

УДК 519.6:504.064

DOI <https://doi.org/10.32689/maup.it.2021.2.2>

Валерія КОВАЧ

доктор наук з державного управління, старший дослідник, провідний науковий співробітник відділу цивільного захисту та інноваційної діяльності, Державна установа «Інститут геохімії навколишнього середовища Національної академії наук України», просп. Академіка Палладіна 34а, Київ, Україна, індекс 03142 (valeriiakovach@gmail.com)

ORCID: <https://orcid.org/0000-0002-1014-8979>

Володимир КУЦЕНКО

молодший науковий співробітник відділу технологій захисту довкілля та радіаційної безпеки, Державна установа «Інститут геохімії навколишнього середовища Національної академії наук України», просп. Академіка Палладіна 34а, Київ, Україна, індекс 03142 (kuts.vo@gmail.com)

ORCID: <https://orcid.org/0000-0002-0577-2056>

Ірина МАРТИНЮК

молодший науковий співробітник відділу технологій захисту довкілля та радіаційної безпеки, Державна установа «Інститут геохімії навколишнього середовища Національної академії наук України», просп. Академіка Палладіна 34а, Київ, Україна, індекс 03142 (mira_27@ukr.net)

ORCID: <https://orcid.org/0000-0003-4654-7071>

Олександр КОВАЛЕНКО

молодший науковий співробітник відділу технологій захисту довкілля та радіаційної безпеки, Державна установа «Інститут геохімії навколишнього середовища Національної академії наук України», просп. Академіка Палладіна 34а, Київ, Україна, індекс 03142 (whitewavex@gmail.com)

ORCID: <https://orcid.org/0000-0003-4798-7722>

Valeriia KOVACH

Doctor of Science in Public Administration, Senior researcher, Leading Researcher of the Department of Civil Protection and Innovation, State Institution "The Institute of Environmental Geochemistry of National Academy of Sciences of Ukraine", 34a Palladin Ave., Kyiv, Ukraine, postal code 03142 (valeriiakovach@gmail.com)

Volodymyr KUTSENKO

Junior Researcher of the Department of environmental protection technologies and radiation safety, State Institution "The Institute of Environmental Geochemistry of National Academy of Sciences of Ukraine", 34a Palladin Ave., Kyiv, Ukraine, postal code 03142 (kuts.vo@gmail.com)

Ірина MARTYNIUK

Junior Researcher of the Department of environmental protection technologies and radiation safety, State Institution "The Institute of Environmental Geochemistry of National Academy of Sciences of Ukraine", 34a Palladin Ave., Kyiv, Ukraine, postal code 03142 (mira_27@ukr.net)

Oleksandr KOVALENKO

Junior Researcher of the Department of environmental protection technologies and radiation safety, State Institution "The Institute of Environmental Geochemistry of National Academy of Sciences of Ukraine", 34a Palladin Ave., Kyiv, Ukraine, postal code 03142 (whitewavex@gmail.com)

Бібліографічний опис статті: Ковач В., Куценко В., Мартинюк І., Коваленко О. Особливості використання методів інтелектуального аналізу даних для вирішення задач екологічної безпеки атмосферного повітря. *Інформаційні технології та суспільство*. 2021. Вип. 2. С. 15–25. DOI: <https://doi.org/10.32689/maup.it.2021.2.2>

Bibliographic description of the article: Kovach, V., Kutsenko, V., Martyniuk, I., Kovalenko, O. (2021). Osoblyvosti vykorystannia metodiv intelektualnoho analizu danykh dlia vyrishennia zadach ekolohichnoi bezpeky atmosferneho povitria [Peculiarities of using methods of intellectual data analysis to solve the problems

of environmental safety of atmospheric air]. *Informatsiini tekhnolohii ta suspilstvo – Information technology and society*, 2, 15–25. DOI: <https://doi.org/10.32689/maup.it.2021.2.2>

ОСОБЛИВОСТІ ВИКОРИСТАННЯ МЕТОДІВ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ ДЛЯ ВИРІШЕННЯ ЗАДАЧ ЕКОЛОГІЧНОЇ БЕЗПЕКИ АТМОСФЕРНОГО ПОВІТРЯ

Анотація. Описано процес управління екологічною безпекою в контексті зменшення негативних впливів на навколишнє середовище. Запропоновано узагальнену структурну модель процесу управління екологічною безпекою, засновану на методах та технологіях інтелектуального аналізу даних моніторингу. Досліджено можливості адаптації та удосконалення ряду найбільш відомих алгоритмів інтелектуального аналізу даних, таких як C4.5, K-means, методу опорних векторів (SVM), kNN, наївного байєсового класифікатора, алгоритму Apriori, для задач аналізу даних мережі моніторингу атмосферного повітря. На прикладі даних щодо концентрацій забруднюючих речовин в атмосферному повітрі міста Кривий Ріг (Дніпропетровська область) побудовано діаграми: розсіювання концентрацій пилу та діоксиду азоту; розсіювання температури повітря і концентрації сірчаного ангідриду. Наведено приклади практичного використання окремих методів з метою виявлення небезпечних ситуацій. **Метою** статті є аналіз принципів та методів управління екологічною безпекою на основі інтелектуального аналізу даних мережі моніторингу атмосферного повітря. **Наукова новизна.** Запропоновано концептуальну модель управління екологічною безпекою урбанізованих територій за даними екологічного моніторингу, що відрізняється від своїх аналогів новими можливостями, базованими на використанні сучасних інформаційних технологій інтелектуального аналізу даних. У **висновках** наголошено, що авторами публікації визначено перспективи застосування методів та засобів інтелектуального аналізу даних щодо інформаційної підтримки прийняття рішень, спрямованих на оцінювання наслідків техногенного впливу і зменшення обсягу навантажень на довкілля. Запропоновано ряд методів та алгоритмів, які дають можливість оцінювати значення невідомих характеристик і параметрів за відомими даними, та показано приклади їх використання.

Ключові слова: екологічна безпека, управління, інтелектуальний аналіз даних, атмосферне повітря.

