ОСОБЛИВОСТІ ВИКОРИСТАННЯ АНАЛІТИЧНИХ ТЕХНОЛОГІЙ ПРИ ПРОВЕДЕННІ УПРАВЛІННЯ РИЗИКАМИ

Олександр Юрійович Стеценко
ВНЗ "Університет економіки та права "КРОК
Аспірант кафедри економіки та фінансів

У сучасному світі, який характеризується високим рівнем невизначеності та негативний вплив зовнішніх чинників, для подолання актуальних криз та забезпечення сталого розвитку підприємства вкрай важливим є впровадження ефективних ризик-орієнтованих підходів до корпоративного управління. З іншої сторони, активний розвиток технологій штучного інтелекту та аналізу великих даних відкриває для бізнесу нові можливості оптимізації операційної та управлінської діяльності за рахунок цифровізації окремих процесів та цілих галузей. Саме тому актуальним є застосування технологій штучного інтелекту та дата аналітики до такої галузі управління як ризик-менеджмент.

Перш ніж розглянути особливості використання аналітичних технологій при оцінці ризиків, слід відмітити, що ідентифікація ризиків – це ітеративний процес пошуку нових типів ризиків та профілювання їх основних характеристик для подальшої смислової інтерпретації, аналізу та обробки [1]. З боку дата-аналітики завдання ідентифікації ризиків може вирішуватися завдання пошуку аномалій в історичних масивах даних про діяльність, що відноситься до галузі застосування ризик-менеджменту. Аномальні спостереження в таких даних можуть пояснюватися в тому числі наявністю взаємозв'язків і взаємодій між об'єктами і суб'єктами діяльності, що вже призводять до настання прихованих (ще не ідентифікованих) ризикових ситуацій і відповідних наслідків, або є потенційними джерелами виникнення таких ситуацій у майбутньому.

Для виділення аномалій можуть використовуватися різні методи інтелектуального аналізу, в результаті застосування яких виявляються рідкісні об'єкти або події, що значно відрізняються від більшості об'єктів, що спостерігаються. Відокремлено дані методи мають слабкі систематичні переваги одного методу перед іншими, також їх ефективність сильно залежить від набору навчальних даних [2]. Отже, оптимальний підхід до виявлення аномалій повинен поєднувати в собі різні комбінації відомих методів.

Для формування однорідних кластерів аномалій пропонується використовувати кластерний аналіз, який реалізується за допомогою

інтелектуальних алгоритмів машинного навчання без вчителя: k-means, ієрархічний алгоритм кластеризації, Birch. Оскільки в задачі поділу аномалій на однорідні групи немає апріорно відомої кількості кластерів та розподілу за

кластерами, для вибору кращого методу кластерного аналізу та налаштування його параметрів можуть використовуватися внутрішні метрики якості кластеризації: силует, індекс Девіса-Болдуїна, індекс Калинського-Харабаша [3].

Завдання оцінки ймовірності настання ризику може розглядатися як окремий випадок класифікації і вирішуватися за допомогою моделей штучного інтелекту, що дозволяють визначати імовірність приналежності об'єкта до того чи іншого класу. Таким чином, завдання оцінки ймовірності настання ризику зводиться до завдання бінарної класифікації. Як вибірки даних для навчання класифікаційної моделі використовуються накопичені історичні масиви інформації про зафіксовані факти настання ризиків у минулому, а також про характеристики суб'єктів і об'єктів, що мають відношення до ризиків. Навчена на таких даних класифікаційна модель визначає ймовірність настання ризику певного типу в майбутньому за сукупністю характеристик суб'єктів та об'єктів, що належать до ситуації, що оцінюється. При цьому імовірність настання ризику фактично є ймовірністю приналежності об'єкта до класу А, де А - клас наявності ризику, а В - клас відсутності ризику. У випадках, коли ризики різнорідні за своєю природою та залежать від різного набору ознак, найбільш ефективним є спосіб навчання кількох бінарних класифікаторів з подальшим обчисленням інтегральної оцінки ймовірності настання ризику.

