

12. D. Stokes, *MySQL and JSON: a practical programming guide*. New York, USA: McGraw Hill Publ., 2018.
13. E. Hanchett, and B. Listwon, *Vue.js in Action*. Shelter Island, USA: Manning Publ., 2018.

Рецензент: О.І. Проніна
канд. техн. наук, доц., ДВНЗ «ПДТУ»

Стаття надійшла 18.10.2023

Стаття прийнята 23.11.2023

УДК 004.056.53:519.17

doi: 10.31498/2225-6733.47.2023.299981

© Кажан Д.В.¹, Левицька Т.О.², Кривенко О.В.³

ВИЗНАЧЕННЯ ТИПУ ВИБУХОНЕБЕЗПЕЧНОГО ПРЕДМЕТУ ЗА ДОПОМОГОЮ ДЕРЕВА РІШЕНЬ

В роботі наведено метод дерева рішень для класифікації типу вибухонебезпечних предметів. Це дозволяє робити висновки відносно типу вибухонебезпечного предмету, що заховано в залежності від типу ґрунту. Дана проблематика є актуальною, оскільки дозволяє обґрунтовано пояснювати тип небезпеки в залежності від різних зовнішніх факторів. В роботі запропоновано загальне представлення побудови дерева рішень для визначення виду вибухонебезпечних предметів. Крім цього в роботі надано візуалізацію розбиття на вузли для наочного представлення побудови дерева рішень з урахуванням розбиття на піддерева та вузли. У якості вузлів виступають основні атрибути: напруга, висота датчика над рівнем ґрунту, тип ґрунту. Наведено алгоритм побудови дерева рішень та структурну блок-схему алгоритму ID3. Було проведено дослідження, на основі якого було обрано для вимірювання хаосу розбиття дерева рішень ентропію. Були перелічені основні характеристики алгоритму ID3 з урахуванням предметної галузі, в якій він застосовується. В роботі наведено одне з побудованих дерев рішень, що було побудовано в рамках досліджень для визначення типу вибухонебезпечного предмета в залежності від можливого типу ґрунту, на прикладі це сухий та піщаний ґрунт.

Ключові слова: дерева рішень, візуалізація даних, вибухонебезпечні предмети, алгоритм ID3.

D.V. Kazhan, T.O. Levytska, O.V. Kryvenko. Determining the type of explosive item using a decision tree. The present circumstances in Ukraine delineate specific areas of study. The imperative to identify an explosive device is increasingly paramount for adults and particularly for children. Failing to recognize this danger poses a threat to health and life. The paper presents a decision tree method for classifying the type of explosive objects. This allows you to draw conclusions about the type of explosive object hidden depending on the type of soil. This problem is relevant, as it allows to reasonably explain the type of danger depending on various external factors. The paper offers a general representation of the construction of a decision tree for determining the type of explosive objects. In addition, the work provides a visualization of the division into nodes for a visual representation of the construction of a decision tree, taking into account the division into subtrees and nodes. The nodes are the main attributes: voltage, sensor height above ground level, soil type. The

¹ магістрант, ДВНЗ «Приазовський державний технічний університет», м. Дніпро

² канд. техн. наук, доцент, ДВНЗ «Приазовський державний технічний університет», м. Дніпро, ORCID: 0000-0003-3359-1313, levitskaya_t_a@pstu.edu

³ канд. техн. наук, доцент, ДВНЗ «Приазовський державний технічний університет», м. Дніпро, ORCID: 0009-0006-2860-6575, krivenko_o_v@pstu.edu

decision tree construction algorithm and the structural block diagram of the ID3 algorithm are given. A study was conducted based on who was chosen to measure the entropy decision tree split chaos. The main characteristics of the ID3 algorithm were listed, taking into account the subject area in which it is used. The paper presents one of the built decision trees, which was built as part of research to determine the type of explosive object depending on the possible type of soil, for example, it is dry and sandy soil. Moreover, the paper provides a comparative analysis of the decision tree model against existing methodologies, shedding light on its efficacy, accuracy, and potential advancements in enhancing threat detection capabilities. This comprehensive analysis not only showcases the model's strengths but also offers insights into potential areas for refinement and improvement.

