

УДК 004.94:628.4:504.064

<https://doi.org/10.15407/pp2026.01.025>*А. М. Тригуба, Н. Я. Коваль, І. Л. Тригуба, І. Р. Фірман, В. Ю. Фамуляк*

АРХІТЕКТУРА ПРОГРАМНОЇ СИСТЕМИ БАГАТОМОДЕЛЬНОГО ПРОГНОЗУВАННЯ ЕКОЛОГІЧНИХ РИЗИКІВ ДЛЯ ЦИФРОВОГО УПРАВЛІННЯ МУНІЦИПАЛЬНИМИ ОРГАНІЧНИМИ ВІДХОДАМИ

У статті досліджується проблема проектування програмної системи багатомодельного прогнозування екологічних ризиків для цифрового управління муніципальними органічними відходами в умовах зростання обсягів значущих даних і потреби оперативного ухвалення рішень на локальному рівні. Проведено аналіз сучасних підходів до прогнозування екологічних ризиків у системах муніципального управління. Встановлено, що використання окремих моделей не забезпечує достатньої стійкості результатів за умов неоднорідності вхідних даних і нелінійності факторів впливу. Обґрунтовано доцільність побудови програмної системи на основі багаторівневої архітектури, яка включає модулі підготовки даних, прогнозного аналізу, оцінювання моделей, інтерпретації ознак і підтримки цифрових управлінських рішень. Запропоновано багатомодельний алгоритм, до складу якого включено Random Forest, Gradient Boosting та XGBoost, що дозволяє автоматично визначати найефективнішу модель за критеріями MAE, RMSE і R^2 . Розроблено UML-архітектуру взаємодії класів DataManager, RiskModelEngine, BenchmarkController і DecisionSupport. Для пояснення результатів використано метод перестановочного оцінювання важливості ознак, що дозволило визначити домінування просторових та інфраструктурних факторів у формуванні ризику. Результати експериментального тестування показали, що найкращу точність забезпечує модель XGBoost (MAE=0.039, RMSE=0.055, $R^2=0.930$), що підтверджує ефективність запропонованого підходу.

Ключові слова: екологічний ризик, багатомодельне прогнозування, муніципальні органічні відходи, машинне навчання, Random Forest, Gradient Boosting, XGBoost, UML-архітектура, permutation importance, цифрове управління, інформаційна система, підтримка ухвалення рішень.

А.М. Tryhuba, N.Ya. Koval, I.L. Tryhuba, I.R. Firman, V.Yu. Famuliak

ARCHITECTURE OF A SOFTWARE SYSTEM FOR MULTI-MODEL ECOLOGICAL RISK FORECASTING FOR DIGITAL MANAGEMENT OF MUNICIPAL ORGANIC WASTE

The article investigates the problem of designing a software system for multi-model ecological risk forecasting for digital management of municipal organic waste under conditions of increasing volumes of significant data and the need for prompt decision-making at the local level. An analysis of modern approaches to ecological risk forecasting in municipal management systems has been carried out. It has been concluded that the use of individual models does not ensure sufficient stability of results under conditions of heterogeneous input data and nonlinear influence factors. The expediency of constructing a software system based on a multi-level architecture has been substantiated, including modules for data preparation, predictive analysis, model evaluation, feature interpretation, and digital decision support. A multi-model algorithm has been proposed, incorporating Random Forest, Gradient Boosting, and XGBoost, which makes it possible to automatically determine the most effective model according to the criteria MAE, RMSE, and R^2 . A UML architecture of interaction between the classes DataManager, RiskModelEngine, BenchmarkController, and DecisionSupport has been developed. To explain the results, the permutation feature importance method has been applied, which made it possible to determine the dominance of spatial and infrastructural factors in risk formation. Experimental testing results showed that the highest accuracy is provided by the XGBoost model (MAE=0.039, RMSE=0.055, $R^2=0.930$), confirming the effectiveness of the proposed approach.

Keywords: ecological risk, multi-model forecasting, municipal organic waste, machine learning, Random Forest, Gradient Boosting, XGBoost, UML architecture, permutation importance, digital management, information system, decision support.

Вступ

Цифрова трансформація муніципального управління відходами поступово переходить від рівня електронного обліку до використання інтелектуальних систем, здатних прогнозувати екологічні наслідки функціонування міської інфраструктури в умовах зростаючого антропогенного навантаження. Особливо це стосується органічних відходів, які у структурі міських потоків формують найбільш динамічну частину біологічно активної маси, що супроводжується утворенням метану, зміною температурних режимів, накопиченням фільтрату та підвищенням ризику локального екологічного перевантаження. У таких умовах управлінські рішення повинні ґрунтуватися не лише на фактичних показниках утворення відходів, а й на прогнозі розвитку екологічних ризиків у короткотерміновій перспективі [1].

Формування цифрових систем управління муніципальними органічними відходами ускладнюється тим, що екологічний ризик має багатофакторну природу і виникає під впливом одночасної дії демографічних, просторових, інфраструктурних, сезонних і технологічних чинників. Підвищення щільності забудови, нерівномірність логістичних потоків, зміна складу органічної фракції та перевантаження переробних потужностей створюють нелінійні залежності, які адекватно не описуються традиційними аналітичними моделями. Саме тому сучасні цифрові платформи управління відходами потребують програмних архітектур, здатних інтегрувати кілька моделей прогнозування, порівнювати їхню ефективність та формувати адаптивні сценарії підтримки управлінських рішень [2].

Складність задачі полягає також у тому, що екологічні процеси в системах органічного поводження з відходами характеризуються різною швидкістю розвитку, неоднорідністю просторового розподілу та високою залежністю від зовнішніх факторів середовища. Наприклад, збільшення частки урбанізованої території одночасно впливає на обсяги накопичення органічної маси, інтенсивність транспортного наван-

таження та рівень локальних викидів, що унеможливує використання лише одного прогнозного алгоритму для всіх типів територій. За таких умов багатомодельний підхід стає не просто альтернативою, а необхідною архітектурною основою цифрової системи прогнозування [3].

У сучасних умовах екологічного моніторингу особливої ваги набуває здатність програмної системи працювати з різномірними даними: статистичними показниками, геопросторовою інформацією, часовими рядами, результатами сенсорного моніторингу та інфраструктурними характеристиками територій. Це вимагає багаторівневої архітектури, в якій окремі модулі виконують функції збору, нормалізації, моделювання, оцінювання та інтерпретації результатів [4].

Стан предметної області

Упродовж останніх років у міжнародній практиці спостерігається активне впровадження підходів, орієнтованих на дані у сфері поводження з муніципальними відходами, де прогнозування виконується на основі алгоритмів машинного навчання, просторової аналітики та цифрових двійників міської інфраструктури [5]. Водночас більшість досліджень орієнтована або на прогнозування кількості відходів, або на оптимізацію маршрутів перевезення, тоді як екологічний ризик як інтегральний показник розглядається значно рідше.

