

УДК 004.8:37.018.43

DOI [https://doi.org/10.24144/2616-7700.2025.46\(1\).218-225](https://doi.org/10.24144/2616-7700.2025.46(1).218-225)**І. М. Лях¹, В. В. Дудник², Ю. М. Ціпінью³, А. Ю. Ціпінью⁴**

¹ ДВНЗ «Ужгородський національний університет»,
професор кафедри інформатики та фізико-математичних дисциплін,
доктор технічних наук, доцент
igor.lyah@uzhnu.edu.ua
ORCID: <https://orcid.org/0000-0001-5417-9403>

² Полтавський державний аграрний університет,
доцент кафедри механічної та електричної інженерії,
кандидат технічних наук, доцент
volodymyr.dudnyk@pdau.edu.ua
ORCID: <https://orcid.org/0000-0002-6553-2951>

³ Викладач-методист закарпатського політехнічного фахового коледжу,
бакалавр кафедри інформатики та фізико-математичних дисциплін
yuratsipino@ukr.net
ORCID: <https://orcid.org/0009-0008-5494-8418>

⁴ ДВНЗ «Ужгородський національний університет»,
студент 4-ого курсу спеціальності 126 «Інформаційні системи та технології»
tsipino.artemii@student.uzhnu.edu.ua
ORCID: <https://orcid.org/0009-0007-1704-2732>

ПРОГНОЗУВАННЯ ЗАЛУЧЕНОСТІ КОРИСТУВАЧІВ ОСВІТНІХ ВЕБПЛАТФОРМ ЗА ДОПОМОГОЮ АЛГОРИТМІВ МАШИННОГО НАВЧАННЯ

У статті розглянуто технологічні підходи до реалізації навчальних онлайн-платформ із елементами гейміфікації. Сучасний розвиток цифрових технологій висуває нові вимоги до вебзастосунків, зокрема в контексті інтерактивного відображення складних даних і статистичних результатів у режимі реального часу. Це стає особливо важливим у сфері освіти, де актуальними є платформи для дистанційного навчання, які потребують не лише зручного інтерфейсу, а й ефективних інструментів для аналізу та візуалізації навчальних результатів. У статті було розглянуто використання різноманітних методів і алгоритмів для обробки даних, зокрема для прогнозування та аналізу поведінки користувачів онлайн-платформ.

У процесі дослідження проведено аналіз доступних інструментів для інтеграції моделей машинного навчання та аналітичних інструментів у вебзастосунки. Проведено порівняння кількох алгоритмів машинного навчання, зокрема таких як: Random Forest, KNN-регресії, гребеневої регресії та еластичної мережі, для аналізу залученості користувачів на платформі Prometheus. Для прогнозування активності користувачів використовувались реальні аналітичні дані, що дозволило побудувати моделі, які здатні враховувати змінні залежності та обробляти аномальні дані, пов'язані з пандемією COVID-19 та іншими соціально-економічними чинниками. Результати показують, що використання різних методів машинного навчання дозволяє здійснити точний прогноз щодо залученості користувачів, зокрема з урахуванням змін у соціально-демографічному контексті.

Дослідження показало, що інтеграція таких інструментів дозволяє значно покращити ефективність взаємодії користувачів з онлайн-платформами, а також створює умови для подальшої оптимізації навчальних процесів, зокрема через персоналізацію контенту та методів взаємодії.

Ключові слова: онлайн-платформи, машинне навчання, регресія, прогнозування, користувацька активність, цифрові технології.

1. Вступ. З розвитком цифрових технологій постає потреба не лише в створенні інтерактивних вебзастосунків, а й у забезпеченні їхньої здатності до обробки та візуального подання даних у режимі реального часу. Відображення складних статистичних залежностей, аналітичних висновків та поведінкових моделей у зручній формі є критичним для прийняття рішень як на рівні користувача, так і на рівні розробника.

У роботі особливу увагу приділено практичному аналізу даних за допомогою побудови інтерактивних графіків. Аналіз охоплював класифікацію даних, виявлення закономірностей і розглядалися інструменти для інтеграції аналітичних моделей у вебінтерфейси, що дозволяють користувачам без технічного бекграунду отримувати доступ до складної інформації у зручному форматі.

Практичний аспект включає реалізацію вебзастосунку, який надає можливість інтерактивної роботи з результатами класифікації, аналізу поведінки користувачів, а також динамічної побудови графіків відповідно до введених параметрів. Такий підхід дозволяє поєднати автоматизовану обробку даних із високим рівнем користувацької візуалізації.