Key words: *decision trees, data visualization, explosive objects, ID3 algorithm.*

Постановка проблеми. Сьогоднішня ситуація в Україні диктує свої напрямки тематики для вивчення. Необхідність уміти розпізнати вибухонебезпечний пристрій стає першочерговою задачею як для дорослих, так і особливо для дітей. Бо невміння розпізнати небезпеку несе в собі загрозу здоров'ю та життю. Окрім цього міни виявилися серйозною перешкодою для українських військових, сповільнюючи наступальні дії, пошкоджуючи бронетехніку. Для подолання мінних полів у Запорізькій і Донецькій області необхідні масштабні військово-інженерні операції, які є непростими навіть для найкраще оснащених армій.

Таким чином, в країні стає гостре питання щодо виявлення мін та їх розмінування. Основна загроза, яка стає перед цивільним населенням, полягає в тому, щоб вчасно помітити міну та проінформувати служби, що займаються розмінуванням. Це значить, що діти та дорослі повинні бути обізнаними у тому, яким чином може їм погрожувати небезпека. Тому навчання з попередження ризиків від вибухонебезпечних предметів є актуальним питанням сьогоднішня в Україні.

Аналіз останніх досліджень і публікацій. В роботі [1] представлено запровадження дистанційного навчання в українських школах під час карантину через пандемію COVID-19. В ній представлено дослідження з використанням різних видів електронної освіти, ресурсів, що допомагають викладачам забезпечити якісне дистанційне навчання. В роботі представлено онлайн-опитування 576 вчителів, на основі якого можна зробити висновки, що 12% опитаних вчителів почали користуватися електронними освітніми ресурсами для забезпечення дистанційного навчання студентів під час карантину.

В роботі [2] розглядається процес реформування освіти, спрямований на перехід вищої освіти до компетентнісних та індивідуальних підходів. Автори зазначають, що відбувається стикування з необхідністю побудови індивідуальних навчальних шляхів для кожного майбутнього морського фахівця. У цьому відношенні технологія адаптації навчання на основі сучасних інформаційно-комунікаційних технологій (ІКТ) набуває великого значення. Автори піднімають питання створення адаптивного інформаційного середовища, актуального для підготовки сучасних і конкурентоспроможних спеціалістів. Автори вважають, що це середовище має базуватися на впровадженні адаптивних технологій навчання та підготовки студентів морського факультету, тому в статті досліджено педагогічну проблему майбутніх мореплавців.

У статті [3] розглядається доповнена реальність, як одна з технологій, що найбільш активно розвивається, яка знайшла своє застосування у сфері освіти. Аналіз різноманітних публікацій підтвердив, що технологія AR відкриває нові можливості для викладачів і підвищує привабливість навчання для студентів різного віку. Особливо це стосується маленьких дітей. У статті визначено види діяльності, які надають дитині особистий досвід використання технології AR. Дается порівняння існуючих додатків AR для вивчення алфавіту. Описано нову програму AR, розроблену за допомогою Unity, C#, Vuforia. Розроблений мобільний AR-додаток дає можливість вивчати український алфавіт, назви цифр, звуки тварин. Завдяки цьому додатку процес навчання супроводжується тривимірною візуалізацією та озвученням кожної букви та цифри. Аналіз опитування вчителів і батьків показав, що при використанні AR-додатку інтерес і самоефективність дітей у вивченні літер і цифр значно зростає. Використання додатку AR збільшили швидкість запам'ятовування матеріалу та допомогли утримати увагу дитини під час навчання нового матеріалу.

Мета дослідження. Метою роботи є підвищення ефективності визначення вибухонебезпечних предметів за допомогою алгоритму ID3.

Виклад основного матеріалу. Дерево рішень – це контрольовані алгоритми машинного навчання, які використовуються як для регресії, так і для формулювання проблеми класифікації. Вони використовують подання дерева для вирішення проблеми, в якій кожен вузол представляє атрибут, кожне посилення представляє правило прийняття рішень, а кожен листок представляє результат (категоричне або безперервне значення) [4].

