

DOI: <https://doi.org/10.36910/6775-2524-0560-2024-54-14>

УДК 004.912

Котлярський Алекс Олександрович, магістрант

Петрашенко Андрій Васильович, к.т.н, доцент

<https://orcid.org/0000-0003-0239-1706>

Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», м. Київ, Україна

СПОСІБ ПІДВИЩЕННЯ ЕФЕКТИВНОСТІ ВИКОРИСТАННЯ ХМАРНИХ РЕСУРСІВ

Котлярський А.О., Петрашенко А.В. Спосіб підвищення ефективності використання хмарних ресурсів. У статті досліджено сферу оптимізації хмарних ресурсів. Досліджується система, побудована на машинному навчанні, спрямована на точне прогнозування робочих навантажень і автоматичне масштабування ресурсів, що забезпечує ефективне використання хмарних служб і зниження витрат. Машинне навчання дає змогу системі адаптуватися до коливань робочого навантаження, розпізнаючи закономірності та тенденції, забезпечуючи гнучкість у використанні ресурсів і швидку адаптацію до нових умов.

Ключові слова: Хмарні ресурси, оптимізація, машинне навчання, прогнозування робочих навантажень, автоматичне масштабування, ресурси обчислювальної хмари, точність прогнозування.

Kotliarskyi A., Petrashenko A. Method of increasing the efficiency of using cloud resources. The article examines the field of optimization of cloud resources. A system built on machine learning is being investigated, aimed at accurately predicting workloads and automatically scaling resources, which ensures efficient use of cloud services and lower costs. Machine learning enables the system to adapt to fluctuations in the workload, recognizing patterns and trends, providing flexibility in the use of resources and rapid adaptation to new conditions.

Keywords: Cloud resources, optimization, machine learning, workload forecasting, automatic scaling, computing cloud resources, forecasting accuracy.

Вступ. У сучасному цифровому світі використання хмарних ресурсів стає ключовим елементом для багатьох компаній та організацій, які прагнуть досягти оптимальної ефективності та знизити витрати на обчислювальні потужності. Проте, зростання складності завдань та змінність робочих навантажень ставлять перед ними виклик ефективного управління цими ресурсами. У зв'язку з цим актуальним стає розробка способів, які б забезпечили автоматизоване та точне управління хмарними ресурсами з метою максимізації їхньої використовуваності при мінімізації витрат. У статті досліджується підхід, побудований на основі машинного навчання, який спрямований на оптимізацію використання хмарних служб шляхом прогнозування робочих навантажень та автоматичного масштабування ресурсів. Цей підхід надає системі гнучкість у адаптації до змінних умов та швидку реакцію на зміни, що забезпечує оптимальне використання хмарних ресурсів і підвищує загальну ефективність обчислювальної інфраструктури.

Постановка задачі. Мета дослідження - оптимізація використання хмарних ресурсів через автоматичне масштабування за допомогою штучного інтелекту. Задачі включають аналіз навантажень, розробку алгоритмів масштабування, їх інтеграцію з інструментами штучного інтелекту, тестування та впровадження системи.

Проблема: ефективне використання хмарних ресурсів.

Метод: автоматичне масштабування з використанням штучного інтелекту.

Мета: оптимізація використання ресурсів.

Задачі: аналіз навантажень, розробка алгоритмів масштабування, їх інтеграція з інструментами штучного інтелекту, тестування та впровадження системи.

Аналіз існуючих підходів та алгоритмів

Актуальність проблеми ефективного використання хмарних ресурсів обумовлена зростанням популярності хмарних обчислювань серед організацій та підприємств у сучасному цифровому середовищі. З метою забезпечення оптимального використання цих ресурсів і підвищення продуктивності обчислювальних процесів виникає потреба в аналізі та вдосконаленні існуючих підходів та алгоритмів [1].

Одним із ключових підходів є планування задач, що полягає у розподілі робочих завдань між різними вузлами хмарної інфраструктури з урахуванням параметрів, таких як час виконання та наявність ресурсів. Крім того, важливим елементом є розробка алгоритмів автоматичного масштабування, які забезпечують адаптацію кількості ресурсів залежно від навантаження [1].

Використання методів прогнозування для передбачення змін в навантаженні також має суттєве значення, оскільки дозволяє системі планувати та адаптуватися до майбутніх змін.