Проведений аналіз публікацій показав, що сучасні системи прогнозування екологічного ризику використовують переважно окремі алгоритми, серед яких найчастіше застосовуються Random Forest, Gradient Boosting, Support Vector Regression та XGBoost [6]. Однак у більшості випадків відсутня програмна архітектура, яка дозволяє виконувати системний бенчмаркінг моделей, автоматично оцінювати якість прогнозу та забезпечувати адаптивний вибір алгоритму залежно від структури вхідних даних.

Виявлено, що існуючі цифрові рішення не забезпечують повної інтеграції

просторових, інфраструктурних та часових змінних у межах єдиної програмної системи. Частина платформ працює лише з табличними даними, інші – лише з GIS-компонентами, а модулі пояснюваності моделей часто взагалі відсутні [7].

Аналіз сучасних підходів

Сучасні технології прогнозування екологічних процесів базуються на кількох групах методів. Першу групу становлять класичні статистичні підходи, зокрема, регресійні моделі та часові ряди, які дозволяють оцінювати середні тенденції, однак мають обмежену здатність описувати складні нелінійні взаємодії [8-10].

Другу групу формують ансамблеві алгоритми машинного навчання, які забезпечують кращу стійкість до шуму даних і здатні працювати з високою розмірністю ознак. Зокрема, Random Forest добре виявляє себе у задачах екологічного ризику через здатність до автоматичного виявлення значущих взаємодій між змінними [11-13].

Третя група пов'язана з використанням пояснювального штучного інтелекту, де поряд із прогнозом формується пояснення впливу кожної ознаки на результат моделі. Це особливо важливо для муніципального управління, де рішення мають бути інтерпретованими для органів місцевого самоврядування [14-15].

Виявлені протиріччя. Аналіз літератури дозволив виявити три основні протиріччя. Перше протиріччя виникає між багатофакторністю екологічного ризику та використанням у більшості систем лише одного алгоритму прогнозування. Друге протиріччя виникає між динамічністю екологічних процесів та статичністю існуючих програмних архітектур, які не підтримують автоматичне оновлення моделей. Третє протиріччя виявляється між високою точністю сучасних ансамблевих моделей і недостатнім рівнем їх інтерпретованості у прикладних муніципальних інформаційних системах.

Мета статті полягає у розробленні архітектури програмної системи багатомодельного прогнозування екологічних ризи-

ків для цифрового управління муніципальними органічними відходами на основі інтеграції ансамблевих методів машинного навчання, модулів оцінювання якості прогнозу та засобів інтерпретації результатів.

Виклад основного матеріалу дослідження

Формування архітектури програмної системи багатомодельного прогнозування екологічних ризиків. Цей етап є базовим. Він забезпечує побудову інтелектуальної інструментарію, здатного забезпечувати адаптивне аналітичне оцінювання стану екологічних систем в умовах динамічної зміни зовнішнього середовища. Особливість такої програмної системи полягає у необхідності одночасного використання кількох математичних моделей прогнозування, що функціонують паралельно, забезпечуючи підвищення точності оцінювання ризиків за рахунок комбінування результатів різних алгоритмічних підходів.

У сучасних інформаційних системах екологічного аналізу архітектурна побудова має ґрунтуватися на принципі багатосферності, який забезпечує незалежність функціональних компонентів, масштабованість, можливість модифікації окремих модулів без порушення роботи всієї системи, а також адаптацію до нових джерел даних і нових моделей прогнозування. У запропонованій структурі архітектури система поділяється на чотири основні функціональні рівні: 1) рівень збору та підготовки даних; 2) аналітичний рівень; 3) рівень бізнес-логіки; 4) рівень представлення результатів.

Перший рівень формують джерела первинної інформації, які включають результати сенсорного моніторингу, геопросторові набори даних, статистичну екологічну інформацію, метеорологічні параметри, характеристики техногенного навантаження території, а також часові ряди спостережень. Формально множину вхідних даних системи доцільно подати як:

$$D = \{(x_i, y_i)\}_{i=1}^N, \quad (1)$$

де x_i – вектор факторних ознак, що характеризують екологічний стан території; y_i – цільове значення показника ризику; N – кількість спостережень у навчальній вибірці.

Перед надходженням до аналітичного рівня дані проходять процедури очищення, нормалізації, усунення пропусків та приведення до єдиного масштабу. Нормалізація ознак здійснюється за виразом:

$$x_i^{norm} = \frac{x_i - x_{min}}{x_{max} - x_{min}}, \quad (2)$$

де x_i – поточне значення ознаки; x_{min} – мінімальне значення ознаки у вибірці; x_{max} – максимальне значення ознаки у вибірці.

Нормалізація ознак дозволяє забезпечити коректність навчання різнорідних моделей.

Другий рівень архітектури становить аналітичне ядро системи, в якому функціонують незалежні програмні модулі для прогнозування. Кожен модуль реалізує окрему модель варіативності – нейромережу, ансамблю, регресійну або статистичну. Відповідно до принципу багато-модельності результати окремих моделей об'єднуються у єдиний інтегрований прогноз рівня екологічного ризику шляхом агрегування виходів окремих прогнозних модулів:

$$R(t) = \sum_{k=1}^M w_k f_k(x_t), \quad (3)$$

де $f_k(x_t)$ – результат k -ої моделі прогнозування рівня екологічного ризику; w_k – ваговий коефіцієнт моделі; M – кількість моделей у системі.

Використання кількох моделей одночасно дозволяє мінімізувати похибку окремих алгоритмів та підвищити стійкість прогнозу рівня екологічного ризику до нестабільності даних.

У межах аналітичного рівня особливу роль виконує модуль аналізу поведінкової інформації, який акумулює результати

функціонування окремих моделей і виконує оцінювання їхньої адаптивності до поточних змін проектного середовища. Саме цей модуль забезпечує перехід від статичної моделі прогнозування до динамічної архітектури, в якій система здатна змінювати ваги моделей залежно від поточної ефективності.

Третій рівень формує бізнес-логіку системи. Тут реалізуються правила інтерпретації прогнозних результатів, порогові значення небезпеки, механізми генерації попереджень та сценарії ухвалення рішень. Для цього вводиться функція ризикової інтерпретації:

$$Z = \begin{cases} 1, & R(t) \geq R_{crit} \\ 0, & R(t) < R_{crit} \end{cases}, \quad (4)$$

де R_{crit} – порогове значення ризику.

Якщо прогнозований рівень ризику перевищує порогове значення, система формує сигнал для управлінського реагування.

