

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

یادگیری ماشین در تخمین سرمایه پوششی ریسک عملیاتی بانک‌ها با رویکرد توزیع زیان

مهدی اکبری*، احمدرضا یزدانیان**

چکیده

مقاله پژوهشی

BY

CC

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 توجه 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 ضرر ساز نیز مورد بررسی قرار گرفته‌است. برای این پژوهش داده‌های زیان عملیاتی مربوط به یک مجموعه بانکداری شامل چند بانک آسیایی، اروپایی و آمریکایی به‌کار گرفته شده‌است. نتایج پژوهش حاکی از آن است که رویکرد توزیع زیان با ترکیب تئوری مقدار بحرانی و الگوریتم‌های یادگیری ماشین (خوشه‌بندی)، همچنین رویکرد توزیع زیان با ترکیب الگوریتم‌های یادگیری 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 در ۲۶ ژوئن ۱۹۷۴ بانک هرستات^۱ را منحل کردند و مدتی پس از آن کمیته نظارت بر امور بانکی بازل^۲ 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 چه بسا به ورشکستگی مؤسسه مالی بیانجامد [۹]. برای داشتن مقایسه‌ای بین ریسک‌های عملیاتی، بازار و اعتباری می‌توان گفت ریسک عملیاتی با عناصری مانند افراد، فرآیندها، سیستم‌ها، ریسک بازار با عناصری مانند اوراق بهادار 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. اما با توجه به افزایش گروه‌های حوادث در ریسک عملیاتی حداکثر زیان ناشی از این دسته ریسک نامشخص و نامحدود است [۹].



شکل ۱. ریسک عملیاتی در بازل ۲

از شکل (۱) مشاهده می‌گردد که طبق بیانیه‌های کمیته‌بال، بانک‌ها می‌توانند الزامات کفایت سرمایه خود را با توجه به ریسک‌های عملیاتی، بازار، نقدینگی و اعتباری تعیین کنند. این کمیته برای محاسبه ریسک عملیاتی روش‌های شاخص پایه^۳، استاندارد^۴ و پیشرفته^۱ را پیشنهاد می‌دهد

^۱ Herstatt

^۲ Basel

^۳ Basic Indicator Approach

^۴ Standardized Approach

[۹]. رویکرد توزیع زیان در سال ۲۰۰۱ توسط کمیته بازل ارائه شد که در آن با استفاده از ماتریس ریسک عملیاتی^۱، توزیع‌های آماری فراوانی و شدت زیان سالانه به‌طور جداگانه تخمین زده می‌شود، سپس با ترکیب این دو تابع توزیع، تابع توزیع زیان تجمعی به‌دست می‌آید. با کم کردن ارزش در معرض خطر ۹۹/۹ درصد از میانگین تابع توزیع زیان تجمعی، ذخیره سرمایه برای هر سلول به‌دست می‌آید، سپس با در نظر 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. در مدل‌سازی 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 به سوی 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 طور خلاصه می‌توان ضرورت و اهمیت بررسی ریسک عملیاتی را «جهت‌گیری فعالیت‌های بانکی به سمت فعالیت‌های بانکی الکترونیک»، «دستورات مدیریت ریسک» و «ادغام و قبضه مالکیت» دانست [۸].

^۱Advanced Measurement Approach

^۲BLETM

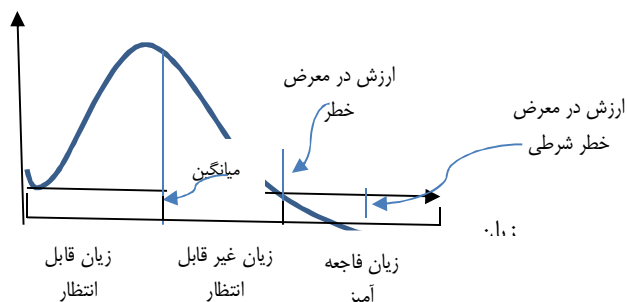
^۳Critical value theory

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 به دقت تخمین زده نشود. اکثر تحقیقات قبلی از روش‌های شاخص پایه و استاندارد برای اندازه‌گیری ریسک عملیاتی استفاده کرده‌اند که در نهایت ذخیره سرمایه را به طور دقیق تخمین ندهاند. در این پژوهش درصد پیاده‌سازی رویکرد توزیع زیان بر مبنای الگوریتم‌های یادگیری ماشین هستیم.