Додатково, розробка методів машинного навчання для аналізу та оптимізації роботи хмарних систем є актуальною задачею. Такі методи можуть використовуватися для розпізнавання патернів в робочих навантаженнях, автоматичного виявлення аномалій та прийняття рішень щодо оптимального розподілу ресурсів [1]. Крім того, важливим елементом є впровадження систем моніторингу та контролю за ефективністю використання ресурсів, які дозволяють виявляти та виправляти проблеми з використанням ресурсів у реальному часі.

Огляд та аналіз існуючих підходів та алгоритмів на тему підвищення ефективності використання хмарних ресурсів є необхідним для подальшого вдосконалення та оптимізації процесів хмарного обчислення.

Проблема перевантаження хмарних ресурсів. Проблема ефективності використання хмарних ресурсів виникає з необхідності оптимального розподілу і використання доступних обчислювальних ресурсів у хмарному середовищі. Незважаючи на можливості масштабування та гнучкості, хмарні обчислення часто стикаються з проблемами. У багатьох випадках, хмарні ресурси можуть бути виділені неправильно або недостатньо ефективно, що може призвести до перевитрат або недостатньої продуктивності системи [2]. В деяких сценаріях роботи, системи хмарних обчислень можуть стикатися зі значними періодичними піками навантаження, які важко передбачити та обробити, що може призвести до непрацездатності системи у критичних моментах. Нестабільність хмарних середовищ та можливість витоку конфіденційної інформації можуть становити серйозну загрозу для безпеки та конфіденційності даних користувачів [2]. Неправильне розподілення ресурсів може призвести до надмірних витрат на інфраструктуру хмарних обчислень або, навпаки, до недостатнього використання можливостей, що може позначитися на фінансовій ефективності проекту.

Використання штучного інтелекту. Один з найбільш підходящих для використання в системі оптимізації хмарних ресурсів методів машинного навчання є метод класифікації з використанням алгоритму «Random Forest» (Випадковий ліс).

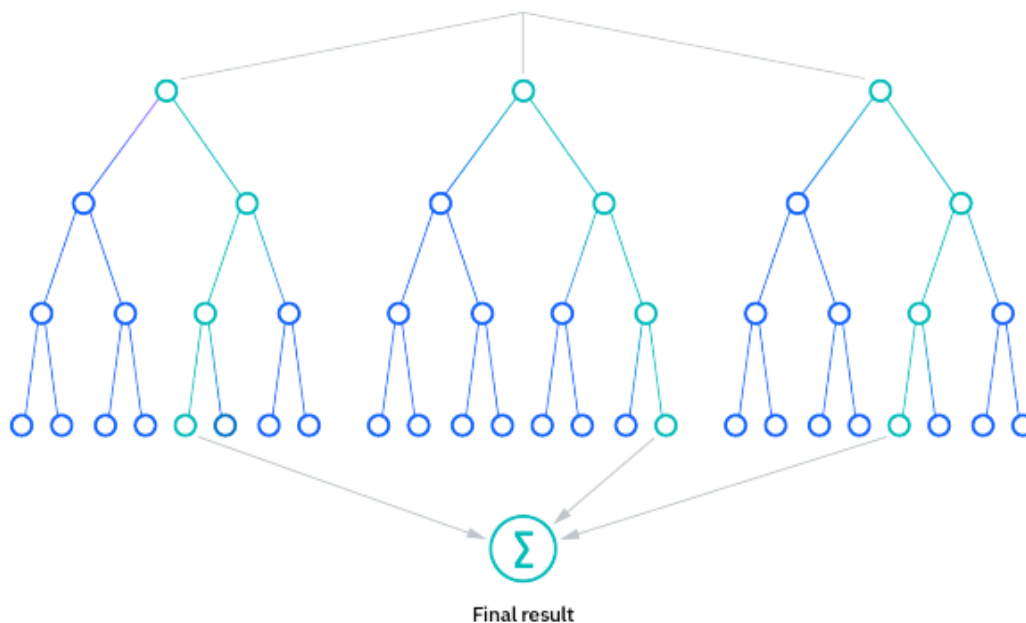


Рис.1 - Random Forest

Алгоритм «Random Forest» відноситься до класу ансамблевих методів машинного навчання, що базуються на рішеннях багатьох дерев прийняття рішень. Основна ідея полягає у створенні великої кількості різних дерев рішень та агрегації їх прогнозів для отримання більш точного та стійкого результату [3].