Використовуються дерева рішень в даній роботі в якості класифікатора для визначення типу мін. Тобто у якості листка виступає тип міни: протитанкова, протипіхотна міна-пастка, М-14 протипіхотна або відсутність міни. Дане розбиття обрано з погляду на реальні дані, що використовуються як тестовий набір даних для визначення типу міни.

У якості вузлів виступають основні атрибути: напруга, висота датчика над рівнем ґрунту, тип ґрунту. В якості правил виступають правила продукційного типу.

Візуальне представлення побудови дерева рішень з урахуванням розбиття на піддерева та вузли наведено на рис. 1.

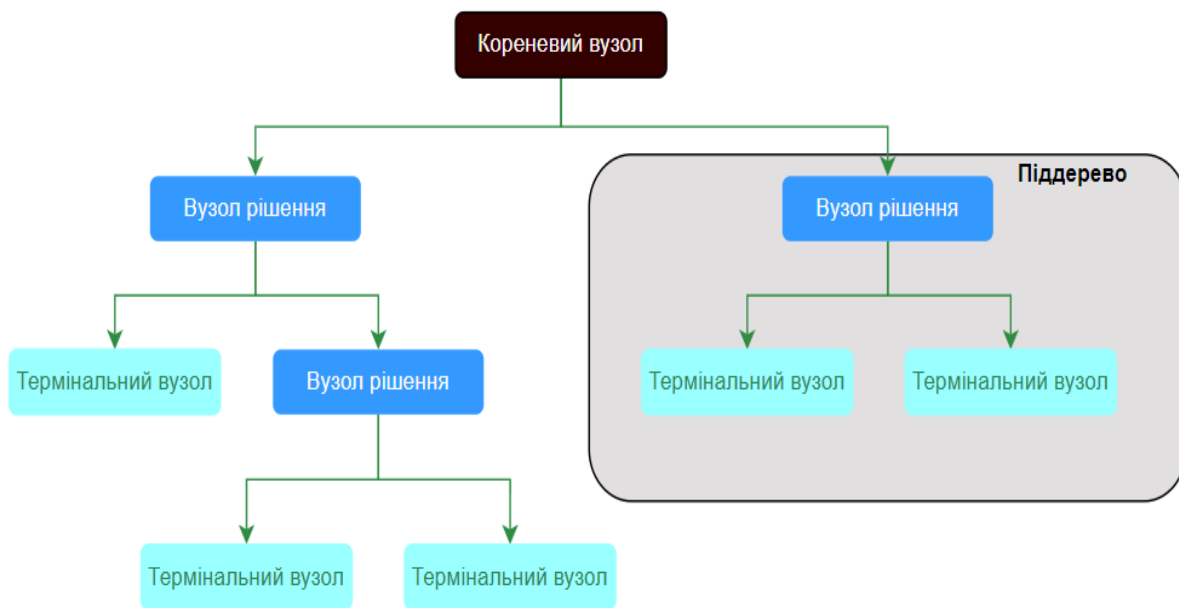


Рис. 1 – Візуалізація розбиття на вузли

Було вирішено використовувати алгоритм ID3, він є алгоритмом класифікації, який використовує жадібний підхід побудови дерева рішень шляхом вибору найкращого атрибута, який дає максимальний приріст інформації (IG) або мінімальну ентропію (H) [5].

Основні складники алгоритму: вузол – функція (атрибут); відділення – рішення (правило); листок – підсумок (категоричний або суцільний).

Кореневий вузол – це найвищий вузол у дереві, який представляє повний набір даних. Також можна сказати, що це початкова точка процесу прийняття рішень.

Рішення / внутрішній вузол. Вузли рішення є нічим іншим, як результатом поділу даних на кілька сегментів даних, і головною метою є створення дочірніх вузлів із максимальною однорідністю або чистотою (тобто всі одного типу).

Ліцевий / кінцевий вузол – цей вузол представляє розділ даних із найвищою однорідністю (означає, що всі вони одного типу).