۲. مبانی نظری و پیشینه پژوهش

اندازه‌گیری ریسک عملیاتی بر اساس معیار ارزش در معرض خطر انجام می‌شود. ارزش در معرض خطر همان چندک یا کوانتایل ۹/۹۹ درصد تابع توزیع زیان تجمعی در یک سطح اطمینان مشخص است. مطابق استانداردهای بال این دوره زمانی یک سال و سطح اطمینان برابر ۹۹/۹ درصد تعریف شده است. 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 زبانی بیشتر از زیان غیرقابل انتظار [۲،۳،۴،۵،۳۹].

توزیع آمیخته



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

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

پس از جمع‌آوری داده‌های زیان عملیاتی، برای هر سلول ماتریس ریسک عملیاتی، باید یک توزیع پیوسته برای داده‌های شدت زیان و یک 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 به کل داده‌های شدت زیان عملیاتی.

۲. استفاده از نظریه مقدار بحرانی^۱.

برای تخمین پارامترهای توزیع نیز روش‌های زیر وجود دارد:

۱. روش گشتاورها^۲.

۲. روش حداکثر درست‌نمایی^۳.

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 $(Q-Q)$ ^۴.

¹ Critical value theory

² Moments method

³ Maximum likelihood method

⁴ Quartile-Quartile

گاهی اوقات مدیران ریسک برای مدل سازی داده های شدت از نظریه مقدار بحرانی استفاده می کنند. نظریه مقدار بحرانی، همانطور که از نام آن پیدا است، شاخه ای از علم آمار می باشد که به بررسی و مطالعه پدیده های بحرانی می پردازد. در مبحث ریسک عملیاتی، فرض این نظریه این است که می توان توزیع مربوط به داده های شدت زیان عملیاتی را به دو دسته بدنه و دنباله تقسیم بندی کرد. در واقع پدیده های بحرانی، حوادث نادری هستند که در دنباله توزیع قرار می گیرند. عقیده و تمرکز اصلی این نظریه بر این است که داده های شدت مربوط به حوادث زیان بار عملیاتی که حیات یک مؤسسه مالی را تهدید می کنند، در دنباله توزیع قرار دارند. برای جداسازی داده های 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 یک سال) به زیر بازه های مجزا تقسیم می شود و برای هر زیر بازه بزرگترین مقدار به عنوان مقدار بحرانی در نظر گرفته می شود.
۲. **روش آستانه:** یک مقدار به عنوان آستانه یا مرز انتخاب می شود و مقادیر بیشتر از آستانه به توزیع پارتو عمومی برازش می شوند. در حقیقت مقادیر کمتر از آستانه به عنوان بدنه توزیع و مقادیر بیشتر از آستانه به عنوان دنباله توزیع هستند.

از نکات حائز اهمیت انتخاب آستانه مناسب است که در تخمین سرمایه پوششی مورد نیاز برای ریسک عملیاتی نقش به سزایی دارد. معمولاً این آستانه به صورت تجربی انتخاب می شود. مفروضات زیر را در مدل سازی داده های فرکانس و شدت در نظر می گیریم:

۱. داده های فرکانس از توزیع های آماری گسسته نظیر پواسن یا دو جمله ای منفی پیروی می کنند.
۲. داده های شدت از توزیع های آماری پیوسته نظیر نمایی، لاگ نرمال، لجستیک و... پیروی می کنند.
۳. بین سلول های ماتریس خطوط کسب و کار-رویدادهای زیان بار همبستگی وجود دارد.
۴. توزیع های شدت و زیان از یکدیگر مستقل هستند.

پس از مدل سازی داده های فرکانس و شدت، باید تابع توزیع زیان تجمیعی را به دست آوریم. این تابع از درهم آمیختن تابع توزیع های شدت و فرکانس به دست می آید. مناسب ترین فرم برای رسیدن به تابع توزیع زیان تجمیعی استفاده از شبیه سازی مونت کارلو است [۱۷]. اگر برای مدل سازی داده های شدت زیان از برازش توزیع های پیوسته به کل داده های شدت زیان استفاده

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

ابتدا فرکانس X را از توزیع‌های گسسته نظیر پواسن یا دوجمله‌ای منفی به دست 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 به صورت غیر نزولی مرتبط می‌کنیم و زیان مرتبط با سطح اطمینان ۹۹/۹ درصد را به دست 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 به دست می‌آوریم. سپس توزیع پارتو عمومی (GPD)^۱ را با استفاده از روش‌های تخمین پارامتر نظیر حداکثر درست‌نمایی به مقادیر دم توزیع یا مقادیر بزرگتر از مقدار آستانه برازش می‌کنیم. فرکانس X را از توزیع‌های گسسته نظیر پواسن یا دوجمله‌ای منفی به دست می‌آوریم. درصدی از X نشان‌دهنده بخشی از داده‌ها است که در بدنه توزیع قرار دارند ($X_{بدنه}$). درصدی از X نشان‌دهنده بخشی از داده‌ها است در دنباله توزیع هستند ($X_{دنباله}$). با جمع تمامی $X_{بدنه}$ ها و $X_{دنباله}$ زیان کل سالانه به دست 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 به کل 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 ابزارهای یادگیری ماشین ارائه می‌دهیم تا