Переваги використання алгоритму «Random Forest» включають [3]:

- Random Forest зазвичай демонструє високу точність прогнозів завдяки агрегації результатів багатьох дерев рішень.
- Цей метод має вбудовану захист від перенавчання, оскільки він базується на випадковому виборі підмножин даних для побудови кожного дерева.
- "Random Forest" добре працює з великими обсягами даних та великою кількістю ознак, що робить його відмінним вибором для систем оптимізації хмарних ресурсів.
- Алгоритм «Random Forest» надає можливість оцінювати важливість кожної ознаки у прогнозуванні, що може бути корисним для подальшого аналізу та вдосконалення системи.

Отже, «Random Forest» є одним з найбільш підходящих методів машинного навчання для використання у системі оптимізації хмарних ресурсів, оскільки він поєднує високу точність прогнозів з ефективністю обробки великих обсягів даних та стійкістю до перенавчання.

Опис запропонованого методу

Реалізація системи оптимізації використання хмарних ресурсів шляхом використання методів машинного навчання включає послідовність ключових етапів, що ретельно розроблені для досягнення ефективності та оптимальності в процесі використання доступних обчислювальних потужностей у хмарному середовищі [4].

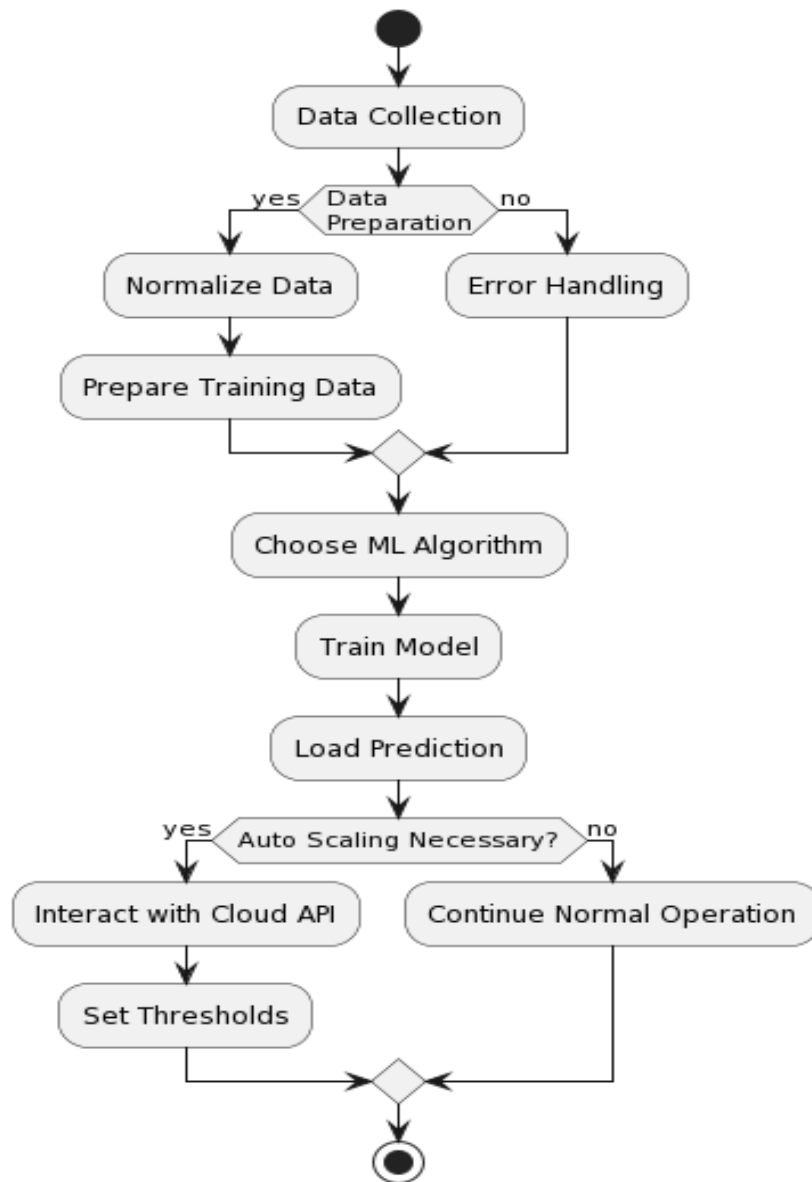


Рис.2 – Блок – схема алгоритму

Перший етап – це збір даних та їх аналіз. На цьому етапі здійснюється збір різноманітної інформації про використання ресурсів, що включає обсяги обробленої інформації, кількість активних користувачів, навантаження на сервери тощо. Крім того, проводиться збір додаткових даних про конкретні сервіси, таких як робочі години, тип подій та академічні періоди, що можуть впливати на роботу системи.

