

ВИКОРИСТАННЯ DATA SCIENCE ДЛЯ ОЦІНЮВАННЯ РИЗИКІВ USING DATA SCIENCE TO RISK ASSESSMENT

УДК 330.33

DOI: <https://doi.org/10.32843/infrastruct47-24>

Чубукова О.Ю.

д.е.н., професор, завідувач кафедри економічної кібернетики та маркетингу Київський національний університет технологій та дизайну

Пономаренко І.В.

к.е.н., доцент, доцент кафедри економічної кібернетики та маркетингу Київський національний університет технологій та дизайну

Домантович О.П.

магістр кафедри економічної кібернетики та маркетингу Київський національний університет технологій та дизайну

Chubukova Olha

Kyiv National University of Technologies and Design

Ponomarenko Ihor

Kyiv National University of Technologies and Design

Domantovych Oksana

Kyiv National University of Technologies and Design

У статті розглянуто специфіку розвитку сучасних підходів до машинного навчання, які є невід'ємним складником Data science. Визначено особливості розповсюдження зазначених технологій завдяки наявності відповідного програмного та комп'ютерного забезпечення. Висвітлено специфіку генерування великих обсягів даних на постійній основі, що дає змогу проводити комплексний аналіз наявної інформації та розробляти відповідні управлінські рішення. Наведено специфіку застосування методів машинного навчання у банківській сфері завдяки стабільному попити на послуги комплексної обробки великих масивів даних. Доведено доцільність комплексного дослідження особливостей ідентифікації різноманітних ризиків у банківській сфері за допомогою алгоритмів машинного навчання. Представлено приклади застосування Data science у банківській сфері: виявлення шахрайства у банківській сфері та моделювання ризиків для інвестиційних установ.

Ключові слова: Data science, Python, банківська сфера, ефективність, машинне навчання, нейронні мережі, програмне забезпечення, ризики.

В статті розглянуто специфіку розвитку сучасних підходів до машинного

обучению, которые являются неотъемлемой составляющей Data science. Определены особенности распространения указанных технологий благодаря наличию соответствующего программного и компьютерного обеспечения. Раскрыта специфика генерирования больших объемов данных на постоянной основе, что позволяет проводить комплексный анализ имеющейся информации и разрабатывать соответствующие управленческие решения. Приведена специфика применения методов машинного обучения в банковской сфере благодаря стабильному спросу на услуги комплексной обработки больших массивов данных. Доказана целесообразность комплексного исследования особенностей идентификации разнообразных рисков в банковских сфере с помощью алгоритмов машинного обучения. Представлены примеры применения Data science в банковской сфере: выявление мошенничества в банковской сфере и моделирование рисков для инвестиционных учреждений.

Ключевые слова: Data science, Python, банковская сфера, эффективность, машинное обучение, нейронные сети, программное обеспечение, риски.

The article discusses the specifics of the development of modern approaches to machine learning, which are an integral part of Data science. The features of the distribution of these technologies are determined due to the availability of appropriate software and computer support. The specifics of generating large amounts of data on an ongoing basis are disclosed, which allows for a comprehensive analysis of the available information and develop appropriate management decisions. The specificity of the application of machine learning methods in the banking sector is given due to the stable demand for services of complex processing of large data arrays. The expediency of a comprehensive study of the features of identification of various risks in the banking sector using machine learning algorithms has been proved. Examples of data science applications in banking are presented: fraud detection in the banking sector and risk modelling for investment institutions. Detection of fraud in the banking sector is associated with the need to protect the interests of both customers and specialized financial institutions. Through the use of machine learning algorithms, it is possible to address the issue of identifying suspicious transactions in a large daily flow of financial transactions. In addition, the use of various methods of classification based on a scientifically sound system of indicators makes it possible to identify individual customers by assessing the level of their solvency and the probability of return of credit resources. Risk modeling for investment institutions, which is explained by the need to protect financial resources involved in the implementation of various projects. Investment banking institutions give companies the opportunity to receive a variety of financial instruments, the value of which correlates with the specifics of the individual legal entity, the peculiarities of the project, and so on. Investment banking assesses the value of companies to create capital for corporate financing, facilitate mergers and acquisitions, conduct corporate restructuring or reorganization, as well as for investment purposes.

Key words: Data science, Python, banking sector, efficiency, machine learning, neural networks, software, risks.

