

Примостка Л.О

доктор економічних наук, професор
Київський національний економічний університет імені Вадима Гетьмана
м. Київ, Україна

Краснова І.В.

доктор економічних наук, професор
Київський національний економічний університет імені Вадима Гетьмана
м. Київ, Україна

Кисіль Т.М.

аспірант
Київський національний економічний університет імені Вадима Гетьмана
м. Київ, Україна

ЗАСТОСУВАННЯ МЕТОДІВ МАШИННОГО НАВЧАННЯ В ІНТЕЛЕКТУАЛЬНИХ ІНФОРМАЦІЙНИХ СИСТЕМАХ БАНКІВСЬКОГО СЕКТОРУ

Банківський сектор є однією з галузей, в якій вдосконалено та запроваджено методи машинного навчання (Machine Learning, ML). Швидкий розвиток технологій та диджиталізації сучасного світу створили інноваційні можливості та виклики для банківських установ. Застосування ML в банківському секторі перетворило традиційну модель фінансових послуг та стало вагомим складовою діяльності банків. Від процесів кредитування та ризик-аналізу до особистого обслуговування клієнтів та розвитку нових фінансових ML-продуктів сприяє оптимізації процесів та розвитку інновацій.

Технології аналізу даних, які базуються на класичних статистичних методах мають низку недоліків. Такі методи як статична перевірка гіпотез, кореляції та регресії [1] ґрунтуються на використанні усереднених показників, що значно ускладнює процес аналізу даних досліджуваної області. Стандартні статистичні методи нехтують нетиповими спостереженнями (сплесками, розкидами, критичними станами), проте, саме вони, представляють інтерес в дослідженнях. Ідентифікація таких спостережень є впливовою при подальшому аналізі і може бути корисною для розуміння сутності досліджуваних об'єктів та/або явищ. Методи математичної статистики корисні при перевірці заздалегідь сформульованих гіпотез та допустимого ймовірного аналізу, що є основою оперативної аналітичної обробки даних (OLAP). Як показують сучасні дослідження, саме такі події можуть стати вирішальними щодо майбутнього поведінки та розвитку складних систем. Саме ці недоліки статистичних методів спонукали до розвитку інноваційних методів дослідження складних інтелектуальних систем, що базуються на методах машинного навчання.

Водночас постала нагальна потреба в такій технології, яка автоматично видобуває із аналітичних даних нетривіальні знання у формі моделей, залежностей, законів тощо, гарантуючи, при цьому, їх статистичну значущість. Згадані інноваційні підходи отримали назву технологій інтелектуального аналізу даних. Важливою особливістю даної технології постає нетривіальність відшукуваних

шаблонів (патернів), які відображають неочевидні, несподівані регулярності у наборі даних на основі прихованих знань.

Завдяки технології Knowledge Discovery in Databases (KDD) запроваджено аналітичний процес дослідження великих обсягів даних із залученням засобів автоматизації, що має на меті виявлення прихованих їх структур, залежностей та взаємозв'язків. KDD-технологія передбачає, що користувач попередньо осмислює задачу й подає неповні (цільові змінні) її формулювання, перетворює дані до формату придатного для їх автоматизованого аналізу та попередньої обробки. Технологія KDD дає змогу не лише підтверджувати (відхиляти) емпіричні результати, а й будувати нові, невідомі раніше моделі. Знайдена модель не зможе здебільшого претендувати на абсолютне знання, але вона надає аналітичні переваги вже завдяки факту виявлення альтернативної статистично значущої моделі. Звідси з точністю можна сказати, що KDD — це синтетична технологія, яка поєднує в собі евристичні підходи, статистичні методи, чисельні математичні методи, а також методи штучного інтелекту. До найуживаніших методів штучного інтелекту можна віднести [4]:

- Об'єднання (association) — виокремлення структур, що повторюються в часовій послідовності. Даний метод визначає правила, за якими можна встановити, що один набір елементів корелює з іншим.

- Аналіз часових рядів (sequence-based analysis) дає змогу відшукати часові закономірності між даними (транзакціями). Метод застосовується при аналізі цільових ринків, керування гнучкістю цін або циклом роботи із замовником (Customer Lifecycle Management);

- Кластеризація (clustering) — групування даних за однаковими характеристиками (близькістю даних в полях БД). Використовується для сегментації ринку та/або замовників, тощо.