В алгоритмі ID3 було вирішено використовувати як критерій визначення хаосу – ентропію. Ентропія є мірою кількості невизначеності в наборі даних S [6]. Математичне представлення ентропії:

$$H(S) = \sum_{c \in C} -p(c) \log_2 p(c), \quad (1)$$

де S – поточний набір даних, для якого обчислюється ентропія (змінюється на кожній ітерації алгоритму ID3);

C – набір класів у S {приклад – $C = \{yes, no\}$ };

$p(c)$ – пропорція кількості елементів у класі c до кількості елементів у наборі S .

Ентропія в ID3 ентропія обчислюється для кожного атрибута, що залишився. Атрибут із найменшою ентропією використовується для розбиття набору S на цій конкретній ітерації [7].

Ентропія = 0 означає, що це чистий клас, тобто всі належать до однієї категорії. В цілому це майже не досягається, бо якщо поставити обмеження складання дерева саме по фінальному значенню ентропії, то буде виконано перенавчання. Тому що коли ентропія дорівнює нулю, то позначає, що розбиття зупинилось на поділені на кожний окремий лист, а це не є метою побудови дерева рішень. Крім цього, це в свою чергу є серйозною проблемою.

Приріст інформації $IG(A)$ повідомляє нам, наскільки похибка в S була зменшена після розбиття набору S на атрибут A . Математичне представлення приросту інформації:

$$IG(A, S)_i = H(S) - \sum_{t \in T} p(t)H(t), \quad (2)$$

де $H(S)$ – ентропія множини S ;

T – підмножини, створені в результаті розбиття множини S за атрибутом A .

Підмножина, яка створюється у результаті розбиття множини S за атрибутом A , розраховується наступним чином:

$$S = \cup_{t \in T} H(t), \quad (3)$$

де $p(t)$ – пропорція кількості елементів у t до кількості елементів у наборі S ;

$H(t)$ – ентропія підмножини t .

В цілому в алгоритмі ID3 можна розрахувати приріст інформації (замість ентропії) для кожного атрибута, що залишився. Його можна використовувати для відбору ознак шляхом оцінки інформаційного приросту кожної змінної у контексті цільової змінної. Атрибут із найбільшим приростом інформації використовується для розбиття набору S на цій конкретній ітерації.

Критерій приросту інформації використовується здебільшого для вибору кращого розбиття підмножин у вузлах дерев рішень в алгоритмах навчання ID3. Але для задачі визначення типу вибухонебезпечного об'єкта при побудові дерева рішень було вирішено використовувати ентропію, оскільки саме визначення хаосу в кожному піддереві є поточною актуальною задачею.

Алгоритм ID3 включає в себе наступні основні кроки:

Крок 1: Попередня обробка даних – очищення і попередня обробка даних, створення множини V . Необхідно обробити відсутні значення та перетворити категоричні змінні в числові представлення, якщо це необхідно.

Крок 2: Вибір кореневого вузла – обчислення ентропії цільової змінної (мітки класу) на основі набору даних, згідно з формулою (1).

Крок 3: Розрахунок приросту інформації – для кожного атрибута в наборі даних обчислюється приріст інформації, згідно з формулою (2), коли набір даних розділено на цей атрибут, кожен атрибут має номер з множини i .

Крок 4: Вибір найкращого атрибута – вибір атрибута (A) із найбільшим приростом інформації як вузла прийняття рішення для дерева.

Крок 5: Розділення набору даних – розділення набору даних на основі значень вибраного атрибута ($A_i \subset S$).

Крок 6: Повторіть процесу – рекурсивно повторюються кроки 2-5 для кожної підмножини (A_i), доки не буде виконано критерію зупинки.

Критерієм зупинки може бути – глибина дерева досягне максимальної межі, всі екземпляри в підмножині належатимуть до одного класу.

Для обраної задачі визначення типу вибухонебезпечного предмету обрано максимальну глибину розбиття $l = 3$.

Алгоритм визначення типу вибухонебезпечного предмета за допомогою дерева рішень представлено на рис. 2.

