

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ БІОРЕСУРСІВ І
ПРИРОДОКОРИСТУВАННЯ УКРАЇНИ

Кваліфікаційна наукова праця
на правах рукопису

КЛИМЕНКО ЄВГЕНІЙ ОЛЕГОВИЧ

УДК 004.93:37.018.43

ДИСЕРТАЦІЯ

ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ ОСВІТНЬОЇ АНАЛІТИКИ НА ОСНОВІ
МЕТОДІВ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ

122 «Комп'ютерні науки»

12 «Інформаційні технології»

Подається на здобуття наукового ступеня доктора філософії

Дисертація містить результати власних досліджень.

Використання ідей, результатів і текстів інших авторів мають посилання на
відповідне джерело

Є.О.Клименко

Науковий керівник
ГЛАЗУНОВА Олена Григорівна,
доктор педагогічних наук, професор

КИЇВ - 2026

АНОТАЦІЯ

Клименко Є.О. Інформаційна технологія освітньої аналітики на основі методів інтелектуального аналізу даних. - Кваліфікаційна наукова праця на правах рукопису.

Дисертація на здобуття наукового ступеня доктор філософії за спеціальністю 122 «Комп'ютерні науки». - Національний університет біоресурсів і природокористування України, Київ, 2026.

У дисертаційній роботі представлені результати досліджень, проведених здобувачем, які зосереджені на розробці та впровадженні інформаційної технології освітньої аналітики, яка забезпечує підвищення обґрунтованості управлінських рішень шляхом застосування методів інтелектуального аналізу освітніх даних.

Дослідження зосереджене на розробці та впровадженні інформаційної технології дозволить суттєво підвищити точність прогнозування навчальної успішності, виявляти приховані закономірності в освітніх даних для прийняття обґрунтованих управлінських рішень на рівні ЗВО.

У сучасних умовах цифрової трансформації освіти та зростання обсягів освітніх даних (результати успішності, цифрова активність, відвідуваність, взаємодія у віртуальному середовищі тощо) особливої важливості набуває використання інтелектуальних інформаційних технологій для підвищення ефективності управління освітнім процесом. Освітня аналітика сьогодні виступає потужним інструментом прийняття рішень, що ґрунтуються на даних.

Актуальною є потреба в інформаційній технології, яка б поєднувала методи машинного навчання, прогнозування, інтерпретації моделей та візуалізації, і могла б ефективно виявляти ризики академічної неуспішності, підтримувати персоналізацію навчання та формувати візуальні аналітичні звіти.

Метою даного дисертаційного дослідження є обґрунтування та розробка інформаційної технології освітньої аналітики, яка забезпечує ефективне виявлення закономірностей у даних про освітній процес на основі методів

інтелектуального аналізу даних з метою підвищення якості управлінських рішень та прогнозування результатів навчання здобувачів освіти.

Об'єкт дослідження – інформаційні процеси збору, представлення, обробки та інтерпретації даних про освітню діяльність у цифровому середовищі ЗВО.

Предмет дослідження – алгоритми, моделі, методи та програмні засоби реалізації інформаційної технології освітньої аналітики, що ґрунтується на методах інтелектуального аналізу даних у цифрових освітніх середовищах.

У першому розділі виконано комплексний теоретико-аналітичний огляд предметної області освітньої аналітики та інтелектуального аналізу освітніх даних. Проаналізовано сучасний стан розвитку Learning Analytics і Education Data Mining в Україні та за кордоном, уточнено їх понятійно-термінологічний апарат і визначено концептуальні відмінності між цими напрямками. Узагальнено сучасні задачі та методи освітньої аналітики, зокрема класифікацію, кластеризацію, регресійне прогнозування, аналіз поведінкових патернів і соціальних взаємодій. Проведено аналіз наукового ландшафту за публікаціями, індексованими у наукометричних базах Scopus та Dimensions із використанням платформи VOSviewer, що дозволило виявити ключові дослідницькі кластери та тенденції розвитку галузі. Обґрунтовано необхідність застосування методів інтелектуального аналізу даних в освіті та сформульовано принципи й функціональні вимоги до аналітичних систем закладів вищої освіти.

У другому розділі здійснено обґрунтування методичних і алгоритмічних засад побудови інформаційної технології освітньої аналітики. Систематизовано джерела освітніх даних і сформовано структуру інформаційної системи, що охоплює підсистеми збору, інтеграції, обробки, зберігання та аналізу даних. Запропоновано методи інтеграції та підготовки даних з ЄДЕБО, LMS Moodle та інформаційної системи «Деканат», що дозволило сформувати багатовимірний профіль здобувача освіти. Розроблено алгоритми очищення, нормалізації та формування аналітичних вибірок. На основі порівняльного аналізу обґрунтовано

вибір моделей Data Mining (дерева рішень, Random Forest, LightGBM, логістична регресія) для задач прогнозування успішності, а також методів кластеризації та регресійного аналізу для дослідження освітніх процесів.

У третьому розділі розроблено та обґрунтовано інформаційну технологію освітньої аналітики як інтегроване програмно-алгоритмічне рішення, що забезпечує повний цикл аналітичної обробки освітніх даних у закладі вищої освіти. Запропоновано архітектуру системи освітньої аналітики, яка включає модулі збору даних, формування ознак, моделювання, оцінювання результатів та візуалізації. Формалізовано атрибути, що характеризують демографічні, освітні та поведінкові характеристики студентів, а також алгоритмічні процедури їх перетворення у структурований формат, придатний для інтелектуального аналізу. Реалізовано математичні моделі для оцінки та прогнозування успішності здобувачів освіти з використанням інтерпретованих моделей машинного навчання.