При оцінці ймовірності настання ризиків з використанням класифікаторів слід враховувати низку ключових особливостей, притаманних даного класу завдань. Перша особливість полягає в тому, що для навчальних вибірок, що використовуються в задачах прогнозування ймовірності ризику, характерна крайня незбалансованість, наявність ризику) зазвичай знаходиться в діапазоні між 100:1 та 10000:1 [3]. Цей фактор вказує на необхідність приділити особливу увагу балансуванню навчальної вибірки, тобто співвідношення об'єктів, що належать до класу В та об'єктів класу А. Для балансування навчальної вибірки може бути використаний метод «under-sampling», що ґрунтується на вибірковому виключенні з набору даних об'єктів переважаючого класу, або метод «over-sampling», що базується на генерації синтетичних об'єктів мінорних класів і їх додаванні в навчальний набір даних.

Ще однією особливістю, яку необхідно враховувати, є можливість випадкових і навмисних, обумовлених людським фактором помилок при

фіксації в історичному масиві даних фактів настання ризиків. Найбільшу проблему тут становлять ситуації, коли з метою навмисного приховання факту настання ризику (порушення) об'єкт відзначається людиною (оператором) як об'єкт без ризику. Це призводить до ситуації прихованих не ідентифікованих ризиків. В історичному масиві накопичується спотворена недостовірна інформація, що суттєво знижує якість моделей оцінки ймовірності настання ризиків, які навчаються на даному масиві. Одним з ефективних заходів щодо нівелювання даного фактора є автоматизована ідентифікація нових типів ризиків через виявлення аномалій. Аномальні об'єкти можуть

виключатися з навчальної вибірки для класифікаторів за відомими (ідентифікованими) типами ризиків або для однорідних груп (кластерів) виявлених аномалій як потенційно нових типів ризиків можуть навчатися окремі класифікатори (у другому варіанті результати ідентифікації нових типів ризиків фактично є додатковою розміткою набору даних навчання класифікаційних моделей) [4].

Для оцінки ймовірності настання ризику можна використовувати як «класичні методи» машинного навчання так і глибокі нейронні мережі. Оптимальна архітектура нейронної мережі підбирається відповідно до природних даних із навчальної вибірки: з метою оцінки ймовірності настання ризиків можуть використовуватися як пов'язані нейронні мережі так і згорткові чи рекурентні. Наступним пропонується виділити найпоширеніші і застосовні моделі класифікаторів [5], які дозволяють визначити не тільки належність об'єкта, що оцінюється, до класу ризику, але і ймовірність приналежності до цього класу: логістична регресія, найближчі сусіди, вирішальні дерева, випадковий ліс та градієнтний бустинг.

Практика показує, що ефективність перерахованих моделей класифікації залежить від специфіки конкретної задачі, яка вирішується та природи даних. Тому в більшості випадків доцільно брати декілька різних моделей з різними конфігураціями, а вибір найбільш оптимальної моделі та її гіперпараметрів робити за метрикою якості, отриманої на тестовій вибірці даних.

В підсумку слід відзначити, що методи дата аналітики, застосовні в завданнях оцінки ймовірності настання ризиків мають потенціал більш ефективного і якісного вирішення цих завдань за рахунок зниження навантаження на експертів і зниження впливу людського фактора на процес та результат оцінки ризиків.

Цитування

1. Русак О. П., Паламарчук Т. М. Ідентифікація ризиків в умовах забезпечення сталого розвитку аграрних підприємств. Науковий вісник Ужгородського національного університету. 2017. Вип. 12. С. 103–106.

2. Campos G.O., Zimek A., Sander J., Campello R., Micenkova B., Schubert E., Assent I., Houle M.E. On the evaluation of unsupervised outlier detection: measures, datasets, and an empirical study. Data Mining and Knowledge Discovery. 2016. v. 30, No. 4.

3. Leevy, J.L., Khoshgoftaar, T.M., Bauder, R.A. et al. A survey on addressing high-class imbalance in big data. J Big Data 5, 42 (2018). URL: https://doi.org/10.1186/s40537-018-0151-6 (дата звернення 31.10.2024)

4. Westerlind S. Anomaly Detection for Portfolio Risk Management. KTH Industrial Engineering and Man-agement Industrial Management. Stockholm, 2018.

5. Sarker, I.H. Machine Learning: Algorithms, Real-World Applications and Research Di-rections. SN COMPUT. SCI. 2, 160 (2021).