Четвертий рівень архітектури забезпечує представлення результатів користувачу. Він включає інтерфейс візуалізації, аналітичні панелі, графіки часової динаміки, карти ризиків і таблиці показників. Саме цей рівень забезпечує інтеграцію системи з кінцевим користувачем і дозволяє застосовувати результати прогнозування у практичній діяльності.

Запропонована архітектура програмної системи побудована за принципом модульної декомпозиції, що дозволяє незалежно розширювати кількість моделей прогнозування без перебудови всієї системи. Кожен новий програмний модуль інтегрується у загальну архітектуру через уніфікований інтерфейс передачі функціональних потоків даних.

На відміну від класичних однотипних систем прогнозування, багатомодельна архітектура забезпечує одночасне врахування різних закономірностей у структурі екологічних даних, що особливо важливо в умовах нелінійності екологічних процесів, сезонної варіативності та багатфакторності ризиків.

Багатомодельний алгоритм прогнозування екологічного ризику. Основною особливістю запропонованої програмної системи є реалізація багатомодельного алгоритму прогнозування екологічного ризику, який ґрунтується на паралельному використанні кількох ансамблевих методів машинного навчання. Такий підхід зумовлений тим, що екологічні ризики у системах управління муніципальними органічними відходами формуються під впливом великої кількості неоднорідних факторів, між якими існують складні нелінійні зв'язки. За цих умов використання лише одного алгоритму не завжди забезпечує достатню стійкість прогнозу, тоді як порівняльне застосування кількох моделей дає змогу виявити ту архітектуру, яка найкраще відображає структуру даних і забезпечує найменшу похибку.

У межах дослідження до складу багатомодельного прогнозного контуру включено три ансамблеві алгоритми: Random Forest, Gradient Boosting та XGBoost. Їх вибір обумовлений тим, що всі три моделі добре працюють із багатовимірними наборами даних, здатні враховувати нелінійність взаємозв'язків між ознаками та ці-

лювою змінною, а також демонструють високу ефективність у задачах регресійного прогнозування зі складною внутрішньою структурою факторів.

У загальному вигляді множину моделей, що входять до складу системи, представлено виразом:

$$M = \{M_{RF}, M_{GB}, M_{XGB}\}, \quad (5)$$

де M_{RF} – модель Random Forest; M_{GB} – модель Gradient Boosting; M_{XGB} – модель XGBoost.

Для кожної моделі прогнозне значення інтегрального екологічного ризику визначається залежністю:

$$\hat{y}_i^{(k)} = M_k(x_i), \quad (6)$$

де x_i – вектор ознак для i -го об'єкта спостереження; M_k – k -та модель із множини M ; $\hat{y}_i^{(k)}$ – прогнозоване значення ризику.

Алгоритм Random Forest належить до методів bagging і формує підсумковий прогноз шляхом усереднення результатів великої кількості дерев рішень, побудова-

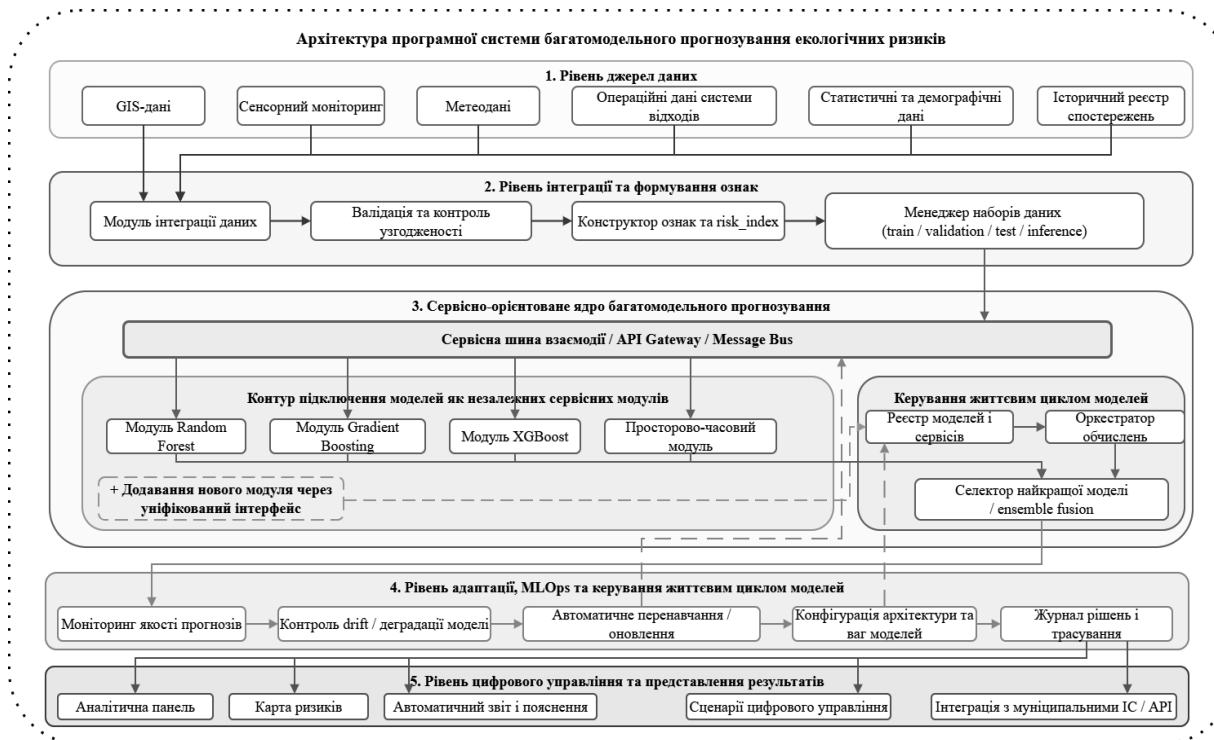


Рис. 1. Схема архітектури програмної системи багатомодельного прогнозування

них на випадкових підвбірках даних та випадкових підмножинах ознак. Завдяки цьому знижується дисперсія прогнозу, підвищується стійкість моделі до шуму та зменшується ризик перенавчання. У задачах екологічного прогнозування така модель є особливо корисною тоді, коли фактори ризику мають складну структуру, а частина ознак є слабоформалізованими або корельованими між собою.

Модель Gradient Boosting реалізує іншу логіку побудови ансамблю. На відміну від Random Forest, у цій архітектурі дерева формуються послідовно, причому кожне наступне дерево намагається компенсувати помилки попередніх. Такий механізм дозволяє краще вловлювати тонкі нелінійні закономірності у даних та забезпечує вищу чутливість до складних локальних залежностей. Водночас ця модель потребує ретельнішого налаштування глибини дерев, швидкості навчання та кількості ітерацій, оскільки надмірне підсилення призводить до погіршення узагальнювальної здатності.