Постановка проблеми. Активний розвиток інноваційних технологій у сучасних умовах призвів до комерціалізації передових досягнень, що дає змогу отримати певним компаніям конкурентні переваги порівняно з іншими суб'єктами підприємницької діяльності. Важливим ресурсом для компаній виступають дані, що генерується з різних джерел інформації [1]. Обробка отриманих даних здійснюється в рамках реалізації підходів Data science та передбачає залучення висококваліфікованих фахівців, які за допомогою методів машинного навчання намагаються виявити приховані взаємозалежності, оцінити наявність зв'язків, застосовуючи відповідні математичні моделі. Завдяки комплексному використанню методів машинного навчання можливо з високим рівнем імовірності ідентифікувати наявні ризики

та розробити відповідні управлінські рішення, які дадуть змогу мінімізувати негативний вплив деструктивних чинників на різноманітні процеси в компанії на різних рівнях [2].

Аналіз останніх досліджень і публікацій. Дослідженню питань оцінювання ризиків за допомогою методів Data science присвячено праці таких учених, як Г. Дікуонцо, Г. Галеоне, Е. Заппімбурсо, В. Делл'Атті, Д. Віксом, Дж. Шоріш та ін. Поряд із цим активний розвиток методів машинного навчання передбачає постійне дослідження окресленої наукової проблематики.

Постановка завдання. Метою дослідження є розгляд питань, пов'язаних із ризик-менеджментом, а також привернення уваги науковців та практиків до необхідності поглиблення теоретичного обґрунтування і розроблення практичних рекомендацій із

даної тематики. Вирішення поставлених завдань передбачає проведення комплексного дослідження особливостей застосування методів машинного навчання для оцінювання ризиків у цілому та у банківській сфері зокрема.

Виклад основного матеріалу дослідження. Особливості функціонування будь-якої системи передбачають наявність нескінченної кількості чинників, дію більшості з яких неможливо передбачити. Поряд із цим окремі чинники можливо ідентифікувати та за умови застосування науково обґрунтованих підходів змоделювати розвиток подій на майбутні періоди часу з високим рівнем імовірності. У процесі розроблення стратегії розвитку компанії на перспективу необхідно враховувати невизначеність, яка може призвести до негативних проявів певних явищ та процесів. У сучасному світі спостерігається динамічна трансформація економічного, політичного та соціального середовища, що призводить до генерації різноманітних ризиків для компаній [3]. Імовірність появи певних ризиків тісно пов'язана безпосередньо з конкуренцією між певними суб'єктами підприємницької діяльності. Простежується кореляція між рівнем прибутковості операцій та їх ризиковістю. Відповідно, ризик обов'язково повинен бути розрахований до максимальної припустимої межі. Управління ризиками завдяки застосуванню кількісних підходів дає можливість мінімізувати негативні ефекти в банківській, страховій, управлінській та інших сферах діяльності. Важливо, щоб аналітики, регулятори та актуарії фінансових ризиків могли кількісно збалансувати витрачені на оцінювання ризиків ресурси та отриманий у кінцевому підсумку позитивний ефект. Отже, управління ризиками – це процес прийняття та виконання управлінських рішень, спрямованих на зниження ймовірності виникнення несприятливих результатів і мінімізацію втрат, що можуть виникнути під час їх реалізації [4; 5].

Слід відзначити, що в процесі дослідження ризиків необхідно використовувати декілька джерел для отримання даних про досліджувані явища та процеси. У процесі формування баз даних необхідно знайти оптимальне співвідношення між вартістю зібраних баз даних та отриманим економічним ефектом. Сьогодні виділяють такі найменш витратні та найбільш ефективні джерела отримання інформації, що використовуються під час аналізу ризиків:

- аналіз проблем та результатів попереднього аналізу ризиків;
- консультації з кваліфікованими фахівцями, спрямовані на здійснення управління ризиками;
- результати маркетингових досліджень та SWOT-аналіз;
- результати внутрішніх та зовнішніх аудитів;
- результати аналізу реалізації бізнес-процесів;
- звіти за страховими випадками;
- досвід інших організацій та конкурентів.

Ураховуючи істотний вплив сучасної банківської системи як важливої компоненти глобального фінансового простору на функціонування національних та міжнародної економічних систем, розглянемо більш детально особливості застосування ключових алгоритмів Data science у даній сфері [6]. Передумовою застосування сучасних алгоритмів дослідження даних є генерування у банківській сфері великих обсягів інформації на постійній основі.

