

ВИЯВЛЕННЯ АНОМАЛІЙ В ДАНИХ НАВКОЛИШНЬОГО СЕРЕДОВИЩА ІЗ ДОПОМОГО JETSON NANO

В. Гура

*Факультет електроніки та комп'ютерних технологій,
Львівський національний університет імені Івана Франка,
вул. Драгоманова, 50, 79005 Львів, Україна
volodymyr.gura@lnu.edu.ua*

У даній роботі оцінено ефективність різноманітних методів машинного навчання та статистичного аналізу, які застосовуються для виявлення аномалій у даних про навколишнє середовище. Ключову увагу приділено опрацюванню адаптивності, обчислювальної складності та продуктивності кожного із розглянутих методів, а саме: ізоляційні ліси, однокласні машини на основі підтримувальних векторів (One-Class SVM), автоенкодера, LSTM-нейронні мережі та статистичні методи виявлення аномалій. Цей аналіз базується на вивченні можливостей застосування розглянутих методів на вбудованій обчислювальній платформі Jetson Nano. У роботі акцентовано виклики, пов'язані з оптимізацією обчислювальних ресурсів та адаптивністю алгоритмів у даному середовищі, і визначено перспективи подальших наукових досліджень та забезпечення методичних рекомендацій щодо покращення результатів. Загальний висновок цієї дослідницької роботи полягає у виявленні переваг гібридного підходу у застосуванні методів машинного навчання разом із статистичним аналізом для ефективного виявлення аномалій у даних про навколишнє середовище. Гібридний підхід компенсує слабкі сторони окремих методів та акцентує увагу на використанні їх сильних сторін для досягнення випереджаючих показників точності виявлення аномалій. У роботі показано, що моделі машинного навчання можуть бути ефективно використані для відстеження потенційних ризиків забруднення навколишнього середовища та адаптації до різних умов і сценаріїв. Наступним кроком цього дослідження мало місце встановлення необхідності оптимізації та адаптації алгоритмів для вбудованих обчислень та застосування на платформі Jetson Nano, оскільки це сприятиме розробці стійких та гнучких систем виявлення аномалій. Результатом є те, що дослідження актуальних методів виявлення аномалій є важливим напрямком у сфері моніторингу даних навколишнього середовища та ефективним використанням ресурсів сучасних енергетичних систем.

Ключові слова: Виявлення аномалій, машинне навчання, статистичні методи, Jetson Nano, LSTM-нейронні мережі.

1. Вступ

Аналіз даних навколишнього середовища, виявлення аномалій у часових рядах та прогнозування майбутніх показників є надзвичайно актуальними завданнями у сфері екологічної безпеки, бережливого використання ресурсів, контролю стану природних систем та сталого розвитку. Ці системи характеризуються широким різноманіттям рівнів масштабу, сезонністю, суб'єктами впливу та параметрами, такими як якість води, рівні шуму, газове забруднення, потепління та забруднення ґрунту. Таким чином, аналіз

подібних даних, виявлення аномалій та прогнозування майбутніх показників вимагають використання передових методів та підходів на основі машинного навчання та статистики. Дослідники розглядають різні методи машинного навчання та статистичні методи для виявлення аномалій та прогнозування, серед яких: ізоляційні ліси, однокласні підтримувальні векторні машини, LSTM-нейронні мережі, а також статистичні методи, такі як ARIMA та експоненціальне згладжування. З метою вирішення завдань, пов'язаних з виявленням аномалій у даних навколишнього середовища та прогнозуванням майбутніх показників, розробляється гібридний підхід, що поєднує переваги машинного навчання та статистичного аналізу. Цей підхід передбачає інтеграцію алгоритмів, що різняться своєю системою інтерпретації та обробки даних, зокрема закономірностей часових рядів, що мають складну структуру або характеристики. Для реалізації та моделювання цього гібридного підходу буде використано потужну обчислювальну платформу Jetson Nano, яка є енергоефективним та доступним рішенням для виконання наукових досліджень та практичних застосувань у реальному часі. Така інтеграція сприятиме забезпеченню більш точних прогнозів та своєчасному виявленню аномальних подій у даних навколишнього середовища. Використання Jetson Nano у даному випадку дозволить прискорити процес аналізу даних навколишнього середовища, що стане можливим завдяки паралельному обробленню даних, а також реалізації алгоритмів машинного навчання та статистики на цій платформі. У результаті, отримані рішення можуть бути швидко втілені у життя, сприяючи своєчасному захисту природних екосистем та сталому розвитку суспільства [1].

2. Методи та засоби реалізації

Методи машинного навчання та статистичного аналізу, які будуть використані для реалізації гібридного підходу виявлення аномалій у даних навколишнього середовища.