2. Аналіз останніх досліджень та публікацій. Дослідженням даного питання займалися вітчизняні дослідники, такі як: Савенко О. І., Божок К. Ю. [1, 2], а також зарубіжні дослідники, такі як: Б. Потер, Б. Божкая та інші [3].

Сучасні дослідження акцентують на важливості багаторівневої трансформації освітнього середовища, що включає інфраструктурні, організаційні, програмно-методичні та особистісні компоненти.

У дослідженнях також відзначено, що залучення користувача значною мірою залежить від якості інтерфейсу платформи, її доступності, зручності та рівня інтерактивності. Зокрема, наявність чату, форуму, вікторин, модулів для обговорення та інтерактивних вправ корелює з підвищенням частоти повернень до платформи, а також зі збільшенням середньої тривалості навчальної сесії. Статистично підтверджено, що рівень повторного входу користувачів на платформи з активними елементами взаємодії є на понад 30% вищим, а показники завершення курсів збільшуються при наявності персоналізованих навчальних маршрутів.

У низці публікацій подано порівняльний аналіз функціональностей популярних освітніх платформ, таких як: Khan Academy, Duolingo, Coursera, Edmodo. Виявлено, що найбільше позитивний ефект демонструють системи, які поєднують зовнішню мотивацію (наприклад, винагороди) із внутрішньою (наприклад, інтерес, саморозвиток), а також надають користувачеві свободу у виборі навчального маршруту.

Окрему увагу приділено віковим характеристикам користувачів: елементи мають найвищу ефективність серед користувачів віком 15–35 років, у той час як для старших вікових груп важливішими виявились простота інтерфейсу, стабільність платформи та наявність підтримки.

Вивчення механізмів залученості користувачів на онлайн-платформах є неможливим без урахування соціально-демографічного контексту, в якому функціонує цифрове середовище. Одним із ключових чинників, що впливає як на ефективність комунікації з користувачем, так і на вибір інструментів є вікова структура населення. Адже саме вік часто визначає цифрову компетентність, уподобання щодо форм подачі контенту та рівень адаптивності до нових техно-

логічних рішень.

З огляду на це, доцільним є звернення до актуальних демографічних даних, що дають змогу оцінити вікові особливості населення країни в динаміці. Статистичні дані останніх років дозволяють простежити стійкі тенденції до старіння населення, зменшення питомої ваги молодших вікових груп та скорочення загальної чисельності. Такий аналіз виступає підґрунтям для подальшого дослідження потенціалу інтерактивних вебзастосунків у контексті конкретних цільових аудиторій [4–6].

Для розгляду даних рішень цільової аудиторії, буде використано наступні методи, такі як:

- Random Forest.
- KNN-регресія.
- Гребенева регресія.
- “Еластична мережа”.

Random Forest є одним з найбільш гнучких та потужних ансамблевих методів машинного навчання, що демонструє високу ефективність у задачах регресії. Його суть полягає у побудові множини дерев рішень на випадкових підвбірках з навчальної вибірки, де для кожного дерева випадковим чином відбирається підмножина ознак для розбиття. Підсумковий прогноз формується шляхом усереднення результатів усіх дерев, що забезпечує підвищену стійкість моделі до шумів та перенавчання. Метод об’єднує кілька дерев рішень (ансамбль), де результат — це середнє всіх передбачень:

$$\hat{f} = \frac{1}{B} \sum_{b=1}^1 f_b(x'), \quad (1)$$

де $f_b(x')$ — передбачення b -го дерева;

B — кількість дерев у лісі.

Серед переваг цього методу — здатність ефективно працювати з великою кількістю ознак, нечутливість до масштабування даних, а також можливість реалізації у багатопотоковому середовищі. Попри деякі обмеження, зокрема великий розмір створюваних моделей, Random Forest залишається одним із провідних підходів до побудови прогнозних моделей у контексті складних соціально-демографічних задач.

KNN-регресія (англ. *k*-nearest neighbors) є одним із найпростіших методів машинного навчання, який не потребує попереднього навчання моделі у традиційному розумінні. Алгоритм базується на припущенні, що схожі об’єкти розташовані близько один до одного у багатовимірному просторі ознак. У цьому методі вся навчальна вибірка зберігається безпосередньо, а для прогнозування значення цільової змінної нової точки обираються k найближчих точок із навчального набору відповідно до обраної метрики відстані.

$$\hat{y} = \frac{1}{k} \sum_{i \in N_k(x)} y_i, \quad (2)$$

де $N_k(x)$ — індекси k найближчих сусідів;

y_i — реальне значення мітки (таргета) для i -того сусіда.