PECULIARITIES OF USING METHODS OF INTELLECTUAL DATA ANALYSIS TO SOLVE THE PROBLEMS OF ENVIRONMENTAL SAFETY OF ATMOSPHERIC AIR

Abstract. The process of environmental safety management in the context of reducing negative impacts on the environment is described. A generalized structural model of the environmental safety management process based on methods and technologies of intellectual analysis of monitoring data is proposed. Possibilities of adaptation and improvement of a number of the most known algorithms of data mining, such as C4.5, K-means, method of reference vectors (SVM), kNN, naive Bayesian classifier, Apriori algorithm, for data analysis of atmospheric air monitoring network are investigated. On the example of data on concentrations of pollutants in the air of Kryvyi Rih city (Dnipropetrovsk region) diagrams are constructed: scattering of concentrations of dust and nitrogen dioxide; dissipation of air temperature and concentration of sulfur dioxide. Examples of practical use of separate methods for the purpose of dangerous situations detection are resulted. **The aim** of the article is to analyze the principles and methods of environmental safety management based on the intellectual analysis of atmospheric air monitoring network data. **Scientific novelty.** A conceptual model of ecological safety management of urban areas based on ecological monitoring data is proposed, which differs from its analogues by new possibilities based on the use of modern information technologies of data mining. **The conclusions** emphasize that the authors of the publication identify prospects for the application of methods and tools of data mining for information support of decision-making aimed at assessing the effects of man-made impact and reducing the burden on the environment. A number of methods and algorithms are proposed, which make it possible to estimate the values of unknown characteristics and parameters according to known data, and examples of their use are shown.

Key words: ecological safety, management, data mining, atmospheric air.

Актуальність проблеми. Інтенсивний розвиток науково-технічного прогресу в останні десятиріччя супроводжується збільшенням техногенного впливу на атмосферу, гідросферу та літосферу, зростанням обсягів відходів виробничої діяльності й транспортних засобів. Серед глобальних екологічних проблем найбільш пріоритетними є забруднення повітря, дефіцит прісної води, зменшення видового й ландшафтного різноманіття планети, парниковий ефект, озонові діри, кислотні дощі, масові захворювання людей, загибель лісів тощо. Зменшення рівня антропогенного впливу на біосферу можна досягти якісним управлінням екологічною безпекою на всіх рівнях, забезпечивши стратегічну орієнтацію на принципи стійкого (гармонійного) розвитку.

На даний час існує багато визначень і трактувань щодо поняття «екологічної безпеки». Згідно ст. 50 Закону України «Про охорону навколишнього природного середовища», термін «екологічна безпека» визначає такий стан навколишнього природного середовища, у якому реалізовано реальні можливості для попередження погіршення екологічної обстановки та виникнення небезпеки для здоров'я людей. Як зазначено в Енциклопедії Сучасної України [esu.com.ua], екологічну безпеку можна розглядати як рівень захищеності життєво важливих інтересів людини, а також суспільства, довкілля та держави від реальних або потенційних загроз, зумовлених антропогенними або природними чинниками.

Головним критерієм для оцінювання екологічної безпеки певної екосистеми є якість життя і здоров'я населення. Отже, виникає необхідність цілеспрямованого управління екологічною системою з

метою підвищення її організованості та досягнення певного ефекту щодо зменшення негативного впливу на населення.

Аналіз останніх досліджень і публікацій. За визначенням американських учених Д. Кліланда і В. Кінга управління можна розглядати як процес, орієнтований на досягнення певних цілей [1]. Для вирішення проблеми необхідно змінити існуючий стан речей (або подій), щоб досягти своєї мети. При всьому різноманітті їх форм вплив можна поділити на два протилежні класи: впливи, що призводять до деградації, руйнування екосистеми, зменшення ступеня її стабільності, та впливи, які сприятимуть розвитку екосистеми, збільшенню ступеня її організованості.

Отже, процес управління екологічною безпекою на будь-якому рівні включає певну сукупність послідовних дій, а саме:

- збір інформації;
- передача інформації в пункти збереження та її обробка;
- аналіз збереженої та довідкової інформації;
- прийняття рішення на основі проведеного аналізу;
- створення відповідного керуючого впливу;
- доведення цього впливу до об'єкта управління;
- перевірка нового стану об'єкта управління.

Управління екологічною безпекою та раціональним використанням природних ресурсів передбачає аналіз потреб людини в природних ресурсах і з'ясування можливостей природи щодо задоволення цих потреб. Однак необхідно не тільки виявити оптимальне поєднання потреб людини та можливостей природи, але й забезпечити оптимальні умови для переходу від нинішнього далеко не ідеального стану до більш ефективного варіанту такого поєднання.

На підставі власного досвіду роботи в галузі екологічної безпеки та на основі аналізу наукових джерел [2; 3, с. 85–102; 4, с. 96–115; 5] визначено, що основною метою досліджень в даному напрямі є розробка і впровадження сучасних механізмів управління екологічною безпекою та станом навколишнього середовища. Під словами «механізми управління» будемо розуміти сукупність певних методів та засобів управління екологічною безпекою й природокористуванням. У свою чергу, метод управління визначає набір способів, прийомів, засобів впливу на керований об'єкт.

Метою статті є аналіз принципів та методів управління екологічною безпекою на основі інтелектуального аналізу даних мережі моніторингу атмосферного повітря.

Результати дослідження. У нашій країні прийняття управлінських рішень та реалізація конкретних регулюючих заходів, які необхідно здійснити щодо збереження навколишнього середовища, відбуваються із затримкою. Час затримки $\tau_{зат}$ може бути значним і залежить від багатьох факторів. За цей час система може вийти за межі області допустимих станів, які відповідають прийнятному рівню якості навколишнього середовища, тобто нормальним умовам для проживання та діяльності людей на досліджуваній території. Більше того, після досягнення $\tau_{зат}$ певного критичного значення екологічна система може перейти в область критичного відхилення, де незначні зміни призведуть до незворотних процесів деградації.

Представимо об'єкт управління екологічною безпекою і станом оточуючого середовища у вигляді двох взаємодіючих підсистем:

- 1) природно-територіальний комплекс, тобто компоненти, на які впливають об'єкти промисловості й транспорту на досліджуваній території, де поширюються та накопичуються забруднюючі речовини;
- 2) безпосередньо промислові об'єкти і транспорт (ПОТ), які функціонують на досліджуваній території і належать підсистемі стаціонарних джерел забруднення та підсистемі пересувних джерел забруднення.

