

О.В. Кучерук, магістр
В.В. Желізко, асистент
Г.В. Марчук, ст. викладач
О.В. Коротун, к.пед.н., доц.

Державний університет «Житомирська політехніка»

Системи підбору кваліфікованих медичних фахівців з онкології на основі аналізу медичних даних

За останні роки системи підтримки прийняття рішень стали критично важливими в сфері охорони здоров'я, надаючи допомогу працівникам медицини в прийнятті обґрунтованих рішень. У статті досліджується роль алгоритмів дерева рішення як одного з провідних методів у системах підтримки прийняття рішень, наголошуючи на їх ефективності в задачах класифікації та прогнозування. Дерева рішень мають кілька переваг, враховуючи інтуїтивну структуру, простоту інтерпретації та можливість надавати чіткі пояснення поряд із точними результатами. Також дерева рішень відзначаються здатністю видобувати значущі правила з комплексних наборів даних, що сприяє виявленню прихованих закономірностей, які можуть бути неочевидними за допомогою традиційних аналітичних методів.

У статті підкреслюється універсальність дерев рішень, що робить їх особливо придатними для широкого спектра медичних застосувань, таких як пошук найбільш підходящого лікаря, встановлення діагнозу та методів лікування. Така можливість значно покращує процеси прийняття клінічних рішень, дозволяючи забезпечити більш індивідуалізовану та ефективну медичну допомогу.

Оскільки охорона здоров'я дедалі більше впроваджує рішення, засновані на даних, інтеграції алгоритмів дерева рішень у системі підтримки прийняття рішень, ймовірно, принесе значні переваги, покращуючи якість медичної практики та, врешті-решт, призводячи до кращих показників здоров'я пацієнтів. У статті робиться висновок про необхідність подальшого розвитку та застосування алгоритмів дерева рішень у медичних установах, підкреслюючи їх потенціал революціонізувати прийняття медичних рішень і системи підтримки в досягненні поліпшення показників здоров'я та вибору лікаря-онколога.

Ключові слова: *аналіз даних; машинне навчання; дерева рішень; штучний інтелект; охорона здоров'я; алгоритми.*

Актуальність теми. Наразі галузь охорони здоров'я потребує пошуку шляхів для покращення підбору висококваліфікованих медичних фахівців, особливо в таких сферах, як онкологія. Лікарі-онкологи відіграють надзвичайно важливу роль під час діагностування та лікування новоутворень та подальшого скерування до необхідного спеціаліста. Сучасні методи підбору лікарів часто базовані на обмежених даних, таких як рекомендації чи географічне розташування і не використовують більшість доступної інформації. Оскільки заклади охорони здоров'я все частіше впроваджують електронні медичні записи та приймають рішення на основі даних, стає очевидним, що існує необхідність у створенні автоматизованої системи, здатної аналізувати медичні дані та надавати об'єктивні рекомендації при виборі лікарів-онкологів.

Один із найбільших викликів під час розробки описаного вище програмного забезпечення полягає в обробці складних та різноманітних медичних даних, що включає в себе не лише базову демографічну інформацію, а й професійний досвід, сертифікації, відгуки пацієнтів та показники клінічної ефективності. Таке програмне забезпечення має гарантувати захист даних та постійно їх оновлювати для надання найбільш точних рекомендацій. З огляду на високі ризики, пов'язані з онкологічною допомогою, помилки при підборі лікарів можуть мати досить серйозні наслідки для пацієнтів. Описані вище виклики підкреслюють необхідність прецизійної розробки автоматизованої системи підбору онкологів з використанням методів штучного інтелекту, таких як дерева рішень.

Підхід покликаний підвищити ефективність та точність процесу підбору лікарів на основі ШІ, яке використовує алгоритми дерева рішень, пропонуючи масштабовану модель, яку можна адаптувати до інших спеціальностей у медицині в майбутньому.

