱۰۷/۸۳	۱۴۵۱۱۶۹/۵۸	۱/۳۷	۰/۵۵

در جدول۴ ویژگی آمار توصیفی داده‌های نوسان زیان عملیاتی آورده شده‌است. همانطور که از نام داده‌های نوسان پیدا است، این داده‌ها نمایانگر تعداد رخ داد زیان عملیاتی در خطوط کسب‌وکار بانک هستند و همگی مثبت هستند و واحد اندازه‌گیری ندارند.

جدول ۵. ویژگی آمار توصیفی داده‌ها شدت

گروه حوادث	*مجموع*	*مینیمم*	*ماکزیمم*	*میانگین*	*واریانس*	*چولگی*	*کشیدگی*
سرقت داخلی	۶۰۰/۵	۲/۵	۱۹۸/۵	۶۶/۷۳	۴۷۶۷/۸۵	۰/۸۲	-۰/۶
سرقت خارجی	۷۸۰	۲/۳	۶۰۷/۹	۸۶/۶۶	۳۹۴۲۷/۹	۲/۳۵	۳/۷
استخدام و ایمنی محیط کار	۵۸۳/۴	۱/۷	۳۰۵/۶	۶۴/۸۴	۱۰۷۷۶/۲۹	۱/۶۷	۱/۲۷
مشتریان، محصولات و شیوه کسب‌وکار	۵۱۲۳/۱	۱۸/۷	۲۵۶۵/۱	۵۶۹/۲۴	۷۰۰۷۶۵/۲۴	۱/۸	۱/۶۹
های آسیب به دارایی فیزیکی	۱۳۴	۰/۱	۴۶/۷	۱۴/۹	۳۶۱/۰۵	۰/۷۶	-۱/۲۴
اختلال در کسب‌وکار و هاخرابی سیستم	۱۱۸/۴	۰/۶	۴۸	۱۳/۱۷	۲۲۶/۸	۱/۴۹	۱/۲۳
تحویل، اجرا و مدیریت فرآیندها	۲۴۲۹/۲	۷۱/۵	۷۳۲/۶	۲۶۹/۹	۶۲۷۰۴/۴۹	۱/۲	-۰/۳۲

در جدول۵ ویژگی آمار توصیفی داده‌های شدت زیان عملیاتی آورده شده‌است که این داده‌ها نمایانگر شدت زیان عملیاتی برحسب میلیون یورو و همگی مثبت هستند.

ماتریس تشابه

در الگوریتم‌های خوشه‌بندی، موضوع مهم تشابه[1]یا عدم تشابه[2] داده‌ها است. در واقع داده‌هایی داده‌هایی که باهم شباهت زیادی داشته باشند در یک خوشه قرار می‌گیرند. در این نوع الگوریتم‌ها ماتریسی تحت عنوان ماتریس تشابه وجود دارد که برای شناسایی تشابه یا عدم تشابه خوشه‌ها استفاده می‌شود. ماتریس تشابه یک ماتریس متقارن است زیرا تشابه بین دو نمونه الف و ب تفاوتی با تشابه بین دو نمونه ب و الف ندارد. بنابراین همواره ماتریس تشابه یک ماتریس مثلثی است [۲۶]. شکل بعد نمونه‌ای ازیک ماتریس تشابه را نشان‌می‌دهد.

$$\begin{bmatrix} ۰ & & & & \\ d(۲,۱) & ۰ & & & \\ d(۳,۱) & d(۳,۲) & ۰ & & \\ \dots & \dots & \dots & \dots & \\ d(n,۱) & d(n,۲) & d(n,۳) & \dots & ۰ \end{bmatrix}$$

شکل ۳. ماتریس تشابه

$d(i,j)$ فاصله‌ی دو نمونه i اُم و j اُم را نشان‌می‌دهد. عدد کوچک‌تر نشان‌دهنده تشابه بالا دو نمونه است.

الگوریتم k-میانگین[3]

این الگوریتم از نمونه الگوریتم‌های خوشه‌بندی است. در این الگوریتم با فرض وجود n نمونه و ورودی k به‌عنوان تعداد خوشه‌های نهایی، نمونه‌ها در k خوشه قرار داده می‌شوند. در ابتدا k نمونه به‌صورت اتفاقی از میان کل نمونه‌ها انتخاب می‌شوند. این نمونه‌ها به‌عنوان نماینده k۴ خوشه در نظر گرفته می‌شوند. گاهی‌اوقات به آن‌ها مرکز ثقل۵یا مرکز خوشه نیز گفته‌می‌شود. به کمک معیارهایی مانند فاصله اقلیدسی تشابه هریک از نمونه‌های باقی‌مانده را با k نماینده محاسبه می‌شود و نمونه مورد نظر به هریک نزدیک‌تر بود، به عضویت آن خوشه درمی‌آید [۲۸].