Прогноз формується як середнє значення цільової змінної знайдених сусідів, або у зваженій версії — з урахуванням ваг, що залежать від відстані. KNN залишається корисним у випадках, коли важлива локальна структура даних, а модель має бути максимально простою та інтерпретованою.

Гребенева регресія (англ. Ridge Regression) є варіацією лінійної регресії, призначеною для вирішення проблеми мультиколінеарності, яка виникає за наявності сильного кореляційного зв'язку між незалежними змінними. На відміну від звичайної лінійної регресії, гребенева регресія додає до цільової функції додатковий регуляризаційний член, який накладає штраф на величини коефіцієнтів. Це дозволяє зменшити їх дисперсію та уникнути переобучення моделі. Метод доцільно застосовувати, коли кількість предикторів велика, існує високий рівень колінеарності, а також у випадках, коли необхідно зменшити чутливість моделі до варіацій у навчальних даних, забезпечуючи при цьому більш стабільні та інтерпретовані результати.

“Еластична мережа” (Elastic Net) поєднує в собі сильні сторони методів Lasso та гребеневої регресії, застосовуючи як L1-, так і L2-регуляризацію. Такий підхід дозволяє моделі зберігати властивості обох методів: здатність Lasso зануляти деякі коефіцієнти і тим самим здійснювати вибір змінних, а також стабільність та контроль над мультиколінеарністю, притаманну гребеневій регресії. “Еластична мережа” є особливо ефективною у випадках, коли незалежні змінні мають високу кореляцію між собою. На відміну від останнього, Elastic Net не накладає жорстких обмежень на кількість обраних предикторів, що забезпечує більшу гнучкість та адаптивність моделі до різних типів даних [7, 8].

Після розгляду теоретичних аспектів моделей регресійного аналізу, було здійснено практичне застосування обраних методів для прогнозування рівня залученості користувачів освітньої платформи Prometheus. Враховуючи оприлюднені аналітичні дані щодо активності користувачів у 2023 та 2024 році, здійснено оцінку можливих сценаріїв розвитку залученості на 2025 рік [9].

Слід зазначити, що дані, починаючи з 2020 року, демонструють аномальні зміни у динаміці користувацької активності, що пов'язано із переходом на онлайн-формат навчання під час пандемії COVID-19. Показники після 2022 року мають окремі спотворення, зумовлені повномасштабною військовою агресією проти України, яка вплинула на доступ до інфраструктури і на поведінку користувачів у цифровому середовищі.

Першим було застосовано метод Random Forest, що дозволяє враховувати нелінійні залежності між змінними та є стійким до перенавчання. На графіку нижче представлено результат прогнозу, отриманого за допомогою цього методу.

На зображенні видно загальну динаміку зростання та реакцію моделі на зміни в історичних даних. Приріст нових користувачів 2024–2025 рр. у відсотковому співвідношенні складає 0.41%.

Наступним підходом стала KNN-регресія, яка враховує подібність між точками у просторі ознак і базується на прогнозуванні значення цільової змінної з урахуванням найближчих сусідів.

Далі було протестовано гребеневу регресію, яка додає до цільової функції штраф за L2-норму коефіцієнтів. Приріст нових користувачів 2024–2025 рр. у відсотковому співвідношенні складає близько 0.001%, що є не значним.

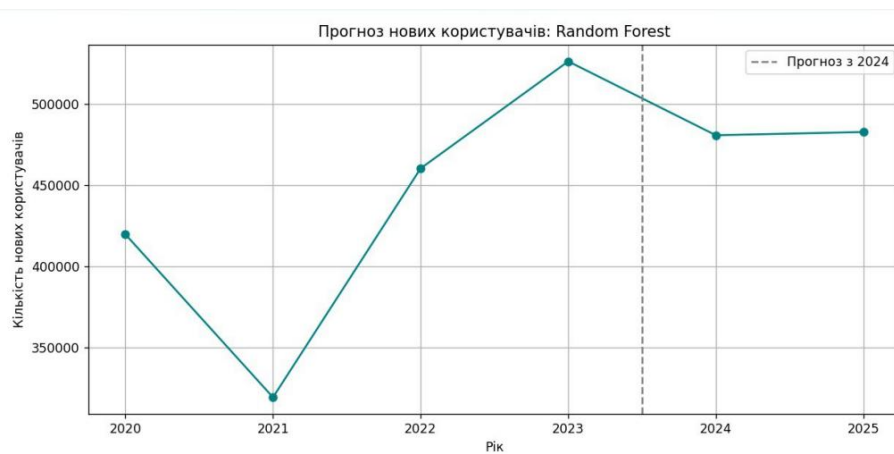


Рис. 1. Графік прогнозування нових користувачів за Random Forest.