Аналіз останніх досліджень та публікацій, на які спираються автори. Існує значна кількість наукових досліджень, присвячених системам підбору лікарів та програмному забезпеченню, що допомагають підібрати необхідних фахівців. Вони окреслюють такі питання: довіра до штучного інтелекту (ШІ) та вплив на сприйняття громадськістю ризиків і переваг його використання у галузі охорони здоров'я [1]; використання ШІ у медицині щодо зменшення відсотковості людських помилок та зберігання часу

лікарів [2]; узагальнення принципів застосування технологій штучного інтелекту у медицині [3], критерії вибору лікаря за допомогою мобільної онлайн-системи [4], проблема недостатньої ефективності традиційних методів рекомендацій у медичній сфері та пропозиції щодо нових підходів, що враховує географічне положення пацієнтів [5], питання розробки та впровадження системи, яка допомагає користувачам знаходити лікаря-спеціаліста на основі симптомів хвороби [6].

Дослідження сайту «helsi.me» [7], спрямованого на розвиток медичних технологій та допомогу пацієнтам знайти необхідного лікаря за різними параметрами (спеціальність, рейтинг, тип клініки, стаж та стаж лікаря, варіанти оплати послуг), виявило основний його недолік, а саме відсутність вибору лікаря у невеликих містах, враховуючи Житомир. Але такий сайт надає можливість пацієнтам записатися на прийом онлайн.

Також чимала кількість наукових робіт вітчизняних та закордонних науковців присвячена успішній реалізації програм щодо пошуку лікарів. Для вирішення цього завдання найчастіше використовують метод штучного інтелекту як алгоритм дерева рішень. У [8] зазначено, Data Mining – це термін, що був створений для опису процесу перегляду великих баз даних з метою пошуку цікавих і раніше невідомих шаблонів. Сфера Data Mining надає технології та засоби, за допомогою яких величезні обсяги даних можуть бути проаналізовані автоматично. Таким чином, використання алгоритмів дерева рішень під час реалізації програми підбору онкологів може значно допомогти із аналізом даних пацієнтів, їх історією лікування, кваліфікацією лікарів, методів лікування тощо, а також спрямована на покращення показників лікування.

Метою статті є дослідження систем підбору лікарів з використанням штучного інтелекту на мові Python.

Викладення основного матеріалу. Системи підтримки прийняття рішень, які значно допомагають лікарям, стають дуже важливою частиною прийняття медичних рішень. Вони базуються на різних моделях, і найкращі з них надають пояснення разом із точною, надійною та швидкою відповіддю. Однією з найбільш життєздатних серед моделей є алгоритми дерева рішень, які вже успішно впроваджені для прийняття багатьох медичних рішень.

Одним із найпопулярніших методів розв'язання задач, класифікації та прогнозування є алгоритм дерева рішень. Іноді цей метод Data Mining також називають деревами вирішальних правил, деревами класифікації і регресії.

Виокремимо основні переваги дерев рішень – це їх інтуїтивність (спрощує розуміння розв'язуваної задачі); результат роботи алгоритмів їх конструювання легко інтерпретується користувачем; можливість витягати правила з бази даних; вони дозволяють створювати класифікаційні моделі в тих сферах, де аналітики досить складно формалізувати знання; алгоритм їх конструювання не вимагає від користувача вибору вхідних атрибутів (незалежних змінних); точність моделей, створених за їх допомогою, порівняно з іншими методами побудови класифікаційних моделей; швидкий процес їх навчання.

Наведемо опис двох типів дерев рішень для Data Mining, а саме: аналіз дерева класифікації, коли результат, що передбачається, є класом, до якого належать дані; регресійний аналіз дерева, коли результат, що передбачається, можна розглядати як дійсне число. Основними компонентами дерева рішень є: вузол, що представляє атрибут (характеристику) даних, за яким приймається рішення про подальше розгалуження; гілка, що з'єднує вузол з іншими вузлами або листом і представляє можливі значення атрибута; лист – кінцевий вузол, який відповідає прогнозу або класу. Математичну модель дерева рішення наводять у вигляді функції, що відображає вхідні дані в прогноз таким чином за формулою (1):

$$y = f(x_1, x_2, \dots, x_n), \quad (1)$$

де y – прогнозне значення (клас або числове значення); x_1, x_2, \dots, x_n – вхідні атрибути та f – функція, визначена структурою дерева рішень.