Модель XGBoost є розвитком ідеології бустингу та поєднує високу прогностичну потужність із засобами регуляризації, контролем складності дерев і ефективною обробкою великих масивів даних. Для задач прогнозування екологічного ризику ця модель є особливо цінною, оскільки дозволяє не лише підвищити точність прогнозу, а й забезпечити гнучке врахування складних комбінацій ознак, які формують інтегральний показник ризику. Саме тому XGBoost доцільно розглядати як одну з базових архітектур у системі багатомодельного прогнозування.

Оцінювання ефективності моделей здійснюється за допомогою метрик, що характеризують середнє відхилення прогнозу від фактичного значення цільової змінної. Однією з базових метрик є середня абсолютна похибка:

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|, \quad (7)$$

де y_i – фактичне значення інтегрального ризику; \hat{y}_i – прогнозоване значення інтег-

рального ризику; N – кількість спостережень.

Середньоквадратична похибка визначається за формулою:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}. \quad (8)$$

На відміну від MAE, ця метрика сильніше реагує на великі відхилення, тому є особливо корисною у тих випадках, коли необхідно контролювати точність прогнозів у сегментах підвищеного екологічного ризику.

Для повнішого аналізу якості прогнозу доцільно також використовувати коефіцієнт детермінації:

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2}. \quad (9)$$

де \bar{y} – середнє значення цільової змінної у вибірці.

Метрика (9) відображає, яку частку варіації інтегрального індексу ризику пояснює модель.

Багатомодельний алгоритм у межах програмної системи реалізується як послідовність взаємопов'язаних операцій. Спочатку підготовлений набір ознак надходить на вхід усіх трьох моделей. Далі кожна з них виконує незалежне навчання та формує прогнозні значення. Після цього розраховуються метрики якості, на підставі яких виконується порівняння моделей і визначається та архітектура, що найкраще відображає структуру конкретних даних. У такий спосіб система не фіксує одну “жорстко задану” модель, а працює в адаптивному режимі вибору найефективнішого інструмента прогнозування.

З методологічної точки зору такий підхід є виправданим, оскільки екологічний ризик не має універсальної аналітичної форми. Для одних територіальних сукупностей кращі результати забезпечують моделі типу bagging, для інших – boosting-підходи. Саме тому багатомодельність у

даній роботі розглядається не як дублювання алгоритмів, а як спосіб підвищення достовірності прогнозного контуру та зме-

ншення залежності системи від структурних обмежень одного методу.

Таблиця 1. Порівняння моделей для прогнозування екологічних ризиків

Модель	Принцип побудови	Переваги	Обмеження	Доцільність у задачі прогнозування екологічного ризику
Random Forest	Bagging ансамбль дерев рішень	Стійкість до шуму, низька чутливість до перенавчання, добра робота з великою кількістю ознак	Менша чутливість до слабких локальних залежностей	Ефективна для стабільного прогнозу за неоднорідних і частково зашумлених даних
Gradient Boosting	Послідовне підсилення дерев	Висока точність, здатність моделювати складні нелінійні залежності	Чутливість до параметрів і ризик перенавчання	Доцільна для виявлення складних взаємозв'язків між чинниками ризику
XGBoost	Регуляризований boosting	Висока прогностична потужність, швидкодія, контроль складності моделі	Вища складність налаштування та інтерпретації	Найбільш придатна для побудови точного прогнозу інтегрального індексу ризику

Таким чином, багатомодельний алгоритм прогнозування екологічного ризику формує аналітичне ядро всієї програмної системи. Його використання дозволяє не лише порівнювати різні ансамблеві архітектури, а й обґрунтовано обирати ту модель, яка забезпечує найкращу якість прогнозування для конкретної структури екологічних, просторових та інфраструктурних даних. Саме це створює передумови для подальшого формування адаптивної системи цифрового управління муніципальними органічними відходами на основі надійних прогнозних оцінок.

UML-архітектура програмної взаємодії модулів. Для забезпечення структурованої реалізації багатомодельного прогнозування екологічного ризику програмну систему доцільно організувати за модульним принципом, у якому окремі

функціональні блоки реалізують незалежні, але логічно пов'язані етапи обробки даних, навчання моделей, оцінювання результатів та формування управлінських висновків. Такий підхід дозволяє підтримувати масштабованість архітектури, розширювати систему новими аналітичними компонентами без зміни базової логіки та забезпечувати повторне використання окремих програмних модулів у суміжних задачах екологічного аналізу.

У структурі програмної архітектури центральне місце займає клас DataManager, який відповідає за організацію вхідних даних. Його функціональна роль полягає у завантаженні даних із різних джерел, приведенні їх до єдиного формату, виконанні нормалізації, перевірки повноти записів, формуванні навчальної та тестової вибірок, а також підготовці маси-

вів ознак для передачі у прогностні модулі. Саме через цей клас реалізується первинна взаємодія між табличними даними, геопросторовими характеристиками територій та інтегральними показниками, що використовуються в моделюванні ризику.

Після завершення підготовчого етапу інформаційний потік передається до класу RiskModelEngine, який реалізує ядро багатомодельного прогнозування. У межах цього класу відбувається ініціалізація моделей машинного навчання, їх навчання, тестування, генерація прогнозів та збереження результатів. Архітектурно цей модуль повинен забезпечувати можливість підключення різних алгоритмів через уніфікований інтерфейс, що дозволяє легко додавати нові методи без зміни загальної логіки системи. У цьому класі формується множина прогностних моделей, описаних виразом (5). Водночас кожна модель реалізується як окремий внутрішній об'єкт прогностного контуру.

Для порівняння результатів різних моделей використовується клас BenchmarkController, який виконує аналітичну функцію оцінювання точності про-

гнозів. У цьому модулі обчислюються метрики якості, формується ранжування моделей за ефективністю, визначається модель з мінімальною похибкою та здійснюється формування порівняльної таблиці результатів. Саме цей клас забезпечує адаптивність системи, оскільки дозволяє автоматично вибирати найкращу модель для конкретної структури вхідних даних.

Остаточний етап програмної логіки реалізується класом DecisionSupport, який перетворює числовий прогноз ризику на управлінське рішення. Його функціональна роль полягає у класифікації територій за рівнями ризику, генерації текстових рекомендацій, підготовці аналітичних повідомлень та передачі результатів у зовнішні цифрові сервіси.

Таким чином, UML-архітектура системи відображає логіку послідовної взаємодії між чотирма базовими класами: 1) підготовка даних; 2) прогнозування; 3) оцінювання; 4) підтримка управлінських рішень. Саме така структурна декомпозиція забезпечує оновлення окремих частин системи без зміни її загальної функціональної логіки.