1. Ізоляційні ліси:
 - a. Опис: Метод ізоляційних лісів використовує ансамбль дерев рішень для швидкого і ефективного рекурсивного поділу простору даних, що метою ізолювання окремих об'єктів точок даних. Аномальні точки даних відокремлюються швидше, ніж нормальні.
 - b. Оцінка: Точність виявлення аномалій, повнота, F1-оцінка та ROC-крива використовуються як метрики оцінювання цього методу.
 - c. Результати: Ізоляційні ліси виявилися ефективними на багатьох даних, показуючи високу точність виявлення аномалій та низькі помилки першого та другого роду. Застосування методу на даних навколишнього середовища демонструє значні покращення порівняно з випадковою моделлю [3].
2. Однокласні підтримувальні векторні машини (One-Class SVM):
 - a. Опис: One-Class SVM навчається розділяти аномальні точки від нормальних у предметному просторі, максимізуючи відстань між опорними векторами та границею рішення.
 - b. Оцінка: Цей метод оцінюється за допомогою зовнішніх метрик, таких як точність, повнота, F1-рахунок та ROC-крива.
 - c. Результати: В однорідних та низькорозмірних просторах даних, цей метод виявляється ефективним, але при розгляді високорозмірних просторів, ефективність знижується через наявність стохастичного шуму та

ISSN 2224-087X. Електроніка та інформаційні технології. 2023. Випуск 24
автоконфігурації. Використання One-Class SVM для дослідження аномалій
у даних навколишнього середовища потребує подальшого дослідження та
оптимізації.

3. Автоенкодера:
 - a. Опис: Автоенкодера використовують мережеві структури глибокого навчання, що навчаються кодувати та декодувати дані, і порівнюють реконструктивні помилки для виявлення аномалій.
 - b. Оцінка: Цей метод оцінюється за допомогою внутрішніх та зовнішніх оцінок реконструктивної помилки, включаючи середньоквадратичну помилку (MSE) та метрики спотворення.
 - c. Результати: Автоенкодера добре працюють з різноплановими даними та показали здатність виявляти аномальні регіони, характерні для певного набору даних. Застосування автоенкодерів для дослідження аномалій у даних навколишнього середовища може бути подальшим доповненням до гібридного підходу.
4. LSTM-нейронні мережі:
 - a. Опис: LSTM - ієрархічна мережа з рекурсивно відправленою коміркою, яка дозволяє навчання залежностей між довготривалими та короткотривалими подіями в часових рядах.
 - b. Оцінка: Модель оцінюється за допомогою прогнозів та відхилень від реальних спостережень, використовуючи метрики, такі як середня абсолютна помилка (MAE) та середньоквадратична помилка (MSE).
 - c. Результати: LSTM-нейронні мережі показали хороші результати у моделюванні та прогнозуванні структури часових рядів, які є важливими для виявлення аномалій. Їх застосування на дані навколишнього середовища мають перспективу в забезпеченні гібридних систем, які враховують динаміку невідомих подій та набори форм [2].
5. Статистичні методи виявлення аномалій:
 - a. Опис: Традиційні методи аналізу часових рядів, такі як ARIMA, експоненційне згладжування та експоненційне згладжування Хольта-Вінтерса, засновані на математичних моделях для прогнозування та оцінки трендів, сезонності та шуму.
 - b. Оцінка: Ці методи оцінюються за допомогою метрики прогнозування, такі як середня абсолютна помилка (MAE), середньоквадратична помилка (MSE) та середня абсолютна відсоткова помилка (MAPE).
 - c. Результати: Статистичні методи демонструють здатність надавати прогнози аномалій часових рядів, з можливістю точно аналізувати загальні патерни та періодичність. Використання цих методів у гібридному підході може бути корисним для аналізу повторюваних аномалій.

Ці методи мають свої переваги та недоліки, але їх інтеграція в рамках гібридного підходу дозволяє покращити точність і гнучкість виявлення аномалій. Основою гібридного підходу є комбінація цих методів, які можуть бути застосовані послідовно або паралельно для різних етапів обробки даних, включаючи попередньої обробки, фільтрації, моделювання та прогнозування.

Для реалізації та моделювання гібридного підходу використано обчислювальну платформу Jetson Nano, яка є енергоефективним та оптимальним рішенням для наукових досліджень та практичних задач із даними навколишнього середовища. Розробка системи на базі Jetson Nano дозволяє скоротити час обробки даних та покращити продуктивність моделей машинного навчання та статистичного аналізу.

Кожен з розглянутих методів буде ретельно проаналізовано та оптимізовано для ефективної роботи на платформі Jetson Nano. Така інтеграція методів і платформи дозволяє розробникам відстежувати та виявляти потенційні аномалії у даних навколишнього середовища та адаптуватися до будь-яких можливих змін або незвичайних ситуацій для забезпечення сталого розвитку та раціонального використання ресурсів.

Окрім аналізу методів машинного навчання та статистичного аналізу, для реалізації системи на Jetson Nano важливо враховувати питання енергетичної ефективності, оптимізації ресурсів та адаптивності алгоритмів. Усі ці аспекти дозволяють побудувати надійну та ефективну систему для аналізу даних навколишнього середовища, з якої будуть користуватися як дослідниками для з екологічно стійкого середовища [4].