^۱ 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 ابزارهای یادگیری ماشین ارائه می‌دهیم تا داده‌های شدت زیان عملیاتی را با استفاده از الگوریتم‌های طبقه‌بندی در طبقه‌های 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 اهمیت 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 بی‌توجه است [۸]. بنابراین کمیته بین‌المللی بال در سپتامبر ۲۰۰۱ تغییرات الزامات کفایت سرمایه را در بیانیه مهمی اعلام نمود^۱. به‌طور کلی می‌توان گفت دو دسته تحقیق در خصوص ریسک عملیاتی تاکنون انجام شده است:

۱. دسته اول مقالات و استانداردهای کمیته بال

۲. دسته دوم مقالات علمی

مقاله فارچوت و همکاران^۲ (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 فعالیت‌های بانک بی‌اعتنا است. اگر چه نسبت بزرگتر سرمایه به دارایی به معنی پوشش بالاتر دارایی‌ها توسط سرمایه است اما استفاده از این معیار محدودیت‌هایی دارد. برای مثال دو صنعت بانکداری را ملاحظه بفرمایید که صنعت اول منابع خود را در دارایی‌های نقدی سرمایه‌گذاری کرده است و صنعت دوم منابع خود را به وام اختصاص داده است. اگر فرض شود نسبت دارایی به سرمایه دو صنعت برابر باشد، در این صورت مخاطره‌های احتمالی برای دارایی‌های صنعت دوم بیشتر است در واقع کفایت سرمایه اولی در مقابله با مخاطرات، بیشتر از دومی است.

^۲ کمیته بین‌المللی بازل (بال) کمیته‌ای است که به تعیین مقررات و نظارت بر فعالیت‌های بانکی می‌پردازد. وظیفه اصلی این کمیته تدوین استانداردها نظیر نگهداری سرمایه مناسب در بانک‌ها است.

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

فونتنوول و همکاران^۱ (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. در نهایت 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^۲ (۲۰۰۵) به بررسی مسأله مقدار بحرانی پرداختند و در سطح اطمینان ۹۹/۹ درصد ذخیره سرمایه را اندازه‌گیری کردند. آن‌ها نشان دادند به دلیل عدم دسترسی به فرم بسته تابع توزیع تجمعی این ذخیره سرمایه کمتر برآورد می‌شود [۱۲].

جامع‌ترین و کامل‌ترین پژوهش با محوریت دم توزیع با مطالعه‌ی دوتا و پری^۳ (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 داده‌ها 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 به سمت خانواده توزیع‌های «جی-اچ»^۴ شدند [۱۷].

کاپل و همکاران^۵ (۲۰۰۸) روشی را معرفی کردند که چگونه می‌توان از ترکیب داده‌های زیان داخلی و خارجی برای تخمین سرمایه مورد نیاز برای ریسک عملیاتی استفاده کرد. در همین سال واله و همکاران^۶ با در نظر 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^۷ (۲۰۱۱) با در نظر گرفتن همبستگی موجود از دو توزیع برای مدل کردن داده‌های شدت استفاده کرد (یکی برای داده‌ها با شدت معمولی و دیگری شدت بیش‌ازحد). در همین سال لی و همکاران^۸ 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 مدل سازی کردند [۲۷].

دمتیس^۱ (۲۰۱۸) در حال ایجاد یک سیستم یادگیری ماشینی برای اسکن تراکنش های 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 انجام گرفته 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 مهم ترین این نتایج ایجاد و توسعه نرم افزار مدیریت ریسک عملیاتی بود که پس از بهره برداری از آن، نتایج را در قالب کتاب های مذکور ارائه نمودند [۲۱].

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

در سال ۱۳۹۲ کامران پاکیزه و همکاران به بررسی رویکرد توزیع زیان در بانک مفروض پرداختند و با استفاده از توزیع های آلفا پایدار و دنباله پهن به مدل سازی داده های شدت پرداختند و نتایج را در نرم افزار R با در نظر گرفتن ضرایب همبستگی اعلام نمودند. طبق نتایج این پژوهش در نظر گرفتن ضریب همبستگی بین گروه حوادث، ذخیره سرمایه را بین ۲۰ درصد تا ۳۰ درصد کاهش می دهد [۲۵].