На рис. 1 наведено узагальнену структурну модель об'єкта управління екологічною безпекою [3, с. 85–102], побудовану з урахуванням підходу, викладеного в роботі [4, с. 96–115]. На схемі позначено окремі інформаційні потоки, які визначають взаємодію між параметрами: X – вектор стану (якості) компонентів природного середовища (концентрації забруднень, рівень енергетичних впливів тощо), тобто це контрольовані показники якості компонентів природного середовища; Z – параметри ПОТ, які визначають умови і результат функціонування окремих об'єктів як з економічного (обсяги, якість продукції та послуг), так і з екологічного погляду (потужність негативного впливу на природне середовище). Для даної моделі $Z = \{Z_c, Z_n\}$, де Z_c, Z_n – множина станів відповідно стаціонарних та пересувних об'єктів ПОТ. При цьому x_c, x_n – множина впливів природного середовища на підсистему ПОТ ($x_c, x_n \subset X$); z'_c, z'_n – множина результатів економічної діяльності об'єктів ПОТ ($z'_c \subset Z_c, z'_n \subset Z_n$); $\omega_{пс}$ – множина зовнішніх впливів на компоненти природного середовища (ПС). У такому вигляді мож-

на врахувати природні та кліматичні параметри, особливості інфраструктури регіону тощо. Враховано також додаткові зовнішні впливи, пов'язані з ПОТ, а саме: $\omega_{\text{ПОТ}}$ – множина зовнішніх впливів на ПОТ, які впливають на результати їх діяльності, виробничі й транспортні параметри, що визначають рівень екологічної небезпеки; $\omega_{\text{ЗС}}$ – множина впливів на зовнішнє середовище через компоненти природного середовища, що надходять від ПОТ; $U_c, U_{\text{П}}$ – множина керуючих впливів відповідно на стаціонарні та пересувні ПОТ.



Рис. 1. Узагальнена модель управління екологічною безпекою територій, що зазнають впливу промислових об'єктів і транспорту

Особам, що приймають рішення (ОПР) щодо виникнення в момент t несприятливої екологічної ситуації, необхідно завчасно надати інформацію для розробки керуючих впливів $U_c(t)$ або $U_p(t)$, за допомогою яких можна контролювати та зменшувати негативний вплив на компоненти природного середовища: $\Delta X(t) < \varepsilon$, $\varepsilon \rightarrow 0$.

Розглянемо детальніше поняття стратегічного і тактичного управління екологічною безпекою, оскільки вони мають вирішальне значення в процесах управління. Стратегічне і тактичне управління екологічною безпекою мають різні цілі та засоби їх досягнення. Якщо стратегічне управління являє собою пошук і реалізацію нових можливостей при змінах зовнішнього середовища, то тактичне управління є процесом створення передумов для реалізації певних нових можливостей.

Тактичне управління можна визначити як засіб реалізації та деталізації відповідної стратегії. Воно розраховано на короткостроковий та середньостроковий періоди. Головною метою тактичного управління екологічною безпекою є вплив на життєдіяльність досліджуваної екосистеми, спрямований на досягнення стратегічних цілей та завдань при найбільш повному і найбільш раціональному використанні доступних ресурсів. Зрештою, засобом реалізації стратегічних планів є тактичне планування. Якщо основна мета стратегічного плану полягає у визначенні щодо перспективи на майбутнє, то тактичне планування допомагає знайти конкретні відповіді на питання, як поетапно досягти бажаного стану.

На рис. 2 наведено концептуальну схему системи управління екологічною безпекою урбанізованої території за даними екологічного моніторингу з використанням сучасних методів і технологій інтелектуального аналізу даних [3, с. 85–102]. Дано пояснення основних блоків, зазначених на схемі.

Блок «Бази даних та алгоритми оцінювання ситуації і ухвалення управлінських рішень в галузі екологічної безпеки» складається з блоків «Управління даними» та «Система підготовки рішень (на основі економічної оцінки ризик-ціна-ефект)». У свою чергу, блок «Система підготовки рішень» включає «Інтелектуальний аналіз даних» і «Систему підтримки прийняття рішень (СППР)».

Блок «Управління даними» забезпечує збереження даних і доступ до них. До таких даних відносять дані моніторингу, розподілені бази даних (БД) аналітичної та звітної інформації та розподілені БД дистанційного зондування Землі (ДЗЗ), які перебувають у сховищі даних. Під розподіленими БД розуміємо сукупність взаємопов'язаних на логічному рівні БД, розподілених у комп'ютерній мережі. Інтелекту-

