

**Кириченко Є.О.**

Чернівецький національний університет імені Юрія Федьковича

**Кириченко О.Л.**

Чернівецький національний університет імені Юрія Федьковича

## ПОБУДОВА МОДЕЛЕЙ НЕЙРОННИХ МЕРЕЖ У ХМАРНОМУ СЕРЕДОВИЩІ ЗАСОБАМИ AWS EMR

Стаття присвячена детальному аналізу можливостей використання хмарної технології AWS EMR (Amazon Web Services Elastic MapReduce) для побудови моделей нейронних мереж. У статті розкрито особливості розгортання моделей машинного навчання, а також визначено підходи до вибору відповідної інфраструктури, що безпосередньо впливає на продуктивність, масштабованість та економічну ефективність розроблених рішень. Розкрито специфіку навчання моделей нейронних мереж на великих масивах даних та здійснено порівняльний аналіз рішень для розгортання нейронних мереж на основі стохастичних моделей за допомогою хмарних сервісів Amazon EMR Serverless. З'ясовано, що нейронну мережу, розроблену на основі стохастичної моделі для передбачення флуктуацій цін на ринку акцій, можна ефективно інтегрувати у веб-додаток, який може бути легко розгорнутий в хмарному середовищі. Навчання моделі здійснювалося на даних, отриманих з Google Finance. Здійснено оцінку якості побудованої моделі та ефективності передбачення на основі середньоквадратичної помилки (RMSE). Встановлено, що вартість використання ресурсів для цієї моделі в середовищі Amazon EMR Serverless склала 0.018\$, що є економічно вигіднішим порівняно з аналогічною конфігурацією на Amazon EMR на EC2, яка коштує приблизно 0.058\$ (економія становить 31%). Зроблено висновки щодо основних переваг та недоліків Amazon EMR Serverless у застосуванні для обробки великих даних у задачах машинного навчання. Наукова новизна роботи полягає у розробці стратегій впровадження моделей нейронних мереж у хмарному середовищі за допомогою AWS EMR. Висновки свідчать про ефективність використання хмарних технологій для підвищення продуктивності та масштабованості моделей машинного навчання, а також про актуальність вибору компонентів AWS для таких рішень у майбутньому.

**Ключові слова:** хмарні обчислення, аналітичні дослідження, хмара, хмарні технології, безсерверна архітектура, AWS EMR (Amazon Web Services Elastic MapReduce), моделі машинного навчання, нейронні мережі, Байєсівська нейронна мережа, розподілені обчислення, великі дані (Big Data).

**Постановка проблеми.** В останні роки штучний інтелект (ШІ) і, зокрема, нейронні мережі стали невід'ємною частиною багатьох сфер життя та бізнесу. Вони використовуються для розпізнавання образів, обробки природної мови, передбачення трендів, аналізу поведінки клієнтів та багато іншого. Одним з ключових факторів, що сприяють швидкому розвитку цих технологій, є використання хмарних обчислень.

**Аналіз останніх досліджень і публікацій.** Хмарні обчислення дозволяють динамічно масштабувати обчислювальні ресурси відповідно до потреб. Це особливо важливо при тренуванні великих моделей нейронних мереж, які потребують значних обчислювальних потужностей. Хмарні платформи надають доступ до сучасного апаратного забезпечення, такого як GPU та TPU, що значно прискорює процес тренування моделей. Це дозволяє суттєво зменшити час, необхідний для отримання результатів [1].

У сучасному динамічному технологічному ландшафті вибір інструментів для розгортання моделей машинного навчання є важливим рішенням, яке може суттєво вплинути на успіх проєктів. Обрання між різними підходами стає все більш складним завданням, оскільки розробники та спеціалісти з обробки даних прагнуть задовольнити специфічні вимоги проєктів. Вибір відповідної інфраструктури безпосередньо впливає на продуктивність, масштабованість та економічну ефективність програм машинного навчання. Завдяки всебічному аналізу розробники можуть узгодити свій вибір із унікальними потребами своїх проєктів, що в кінцевому підсумку оптимізує процес розгортання. Це особливо актуально у сфері машинного навчання, де відмінності в інфраструктурі, інструментах і підтримці спеціалізованих алгоритмів можуть суттєво вплинути на результати [2].