۳. روش شناسی پژوهش

داده های پژوهش

داده های مورد استفاده در این پژوهش، مربوط به چند بانک استرالیایی، اروپایی، آمریکای شمالی، برزیلی و هندی به عنوان یک سبد مربوط به سال های ۲۰۰۲ تا ۲۰۰۸ جمع آوری شده است [۱۱]. در ادامه ویژگی آمار توصیفی داده ها و خصوصیات آن ها بیان می شود.

^۱ 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

در جدول ۲ نام بانک‌های شرکت‌کننده به تفکیک استفاده از روش‌های اندازه‌گیری ریسک عملیاتی آورده شده‌است.

جدول ۳. داده‌های ضرر همراه با سال وقوع آن‌ها

تمام اعضا	قبل ۲۰۰۲	۲۰۰۲	۲۰۰۳	۲۰۰۴	۲۰۰۵	۲۰۰۶	۲۰۰۷	۲۰۰۸
تعداد ضررها	۱۴۰۱۷	۱۰۲۱۶	۱۳۶۹۱	۲۲۱۵۲	۳۳۲۱۶	۳۶۳۸۶	۳۶۶۲۲	۵۵۸۲
مبلغ کل ضرر (میلیون یورو)	۳۸۲۹	۱۲۰۶۹	۴۵۶۲	۷۲۱۲	۹۷۴۰	۷۴۴۶	۷۸۱۷۵	۹۷۱
تعداد بانک‌های گزارشگر	۲۴	۳۵	۵۵	۶۸	۱۰۸	۱۱۵	۱۱۷	۸۴
تعداد زیان در هر بانک	۵۸۴	۲۹۲	۲۴۹	۳۳۶	۳۰۸	۳۱۶	۳۱۳	۶۶
مبلغ کل ضرر برای هر بانک (میلیون یورو)	۱۶۰	۳۴۵	۸۳	۱۰۶	۹۰	۶۵	۶۷	۱۲
میانگین ضرر در هر بانک (میلیون یورو)	-/۳۷	۱/۱۸	-/۳۳	۰/۳۳	-/۲۹	۰/۲۰	-/۲۲	-/۱۷
استفاده از روش‌های پیشرفته	قبل ۲۰۰۲ <th>۲۰۰۲</th> <th>۲۰۰۳</th> <th>۲۰۰۴</th> <th>۲۰۰۵</th> <th>۲۰۰۶</th> <th>۲۰۰۷</th> <th>۲۰۰۸</th>	۲۰۰۲	۲۰۰۳	۲۰۰۴	۲۰۰۵	۲۰۰۶	۲۰۰۷	۲۰۰۸
تعداد ضررها	۹۷۲۱	۸۸۳۵	۱۱۳۰۱	۱۷۰۷۰	۲۲۴۷۶	۲۴۷۵۵	۲۴۱۸۲	۳۵۶۸
مبلغ کل ضرر (میلیون یورو)	۳۵۱۲	۱۱۹۶۷	۴۲۷۱	۶۶۱۳	۷۵۱۴	۵۹۱۰	۵۹۹۷	۷۲۰
تعداد بانک‌های گزارشگر	۱۶	۲۶	۳۱	۳۶	۴۰	۴۱	۴۱	۳۲
تعداد زیان در هر بانک	۶۰۸	۳۴۰	۳۶۵	۴۷۴	۵۶۲	۶۰۴	۵۹۰	۱۱۲
مبلغ کل ضرر برای هر بانک (میلیون یورو)	۲۱۹	۴۶۰	۱۳۸	۱۸۴	۱۸۸	۱۴۴	۱۴۶	۲۳
میانگین ضرر در هر بانک (میلیون یورو)	-/۳۶	۱/۳۵	-/۳۸	-/۳۹	-/۳۸	۰/۲۴	-/۲۵	-/۲۰
استفاده از سایر روش‌ها	قبل ۲۰۰۲ <th>۲۰۰۲</th> <th>۲۰۰۳</th> <th>۲۰۰۴</th> <th>۲۰۰۵</th> <th>۲۰۰۶</th> <th>۲۰۰۷</th> <th>۲۰۰۸</th>	۲۰۰۲	۲۰۰۳	۲۰۰۴	۲۰۰۵	۲۰۰۶	۲۰۰۷	۲۰۰۸
تعداد ضررها	۴۲۹۶	۱۳۸۱	۲۳۹۰	۵۰۸۲	۱۰۷۴۰	۱۱۶۳۱	۱۲۴۴۰	۲۰۱۴
مبلغ کل ضرر (میلیون یورو)	۳۱۷	۱۰۱	۲۹۲	۵۹۹	۲۲۲۶	۱۵۳۶	۱۸۷۸	۲۵۱
تعداد بانک‌های گزارشگر	۸	۹	۲۴	۳۲	۶۸	۷۴	۷۶	۵۲
تعداد زیان در هر بانک	۵۳۷	۱۵۳	۱۰۰	۱۵۹	۱۵۸	۱۵۷	۱۶۴	۳۹
مبلغ کل ضرر برای هر بانک (میلیون یورو)	۴۰	۱۱	۱۲	۱۹	۳۳	۲۱	۲۵	۵
میانگین ضرر در هر بانک (میلیون یورو)	-/۰۷	-/۰۷	-/۱۲	-/۱۲	-/۲۱	۰/۱۳	-/۱۵	-/۱۲