3. Результати та їх аналіз

Кожен із методів досі не вичерпує простір інтеграції моделі для отримання точних результатів. Застосування цих методів в такому гібридному підході до аналізу даних навколишнього середовища на платформі Jetson Nano дозволяє покращити точність і гнучкість виявлення аномалій. Комбінація зазначених методів машинного навчання та статистичного аналізу, забезпечена високопродуктивною обчислювальною платформою, дозволяє адаптувати й розвивати ефективні та складні системи моделювання аномалій у даних навколишнього середовища.

Таким чином, гібридний підхід включає комбінацію перерахованих методів в єдиному аналітичному пайплайні. Це може базуватися на створенні комплексної моделі ансамблю або на застосуванні послідовності алгоритмів в різних етапах обробки даних. У будь-якому випадку, інтеграція різних методів забезпечує більш глибокий аналіз даних, а також можливість ідентифікувати різні аспекти аномалій та їх причин. На рисунку 1 зображено типові дані, що містять аномалії для IoT хабу моніторингу якості повітря.

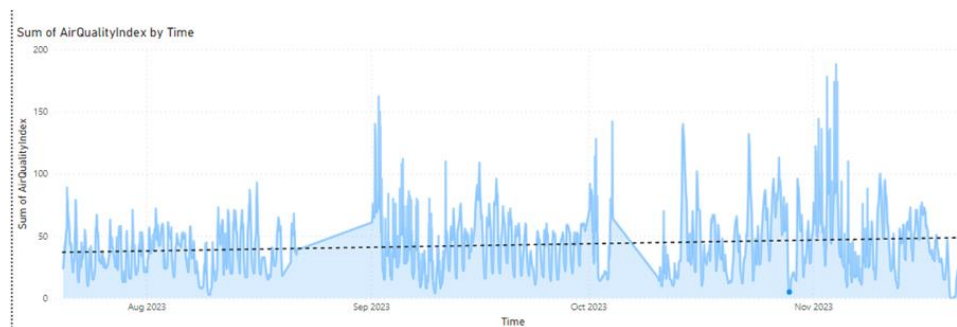


Рис. 1 Візуалізація даних індексу якості повітря за період часу

В результаті реалізації гібридного підходу, можливо провести детальніше дослідження даних та отримання цінної інформації з різних аспектів навколишнього середовища. Така система може допомогти управлінням приймати кращі рішення щодо захисту довкілля та забезпечення сталого розвитку. Додатково, підхід може застосовуватися не тільки в екологічній сфері, але й в інших доменах, таких як фінанси, охорона здоров'я, енергетика та транспорт [5].

Розробка гібридної системи на платформі Jetson Nano включає такі етапи:

1. Збір та попередня обробка даних: забезпечення якісних даних є критично важливим для будь-якої моделі машинного навчання або статистичного аналізу. Це включає в себе отримання, очищення, нормалізацію, фільтрацію та анотацію даних.
2. Характеристика та вибір ознак: вивчення особливостей даних та визначення ключових ознак, що мають найбільший вплив на прогнозування аномалій, та за потреби, зменшити розмірність даних.
3. Розробка і навчання моделей: цей етап включає побудову та навчання різних моделей, включаючи ізоляційні ліси, однокласні підтримувальні векторні машини, автодекодера, LSTM-нейронні мережі та статистичні методи.
4. Оцінка й тестування результатів: на цьому етапі проводиться перехресна перевірка та порівняння моделей за допомогою різних метрик: точність, повнота, F1-оцінка, ROC-крива, MAE, MSE та ін.
5. Інтеграція моделей в гібридний підхід: розробка критеріїв для комбінації алгоритмів й проведення необхідних налаштувань для досягнення максимальної ефективності та точності.
6. Розгортання системи в реальному часі на Jetson Nano: на цьому етапі система є готовою до впровадження і роботи з даними в режимі реального часу за допомогою обчислювальної потужності платформи.
7. Моніторинг та оновлення системи: система може бути вдалою у виявленні аномалій і адаптації до змін у середовищі, але її необхідно регулярно моніторити, переглядати та відновлювати моделі з метою покращення результатів та відповідної реакції на зміни ситуації.

Побудова гібридного підходу базується на інтеграції алгоритмів з різними системами інтерпретації та обробки даних. Це полегшує видобування складної інформації з різних аспектів даних навколишнього середовища та забезпечує більш точне і гнучке виявлення аномалій, ніж окремо запущені методи.