Другий етап передбачає підготовку даних, що включає створення тренувальної вибірки, що містить інформацію про навантаження та використання ресурсів, необхідну для тренування моделі машинного навчання.

Третій етап включає вибір відповідного алгоритму машинного навчання, його навчання на підготовлених даних та підключення навченої моделі до системи.

На четвертому етапі розробляються інтеграції, що дозволяють системі взаємодіяти з API хмарних платформ для автоматичного масштабування ресурсів. Визначаються порогові значення навантаження, при перевищенні яких система автоматично масштабує ресурси.

На п'ятому етапі використовуються додаткові характеристики сервісів для їх класифікації та розробки індивідуальних стратегій масштабування, враховуючи унікальні особливості кожного сервісу.

Останній етап передбачає тестування та вдосконалення системи на основі тестових даних та реальних умов експлуатації, що дозволяє підтвердити ефективність та надійність розробленої системи.

У разі потреби взаємодіє з хмарним API для налаштування ресурсів та встановлює порогові значення, інакше продовжує звичайну роботу.

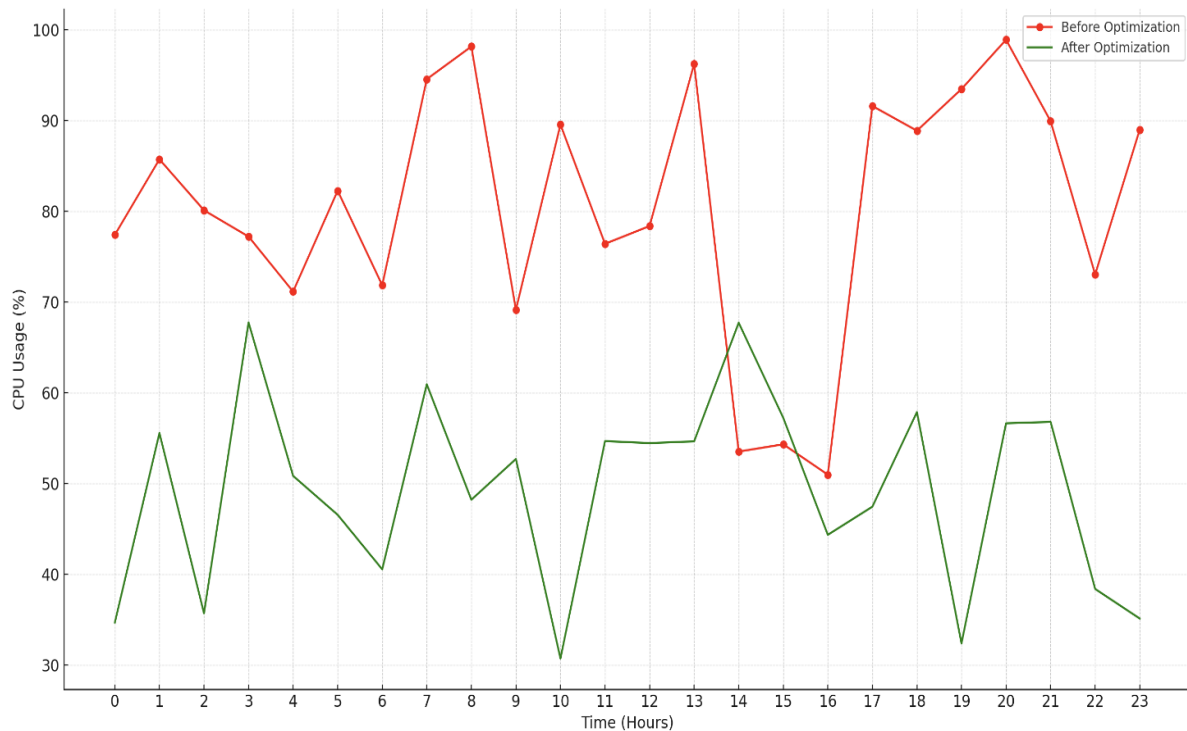


Рис.3 – Завантаження системи (використання ЦП) до та після оптимізації

До оптимізації: у системі спостерігалася більша варіативність у використанні ЦП, коли відсотки коливалися протягом дня. Це вказує на неефективне використання ресурсів, що призводить до можливого надлишку в періоди низького попиту та недостатнього в періоди пік.

Після оптимізації: впровадження алгоритму оптимізації призвело до помітного зниження варіабельності використання ЦП і загального зниження відсотка використання. Це передбачає більш ефективний розподіл ресурсів, динамічну адаптацію до робочого навантаження та потенційне зниження операційних витрат.

Таблиця 1 - Зміни у використанні CPU

Time (Hours)	Load Before Optimization (%)	Load After Optimization (%)
0	77.44	34.73
1	85.76	55.60
2	80.14	35.73
3	77.24	67.79
4	71.18	50.87

Ці дані показують, як оптимізація може вплинути на рівномірніше та ефективніше використання хмарних ресурсів, знижуючи загальні витрати та покращуючи продуктивність системи. Після оптимізації спостерігається значне зниження використання CPU у більшості випадків. Це свідчить про більш ефективне використання ресурсів, де система адаптується до поточних потреб без надмірного резервування потужностей, які насправді не використовуються.