У четвертому розділі здійснено практичну реалізацію та апробацію розробленої інформаційної технології освітньої аналітики на реальних даних НУБіП України. Реалізовано сценарійний підхід до прогнозування успішності здобувачів освіти («Стартовий», «Проміжний», «Прикінцевий»), що дозволило оцінити ефективність прогнозування на різних етапах навчання. Експериментальні результати підтвердили високу прогностичну здатність розроблених моделей, зокрема для завершального сценарію досягнуто значень Balanced Accuracy у межах 0,77–0,82, а також стабільність моделей Random Forest і LightGBM. Доведено статистичну надійність і достовірність отриманих результатів. Показано практичну цінність технології для раннього виявлення студентів групи ризику, моніторингу динаміки успішності та підтримки управлінських рішень. Окрему увагу приділено візуалізації даних, яка в межах запропонованої технології виступає ключовим етапом аналітичного процесу та реалізується за допомогою інтерактивних дашбордів Power BI.

Запропоновані методи та розроблену інформаційну технологію застосовано для прогнозування успішності студентів факультету інформаційних технологій НУБІП України та візуалізації відкритих даних про освітню діяльність університету на його цифровій платформі. Результати досліджень використовувались в Національному університеті біоресурсів і природокористування України при викладанні дисципліни «Business intelligence системи». Результати роботи впроваджені в Національному університеті біоресурсів і природокористування України (НУБІП України). Акти про впровадження представлені в додатку А та Б.

У підсумку дисертаційна робота розв'язує актуальну науково-прикладну задачу створення інформаційної технології освітньої аналітики на основі методів інтелектуального аналізу даних та має теоретичну й практичну значущість для розвитку цифрової трансформації закладів вищої освіти України.

Ключові слова: освітня аналітика; інформаційна технологія; інтелектуальний аналіз даних; прогнозування успішності; машинне навчання; модель; алгоритми; оптимізація; залученість; візуалізація даних; електронні курси; користувачі; LMS Moodle; освітній процес; здобувачі вищої освіти; цифрові сліди; цифрове освітнє середовище.

ANNOTATION

Klymenko Ye.O. Information technology of educational analytics based on methods of intelligent data analysis. - Qualification scientific work as a manuscript.

Thesis for the degree of Doctor of Philosophy in the specialty 122 “Computer Science”. - National University of Life and Environmental Sciences of Ukraine, Kyiv, 2026.

The dissertation presents the results of research conducted by the applicant, which focuses on the development and implementation of information technology for educational analytics, which ensures the improvement of the validity of management decisions through the application of methods of intelligent analysis of educational data.

In the context of the digital transformation of education and the rapid growth of educational data volumes (academic performance results, digital activity, attendance, interactions in virtual learning environments, etc.), the application of intelligent information technologies becomes critically important for improving the effectiveness of educational process management. Educational analytics today acts as a powerful data-driven decision-making tool.

There is an urgent need for an information technology that integrates machine learning methods, predictive modeling, model interpretability, and data visualization to effectively identify risks of academic failure, support personalized learning trajectories, and generate visual analytical reports for various stakeholders.

The aim of this dissertation research is to substantiate and develop information technology for educational analytics that ensures effective identification of patterns in data on the educational process based on intelligent data analysis methods in order to improve the quality of management decisions and predict the learning outcomes of students.

The object of the research is the information processes for collecting, presenting, processing and interpreting data on educational activities in the digital environment of higher education institutions

The subject of the research comprises algorithms, models, methods, and software tools for implementing an educational analytics information technology based on data mining methods in digital educational environments. The research focuses on the development and implementation of an information technology that enables a significant increase in the accuracy of academic performance prediction, the identification of hidden patterns in educational data, and the support of evidence-based managerial decision-making at the HEI level.

In Chapter 1, a comprehensive theoretical and analytical review of the subject area of educational analytics and educational data mining is conducted. The current state of development of Learning Analytics and Educational Data Mining in Ukraine and internationally is analyzed, their conceptual and terminological frameworks are refined, and the conceptual differences between these domains are identified. Modern tasks and methods of educational analytics are summarized, including classification, clustering, regression-based prediction, analysis of behavioral patterns, and social interaction analysis. A scientific landscape analysis based on publications indexed in the Scopus and Dimensions databases is performed using the VOSviewer platform, which enables the identification of key research clusters and development trends. The necessity of applying data mining methods in education is substantiated, and principles and functional requirements for analytical systems in higher education institutions are formulated.

In Chapter 2, the methodological and algorithmic foundations for constructing an educational analytics information technology are substantiated. Educational data sources are systematized, and the structure of the information system is developed, covering subsystems for data collection, integration, processing, storage, and analysis. Methods for data integration and preprocessing from the Unified State Electronic Database on Education, the Moodle Learning Management System, and the “Dean’s Office” information system are proposed, enabling the formation of a multidimensional student profile. Algorithms for data cleaning, normalization, and analytical dataset formation are developed. Based on comparative analysis, the selection of data mining

models (decision trees, Random Forest, LightGBM, logistic regression) for academic performance prediction tasks, as well as clustering and regression methods for analyzing educational processes, is justified.