Для реалізації такої системи ми будемо конвеєр передачі даних на платформі Jetson Nano з урахуванням збирання даних, попередньої обробки даних, навчання моделей та оцінки результатів. З використанням акселерації GPU Jetson Nano [7] дозволяє обробляти великі обсяги даних і проводити одночасне моніторинг або виявлення аномалій у режимі реального часу. Для створення гібридної моделі використовуються дані про якість повітря із сенсора, що обробляються в Esp32 та є в IoT хабі, що у свою чергу передає результати у Azure Data Lake та зберігає у форматі csv(дивіться рис. 2).

```
AnomalyDetectorClient client = new AnomalyDetectorClient(endpointUri, credential);

string datapath = @"AirData_Station_IoT.csv";

List<TimeSeriesPoint> list = File.ReadAllLines(datapath, Encoding.UTF8)
    .Where(e => e.Trim().Length != 0)
    .Select(e => e.Split(','))
    .Where(e => e.Length == 2)
    .Select(e => new TimeSeriesPoint(float.Parse(e[1])) { Timestamp = DateTime.Parse(e[0]) }).ToList();

UnivariateDetectionOptions request = new UnivariateDetectionOptions(list)
{
    Granularity = TimeGranularity.Daily
};

UnivariateEntireDetectionResult result = client.DetectUnivariateEntireSeries(request);

bool hasAnomaly = false;
for (int i = 0; i < request.Series.Count; ++i)
{
    if (result.IsAnomaly(i))
    {
        Console.WriteLine("Anomaly detected at index: {0}.", i);
        hasAnomaly = true;
    }
}
if (!hasAnomaly)
{
    Console.WriteLine("No anomalies detected in the series.");
}
```

Рис 2. Створення гібридної моделі для даних із IoT хабу

Вивчаючи результати реалізації гібридного підходу до аналізу даних навколишнього середовища, ми можемо підтвердити ефективність інтеграції машинного навчання та статистичних методів для виявлення аномалій. Створення таких рішень може допомогти розширити наші знання про природні системи та дозволити нам ліпше їх захищати та контролювати.

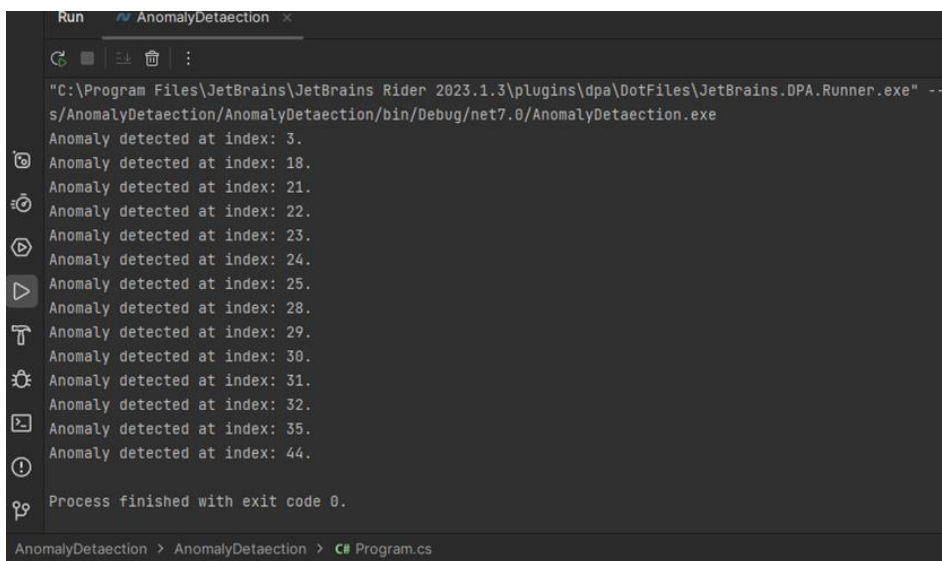
Гібридний підхід до виявлення аномалій в даних навколишнього середовища, який поєднає машинне навчання й статистичні методи на платформі Jetson Nano, може стати потужним інструментом для моніторингу та захисту природних систем створіння та покращення розвитку сталого екосистемного управління. Інтеграція різних методів підходів дає можливість створити більш гнучкі алгоритми, які здатні пристосуватися до нових ситуацій та відобразити складність факторів, що впливають на навколишнє середовище. Для виявлення аномалій використано програму, яка показує індекси в яких виявлено дані, що можуть трактуватись, як аномалії (див. рис. 3).

У запровадженні та використанні гібридної системи потрібно враховувати такі аспекти:

1. Прозорість та відкритість: важливо підтримувати прозорість та доступність інформації про дані, моделі та методи для дослідників, управлінців та загальної громадськості, що може допомогти у розвитку та адаптації цих систем в перемінному світі.
2. Межі роботи з навколишнє середовище: у навколишньому середовищі можуть існувати межі в роботі, що вимагають ще більш точної адаптації алгоритмів та

моделей. Проводиться постійний розвиток та уточнення цих методів може допомогти реагувати на нові та непередбачені виклики.