Одним із ключових під час побудови дерев рішень є ентропія. Вона є мірою невизначеності чи випадковості у даних. У контексті дерев рішень її використовують для оцінки чистоти вузла. Чим вища ентропія, тим менш однорідні дані у цьому вузлі. Ентропію визначають за формулою (2):

$$H(S) = - \sum_{i=1}^n p(i) \cdot \log_2 p(i), \quad (2)$$

де $H(S)$ – ентропія множини S ; $p(i)$ – частка елементів класу i в множині S .

Обчислення середньозваженої ентропії після розбиття відбувається за формулою (3) як сума добутків ентропії кожної підмножини на її відносну частоту в початковому наборі:

$$H(S, A) = \sum_{i=1}^n \frac{|Sv|}{|S|} \cdot H(Sv), \quad (3)$$

де $H(S, A)$ – середньозважена ентропія після розбиття за атрибутом A ; Sv – підмножина даних, отримана після розбиття за значенням атрибута A ; $|Sv|$ – кількість елементів у підмножині Sv ; $|S|$ – загальна кількість елементів у наборі даних.

Також здійснюється розрахунок інформаційного приросту за формулою (4), що показує, наскільки зменшується невизначеність після розбиття даних за певним атрибутом:

$$\text{Gain}(S, A) = H(S) - H(S, A), \quad (4)$$

де $\text{Gain}(S, A)$ – інформаційний приріст при розбитті за атрибутом A ; $H(S)$ – ентропія вихідного набору; $H(S, A)$ – середньозважена ентропія після розбиття.

Зазвичай дерева рішень використовуються для того, щоб уточнити певні деталі, скласти карти чи знайти відповіді на складну проблему. Для побудови дерева рішень важливо імпортувати певні бібліотеки, які містять необхідні функції та інструменти і зчитати дані з наданого набору. Зазвичай дані наводяться у вигляді таблиці, де кожен рядок відповідає окремому прикладу, а стовпці – ознакам або змінним. У Python можна використовувати бібліотеку pandas для зчитування даних з файлу CSV з вказанням назви набору даних (рис. 1)

```
print(data)
```

	age	gender	experience	education	classification	city
0	50	female	13	Internship	Surgical Oncologist	Zhytomyr
1	60	male	32	Internship	Surgical Oncologist	Zhytomyr
2	55	male	28	Internship	Medical Oncologist	Zhytomyr
3	60	male	32	Internship	Mammologist	Zhytomyr
4	40	female	15	Internship	Gynecologic Oncologist	Kyiv
..
495	47	male	22	Internship	Gynecologic Oncologist	Kyiv
496	52	male	27	Internship	Mammologist	Kyiv
497	65	male	26	Internship	Mammologist	Chernivtsi
498	35	female	9	Internship	Mammologist	Kharkiv
499	52	male	27	Internship	Oncologist	Lviv

	courses	visits_per_month	rating
0	no	122	0.953
1	yes	127	0.717
2	no	149	0.877
3	yes	154	0.830
4	yes	191	0.892
..
495	no	195	0.975
496	no	195	0.799
497	yes	121	0.790
498	no	138	0.742
499	no	185	0.730

[500 rows x 9 columns]

Рис. 1. Набір даних data

Отримані дані необхідно розділити на тестові та тренувальні: 80 % даних використаємо для тренування моделі дерева рішень, 20 % даних для тестування та отримання результату. Проведемо налаштування моделі дерева рішень, її тренування на навчальних даних та прогнозування значень на тестовому наборі.