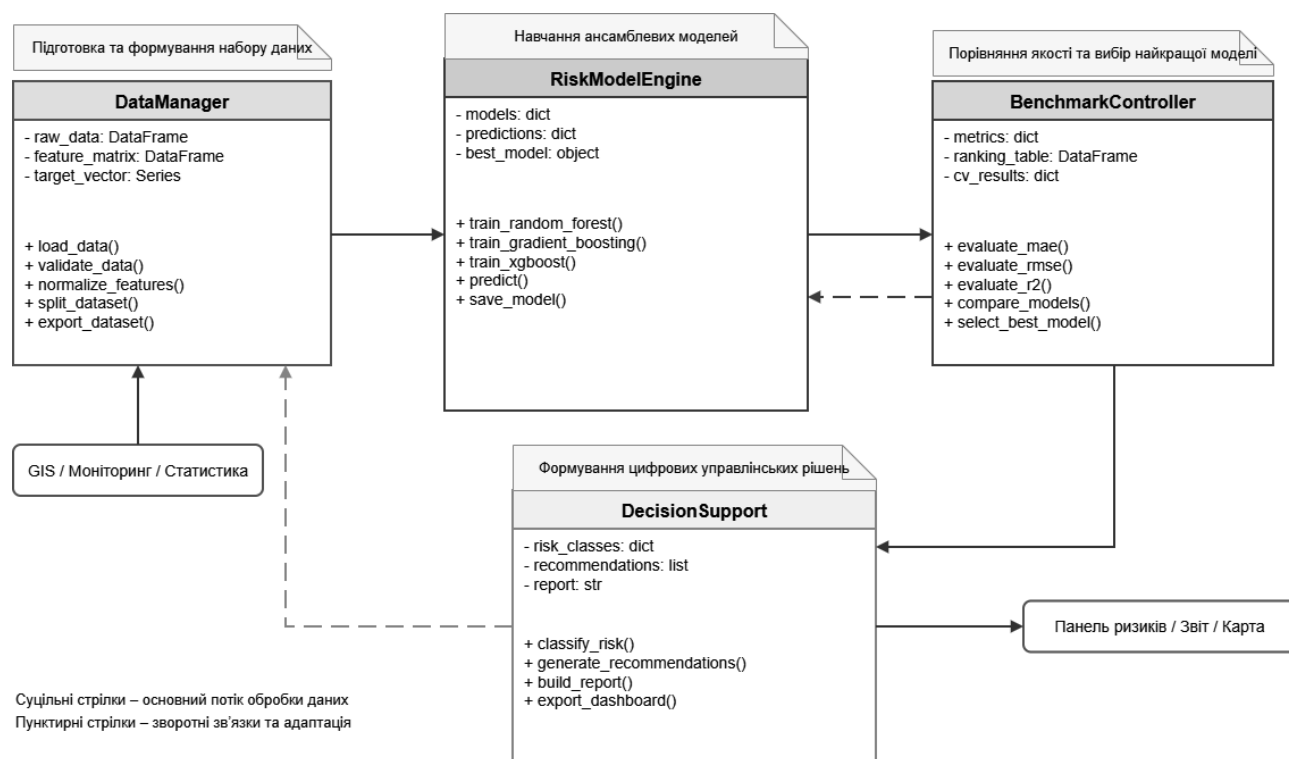


Рис. 2. UML-архітектура програмної системи

Модуль інтерпретації важливості ознак. Однією з принципових вимог до сучасних інтелектуальних систем прогнозування є не лише отримання точного результату, а й пояснення причин, які формують конкретне прогнозне значення. У задачах екологічного ризику це особливо важливо, оскільки управлінські рішення повинні спиратися на зрозумілу аналітичну аргументацію. Для цього у програмній системі реалізовано модуль інтерпретації важливості ознак, який дозволяє оцінити внесок кожного фактора у формування інтегрального індексу ризику.

В межах дослідження використовується підхід *permutation importance*, що базується на вимірюванні зміни функції втрат після випадкового перемішування значень окремої ознаки:

$$PI_j = E \left[L(y, \hat{f}(X_j^{perm})) - L(y, \hat{f}(X)) \right] \quad (10)$$

де PI_j – відображає важливість j -ї ознаки; $L(\cdot)$ – функція втрат моделі; $\hat{f}(X)$ – базовий прогноз моделі; X_j^{perm} – набір даних із випадково переставленими значеннями j -ю ознаки.

Ідея цього підходу полягає в тому, що у разі, якщо після перемішування значень конкретної змінної якість прогнозу екологічного ризику різко погіршується, це означає високу аналітичну значущість відповідного фактора.

У прикладних умовах екологічного прогнозування система часто виявляє домінування просторових ознак. Це пояснюється тим, що територіальні параметри безпосередньо визначають доступність інфраструктури, логістику збору відходів, щільність забудови, близькість до природних об'єктів і транспортних коридорів. Саме тому змінні типу площі території, відстані до водних об'єктів, щільності забудови або просторової структури населеного пункту часто демонструють найбільші значення *permutation importance*.

Водночас інфраструктурні змінні також мають високу аналітичну вагу. Наприклад, показники інтенсивності транс-

портної доступності, наявності полігонів, кількості логістичних вузлів чи рівня комунального забезпечення визначають потенційну здатність території до ефективного управління органічними потоками. У цьому випадку високий внесок змінної означає, що саме вона формує найбільший вплив на зміну інтегрального ризику.

Отож, модуль інтерпретації дозволяє перейти від “чорної скриньки” прогнозу до аналітично поясненого механізму ухвалення рішень.

Формування цифрових управлінських рішень. Отримане числове значення прогнозного ризику має бути перетворене на форму, придатну для практичного використання в системах муніципального управління. Для цього у програмній системі реалізовано механізм класифікації прогнозу за трьома рівнями ризику:

$$C(x) = \begin{cases} C_1, & R < 0.33 \\ C_2, & 0.33 \leq R < 0.66, \\ C_3, & R \geq 0.66 \end{cases} \quad (11)$$

де R – прогнозне значення інтегрального індексу ризику.

Клас C_1 відповідає низькому рівню ризику. У цьому випадку система фіксує стабільний стан екологічної ситуації, відсутність критичних тенденцій накопичення негативних факторів та достатню ефективність існуючих управлінських механізмів. Клас C_2 характеризує середній рівень ризику. Такий стан означає наявність факторів, які у середньостроковій перспективі призводять до погіршення ситуації. У цьому випадку система формує рекомендації щодо посиленого моніторингу, уточнення логістичних маршрутів, корекції графіків вивезення відходів або підвищення частоти контролю окремих зон.

Клас C_3 відображає високий рівень ризику. Для таких територій система генерує рекомендації щодо пріоритетного управлінського втручання: перегляду інфраструктурної конфігурації, коригування просторової схеми збору відходів, підсилення контролю критичних ділянок, зміни сценарію управління ресурсами.

Цифрові рекомендації формуються автоматично через набір правил, які пов'язують клас ризику з типовими сценаріями реагування. Таким чином прогноз переходить у форму готового управлінського інструмента.

