

УДК 004.89:37.018.43

DOI <https://doi.org/10.32689/maup.it.2025.2.10>

Євгеній КЛИМЕНКО

здобувач вищої освіти за спеціальністю «Комп'ютерні науки»,
Національний університет біоресурсів і природокористування України,
ye.klymenko@nubip.edu.ua
ORCID: 0009-0006-6353-6015

Олена ГЛАЗУНОВА

доктор педагогічних наук, професор,
проректор з науково-педагогічної роботи та цифрової трансформації,
Національний університет біоресурсів і природокористування України,
o-glazunova@nubip.edu.ua
ORCID: 0000-0002-0136-4936

АРХІТЕКТУРА ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ ОСВІТНЬОЇ АНАЛІТИКИ З ВИКОРИСТАННЯМ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ

Анотація. У статті розглянуто підходи до побудови архітектури інформаційної технології освітньої аналітики з інтеграцією методів інтелектуального аналізу даних. Запропоновано концептуальну модель аналітичної системи, яка дозволяє виявляти залежності між активністю студентів у навчальному середовищі та їхньою академічною успішністю, а також визначати часові патерни відвідуваності. Застосування кластерного аналізу дало змогу виокремити типові поведінкові групи студентів, що відкриває перспективи для персоналізації навчання та раннього виявлення ризиків. Результати дослідження можуть бути використані для розробки ефективних систем підтримки прийняття рішень у закладах вищої освіти.

Мета роботи. Розробка концептуальної архітектури інформаційної технології освітньої аналітики, яка інтегрує методи інтелектуального аналізу даних для виявлення прихованих закономірностей у поведінці студентів, прогнозування академічної успішності та підтримки прийняття управлінських рішень у сфері вищої освіти. Особливу увагу приділено застосуванню алгоритмів машинного навчання, кластерного аналізу, регресійного моделювання та візуалізації даних у контексті побудови ефективної, масштабованої та адаптивної аналітичної системи.

Методологія. Методологічну основу дослідження становить міждисциплінарний підхід, що поєднує принципи освітньої аналітики, інтелектуального аналізу даних та архітектурного проектування інформаційних систем.

Наукова новизна дослідження полягає в інтеграції модулів візуалізації, прогнозування та підтримки прийняття рішень у єдину архітектуру, орієнтовану на потреби закладів вищої освіти та практичній апробації моделі на реальних даних з LMS, що підтверджує її ефективність та адаптивність до умов сучасного освітнього середовища.

Висновки. Доведено, що розробка інформаційних технологій на основі використання методів інтелектуального аналізу даних при впровадженні систем електронного навчання сприяє вирішенню задач, пов'язаних із розумінням поведінки студентів, поліпшенням якості електронних курсів, вдосконаленням методик навчання, зменшенням витрат на організацію процесу навчання та визначає подальші напрями освітньої аналітики відповідно до загально-світових тенденцій.

Ключові слова: освітня аналітика, інтелектуальний аналіз даних, архітектура інформаційної системи, машинне навчання, кластеризація, прогнозування успішності, навчальна активність, системи підтримки прийняття рішень, візуалізація освітніх даних, Learning Management System (LMS).

Yevhenii KLYMENKO, Olena HLAZUNOVA. ARCHITECTURE OF EDUCATIONAL ANALYTICS INFORMATION TECHNOLOGY USING DATA MINING METHODS

Abstract. The article explores approaches to designing the architecture of educational analytics information technology with the integration of data mining methods. A conceptual model of an analytical system is proposed, enabling the identification of relationships between student activity in the learning environment and their academic performance, as well as the detection of temporal attendance patterns. The application of cluster analysis made it possible to distinguish typical behavioral groups of students, opening prospects for personalized learning and early risk detection. The research results can be used to develop effective decision support systems in higher education institutions.

Purpose of the Study. To develop a conceptual architecture of educational analytics information technology that integrates data mining methods to identify hidden patterns in student behavior, predict academic performance, and support decision-making in higher education. Special attention is paid to the application of machine learning algorithms, cluster analysis, regression modeling, and data visualization in the context of building an effective, scalable, and adaptive analytical system.

Methodology. The methodological basis of the study is an interdisciplinary approach that combines the principles of educational analytics, data mining, and information systems architecture design.