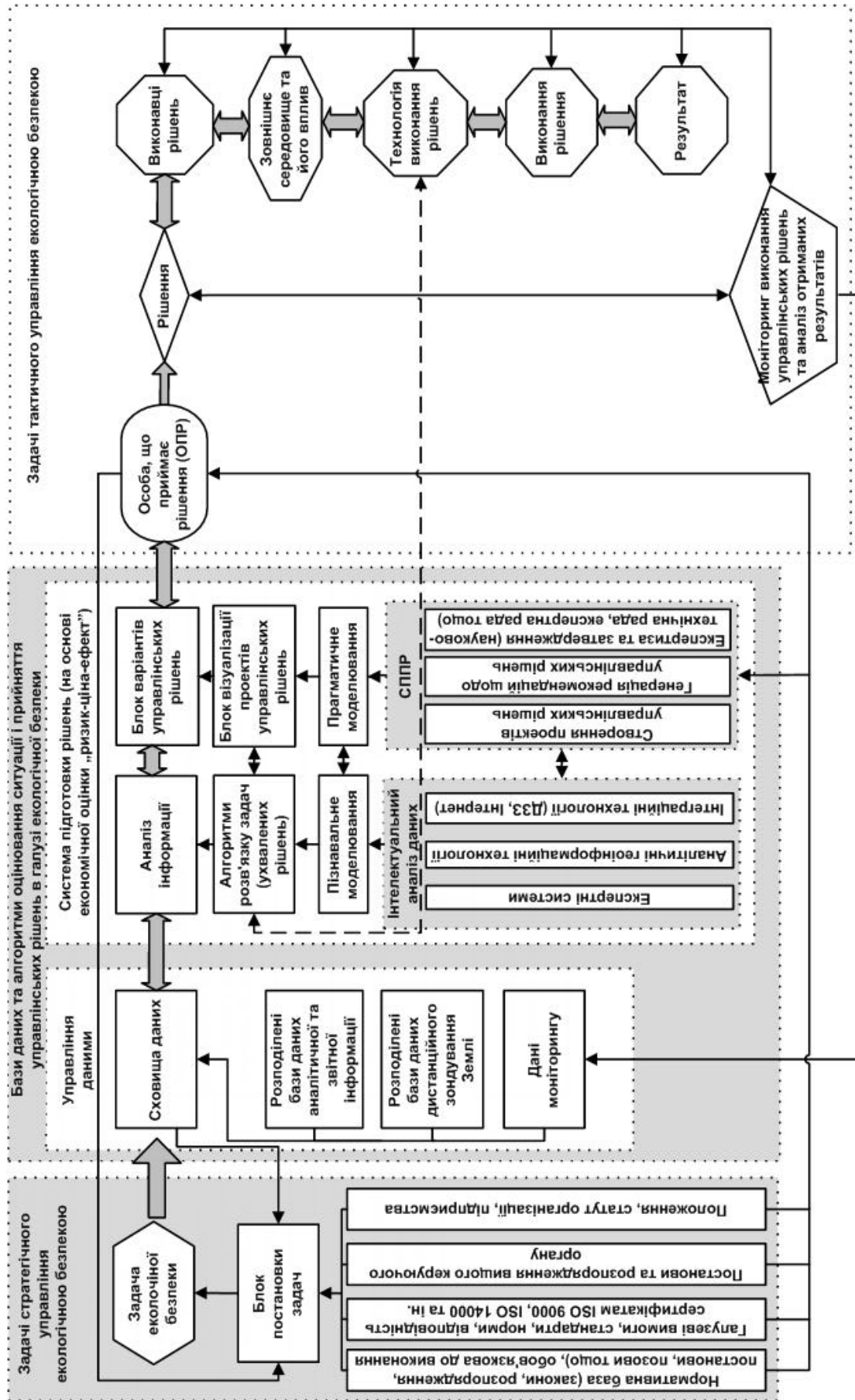


Рис. 2. Концептуальна схема системи управління екологічною безпекою за даними екологічного моніторингу

альний аналіз даних у наведеній схемі включає експертні системи, геоінформаційні та інтеграційні технології (ДЗЗ, Інтернет).

В цілому інтелектуальний аналіз даних у задачах екологічної безпеки можна визначити як метод підтримки прийняття рішень, заснований на дослідженні певних закономірностей щодо вихідних даних. Він включає методики та засоби, які на основі будь-яких моделей, алгоритмів або математичних теорем дають можливість оцінювати значення невідомих характеристик і параметрів за відомими даними.

Алгоритми інтелектуального аналізу. На основі публікацій [6, с. 1–21; 7, с. 3558–3571; 8, с. 1–15; 9; 10; 11, с. 62–75; 12; 13, с. 139–151] визначено найпоширеніші алгоритми інтелектуального аналізу даних, спрямовані на виявлення прихованих закономірностей або взаємозв'язків між змінними у великих масивах необроблених даних.

Класифікатор C4.5 створює класифікацію у вигляді дерева рішень. Для цього задається набір даних, який являє собою вже класифіковані речі. Алгоритм C4.5, розроблений J.R. Quinlan [10], є вдосконаленою версією алгоритму ID3 того ж автора. Зокрема, в C4.5 додано відсікання гілок, можливість роботи з числовими атрибутами, а також можливість побудови дерева з неповною навчальною вибіркою, в якій відсутні значення деяких атрибутів.

Для того щоб за допомогою C4.5 побудувати дерево рішень і застосовувати його, дані повинні задовольняти декільком умовам, а саме:

- інформація про об'єкти, що необхідно класифікувати, повинна бути представлена у вигляді кінцевого набору ознак (атрибутів), кожен з яких має дискретне або числове значення. Якщо такий набір атрибутів назвати прикладом, тоді для всіх прикладів кількість атрибутів і їх склад повинні бути постійними;
- кожен із класів, на які будуть розбиватися приклади, повинен мати кінцеве число елементів, а кожен приклад повинен однозначно відноситись до конкретного класу;
- для випадків з нечіткою логікою, тобто коли приклади можна віднести до одного з класів з певною ймовірністю, C4.5 непридатний;
- у навчальній вибірці кількість прикладів має бути значно більше кількості класів і кожен приклад повинен бути заздалегідь асоційований зі своїм класом. За цими ознаками C4.5 віднесено до методів машинного навчання з учителем.

Алгоритм кластерного аналізу K-means виділяє певну кількість груп k з набору об'єктів таким чином, щоб члени кожної групи були максимально схожими між собою. Це досить популярний метод кластерного аналізу для вивчення набору даних. Нагадаємо, що кластерний аналіз – це сімейство алгоритмів, призначених для формування груп, де в кожній групі члени даної групи мають більше схожих рис з представниками своєї групи, ніж з тими, хто в цій групі не перебуває (в даному контексті кластери та групи є синонімами) [3, с. 85–102].

Метод K-means – це метод, метою якого є поділ m спостережень (з простору даних) на k кластерів таким чином, що кожне спостереження буде віднесено до того кластеру, до центру (центроїду) якого воно буде найближчим. Як міра близькості використовується Евклідова відстань:

$$\|x - y\| = \sqrt{\sum_{p=1}^n (x_p - y_p)^2},$$

де $x, y \in R^n$.

Для ряду спостережень (x^1, x^2, \dots, x^m) , $x^j \in R^n$, метод K-means мінімізує сумарне квадратичне відхилення точок кластерів від центроїдів цих кластерів:

$$\min \left[\sum_{i=1}^k \sum_{x^{(j)} \in S_i} \|x^{(j)} - \mu_i\|^2 \right],$$

де $x^{(j)} \in R^n$, $\mu_i \in R^n$, μ_i – центр ваги для кластера S_i .

Зауважимо, що при використанні даного методу неправильний вибір початкового числа кластерів k може призвести до некоректних результатів. Саме тому при використанні методу K-means на початковому етапі необхідно визначити відповідне число кластерів для визначеного набору даних.

Метод опорних векторів (SVM) знаходить гіперплощину для класифікації даних на два класи. На відміну від C4.5 в ньому не використовуються дерева рішень.

SVM – це лінійний алгоритм, який використовується в задачах класифікації та регресії. Він широко застосовується на практиці і за його допомогою можна вирішувати як лінійні, так і нелінійні задачі. Суть цього методу полягає у створенні лінії або гіперплощини, яка розділяє дані на класи. Якщо є два

класи спостережень і передбачається лінійна форма кордону між класами, то можливі два випадки.

Перший з них пов'язаний з можливістю ідеального поділу даних за допомогою деякої гіперплощини $z_k(x) = \sum_{i=1}^p \beta_i x_i + \beta_0$.

На рис. 3, а наведено двовимірний варіант. Оскільки таких гіперплощин може бути безліч, оптимальною буде та з них, яка максимально віддалена від навчальних точок, тобто має максимальний проміжок (зазор) M [12].