Особливості програмної реалізації системи. Програмна реалізація системи виконана мовою Python, що забезпечує високу гнучкість побудови аналітичних модулів, широкі можливості інтеграції з бібліотеками машинного навчання та просторового аналізу.

Основним середовищем експериментальної реалізації виступає Google Colab, що дозволяє організувати повний цикл дослідження без локального розгортання серверної інфраструктури, виконувати навчання моделей у хмарному середовищі та зберігати результати експериментів у єдиному репозиторії.

Для табличної обробки використовуються бібліотека Pandas, через яку реалізуються операції очищення, агрегації, нормалізації та трансформації ознак. Саме на цьому рівні формується аналітичний датафрейм, який далі передається у модельний контур.

Моделі Random Forest і Gradient Boosting реалізовані через бібліотеку Scikit-learn, що дозволяє використовувати стандартизовані механізми навчання, кросвалідації, оцінювання та побудови інтерпретаційних характеристик.

Для XGBoost використовується окремий пакет XGBoost, який забезпечує ефективне регуляризоване навчання ансамблевих дерев та високу швидкість у роботі з багатовимірними наборами даних.

Особливістю системи є інтеграція з GIS-даними, що реалізується через сумісність із геопросторовими структурами даних. Це дозволяє використовувати координатні характеристики, просторові індекси, територіальні відстані та інфраструктурні параметри як повноцінні ознаки моделі.

У результаті сформована програмна система поєднує аналітичну гнучкість Python, обчислювальні можливості ансамблевих алгоритмів та просторову логіку цифрового управління, що створює основу

для масштабованого використання у задачах екологічного прогнозування.

Результати експериментального тестування системи. Експериментальне тестування розробленої програмної системи виконувалося з метою перевірки здатності багатомодельного алгоритму забезпечувати достовірне прогнозування інтегрального екологічного ризику в умовах неоднорідності просторових, інфраструктурних та соціально-економічних характеристик територій. Основна увага приділялася порівнянню точності ансамблевих моделей, аналізу стійкості прогнозу та визначенню тієї архітектури, яка найкраще відображає закономірності формування ризику в умовах муніципального управління органічними відходами.

Для проведення тестування сформовано експериментальний масив даних, що включав територіальні одиниці з різними характеристиками щільності населення, площі, транспортної доступності, близькості до водних об'єктів, рівня урбанізації, доходів населення та показників утворення органічних відходів. Перед початком навчання всі змінні були приведені до безрозмірної форми, а вибірка поділена на навчальну та тестову частини у співвідношенні 80:20. Такий підхід дозволив забезпечити коректну перевірку узагальнюючої здатності моделей.

На першому етапі проведено навчання трьох моделей: Random Forest, Gradient Boosting та XGBoost. Для кожної з них виконано однакову процедуру побудови прогнозу, після чого здійснено оцінювання за критеріями середньої абсолютної похибки, середньоквадратичної похибки та коефіцієнта детермінації. Порівняльний аналіз показав, що всі три алгоритми забезпечують прийнятну якість прогнозування, однак між ними спостерігається суттєва різниця в точності.

У випадку моделі Random Forest отримано стабільний результат із низькою чутливістю до локальних коливань вибірки. Середня абсолютна похибка цієї моделі склала 0.061, середньоквадратична похибка – 0.082, а коефіцієнт детермінації становив 0.881. Такі значення свідчать про достатньо високу узагальнювальну здат-

ність алгоритму, однак модель менш чутливо відображає слабкі нелінійні залежності між окремими просторовими параметрами.

Модель Gradient Boosting продемонструвала вищу адаптивність до складних нелінійних залежностей. Отримане значення середньої абсолютної похибки становило 0.048, а RMSE – 0.069. Коефіцієнт детермінації досяг 0.907, що підтверджує кращу здатність моделі пояснювати варіацію інтегрального індексу ризику порівняно з Random Forest.

Найвищу точність прогнозування забезпечила модель XGBoost. Для неї зафіксовано мінімальне значення середньої абсолютної похибки $MAE = 0.039$ та найменше значення середньоквадратичної похибки – $RMSE = 0.055$. Коефіцієнт детермінації досяг рівня – $R^2 = 0.93$.

Отримані результати свідчать про найвищу прогностичну стійкість саме цієї архітектури. Це пояснюється тим, що XGBoost поєднує механізм послідовного підсилення слабких моделей із регуляризацією, яка стримує перенавчання та забезпечує кращу узагальнюючу здатність.

Таблиця 2. Результати експериментального тестування моделей

Модель	MAE	RMSE	R ²	Узагальнена оцінка
Random Forest	0.061	0.082	0.881	Стабільний прогноз за високої неоднорідності даних
Gradient Boosting	0.048	0.069	0.907	Висока чутливість до нелінійних залежностей
XGBoost	0.039	0.055	0.930	Найвища точність прогнозування

Порівняння прогнозних значень із фактичними показало, що найбільші відхилення виникають у територіальних одиницях із комбінованим впливом просторових та інфраструктурних факторів. Саме у цих випадках роль складних взаємодій між ознаками найбільш виражена, а перевага XGBoost проявляється найчіткіше.

Додатковий аналіз permutation importance підтвердив, що найбільший внесок у формування прогнозу мають змінні, пов'язані з просторовою структурою території. Найвищі значення важливості було зафіксовано для ознак, що характеризують площу території, частку урбанізованої забудови, відстань до водних об'єктів і рівень транспортної доступності. Це свідчить про те, що геопросторові параметри визначають базову архітектуру формування екологічного ризику.

Окремо проведено аналіз поведінки системи у процесі класифікації прогнозів за рівнями ризику. Встановлено, що більшість територій із високими значеннями

індексу ризику належать до класу C_3 , де необхідне активне управлінське втручання. Для середнього рівня ризику система формує сценарії посиленого моніторингу, тоді як для низького рівня ризику пропонується збереження чинної організації управління.

Із технологічної точки зору тестування підтвердило стабільність програмної реалізації у середовищі Python та відсутність критичних обмежень при масштабуванні набору даних. Навіть у разі збільшення кількості спостережень система зберігала стійкість виконання, а час побудови прогнозу залишався прийнятним для практичного використання.

Таким чином, результати експериментального тестування підтвердили, що запропонована програмна система здатна забезпечувати високоточне прогнозування екологічного ризику, а багатомодельний підхід створює аналітичну основу для адаптивного цифрового управління муніципальними екологічними процесами.

Висновки

У роботі розроблено архітектуру програмної системи багатомодельного прогнозування екологічного ризику для вирішення задачі цифрового управління муніципальними органічними відходами. Вона базується на інтеграції методів машинного навчання, модульної програмної організації, механізмів інтерпретації важливості ознак та цифрової підтримки управлінських рішень. Запропонований підхід орієнтований на обробку неоднорідних просторових, інфраструктурних і соціально-економічних даних, що дозволяє формувати аналітично обґрунтовані прогнози інтегрального екологічного ризику.