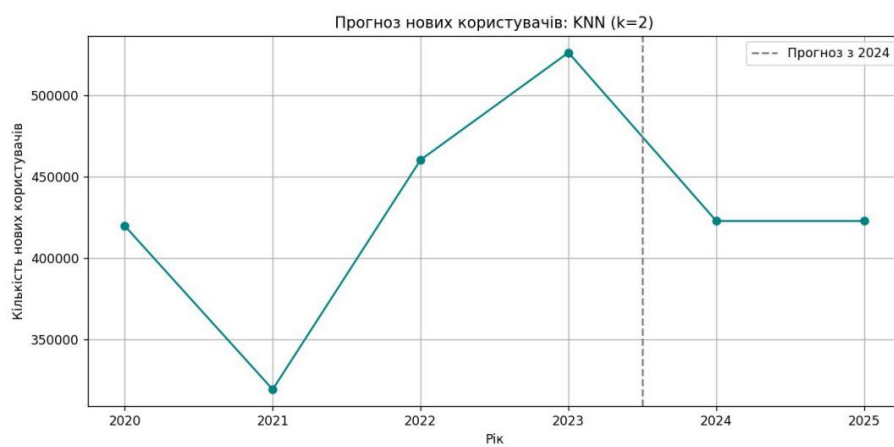


Рис. 2. Графік прогнозування нових користувачів за KNN-регресією.

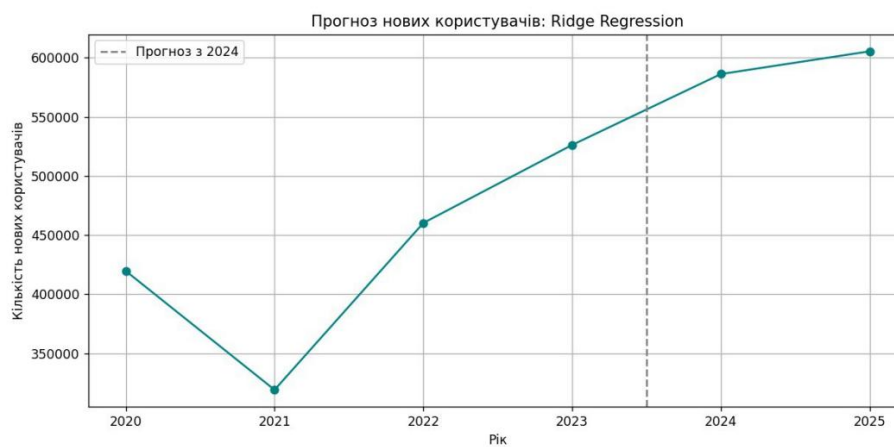


Рис. 3. Графік прогнозування нових користувачів за гребеневою регресією.

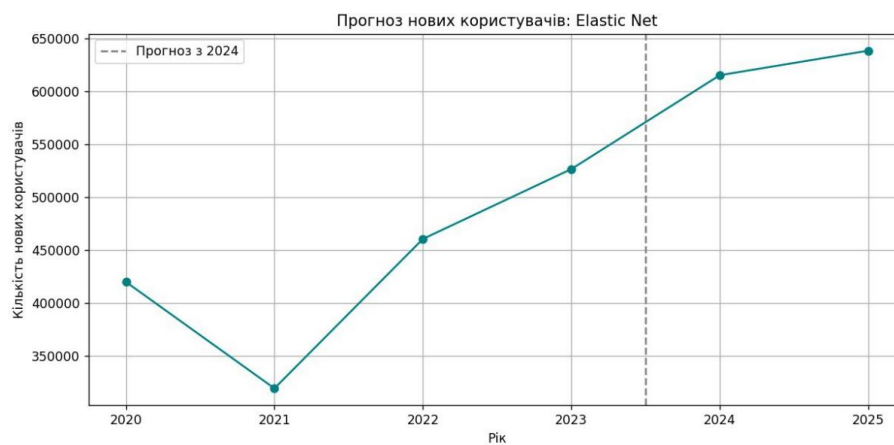


Рис. 4. Графік прогнозування нових користувачів за еластичною мережею.

