

Густера Олег Михайлович
*кандидат економічних наук,
асистент кафедри статистики,
інформаційно-аналітичних систем та демографії,
Київський національний університет імені Тараса Шевченка*

DOI: <https://doi.org/10.36059/978-966-397-481-1-5>

ВИКОРИСТАННЯ МЕТОДІВ МАШИННОГО НАВЧАННЯ ДЛЯ СТАТИСТИЧНОЇ ОЦІНКИ РИЗИКУ

Незважаючи на те що окремі програмно-аналітичні продукти оцінки та управління ризиком не є досить поширеними, особливо в малому та середньому бізнесі, безпосередньо завдання оцінки ризику та зниження ступеня невизначеності в тій чи іншій мірі включається в більшість систем прийнятті рішень та базується здебільшого на використанні імовірнісних моделей та статистичних даних.

Статистична оцінка ризику як один з найбільш об'єктивних та точних методів найчастіше використовується за умови відповідності вхідних даних якісним та кількісним критеріям. Перевага даного методу визначається відсутністю спотворень результатів що можуть бути викликані суб'єктивною оцінкою та людським фактором. При цьому, використання статистичного методу вимагає одночасно з відповідністю вхідних даних побудування адекватної моделі та оцінки прогностичної придатності побудованої моделі. В таких умовах використання сучасних методів машинного навчання та програмних інструментів дозволяє підвищити ефективність процесу оцінювання і точність оцінки за рахунок оптимізації рутинних операцій та використання обчислювальних можливостей сучасного апаратного забезпечення [1].

Машинне навчання може розглядатись як окремий інструмент або доповнювати статистичну оцінку ризику завдяки можливості обробки великих об'ємів даних, що надає якісно нові можливості для обробки

неструктурованих та нерозмічених даних, виявлення нових взаємозв'язків та підвищення точності розрахунків.

Так, наприклад, для оцінки ризиків у фінансово-кредитних та страхових організаціях одним з найбільш важливих факторів є наявність інформації про об'єкт оцінки – страхувальника чи отримувача кредиту. При цьому дуже часто необхідно приймати рішення при відсутності або неповноті даних про кредитну або страхову історію потенційного клієнта. На сьогоднішній день навіть у країнах з розвинутою банківською системою та максимальною взаємодією між кредитно-фінансовими організаціями більше ніж 10% населення не отримували кредити чи не користувались страховими послугами, і, відповідно, не мають жодної кредитної чи страхової історії яка б могла використовуватись для оцінки. В деяких країнах цей показник перевищує 50%, що суттєво підвищує рівень невизначеності.

В таких умовах отримання додаткових даних чи оцінок з інших джерел може суттєво підвищити якість оцінки та використовуватись при прийнятті рішень.

За умови попередньої розмітки та обробки інструменти машинного навчання дозволяють використовувати дані не лише у числовому форматі, але й в інших форматах, не призначених для автоматизованої обробки інформаційними системами – наприклад текстової або графічної [2].

При цьому, окрім того що зменшується рівень невизначеності та виникає можливість отримати нові дані, збільшується розмірність моделей за рахунок додавання нових параметрів. Як результат, це дозволяє враховувати нові фактори, які пояснюють ту частину поведінки системи що до цього визначалась як випадкові коливання. Наприклад, на платоспроможність як головний фактор ризику при видачі кредиту можуть впливати інші фактори, що напряму не формують дохід людини, такі як: вік, стать, сімейний стан, рівень освіти та інші. Використання моделей класифікації дозволяє відносити клієнтів до різних категорій – за рівнем ризику невиклати кредиту та платоспроможністю для банківської галузі чи за імовірністю страхового випадку для страхових компаній відповідно. Модель

регресії дозволяє визначити наявність, напрямок та тісноту взаємозв'язку між різними факторами, які до цього не були враховані або неможливо було отримати точні параметри моделі [3].

Перспективи використання та переваги методів машинного навчання забезпечуються, в першу чергу, наявністю якісних та розмічених даних, що можуть бути використані для побудування моделей. Велика кількість доступних даних сама по собі не дозволяє їх використовувати без попередньої підготовки та приведення до відповідної форми. При цьому якісна розмітка даних має здійснюватися із залученням фахівців у предметній області, або за підтримки експертів у конкретній галузі.

Таким чином, збільшення розмірності моделі оцінки ризику за рахунок введення нових параметрів дозволяє підвищити точність отриманих результатів та водночас використовувати статистичне оцінювання ризику для оцінки нових об'єктів або елементів сукупності, що раніше не враховувались або не підлягали оцінці. Використання сучасних методів машинного навчання за рахунок збільшення розмірності і врахування нових параметрів та опрацювання великих об'ємів даних дозволяє застосовувати моделі оцінки ризику в нових галузях.

Література:

1. Ameisen, E. Building Machine Learning Powered Applications: Going from Idea to Product: Sebastopol, CA : O'Reilly Media, 2020. 260 с.
2. Koul, A., Ganju, S., Kasam, M. Practical Deep Learning for Cloud, Mobile, and Edge: Real-World AI & Computer-Vision Projects Using Python, Keras & TensorFlow: Sebastopol, CA : O'Reilly Media, 2019. 620 с.
3. Morales, M. Grokking Deep Reinforcement Learning: Shelter Island, NY: Manning Publications, 2019. 450 с.
4. Hamdi M, Mestiri S, Arbi A. Artificial intelligence techniques for bankruptcy prediction of Tunisian companies: an application of machine learning and deep learning-based models. *Journal of Risk and Financial Management*. 2024. № 17(4).