У межах дослідження сформовано багатомодельний алгоритм прогнозування екологічного ризику, до складу якого включено ансамблеві методи Random Forest, Gradient Boosting та XGBoost. Реалізовано формальну схему побудови прогнозу екологічного ризику, в якій кожна модель функціонує як окремий аналітичний модуль, а підсумкова оцінка якості здійснюється за критеріями MAE, RMSE та R^2 . Це дозволило забезпечити не лише порівняння моделей, а й забезпечити адаптивний вибір тієї архітектури, яка найкраще відповідає структурі конкретних даних.

Розроблено UML-архітектуру програмної системи, у якій передбачена функціональна взаємодія між класами DataManager, RiskModelEngine, BenchmarkController та DecisionSupport, що забезпечує повний цикл обробки – від завантаження і нормалізації даних до формування цифрових управлінських рекомендацій. Така модульна побудова уможливує масштабування системи, підключення нових моделей та інтеграції із зовнішніми аналітичними сервісами.

Важливою складовою системи є модуль інтерпретації важливості ознак, реалізований на основі підходу permutation importance, що дозволяє оцінювати реальний внесок окремих факторів у формування прогнозу. Це забезпечує пояснюваність результатів машинного навчання та дозволяє виявляти фактори, які найбільше впливають на зміну екологічного ризику. В

результаті досліджень встановлено, що найбільшу аналітичну вагу мають просторові характеристики територій, показники транспортної доступності, рівень урбанізації та інфраструктурні параметри.

Експериментальне тестування програмної системи багатомодельного прогнозування екологічних ризиків підтвердило високу ефективність запропонованої архітектури. Найкращі результати продемонструвала модель XGBoost, для якої отримано MAE = 0.039, RMSE = 0.055 та $R^2 = 0.930$, що свідчить про високий рівень узагальнювальної здатності моделі та її придатність до використання у прикладних задачах цифрового екологічного управління.

Запропонований підхід дозволяє перекладати числові результати прогнозування екологічного ризику у цифрові управлінські рішення шляхом автоматичного віднесення територій до класів ризику, що створює основу для формування рекомендацій щодо організації логістики, моніторингу, просторового планування та екологічного реагування на муніципальному рівні.

Подальші дослідження доцільно проводити у напрямку розширення набору моделей за рахунок нейромережевих архітектур, інтеграцію потокових сенсорних даних у реальному часі, використання геоінформаційних платформ для автоматичного оновлення просторових ознак, а також розроблення цифрового двійника екологічної системи муніципального управління органічними відходами. Перспективним напрямом також є інтеграція системи з міськими інформаційними платформами для підтримки стратегічного управління сталим розвитком територій міських громад.

Література

1. P. Xu, H. Zheng, A multi-AI approach to predicting municipal solid waste generation and recycling demand in Hong Kong, *Resources, Conservation and Recycling*. 2026. Vol. 225. Art. 108590. DOI: <https://doi.org/10.1016/j.resconrec.2025.108590>

2. W. Xia, J. Chen, X. Li, Application of machine learning algorithms in municipal solid waste management: a review, *Waste Management & Research*. 2022. Vol. 40, № 1. P. 7–19. DOI: <https://doi.org/10.1177/0734242X211033716>
3. S. D. Latif, N. A. B. Hazrin, M. K. Younes, A. N. Ahmed, A. Elshafie, Evaluating different machine learning models for predicting municipal solid waste generation: A case study of Malaysia, *Environment, Development and Sustainability*. 2024. Vol. 26, № 5. P. 12875–12898. DOI: <https://doi.org/10.1007/s10668-023-03882-x>.
4. K. Ameri, M. Hempel, H. Sharif, J. Lopez, K. Perumalla, Design of a novel information system for semi-automated management of cybersecurity in industrial control systems, *ACM Transactions on Management Information Systems*. 2022. Vol. 14, № 1. P. 4–35. DOI: <https://doi.org/10.1145/3546580>.
5. A. Tryhuba, T. Hutsol, S. Glowacki, European Green Deal: Justification of the relationships between the functional indicators of bioenergy production systems using organic residential waste, *Energies*. 2024. Vol. 17, № 6. Art. 1461. DOI: <https://doi.org/10.3390/en17061461>
6. A. N. Ahmed, A. F. Ghani, N. S. A. Wahid, Comparative analysis of machine learning algorithms for municipal waste prediction, *Science of the Total Environment*. 2023. Vol. 857. Art. 159493. DOI: <https://doi.org/10.1016/j.scitotenv.2022.159493>.
7. M. S. Islam, R. Hyder, Explainable artificial intelligence in environmental decision systems: limitations and future challenges, *Environmental Modelling & Software*. 2024. Vol. 176. Art. 106008. DOI: <https://doi.org/10.1016/j.envsoft.2024.106008>.
8. H. Zheng, P. Xu, Statistical forecasting approaches for municipal waste generation under urban dynamics, *Waste Management*. 2022. Vol. 138. P. 1–10. DOI: <https://doi.org/10.1016/j.wasman.2021.11.018>.
9. B. Liu, J. Chen, Forecasting municipal solid waste generation using long short-term memory neural networks, *Environmental Earth Sciences*. 2024. Vol. 83. Art. 399. DOI: <https://doi.org/10.1007/s12665-024-11702-2>.
10. J. Kaza, L. Yao, S. Bhada-Tata, F. Van Woerden, What a Waste 2.0: A Global Snapshot of Solid Waste Management to 2050. World Bank, 2023 update. Accessed: 28.01.2026. URL: <https://www.worldbank.org/en/publication/wh-at-a-waste>
11. J. Chen, Y. Yang, Z. Feng, R. Huang, G. Zhou, H. You, X. Han, Ecological Risk Assessment and Prediction Based on Scale Optimization - A Case Study of Nanning, a Landscape Garden City in China, *Remote Sensing*. 2023. Vol. 15, № 5. Art. 1304. DOI: <https://doi.org/10.3390/rs15051304>.
12. D. Zhou, Y. Jia, N. Zhang, H. Hou, F. Wang, Z. Wang, An ensemble machine-learning model with online learning strategy for building load forecasting, *Applied Thermal Engineering*. 2026. Vol. 288, Part 1. Art. 129434. DOI: <https://doi.org/10.1016/j.applthermaleng.2025.129434>.
13. A. Tryhuba, T. Hutsol, S. Glowacki, V. Tryhuba, Prediction of biogas production volumes from household organic waste based on machine learning, *Energies*. 2024. Vol. 17, № 7. Art. 1786. DOI: <https://doi.org/10.3390/en17071786>.
14. А. М. Тригуба, О. Я. Андрушків, І. Л. Тригуба, Циркуляційно-ціннісна модель управління проектами енергозабезпечення житлових масивів, Управління розвитком складних систем. 2025. № 64. С. 138–152. DOI: <https://doi.org/10.32347/2412-9933.2025.64.138-152>
15. А. Тригуба, Р. Шолудько, О. Андрушків, Р. Олійник, М. Коциловський, Інтелектуальні моделі управління інфраструктурними проектами розвитку громад в умовах багаторівневих ризиків, Вісник Львівського державного університету безпеки життєдіяльності. 2025. № 31. С. 213–226. DOI: <https://doi.org/10.32447/20784643.31.2025.21>