به حداقل رساندن درصد خطا از فرمول زیر محاسبه می‌شود:

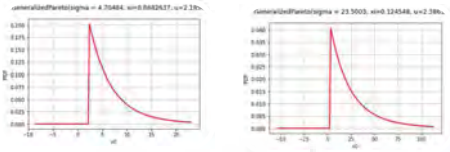
رابطه (۱)

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

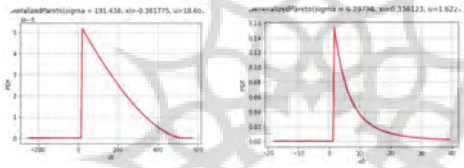
[1] Similarity
[2] Dissimilarity(Distance)
[3] K-Mean Algorithm
[4] Centroid
[5] Center Of Gravity

$$E = \sum_{i=1}^{k} \sum_{p \in c_i} (p - m_i)^2$$

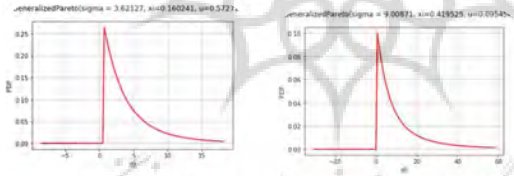
در عبارت فوق E مجذور مربعات خطا برای کلیه نمونه‌ها، p یک نمونه از داده‌ها و m_i نماینده‌یا مرکز ثقل خوشه c_i است. قابل‌ذکر است که در این الگوریتم نمونه‌ها فقط به‌یک خوشه تعلق دارند و نمی‌توانند همزمان متعلق به چند خوشه باشند. در نظریه مقدار بحرانی برای انتخاب آستانه مناسب روش خاص و مناسبی وجود ندارد و این‌کار تجربی انجام‌می‌شود. اکنون با الگوریتم k–میانگین داده‌های هر گروه حوادث ضررساز را به دو دسته تقسیم کنیم و برای هر گروه حوادث یک آستانه مناسب انتخاب کنیم. پس از تقسیم داده‌های شدت به دنباله و بدنه توزیع، داده‌های دنباله را با روش بوت استرپ افزایش می‌دهیم و به توزیع پارتو عمومی برازش می‌کنیم.



سرقت خارجی **سرقت داخلی**

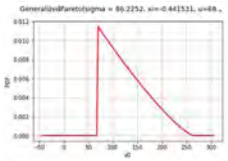


مشتریان، محصولات و شیوه کسب وکار **استخدام و ایمنی محیط کار**



اختلال در کسب وکار و خرابی سیستم ها **آسیب به دارایی های فیزیکی**

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی



تحویل، اجرا و مدیریت
فرآیندها

شکل ۴. برازش داده‌های شدت گروه حوادث زیان بار به توزیع پارتو عمومی

در شکل (۴) برای هر دسته از داده‌های شدت گروه حوادث‌ساز یک آستانه مناسب انتخاب می‌شود، بدین منظور که داده‌ها با الگوریتم k-میانگین به دو دسته بدنه و دنباله جداسازی شده و مقادیر دنباله به توزیع پارتو عمومی برازش می‌شوند. تمامی مراحل در قالب یک فرآیند در کد (۱) قابل دسترسی است. بعد از برازش داده‌ها به توزیع پارتو عمومی[1]، ارزش در معرض خطر را در سطح اطمینان ۹۹/۹درصد محاسبه می‌کنیم[2].

جدول ۶. ارزش در معرض خطرتوزیع پارتو

گروه حوادث	$Var_{99.9\%}$
سرقت داخلی	۱۱۶/۰۸۱
سرقت خارجی	۲۶/۷۲۹
استخدام و ایمنی محیط کار	۶۸/۵۱۱
مشتریان، محصولات و شیوه کسب‌وکار	۵۶۹/۲۳۳
های فیزیکی‌آسیب به دارایی	۶۰/۶۱۳
هااختلال در کسب‌وکار و خرابی سیستم	۱۹/۶۲۵
تحویل و اجرا و مدیریت فرآیندها	۲۴۸/۶۸۹

در جدول (۶) با استفاده از مفهوم ارزش در معرض خطر که در ابتدا توضیح داده‌شد برای هر یک از داده‌های شدت گروه حوادث‌ساز با استفاده ازتوزیع پارتو عمومی، ارزش در معرض خطر محاسبه می‌شود که مقادیر بر حسب میلیون یورو می‌باشند.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

[1] Generalized pareto distribution

[2] برای مشاهده قطعه کد این قسمت می توان به پیوست مراجعه نمود.(قطعه کد۱)

طبقه‌بندی خطوط کسب‌وکار داده‌های شدت[۱]

بعد از تئوری مقدار بحرانی، اکنون با استفاده از روش‌های مبنی بر یادگیری ماشین، خطوط کسب‌وکار را در کلاس‌های مشخصی طبقه‌بندی می‌شود تا سرمایه لازم برای کلاس‌های مشخص‌شده تخمین زده‌شود [۲۶،۲۸].