© Є. Клименко, О. Глазунова, 2025

Стаття поширюється на умовах ліцензії CC BY 4.0

Scientific Novelty. The scientific novelty of the study lies in the integration of visualization, forecasting, and decision support modules into a unified architecture tailored to the needs of higher education institutions, as well as in the practical validation of the model using real LMS data, which confirms its effectiveness and adaptability to the conditions of the modern educational environment.

Conclusions It has been proven that the development of information technologies based on data mining methods in the implementation of e-learning systems contributes to solving tasks related to understanding student behavior, improving the quality of online courses, enhancing teaching methods, reducing the cost of organizing the learning process, and defining future directions for educational analytics in line with global trends.

Key words: educational analytics, data mining, information system architecture, machine learning, clustering, performance prediction, learning activity, decision support systems, educational data visualization, Learning Management System (LMS).

Постановка проблеми. У сучасних умовах цифрової трансформації освіти зростає потреба у впровадженні інноваційних інформаційних технологій, здатних забезпечити ефективне управління освітнім процесом, моніторинг академічної успішності та прийняття обґрунтованих рішень на основі даних. Одним із ключових напрямів у цьому контексті є освітня аналітика, яка поєднує методи збору, обробки та інтерпретації освітніх даних з метою підвищення якості навчання та управління освітніми системами. Особливої актуальності набуває розробка архітектури інформаційної технології освітньої аналітики, яка б інтегрувала інструменти інтелектуального аналізу даних у єдину систему підтримки прийняття рішень. Такий підхід дозволяє не лише виявляти приховані закономірності у поведінці студентів, а й прогнозувати ризики академічної неуспішності, оптимізувати розклад занять, персоналізувати навчальні траєкторії та підвищувати ефективність освітніх стратегій.

З огляду на стрімке зростання обсягів освітніх даних, що генеруються в системах управління навчанням (LMS), електронних журналах, платформах дистанційного навчання тощо, постає необхідність у створенні гнучкої, масштабованої та безпечної архітектури, здатної забезпечити повний цикл аналітичної обробки даних – від збору до візуалізації результатів. Таким чином, дослідження у сфері архітектури інформаційних технологій освітньої аналітики є надзвичайно актуальним і має вагоме практичне значення для розвитку сучасної освіти.

Аналіз останніх досліджень та публікацій. Упродовж останнього десятиліття освітня аналітика (Learning Analytics, LA) та інтелектуальний аналіз освітніх даних (Educational Data Mining, EDM) стали ключовими напрямками досліджень у сфері цифрової трансформації освіти. Їх мета полягає у виявленні закономірностей у поведінці студентів, прогнозуванні академічної успішності та підтримці прийняття рішень на основі даних. Так, Romero і Ventura [4] у своєму огляді підкреслюють, що сучасні системи освітньої аналітики дедалі частіше інтегрують алгоритми машинного навчання, зокрема кластеризацію, регресійне моделювання та нейронні мережі, для аналізу великих обсягів освітніх даних. Вітчизняні науковці також акцентують увагу на важливості побудови адаптивних архітектур, здатних до масштабування та інтеграції з LMS-платформами [1-3] Papamitsiou та Economides [9] провели систематичний аналіз емпіричних досліджень у галузі LA та EDM, виявивши, що більшість робіт зосереджені на аналізі відвідуваності, активності в LMS та оцінювання. Вони наголошують на потребі у створенні гнучких архітектур, які дозволяють поєднувати різні джерела даних та забезпечують візуалізацію результатів для викладачів і адміністраторів. Інші дослідники, зокрема [7-8; 10-11] розглядають використання аналітики для підтримки академічної успішності у вищій освіті. У своїх роботах вони наводять приклади використання інтелектуального аналізу освітніх даних, пропонують концептуальну модель архітектури освітньої аналітики, яка включає модулі збору, обробки, аналізу та візуалізації даних. Особливу увагу приділено питанням етичності, конфіденційності та інтерпретованості моделей. Крім того, IEEE та Springer [5; 10] регулярно публікують результати досліджень, присвячених архітектурним рішенням у сфері освітньої аналітики. Зокрема, у матеріалах конференцій ICDM [6] представлено приклади реалізації аналітичних платформ, що використовують алгоритми класифікації, кластеризації та прогнозування для аналізу поведінки студентів.

