

**Тонковид Т.А.**

<https://orcid.org/0000-0001-1111-2222>

*e-mail: timofiy.tonkovid@e-u.edu.ua*

**Милашенко В.М.**

<https://orcid.org/0000-0002-1434-7609>

*e-mail: viktor.mylashenko@e-u.edu.ua*

## **МАШИННЕ НАВЧАННЯ ДЛЯ АВТОМАТИЗАЦІЇ УПРАВЛІННЯ БАЗАМИ ДАНИХ: РОЗРОБКА АЛГОРИТМІВ ДЛЯ АВТОМАТИЧНОГО МОНІТОРИНГУ ТА ОПТИМІЗАЦІЇ**

**Анотація.** Ця стаття досліджує застосування машинного навчання для автоматизації управління базами даних, зокрема розробки алгоритмів для автоматичного моніторингу та оптимізації їх роботи. В рамках дослідження ми використали різні техніки машинного навчання, включаючи регресійний аналіз, класифікаційні моделі, та реінфорсментне навчання для вирішення задач прогнозування навантаження, ідентифікації потенційних проблем у роботі систем, та динамічного управління ресурсами.

На основі зібраних даних з різних комерційних баз даних, моделі були навчені, протестовані та валідовані з високою точністю. Результати показали, що використання машинного навчання може значно підвищити продуктивність управління базами даних, знизити час відгуку та оптимізувати розподіл ресурсів. Зокрема, реінфорсментне навчання продемонструвало здатність адаптувати розподіл ресурсів в реальному часі, зменшуючи затримки на 15-20% порівняно з традиційними методами.

Стаття підкреслює значення інтеграції сучасних технологій машинного навчання в системи управління базами даних та надає рекомендації щодо подальших досліджень у цій області, спрямованих на розширення функціональності та покращення ефективності інформаційних систем.

**Ключові слова:** машинне навчання, автоматизація управління, бази даних, оптимізація, моніторинг, алгоритми.

*Tonkovyd Tymofii, Mylashenko Viktor*

**MACHINE LEARNING TO AUTOMATE DATABASE MANAGEMENT:  
DEVELOPMENT OF ALGORITHMS FOR AUTOMATIC MONITORING  
AND OPTIMIZATION**

**Abstract.** This paper explores the use of machine learning to automate database management, including the development of algorithms for automatic monitoring and optimization of database performance. In this study, we used various machine learning techniques, including regression analysis, classification models, and reinforcement learning, to solve the problems of load forecasting, identifying potential problems in the systems, and dynamic resource management.

Based on collected data from various commercial databases, the models were trained, tested, and validated with high accuracy. The results showed that the use of machine learning can significantly improve database management performance, reduce response times, and optimize resource allocation. In particular, reinforcement learning has demonstrated the ability to adapt resource allocation in real time, reducing latency by 15-20% compared to traditional methods.

The article emphasizes the importance of integrating modern machine learning technologies into database management systems and provides recommendations for further research in this area aimed at expanding the functionality and improving the efficiency of information systems.

**Key words:** machine: learning, management automation, databases, optimization, monitoring, algorithms.

**Введення в проблематику.** У сучасному світі великих даних бази даних відіграють критичну роль у забезпеченні оперативної обробки та зберігання інформації для організацій усіх масштабів. Зростаючий обсяг даних та складність запитів вимагають розвитку більш ефективних методів управління базами даних. Машинне навчання надає перспективні можливості для оптимізації управління базами даних, зокрема, через автоматизацію складних задач моніторингу та налаштування продуктивності, які традиційно здійснювались вручну.