- Класифікація (classification) — віднесення даних до одного із заздалегідь визначених класів, як наприклад оцінювання ризиків платоспроможності, видачею кредитів, тощо.

Часто в інтелектуальних інформаційних системах, при аналізі даних, застосовують методи: оцінювання (estimation), нечітку логіку (fuzzy logic); нейронні мережі (neural networks) [2], фрактальні перетворення (fractal-based transforms), генетичні алгоритми (genetic algorithms). До методів інтелектуального аналізу даних можна віднести візуалізацію даних [3] — побудову графічного образу даних, що допомагає у процесі загального аналізу даних вбачати аномалії, структури, тренди.

При інтелектуальному аналізі даних можна виділити два основних типи задач, які ефективно вирішуються за методами машинного навчання [5]:

1. Задачі, які базуються на наявних даних різних моделей з метою прогнозування та прийняття управлінських рішень, з врахуванням схожих ситуаційних сценаріїв.

2. Задачі, які з'ясовують властивості відповідних залежностей у наборі даних, а також їх взаємовплив. При цьому не важливо те, як інтелектуальна система передбачує прогноз, але надає повне розуміння взаємовпливу досліджуваних закономірностей (специфічних особливостей). Саме такі аспекти будуть впливати

на формування подальших ситуаційних сценаріїв прийняття управлінських рішень.

При інтелектуальному аналізі даних можна виділити основні етапи пошуку нових знань та його оцінювання, а саме:

- зведення неструктурованих даних до форми, придатної для застосування конкретних реалізацій систем KDD. Вхідна інформація, яка являє собою, в більшості, текстову інформацію повинна бути приведена до нормалізації, тобто повинен бути сформований дата-сет кількісних та/або якісних даних для подальшого аналізу;

- в процесі попередньої обробки даних необхідно розглянути набір даних за всіма можливими ознаками, що стосуються шуканої залежності, вибрати з неї ті, які явно не придатні для подальшого дослідження, та виокремити ті, що найімовірніше ввійдуть до результативної залежності;

- застосування методів KDD за різними сценаріями, що містять складні комбінації тих методів, які допомагають аналізувати дані з різних поглядів.

- верифікація та перевірка результатів здійснюється тоді, коли усі наявні дані, які мають бути проаналізовані, розділяються на навчальну та тестову вибірки, які дають змогу оцінити ефективність та точність сформованих моделей;

- на етапі інтерпретації результатів пропонуються найефективніші експертні результати, які можуть застосовувати менеджери в банківській діяльності.

Саме дотримання такого процесу при інтелектуальному аналізі даних в банківському секторі надає великий потенціал для трансформації і вдосконалення різних аспектів фінансової діяльності та обслуговування клієнтів, допомагають банкам підвищити ефективність та надійність банківської діяльності, знизити ризики та підвищити стійкість та стабільність функціонування. Звідси можна зробити висновок, що методи машинного навчання допомагають банкам приймати більш точні та обґрунтовані рішення, зменшувати ризики та покращувати взаємодію з клієнтами.

Література:

1. Гороховатський В.О., Творошенко І.С., Методи інтелектуального аналізу та оброблення даних: навч. посібник. – Харків: ХНУРЕ, 2021. – 92 с.

2. Домінова І.В., Кисіль Т. М., Прогнозування оцінки ймовірності банкрутства банківських установ на основі адаптованої нормативно-індексної моделі // Ефективна економіка. - 2021. - Вип. №1, (дата звернення: 28.01.2021)

3. Кисіль Т. М., Алгоритм функціонування інтелектуальної банківської системи // Моделювання та інформаційні системи в економіці: зб. наук. пр. – Київ: КНЕУ, 2020. – Вип. №100, – С. 93-103

4. Колодчак О.М., Інтелектуальний аналіз даних // Національний університет “Львівська політехніка”, 2021, № 54 (2), с. 49-58

5. Круковець Д.О., Можливості Data Science в центральних банках // Visnyk of the national bank of ukraine, 2020, № 249 с.14-26