Таким чином, аналіз літератури свідчить про зростаючий інтерес до побудови архітектур освітньої аналітики, що базуються на методах інтелектуального аналізу даних. Проте, залишається відкритим питання щодо уніфікації підходів до архітектурного проектування таких систем, що й обумовлює актуальність подальших досліджень у цьому напрямі.

Метою статті є розробка концептуальної архітектури інформаційної технології освітньої аналітики, яка інтегрує методи інтелектуального аналізу даних для виявлення прихованих закономірностей у поведінці студентів, прогнозування академічної успішності та підтримки прийняття управлінських рішень у сфері вищої освіти. Особливу увагу приділено застосуванню алгоритмів машинного навчання, кластерного аналізу, регресійного моделювання та візуалізації даних у контексті побудови ефективної, масштабованої та адаптивної аналітичної системи.

Виклад основного матеріалу. Відповідно до сучасних міжнародних тенденцій до більшої автономії закладів освіти, використання освітніх даних для інформування про прийняття рішень щодо

прогнозування освітніх траєкторій здобувачів освіти, підзвітності та самовдосконалення є критичним питанням. Інтелектуальний аналіз освітніх даних став обов'язковим напрямом дослідження завдяки багатьом перевагам, яких можуть досягти навчальні заклади, використовуючи інформаційні технології освітньої аналітики.

Завдяки появі нових систем, технологій, пристроїв і засобів зв'язку кількість даних, що виробляються, стрімко зростає з кожним роком. Великі дані включають дані, створені різними пристроями та програмами, як-от дані освітніх сервісів та платформ управління навчанням або дані пошукових систем. Зокрема, джерела надходження даних про освітній процес узагальнено на (рис. 1).

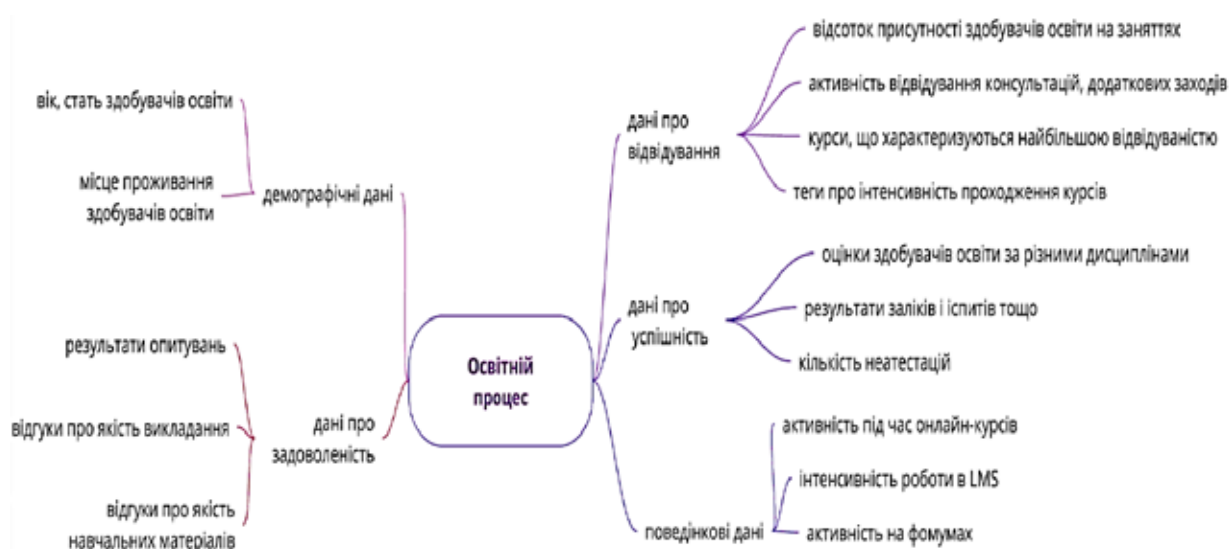


Рис. 1. Канали надходження даних про освітній процес

У нашому дослідженні для аналітики освітнього процесу використовувались дані з 3 джерел:

- Програмний комплекс ЄДЕБО містить дані про здобувача освіти (персональні дані особи – ПІБ та дата народження, дані документа, що посвідчує особу, документів про освіту, реєстраційний номер облікової картки платника податків (РНОКПП), строки навчання, документи про освіту, послідовність здобуття освіти.
- LMS закладу освіти (Learning Management system – moodle) збирають і зберігають дані, пов'язані з опануванням студентами змісту курсів, оцінюванням і обговореннями, надаючи дані про залученість студентів, результативність освітнього процесу і навчальну поведінку.
- Інформаційна система «Деканат» – аналітика індивідуальної освітньої траєкторії та успішності студентів.