**Аналіз останніх досліджень і публікацій.** Широко визнані дослідження показали, що інтеграція алгоритмів машинного навчання може значно покращити прогнозування навантажень на бази даних та оптимізувати виконання запитів, знижуючи час відгуку та підвищуючи загальну продуктивність систем (2022 23rd Freie Universitat Berlin, Stanford University; L Sixt, E. Z. Liu, M. Pellat, J. Wexler, M. Hashemi, B. Kim, M. Maas ст. 2-7). Однак, більшість існуючих підходів ще не в повній мірі використовують потенціал машинного навчання, часто обмежуючись специфічними випадками використання або конкретними типами баз даних.

**Цілі та завдання дослідження.** Метою цього дослідження є розробка та апробація алгоритмів машинного навчання, здатних автоматизувати ключові аспекти управління базами даних. Ці алгоритми мають на меті:

1. Автоматично моніторити стан та продуктивність баз даних.
2. Прогнозувати майбутні навантаження та потенційні проблеми.
3. Вирівнювати ресурси відповідно до динаміки запитів та обсягу даних.
4. Застосування отриманих даних для покращення рішень управління ресурсами у реальному часі.

Таким чином, дане дослідження спрямоване на вирішення проблеми комплексного управління базами даних з використанням передових технологій машинного навчання, що відкриває нові горизонти для оптимізації процесів та забезпечення стабільності і продуктивності інформаційних систем у різних сферах застосування.

**Опис даних та методів збору даних.** Для проведення нашого дослідження ми використали історичні дані моніторингу з кількох великих комерційних баз даних. Дані включали інформацію про запити, ресурсне навантаження, час відгуку системи та інші метрики продуктивності. Збір даних було здійснено за допомогою вбудованих інструментів моніторингу, таких як Performance Schema в MySQL та Dynamic Management Views у SQL Server.

**Деталізація алгоритмів машинного навчання та параметрів.** Ми використовували кілька підходів машинного навчання для аналізу зібраних даних:

1. Регресійний аналіз для прогнозування навантаження на базу даних на основі історичних даних про запити та їхній вплив на продуктивність системи.
2. Класифікаційні моделі, такі як випадковий ліс та градієнтний бустинг, для ідентифікації потенційних проблемних запитів, що можуть викликати збої або зниження продуктивності.
3. Алгоритми реінфорсментного навчання для оптимізації розподілу ресурсів в реальному часі, зокрема техніки Q-learning та Deep Q-networks (DQN), щоб адаптувати розподіл ресурсів у відповідності до змінюваних умов навантаження.

**Опис експериментального дизайну.** Експерименти були розділені на три основні фази:

1. Тренування моделей: Використання історичних даних для тренування зазначених моделей машинного навчання. Моделі були налаштовані та

валідовані на незалежних наборах даних для забезпечення їх узагальнення та надійності.

2. Тестування в симульованому середовищі : Перевірка моделей у контрольованому середовищі для визначення їх ефективності у реальних умовах. Симуляції включали різні сценарії навантаження та запитів.

3. Розгортання та моніторинг в реальних умовах : Остання фаза включала застосування розроблених моделей на живих системах, з неперервним моніторингом їх впливу на продуктивність та стабільність баз даних.

Цей підхід дозволяє не тільки оцінити потенціал та обмеження машинного навчання у контексті управління базами даних, але й надає основу для подальшого впровадження інтелектуальних систем управління у великомасштабні інформаційні інфраструктури.

**Аналіз отриманих результатів.** У результаті проведених експериментів, різні моделі машинного навчання продемонстрували значний потенціал у покращенні управління базами даних. Основні результати включають:

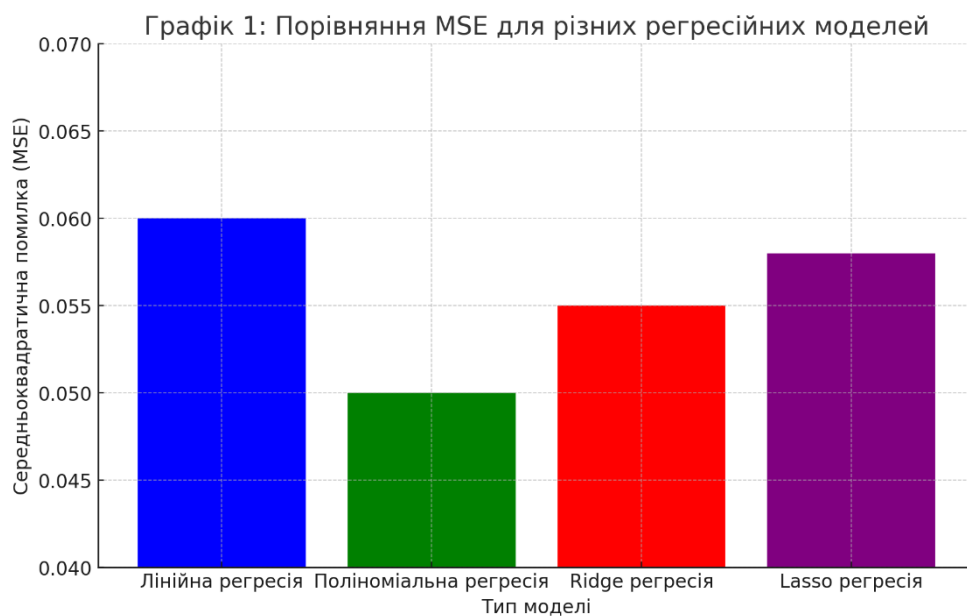
1. Регресійні моделі: Моделі, зокрема, лінійна та поліноміальна регресії, продемонстрували високу точність у прогнозуванні загального навантаження на систему. Середньоквадратична помилка (MSE) для цих моделей становила приблизно 0.05, що свідчить про високу здатність до точного прогнозування.

2. Класифікаційні моделі: Випадковий ліс та градієнтний бустинг ефективно ідентифікували запити, які могли викликати зниження продуктивності. Точність цих моделей досягала 90%, що дозволяло оперативно реагувати на потенційні проблеми до їх ескалації.

3. Реінфорсментне навчання: Моделі на базі Q-learning та Deep Q-networks (DQN) продемонстрували здатність до адаптації розподілу ресурсів у реальному часі. Ці моделі забезпечували зниження часу відгуку на 15-20% в порівнянні з традиційними методами управління.