درعلم یادگیری‌ماشین جداسازی و تفکیک داده‌ها به دو مجموعه داده‌های آموزشی[۲] و داده‌های آزمایشی[۳]، موضوع مهمی برای سنجش عملکرد یک الگوریتم به حساب می‌آید. در مسائل طبقه‌بندی[۴] این موضوع بسیار قابل توجه‌است. لذا قبل از طبقه‌بندی داده‌های شدت زیان عملیاتی، این داده‌ها به دو مجموعه آموزشی و آزمایشی تفکیک شده‌اند. برای جداسازی داده‌ها به دو مجموعه داده‌های آموزشی و داده‌های آزمایشی، ابتدا یک سطر ماتریس ریسک عملیاتی ، به عنوان داده‌های آزمایشی در نظر گرفته شده‌است و بقیه سطرها به عنوان داده‌های آموزشی در نظر گرفته شده‌است. به دلیل این که ممکن است این داده‌ها با ترتیب خاصی بسته به معیارهایی همچون زمان، مرتب شده باشند، برای کنترل واریانس کار این انتخاب را تصادفی انجام داده ایم. گفتنی است در هر مرحله انتخاب تصادفی، مدل انتخاب شده با استفاده از معیار R-squared و تابع Model.Score مورد سنجش قرار گرفته است و مدلی انتخاب شده‌است که خطای کمتری داشته باشد[۵]. در هنگام استفاده از الگوریتم‌های یادگیری‌ماشین هایپر پارامترها[۶] قبل از شروع و اجرای فرآیند مشخص می‌شوند و پارامترهای الگوریتم در طول فرآیند مشخص‌می شوند. اگر در تحقیق حاضر از داده‌های شبیه‌سازی شده استفاده می‌کردیم یا حجم داده‌ها بسیار بود، از روش‌هایی مانند RandomizedSearchCV برای مشخص کردن هایپر پارامترها استفاده می‌شد، بدین منظور که یک نمونه برای هایپرپارامترها انتخاب می‌شد و با محدود کردن فضای جواب بهترین حالت ممکن انتخاب می‌شد. اما باتوجه به محدود بودن داده‌ها، با آزمون و خطا برای الگوریتم k-نزدیک‌ترین همسایه هایپر پارامتر که همان تعداد درختان تصمیم جنگل هستند برابر تعداد سطرهای ماتریس ریسک عملیاتی در نظر گرفته شده‌است[۷].

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

---

[۱] برای مشاهده قطعه کد این قسمت می توان به پیوست مراجعه نمود.(قطعه کد۲)

[2] Train Set

[3] Test Set

[4] Classification

[۵] قطعه کد۲ گویای این مطلب است. گفتنی است نتیجه ارزیابی ۶۷ درصد قابل قبول است.

[6] hyper parameters

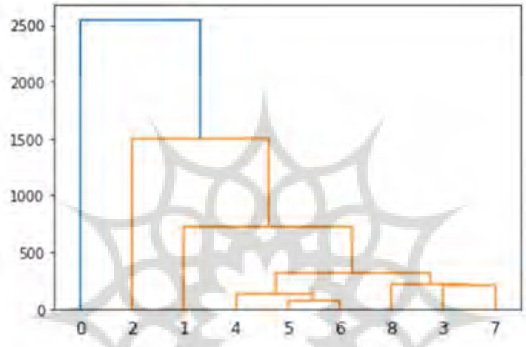
[۷] برای مشاهده این مطلب می توان به پیوست، قطعه کد ۲ رجوع کرد.

الگوریتم k-نزدیک‌ترین همسایه (knn)[1]

یکی از رایج‌ترین کاربردهای الگوریتم knn، تشخیص الگو است. برای یک داده‌ی آزمایشی الگوریتم به‌دنبال k نمونه مشابه از بین نمونه‌ها می‌گردد. نزدیکی دو نمونه با به‌دست‌آوردن تشابه یا فاصله دو نمونه محاسبه می‌شود. پس از یافتن k داده‌ی مشابه با نمونه‌ی آزمایشی، برچسب کلاس داده‌ی آزمایشی مشخص می‌شود [۲۶،۲۸].