References

1. P. Xu, H. Zheng, A multi-AI approach to predicting municipal solid waste generation and recycling demand in Hong Kong, in: Resources, Conservation and Recycling 225 (2026) 108590. doi: 10.1016/j.resconrec.2025.108590.
2. W. Xia, J. Chen, X. Li, Application of machine learning algorithms in municipal solid waste management: a review, in: Waste Management & Research 40(1) (2022) 7–19. doi: 10.1177/0734242X211033716.
3. S. D. Latif, N. A. B. Hazrin, M. K. Younes, A. N. Ahmed, A. Elshafie, Evaluating different machine learning models for predicting municipal solid waste generation: a

- case study of Malaysia, in: *Environment, Development and Sustainability* 26(5) (2024) 12875–12898. doi: 10.1007/s10668-023-03882-x.
4. K. Ameri, M. Hempel, H. Sharif, J. Lopez, K. Perumalla, Design of a novel information system for semi-automated management of cybersecurity in industrial control systems, in: *ACM Transactions on Management Information Systems* 14(1) (2022) 4–35. doi: 10.1145/3546580.
 5. A. Tryhuba, T. Hutsol, S. Glowacki, European Green Deal: justification of the relationships between the functional indicators of bioenergy production systems using organic residential waste, in: *Energies* 17(6) (2024) 1461. doi: 10.3390/en17061461.
 6. A. N. Ahmed, A. F. Ghani, N. S. A. Wahid, Comparative analysis of machine learning algorithms for municipal waste prediction, in: *Science of the Total Environment* 857 (2023) 159493. doi: 10.1016/j.scitotenv.2022.159493.
 7. M. S. Islam, R. Hyder, Explainable artificial intelligence in environmental decision systems: limitations and future challenges, in: *Environmental Modelling & Software* 176 (2024) 106008. doi: 10.1016/j.envsoft.2024.106008.
 8. H. Zheng, P. Xu, Statistical forecasting approaches for municipal waste generation under urban dynamics, in: *Waste Management* 138 (2022) 1–10. doi: 10.1016/j.wasman.2021.11.018.
 9. B. Liu, J. Chen, Forecasting municipal solid waste generation using long short-term memory neural networks, in: *Environmental Earth Sciences* 83 (2024) 399. doi: 10.1007/s12665-024-11702-2.
 10. J. Kaza, L. Yao, S. Bhada-Tata, F. Van Woerden, *What a Waste 2.0: A Global Snapshot of Solid Waste Management to 2050*, World Bank (2023 update). URL: <https://www.worldbank.org/en/publication/wh-at-a-waste>.
 11. J. Chen, Y. Yang, Z. Feng, R. Huang, G. Zhou, H. You, X. Han, Ecological risk assessment and prediction based on scale optimization: a case study of Nanning, a landscape garden city in China, in: *Remote Sensing* 15(5) (2023) 1304. doi: 10.3390/rs15051304.
 12. D. Zhou, Y. Jia, N. Zhang, H. Hou, F. Wang, Z. Wang, An ensemble machine-learning model with online learning strategy for building load forecasting, in: *Applied Thermal Engineering* 288 (2026) 129434. doi: 10.1016/j.applthermaleng.2025.129434.
 13. A. Tryhuba, T. Hutsol, S. Glowacki, V. Tryhuba, Prediction of biogas production volumes from household organic waste based on machine learning, in: *Energies* 17(7) (2024) 1786. doi: 10.3390/en17071786.
 14. A. M. Tryhuba, O. Ya. Andrushkiv, I. L. Tryhuba, Circular-value model of management of energy supply projects for residential areas, in: *Management of Development of Complex Systems* 64 (2025) 138–152. doi: 10.32347/2412-9933.2025.64.138-152. [in Ukrainian]
 15. A. Tryhuba, R. Sholudko, O. Andrushkiv, R. Oliinyk, M. Kotsylovskiyi, Intelligent models of management of infrastructure development projects of communities under multilevel risks, in: *Bulletin of Lviv State University of Life Safety* 31 (2025) 213–226. doi: 10.32447/20784643.31.2025.21. [in Ukrainian]

Дата першого надходження до видання:
29.01.2026

Внутрішня рецензія отримана: 06.02.2026

Зовнішня рецензія отримана: 08.02.2026

Дата прийняття статті до друку: 19.03.2026

Дата публікації: 16.04.2026

Про авторів:

¹Тригуба Анатолій Миколайович,
доктор технічних наук, професор,
завідувач кафедри інформаційних
технологій

¹Tryhuba Anatoliy,
Ph.D (doctor, technical sciences), professor,
head of department
<https://orcid.org/0000-0001-8014-5661>

²Коваль Назарій Ярославович,
доктор філософії, проректор

²Koval Nazariy,
Ph.D, deputy president
<https://orcid.org/0000-0001-7846-2924>

¹Тригуба Інна Леонтіївна,
кандидат сільськогосподарських наук,
доцент, завідувач кафедри генетики,
селекції та захисту рослин

¹Tryhuba Inna,
Ph.D (agricultural sciences),
associate professor, head of department
<https://orcid.org/0000-0002-5239-5951>

²Фірман Ігор Романович,
здобувач кафедри інформаційних
технологій та систем електронних
комунікацій

²Firman Igor,
Candidate for a scientific degree
<https://orcid.org/0009-0005-5840-9815>

¹Фамуляк Володимир Юрійович,
аспірант кафедри інформаційних
технологій

¹Famuliak Volodymyr,
Post-graduate student
<https://orcid.org/0009-0005-5347-9427>

Місце роботи авторів:

¹ Львівський національний університет
ветеринарної медицини та біотехнологій
імені Степана Гжицького

¹ Lviv National University of Veterinary
Medicine and Biotechnology named
after S. Gzhytsky
E-mail: trianamik@gmail.com,
trinle@ukr.net, vovfam@gmail.com
Сайт: <https://lvet.edu.ua/>

² Львівський державний університет
безпеки життєдіяльності

² Lviv State University of Life Safety
E-mail: kovaln870@gmail.com,
firmanigorromanovuch@gmail.com
Сайт: <https://ldubgd.edu.ua>