از جدول ۳ 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
تحويل، اجرا و مدیریت فرآیندها	۹۹۷۰/۵	۶۹/۱	۳۷۴۳/۴	۱۱۰۷/۸۳	۱۴۵۱۱۶۹/۵۸	۱/۳۷	۰/۵۵

در جدول ۴ 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.

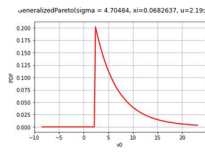
جدول ۵. ویژگی آمار توصیفی داده‌ها شدت

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	۵۱۳۳/۱	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 هاخرایی سیستم	۱۱۸/۴	۰/۶	۴۸	۱۳/۱۷	۲۲۶/۸	۱/۴۹	۱/۲۳
تحويل، اجرا و مدیریت فرآیندها	۲۴۲۹/۲	۷۱/۵	۷۳۲/۶	۲۶۹/۹	۶۲۷۰۴/۴۹	۱/۲	-۰/۳۲

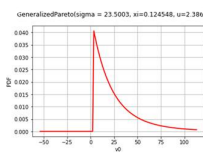
در جدول ۵ ویژگی آمار توصیفی داده‌های شدت زیان عملیاتی آورده شده‌است که این داده‌ها نمایانگر شدت زیان عملیاتی برحسب میلیون یورو و همگی مثبت هستند.

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

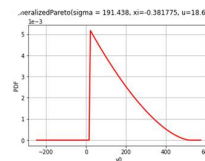
در عبارت فوق E مجذور مربعات خطا برای کلیه نمونه‌ها، p یک نمونه از داده‌ها و m_i نماینده یا مرکز ثقل خوشه C_i است. قابل ذکر است که در این الگوریتم نمونه‌ها فقط به یک خوشه تعلق دارند و نمی‌توانند همزمان متعلق به چند خوشه باشند. در نظریه مقدار بحرانی برای انتخاب آستانه مناسب روش خاص و مناسبی وجود ندارد و این کار تجربی انجام می‌شود. اکنون با الگوریتم k -میانگین داده‌های هر گروه حوادث ضرر ساز 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.



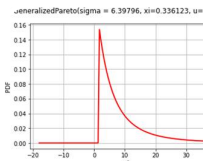
سرقت خارجی



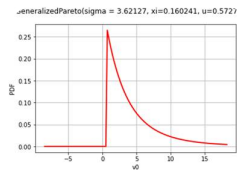
سرقت داخلی



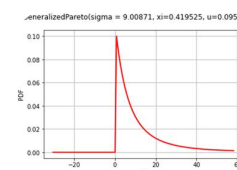
مشتریان، محصولات و شیوه کسب و کار



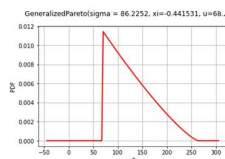
استخدام و ایمنی محیط کار



اختلال در کسب و کار و خرابی سیستم‌ها



آسیب به دارایی‌های فیزیکی



تحويل، 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^۱، ارزش در معرض خطر را در سطح اطمینان ۹۹/۹ درصد محاسبه می‌کنیم.^۲

جدول ۶. ارزش در معرض خطر توزیع 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 فیزیکی آسیب به دارایی	۶۰/۶۱۳
هاختلال در کسب‌وکار و خرابی سیستم	۱۹/۶۲۵
تحويل و اجرا و مدیریت فرآیندها	۲۴۸/۶۸۹

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

^۱ Generalized pareto distribution

^۲ برای مشاهده قطعه کد این قسمت می‌توان به پیوست مراجعه نمود. (قطعه کد ۱)

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 مشخص شده تخمین زده شود [۲۶،۲۸].