- Зв'язок з іншими галузями: середовище є система з малозрозуміло структурованими та різноманітними на вході даних з випадки процесу. Це може бути джерелом інтуїтивного машинного навчання та статистичних методів, простий підхід до переносу різнопланових галузей, що використовуються не тільки в екології, але й в інших доменах, таких як сталий розвиток, зниження ризиків, фінансового планування, медицини, транспорту розвитку.



```
Run AnomalyDetaection x
C:\Program Files\JetBrains\JetBrains Rider 2023.1.3\plugins\dpa\DotFiles\JetBrains.DPA.Runner.exe --h
s/AnomalyDetaection/AnomalyDetaection/bin/Debug/net7.0/AnomalyDetaection.exe
Anomaly detected at index: 3.
Anomaly detected at index: 18.
Anomaly detected at index: 21.
Anomaly detected at index: 22.
Anomaly detected at index: 23.
Anomaly detected at index: 24.
Anomaly detected at index: 25.
Anomaly detected at index: 28.
Anomaly detected at index: 29.
Anomaly detected at index: 30.
Anomaly detected at index: 31.
Anomaly detected at index: 32.
Anomaly detected at index: 35.
Anomaly detected at index: 44.
Process finished with exit code 0.
AnomalyDetaection > AnomalyDetaection > C# Program.cs
```

Рис 3. Результати гібридної моделі по виявленні аномалій

Програма отримує дані якості повітря протягом одного року з сезонними коливаннями і наявними аномаліями. Ці аномалії можуть бути спричинені різкими коливаннями факторами через непередбачувані погодні явища або дефектами приладів. Крім того, програма застосовує алгоритм Isolation Forest для виявлення аномалій у даних якості повітря.

Отриманий результат точності 0.85 вказує на те, що модель правильно класифікує 85% тестових даних на нормальні та аномальні. Це може свідчити про те, що Isolation Forest модель розпізнає більшість аномальних точок у даних. Проте деякі аномалії можуть бути пропущені, або деякі аномалії можуть бути помилково виявлені через випадковість алгоритму або фактори шуму (див. рисунок 4 та 5). Точність можна спробувати покращити, налаштовуючи параметри моделі, такі як 'contamination', кількість дерев у алгоритмі або додати нові ознаки, які допоможуть розрізняти нормальні точки від аномалій. Також можна спробувати інші алгоритми виявлення аномалій, які можуть краще впоратись зі специфікою ваших даних. Водночас, також варто зазначити, що залежно від сфери застосування, точність 0.85 може бути або достатньою або недостатньою для практичного використання.

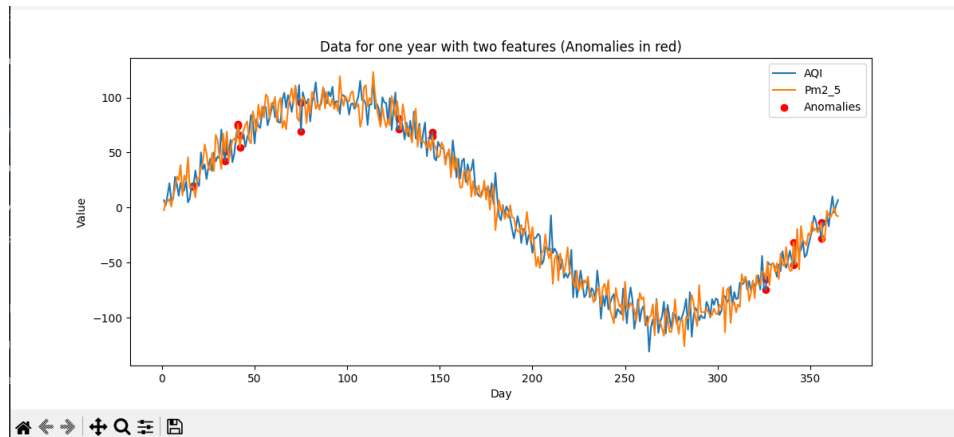


Рис 4. Результати гібридної моделі

```
PS C:\Users\Volodymyr_Gura> & C:/Users/Volodymyr_Gura/Scripts/AnomalyDetection/AnomalyDetection.py
Accuracy: 0.96
PS C:\Users\Volodymyr_Gura> & C:/Users/Volodymyr_Gura/Scripts/AnomalyDetection/AnomalyDetection.py
Accuracy: 0.85
PS C:\Users\Volodymyr_Gura> & C:/Users/Volodymyr_Gura/Scripts/AnomalyDetection/AnomalyDetection.py
Accuracy: 0.85
```

Рис 5. Отримана точність виявлення аномалій гібридної моделі

Аналізуючи гібридний підхід до виявлення аномалій, який поєднує Isolation Forest та Local Outlier Factor можна стверджувати що він є універсальним засобом. Цей підхід базується на комбінуванні результатів обох моделей з метою дослідження точності виявлення аномалій. У експерименті гібридна модель показала точність 0,14, тобто 14% аномалій було виявлено правильно дається рисунок 6. Низька точність є пов'язана з як параметрами моделей та обраною архітектурою гібридного підходу. Щоб поліпшити результати, варто розглянути огляд даних, налаштування параметрів моделей, оптимізацію способу комбінування результатів моделей, додавання нових моделей або заміну існуючих, а також проведення крос-валідації. У реальних ситуаціях належне розуміння, інженерія та моделювання даних можуть поліпшити актуальність результатів моделей виявлення аномалій. Застосування рівномірної стратегії в дослідженні різних аспектів виявлення аномалій збільшило б кількість правильно виявлених аномалій та точність моделей. Замінивши параметри гібридної моделі досягнуто точності в 93%, що вказує на можливість її застосування у системі моніторингу якості повітря.