درخت تصمیم[2]

درخت تصمیم برای طبقه‌بندی استفاده می‌شود. همانطور که از نام آن مشخص است، این درخت از تعدادی شاخه و گره تشکیل‌شده‌است. برگ‌ها بیانگر کلاس هستند. در هر یک از گره های دیگر بر اساس صفات خاصه تصمیم‌گیری به عمل می‌آید. درخت تصمیم یک روش ساده و مهم در طبقه‌بندی است زیرا شکل درخت تصمیم گویای همه‌چیز است [۱۴،۲۶،۲۸].



شکل ۵. درخت تصمیم برای خطوط کسب‌وکار داده‌های شدت

در شکل۵ خطوط کسب‌وکار داده‌های شدت زیان عملیاتی با استفاده از درخت تصمیم در پنج طبقه قرار گرفته اند که در درخت تصمیم فوق، عدد صفر نمایانگر خط کسب‌وکار تأمین مالی، عدد یک نمایانگر خط کسب‌وکار خرید و فروش، عدد دو نمایانگر خط کسب‌وکار بانکداری خرده فروشی، عدد سه نمایانگر خط کسب‌وکار پرداخت و تسویه، عدد چهار نمایانگر خط کسب‌وکار خدمات نمایندگی، عدد پنج نمایانگر خط کسب‌وکار بانکداری تجاری، عدد شش نمایانگر خط کسب‌وکار مدیریت دارایی‌ها، عدد هفت نمایانگر خط کسب‌وکار کارمزد خرده‌فروشی و عدد

[1] k-Nearest Neighbors
[2] Decision Tree

هشت نمایانگر خط کسب‌وکار غیر مرتبط است. در الگوریتم درخت تصمیم هایپر پارامتر برابر تعداد سطرهای ماتریس ریسک عملیاتی در نظر گرفته شده‌است.

رگرسیون لجستیک[1]

رگرسیون لجستیک یک مدل خاص از مدل رگرسیون خطی عمومی است. رگرسیون‌لجستیک توسط تابع لجستیک تعریف می‌شود. تابع لجستیک به فرم زیر است:

رابطه (۲)

$$\sigma: \mathbb{R} \to \left(۰,۱\right)$$

$$\sigma(t) = \frac{۱}{۱ + e^{t}}$$

که در آن $\mathbb{R}$ مجموعه اعداد حقیقی و $t\epsilon\mathbb{R}$.

مدل خطی رگرسیون به فرم رابطه (۳) است.

رابطه (۳)

$$y_i = \beta_۰ + \beta_۱ x_{۱,i} + \cdots + \beta_k x_{k,i}$$

که در آن $\beta_۰$ عرض ازمبدا، β_i شیب خط و $x_{k,i}$ متغیرهای تصادفی می‌باشند. با در نظر گرفتن تابع لجستیک رگرسیون لجستیک به فرم زیر تعریف می‌شود [۲۸]:

رابطه (۴)

$$\Pr\left(y_i = ۱ \middle| x^{\rightarrow}{}_i; \beta^{\rightarrow}\right) = \sigma(\beta_۰ + \beta_۱ x_{۱,i} + \cdots + \beta_k x_{k,i})$$

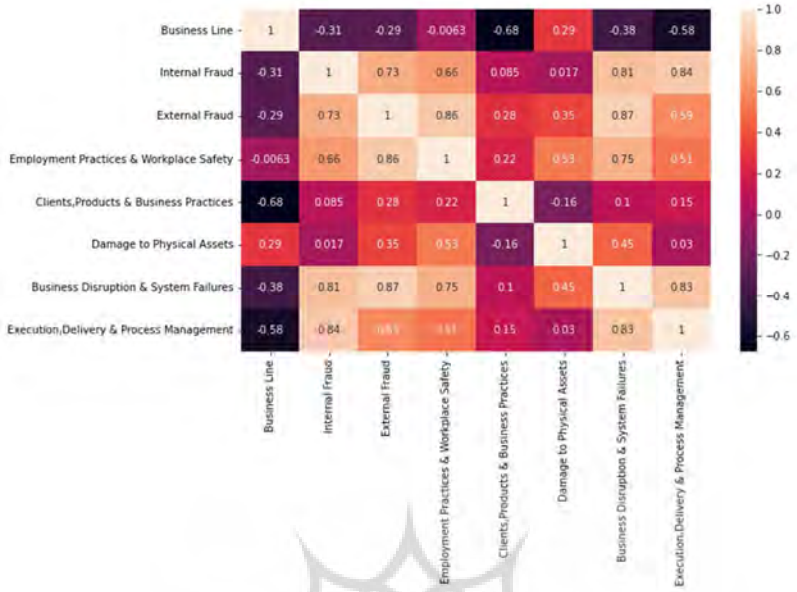
ماتریس همبستگی[2]

برای تشکیل ماتریس همبستگی، متغیرها همان ویژگی‌های مجموعه داده هستند. تعداد سطرها و ستون‌های این ماتریس برابر تعداد ویژگی‌های مجموعه داده‌است. هر سلول این ماتریس دارای رنگ مشخصی است که در بازه منفی یک تا یک قرار دارند. هرچقدر عدد رنگ به منفی یک نزدیک‌تر باشد، یعنی در محل تقاطع، دو ویژگی با هم ارتباط معکوس دارند و بالعکس.