در علم یادگیری ماشین 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 توجه است. 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 یادگیری ماشین هاپیر پارامترها^۶ قبل از شروع و اجرای فرآیند مشخص می‌شوند و پارامترهای الگوریتم در طول فرآیند مشخص می‌شوند. اگر در تحقیق حاضر از داده‌های شبیه‌سازی شده استفاده می‌کردیم یا حجم داده‌ها بسیار بود، از روش‌هایی مانند RandomizedSearchCV برای مشخص کردن هاپیر پارامترها استفاده می‌شد، بدین منظور که یک نمونه برای هاپیر پارامترها انتخاب می‌شد و با محدود کردن فضای جواب بهترین حالت ممکن انتخاب می‌شد. اما با توجه به محدود بودن داده‌ها، با آزمون و خطا برای الگوریتم k -نزدیک‌ترین همسایه هاپیر پارامتر که همان تعداد درختان تصمیم جنگل هستند برابر تعداد سطرهای ماتریس ریسک عملیاتی در نظر گرفته شده‌است^۷.

^۱ برای مشاهده قطعه کد این قسمت می‌توان به پیوست مراجعه نمود. (قطعه کد ۲)

^۲ Train Set

^۳ Test Set

^۴ Classification

^۵ قطعه کد ۲ گویای این مطلب است. گفتنی است نتیجه ارزیابی ۶۷ درصد قابل قبول است.

^۶ hyper parameters

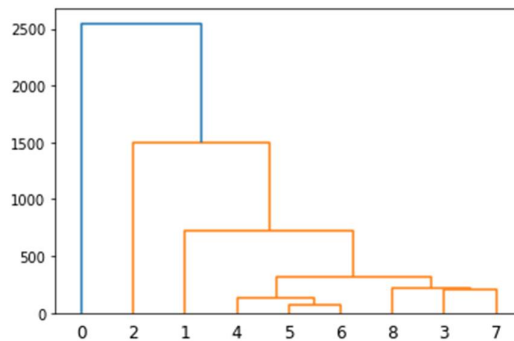
^۷ برای مشاهده این مطلب می‌توان به پیوست، قطعه کد ۲ رجوع کرد.

الگوریتم k -نزدیک‌ترین همسایه (knn) ^۱

یکی از رایج‌ترین کاربردهای الگوریتم knn ، تشخیص الگو است. برای یک داده‌ی آزمایشی الگوریتم به دنبال 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^۲

درخت تصمیم برای طبقه‌بندی استفاده می‌شود. همانطور که از نام آن مشخص است، این درخت از تعدادی شاخه و گره تشکیل شده است. برگ‌ها بیانگر کلاس هستند. در هر یک از گره‌های 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 شکل ۵ 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

^۱ k-Nearest Neighbors

^۲ Decision Tree

هشت نمایانگر خط کسب و کار غیر مرتبط است. در الگوریتم درخت تصمیم هایپر پارامتر برابر تعداد سطرهای ماتریس ریسک عملیاتی در نظر گرفته شده است.

رگرسیون لجستیک^۱

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

رابطه (۲)

$$\sigma: \mathbb{R} \rightarrow (0, 1)$$

$$\sigma(t) = \frac{1}{1 + e^{-t}}$$

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

مدل خطی رگرسیون به فرم رابطه (۳) است.

رابطه (۳)

$$y_i = \beta_0 + \beta_1 x_{1,i} + \dots + \beta_k x_{k,i}$$

که در آن β_0 عرض از مبدا، β_i شیب خط و $x_{k,i}$ متغیرهای تصادفی می باشند. با در نظر گرفتن تابع لجستیک رگرسیون لجستیک به فرم زیر تعریف می شود [۲۸]:

رابطه (۴)

$$\Pr(y_i = 1 | \vec{x}_i; \vec{\beta}) = \sigma(\beta_0 + \beta_1 x_{1,i} + \dots + \beta_k x_{k,i})$$