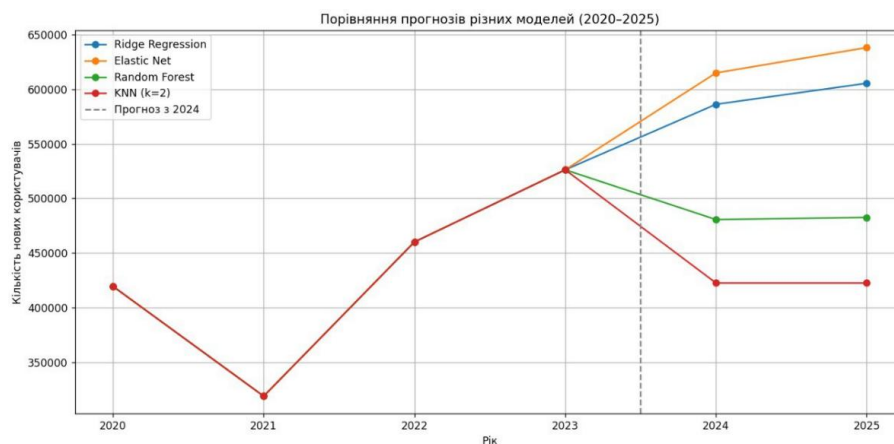


Рис. 5. Загальний графік прогнозування нових користувачів.

Графік демонструє згладжений тренд, модель менш чутлива до шуму, однак може недооцінювати пікові значення. Останнім методом стала еластична мережа, що об'єднує переваги L1 та L2 регуляризацій. Приріст нових користувачів 2024–2025 рр. у відсотковому співвідношенні складає близько 3.28%.

Для наочного порівняння результатів усіх методів нижче наведено підсумковий графік, на якому візуалізовано прогнози, отримані кожним із підходів.

Серед протестованих методів найбільш точним виявився Random Forest. Попри обмеженість і відносну вузькість доступного набору даних, ця модель продемонструвала високу стійкість до варіативності, зберігаючи здатність коректно виявляти приховані залежності між ознаками. Її ансамблева природа дозволила уникнути перенавчання, що особливо важливо в умовах обмеженої кількості спостережень, та забезпечити найбільш стабільний і достовірний прогноз.

3. Висновки. У ході дослідження було встановлено, що ефективність онлайн-платформ значною мірою залежить від рівня інтерактивності, персоналізації контенту та можливості адаптації до соціально-демографічних характеристик користувачів. Аналіз сучасних досліджень підтвердив актуальність

застосування моделей машинного навчання для прогнозування користувацької залученості, зокрема в умовах цифрових трансформацій та змін у структурі цільової аудиторії. Особливої уваги заслуговує метод Random Forest, який продемонстрував найвищу точність прогнозу завдяки своїй здатності ефективно обробляти великі обсяги даних із нелінійними залежностями.

Практичне застосування чотирьох методів регресійного аналізу — Random Forest, KNN-регресії, гребеневої регресії та еластичної мережі — засвідчило різний рівень ефективності кожного підходу щодо прогнозування динаміки користувачів освітньої платформи Prometheus. Найвищий приріст прогнозованої залученості у 2025 році показала модель еластичної мережі, однак з точки зору стабільності та стійкості до змін у структурі даних, найнадійнішими виявилися результати, отримані за допомогою Random Forest. Таким чином, подальші дослідження доцільно спрямовувати на вдосконалення гібридних моделей із урахуванням соціально-демографічного контексту та специфіки поведінки користувачів у динамічному цифровому середовищі.