Одним з головних критеріїв при виборі хмарних технологій для моделей нейронних мереж є обсяг даних з якими потрібно працювати. Великі дані вимагають значних обчислювальних ресурсів для зберігання та обробки. Традиційні системи зберігання даних часто не можуть забезпечити необхідну швидкість доступу та обробки даних. Ефективне масштабування обчислювальних ресурсів для тренування нейронних мереж є серйозним викликом, що потребує паралельного обчислення та розподілу навантаження між багатьма процесорами або графічними процесорами (GPU) [3]. Навчання моделей нейронних мереж на великих даних може займати значний час, що вимагає оптимізації процесу навчання та використання ефективних алгоритмів [4].

**Постановка завдання. Метою дослідження** є вивчення можливостей та переваг використання AWS EMR для побудови та тренування моделей нейронних мереж у хмарному середовищі.

**Об'єкт дослідження:** процес розгортання моделей нейронних мереж у хмарному середовищі.

**Предмет дослідження:** використання технології AWS EMR для розробки ефективних методів розподіленого обчислення, що дозволять значно прискорити процес навчання моделей та покращити їх продуктивність.

**Виклад основного матеріалу.** У даній статті проведемо порівняльний аналіз рішень для створення і розгортання нейронних мереж на основі стохастичних моделей за допомогою хмарних сервісів.

Розглянемо нейро-мережу на основі стохастичної моделі для вирішення задачі передбачення флуктуацій цін на ринку акцій. Архітектура розробленого додатку (рис. 1) складається з веб-серверу Flask, що обробляє запити від клієнтів, взаємодіє з кластером EMR для отримання передбачень та повертає результати клієнту; кластеру Amazon

EMR, який зберігає побудовану модель нейронної мережі, обробляє великі обсяги даних та забезпечує масштабованість; та сховища Amazon S3 для збереження побудованої моделі нейронної мережі та інших (вхідних, вихідних) даних.

Де Amazon EMR (Elastic MapReduce) – це керувана платформа для обробки великих даних, яка дозволяє швидко та економічно обробляти і аналізувати великі обсяги даних використовуючи різноманітні фреймворки з відкритим кодом, такі як Apache Spark, Apache Hive і Presto [5].

Amazon EMR призначений для зниження витрат при обробці великих обсягів даних та пропонує численні переваги, включаючи масштабованість, високу продуктивність, економічність з оплатою лише за використані ресурси, інтеграцію з іншими сервісами AWS для легкого обміну даними, простоту налаштування та управління кластером, високу безпеку з можливістю шифрування даних, автоматизацію та управління завданнями через Step Functions і Data Pipelines, а також підтримку широкого спектру аналітичних інструментів, таких як Apache Hive, Pig, HBase та Presto [6].

Ще однією опцією є можливість використання Amazon EMR Serverless, що дозволяє запускати додатки для обробки великих даних без необхідності керування кластерами (рис. 2) [7].

Сервіс допомагає уникнути надмірного або недостатнього виділення ресурсів для завдань обробки даних. Він автоматично визначає ресурси, які потрібні програмі, отримує ці ресурси для обробки завдань і звільняє ресурси, коли завдання завершуються [8].

Побудова моделі нейронної мережі на Amazon EMR Serverless починається з підготовки даних (рис. 3). Цей етап включає збір даних, необхідних для навчання нейронної мережі, та їх збереження у Amazon S3. Потім виконується попередня обробка даних, що включає очищення, нормалі-

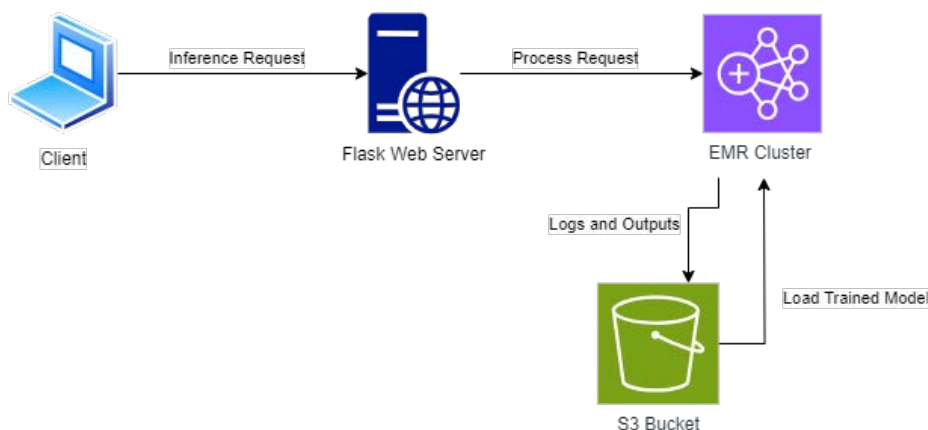


Рис. 1. Додатку для передбачення флуктуацій цін на ринку акцій

зацію та трансформацію даних для забезпечення їхньої готовності до навчання.