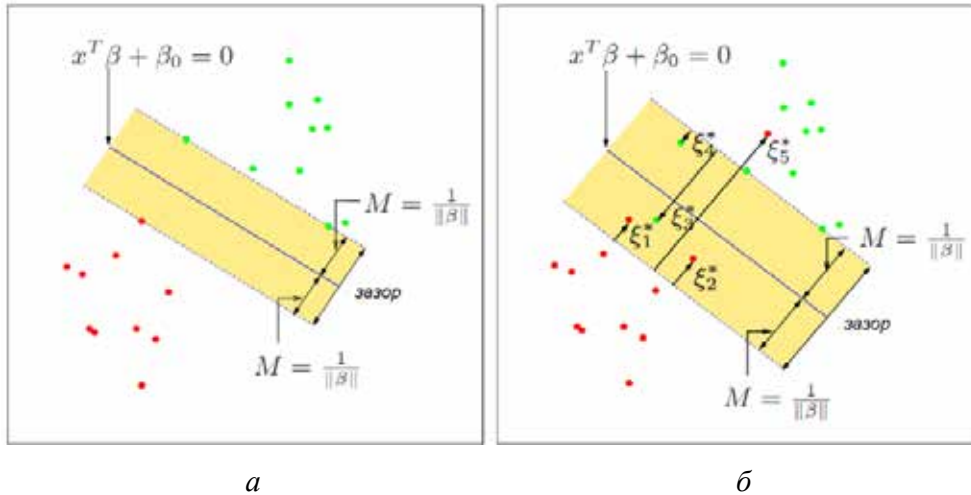


Рис. 3. Класифікатори з мінімальним зазором (а) і на опорних векторах (б)

Другий випадок показано на рис. 3, б, коли множина точок перекривається і обидва класи лінійно нероздільні. Власне опорними векторами називаються значення, які лежать безпосередньо на кордоні, що розділяє смуги, або на неправильній для свого класу стороні щодо кордонів зазору (такі точки позначені ξ_j^*). Для граничних і всіх інших точок прийнято $\xi_j^* = 0$.

Оптимальну розділяючу гіперплощину такого класифікатора $z_k(x)$ можна визначити за умови максимізації ширини зазору M , але при цьому дозволено невірно класифікувати деяку невелику групу спостережень, що відносяться до опорних векторів. Математично пошук рішення зведено до задачі квадратичної оптимізації з лінійними обмеженнями, яка гарантовано сходиться до одного глобального мінімуму.

Алгоритм класифікації kNN (k найближчих сусідів) відрізняється від раніше описаних тим, що він нагадує «лінивого учня». В основі методу kNN-класифікатора лежить гіпотеза компактності, яка передбачає, що тестований об'єкт d матиме таку ж мітку класу, як і навчальні об'єкти в локальній області його найближчого оточення. У варіанті kNN кожен об'єкт відносимо до пріоритетного класу найближчих сусідів, де k – параметр алгоритму.

Вирішальні правила в методі kNN визначаються межами суміжних сегментів діаграми Вороного, що розділяє площину на n опуклих багатокутників, кожен з яких містить один і тільки один об'єкт навчальної вибірки (рис. 4). В p -мірних просторах границі розв'язків складаються вже з сегментів $(p-1)$ -мірних напівплощин, утворених опуклими многогранниками Вороного [12].

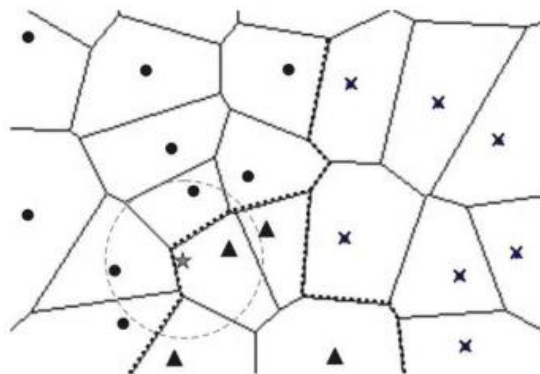


Рис. 4. Приклад роботи алгоритму класифікації найближчих сусідів

Алгоритм передбачення будується за принципом «більшості голосів», тобто за результатами голосування вибирається мітка класу-переможця. На рис. 4 тестований об'єкт «зірочка» потрапляє в осередок об'єктів класу «трикутник» і при $k = 1$ буде віднесений до цього класу. Однак при $k = 3$ за «голосами» двох найближчих сусідів з трьох екзаменованих цей об'єкт буде віднесено до класу «кружечків». Імовірнісний варіант методу kNN використовує для ранжирування передбачуваних класів суму голосів сусідів з урахуванням їх ваг, зокрема, евклідової міри відстані між тестованим об'єктом і кожним із сусідів.

Варіант 1NN завжди забезпечує 100% правильного розпізнавання прикладів навчальної вибірки (самий найближчий сусід – він сам), проте часто помиляється на невідомих йому даних. При збільшенні k від одиниці до деяких меж якість розпізнавання на контрольній вибірці буде зростати. Оптимальне щодо точності прогнозів значення k може бути знайдено з використанням перехресної перевірки. Для цього по фіксованому значенню k будується модель k найближчих сусідів і оцінюється CV-помилка класифікації. Ці дії повторюються для різних значень k і значення з найменшою помилкою розпізнавання приймається як оптимальне.

Наївний байєсів класифікатор включає сімейство алгоритмів класифікації, які поділяють одне спільне припущення. В основі байєсівської класифікації лежить гіпотеза максимальної ймовірності, тобто об'єкт d належить класу c_j ($c_j \in C$), і при цьому досягається найбільша апостеріорна ймовірність $\max P(c_j|d)$. За формулою Байєса:

$$P(c_j|d) = \frac{P(c_j)P(d|c_j)}{P(d)} \approx P(c_j)P(d|c_j),$$

де $P(d|c_j)$ – ймовірність зустріти об'єкт d серед об'єктів класу c_j ; $P(c_j)$ і $P(d)$ – апіорні ймовірності класу c_j і об'єкта d (остання не впливає на вибір класу і її можна не враховувати).

Якщо зробити «наївне» припущення, що всі ознаки, за якими класифікуються об'єкти, абсолютно рівноправні між собою і не пов'язані одна з одною, то $P(d|c_j)$ можна обчислити як добуток ймовірностей зустріти ознаку x_i ($x_i \in X$) серед об'єктів класу c_j :

$$P(d|c_j) = \prod_{i=1}^{|X|} P(x_i|c_j),$$

де $P(x_i|c_j)$ – ймовірнісна оцінка вкладу ознаки x_i в те, що $d \in c_j$.