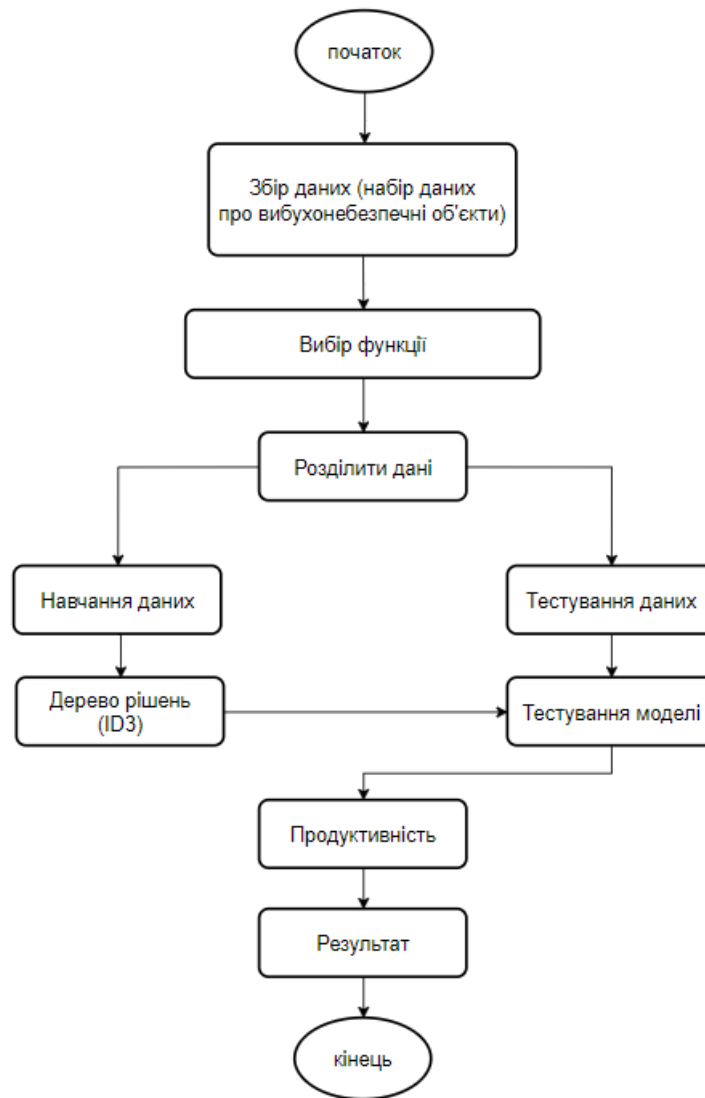


Рис. 2 – Блок схема алгоритму ID3

Основні характеристики алгоритму ID3 такі:

- 1) ID3 використовує жадібний підхід, тому він не гарантує оптимального рішення; він може застрягти в локальних оптимумах.
- 2) ID3 може переобладнати навчальні дані (щоб уникнути переобладнання, слід віддавати перевагу меншим деревам рішень перед більшими).
- 3) Цей алгоритм зазвичай створює невеликі дерева, але він не завжди створює найменше можливе дерево, що в обраній задачі визначення вибухонебезпечних предметів є головною задачею.
- 4) ID3 важче використовувати для безперервних даних (якщо значення будь-якого заданого атрибута є безперервним, тоді є багато місць для розділення даних на цьому атрибуті, і пошук найкращого значення для розділення може зайняти багато часу), з поглядом на обрану предметну галузь даних мінус не є критичним.

Оскільки за допомогою класифікації можна визначати схожі характеристики та можливе передбачення розташування вибухонебезпечних предметів в залежності від типу ґрунту.

В виборці, на якій проводились дослідження та яка стала основою для побудови дерев рішень, використовуються наступні показники: напруга, висота, тип ґрунту, тип міни. Детальніше наведено визначення кожного показника, що використовується у вибірці.

Напруга – Значення вихідної напругу датчика FLC завдяки магнітного скривлення.

Висота – висота датчика від землі.

Тип ґрунту – тип ґрунту, де може бути встановлена міна, поділ іде на шість різних типів ґрунту в залежності від вологості стану ґрунту.

Типи мін – це міни, що найчастіше зустрічаються на суші, у виборці розглянуто п'ять різних класів.