Список використаної літератури

1. Лях І. М., Савенко О. І., Чобаль В. В., Гладкий К. І. Використання інтерактивних методів навчання: шлях до ефективної освіти. Актуальні проблеми економіки. 2025. Т. 286, № 4. С. 14–20.
2. Божок К. Ю. Огляд цифрових платформ для надання освітніх послуг з елементами гейміфікації. *Наукові барви* : Збірник тез доповідей учасників III Всеукраїнської наукової Інтернет-конференції. 8 листопада, 2024. Київ : КНЕУ, 2024. С. 26–28.
3. Porter B., Bozkaya B. Assessing the Effectiveness of Using Live Interactions and Feedback to Increase Engagement in Online Learning. 2020. arxiv. DOI: <https://doi.org/10.48550/arXiv.2008.08241>
4. Статистичний збірник наявного населення 2022. URL: <http://db.ukrcensus.gov.ua/PXWEB2007> (дата звернення: 25.02.2025).
5. Національні демографічні прогнози. Національної академії наук України. URL: https://idss.org.ua/forecasts/nation_pop_proj (дата звернення: 23.02.2025).
6. Ланда В., Ревук М. Три важливі демографічні показники. 2023-го народилося найменше дітей за останні 300 років. 2024. URL: <https://texty.org.ua/articles/112194/try-vazhlyvi-demohrafichni-pokaznyky-2023-ho-narodylosya-najmenshe-ditej-za-ostanni-300-rokiv> (дата звернення: 25.02.2025).
7. Noi P. T., Kappas M. Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery. *Multidisciplinary Digital Publishing Institute*. 2017. Vol. 18, No. 1. 18. DOI: <https://doi.org/10.3390/s18010018>
8. What is lasso regression? IBM URL: <https://www.ibm.com/think/topics/lasso-regression> (date of access: 24.02.2025).
9. Сабадишина Ю. Лише третина учасників завершує безплатні курси, найбільше слухачів — віком 30+. Prometheus про онлайн-навчання українців. 2024. DOU. URL: <https://dou.ua/lenta/news/prometheus-in-2023> (дата звернення: 25.02.2025).

Liakh I. M., Dudnyk V. V., Tsipino Y. M., Tsipino A. Y. Forecasting user engagement in educational web platforms using machine learning algorithms.

The article examines technological approaches to the implementation of educational online platforms incorporating gamification elements. The current development of digital technologies imposes new requirements on web applications, particularly in the context of interactive visualization of complex data and statistical results in real time. This is especially relevant in the field of education, where distance learning platforms must offer not only user-friendly interfaces but also effective tools for analyzing and visualizing learning outcomes. The article explores the use of various methods and algorithms for data

processing, specifically for predicting and analyzing user behavior on online platforms.

The study involved an analysis of available tools for integrating machine learning models and analytical instruments into web applications. Several machine learning algorithms, including Random Forest, KNN regression, Ridge regression, and Elastic Net, were compared in the context of analyzing user engagement on the Prometheus platform. Real analytical data were used to forecast user activity, enabling the development of models capable of accounting for variable dependencies and handling anomalous data caused by the COVID-19 pandemic and other socio-economic factors. The results demonstrate that the application of diverse machine learning methods enables accurate forecasting of user engagement, particularly when considering shifts in the socio-demographic landscape.

The research indicates that integrating such tools significantly enhances the effectiveness of user interaction with online platforms and establishes a foundation for further optimization of learning processes, particularly through content personalization and adaptive interaction methods.

Keywords: online platforms, machine learning, regression, forecasting, user activity, digital technologies.

References

1. Lyakh, I. M., Savenko, O. I., Chobal, V. V., & Gladky, K. I. (2025). Using interactive teaching methods: the way to effective education. *Current problems of economics*, 286(4), 14–20.
2. Bozhok, K. Yu. (November 8, 2024). Review of digital platforms for providing educational services with elements of gamification. In *Scientific colors: Collection of abstracts of reports of participants of the III All-Ukrainian Scientific Internet Conference*. Kyiv: KNEU, 26–28.
3. Porter, B., & Bozkaya, B. (2020). Assessing the Effectiveness of Using Live Interactions and Feedback to Increase Engagement in Online Learning. arxiv. <https://doi.org/10.48550/arXiv.2008.08241>
4. Statistical collection of the current population. (2022). Retrieved from <http://db.ukrcensus.gov.ua/PXWEB2007>
5. National Demographic Forecasts. National Academy of Sciences of Ukraine. Retrieved from https://idss.org.ua/forecasts/nation_pop_proj
6. Landa, V., & Revuk, M. (2024). Three important demographic indicators. In 2023, the fewest children were born in the last 300 years. Retrieved from <https://texty.org.ua/articles/112194/try-vazhlyvi-demohrafichni-pokaznyky-2023-ho-narodylosya-najmenshe-ditej-za-ostanni-300-rokiv>
7. Noi, P. T., & Kappas, M. (2017). Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery. *Multidisciplinary Digital Publishing Institute*, 18(1), 18. <https://doi.org/10.3390/s18010018>
8. What is lasso regression? IBM Retrieved from <https://www.ibm.com/think/topics/lasso-regression>
9. Sabadyshyna, Yu. (2024). Only a third of participants complete free courses, most of the students — aged 30+. Prometheus about online learning of Ukrainians. DOU. Retrieved from <https://dou.ua/lenta/news/prometheus-in-2023>

Одержано 18.03.2025