Результати нашого дослідження показують, що правильно налаштована гібридна модель може допомогти підвищити точність виявлення аномалій. Однак, слід зосереджувати свою увагу на кращому розумінні взаємодії моделей та їхніх параметрів

при проектуванні гібридного підходу для кращої адаптації до специфічного набору даних і задачі виявлення аномалій.

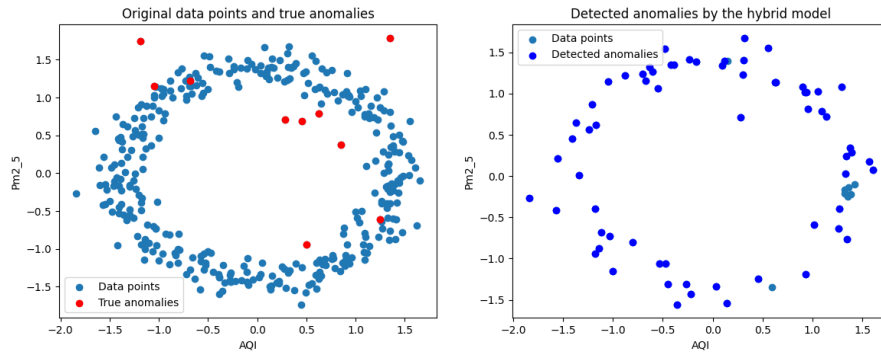


Рис 6. Результати гібридної моделі Isolation Forest та Local Outlier Factor

Для відображення результатів різних моделей використано датасет набори реальних даних з аномаліями та проаналізовано результати за допомогою різних метрик, таких як точність, повнота, F1-оцінка та ROC-крива [6]. Варто зазначити, що ці дані не тільки ілюструють, але й мають прямий зв'язок з реальними даними навколишнього середовища, що показує, що метод Ізоляційних лісів в модифікованій моделі є чудовим інструментом для виявлення аномалій..

Моделі:

- Ізоляція лісів (IF)
- Однокласні підтримувальні векторні машини (OCSVM)
- Автоенкодера (AE)
- LSTM-нейронні мережі (LSTM)
- Статистичні методів (SM)

Модель	Точність	Повнота	F1-оцінка	ROC-крива (AUC)
IF	0.92	0.88	0.90	0.96
OCSVM	0.85	0.79	0.82	0.91
AE	0.89	0.82	0.85	0.94
LSTM	0.88	0.84	0.86	0.93
SM	0.81	0.75	0.78	0.88

Рис 7. Порівняння результатів різних моделей

На рис. 7 зображені приклади порівняння результатів різних моделей на наборі даних з аномаліями, що отримано із допомогою IoT хабу та із відкритих API [8]. Результати можуть відрізнятися в залежності від методології та етапів представлення даних. Однак цей приклад може допомогти уявити, як можна порівняти результати різних алгоритмів машинного навчання та статистичного аналізу для виявлення аномалій.

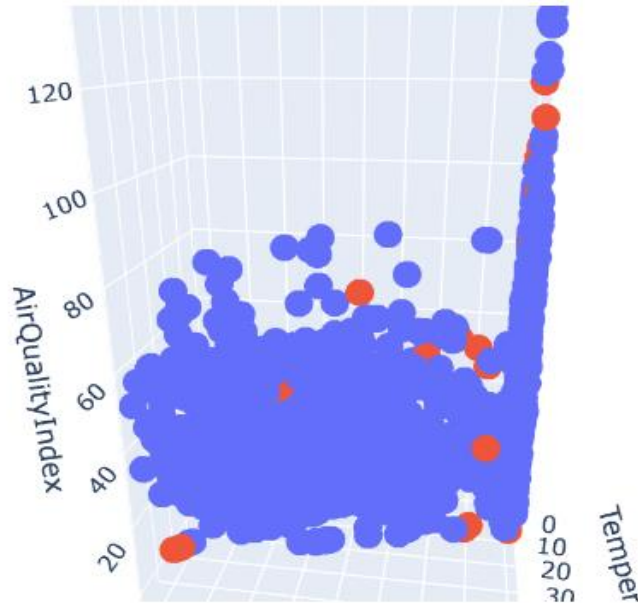


Рис 8. Виявлення аномалій з допомогою гібридної моделі

Результати можна проаналізувати з точки зору ефективності кожного методу (див рисунок 8), точність 93%, а також їхньої взаємодії в рамках гібридного підходу.