На практиці при перемноженні дуже малих умовних ймовірностей може спостерігатися втрата значущих розрядів. У зв'язку з цим замість оцінок ймовірностей $P(x_i|c_j)$ застосовують логарифми цих ймовірностей. Оскільки логарифм є монотонно зростаючою функцією, то клас c_j з найбільшим значенням логарифма ймовірності залишиться найбільш ймовірним. Тоді вирішальне правило наївного байєсового класифікатора (Naive Bayes Classifier) приймає наступний остаточний вигляд [12]:

$$c^* = \arg_{c_j \in C} \max \left[\log P(c_j) + \sum_{i=1}^X P(x_i|c_j) \right].$$

Алгоритм Apriori належить до масштабованих алгоритмів, спрямованих на пошук асоціативних правил в БД. Сучасні БД мають дуже великі розміри, а також тенденцію щодо подальшого їх зростання. Пошук асоціативних правил в таких БД спрямований на отримання даних для вивчення кореляцій і взаємозв'язків між змінними. Алгоритм Apriori реалізує такий пошук в два етапи. На першому етапі здійснюється пошук вибірок, які часто зустрічаються в даній базі, а на другому – виявляються асоціативні правила, що є достовірними для цих вибірок.

Приклади інтелектуального аналізу даних. Інтелектуальний аналіз даних моніторингу стану атмосферного повітря (АП) урбанізованих територій проведено в рамках інформаційно-аналітичної системи еколого-енергетичного моніторингу AISEEM [3, с. 85–102] та програми Orange [14], яка є сучасним інструментом для візуалізації даних, машинного навчання та інтелектуального аналізу даних. Взаємодія між цими програмами здійснювалась за допомогою інтеграції методів та засобів Orange в систему AISEEM.

Дані щодо забруднюючих речовин завантажувались з автоматизованої системи моніторингу за станом АП м. Кривий Ріг (Дніпропетровська область). Основною метою системи моніторингу є безперервний вимір концентрацій забруднюючих речовин і метеорологічних параметрів АП жилої зони м. Кривий Ріг [15].

На рис. 5 темно-зеленим кольором позначені міські автоматизовані пости, темно-синім – пости підприємства ПАТ «Південний гірничо-збагачувальний комбінат», темно-червоним – ПРАТ «СУХА БАЛКА» та помаранчевим – ПАТ «Кривбасзалізрудком». Спостереження здійснено за шістьма забруднюючими речовинами: діоксид азоту (NO₂), оксид азоту (NO), сірчаний ангідрид (SO₂), оксид вуглецю (CO), аміак (NH₃), сірководень (H₂S), пил. Крім того, на міських автоматизованих постах спостережень (ПАС) № 3

та № 5 здійснено спостереження за етаном (C_2H_6) та озоном (O_3), а метеорологічними параметрами є середня швидкість вітру, напрямок вітру, температура, відносна вологість та тиск.

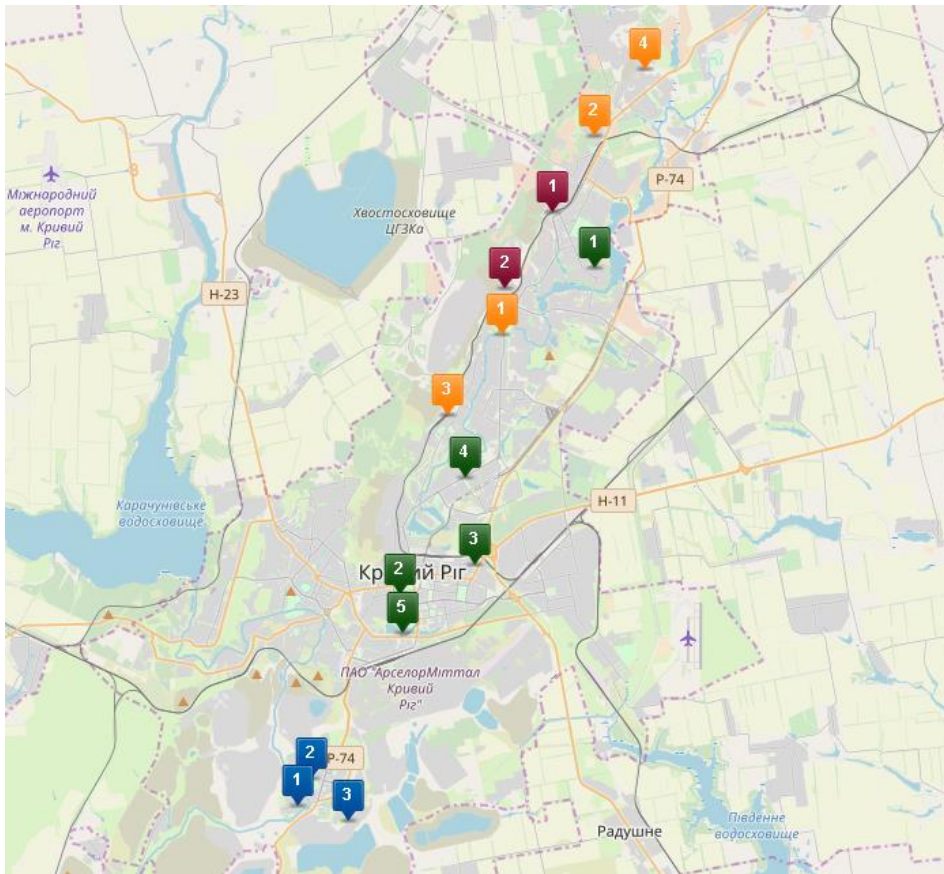


Рис. 5. Автоматизовані пости спостереження за станом АП

Варто також зазначити, що на території міста виробництво теплової енергії здійснює ПАТ «Криворізька теплоцентраль», на балансі якої є шість опалювальних котельнь, обладнаних 31 паровими та водогрійними котлами з установленою та наявною потужностями відповідно 1406 Гкал/год та 1081 Гкал/год. Найшкідливіші викиди цього енергетичного об'єкта – це сірчаний ангідрид, оксиди азоту та пил.

При аналізі даних було виключено інформацію щодо напрямку вітру, оскільки напрямок вітру змінюється від 0° до 360° , і для 358° та 2° фактично отримуємо північний напрямок вітру згідно з даними [15] з зовсім різними значеннями, що ускладнює їх подальше трактування.

Наведемо приклади інтелектуального аналізу даних для погодинних максимальних разових концентрацій забруднюючих речовин та метеорологічних даних на міському ПАС № 1 з 10 по 16 грудня 2018 року.