In Chapter 3, an educational analytics information technology is developed and substantiated as an integrated software and algorithmic solution that ensures a full analytical processing cycle of educational data within a higher education institution. The architecture of the educational analytics system is proposed, including modules for data collection, feature engineering, modeling, result evaluation, and visualization. Attributes characterizing students' demographic, educational, and behavioral characteristics are formalized, along with algorithmic procedures for their transformation into a structured format suitable for intelligent analysis. Mathematical models for assessing and predicting students' academic performance using interpretable machine learning models are implemented.

In Chapter 4, the practical implementation and validation of the developed educational analytics information technology are carried out using real data from the National University of Life and Environmental Sciences of Ukraine. A scenario-based approach to academic performance prediction ("Initial," "Intermediate," and "Final" scenarios) is implemented, allowing the evaluation of predictive effectiveness at different stages of the learning process. Experimental results confirm the high predictive capability of the developed models; in particular, for the final scenario, Balanced Accuracy in the range of 0.77–0.82 are achieved, demonstrating the stability of the Random Forest and LightGBM models. The statistical reliability and validity of the obtained results are proven. The practical value of the technology for early identification of at-risk students, monitoring academic performance dynamics, and supporting managerial decision-making is demonstrated. Special attention is given to data visualization, which within the proposed technology serves as a key stage of the analytical process and is implemented using interactive Power BI dashboards.

The proposed methods and the developed information technology are applied to predict academic performance of students of the Faculty of Information Technologies

at the National University of Life and Environmental Sciences of Ukraine and to visualize open data on the university's educational activities on its digital platform. The research results are used in teaching the course "Business Intelligence Systems" at the National University of Life and Environmental Sciences of Ukraine. The results of the work have been implemented at the National University of Life and Environmental Sciences of Ukraine (NULES of Ukraine). The implementation reports are presented in Appendices A and B.

In conclusion, the dissertation solves a relevant scientific and applied problem of developing an educational analytics information technology based on data mining methods and has significant theoretical and practical value for the digital transformation of higher education institutions in Ukraine.

Keywords: educational analytics; information technology; intellectual data analysis; academic performance prediction; machine learning; model; algorithms; optimization; engagement; data visualization; e-courses; users; LMS Moodle; learning process; higher education students; digital footprints; digital learning environment.

СПИСОК ОПУБЛІКОВАНИХ ПРАЦЬ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

Статті у наукових виданнях,

включених до міжнародних наукометричних баз даних

Scopus та/або Web of Science Core Collection

1. Hlazunova O., Klymenko N., Mokriiev M., Nehrey M., **Klymenko Ye.** 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. P. 670–682. *(Глазуновою О.Г. визначено загальну концепцію дослідження, наукове консультування, участь у формуванні методологічних засад, наукової новизни роботи та узагальненні результатів. Клименко Н.А. проведено літературний науковий огляд та запропоновано ідею дослідження щодо застосування технологій освітньої аналітики в системах дистанційного навчання. Мокрієвим М.В. здійснена технічна реалізація збору даних із LMS Moodle, налаштування інструментів аналітики та забезпечення інтеграції програмних засобів аналізу даних. Негрей М.В. здійснила проведення статистичної валідації та інтерпретації результатів отриманих аналітичних моделей. Клименко Є.О. представив результати технології аналізу даних, та сформував рекомендації щодо впровадження систем аналізу великих даних у закладах вищої освіти).*

Статті у наукових виданнях,

включених до Переліку наукових фахових видань України

2. 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. С. 49–59. *(Глазуновою О.Г. сформовано концептуальні засади дослідження, визначено стратегічні напрями цифрової трансформації освітнього середовища, обґрунтовано необхідність використання цифрових інструментів освітньої аналітики для підтримки управлінських рішень у закладах вищої освіти. Клименком Є.О. розроблено структуру та логіку реалізації процесу аналізу*

великих масивів освітніх даних та здійснено експериментальне дослідження інформаційної технології освітньої аналітики на прикладі навчального порталу Національного університету біоресурсів і природокористування України. **Волошина Т.В.** провела аналіз груп цифрових інструментів освітньої аналітики, систематизувала підходи до класифікації аналітичних засобів LMS Moodle. Мокрієвим М.В. сформовано систему освітніх індикаторів для оцінювання активності та залученості здобувачів освіти. **Вороненко О.В.** провів експертну оцінку використання результатів освітньої аналітики та аналіз практичної значущості отриманих результатів для освітніх програм.)

3. **Клименко Є.,** Глазунова О. Методи інтелектуального аналізу освітніх даних у системах електронного навчання. Інформаційні технології та суспільство. 2024. № 2 (13). С. 34–40. (Клименко Є.О. представив теоретичне дослідження розвитку, основних задач і методів інтелектуального аналізу для застосування в інформаційних системах і технологіях, обґрунтував можливості використання методів інтелектуального аналізу для прогнозування результатів навчання та підтримки прийняття управлінських рішень Глазунова О.Г. здійснювала наукове консультування щодо методології освітньої аналітики та визначила перспективи застосування методів інтелектуального аналізу даних в інформаційних системах).