Наступним кроком є конфігурація Amazon S3 для завантаження оброблених даних у бакет, звідки вони будуть доступні для подальшої обробки в Amazon EMR Serverless. Після цього створюється кластер EMR Serverless, де налаштовуються необхідні фреймворки та версії програмного забезпечення.

Після налаштування кластеру розробляються Python або Scala скрипти для обробки даних та навчання моделі нейронної мережі. Ці скрипти зберігаються у бакеті S3 для подальшого використання кластером EMR. Далі виконується запуск кластеру, на якому обробляються дані та запускаються задачі.

Під час навчання моделі нейронної мережі здійснюється обчислювальний процес навчання на підготовлених даних. Після навчання моделі її зберігають у S3 для подальшого використання. Потім проводиться оцінка точності моделі на тестових даних, після чого, за необхідності, проводиться оптимізація та повторне навчання для покращення результатів.

Після завершення навчання модель розгортається для використання у реальних сценаріях [9].

Для навчання моделі використовувались дані, отримані з Google Finance [10]. Цей сервіс дозволяє отримати широкий спектр фінансової

інформації за вказаними фільтрами, включаючи в себе історичні дані про ціни акцій, об'єми торгів, фінансові звіти, та інші показники. Структура завантаженого файлу з Google Finance наведена в таблиці 1.

Таблиця 1  
Структура завантаженого файлу з Google Finance з цінами акцій

| Назва поля                           | Опис  |
|--------------------------------------|---|
| Відкриття (Open)                     | ціна акції на початку торгового дня                               |
| Максимум (High)                      | найвища ціна, за якою торгувалась акція протягом дня              |
| Мінімум (Low)                        | найнижча ціна, за якою торгувалась акція протягом дня             |
| Закриття (Close)                     | ціна акції на момент закриття торгів                              |
| Кориговане закриття (Adjusted Close) | ціна закриття, скоригована на корпоративні дії, такі як дивіденди |
| Обсяг торгів (Volume)                | загальна кількість акцій, якими торгували протягом дня.           |

Також є можливість вибору для яких саме часових інтервалів необхідно отримати дані. Для тренування нашої нейронної мережі використову-



Рис. 2. Принципи роботи Amazon EMR Serverless



Рис. 3. Процес побудови моделі нейронної мережі на Amazon EMR Serverless

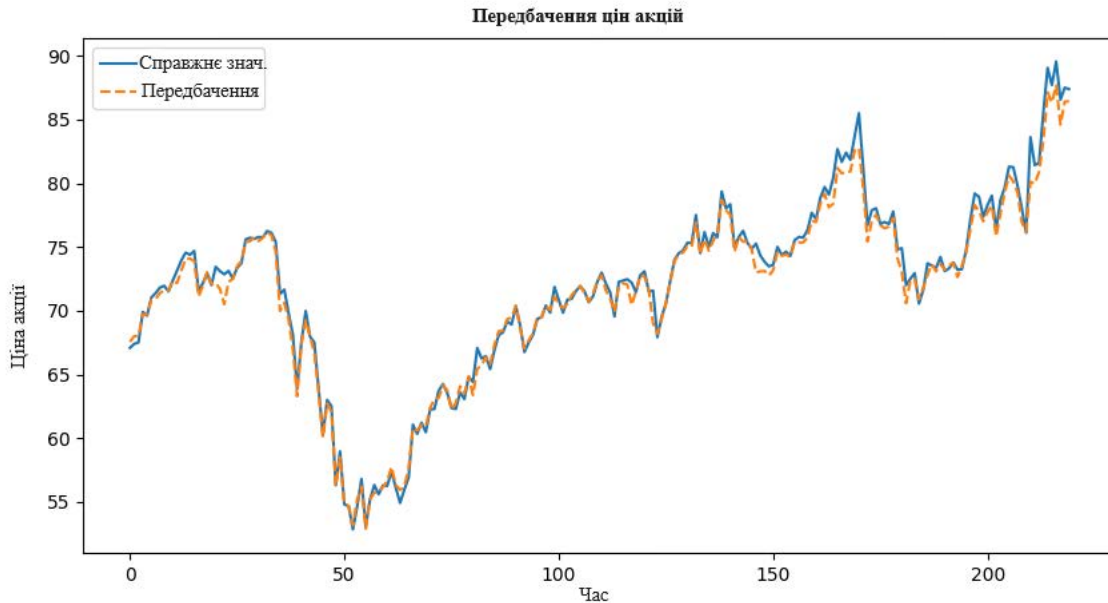


Рис. 4. Прогнозування цін акцій

валися дані за день, що дозволяє отримати більш точні прогнози та детально аналізувати короткочасові тенденції [11].