На рис. 6 зображено діаграму розсіювання, де на осі абсцис показано рівні концентрацій пилу, а на осі ординат – діоксиду азоту. Як бачимо, концентрації пилу на цьому часовому інтервалі змінювались від 0,156 до 0,284, а концентрації діоксиду азоту – від 0,027 до 0,091. На рис. 7 на осі абсцис показано температуру повітря ($^\circ C$), а на осі ординат – концентрації сірчаного ангідриду. Температура на цьому часовому інтервалі змінювалась від $-4,1^\circ C$ до $+4,6^\circ C$, а концентрації сірчаного ангідриду – від 0,018 до 0,021 mg/m^3 .

Підкреслимо, що на даному етапі дослідження наведено лише попередні результати аналізу даних моніторингу. Адже застосування інтелектуальних технологій для виявлення нових знань та закономірностей потребує значно більшої кількості даних щодо забруднення атмосфери, викидів транспорту та промислових підприємств, захворюваності населення тощо. Маємо надію, що такі дані будуть доступні після реалізації відповідних загальнодержавних програм, що дасть можливість здійснювати більш ґрунтовний аналіз даних моніторингу стану атмосферного повітря.

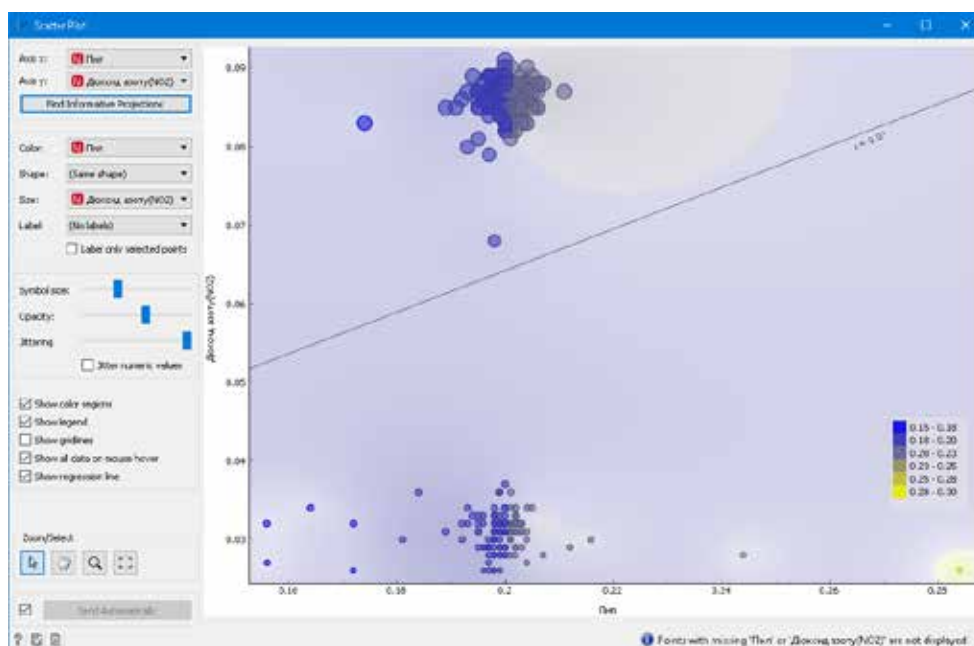


Рис. 6. Діаграма розсіювання концентрацій пилу та діоксиду азоту, (мг/м³)

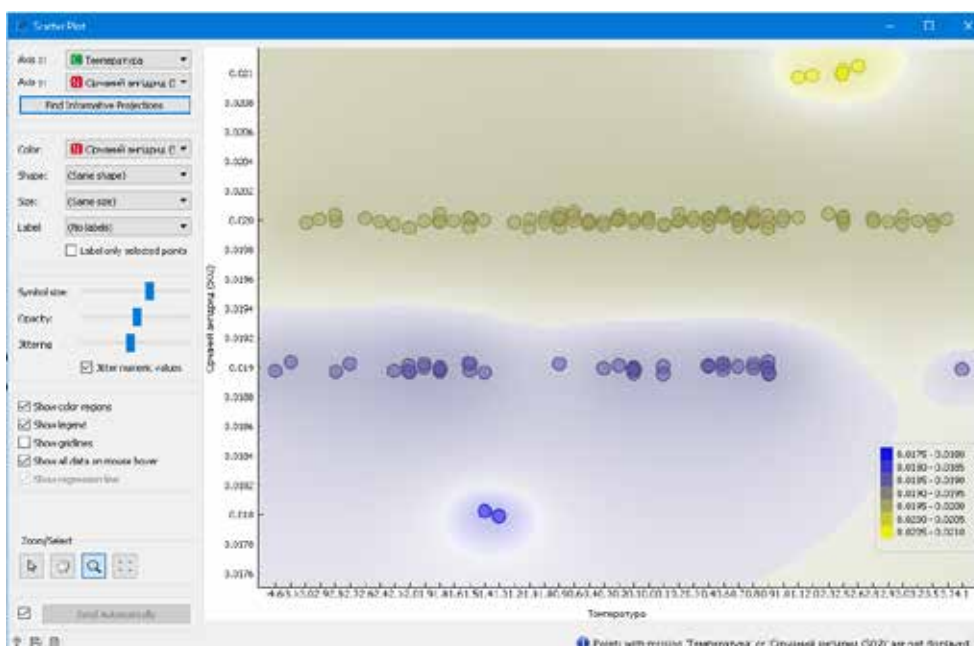


Рис. 7. Діаграма розсіювання температури повітря (°C) і концентрації сірчаного ангідриду

Висновки. Запропонована концептуальна модель управління екологічною безпекою урбанізованих територій за даними екологічного моніторингу, що відрізняється від своїх аналогів новими можливостями, базованими на використанні сучасних інформаційних технологій інтелектуального аналізу даних. Визначено перспективи застосування методів та засобів інтелектуального аналізу даних щодо інформаційної підтримки прийняття рішень, спрямованих на оцінювання наслідків техногенного впливу і зменшення обсягу навантажень на довкілля. Запропоновано ряд методів та алгоритмів, які дають можливість оцінювати значення невідомих характеристик і параметрів за відомими даними, та показано приклади їх використання.