4. Глазунова О., **Клименко Є.** Інформаційна технологія аналітики освітніх даних. Наука і техніка сьогодні. 2025. № 2 (43). С. 1147–1155. (Глазуновою О.Г. надано наукове консультування щодо застосування методів Data Mining в освітньому середовищі, проведено рецензування та загальне керівництво роботою над статтею. Клименко Є.О. визначив та обґрунтував основні складові інформаційної технології аналітики освітніх даних із використанням методів інтелектуального аналізу даних та Big Data на основі журналів активності LMS Moodle).

5. **Клименко Є.,** Глазунова О. Архітектура інформаційної технології освітньої аналітики з використанням інтелектуального аналізу даних.

Інформаційні технології та суспільство. 2025. № 2 (17). С. 69–75. *(Клименко Є.О. обґрунтував результати роботи розробленої інформаційної технології освітньої аналітики на основі використання методів інтелектуального аналізу освітніх даних. Глазуною О.Г. здійснено методологічне обґрунтування дослідження та проведений комплексний аналіз можливостей імплементації інтелектуального аналізу в інформаційну технологію освітньої аналітики).*

Тези наукових доповідей

6. **Клименко Є. О.** Компетентнісний підхід в освітніх траєкторіях випускників ІТ спеціальностей ЗВО на базі платформи Moodle. Економіко-правовий розвиток сучасної України: XI Всеукраїнська наукова конференція студентів, аспірантів та молодих вчених, що присвячена 25-й річниці з дня заснування Національного університету «Одеська юридична академія», м. Одеса, 25 листопада 2022 року: тези доповіді. Одеса, 2022. С. 197–199.

7. **Клименко Є., Глазунова О.** Врахування компетентісного підходу в освітніх траєкторіях здобувачів освіти. Sectoral research XXI: characteristics and features: V International Scientific and Theoretical Conference, Chicago, USA, February 3, 2023: Conference Paper. Chicago, USA, 2023. P. 127–128. *(Клименко Є. Обґрунтував доцільність впровадження інформаційних технологій з використанням модульного об'єктно-орієнтованого динамічного навчального середовища для ефективної персоналізації навчання студентів у вищих навчальних закладах. Глазуною О.Г. визначено концептуальні рамки дослідження та доведено вплив трансформацій на організацію освітнього процесу у закладах вищої освіти).*

8. **Клименко Є., Глазунова О.** Moodle Big Data analytics за допомогою Power BI. Science in Motion: Classic and modern tools and methods in scientific investigations: II International Scientific and Practical Conference, Vinnytsia–Vienna, Ukraine-Austria, January 19, 2024: Conference Paper. Vinnytsia–Vienna, Ukraine-Austria, 2024. P. 201–203. *(Клименком Є.О. розроблені підходи до аналізу великих масивів даних LMS Moodle з інтеграцією з Power BI та надані результати*

аналізу великих даних з платформи електронного навчання Moodle в НУБІП України. Глазуною О.Г сформульовано постановку задачі аналізу великих масивів даних, що зберігаються в СУБД та здійснене наукове консультування щодо застосування цих методів в освітніх інформаційних системах).

9. **Клименко Є.** Аналіз освітніх даних у системах електронного навчання. Інформаційні технології: економіка, техніка, освіта 2024: Міжнародна науково-практична конференція молодих вчених, м. Київ, 7–8 листопада 2024 року: тези доповіді. Київ, 2024. С. 79–81.

10. **Клименко Є.** Розробка інформаційних систем прогнозування успішності здобувачів освіти. Інформаційне суспільство: технологічні, економічні та технічні аспекти становлення: Міжнародна наукова інтернет-конференція, Тернопіль–Ополе, Україна–Польща, 15–16 квітня 2025 року: тези доповіді. Тернопіль, 2025, випуск 98 С.22-25.

11. **Клименко Є.** Прогнозування успішності здобувачів освіти методами інтелектуального аналізу даних. Інформаційні технології: економіка, техніка, освіта '2025: XVI Міжнародна науково-практична конференція молодих вчених, м. Київ, 28–29 жовтня 2025 року: тези доповіді. Київ, 2025. С. 89–90.

12. Hlazunova O., Klymenko N., **Klymenko Ye.** Machine Learning in Educational Analytics Systems. Cyfryzacja Edukacji na Uczelniach Przyrodniczych: IX Międzynarodowa Konferencja, Wrocławiu, Polska, Listopad 20–21, 2025: tezy raportu. (Глазунова О.Г. обгрунтувала доцільність використання методів машинного навчання в освітній аналітиці. Клименко Н.А. провела інтерпретацію результатів машинного навчання в контексті освітнього процесу. Клименком Є.О проведено аналіз сучасних алгоритмів машинного навчання для задач прогнозування успішності здобувачів освіти та розроблено архітектурну модель інформаційної технології освітньої аналітики з використанням ML-алгоритмів).