Серед ключових напрямів застосування Data science у банківській сфері доцільно виокремити таке:

1. Виявлення шахрайства у банківській сфері пов'язане з необхідністю захисту інтересів як клієнтів, так і спеціалізованих фінансових установ. Завдяки застосуванню алгоритмів машинного навчання можливо вирішити питання стосовно ідентифікації підозрілих операцій у великому щоденному потоці фінансових транзакцій. Поряд із цим використання різноманітних методів класифікацій на основі науково обґрунтованої системи показників дає можливість ідентифікувати окремих клієнтів, оцінивши рівень їхньої платоспроможності та ймовірності повернення отриманих кредитних ресурсів [7].

Застосування великих обсягів інформації в процесі розроблення ефективних алгоритмів ідентифікації ризиків потребує комплексних знань у прикладній сфері та вміння застосовувати відповідні методи машинного навчання. Серед методів інтелектуального аналізу необхідно звернути увагу на такі, як об'єднання, кластеризація, прогнозування та класифікація. Застосування різноманітних архітектур нейронних мереж у поєднанні з потужними обчислювальними можливостями сучасних комп'ютерів та завдяки спеціалізованим мовам програмування (Python та R) дає можливість із високим рівнем імовірності ідентифікувати випадки шахрайства у банківській сфері.

2. Моделювання ризиків для інвестиційних установ пояснюється необхідністю захисту фінансових ресурсів, які залучаються для реалізації різноманітних проєктів. Інвестиційні банківські установи дають можливість отримувати компаніям різноманітні фінансові інструменти, вартість яких корелює зі специфікою діяльності окремої юридичної особи, особливостями реалізації відповідного проєкту тощо. Інвестиційно-банківська діяльність оцінює вартість компаній для створення капіталу корпоративного фінансування, сприяння злиттю і поглинанням, проведення корпоративної реструктуризації або реорганізації, а також в інвестиційних цілях. Завдяки застосуванню алгоритмів машинного навчання можливо оцінити потенційні ризики та сформувати ефективні рішення стосовно можливості та умов надання відповідних фінансових ресурсів [8].

Розглянемо більш детально кількісні методи машинного навчання, які застосовуються у банківській

діяльності в процесі ідентифікації підозрілих операцій, що можуть виявитися шахрайськими [9].

Лінійна регресія належить до базових алгоритмів, оскільки широко розповсюджена серед науковців та практиків у зазначеній сфері впродовж тривалого історичного періоду часу. Для застосування даного підходу використовуються великі масиви даних, що характеризують досліджуване явище впродовж тривалого періоду часу. Слід відзначити, що в процесі побудови прогнозних значень дуже часто застосовують підхід Feature engineering, який передбачає формування (розрахунок) нових ознак, передусім йдеться про специфічні індекси. Залежно від характеру взаємозв'язків між факторними та результативною ознаками застосовуються різноманітні види регресійних рівнянь. Здебільшого застосовуються лінійні рівняння, у тому числі за рахунок трансформації первинних даних за відповідними ознаками, що пов'язані між собою нелінійно. У кінцевому підсумку передбачається ідентифікація ризиків та прийняття відповідних управлінських рішень.

Методи Монте-Карло набули значного розповсюдження під час проведення кількісного оцінювання фінансових операцій, у тому числі й деривативного ціноутворення. Зазначений метод характеризується такою перевагою, як застосування параметричного наближення, що дає можливість зафіксувати нестабільність. Згідно з представленим методом, використовується параметричне наближення, яке фіксує волатильність і виступає інструментом вибору для повторюваного моделювання портфелів із використанням стохастичних процесів. Методи Монте-Карло були одними з перших застосовані для фінансового моделювання на основі великих масивів комплексної інформації.

Алгоритми класифікації показали свою ефективність під час оцінювання кредитних ризиків. На противагу регресійним рівнянням зазначений підхід не дає змоги отримати кількісну прогностичну оцінку, а розподіляє досліджувану сукупність на групи, що характеризуються певним чином. Згідно з вибраною шкалою ризиковості, клієнти можуть бути віднесені до певних груп, які характеризуються різним рівнем імовірності неповернення кредитних ресурсів. У процесі створення системи класифікації за допомогою методів машинного навчання необхідно забезпечити достовірність отриманих результатів, оскільки некоректне застосування методів може призвести до хибних оцінок та некоректних управлінських рішень.

Найбільші перспективи для оцінювання ризиків у банківській сфері мають нейронні мережі, що дають змогу динамічно оцінювати комплексну статистичну інформацію. В умовах постійних змін впливу чинників зовнішнього середовища на діяльність банківської установи дуже важливо

оперативно реагувати на відповідні виклики. Використання алгоритму, який постійно навчається завдяки підлаштуванню до існуючої ситуації, дає змогу отримати високоякісні результати та уникнути суб'єктивних оцінок.