Метою закладів освіти в рамках надання освітніх послуг, зокрема і НУБІП України є підвищення якості освіти та успішності здобувачів. Тому підбір інструментів для аналітики освітніх даних з цифрових платформ освітнього середовища є важливим етапом інформаційних технологій та запорукою отримання валідних прогностичних моделей.

Проектування архітектури інформаційної технології освітньої аналітики здійснюється з урахуванням вимог до функціональної повноти, масштабованості, інтероперабельності та зручності обробки даних із різних освітніх джерел. Загальна архітектура системи базується на модульному принципі, що дозволяє незалежно розробляти, впроваджувати та оновлювати її компоненти без порушення цілісності системи.

Система включає наступні основні модулі (рис. 2):

1. **Модуль збору та імпорту даних** забезпечує підключення до зовнішніх джерел даних: систем управління навчанням (LMS, зокрема Moodle); єдиної державної електронної бази з питань освіти (ЄДЕБО); автоматизованої системи управління «Деканат»; відкритих державних або внутрішніх реєстрів ЗВО.

2. **Модуль зберігання та обробки даних** виконує завдання з очищення, нормалізації, агрегації та підготовки даних для подальшого аналізу. Реалізується на базі бази даних типу Data Warehouse.

3. **Модуль інтелектуального аналізу** реалізує алгоритми машинного навчання (класифікація, регресія, кластеризація, асоціативний аналіз) та інтерпретаційні моделі (SHAP) для виявлення закономірностей та прогнозування освітніх результатів.

4. **Аналітичний модуль візуалізації** побудований на основі платформи Power BI. Здійснює побудову дашбордів, інтерактивних графіків, карт, зведених таблиць із показниками успішності, активності та якості взаємодії у навчальному середовищі.

5. **Модуль управління доступом і безпеки** відповідає за автентифікацію користувачів, управління ролями, контроль доступу до окремих показників та персоналізованих даних.

6. **Інтерфейс інтеграції з користувачами (Dashboard/API)** – надає можливості взаємодії з користувачем для зовнішніх систем.

Прототип відповідної архітектури запропонований на (рис. 2).

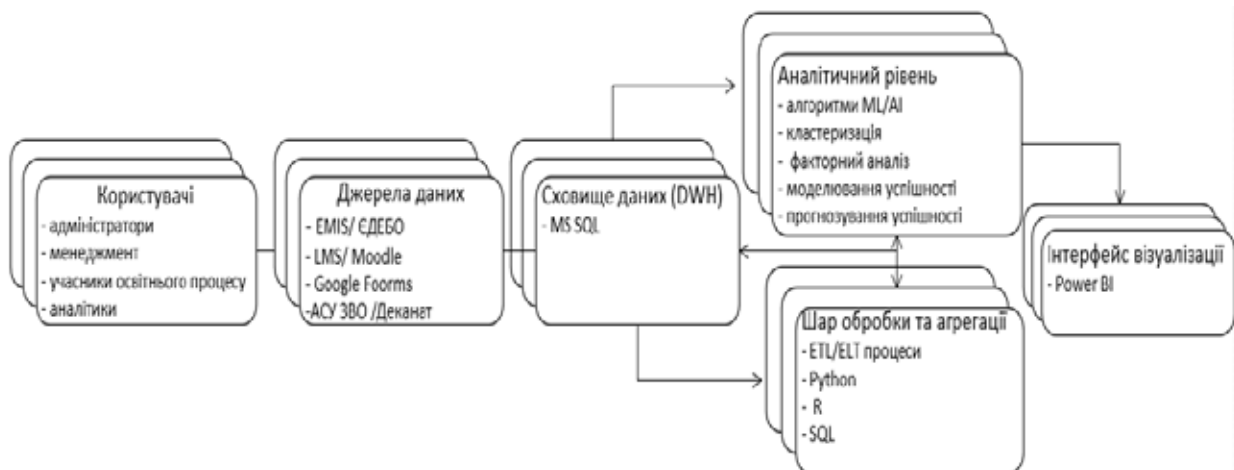


Рис. 2. Прототип інформаційної технології освітньої аналітики

Тут, оскільки нашою відправною точкою є великі дані, зібрані з LMS, ЄДЕБО та інформаційних сервісів та систем НУБІП України, використання інструментів аналітики є неминучим для досягнення нашої кінцевої мети – виявити недоліки в процесах навчання та спробувати надавати звіти та прогнози для здобувачів освіти.