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ ТА СКОРОЧЕНЬ

IT– інформаційна технологія;

ЗВО – заклад вищої освіти;

EDM (Educational Data Mining) – інтелектуальний аналіз освітніх даних;

LA (Learning Analytics) – освітня (навчальна) аналітика;

LMS (Learning Management Systems) – система управління навчанням, онлайн-платформа або програмне забезпечення для створення, організації, зберігання та поширення навчальних матеріалів;

DM (Data Mining)– інтелектуальний аналіз даних, глибинний аналіз даних;

ML(Machine Learning) – машинне навчання, підрозділ штучного інтелекту;

OLAP (Online Analytical Processing) – аналітична обробка по багатовимірних даних у реальному часі;

Moodle (Modular Object-Oriented Dynamic Learning Environment) – відкрита платформа (LMS) для організації дистанційного та змішаного навчання;

EMIS (Education Management Information System) – інформаційна система управління освітою;

ЄДЕБО – єдина державна електронна база з питань освіти;

ІС «Деканат» – інформаційна система обліку освітніх траєкторій здобувачів Національного університету біоресурсів і природокористування України;

UML (Unified Modeling Language) – стандартизована уніфікована мова моделювання, що використовує графічні позначення для створення абстрактних моделей програмних систем, бізнес-процесів та організаційних структур;

BPMN (Business Process Model and Notation) – стандартизована графічна система, призначена для моделювання бізнес-процесів

ЗМІСТ

ВСТУП	17
РОЗДІЛ 1. АНАЛІЗ СУЧАСНОГО СТАНУ, ПІДХОДІВ І ЗАВДАНЬ ОСВІТНЬОЇ АНАЛІТИКИ	26
1.1. Теоретичні підходи до визначення понять освітньої аналітики та інтелектуального аналізу освітніх даних	26
1.2. Основні задачі Learning Analytics у контексті розвитку Data Mining	30
1.3. Обґрунтування доцільності використання інтелектуального аналізу даних у цифровому освітньому середовищі	36
1.4. Структура наукового ландшафту освітньої аналітики на основі наукометричних баз даних	37
1.5. Інструментарій та завдання інтелектуального аналізу даних в освіті	44
Висновки до розділу 1	54
РОЗДІЛ 2. ОБґРУНТУВАННЯ ТА СКЛАДОВІ ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ ОСВІТНЬОЇ АНАЛІТИКИ	56
2.1. Джерела даних і функціональні компоненти інформаційної технології освітньої аналітики	56
2.2. Порівняльний аналіз цифрових платформ управління навчанням	68
2.3. Аналітична та прогностична обробка освітніх даних з використанням LMS Moodle	72
2.4. Методичні, алгоритмічні та модельні засади інформаційної технології освітньої аналітики	80
Висновки до розділу 2	85
РОЗДІЛ 3. РОЗРОБКА ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ ОСВІТНЬОЇ АНАЛІТИКИ	87
3.1. Архітектура інформаційної технології освітньої аналітики	87
3.2. Формалізація атрибутів та ознак успішності здобувачів освіти	96
3.3. Математичні моделі оцінювання успішності здобувачів освіти	99

3.4. Статистичний аналіз поведінкових та академічних предикторів успішності здобувачів освіти	109
Висновки до розділу 3	115
РОЗДІЛ 4. ПРОГНОЗУВАННЯ УСПІШНОСТІ ЗДОБУВАЧІВ ОСВІТИ З ВИКОРИСТАННЯМ АЛГОРИТМІВ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ	117
4.1. Методологія та етапи побудови прогностичної моделі успішності здобувачів освіти	117
4.2. Алгоритмічна та програмна реалізація моделей прогнозування успішності здобувачів освіти	119
4.3. Апробація підходів прогнозування успішності студентів (на базі вибірки здобувачів освіти НУБІП України)	132
4.3.1. Сценарійний підхід 1 «Стартовий»	132
4.3.2. Сценарійний підхід 2 «Проміжний»	142
4.3.3. Сценарійний підхід 3 «Прикінцевий»	159
4.4. Візуалізація як етап аналізу даних в освітній аналітиці	167
Висновки до розділу 4	173
ВИСНОВКИ	175
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ	178
ДОДАТКИ	192

ВСТУП

Актуальність теми дослідження. Цифрова трансформація суспільства та освіти, зокрема, є однією з найважливіших тенденцій сучасного розвитку. Інтенсивне впровадження інформаційно-комунікаційних технологій в освітній процес створює нові можливості для забезпечення якості освіти, підвищення ефективності роботи закладів вищої освіти та індивідуалізації навчання. Водночас величезні масиви даних, що збираються в освітніх інформаційних системах, як і раніше, недостатньо використовуються при прийнятті рішень на рівні менеджменту навчальних закладів та вдосконаленні освітніх траєкторій здобувачів освіти.

Особливу роль у цьому контексті відіграє освітня аналітика, що базується на застосуванні методів інтелектуального аналізу даних, статистичного моделювання та машинного навчання. Цей підхід дозволяє не тільки описати поточний стан освітнього процесу, але й виявити приховані закономірності, створити моделі прогнозування успішності та сформулювати науково обґрунтовані рекомендації для прийняття рішень на різних рівнях освітньої системи.

Протягом останніх двох десятиліть українські навчальні заклади розробили потужну цифрову інфраструктуру: системи управління навчанням, електронні журнали, державні інформаційні системи, а також корпоративні платформи, що підтримують дистанційне та змішане навчання. Використання інтелектуальних методів аналізу даних у сфері освіти відкриває нові можливості. Так, наприклад, кластерний аналіз може бути використаний для визначення типових поведінкових моделей здобувачів освіти у навчальному середовищі, методи прогнозування дозволяють на ранньому етапі виявляти здобувачів, які ризикують отримати незадовільні оцінки або бути відрахованими, а асоціативні правила і послідовні аналізи допомагають встановити зв'язок між навчальною діяльністю і досягнутими результатами.