[1] Logistic regression

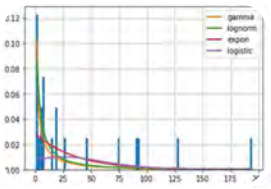
[2] Correlation Matrix

هر ویژگی با خودش همبستگی حداکثری دارد، لذا تمامی عناصر قطر اصلی این ماتریس یک هستند و این مطلب گویای این است که ماتریس همبستگی متقارن است [۱۴،۲۸].

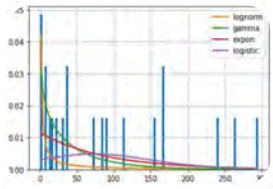


شکل ۶. ماتریس همبستگی برای داده‌های شدت

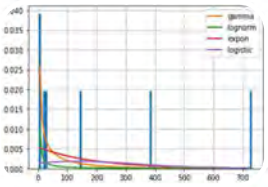
در شکل (۶) مجموعه داده‌های شدت زیان عملیاتی در قالب ماتریس همبستگی پیاده سازی شده‌است. لذا داده‌هایی که همبستگی حداکثری داشته باشند در یک طبقه قرارمی‌گیرند. با اجرای الگوریتم درخت تصمیم والگوریتم knn، ضمن استفاده از ماتریس همبستگی، خطوط کسب‌وکار داده‌های شدت به پنج کلاس تقسیم می‌شوند. کلاس اول شامل«پرداخت و تسویه، کارمزد خرده فروشی و غیر مرتبط»، کلاس دوم شامل «خدمات نمایندگی، بانکداری تجاری و مدیریت دارایی ها»، کلاس سوم شامل «تأمین مالی»، کلاس چهارم شامل «خرید و فروش» و کلاس پنجم شامل «بانکداری خرده‌فروشی». نتیجه طبقه‌بندی را در قالب مدل رگرسیون لجستیک، درخت تصمیم و جنگل تصادفی مورد سنجش قرار می‌دهیم نتایج بر درستی طبقه‌بندی دلالت دارند. برای مشاهده این موضوع قطعه کد (۲) را می‌توان مشاهده نمود. در مرحله‌ی آخر میزان سرمایه مورد نیاز برای هر پنج کلاس محاسبه می‌شود.



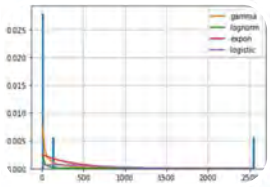
کلاس دوم



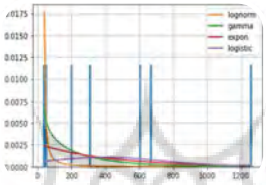
کلاس اول



کلاس چهارم



کلاس سوم



کلاس پنجم

شکل ۷. برازش داده‌های کلاس‌های‌طبقه‌بندی شده به توزیع‌های پیوسته

در شکل (۷) کلاس‌های به وجود آمده از مرحله طبقه‌بندی داده‌های شدت زیان عملیاتی، به توزیع‌های پیوسته آماری برازش‌می‌شوند.

جدول ۷. بهترین توزیع منتخب برای داده‌های کلاس‌های طبقه‌بندی شده

نام گروه کلاس	بهترین توزیع	پارامتر مکان	پارامتر مقیاس	مجذور مربعات خطا
اول	لاگ نرمال	۰/۳۹	۴/۰۳	۰/۰۰۵
دوم	گاما	۰/۵۹	۵۴/۳۸	۰/۰۱
سوم	گاما	۰/۰۹۹	۷۳۷/۴۳۲	۰/۰۰۰۴
چهارم	گاما	۲/۶۹	۱۸۲/۴۳۲	۰/۰۰۱
پنجم	گاما	۳۳/۹۹	۱/۰۳	۰/۰۰۰۷

در جدول (۷) پس از برازش کلاس‌های طبقه‌بندی شده به توزیع‌های پیوسته آماری بهترین توزیع انتخاب می‌شود. (ملاک انتخاب توزیع مناسب‌تر، توزیعی است که مجذور مربعات خطای

کمتری داشته باشد) سپس پارامترهای توزیع با روش حداکثر درست نمایی تخمین زده‌می‌شود (درقسمت پیوست می‌توان به کد (۲) مراجعه نمود).