Спираючись на ці знання, можна навчити уважно відслідковувати навколишнє середовище та виявляти небезпеку. На рис. 3 наведено побудоване дерево рішень для типу ґрунту сухий та піщаний. Таким чином, можна відслідковувати по різному типу ґрунту, який саме тип вибухонебезпечного об'єкта там розташовано.

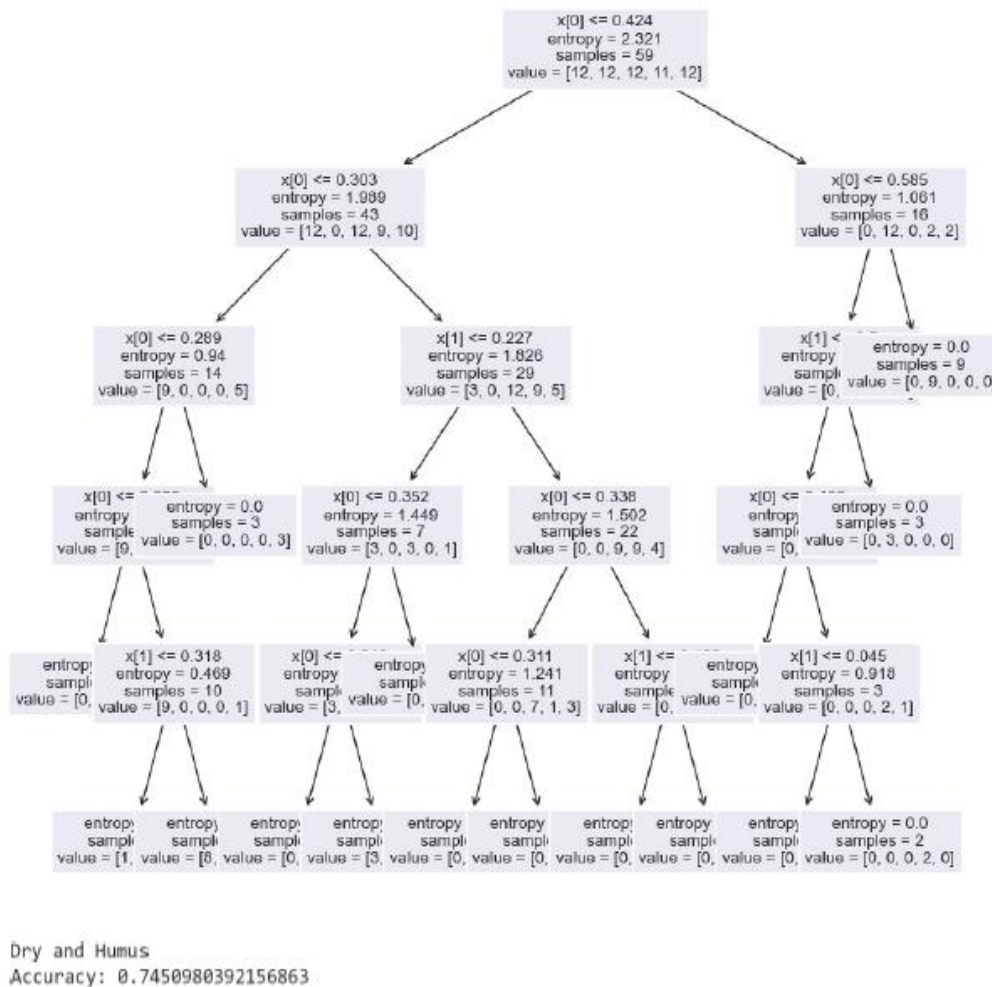


Рис. 3 – Побудоване дерева рішень для типу ґрунту сухий та піщаний

На рис. 3 наведено одно з побудованих дерев рішень, крім нього побудовано ще шість дерев рішень, на яких представлено поєднання різних типів ґрунту. Це підтверджує те, що саме тип ґрунту є одним із головних показників, оскільки різні типи мін будуть встановлені в різному типі ґрунту.