Маємо множину D , що містить N наборів навчальних даних. При цьому кожен i -й набір (A_1, A_2, \dots, A_n) складається з вхідних даних – атрибутів від A_1 до A_n та вихідних даних – атрибуту класу C . Можна припустити, що атрибути A_1, \dots, A_n приймають лише категоріальні значення. Атрибут класу C приймає одне з K дискретних значень: $C \in \{1, \dots, K\}$. Метою є прогнозування класифікаційним правилом значення атрибуту класу C на основі значень атрибутів A_1, \dots, A_n [9].

Проведемо побудову дерева рішень для задачі підбору лікаря з метою визначення значення поля «classification». При цьому використовуватимуться лише вхідні ознаки ['age', 'experience', 'visits_per_month', 'rating']. Для цього скористаємося відповідними алгоритмами та бібліотеками, імпортованими заздалегідь та визначеними у коді, що дозволяє зчитати дані з набору, обмежити їх вхідними ознаками та навчити модель дерева рішень на цих даних. Після його виконання буде побудовано дерево рішень, що допоможе розв'язувати задачу підбору лікаря на основі зазначених ознак (рис. 2).

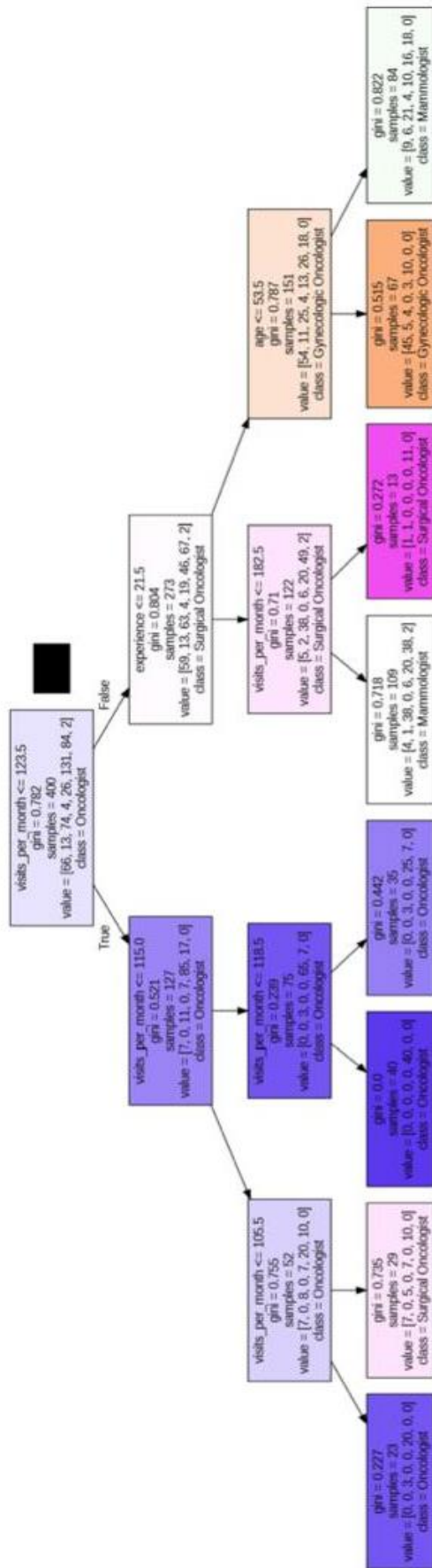


Рис. 2. Дерево рішень

В першу чергу проведемо вибір вхідних та вихідних ознак з набору даних. Категоріальні ознаки закодовано за допомогою методу `pd.get_dummies`. Після цього дані розділено на навчальні та тестові. Проведемо налаштування моделі дерева рішень, її тренування на навчальних даних та прогнозування значень на тестовому наборі. Після проведених кроків побудови та тренування моделі дерева рішень отримано точність прогнозування 45 % при обмеженні глибини дерева на рівні трьох (рис. 3).

```
Точність моделі: 0.45
Обмежившись значенням max_depth: 3
```