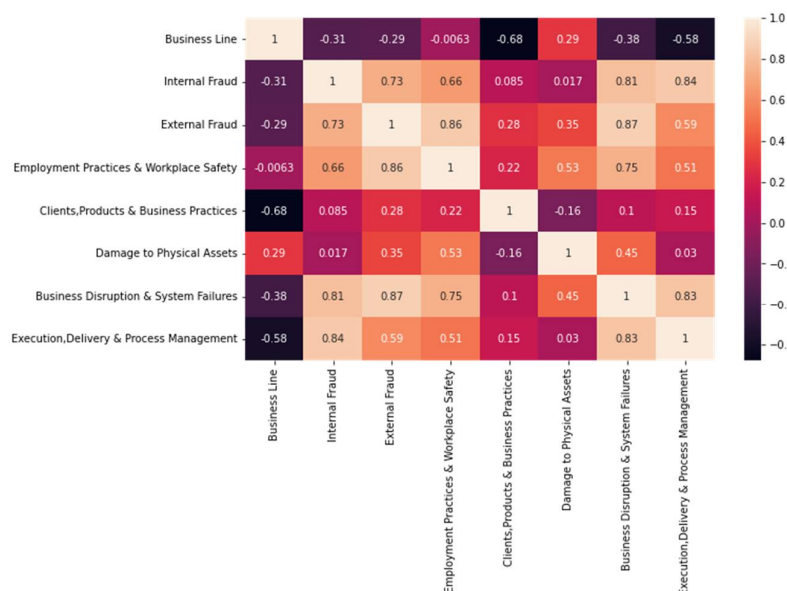
ماتریس همبستگی^۲

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

¹ Logistic regression

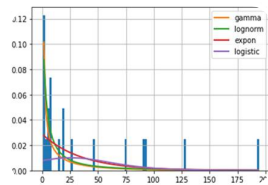
² Correlation Matrix

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].

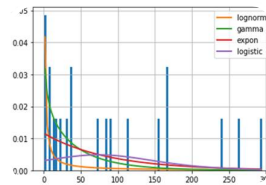


شکل ۶: ماتریس همبستگی برای داده‌های شدت

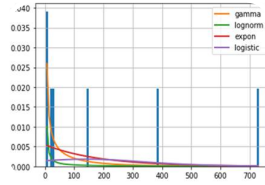
در شکل (۶) مجموعه داده‌های شدت زیان عملیاتی در قالب ماتریس همبستگی پیاده سازی شده است. لذا داده‌هایی که همبستگی حداکثری داشته باشند در یک طبقه قرار می‌گیرند. با اجرای الگوریتم درخت تصمیم و الگوریتم *knn*، ضمن استفاده از ماتریس همبستگی، خطوط کسب و کار داده‌های شدت به پنج کلاس تقسیم می‌شوند. کلاس اول شامل «پرداخت و تسویه، کارمزد خرده فروشی و غیر مرتبط»، کلاس دوم شامل «خدمات نمایندگی، بانکداری تجاری و مدیریت دارایی‌ها»، 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.



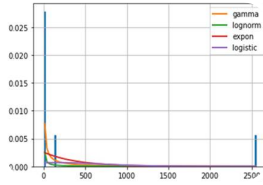
کلاس دوم



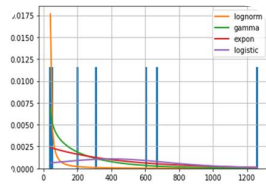
کلاس اول



کلاس چهارم



کلاس سوم



کلاس پنجم

شکل ۷. برازش داده‌های کلاس‌های طبقه‌بندی شده به توزیع‌های پیوسته

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

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 نمای تخمین زده می‌شود (در قسمت پیوست می‌توان به کد (۲) مراجعه نمود).

جدول ۸. محاسبه ارزش در معرض خطر برای کلاس‌های مشخص شده

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 داده شد 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 به همراه الگوریتم‌های یادگیری ماشین (خوشه‌بندی) استفاده شده است. برای هر گروه حوادث ضرر ساز داده‌ها به دو دسته دنباله و بدنه توزیع تفکیک شده و به منظور افزایش داده‌ها از بوت استرپ استفاده شده است و با استفاده از تئوری مقدار بحرانی برای هر گروه حوادث ضرر ساز مقدار سرمایه را تخمین زده شده است. گفتنی است که مقادیر تخمین زده شده برای سرمایه به صورت یک فاصله اطمینان ۹۵ درصد و بر حسب میلیون یورو می‌باشد. همچنین به منظور کاهش واریانس برآوردگر از روش بوت استرپ استفاده شده است.

جدول ۹. مقادیر تخمین زده شده از روش اول

مقدار سرمایه تخمین زده شده (M€)	گروه حوادث
(۷۱/۰۰۳ و ۷۶/۷۸۰)	سرقت داخلی
(۱۸/۷۱۳ و ۱۹/۳۴۰)	سرقت خارجی
(۵۶/۵۵۴ و ۵۸/۵۰۱)	استخدام و ایمنی محیط کار
(۳۰۵/۷۳۴ و ۳۱۷/۷۵۴)	مشتریان، محصولات و شیوه کسب و کار
(۴۵/۲۹۷ و ۴۶/۹۲۸)	آسیب به دارایی‌های فیزیکی
(۱۲/۰۴۳ و ۱۳/۰۲۰)	اختلال در کسب و کار و خرابی سیستم‌ها
(۱۱۵/۸۰۷ و ۱۲۱/۱۹۳)	تحویل و اجرا و مدیریت فرایندها

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

روش دوم

در این روش خطوط کسب و کار با الگوریتم‌های یادگیری ماشین (طبقه بندی) به پنج کلاس تفکیک می شود. گفتنی است که مقادیر تخمین زده شده برای سرمایه به صورت یک فاصله اطمینان ۹۵ درصد و برحسب میلیون یورو می باشد.