В якості моделі було обрано Байєсівську нейронну мережу, яка забезпечує імовірнісну реалізацію стандартної нейронної мережі з тією ключовою відмінністю, що ваги та зміщення представлені через апостеріорні розподіли ймовірностей, а не одноточкові оцінки.

Після імпорту даних першим кроком є попередня обробка. Це включає в собі такі завдання як обробка відсутніх значень і стандартизація функцій за допомогою методів, таких як масштабування. Наприклад, перетворення даних у формат, сумісний з алгоритмами машинного навчання, вимагає їх перетворення у Pandas DataFrames [12]. Ця операція є критичною для подальших етапів моделювання.

Для навчання моделі, що передбачатиме майбутні ціни акцій, важливо ретельно розділити набір даних. У випадку часових рядів це вимагає особливого підходу. Зазвичай дані розбиваються на навчальні та тестові підмножини, дотримуючись хронологічного порядку, що дозволяє моделі вивчати минулі тенденції та закономірності.

Для нашої моделі перші 80% даних використовуються для навчання, а решта – для тестування. Такий розподіл дозволяє ефективно оцінювати точність моделі, забезпечуючи достатню кількість даних для навчання та незалежну перевірку її ефективності [13].

На рисунку 4 показано результати роботи моделі на тестових даних.

Ефективність передбачення вимірюється середньоквадратичною помилкою (RMSE), наведеною нижче.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2},$$

де  $y_i$  – фактичне значення,  $\hat{y}_i$  – прогнозоване значення, а  $N$  – кількість точок даних для одного горизонту прогнозування (кроку). Ми використовуємо RMSE, оскільки це один із ключових показників продуктивності для прогнозування часових рядів.

Середньоквадратична помилка (RMSE) зі значенням близьким до нуля вказує на якість моделі [14]. Для побудованої моделі RMSE дорівнює 0.610548001.

Важливим аспектом при побудові моделі є вартість використаних ресурсів у хмарному середовищі. У випадку Amazon EMR Serverless користувач сплачує за кількість віртуального процесора, пам'яті та ресурсів зберігання, які споживає програмне забезпечення, причому плата нараховується за фактично використані ресурси [15]. Для побудованої моделі вартість використаних ресурсів становить 0.018\$, аналогічна конфігурація Amazon EMR на EC2 коштуватиме приблизно 0.058\$, тобто економія становить 31%.

**Висновки.** Аналіз результатів побудованої моделі свідчить про те, що Amazon EMR Serverless є ідеальним вибором при обробці великих даних у хмарному середовищі. Оскільки даних генерується та збирається все більше і більше, тому для аналізу даних потрібні високопродуктивні, масштабовані та гнучкі інструменти для надання інформації в режимі реального часу. Amazon EMR Serverless є економічно ефективним, легко конфігурується, дозволяє використовувати інструменти з відкритим кодом (Hive, Pig, Spark тощо) та централізовано зберігати на AWS S3. З іншого боку Amazon EMR Serverless створено для масової пара-



лельної обробки даних, тому не варто використовувати цей інструмент, якщо набір даних достатньо малий і легко обробляється на одній машині.

Таким чином поєднання нейронних мереж з хмарними обчисленнями, такими як AWS EMR,

відкриває нові можливості для науковців та розробників, забезпечуючи ефективність, масштабованість та доступ до потужних обчислювальних ресурсів, необхідних для реалізації складних проєктів у галузі штучного інтелекту.