Рис. 3. Точність моделі дерева рішень з глибиною 3

У межах експерименту збільшимо глибину моделі дерева рішень до 6. Після тренування моделі з новим параметром глибини отримали точність прогнозування на рівні 69 % (рис. 4), що свідчить про те, що збільшення глибини дерева дозволило моделі краще враховувати взаємозв'язки між ознаками та покращити її прогностичні можливості. Значення точності 69 % вказує на те, що чим більша глибина моделі, тим більше вона здатна зробити точніші прогнози щодо класифікації.

```
Точність моделі: 0.69
Обмежившись значенням max_depth: 6
```

Рис. 4. Точність моделі дерева рішень з глибиною 6

Інтеграція дерев рішень у системі вибору найбільш підходящого лікаря-онколога забезпечує можливість швидшого та надійнішого вибору спеціалістів на основі надійних даних. Збір даних є ключовим елементом у побудові ефективних систем підтримки прийняття рішень, особливо в галузі медицини. Якість зібраних даних безпосередньо впливає на точність та надійність побудови моделей прийняття рішень, таких як дерева рішень.

За допомогою високоточних та надійних даних алгоритми побудови дерев рішень можуть бути навчені для виявлення закономірностей та прогнозування результатів у сфері охорони здоров'я, покращуючи процеси встановлення діагнозу та лікування. Результати процесу збору даних, опублікованих у праці [10], підкреслює важливість структурованого збору даних для забезпечення систем зі штучним інтелектом, що використовуються і в лікарнях. З розвитком збору даних методи все більше автоматизуються, що підвищує загальну ефективність інструментів прийняття рішень, особливо в тих сферах, де критично важливі точні та швидкі рішення.

Висновки та перспективи подальших досліджень. У статті було проаналізовано останні дослідження, присвячені системам підтримки прийняття рішень та підбору лікарів онлайн. Виявлено, що серед різних методів, що використовуються у системі підтримки прийняття рішень (СППР), алгоритми дерева рішень вважаються найефективнішими завдяки своїй простоті, зрозумілості та надійній продуктивності в задачах класифікації та прогнозування. Алгоритми пропонують точні результати та чіткі пояснення, що особливо важливо при побудові довіри між працівниками медичної сфери та пацієнтами. Деревя рішень є універсальними методами, які продуктивно розв'язують як задачі класифікації, так і регресії, що робить алгоритми придатними для широкого спектра медичних застосувань від діагностики захворювань до прогнозування результатів лікування. Авторами пропонується використовувати метод побудови дерева рішень для пошуку найбільш підходящого лікаря за критеріями пацієнта.

Список використаної літератури:

1. Kerstan S. Choosing human over AI doctors? How comparative trust associations and knowledge relate to risk and benefit perceptions of AI in healthcare / S.Kerstan, N.Bienefeld, G.Grote // Risk Analysis. – 2024. – Vol. 44 (4). – P. 939–957 [Electronic resource]. – Access mode : <https://onlinelibrary.wiley.com/doi/full/10.1111/risa.14216>.
2. Božić V. Artificial Intelligence – powered. Expert Systems in Hospitals / V.Božić. – 2024. – 15 p. [Electronic resource]. – Access mode : https://www.researchgate.net/profile/Velibor-Bozic-2/publication/378943075_Artificial_Intelligence_powered_Expert_Systems_in_Hospitals/links/65f2e25932321b2c5ff78caa5/Artificial-Intelligence-powered-Expert-Systems-in-Hospitals.pdf.
3. Технології штучного інтелекту в медичній практиці / О.В. Вороненко, І.Б. Галелюка, В.О. Романов, О.П. Мінцєр. – 2020. – С. 17–27 [Електронний ресурс]. – Режим доступу : <http://ir.nuozu.edu.ua:8080/handle/lib/1705>.
4. Wankhade M.S. Research Paper on : Doctor Appointment Booking System / M.S. Wankhade. – 2023 [Electronic resource]. – Access mode : https://www.researchgate.net/publication/371171550_Research_Paper_on_Doctor_Appointment_Booking_System.