Центральним елементом системи є власне аналітичний рівень, на якому відбувається вся робота з даними. Аналіз даних – це процес дослідження, очищення, перетворення та моделювання великих наборів даних з метою виявлення прихованих закономірностей, невідомих кореляцій і збору корисної інформації.

Інформаційна технологія аналітики навчання, динамічний інструмент у сфері освіти, швидко розвивається і стає незамінним етапом освітніх стратегій у сучасних закладах освіти. Ця концепція, яка заглиблюється в аналіз і застосування даних у навчанні, є трансформаційним підходом до покращення результатів навчання. Прогнозування академічної успішності студентів є важливим завданням для будь-якої установи для прийняття обґрунтованих рішень щодо прогресу студентів у навчанні. Алгоритми інтелектуального аналізу даних можуть бути використані для прогнозування успішності студентів та визначення факторів, що впливають на точність прогнозу. На (рис. 3) показано методологію реалізації запропонованої роботи.

У межах дослідження, спрямованого на аналіз освітньої діяльності та прогнозування академічної успішності студентів, було використано емпіричні дані, отримані з системи управління навчанням (Learning Management System, LMS) факультету інформаційних технологій Національного університету біоресурсів і природокористування України. Вибірка охоплює студентів з першого по четвертий курс, які навчаються за різними освітніми програмами факультету, і налічує 74 427 записів. Зібраний масив даних включає широкий спектр змінних, зокрема: ідентифікатори студентів та навчальних груп, прізвище, ім'я та по батькові, код і назву дисципліни, дату та номер заняття, інформацію про відвідування, нормалізований ідентифікатор, стать, джерело фінансування навчання (бюджет або контракт), курс навчання, конкурсний бал при вступі, кількість входів до LMS-курсу, а також поточні академічні результати.

Структура та обсяг даних забезпечують можливість проведення як описового статистичного аналізу освітнього процесу, так і побудови моделей прогнозування академічної успішності із застосуванням сучасних методів машинного навчання. У ході дослідження було сформовано першу аналітичну модель, метою якої є виявлення залежності між рівнем активності студентів у системі управління навчанням (визначеним за кількістю входів на курс) та отриманими ними підсумковими балами з дисципліни.