У контексті воєнного стану та відновлення української системи освіти після війни питання ефективного використання цифрових даних набуло особливого значення. Масовий перехід до дистанційного та змішаного навчання, переселення учасників освітнього процесу, нерівний доступ до освітніх ресурсів та необхідність забезпечення безперервності освітнього процесу призвели до різкого збільшення обсягу цифрової освітньої інформації. У цих умовах освітня аналітика є інструментом для своєчасного виявлення проблем у сфері освіти, забезпечення прозорості та гнучкості в управлінні, а також розробки адаптивних освітніх стратегій. Вже є численні приклади використання аналітичних інструментів у вищих навчальних закладах України (активно аналізуються дані системи управління навчанням Moodle для моніторингу навчальної діяльності студентів під час воєнного стану та виявлення груп ризику, впроваджуються аналітичні панелі для адміністраторів та викладачів для відстеження результатів навчання та відвідуваності в режимі реального часу, розробляються інформаційні системи, які можуть стати основою для майбутнього впровадження машинного навчання та прогнозової аналітики на національному рівні.

Крім того, інтеграція України до Європейського освітнього простору вимагає впровадження сучасних підходів до збору, аналізу та інтерпретації освітніх даних відповідно до міжнародних стандартів. Такі організації, як ОЕСР та ЮНЕСКО, наголошують на важливості політики, заснованої на доказах, в якій управлінські рішення приймаються на основі достовірних даних, прогнозів та об'єктивних аналітичних моделей. З огляду на цілі сталого розвитку в освіті (SDG 4 - Sustainable Development Goals Goal 4: Quality Education), глобальні цифрові трансформації останніх років, а також виклики, пов'язані з дистанційною та змішаною формами навчання, зростає потреба у впровадженні інноваційних рішень у сфері освітньої аналітики.

Виходячи з викладеного, розробка інформаційної технології освітньої аналітики, що інтегрує сучасні методи інтелектуального аналізу даних та забезпечує побудову інтерпретованих моделей прогнозування успішності

здобувачів освіти, є своєчасною задачею як з наукової, так і з прикладної точок зору.

Уперше концепт освітньої аналітики як самостійного міждисциплінарного напрямку був окреслений у роботах G. Siemens, D. Gašević та S. Dawson (2011), які започаткували конференцію Learning Analytics and Knowledge і сформували засади використання аналітики для підтримки освітніх рішень. Подальший розвиток теоретичних і методологічних аспектів забезпечили праці R. Ferguson (2012), N. Sclater (2017) та B. Rienties (2018), де наголошено на ролі інструментів візуалізації, прогнозування та етичних засадах аналітичної діяльності у ЗВО. Окрему нішу займають дослідження з інтелектуального аналізу освітніх даних (Educational Data Mining), які систематизовані у роботах R. Baker та K. Yacef (2020). Згадані дослідники акцентують на можливостях класифікації, кластеризації, часових моделей та глибокого навчання для виявлення закономірностей у навчальних даних.

В українському науковому просторі проблема цифровізації освіти, освітньої аналітики та її цифрового інструментарію досліджується у роботах В. Бикова та А. Гуржія, які формують науково-методичне забезпечення цифровізації освіти України та закладають методологічні основи хмарноорієнтованих освітніх середовищ і цифрової освіти; Н. Морзе, яка опікується цифровізацією навчального середовища вищих навчальних закладів та розвитком передових цифрових екосистем в університетах, О. Глазунової, яка досліджує моделі ефективного використання інформаційно-комунікаційних та дистанційних технологій навчання і системи електронного навчання у ЗВО, О. Горошкіної, В. Лахна, О. Криворучко які аналізують цифрову трансформацію українських ЗВО та інформаційні технології формування компетентностей здобувачів освіти. Також важливим джерелом підтвердження актуальності є міжнародні документи, зокрема, UNESCO Guidelines on the Ethics of AI in Education (2022), OECD AI and the Future of Skills (2021), які підкреслюють

стратегічне значення використання освітньої аналітики та штучного інтелекту у сфері освіти.

Таким чином, сукупність наукових і прикладних досліджень у цій галузі, а також потреба українських ЗВО в ефективних інструментах аналізу та прогнозування освітніх результатів, визначають високу затребуваність створення інформаційної технології освітньої аналітики, що базується на методах інтелектуального аналізу даних.

Тому розробка інформаційних технологій для освітньої аналітики на основі інтелектуальних методів аналізу даних є своєчасним науково-практичним завданням, що відповідає стратегічним напрямкам розвитку освіти в Україні. Впровадження такого підходу сприятиме підвищенню ефективності управління навчальними закладами, індивідуалізації навчання, оперативному реагуванню на виклики воєнного та повоєнного періоду, інтеграції України в сучасний європейський цифровий освітній простір.

Дослідження зосереджене на розробці та впровадженні інформаційної технології дозволить суттєво підвищити точність прогнозування навчальної успішності, виявляти приховані закономірності в освітніх даних для прийняття обґрунтованих управлінських рішень на рівні ЗВО.

Зв'язок роботи з науковими програмами, планами, темами, грантами.