5. Chunhua J. Doctor Recommendation Model Based on Ontology Characteristics and Disease Text Mining Perspective / J.Chunhua, Z.Shuangzhu. – 2021 [Electronic resource]. – Access mode : https://www.researchgate.net/publication/353801231_Doctor_Recommendation_Model_Based_on_Ontology_Characteristics_and_Disease_Text_Mining_Perspective.
6. Jagadeesh Kannan R. Doctor Finder: Find doctors on the Go / R.Jagadeesh Kannan. – 2020 [Electronic resource]. – Access mode : https://www.researchgate.net/publication/346206050_Doctor_Finder_Find_doctors_on_the_Go.
7. Медична система для пацієнтів, лікарів, державних та приватних медичних закладів [Електронний ресурс]. – Режим доступу : <https://helsi.me/>.
8. Maimon O.Z. Data mining with decision trees: theory and applications / O.Z. Maimon, L.Rokach // World scientific. – 2014. – Vol. 81 [Electronic resource]. – Access mode : https://books.google.com.ua/books?hl=uk&lr=&id=OVYCCwAAQBAJ&oi=fnd&pg=PR6&dq=5.%09Rokach+L.+DATA+MINING+WITH+DECISION+TREES+Theory+and+Applications&ots=tLieb11fYL&sig=yydce3QYzj8eZmnSW0WV8Y5G08&redir_esc=y#v=onepage&q=5.%09Rokach%20L.%20DATA%20MINING%20WITH%20DECISION%20TREES%20Theory%20and%20Applications&f=false.
9. Андрущак І.С. Розробка та застосування методів прийняття рішень, оптимізації та керування в системних медичних дослідженнях / І.С. Андрущак, В.П. Марценюк. – 2015 [Електронний ресурс]. – Режим доступу : https://scc.knu.ua/upload/iblock/729/dis_Andrushchak%20I.Ye..pdf.
10. PoPu-Data: A Multilayered, Simultaneously Collected Lying Position Dataset / L.Fonseca, F.Ribeiro, J.Metrólho and other // Data. – 2023. – 8 p. [Electronic resource]. – Access mode : <https://www.mdpi.com/2306-5729/8/7/120>.

References:

1. Kerstan, S., Bienefeld, N. and Grote, G. (2024), «Choosing human over AI doctors? How comparative trust associations and knowledge relate to risk and benefit perceptions of AI in healthcare», *Risk Analysis*, Vol. 44 (4), pp. 939–957, [Online], available at: <https://onlinelibrary.wiley.com/doi/full/10.1111/risa.14216>
2. Božić, V. (2024), «Artificial Intelligence – powered. Expert Systems in Hospitals», 15 p., [Online], available at: https://www.researchgate.net/profile/Velibor-Bozic-2/publication/378943075_Artificial_Intelligence_-_powered_Expert_Systems_in_Hospitals/links/65f2e25932321b2cff78caa5/Artificial-Intelligence-powered-Expert-Systems-in-Hospitals.pdf
3. Voronenko, O.V., Haleliuka, I.B., Romanov, V.O. and Mintser, O.P. (2020), «Technologies of artificial intelligence in medical practice», pp. 17–27, [Online], available at: <http://ir.nuozu.edu.ua:8080/handle/lib/1705>
4. Wankhade, M.S. (2023), «Research Paper on: Doctor Appointment Booking System», [Online], available at: https://www.researchgate.net/publication/371171550_Research_Paper_on_Doctor_Appointment_Booking_System
5. Chunhua, J. and Shuangzhu, Z. (2021), «Doctor Recommendation Model Based on Ontology Characteristics and Disease Text Mining Perspective», [Online], available at: https://www.researchgate.net/publication/353801231_Doctor_Recommendation_Model_Based_on_Ontology_Characteristics_and_Disease_Text_Mining_Perspective
6. Jagadeesh Kannan, R. (2020), «Doctor Finder: Find doctors on the Go», [Online], available at: https://www.researchgate.net/publication/346206050_Doctor_Finder_Find_doctors_on_the_Go
7. *Medical system for patients, doctors, public and private medical institutions*, [Online], available at: <https://helsi.me/>
8. Maimon, O.Z. and Rokach, L. (2014), «Data mining with decision trees: theory and applications», *World scientific*, Vol. 81, [Online], available at: https://books.google.com.ua/books?hl=uk&lr=&id=OVYCCwAAQBAJ&oi=fnd&pg=PR6&dq=5.%09Rokach+L.+DATA+MINING+WITH+DECISION+TREES+Theory+and+Applications&ots=tLieb11fYL&sig=yydce3QYzj8eZmnSW0WV8Y5G08&redir_esc=y#v=onepage&q=5.%09Rokach%20L.%20DATA%20MINING%20WITH%20DECISION%20TREES%20Theory%20and%20Applications&f=false
9. Andrushchak, I.Ye. and Martseniuk, V.P. (2015), «Development and Application of Decision-Making, Optimization, and Management Methods in Systematic Medical Research», [Online], available at: https://scc.knu.ua/upload/iblock/729/dis_Andrushchak%20I.Ye..pdf
10. Fonseca, L., Ribeiro, F., Metrólho, J. et al., «PoPu-Data: A Multilayered, Simultaneously Collected Lying Position Dataset», *Data*, 8 p., [Online], available at: <https://www.mdpi.com/2306-5729/8/7/120>