Алгоритми машинного навчання та прикладні моделі Data science допомагають фінансовим організаціям зрозуміти природу ризиків та оптимізувати власну діяльність згідно з отриманими в процесі моделювання результатами. Перевага застосування великих даних полягає у тому, що значна кількість спостережень дає можливість більш точно спрогнозувати досліджуване явище. Поряд з отриманням високоякісних прогностичних моделей доводиться враховувати значну обчислювальну складність, що виражається у необхідності використання сучасного апаратного та програмного забезпечення, а також у значних часових витратах.

Висновки з проведеного дослідження. Отже, для забезпечення високого рівня конкурентоспроможності сучасні компанії активно використовують алгоритми машинного навчання, що дають змогу досягти значних результатів завдяки існуванню потужного комп'ютерного обладнання та відповідних програмних продуктів.

Доведено, що в сучасних умовах активного розповсюдження банківських установ та надання широкого асортименту послуг різноманітним групам користувачів існує потреба в мінімізації відповідних ризиків. Досягнення поставлених завдань реалізується завдяки застосуванню комплексу алгоритмів Data science. Зростання ефективності моделювання ризиків позитивно впливає на діяльність банківської сфери на національному та глобальному рівнях.

БІБЛІОГРАФІЧНИЙ СПИСОК:

1. The Role of Data in Business. URL: <https://smallbusiness.chron.com/role-data-business-20405.html> (дата звернення: 20.09.2020).
2. Role of Data Science in Risk Management. URL: <https://financetrain.com/role-of-data-science-in-risk-management/> (дата звернення: 20.09.2020).
3. Economic Development in the 21st Century: A Review. URL: <https://developingeconomics.org/2018/12/06/economic-development-in-the-xxist-century-a-review/> (дата звернення: 20.09.2020).
4. Risk Management Process. URL: <https://www.360factors.com/blog/five-steps-of-risk-management-process/> (дата звернення: 20.09.2020).
5. What Is Risk Management? URL: <https://corporatefinanceinstitute.com/resources/knowledge/strategy/risk-management/> (дата звернення: 20.09.2020).
6. Role of Data Science in Risk Management. URL: <https://www.henryharvin.com/blog/role-of-data-science-in-risk-management/> (дата звернення: 20.09.2020).
7. Detecting Financial Fraud Using Machine Learning: Winning the War Against Imbalanced Data. URL: <https://towardsdatascience.com/detecting-financial->

fraud-using-machine-learning-three-ways-of-winning-the-war-against-imbalanced-a03f8815c9e9 (дата звернення: 20.09.2020).

8. A Beginner's Guide to Credit Risk Modelling. URL: <https://www.digitalvidya.com/blog/credit-risk-modelling/> (дата звернення: 20.09.2020).

9. How Can Data Science Improve Risk Management? URL: <https://www.toolbox.com/tech/big-data/blogs/how-can-data-science-improve-risk-management-032118/> (дата звернення: 20.09.2020).

REFERENCES:

1. The Role of Data in Business. Available at: <https://smallbusiness.chron.com/role-data-business-20405.html> (accessed 20 September 2020).

2. Role of Data Science in Risk Management. Available at: <https://financetrain.com/role-of-data-science-in-risk-management/> (accessed 20 September 2020).

3. Economic Development in the 21st Century: A Review. Available at: <https://developingeconomics.org/2018/12/06/economic-development-in-the-xxist-century-a-review/> (accessed 20 September 2020).

4. Risk Management Process. Available at: <https://www.360factors.com/blog/five-steps-of-risk-management-process/> (accessed 20 September 2020).

5. What Is Risk Management? Available at: <https://corporatefinanceinstitute.com/resources/knowledge/strategy/risk-management/> (accessed 20 September 2020).

6. Role of Data Science in Risk Management. Available at: <https://www.henryharvin.com/blog/role-of-data-science-in-risk-management/> (accessed 20 September 2020).

7. Detecting Financial Fraud Using Machine Learning: Winning the War Against Imbalanced Data. Available at: <https://towardsdatascience.com/detecting-financial-fraud-using-machine-learning-three-ways-of-winning-the-war-against-imbalanced-a03f8815c9e9> (accessed 20 September 2020).

8. A Beginner's Guide to Credit Risk Modelling. Available at: <https://www.digitalvidya.com/blog/credit-risk-modelling/> (accessed 20 September 2020).

9. How Can Data Science Improve Risk Management? Available at: <https://www.toolbox.com/tech/big-data/blogs/how-can-data-science-improve-risk-management-032118/> (accessed 20 September 2020).