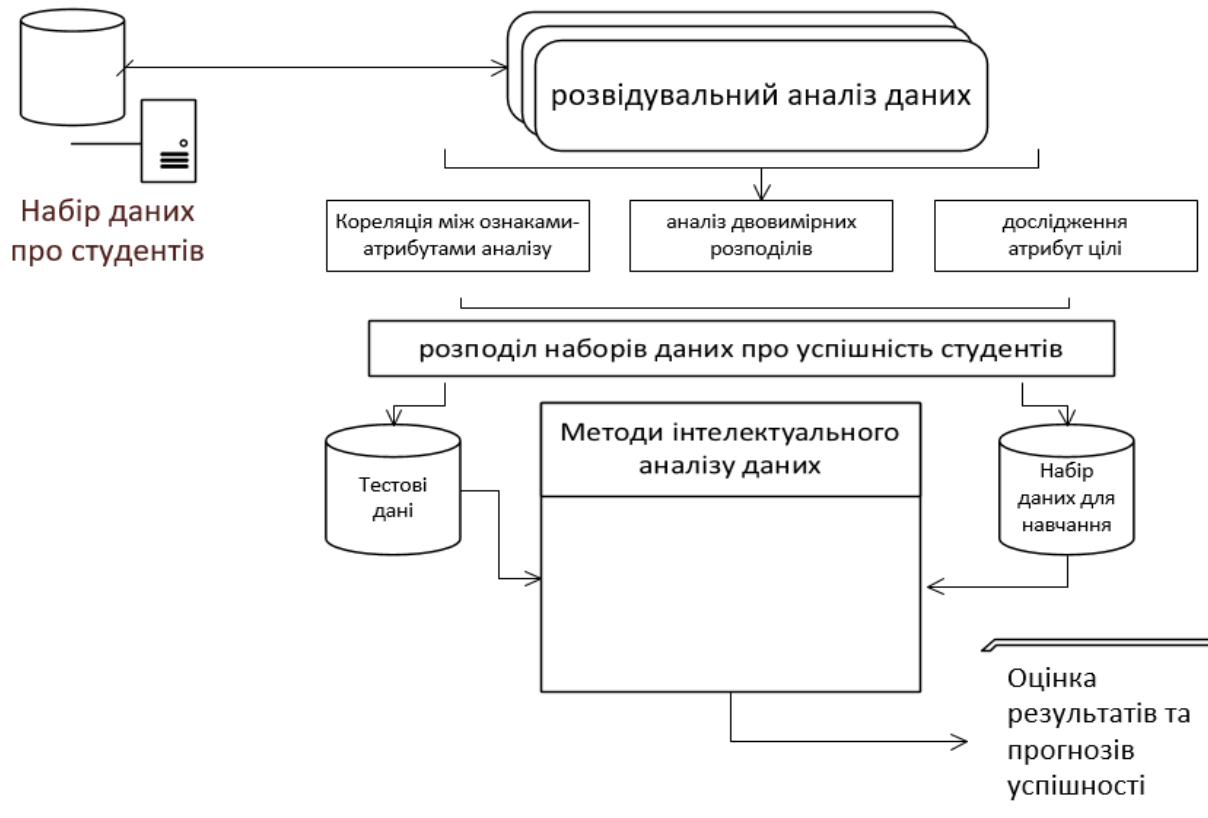


Рис. 3. Блок-схема методології побудови прогностичного моделювання для прогнозування академічної успішності студентів

З метою підвищення точності оцінювання моделі з аналізу були виключені записи студентів із нульовими балами, оскільки такі результати, як правило, є наслідком неявки на підсумковий контроль або нездачі іспиту/заліку. Для візуалізації виявленої залежності було побудовано діаграму розсіювання, де по осі абсцис (X) відображено кількість входів студента на курс у LMS, а по осі ординат (Y) – його підсумковий бал з відповідної дисципліни.

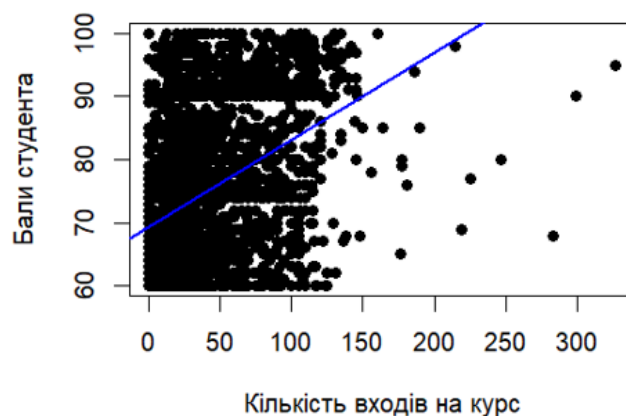


Рис. 4. Характер впливу навчальної активності (кількості входів на відповідний навчальний курс в LMS) на успішність здобувачів освіти

Побудований на основі відібраних даних графік демонструє наявність певної закономірності між активністю студентів у системі управління навчанням (вимірюваною кількістю входів на курс) та їхніми підсумковими балами з дисципліни. На графіку також чітко простежується лінія лінійної регресії з позитивним нахилом, що свідчить про тенденцію до зростання академічної успішності зі збільшенням активності у навчальному курсі (після виключення записів із нульовими балами).

У межах другої аналітичної моделі було досліджено залежність між відвідуваністю студентів та часовими характеристиками проведення занять – зокрема, місяцем та годинаю. Такий підхід дозволив виявити часові патерни, які можуть мати практичне значення при плануванні навчального процесу.

Результати аналізу підтвердили наявність стійких закономірностей у поведінці студентів. Після виключення аномальних, основні тенденції залишилися незмінними, що свідчить про їхню стабільність (рис. 5).