جدول ۸. محاسبه ارزش در معرض خطر برای کلاس‌های مشخص‌شده

گروه کلاس	$Var_{99.9\%}$
اول	۲۰۳۲۷۹۹/۴۳۳
دوم	۲۳۳/۴۴۰
سوم	۹۹۰/۳۶۳
چهارم	۱۵۹۶/۰۵۶
پنجم	۵۶/۴۵۱

در جدول (۸) با استفاده از مفهوم ارزش در معرض خطر که در ابتدا توضیح داده‌شد برای هر یک از کلاس‌های طبقه‌بندی شده، ارزش در معرض خطر محاسبه می‌شود که مقادیر بر حسب میلیون یورو می‌باشند. لازم به ذکر است که در قالب یک فرآیند این مراحل در قطعه کد (۲) قابل دسترسی است.

۴. تحلیل داده‌ها و یافته‌ها

روش اول

در این روش از تئوری مقدار بحرانی به‌همراه الگوریتم‌های یادگیری ماشین (خوشه‌بندی) استفاده شده‌است. برای هر گروه حوادث ضررساز داده‌ها به دو دسته دنباله و بدنه توزیع تفکیک شده و به‌منظور افزایش داده‌ها از بوت استرپ استفاده شده‌است و با استفاده از تئوری مقدار بحرانی برای هر گروه حوادث ضررساز مقدار سرمایه را تخمین زده شده‌است. گفتنی است که مقادیر تخمین زده شده برای سرمایه به‌صورت یک فاصله اطمینان ۹۵درصد و برحسب میلیون یورو می‌باشد. همچنین به منظور کاهش واریانس برآوردگر از روش بوت استرپ استفاده شده‌است.

جدول ۹. مقادیر تخمین زده‌شده از روش اول

گروه حوادث	مقدار سرمایه تخمین زده‌شده($M€$)
سرقت داخلی	(۷۶/۷۸۰ و ۷۱/۰۰۳)
سرقت خارجی	(۱۸/۷۱۳و۱۹/۳۴۰)
استخدام و ایمنی محیط کار	(۵۶/۵۵۴و۵۸/۵۰۱)
مشتریان، محصولات و شیوه کسب‌وکار	(۳۰۵/۷۳۴و۳۱۷/۷۵۴)
آسیب به دارایی‌های فیزیکی	(۴۵/۲۹۷و۴۶/۹۲۸)
اختلال در کسب‌وکار و خرابی سیستم‌ها	(۱۲/۰۴۳ و ۱۳/۰۳۰)
تحویل و اجرا و مدیریت فرآیندها	(۱۱۵/۸۰۷و۱۲۱/۱۹۳)

در جدول (۹) میزان سرمایه لازم برای مقابله با ریسک عملیاتی به تفکیک گروه حوادث ساز محاسبه شده‌است بدین منظور که از تفاضل میانگین توزیع پارتو عمومی و ارزش در معرض خطر این سرمایه مورد نیاز بدست‌می‌آید.

روش دوم

در این روش خطوط کسب‌وکار با الگوریتم‌های یادگیری ماشین (طبقه‌بندی) به پنج کلاس تفکیک می‌شود. گفتنی است که مقادیر تخمین زده‌شده برای سرمایه به‌صورت یک فاصله اطمینان ۹۵درصد و برحسب میلیون یورو می‌باشد.

جدول ۱۰. مقادیر تخمین زده‌شده از روش دوم

نام کلاس	شده(مقدار سرمایه تخمین زده $M€$)
اول	(۲۰۳۲۷۱۴/۰۳۲ و ۲۰۳۲۶۳۷/۹۹۵)
دوم	(۱۹۶/۳۹۱و۲۰۲/۵۷۳)
سوم	(۹۳۶/۳۶۸و۹۵۷/۱۸۹)
چهارم	(۱۰۸۶/۶۱۴و۱۱۲۱/۲۵۳)
پنجم	(۲۰/۳۹۴و۲۱/۱۳۹)

در جدول (۹) میزان سرمایه لازم برای مقابله با ریسک عملیاتی به تفکیک کلاس‌های طبقه‌بندی شده محاسبه شده‌است بدین منظور که از تفاضل میانگین توزیع های برازش شده و ارزش در معرض خطر این سرمایه مورد نیاز به‌دست‌می‌آید.