#### Список літератури:

1. Aach M., Inanc E., Sarma R. et al. Large scale performance analysis of distributed deep learning frameworks for convolutional neural networks. *Journal of Big Data*, 2023. 10(1). 96(2023).
2. Rajendran P., Maloo S., Mitra R., Chanchal A., Aburukba R. Comparison of Cloud-Computing Providers for Deployment of Object-Detection Deep Learning Models. *Applied Sciences*, 2023. 13. 12577.
3. Wu X., Zhu X., Wu G.-Q., & Ding W. Data mining with big data. *IEEE Transactions on Knowledge and Data Engineering*, 2014. 26(1). P. 97-107.
4. Li D., Zhang Q., & Zheng D. Parallel computing frameworks for big data. *IEEE Cloud Computing*, 2017. 4(1). P. 24-33.
5. Amazon EMR: Easily run and scale Apache Spark, Hive, Presto, and other big data workloads. URL: <https://aws.amazon.com/emr/?nc=sn&loc=0> (дата звернення 14.10.2024)
6. Shiyani P. Big Data Analytics Adoption in Cloud (Amazon EMR). 2020. URL: [https://www.researchgate.net/publication/347506371\\_Big\\_Data\\_Analytics\\_Adoption\\_in\\_Cloud\\_Amazon\\_EMR](https://www.researchgate.net/publication/347506371_Big_Data_Analytics_Adoption_in_Cloud_Amazon_EMR) (дата звернення 15.10.2024)
7. R. A. P. Rajan. Serverless architecture—a revolution in cloud computing. *2018 Tenth International Conference on Advanced Computing (ICoAC)*, Chennai, India, 2018. Pp. 88–93. DOI: 10.1109/ICoAC44903.2018.8939081
8. What is Amazon EMR Serverless? URL: <https://docs.aws.amazon.com/emr/latest/EMR-Serverless-UserGuide/emr-serverless.html> (дата звернення 17.10.2024)
9. Nandan Umarji. Databricks to Amazon EMR: Build Serverless Enterprise Data Science Platform on AWS. 2024. URL: <https://mactores.com/blog/databricks-to-amazon-emr-build-serverless-enterprise-data-science-platform-on-aws> (дата звернення 15.10.2024)
10. Stocks API Documentation. URL: <https://polygon.io/docs/stocks/getting-started> (дата звернення 17.10.2024)
11. Dwiandiyanta B.Y., Hartanto R. and Ferdiana R. Deep Learning in Stock Market Prediction: A Five-Year Literature Review on Developments, Challenges, and Future Directions. *Journal of Theoretical and Applied Information Technology*, 2023. 101(21). DOI: 10.3844/jcssp.2022.350.358
12. Rebecca Vickery. Introduction to Pandas DataFrames: Processing, cleaning and selecting data using pandas. URL: <https://towardsdatascience.com/introduction-to-pandas-dataframes-b1b61d2cec35> (дата звернення 17.10.2024)
13. Chandra Rohitash, He Yixuan. Bayesian neural networks for stock price forecasting before and during COVID-19 pandemic. *PLOS ONE*, 2021. 16(7): e0253217. URL: <https://doi.org/10.1371/journal.pone.0253217> (дата звернення 20.10.2024)
14. Noel, D. Stock Price Prediction using Dynamic Neural Networks, 2023. URL: <https://doi.org/10.48550/arXiv.2306.12969> (дата звернення 20.10.2024)
15. Amazon EMR pricing. URL: <https://aws.amazon.com/emr/pricing/> (дата звернення 17.10.2024)

#### Kyrychenko Ye.O., Kyrychenko O.L. BUILDING NEURAL NETWORK MODELS IN A CLOUD BY MEANS OF AWS EMR

The article presents a detailed analysis of the potential of AWS EMR (Amazon Web Services Elastic MapReduce) cloud technology for building neural network models. It examines the specifics of deploying machine learning models and identifies approaches for selecting the appropriate infrastructure that directly impacts the performance, scalability, and cost-effectiveness of developed solutions. The paper explores the specifics of training neural network models on large datasets and provides a comparative analysis of deployment solutions for neural networks based on stochastic models using Amazon EMR Serverless cloud services. The study finds that a neural network developed on a stochastic model for predicting stock market price fluctuations can be effectively integrated into a web application, easily deployable in a cloud environment. The model training was conducted on data sourced from Google Finance, with model quality and prediction efficiency evaluated using the Root Mean Square Error (RMSE). It was determined that the resource usage cost for this model on Amazon EMR Serverless was \$0.018, which is more economically favorable compared to a similar configuration on Amazon EMR on EC2, which costs around \$0.058 (resulting in a 31% savings). Conclusions were drawn regarding the primary advantages and disadvantages of Amazon EMR Serverless for big data processing in machine learning tasks. The scientific novelty of this work lies in the development of strategies for deploying neural network models in the cloud using AWS EMR. The conclusions highlight the efficiency of cloud technologies for enhancing the performance and scalability of machine learning models, as well as the relevance of selecting AWS components for such solutions in the future.

**Key words:** cloud computing, analytical research, cloud, cloud technologies, serverless architecture, AWS EMR (Amazon Web Services Elastic MapReduce), machine learning models, neural networks, Bayesian neural network, distributed computing, Big Data.