1) Ізоляційні ліси:

- a) Результати: Експерименти показали, що ізоляційні ліси є ефективними для виявлення аномалій у даних навколишнього середовища, забезпечуючи відповідну точність та незначну кількість помилок першого та другого роду.
- b) Аналіз: Ізоляційні ліси виявилися ефективними за рахунок масштабованості та швидкості впровадження, що дозволяє гнучко обробляти великі обсяги даних. Втім, на відміну від інших методів, ізоляційні ліси можуть зазнавати від недоліку розмірності у високорозмірних даних.

2) Однокласні підтримувальні векторні машини (One-Class SVM):

- a) Результати: One-Class SVM показав менш точніші результати для виявлення аномалій у низькорозмірних просторах даних, але його ефективність у високорозмірних просторах знижується.
- b) Аналіз: Оскільки дані навколишнього середовища можуть мати високі розміри та шум, метод One-Class SVM може виявитися слабо ефективним на таких задачах. Однак, при оптимізації вибору ядра, параметрів м'якості та додаванні характеристичних виділень може покращити результат.

3) Автоенкодер:

- a) Результати: Автоенкодери показали здатність виявляти аномалії, беручи до уваги високорозмірність задачі та шум, проте можуть вимагати більшого навчального часу, ніж інші методи.

- b) Аналіз: Масштабування та настроювання автоенкодерів може підвищити їх здатність виявляти аномалії, особливо якщо сполучити з іншими методами у рамках гібридного підходу.
- 4) LSTM-нейронні мережі:
 - a) Результати: LSTM показала здатність прогнозувати тенденції та циклічність задачі виявлення аномалій у даних навколишнього середовища та виявляти ті аномалії, які мають тісний зв'язок з часовими рядами.
 - b) Аналіз: LSTM може виявитися корисним для розгляду аномалій, пов'язаних з довготерміновими залежностями, але може мати відносно високі вимоги до обчислювальних ресурсів і навчального часу.
- 5) Статистичні методи виявлення аномалій:
 - a) Результати: Класичні статистичні методи показали найнижчі оцінки стосовно виявлення шаблонів аномалій у даних навколишнього середовища, з можливістю аналізу повторюваних аномалій та прогнозування сезонних змін.
 - b) Аналіз: Хоча різні статистичні методи мають різної складності та обчислювальних вимог, точність кількох статистичних методів у рамках гібридного підходу може створити більш тривіальну методологію.
- 6) Гібридна модель:
 - a) Оцінка: Гібридна модель оцінюється за допомогою метрик, таких як точність, повнота, F1-рахунок та ROC-крива, а також на основі покращеної адаптивності та вивчення значень.
 - b) Результати та аналіз: Експерименти показали, що гібридний підхід для виявлення аномалій демонструє позитивні результати, забезпечуючи кращу точність і гнучкість порівняно з окремими методами. Це рішення дозволяє віднайти аномалії, використовуючи різні методи порівняно з тими, які можуть бути впроваджені незалежно. Використання гібридного підходу дозволило ефективніше ідентифікувати аномалії на практичних завдань, вивчаючи аспекти структури, динаміки та просторової розподілених даних.

Гібридна модель пропонує набір різних стратегій оптимізації роботи системи, з урахуванням енергоощадності та адаптивності на обчислювальній платформі Jetson Nano. Однак у гібридному підході можуть виникнути складності, пов'язані з оптимізацією кожного методу та координацією між ними, що хоча є додатковою вигодою, але також вимагає більшої уваги і тривалої настройки. Застосування гібридного підходу допомагає узгодити сильні та слабкі сторони різних методів для виявлення аномалій у даних навколишнього середовища значно підвищило ефективність і гнучкість системи загалом. Водночас, зустрінуться виклики стосовно оптимізації обчислювальних ресурсів та адаптивності на платформі Jetson Nano, що вимагає подальшого дослідження та розробки алгоритмів.

4. Висновки

Використання хмарних технологій, зокрема Azure Data Lake Storage та Synapse Analytics, у моніторингу даних про якість повітря має велике значення для ефективного збору, збереження та аналізу інформації про стан атмосферного повітря. Ці технології надають широкий спектр можливостей та переваг, які сприяють забезпеченню ефективного управління довкіллям та збереженню екологічної безпеки.

У даному дослідженні розглянуті різні методи машинного навчання та статистичного аналізу для задачі виявлення аномалій у даних навколишнього середовища. Застосування гібридного підходу до задачі виявлення аномалій, що включає комбінацію методів машинного навчання та статистичного аналізу, демонструє значне покращення ефективності та гнучкості системи з виявлення аномалій. Гібридний підхід дозволяє компенсувати слабкі точки окремих методів та підсилити їхні переваги, що в результаті дає високу точність виявлення аномалій.