Тема дисертації «Інформаційна технологія освітньої аналітики на основі методів інтелектуального аналізу даних» безпосередньо пов'язана з пріоритетними напрямами державної політики України. Зв'язок теми дисертації з державними програмами зумовлений необхідністю реалізації стратегічних напрямів цифрової трансформації освіти в Україні та відповідає цілям сталого розвитку в освіті (SDG 4 – Quality Education), а також положенням Концепції цифрової трансформації освіти і науки України, Стратегії розвитку вищої освіти в Україні на 2022–2032 роки та Плану відновлення України. Робота виконувалась в рамках наукової тематики НУБІП України «Стратегія цифрової трансформації економіки України як інструменту забезпечення соціально-економічного

розвитку та національної безпеки» Номер державної реєстрації 0121U110194 № 110/1м-пр-2021. Тема дисертації повністю відповідає науковим напрямкам факультету інформаційних технологій та кафедри інформаційних систем і технологій, узгоджується із загальною дослідницькою стратегією Національного університету біоресурсів і природокористування України, спрямованою на цифрову трансформацію освітньої діяльності та використання сучасних інформаційних технологій в освітньому процесі.

Мета і завдання дослідження. Метою дисертаційного дослідження є обґрунтування та розробка інформаційної технології освітньої аналітики, яка забезпечує багаторівневий аналіз та інтерпретацію освітніх даних на основі методів інтелектуального аналізу даних з метою прогнозування результатів навчання здобувачів вищої освіти.

Для досягнення поставленої мети необхідно розв'язати такі *завдання*:

1. Проаналізувати сучасний стан розвитку освітньої аналітики в Україні та світі, визначити концептуальні засади, наукові підходи, моделі та інструменти, що застосовуються для збору, обробки й аналізу освітніх даних.

2. Обґрунтувати принципові засади побудови, структуру та функції інформаційної технології освітньої аналітики з використанням методів інтелектуального аналізу даних.

3. Обґрунтувати вибір алгоритмів, методів та моделей інтелектуального аналізу освітніх даних для реалізації інформаційної технології освітньої аналітики з метою прогнозування успішності здобувачів освіти та підтримки прийняття управлінських рішень.

4. Розробити інформаційну технологію освітньої аналітики як інтегроване програмно-алгоритмічне рішення, що реалізує методи аналітичної обробки освітніх даних, інтегрованої з наявними освітніми інформаційними сервісами закладу освіти та програмні засоби її реалізації.

5. Здійснити апробацію інформаційної технології освітньої аналітики та оцінити ефективність і практичну цінність розробленої інформаційної технології.

Об'єкт дослідження – інформаційні процеси збору, представлення, обробки та інтерпретації даних про освітню діяльність у цифровому середовищі ЗВО.

Предмет дослідження – алгоритми, моделі, методи, інформаційні технології та програмні засоби реалізації інформаційної технології освітньої аналітики, що ґрунтується на методах інтелектуального аналізу даних у цифрових освітніх середовищах.

Методи дослідження. Теоретичною основою дисертаційного дослідження є наукові публікації з технологій, стандартів та теорії інтелектуального аналізу освітніх даних, формування інформаційних технологій, теорії прийняття рішень, статистичні методи, методи машинного навчання, методи оптимізації. Методологічну основу дослідження становлять системний та процесний підходи, що дозволяють розглядати освітню аналітику як невід'ємний елемент управління якістю освіти. На технологічному рівні використано предиктивний підхід та методи Educational Data Mining та Learning Analytics (LA), які забезпечують перехід від констатації фактів до прогнозування індивідуального розвитку здобувачів освіти.

Наукова новизна отриманих результатів:

вперше:

1) розроблено інформаційну технологію освітньої аналітики, яка інтегрує методи інтелектуального аналізу, прогнозні моделі для прийняття управлінських рішень у єдиному цифровому середовищі закладу вищої освіти;

2) розроблено методи інтеграції даних з різних джерел для формування багатовимірного профілю здобувача освіти;

удосконалено

3) моделі оцінювання ключових характеристик успішності здобувача освіти та алгоритмічне забезпечення процесу освітньої аналітики, яке враховує динамічні зміни навчальної активності студентів у реальному часі та дозволяє здійснювати багаторівневий аналіз результатів навчання;

набули подальшого розвитку:

4) теоретичні засади освітньої аналітики, зокрема визначення структури, функцій та принципів побудови інформаційних аналітичних систем у сфері вищої освіти, що базуються на інтеграції методів інтелектуального аналізу даних із системами управління навчанням;

5) практичні підходи до цифрової трансформації освітнього процесу через впровадження інструментів освітньої аналітики, які сприяють прийняттю обґрунтованих управлінських рішень, персоналізації навчання та підвищенню якості освіти в українських закладах вищої освіти.

Практичне значення отриманих результатів. Практична цінність полягає у тому, що розроблено та впроваджено інформаційну технологію освітньої аналітики, яка забезпечує інтеграцію освітніх даних із різних інформаційних систем закладу вищої освіти, їх аналітичну обробку, прогнозування навчальної успішності здобувачів освіти та підтримку прийняття управлінських рішень. Запропоновані алгоритми, моделі та програмні засоби можуть бути використані у діяльності закладів вищої освіти для моніторингу освітнього процесу, раннього виявлення студентів групи ризику, персоналізації навчання та підвищення якості освітніх послуг. Результати досліджень використовувались в Національному університеті біоресурсів і природокористування України при викладанні дисципліни «Business intelligence системи». Практичні результати роботи впроваджені в Національному університеті біоресурсів і природокористування України як модуль освітньої аналітики в інформаційну систему управління освітнім процесом Nubip Digital.