Кучерук Ольга Віталіївна – здобувач освіти Державного університету «Житомирська політехніка».
<https://orcid.org/0009-0007-0014-4556>.

Наукові інтереси:

- вебтехнології;
- штучний інтелект.

Желізко Віктор Вікторович – асистент кафедри комп'ютерних наук Державного університету «Житомирська політехніка».

<https://orcid.org/0009-0001-4178-6631>.

Наукові інтереси:

- вебтехнології;
- штучний інтелект.

Марчук Галина Вікторівна – старший викладач кафедри комп’ютерних наук Державного університету «Житомирська політехніка».

Наукові інтереси:

- програмування;
- інтелектуальний аналіз баз даних;
- комп’ютерна дискретна математика.

Коротун Ольга Володимирівна – кандидат педагогічних наук, доцент кафедри комп’ютерних наук Державного університету «Житомирська політехніка».

Наукові інтереси:

- бази та сховища даних;
- інтелектуальний аналіз баз даних;
- чисельні методи.

Kucheruk O.V., Zhelizko V.V., Marchuk H.V., Korotun O.V.

Systems for selecting qualified oncology specialists based on medical data analysis

In recent years, decision support systems have become critically important in the healthcare sector, assisting medical professionals in making informed decisions. This article explores the role of decision tree algorithms as one of the leading methods in decision support systems, emphasizing their effectiveness in classification and prediction tasks. Decision trees offer several advantages, including an intuitive structure, ease of interpretation, and the ability to provide clear explanations alongside accurate results. Furthermore, decision trees are noted for their capability to extract meaningful rules from complex datasets, contributing to the discovery of hidden patterns that may be obscured by traditional analytical methods.

The article highlights the versatility of decision trees, making them particularly suitable for a wide range of medical applications, such as finding the most appropriate physician, establishing diagnoses, and determining treatment methods. This capability significantly enhances clinical decision-making processes, enabling more personalized and effective medical care.

As healthcare increasingly adopts data-driven solutions, integrating decision tree algorithms into decision support systems is likely to yield significant benefits, improving the quality of medical practice and ultimately leading to better patient health outcomes. The article concludes with the need for further development and application of decision tree algorithms in medical institutions, emphasizing their potential to revolutionize medical decision-making and support systems in achieving improved health outcomes and physician selection for oncology patients.

Keywords: data analysis; machine learning; decision trees; artificial intelligence; healthcare; algorithms.

Стаття надійшла до редакції 26.09.2024.