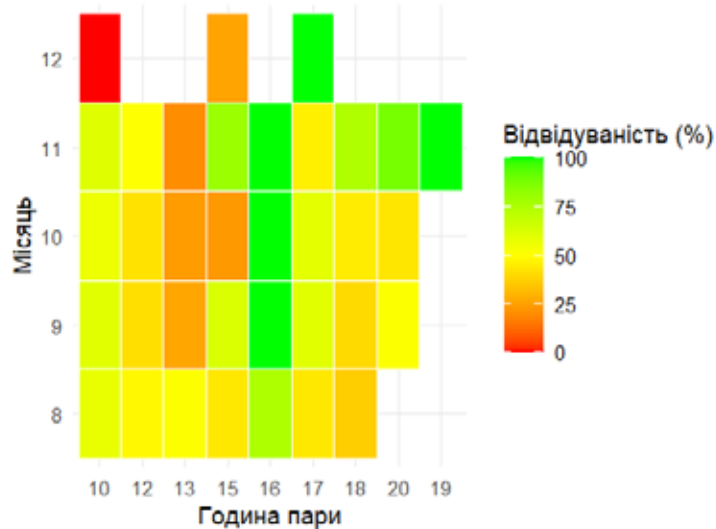


Рис. 5. Рівень відвідуваності ресурсів в LMS

З метою поглибленого аналізу поведінкових моделей студентів у контексті навчальної активності та академічної успішності було застосовано метод кластерного аналізу – алгоритм k-середніх (k-means). Для побудови моделі кластеризації було обрано два ключові параметри: кількість відвідувань занять та середній бал студента за курс. Попередньо з вибірки було виключено записи з нульовими середніми балами, оскільки вони, як правило, пов'язані з академічною заборгованістю або відсутністю оцінювання, що могло б спотворити результати кластеризації (рис. 6).

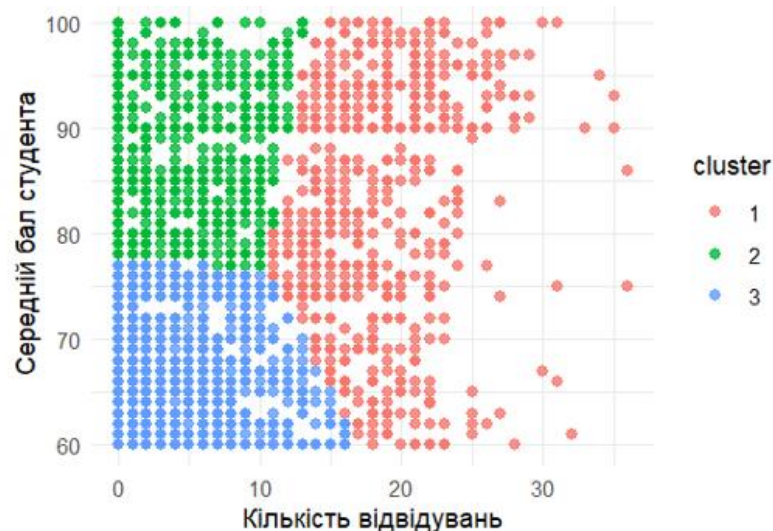


Рис. 6. Кластеризація студентів за відвідуваністю та балами успішності

Аналіз дозволив виокремити три основні кластери студентів, кожен з яких характеризується специфічними особливостями навчальної поведінки:

- Кластер 1 (червоний) – найбільш чисельна група, що включає студентів із високим рівнем відвідуваності та середніми або високими академічними результатами. Це свідчить про позитивну кореляцію між регулярною присутністю на заняттях та успішністю. Водночас у цій групі трапляються

студенти з нижчими балами, що може вказувати на вплив додаткових чинників, таких як складність дисципліни або рівень мотивації.

- Кластер 2 (зелений) – студенти з низькою відвідуваністю, але високими середніми балами. Ця категорія демонструє здатність до самостійного опанування матеріалу, ймовірно, за допомогою альтернативних освітніх ресурсів (електронні курси, відеолекції тощо). Наявність такого кластеру підкреслює необхідність диференційованого підходу до оцінювання ефективності навчання.

- Кластер 3 (синій) – студенти з низькою відвідуваністю та низькими академічними результатами. Ця група є найбільш вразливою з точки зору ризику академічної неупішності та потребує цілеспрямованої педагогічної підтримки.