Список використаних джерел:

1. Клиланд Д., Кинг В. Системный анализ и целевое управление. Москва : Сов. радио, 1974. 280 с.
2. Пампура В.И. Оптимальное управление безопасностью экологически опасных объектов. К. : Наукова думка, 2012. 599 с.
3. Яцишин А.В., Куцан Ю.Г., Артемчук В.О. та ін. Принципи та методи управління екологічною безпекою на основі інтелектуального аналізу даних мережі моніторингу атмосферного повітря. *Електронне моделювання*. 2019. № 4(41). С. 85–102.
4. Ивашук О.А., Константинов И.С. Обеспечение адаптивного управления экологической безопасностью промышленно-транспортного комплекса. *Управление большими системами*. 2009. Вып. 25. С. 96–115.
5. Капралов Е.Г., Кошкарев А.В., Тикунов В.С. Геоинформатика : учебник для вузов. М. : Академия, 2008. 384 с.
6. Chen C., Chuang C., Jiang J. Ecological Monitoring Using Wireless Sensor Networks. Overview, Challenges, and Opportunities. *Advancement in Sensing Technology. Smart Sensors, Measurement and Instrumentation*. 2013. Vol. 1. P. 1–21.
7. Dias D., Tchepel O. Modelling of Human Exposure to Air Pollution in the Urban Environment: a GPS-based approach. *Environmental Science and Pollution Research*. 2014. Vol. 5. P. 3558–3571.
8. Peters D.P.C., Havstad K.M., Cushing J. et al. Harnessing the Power of Big Data: Infusing the Scientific Method With Machine Learning to Transform Ecology. *Ecosphere*. 2014. Vol. 5(6). P. 1–15.
9. Алгоритмы интеллектуального анализа данных. 2015. URL: <https://tproger.ru/translations/top-10-data-mining-algorithms/>.
10. Деревья решений – C4.5 математический аппарат. Часть 1. 2019. URL: <https://basegroup.ru/community/articles/math-c45-part1>.
11. Згуровский М.З., Болдак А.А., Ефремов К.В. Интеллектуальный анализ и системное согласование научных данных в междисциплинарных исследованиях. *Кибернетика и системный анализ*. 2013. № 4. С. 62–75.
12. Шитиков В.К., Мاستицкий С.Э. Классификация, регрессия, алгоритмы Data Mining с использованием R. 2017. URL: <https://github.com/ranalytics/data-mining>.
13. Ashouri M., Haghghat F., Fung B.C.M. et al. Development of Building Energy Saving Advisory: A Data Mining Approach. *Energy and Buildings*. 2018. Vol. 172. P. 139–151.
14. Orange. 2021. URL: <http://orange.biolab.si/download/>.
15. Кривий Ріг. Автоматизовані пости спостереження. 2018. URL: <https://krmisto.gov.ua/ua/rc/ecomon.html>.

References:

1. Cleland, D., King, V. (1974). *Sistemnyy analiz i tselevoye upravleniye [System Analysis and Target Management]*. Moscow: Sov. Radio [in Russian].
2. Pampuro, V.I. (2012). *Optimal'noye upravleniye bezopasnost'yu ekologicheskii opasnykh ob'yektov [Optimal safety management of environmentally hazardous facilities]*. Kiev, Ukraine: Naukova Dumka [in Russian].
3. Iatsyshyn, A.V., Kutsan, Yu.H., Artemchuk, V.O. et al. (2019). Pryntsypy ta metody upravlinnya ekolohichnoyu bezpekoyu na osnovi intelektual'noho analizu danykh merzhi monitorynhu atmosfernoho povitrya. [The principles and methods of ecological safety management through the data of air monitoring network analysis]. *Elektron. Model*, no. 41(4), pp. 85–101 [in Ukrainian].
4. Ivashhuk, O.A., Konstantinov, I.S. (2009). Obespecheniye adaptivnogo upravleniya ekologicheskoy bezopasnost'yu promyshlenno-transportnogo kompleksa. [Guarding of adaptive management by ecological safety of an industrial and transport complex]. *Large-Scale Systems Control*, no. 25, pp. 96–115 [in Russian].
5. Kapralov, E.G., Koshkarev, A.V., Tykunov, V.S. (2008). *Geoinformatika: uchebnik dlya vuzov [Geoinformatics: a textbook for universities]*. Moscow : Academy [in Russian].
6. Chen, C., Chuang, C., Jiang, J. (2013). Ecological Monitoring Using Wireless Sensor Networks. Overview, Challenges, and Opportunities. *Advancement in Sensing Technology. Smart Sensors, Measurement and Instrumentation*, no. 1, pp. 1–21.
7. Dias, D., Tchepel, O. (2014). Modelling of human exposure to air pollution in the urban environment: a GPS-based approach. *Environmental Science and Pollution Research*, no. 5, pp. 3558–3571.
8. Peters, D.P.C., Havstad, K.M., Cushing, J. et al. (2014). Harnessing the power of big data: infusing the scientific method with machine learning to transform ecology. *Ecosphere*, no. 5(6), pp. 1–15.
9. Algoritmy intellektual'nogo analiza dannykh. [Algorithms of the intellectual data analysis] (2015). URL: <https://tproger.ru/translations/top-10-data-mining-algorithms/> [in Russian].
10. Derev'ya resheniy – C4.5 matematicheskii apparat. Chast' 1. [Decision trees – C4.5 mathematical apparatus. Part 1]. (2019). URL: <https://basegroup.ru/community/articles/math-c45-part1> [in Russian].
11. Zgurovsky, M.Z., Boldak, A.A., Yefremov, K.V. (2013). Intellektual'nyy analiz i sistemnoye soglasovaniye nauchnykh dannykh v mezhdistsiplinarynykh issledovaniyakh. [Intelligent analysis and systemic adjustment of scientific data in interdisciplinary research]. *Cybernetics and Systems Analysis*, no. 49, pp. 62–75 [in Russian].
12. Shitikov, V.K., Mastitsky, S.E. (2017). Klassifikatsiya, regressiya, algoritmy Data Mining s ispol'zovaniyem R. [Classification, regression, Data Mining algorithms using R]. URL: <https://github.com/ranalytics/data-mining> [in Russian].
13. Ashouri, M., Haghghat, F., Fung, B.C.M. et al. (2018). Development of building energy saving advisory: A data mining approach. *Energy and Buildings*, no. 172, pp. 139–151.
14. Orange. (2021). URL: <http://orange.biolab.si/download/>.
15. Kryvyy Rih. Avtomatyzovani posty sposterezhennya. (2018). [Kryviy Rih. Automated surveillance posts]. URL: <https://krmisto.gov.ua/ua/rc/ecomon.html> [in Ukrainian].