

**Морозов О.С.
Яровий А.А.
Сімончук С.В.**

ЗАСТОСУВАННЯ МАШИННОГО НАВЧАННЯ ДЛЯ АНАЛІЗУ ЗМІН У ВЕБ-РЕСУРСАХ

Вінницький національний технічний університет

Анотація

Розглянуто застосування машинного навчання для автоматизації аналізу змін у веб-ресурсах, таких як оновлення контенту або редизайн. Традиційні методи, такі як скрипти чи порівняння версій, мають обмеження, зокрема при роботі з великими та складними сайтами. Пропонується використання методів машинного навчання, зокрема класифікації, кластеризації та виявлення аномалій для автоматичного виявлення змін. Описуються алгоритми для обробки текстового контенту (NLP) та зображень (CNN). Переваги машинного навчання включають критерії автоматизації, швидкості та точності.

Ключові слова: машинне навчання, моніторинг веб-ресурсів, згорткові нейронні мережі.

Abstract

Discusses the application of machine learning for automating the analysis of changes on web resources, such as content updates or redesigns. Traditional methods, such as scripts or version comparison, have limitations, particularly when dealing with large and complex websites. The use of machine learning methods, including classification, clustering, and anomaly detection, is proposed for automatically detecting changes. Algorithms for processing textual content (NLP) and images (CNN) are described. The advantages of machine learning include the criteria of automation, speed and accuracy.

Keywords: machine learning, web resource monitoring, convolutional neural networks (cnn).

Вступ

Веб-ресурси є динамічними об'єктами, які постійно зазнають змін. Це можуть бути як дрібні оновлення контенту, так і суттєві зміни, як-от редизайн сайту чи зміна його структури. Моніторинг таких змін є важливим для багатьох сфер, від забезпечення якості контенту до SEO-оптимізації та автоматизації тестування веб-додатків. Проте, традиційні методи, як скрипти або ручна перевірка, мають свої обмеження. Вони не здатні працювати з великими обсягами даних або складними веб-ресурсами [1-3].

Традиційні підходи, зокрема порівняння версій або використання регулярних виразів, мають обмежену гнучкість і точність при аналізі великих, складних веб-ресурсів. Вони також не можуть

справлятися зі складними змінами в даних чи зображеннях і не забезпечують високий рівень автоматизації процесу аналізу [2, 3].

Результати дослідження

Машинне навчання відкриває нові горизонти для автоматизації аналізу змін у веб-ресурсах. Алгоритми машинного навчання можуть обробляти великі обсяги даних, розпізнавати закономірності та автоматично класифікувати типи змін. Це дозволяє значно зменшити навантаження на аналізаторів і зробити моніторинг більш точним і ефективним [4, 5].

Машинне навчання пропонує різні підходи для вирішення цієї задачі, включаючи класифікацію, регресію, кластеризацію та виявлення аномалій. Наприклад:

- класифікація може бути використана для автоматичного визначення типу зміни (оновлення контенту, зміна структури сайту, додавання нових елементів тощо);
- кластеризація дозволяє групувати подібні зміни і визначати їх закономірності;
- виявлення аномалій може допомогти виявляти неочікувані або потенційно небажані зміни в контенті веб-ресурсів.

Моделі нейронних мереж, зокрема рекурентні нейронні мережі (RNN), ефективно працюють з послідовними змінами на веб-ресурсах, дозволяючи відстежувати еволюцію контенту і прогнозувати майбутні зміни на основі історичних даних. Методи обробки природної мови (NLP), такі як Word2Vec або BERT, дають змогу моделювати зміни в текстовому контенті, навіть коли змінюються лише окремі слова, але зміст залишається подібним.

Методи обробки природної мови (NLP): текстовий контент веб-ресурсів зазвичай змінюється через оновлення статей, новини чи оголошення. Для аналізу таких змін можуть бути використані методи обробки природної мови, такі як Word2Vec, GloVe або BERT. Ці методи дозволяють моделювати зміни в контексті, фіксуючи смислові зміни в тексті, навіть якщо словоформи залишаються подібними.

Згорткові нейронні мережі (CNN) для аналізу зображень: веб-ресурси часто містять графічний контент, який також може змінюватися (наприклад, оновлення банерів, зміни дизайну). Для автоматичного виявлення таких змін ефективними є згорткові нейронні мережі, які здатні розпізнавати відмінності між зображеннями і класифікувати їх відповідно до типу зміни.