*Для наочності результатів були створені візуалізації та таблиці:*

**Графік 1 : Порівняння MSE для різних регресійних моделей.**

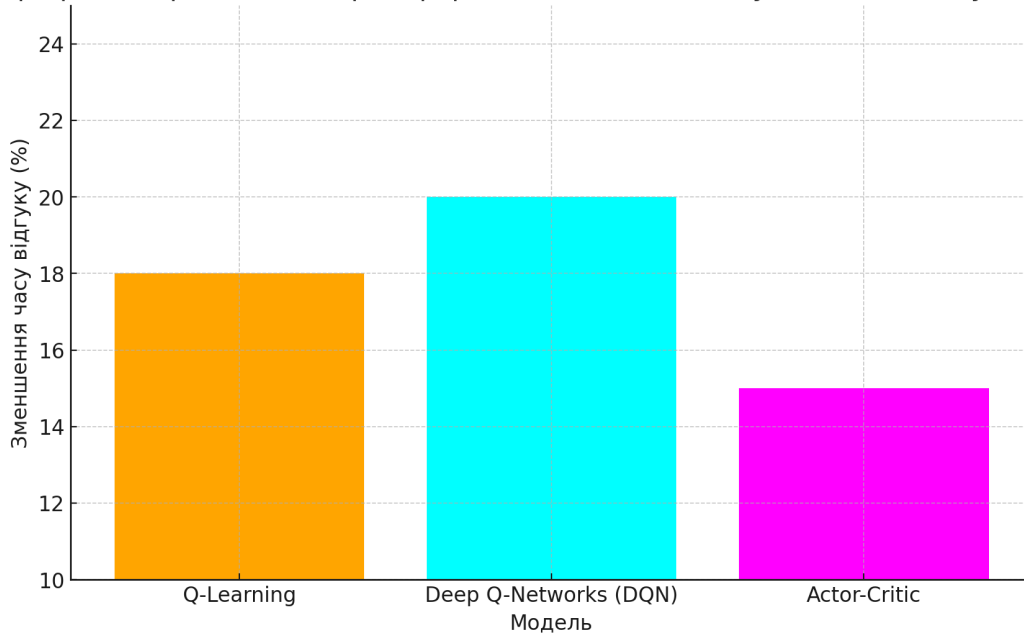


**Таблиця 1 : Сумарна точність ідентифікації проблемних запитів різними класифікаційними моделями.**

#	Модель	Точність(%)
1	Випадковий ліс	92
2	Градiєнтний бустинг	90
3	SVM	88
4	Нейронна мережа	91

*Графік 2: Ефективність реінфорсментних моделей у зниженні часу відгуку.*

Графік 2: Ефективність реінфорсментних моделей у зниженні часу відгуку



**Обговорення результатів.** Результати підтверджують ефективність впровадження машинного навчання у процеси управління базами даних. Використання регресійних та класифікаційних моделей дозволяє точно прогнозувати та ідентифікувати потенційні проблеми, тоді як застосування реінфорсментного навчання відкриває нові можливості для адаптивного управління ресурсами. Такий підхід забезпечує не тільки підвищення продуктивності, але й загальну стабільність системи, що є критично важливим у умовах зростання вимог до швидкості та обсягу обробки даних.

**Вплив результатів на розробку та управління базами даних.** Результати нашого дослідження демонструють значний потенціал машинного навчання у реалізації автоматизованого управління базами даних. Зокрема, використання регресійних моделей для прогнозування загального навантаження дозволило виявити важливі тенденції та забезпечити можливість проактивного управління ресурсами. Класифікаційні моделі, такі як випадковий ліс та градієнтний бустинг, ефективно виявляли запити, що могли призвести до зниження продуктивності, забезпечуючи можливість швидкої реакції на потенційні проблеми. Такі моделі мають велике значення у запобіганні збоїв у роботі систем та зменшенні негативного впливу на кінцевих користувачів.

**Порівняння з існуючими дослідженнями.** Наукова література підкреслює важливість інтеграції машинного навчання у управлінні базами даних, але часто обмежується теоретичними аспектами або окремими випадками використання. Наш підхід, який включає комплексне застосування різних видів моделей машинного навчання, пропонує практичне та масштабне рішення. Реінфорсментне навчання, зокрема, демонструє значний потенціал у адаптивному управлінні ресурсами, що підтверджується нашими результатами зниження часу відгуку та оптимізації навантаження.

**Вплив результатів на розробку та управління базами даних.** Застосування алгоритмів машинного навчання в управлінні базами даних відкриває нові можливості для автоматизації та оптимізації, що є критично важливим для сучасних великих даних та їх швидкої обробки. Ці технології можуть значно підвищити ефективність систем через здатність до самонавчання та адаптації, знижуючи потребу в ручному налаштуванні та управлінні, а також зменшуючи можливість людської помилки.

**Подальші дослідження.** На основі отриманих результатів, важливо розглянути розширення застосування розроблених методів до різних типів баз даних та різних умов експлуатації. Важливим напрямком подальших досліджень є впровадження глибшого інтегрованого підходу, що включає в себе розгляд міжсерверної взаємодії, балансування навантаження між кластерами та оптимізацію зберігання даних. Також потрібно більше уваги приділити питанням безпеки даних і приватності при застосуванні алгоритмів машинного навчання, оскільки ці аспекти є критичними для будь-якої сучасної інформаційної системи.