Висновки

Запропонований алгоритм ID3 використовується для визначення типу вибухонебезпечних предметів, що представляє собою значущий крок у вирішенні актуальної проблеми в Україні. Використання методу дерева рішень для класифікації типів вибухонебезпечних предметів виявляється обґрунтованим підходом, який дозволяє адаптувати тип небезпеки до різних зовнішніх факторів. Подана структурна блок-схема алгоритму ID3.

В подальшому планується використання побудованих дерев рішень в основі інтерактивної навчальної системи. Що є актуальною науково-практичною задачею для вивчення типів небезпеки при її виявленні.

Перелік використаних джерел:

1. Experience of use of electronic educational resources by Ukrainian teachers during the distance learning due to the COVID-19 pandemic (March-May 2020) / Zhenchenko M., Melnyk O., Prykhoda Y., Zhenchenko I. *ICT in education, research and industrial applications. integration, harmonization and knowledge transfer* : Proceedings of the 17th International Conference, Kherson, Ukraine, 28 September - 2 October 2021. Vol. II. Pp. 55-65.
2. Adaptive Learning Environment Design in the System of Future Maritime Specialits' Training / S. Voloshynov, I. Riabukha, O. Dobroshtan, H. Popova, T. Spychak. *ICT in education, research and industrial applications. integration, harmonization and knowledge transfer* : Proceedings of the 17th International Conference, Kherson, Ukraine, 28 September - 2 October 2021. Vol. II. Pp. 66-76.
3. Using augmented reality for early literacy / Piatykop O., Pronina O., Tymofieieva I., Palii I. *ICT in education, research and industrial applications. integration, harmonization and knowledge transfer* : Proceedings of the 17th International Conference, Kherson, Ukraine, 28 September - 2 October 2021. Vol. II. Pp. 221-236.
4. Shai S.-S., Shai B.-D. Understanding machine learning: from theory to algorithms. Cambridge University Press, 2014. 500 p.
5. Gambhir A. Classification problem in data mining using decision trees. LAP LAMBERT Academic Publishing, 2019. 96 p.
6. Maimon O.Z., Rokach L. Data mining with decision trees: theory and applications. 2nd ed. World Scientific Publishing, 2014. 330 p.
7. Rokach L. Ensemble learning: pattern classification using ensemble methods. 2nd ed. World Scientific Publishing, 2019. 300 p.

References:

1. M. Zhenchenko, O. Melnyk, Y. Prykhoda, and I. Zhenchenko, «Experience of use of electronic educational resources by Ukrainian teachers during the distance learning due to the COVID-19 pandemic (March-May 2020)», in Proceedings of the 17th International Conference ICT in education, research and industrial applications. Integration, harmonization and knowledge transfer, Kherson, Ukraine, 2021, vol. II, pp. 55-65.
2. S. Voloshynov, I. Riabukha, O. Dobroshtan, H. Popova, and T. Spychak, «Adaptive Learning Environment Design in the System of Future Maritime Specialits' Training», in Proceedings of the 17th International Conference ICT in education, research and industrial applications. integration, harmonization and knowledge transfer, Kherson, Ukraine, 2021, vol. II, pp. 66-76.
3. O. Piatykop, O. Pronina, I. Tymofieieva, and I. Palii, «Using augmented reality for early literacy», in Proceedings of the 17th International Conference ICT in education, research and industrial applications. Integration, harmonization and knowledge transfer, Kherson, Ukraine, 2021, vol. II, pp. 221-236.
4. S.-S. Shai, and B.-D. Shai, *Understanding machine learning: from theory to algorithms*. Cambridge University Press Publ., 2014.
5. A. Gambhir, *Classification problem in data mining using decision trees*. LAP LAMBERT Academic Publishing, 2019.
6. O.Z. Maimon, and L. Rokach, *Data mining with decision trees: theory and applications*, 2nd ed. World Scientific Publishing, 2014.
7. L. Rokach, *Ensemble learning: pattern classification using ensemble methods*, 2nd ed. World Scientific Publishing, 2019.

Рецензент: О.І. Проніна
канд. техн. наук, доц., ДВНЗ «ПДТУ»

Стаття надійшла 18.10.2023
Стаття прийнята 23.11.2023