Моделі машинного навчання дозволяють автоматизувати процес моніторингу змін, що значно зменшує час, необхідний для виявлення змін у веб-ресурсах. Вони можуть працювати в реальному часі, що забезпечує оперативний доступ до актуальної інформації.

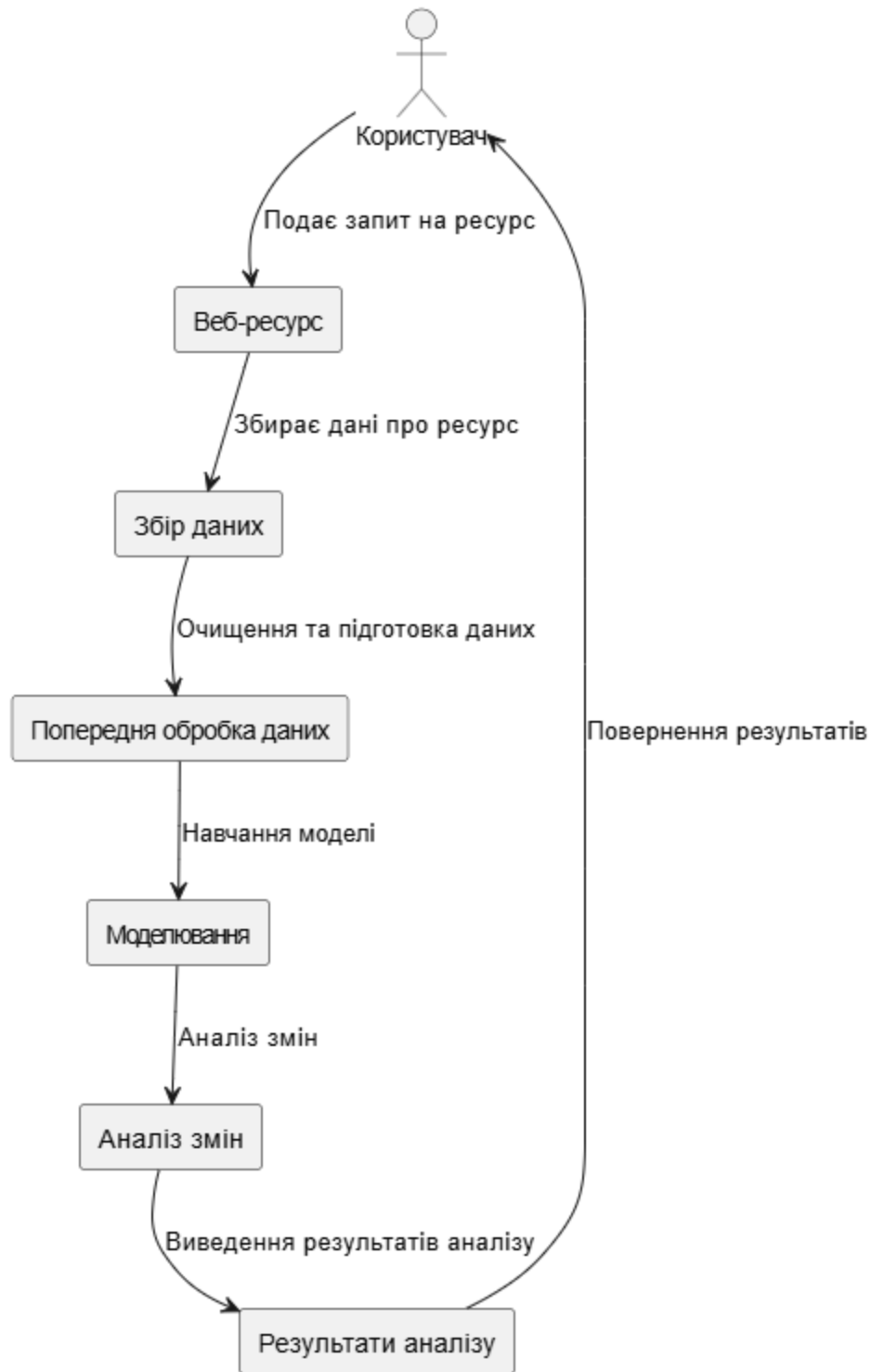


Рисунок 1 – Узагальнена схема процесу застосування машинного навчання для аналізу змін у веб-ресурсах

Моделі машинного навчання можуть адаптуватися до різних типів веб-ресурсів. Це важливо, оскільки різні категорії сайтів (новинні сайти, інтернет-магазини, блоги) можуть мати різні характеристики і потребувати специфічного підходу.