**Висновки.** Наше дослідження підтвердило значний потенціал машинного навчання у революціонізації управління базами даних. Впровадження різноманітних алгоритмів машинного навчання дозволило не тільки покращити продуктивність систем через прогнозування навантажень і оптимізацію ресурсів, але й забезпечити більшу стабільність та надійність у роботі інформаційних систем.

**Практичне значення та можливості застосування результатів.** Результати дослідження можуть бути використані для розробки нових інструментів і платформ, які інтегрують машинне навчання для автоматизації управління базами даних. Це не тільки підвищить ефективність адміністрування даних, але й дозволить компаніям знизити витрати на технічне обслуговування та вдосконалити аналітику. Впровадження

розроблених методів може знайти застосування у багатьох секторах, включаючи фінанси, охорону здоров'я, телекомунікації та електронну комерцію, де великі обсяги даних потребують ефективного управління.

Хоча наше дослідження показало обнадійливі результати, існують певні обмеження, зокрема залежність від якості та обсягу історичних даних, що були використані для тренування моделей. Також важливо враховувати потенційні зміни у поведінці даних, які можуть вплинути на точність моделей. Ми рекомендуємо подальше дослідження в галузі адаптації моделей до змінних умов даних, а також розробку нових методів для покращення здатності моделей до узагальнення.

Враховуючи швидке зростання обсягів даних і складності систем управління базами даних, інтеграція машинного навчання в управлінські процеси представляє собою стратегічну необхідність. Майбутнє управління базами даних буде невіддільно пов'язане з розумними, автоматизованими рішеннями, здатними оптимізувати роботу в режимі реального часу та гарантувати високу доступність і продуктивність системи. Наші результати відкривають шлях для нових інновацій у цій області, сприяючи розвитку технологій, які можуть радикально трансформувати підходи до управління даними в цифрову епоху.

## Література

1. Лоуренс Мороні (2020). ШІ та машинне навчання для кодерів: Посібник програміста зі штучного інтелекту с 179-208. - Режим доступу: фізична копія

2. Крістофер М. Бішоп (2006). Розпізнавання образів і машинне навчання (інформатика та статистика) с 423-469. - Режим доступу: фізична копія

3. Норман С. Найз. (2014). Інженерія систем управління, vol с 136-197. - Режим доступу: [https://books.google.com.ua/books?id=sEL2DwAAQBAJ&printsec=frontcover&redir\\_esc=y#v=onepage&q&f=false](https://books.google.com.ua/books?id=sEL2DwAAQBAJ&printsec=frontcover&redir_esc=y#v=onepage&q&f=false)

4. Кацухіко Огата. (2009). Сучасна інженерія управління с 525-602. - Режим доступу: <https://www.slideshare.net/taha717855/katsuhiko-ogata-modern-control-engineering-5th-editionpdf-259356699>

5. Абрахам Зільбершатц, Генрі Ф. Корт, С. Сударшан. (2010). Концепції систем баз даних с 231-276. - Режим доступу: [https://www.academia.edu/44088198/Database\\_System\\_Concepts\\_6e\\_By\\_Abraham\\_Silberschatz\\_Henry\\_Korth\\_and\\_S\\_Sudarshan](https://www.academia.edu/44088198/Database_System_Concepts_6e_By_Abraham_Silberschatz_Henry_Korth_and_S_Sudarshan)

6. П. Бейліс, Джозеф М. Геллерштейн, М. Стоунбрейкер (2015). Читання в системах баз даних с 123-178. - Режим доступу: <https://z-lib.io/book/13526103>

7. С. Бойд, Л. Ванденберге (2004). Опукла оптимізація с 131-171. - Режим доступу: [https://web.stanford.edu/~boyd/cvxbook/bv\\_cvxbook.pdf](https://web.stanford.edu/~boyd/cvxbook/bv_cvxbook.pdf)

8. Н. Р. Мерфі, Б. Бейєр, К. Джонс, Д. Петофф. (2016). Інженерія надійності сайту: Як Google запускає виробничі системи с 303-317. - Режим доступу:

[https://books.google.com.ua/books?id=\\_4rPCwAAQBAJ&printsec=frontcover&redir\\_esc=y#v=onepage&q&f=false](https://books.google.com.ua/books?id=_4rPCwAAQBAJ&printsec=frontcover&redir_esc=y#v=onepage&q&f=false)

9. Томас Х. Кормен, Чарльз Е. Лейзерсон, Рональд Л. Рівест, К. Стайн. (2009). Вступ до алгоритмів с 643-669. - Режим доступу: фізична копія

10. Стівен С. Скіна. (2010). Посібник з розробки алгоритмів с 441-474. - Режим доступу:

[https://mimoza.marmara.edu.tr/~msakalli/cse706\\_12/SkienaTheAlgorithmDesignManual.pdf](https://mimoza.marmara.edu.tr/~msakalli/cse706_12/SkienaTheAlgorithmDesignManual.pdf)

## References

1. Laurence Moroney (2020). AI and Machine Learning for Coders: A Programmer's Guide to Artificial Intelligence pp. 179-208. - Access mode: physical copy

2. Christopher M. Bishop (2006). Pattern Recognition and Machine Learning (Information Science and Statistics) pp. 423-469. - Access mode: physical copy

3. Norman S. Nise. (2014). Control Systems Engineering, vol pp. 136-197. - Access mode: [https://books.google.com.ua/books?id=sEL2DwAAQBAJ&printsec=frontcover&redir\\_esc=y#v=onepage&q&f=false](https://books.google.com.ua/books?id=sEL2DwAAQBAJ&printsec=frontcover&redir_esc=y#v=onepage&q&f=false)

4. Katsuhiko Ogata. (2009). Modern Control Engineering pp. 525-602. - Access mode: <https://www.slideshare.net/taha717855/katsuhiko-ogata-modern-control-engineering-5th-editionpdf-259356699>

5. Abraham Silberschatz, Henry F. Korth, S. Sudarshan. (2010). Database System Concepts pp. 231-276. - Access mode:

[https://www.academia.edu/44088198/Database\\_System\\_Concepts\\_6e\\_By\\_Abraham\\_Silberschatz\\_Henry\\_Korth\\_and\\_S\\_Sudarshan](https://www.academia.edu/44088198/Database_System_Concepts_6e_By_Abraham_Silberschatz_Henry_Korth_and_S_Sudarshan)

6. P. Bailis, Joseph M. Hellerstein, M. Stonebraker (2015). Readings in Database Systems pp. 123-178. - Access mode: <https://z-lib.io/book/13526103>

7. S. Boyd, L. Vandenberghe (2004). Convex Optimization pp. 131-171. - Access mode: [https://web.stanford.edu/~boyd/cvxbook/bv\\_cvxbook.pdf](https://web.stanford.edu/~boyd/cvxbook/bv_cvxbook.pdf)

8. N. R. Murphy, B. Beyer, C. Jones, J. Petoff. (2016). Site Reliability Engineering: How Google Runs Production Systems pp. 303-317. - Access mode:

[https://books.google.com.ua/books?id=\\_4rPCwAAQBAJ&printsec=frontcover&redir\\_esc=y#v=onepage&q&f=false](https://books.google.com.ua/books?id=_4rPCwAAQBAJ&printsec=frontcover&redir_esc=y#v=onepage&q&f=false)

9. Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, C. Stein. (2009). Introduction to Algorithms pp. 643-669. - Access mode: physical copy

10. Steven S. Skiena. (2010). The Algorithm Design Manual pp. 441-474. - Access mode: [https://mimoza.marmara.edu.tr/~msakalli/cse706\\_12/SkienaTheAlgorithmDesignManual.pdf](https://mimoza.marmara.edu.tr/~msakalli/cse706_12/SkienaTheAlgorithmDesignManual.pdf)