۵. بحث و نتیجه‌گیری

در این قسمت نتایج حاصل از روش‌های گفته‌شده بیان می‌شود و به بحث و مقایسه این روش‌ها پرداخته می‌شود. در اکثر پژوهش‌های پیشین از روش های شاخص پایه و استاندارد برای اندازه‌گیری ریسک عملیاتی استفاده کرده‌اند، پژوهش حاضر با ارائه دو مدل بدون آنکه نیاز به یادگیری داشته باشد اندازه‌گیری ریسک عملیاتی را در قالب یک مدل که شامل چند فرآیند است را محاسبه کرده است. به عبارت دیگردو روش گفته شده در پژوهش حاضر تنها با گرفتن داده‌ها به عنوان ورودی مدل ، عملیات لازم را انجام داده و به تفکیک گروه حوادث ضررساز، سرمایه لازم برای مقابله با ریسک عملیاتی را تخمین می‌زنند. گفتنی است در تمامی روش‌های پژوهش حاضر، از رویکرد ارزش در معرض خطر برای محاسبه سرمایه پوششی لازم استفاده‌شده‌است. فقدان پایگاه داده منسجم از نقاط ضعف رویکرد توزیع زیان است. بنابراین اگر با یک پایگاه داده منسجم روبه‌رو باشیم و تعداد داده‌های زیان عملیاتی زیاد باشند روش‌های اول و دوم (روش‌های مبنی بر یادگیری ماشین) هزینه محاسباتی را کاهش داده و کارآمدتر خواهند بود. در روش اول

چون از ترکیب الگوریتم‌های یادگیری ماشین و تئوری مقدار بحرانی استفاده شده‌است، با افزایش حجم داده‌ها کارایی این الگوریتم‌ها بیشتر می‌شود. به‌طور کلی روش اول در تخمین سرمایه موردنیاز مناسب‌تر و منسجم‌تر است زیرا نسبت به سایر روش‌ها تخمین کمتری را لحاظ می‌کند. در واقع روش اول داده‌های موجود در دم توزیع را نیز لحاظ‌می‌کند و این امر سبب می‌شود که سرمایه مورد نیاز دقت بیشتری داشته باشد. روش دوم با طبقه‌بندی داده‌های شدت زیان عملیاتی منجرمی‌شود تا داده‌ها در کلاس‌هایی قرارگیرند که بیشترین همبستگی را داشته باشند و با لحاظ همبستگی بین داده‌ها سرمایه مورد نیاز برای ریسک عملیاتی را تخمین بزند. باگسترش موسسات مالی اسلامی، استفاده از روش‌های اندازه‌گیری ریسک عملیاتی مورد توجه این موسسات قرار گرفته است. وناردوز[1] اصطلاح بانکداری اسلامی را اصطلاحی مشابه با موافقت نامه‌های بال می‌داند که با تأسیس اولین بانک اسلامی در مصر درسال۱۹۶۳ گسترش یافته و با سرعت بیشتری در کشورهای اسلامی از جمله ایران و جنوب شرقی آسیا در حال گسترش است. در بانکداری ایران اقداماتی بر مبنای بال[1] و بال[2] از سال۱۳۸۲ برای مدیریت ریسک عملیاتی صورت گرفته است. باتوجه به گزارشات پژوهش‌های اتاق ایران مبنی بر این که ریسک عملیاتی بانکداری ایران نسبت به بقیه کشورهای خاورمیانه بالا است، لذا رویکردهای ارائه‌شده در پژوهش حاضر کاربردی و قابل استفاده است.

۶. پیشنهادها و محدودیت‌ها

برای تحقیقات و پژوهش‌های آینده پیشنهاد می‌گردد که:

۱. در روش اول که الگوریتم های‌یادگیری ماشین (خوشه‌بندی) با تئوری مقدار بحرانی ترکیب شد، سایر الگوریتم‌ها نیز بررسی شوند.

۲. در روش دوم به‌جای استفاده از روش پارامتری و ترکیب آن با الگوریتم‌های طبقه‌بندی از روش تئوری مقدار بحرانی استفاده‌کنند و این روش را با الگوریتم‌های طبقه‌بندی ترکیب کنند.

پیوست

کد ۱

https://colab.research.google.com/drive/19llanG16QUAYtI2AC9TcNPpFm2s-IqGQ?usp=sharing

کد ۲

https://colab.research.google.com/drive/1AUPlvHpy1TAfcAjDpSvem5tbYHckRYrL?usp=sharing

سپاسگزاری

از کلیه افرادی که ما را در انجام این پژوهش یاری نمودند تشکر می‌نماییم. در این پژوهش از سازمان، نهاد یا شخصی کمک مالی دریافت نشده است.

[1] Venardos

منابع

1.Frachot, A.,Georges, P.,Roncalli, T. (2001). Loss Distribution Approach for operational risk, National Institute of Statistics and Economic Studies (INSEE) - Center for Research in Economics andStatistics(CREST), Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1032523

2.Farhadian, A., Rostami, M., Nilchi, M. (2021). Compare Canonical Stochastic Volatility Model of focal MSGJR-GARCH to Measure the volatility of Stock Returns and Calculating VaR. *Journal of Financial Management Perspective*,10(32),131-158. In persian.