Використання алгоритму «Random Forest» для оптимізації хмарних ресурсів показало свою ефективність. Цей метод машинного навчання забезпечив системі можливість вчасно адаптуватися до змін у використанні ресурсів, завдяки аналізу та прогнозуванню навантаження.

Висновки. Реалізація системи оптимізації використання хмарних ресурсів з використанням методів машинного навчання є перспективним та ефективним підходом для досягнення максимальної продуктивності та ефективності у хмарних обчисленнях. Послідовний аналіз та обробка даних, вибір оптимальних алгоритмів машинного навчання, а також автоматизоване масштабування ресурсів дозволяють забезпечити оптимальне використання доступних обчислювальних потужностей, зменшуючи витрати та забезпечуючи найвищий рівень сервісу для користувачів. Науково обґрунтований підхід до розробки та вдосконалення системи дозволяє не лише ефективно використовувати хмарні ресурси, а й підтримувати їх адаптацію до змінних умов, забезпечуючи стабільну та надійну роботу системи в усіх ситуаціях експлуатації.

Список бібліографічного опису

1. Cloud Optimization is About Delivering Business Efficiency. URL: <https://www.densify.com/resources/cloud-optimization/>.
2. MAAYAN G. 5 Technologies Powering Cloud Optimization. *GILAD DAVID MAAYAN*. URL: <https://devops.com/5-technologies-powering-cloud-optimization/>.
3. Maayan G. The Science of Cloud Cost Optimization. *Gilad David Maayan*. URL: <https://www.codemotion.com/magazine/devops/cloud/the-science-of-cloud-cost-optimization/>.
4. Gupta P. Machine Learning and Optimization Models for Optimization in Cloud. URL: <https://www.routledge.com/Machine-Learning-and-Optimization-Models-for-Optimization-in-Cloud/Gupta-Goyal-Chakraborty-Elngar/p/book/9781032028200>.

References

1. Cloud Optimization is About Delivering Business Efficiency URL: <https://www.densify.com/resources/cloud-optimization/>.
2. MAAYAN G. 5 Technologies Powering Cloud Optimization. *GILAD DAVID MAAYAN*. URL: <https://devops.com/5-technologies-powering-cloud-optimization/>.
3. Maayan G. The Science of Cloud Cost Optimization. *Gilad David Maayan*. URL: <https://www.codemotion.com/magazine/devops/cloud/the-science-of-cloud-cost-optimization/>.
4. Gupta P. Machine Learning and Optimization Models for Optimization in Cloud. *Elngar*. URL: <https://www.routledge.com/Machine-Learning-and-Optimization-Models-for-Optimization-in-Cloud/Gupta-Goyal-Chakraborty-Elngar/p/book/9781032028200>.