Такий розподіл дозволяє глибше зрозуміти різноманіття навчальних стратегій студентів та ідентифікувати потенційні зони ризику, що є важливим для розробки індивідуалізованих освітніх підходів.

Висновки. У статті було розглянуто концептуальні та практичні аспекти побудови архітектури інформаційної технології освітньої аналітики з інтеграцією методів інтелектуального аналізу даних. Проведений аналіз літературних джерел засвідчив зростаючий інтерес наукової спільноти до застосування алгоритмів машинного навчання, кластеризації, регресійного моделювання та візуалізації даних у сфері освіти.

Запропонована аналітична модель дозволила виявити залежності між активністю студентів у системі управління навчанням та їхньою академічною успішністю, а також часові патерни відвідуваності занять. Застосування кластерного аналізу дало змогу виокремити типові поведінкові групи студентів, що відкриває перспективи для персоналізації навчального процесу та раннього виявлення ризиків академічної неупішності.

Результати дослідження підтверджують доцільність створення гнучкої, масштабованої архітектури освітньої аналітики, яка поєднує збір, обробку, аналіз і візуалізацію освітніх даних. Така система може стати ефективним інструментом підтримки прийняття рішень у закладах вищої освіти, сприяти підвищенню якості навчального процесу та забезпечити умови для впровадження адаптивного навчання.

Подальші дослідження доцільно спрямувати на розширення функціональності аналітичної системи, інтеграцію з іншими освітніми платформами, а також на вивчення етичних аспектів використання освітніх даних.

Список використаних джерел:

1. Глазунова О. Г., Клименко Є. О., Волошина Т. В., Мокрієв М. В., Вороненко О. В. Освітня аналітика в університетах: інструменти для аналізу та прогнозування. *Телекомунікаційні та інформаційні технології*. 2024. № 2(83) с. 49–59 <https://doi.org/10.31673/2412-4338.2024.026171>
2. Скрипник А., Клименко Н., Костенко І. Рівень освіченості населення в галузі цифрових технологій та зростання економік країн. *Інформаційні технології і засоби навчання*. 2020. № 4 с. 278–297 <https://doi.org/10.33407/itl.v78i4.2948>
3. Hlazunova Olena, Klymenko Nataliia, Mokriev Maksym, Nehrey Maryna, Klymenko Yevhenii. Data Analysis Technologies for Enhanced Educational Processes: A Case Study Using the Moodle LMS. *Lecture Notes on Data Engineering and Communications Technologies*. 2025. Vol. 242, Pages 670–682
4. Romero C., Ventura S. Educational data mining and learning analytics: An updated survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2020. 10(3), e1355. <https://doi.org/10.1002/widm.1355>
5. Eighth IEEE International Conference on Data Mining. *IEEE Xplore*. 2008. URL: <https://ieeexplore.ieee.org/xpl/conhome/4781077/proceeding>
6. IEEE International Conference on Data Mining (ICDM). *IEEE Xplore*. 2025. (July 11, 2025), URL: <https://ieeexplore.ieee.org/xpl/conhome/1000179/all-proceedings>
7. Ifenthaler D., Yau J. Y.-K. Utilising learning analytics to support study success in higher education: A systematic review. *Educational Technology Research and Development*, 2020. 68, 1961–1990. <https://doi.org/10.1007/s11423-020-09788-z>
8. Okike E., Morogosi M. Educational Data Mining for Monitoring and Improving Academic Performance at University Levels. *International Journal of Advanced Computer Science and Applications*. 2020. № 11. DOI: 10.14569/IJACSA.2020.0111171.
9. Papamitsiou Z., Economides A. A. Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence. *Educational Technology & Society*, 2014. 17(4), 49–64. URL: <https://www.jstor.org/stable/jeductechsoci.17.4.49>
10. Sinha S., Castro E., Moran C. How artificial intelligence can personalize education. *IEEE Spectrum*. 2023. URL: <https://spectrum.ieee.org/how-ai-can-personalize-education>
11. Williamson B. Introduction: Learning machines, digital data and the future of education. SAGE Publications Ltd. 2017. <https://doi.org/10.4135/9781529714920>

Дата надходження статті: 12.07.2025

Дата прийняття статті: 18.07.2025

Опубліковано: 23.09.2025