3.Shahriari, A.A., Karimzade, S., Bahmanesh, R. (2021). Stock portfolio optimization in fireworks algorithm using risk value and comparison with Particle Swarm Optimization (PSO). *Journal of Financial Management Perspective*.11(35),9-37. In persian.

4.Johnemark, A. (2012). Modeling Operational Risk.(Degree Projectin Mathematical Statistics). Royal Institute of Technology School of Engineering Sciences, Sweden.

5.Afambo,E. (2006). Operational Risk Capital Provisions for Banks and Insurance Companies(Degree of Doctor of Philosophy in the Robinson College of Business). Department of Risk Management and Insurance, Georgia State University,USA.

6.Basel Committee on banking supervision (2005). International Convergence of Capital Measurement and Capital Standards,A RevisedFramework - Comprehensive Version, Retrieved from https://www.bis.org/publ/bcbs128.htm

7.Basel Committee on banking supervision (2003). Sound Practicesfor The Management and Supervision of Operational risk. Retrieved from https://www.bis.org/publ/bcbs96.htm

8.Basel Committee on banking supervision (2001).Working Paperon The Regulatory Treatment of Operational Risk. Retrieved from https://www.bis.org/publ/bcbs_wp8.pdf

9.Basel Committee on Banking Supervision (2009). Bank for International Settlements. Results from the 2008 Loss Data Collection Exercise for Operational Risk. Retrieved from https://www.bis.org/publ/bcbs160.htm

10.Chavez,V., Embrechts,P., Neslehova,J. (2005). Quantitative models for Operational Risk. Journal of Banking and Finaance,30(10),2635-2658.

11.Chapell,A., Crama,Y., Hubner, J., Peters, J. (2004). Operational Risk:Implications for Risk Measurement and Management in the Financial Sector.NBB Working Paper, 51, National Bank of Belgium, Brussels, 17-18May2004, Financial Management.

12.Dhruv, D.T. (2019). Credit Risk Anal Ysis Using Machin Learning And Neural Networksneural Networks. (Submitted in partial fulfillment of the requirements for the degree of master of science in Mathematical Sciences) Open Access Master's Report, Michigan Technological University,USA.

13.Bakassa-Traore, D.E. (2015). Computing Value-At-Risk And Expected Shortfall in Operational Risk. (Master of Science in the Program of Applied Mathematics). Ryerson University,Canada.

14.Demetis, D.S. (2018). Fighting money laundering with technology: A case study of Bank X in the UK. *Decision Support Systems*, 105, 96-107.

15.Dutta, K., Perry, J. (2007). A Tail of Tails: An Empirical Analysis of Loss Distribution Models for Estimating Operational Risk Capital. Working Papers, No. 06-13, Federal Reserve Bank of Boston, Boston, MA.July2007, Financial Management.

16.Dahlberg, E. (2015). Bayesian Inference Methods in Operational Risk. (Master of Mathematical Statistics). School of Engineering Sciences, Sweden.

17.Erfanian, A. (2006). Comparative study and implementation of operational risk measurement models in industry and Mine Bank. *Sharif Scientific Quarterly*. 34,116-143,In persian.

18.Eghtesad Novin Bank (2006). Asset management - debt and liquidity risk in financial institutions. Iran, Financial organizations – Management Books.

19.Fontnouvelle, D.P., Jordan, J., Rosengren, E. (2003, April 29). Using Loss Data to Quantify Operational Risk. Federal Reserve Bank of Boston. Paper SSRN,1-32.

20.Gabriella, L. (2011). Regulation and practice of managing bank's operational risk.(PhD School in Business and Management).earningBudapest University, Magyarorszag.

21.Pakize, K., Nosrati, H. (2014). Estimation of operational risk capital reserve in banking industry with loss distribution approach. *Journal of financial engineering*. 20,1-26. In persian.

22.Bhatia, P. (2019). Data mining and Data Warehousing. India: University books.

23.Modarresi, N., Peymani, M., Darvishi, M. (2021). Estimation of Conditional Value at Risk under Stochastic Volatility Levy Processes for Tehran Stock Market. *Journal of Financial Management Perspective*. 11(34),69-94. In persian.

استناد

اکبری، مهدی و احمدرضا، یزدانیان (۱۴۰۲). یادگیری ماشین در تخمین سرمایه پوششی ریسک‌عملیاتی بانک‌ها با رویکرد توزیع زیان. *چشم‌انداز مدیریت مالی*، ۱۳(۴۲)، ۹-۳۴.

Citation

Akbari, Mahdi & Yazdanian, Ahmadreza (2023). Machine learning in estimating operational risk coverage capital of banks with a loss distribution Approach. *Journal of Financial Management Perspective*, 13(42), 9 - 34. (*in Persian*)