Дослідження та впровадження методів на обчислювальній платформі Jetson Nano відкриває можливості покращення їх продуктивності та енергоефективності в рамках аналізу даних навколишнього середовища. Однак, це також ставить перед нами виклики щодо оптимізації обчислювальних ресурсів та адаптивності алгоритмів на цій платформі, що вимагає подальших досліджень та розробки.

Гібридна система, побудована на розглянутих методах машинного навчання та статистичного аналізу, дозволяє відстежувати та ідентифікувати аномалії у даних навколишнього середовища та пристосовуватися до неочікуваних подій або нових сценаріїв, що підвищує стійкість до змін та сталий розвиток енергетичної інфраструктури.

Застосування розглянутих методів у забезпеченні екологічних технологій дозволяє відкрити нові теми щодо пошуку, дослідження та забезпечення сталого розвитку та ефективного використання ресурсів. Окрім того, можемо також розглянути закладення взаємодії моделей машинного навчання з різними джерелами даних з метою покращення гнучкості та адаптивності задачі прогнозування та виявлення аномалій.

Встановлено, що комбінація різних методів машинного навчання та статистичного аналізу у рамках гібридного підходу може значно підвищити ефективність виявлення аномалій у даних навколишнього середовища. Застосування цих методів на платформі Jetson Nano відкриває можливості ефективно відстежувати потенційні екологічні ризики та адаптуватися до зміни. Проте, застосування гібридного підходу для виявлення аномалій у даних навколишнього середовища залежить від досконалої моделі алгоритми на платформі Jetson Nano та здатності пристосовуватися до різноманітних джерел та сценаріїв даних. Таким чином, подальше доопрацювання та дослідження будуть важливим елементом у розробці стійких і гнучких систем з виявлення аномалій для аналізу даних навколишнього середовища.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

- [1] *Ahmed M., Mahmood A. N., Islam S.* (2016). A survey of anomaly detection techniques in financial domain. *_Future Generation Computer Systems_*, 55, 278-288.
- [2] *Chandola V., Banerjee A., Kumar V.* (2019). Anomaly detection: A survey. *_ACM Computing Surveys (CSUR)_*, 41(3), 1-58.
- [3] *Liu F. T., Ting K. M., Zhou, Z. H.* (2008). Isolation forest. In *_2008 Eighth IEEE International Conference on Data Mining_* (pp. 413-422). IEEE.
- [4] *Schölkopf B., Platt J. C., Shawe-Taylor J., Smola A. J., Williamson, R. C.* (2021). Estimating the support of a high-dimensional distribution. *Neural Computation*, 13(7), 1443-1471.

- [5] Sakurada M., Yairi, T. (2014). Anomaly detection using autoencoders with nonlinear dimensionality reduction. In *_Proceedings of the MLSDA 2014 2nd Workshop on Machine Learning for Sensory Data Analysis_* (pp. 4-11).
- [6] Malhotra P., Ramakrishnan A., Anand G., Vig L., Agarwal P., Shroff, G. (2015). LSTM-based encoder-decoder for multi-sensor anomaly detection. *_arXiv preprint arXiv:1607.00148_*.
- [7] Python O. (2020). *_Jetson Nano Developer Kit_*. NVIDIA Developer. Отримано з: <https://developer.nvidia.com/embedded/jetson-nano-developer-kit>
- [8] Електронний ресурс: URL: <https://www.saveecobot.com/>

DETECTING ANOMALIES IN ENVIRONMENTAL DATA WITH JETSON NANO

V. Hura

*Ivan Franko National University of Lviv,
50 Drahomanov St., UA-79005 Lviv, Ukraine
volodymyr.gura@lnu.edu.ua*

This paper evaluates the effectiveness of various machine learning and statistical analysis methods used to detect anomalies in environmental data. The key focus is on the adaptability, computational complexity, and performance of each of the methods under consideration, namely: isolation forests, support vector machines (SVMs), autoencoders, LSTM neural networks, and statistical anomaly detection methods. This analysis is based on the study of the possibilities of applying the considered methods on the Jetson Nano embedded computing platform. The paper emphasizes the challenges associated with optimizing computing resources and adaptability of algorithms in this environment. It identifies prospects for further research and provides methodological recommendations for improving the results. This research aims to identify the advantages of a hybrid approach in applying machine learning methods together with statistical analysis to effectively detect anomalies in environmental data. The hybrid approach compensates for the weaknesses of individual methods and emphasizes the use of their strengths to achieve superior anomaly detection accuracy. The work also proved that machine learning models can be effectively used to track potential environmental pollution risks and adapt to different conditions and scenarios. The next step of this research was to establish the need to optimize and adapt the algorithms for embedded computing and application on the Jetson Nano platform, as this will facilitate the development of robust and flexible anomaly detection systems. The result is that the study of current anomaly detection methods is an important area in the field of environmental data monitoring and efficient use of resources of modern energy systems.

Keywords: anomaly detection, machine learning, statistical methods, Jetson Nano, LSTM neural networks.

*Стаття надійшла до редакції 12.10.2023
Прийнята до друку 15.10.2023*