جدول ۱۰. مقادیر تخمین زده شده از روش دوم

نام کلاس	$M€$ (مقدار سرمایه تخمین زده)
اول	(۲۰۳۲۷۱۴/۰۳۲ و ۲۰۳۲۶۳۷/۹۹۵)
دوم	(۲۰۲/۵۷۳ و ۱۹۶/۳۹۱)
سوم	(۹۳۶/۳۶۸ و ۹۵۷/۱۸۹)
چهارم	(۱۰۸۶/۶۱۴ و ۱۱۲۱/۲۵۳)
پنجم	(۲۰/۳۹۴ و ۲۱/۱۳۹)

در جدول (۹) میزان سرمایه لازم برای مقابله با ریسک عملیاتی به تفکیک کلاس‌های طبقه بندی شده محاسبه شده است بدین منظور که از تفاضل میانگین توزیع های برازش شده و ارزش در معرض خطر این سرمایه مورد نیاز به دست می آید.

۵. بحث و نتیجه گیری

در این قسمت نتایج حاصل از روش‌های گفته شده بیان می شود و به بحث و مقایسه این روش‌ها پرداخته می شود. در اکثر پژوهش‌های پیشین از روش‌های شاخص پایه و استاندارد برای اندازه گیری ریسک عملیاتی استفاده کرده اند، پژوهش حاضر با ارائه دو مدل بدون آنکه نیاز به یادگیری داشته باشد اندازه گیری ریسک عملیاتی را در قالب یک مدل که شامل چند فرآیند است را محاسبه کرده است. به عبارت دیگر دو روش گفته شده در پژوهش حاضر تنها با گرفتن داده‌ها به عنوان ورودی مدل، عملیات لازم را انجام داده و به تفکیک گروه حوادث ضرر ساز، سرمایه لازم برای مقابله با ریسک عملیاتی را تخمین می زنند. گفتنی است در تمامی روش‌های پژوهش حاضر، از رویکرد ارزش در معرض خطر برای محاسبه سرمایه پوششی لازم استفاده شده است. 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 لحاظ می‌کند و این امر سبب می‌شود که سرمایه مورد نیاز دقت بیشتری داشته باشد. روش دوم با طبقه‌بندی داده‌های شدت زیان عملیاتی منجر می‌شود تا داده‌ها در کلاس‌هایی قرار گیرند که بیشترین همبستگی را داشته باشند و با لحاظ همبستگی بین داده‌ها سرمایه مورد نیاز برای ریسک عملیاتی را تخمین بزنند. با گسترش موسسات مالی اسلامی، استفاده از روش‌های اندازه‌گیری ریسک عملیاتی مورد توجه این موسسات قرار گرفته است. وناردوز^۱ اصطلاح بانکداری اسلامی را اصطلاحی مشابه با موافقت نامه‌های بال می‌داند که با تأسیس اولین بانک اسلامی در مصر در سال ۱۹۶۳ 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^۱ و بال^۲ از سال ۱۳۸۲ 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 الگوریتم‌های یادگیری 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 استفاده کنند و این روش را با الگوریتم‌های طبقه‌بندی ترکیب کنند.

پیوست

کد ۱

<https://colab.research.google.com/drive/19llanG16QUAYt2AC9TcNppFm2s-lqGQ?usp=sharing>

کد ۲

<https://colab.research.google.com/drive/1AUP1vHpy1TAfcAjDpSvem5tbYHekRYrL?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.

^۱ 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 and Statistics (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 Project in 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 Revised Framework - Comprehensive Version, Retrieved from <https://www.bis.org/publ/bcbs128.htm>
7. Basel Committee on banking supervision (2003). Sound Practices for The Management and Supervision of Operational risk. Retrieved from <https://www.bis.org/publ/bcbs96.htm>
8. Basel Committee on banking supervision (2001). Working Paper on 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 Finance*, 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, 17-18 May 2004, Financial Management.
12. Dhruv, D.T. (2019). Credit Risk Analysis Using Machine Learning And Neural 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. July 2007, 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). Earning Budapest University, Magyarország.
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 ریسک عملیاتی بانک‌ها با رویکرد توزیع زیان. *چشم‌انداز مدیریت مالی*، ۱۳(۴۲)، ۹-۳۴.

Citation

Akbari, Mahdi & Yazdaniyan, 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)