Машинне навчання можна ефективно застосувати для автоматичного моніторингу новинних сайтів, виявлення змін у статтях і класифікації типів новин (політичні, економічні, культурні). Це дозволяє не лише відслідковувати зміни, але й забезпечити точну категоризацію контенту. Окрім того, машинне навчання корисне для SEO-аналізу: зміни в метатеггах, структурах URL або наявність нових внутрішніх посилань можуть бути автоматично зафіксовані та проаналізовані для оцінки їх впливу на пошукові рейтинги сайту.

Разом із тим, варто відзначити і проблемні аспекти, які мають місце при застосуванні машинного навчання у сфері автоматизованого тестування програмного забезпечення і аналізу змін у веб-ресурсах, зокрема:

- Якість даних. Для навчання моделей машинного навчання потрібні великі набори даних, які включають різноманітні типи змін. Збір та підготовка таких даних можуть бути ресурсозатратними.
- Інтерпретація результатів. Складність і структура типу "Black-Box" деяких моделей, зокрема глибоких нейронних мереж, можуть ускладнити інтерпретацію результатів і розуміння того, як саме алгоритм прийшов до конкретного висновку.
- Проблеми з адаптацією до нових ресурсів: Моделі машинного навчання можуть потребувати перенавчання, коли вони застосовуються до нових типів веб-ресурсів або змінюються умови, у яких вони функціонують.

Висновки

Машинне навчання відкриває нові горизонти для аналізу змін у веб-ресурсах, забезпечуючи автоматизацію, підвищення точності та швидкості виявлення та класифікації змін. Ці технології штучного інтелекту можуть значно полегшити моніторинг і прогнозування динамічних змін, що робить їх важливими інструментами для сучасних веб-ресурсів, які працюють з великими обсягами даних. Проте, для ефективного застосування таких підходів необхідно вирішити проблеми з якістю даних, інтерпретацією результатів та адаптацією моделей до різноманітних типів сайтів.

Список використаної літератури

1. Технології розробки WEB-ресурсів / В. П. Молчанов, О. К. Пандорін. – Харків : ХНЕУ ім. С. Кузнеця, 2019. – 130 с.
2. Кисюк Д.В., Турко Г.А. Аналіз фреймворку REACT як інструмента для створення динамічних сайтів. : Збірник матеріалів Міжнародної науково-практичної конференції "Молодь в науці: дослідження, проблеми, перспективи (МН-2024)". – В.: ВНТУ, 2024. – С. 1-3. [Електронний ресурс] – Режим доступу: <https://conferences.vntu.edu.ua/index.php/mn/mn2024/paper/viewFile/19897/16478>
3. Миколайчук, Р., Старинський, І. і Миколайчук, В. (2024) «Аналіз технологічних аспектів реалізації веб-скрапінгу статичних і динамічних сайтів», Сучасні інформаційні технології у сфері безпеки та оборони. Київ, Україна, 51(3), с. 80–88. doi: 10.33099/2311-7249/2024-51-3-80-88.
4. Витвицький, Р., Якубовський, В. (2024). Використання штучного інтелекту та машинного навчання для автоматизації процесів тестування програмного забезпечення в Україні. Herald of Khmelnytskyi National University. Technical Sciences, 345(6(2), 21-27.
5. Онищенко, Р., Котенко, Н., Жирова, Т. (2024). Роль та ефективність засобів штучного інтелекту в тестуванні програмного забезпечення. Інформаційні технології та суспільство, (2 (13), 66-70. <https://doi.org/10.32689/maup.it.2024.2.10>

Морозов Олександр Сергійович – аспірант кафедри комп’ютерних наук, факультет інтелектуальних інформаційних технологій та автоматизації, Вінницький національний технічний університет, м. Вінниця.

Яровий Андрій Анатолійович – д.т.н., проф., завідувач кафедри комп’ютерних наук, Вінницький національний технічний університет, м. Вінниця.

Сімончук Сергій Володимирович – асистент кафедри комп’ютерних наук, Вінницький національний технічний університет, м. Вінниця.

Oleksandr S. Morozov – post-graduate student of the Department for Computer Science, Faculty of Intelligent Information Technologies and Automation, Vinnytsia National Technical University, Vinnytsia.

Andrii A. Yarovy – Dr. Sc. (Eng.), Professor, Head of the Department for Computer Science, Vinnytsia National Technical University, Vinnytsia.

Serhii V. Simonchuk – assistant of the Department for Computer Science, Vinnytsia National Technical University, Vinnytsia.