Особистий внесок здобувача. У процесі виконання дисертаційного дослідження всі основні результати, положення та висновки отримані автором самостійно. Здобувач особисто здійснив пошук, аналіз і систематизацію наукових, методичних та інформаційних джерел, що стосуються тематики дослідження. Самостійно розроблено інформаційну технологію, здійснено підбір моделей інтелектуального аналізу освітніх даних. Усі етапи дослідження від теоретичного обґрунтування до практичної реалізації та апробації результатів виконано без залучення сторонніх розробок.

Апробація результатів дослідження. Основні результати та перспективи дисертаційного дослідження обговорювались на міжнародних та всеукраїнських науково-практичних конференціях: Міжнародна науково-практична конференція молодих вчених «Інформаційні технології: економіка, техніка, освіта'2024», 7-8 листопада 2024 року, НУБіП України, м. Київ; XI Всеукраїнська наукова конференція студентів, аспірантів та молодих вчених, що присвячена 25-й річниці з дня заснування Національного університету «Одеська юридична академія», 25 листопада 2022, м. Одеса; V International Scientific and Theoretical Conference «Sectoral research XXI: characteristics and features», February 3, 2023; Chicago, USA; II International Scientific and Practical Conference «Science in motion: classic and modern tools and methods in scientific investigations» 19.01.2024, Vinnytsia, UKR - Vienna, AUT; ICCSEE2024: The 7th International Conference on Computer Science, Engineering and Education Applications April 27 - 28, 2024, Kyiv, Ukraine; Міжнародна наукова інтернет-конференція Інформаційне суспільство: технологічні, економічні та технічні аспекти становлення, 15-16 квітня 2025, м. Тернопіль, Україна, м. Ополе, Польща; IX Międzynarodowa Konferencja «Cyfryzacja Edukacji na Uczelniach Przyrodniczych», Polska, Listopad 20–21, 2025 Wrocławiu.

Публікації. За темою дисертаційного дослідження опубліковано 12 наукових праць, серед яких одна стаття, опублікована в виданні, що входить до наукометричної бази Scopus, 4 статті опубліковані у фахових виданнях України

та 7 тез доповідей, опубліковано у збірниках матеріалів наукових конференцій.

Структура та обсяг дисертації. Дисертаційна робота складається з анотацій, змісту, вступу, чотирьох розділів, висновків, списку використаних джерел та додатків. Загальний обсяг дисертації складає 203 сторінки, з них: 57 рисунків по тексту; 22 таблиці по тексту; список використаних джерел із 138 найменувань на 14-ти сторінках; 4 додатки на 12-ти сторінках.

РОЗДІЛ 1

АНАЛІЗ СУЧАСНОГО СТАНУ, ПІДХОДІВ І ЗАВДАНЬ ОСВІТНЬОЇ АНАЛІТИКИ

1.1. Теоретичні підходи до визначення понять освітньої аналітики та інтелектуального аналізу освітніх даних

Розвиток освітньої аналітики нерозривно пов'язаний із процесами цифрової трансформації освіти, зокрема впровадженням електронного навчання та широким використанням інформаційно-комунікаційних технологій у навчальному процесі. Ще на початку 2000-х років заклади вищої освіти почали активно впроваджувати системи управління навчанням (LMS – Learning Management Systems), які дозволяли організовувати дистанційні курси, автоматизувати облік навчальних дій студентів, здійснювати розклад, ведення електронного журналу та інші адміністративні функції. Ці системи стали стандартом для організації електронного освітнього середовища у багатьох університетах світу. Вони почали накопичувати значні обсяги даних про взаємодію студентів із навчальним контентом – кількість входів у систему, час активності, переглянуті ресурси, участь у форумах, виконання тестів тощо.

Проте в перші роки існування систем управління навчанням (LMS) накопичувані в них дані мали переважно адміністративне значення. Вони використовувалися здебільшого для технічного забезпечення навчального процесу: моніторингу відвідуваності курсів, автоматизованого оцінювання знань, формування академічної звітності, обліку навчального навантаження викладачів, планування ресурсів тощо. Тобто ці дані розглядалися як допоміжний інструмент для оптимізації управлінських та організаційних рішень у закладах освіти, але не як засіб глибшого розуміння самого процесу навчання.

Водночас можливості використання цих даних у педагогічному контексті залишалися майже нерозкритими. Дані про активність студентів у системі – перегляд матеріалів, виконання завдань, участь у форумах, взаємодія з викладачами – або зовсім не аналізувалися, або залишалися у «сирому» вигляді,

незрозумілому для викладача чи методиста. Відсутність відповідних інструментів, знань і практик обмежувала потенціал цих цифрових слідів як джерела для аналізу освітньої взаємодії чи індивідуальних траєкторій навчання.

З часом науковці та фахівці в галузі освіти акцентували увагу на те, що цифрові сліди студентів у LMS – це не просто технічні записи, а своєрідне відображення їх навчальної поведінки. З них можна отримати цінну інформацію про мотивацію до навчання, стиль засвоєння матеріалу, труднощі з окремими темами, рівень залученості, емоційний стан, а іноді – навіть передбачити ризик відрахування чи низьких результатів навчання.

Це усвідомлення стало поштовхом до формування нової міждисциплінарної галузі освітньої аналітики (*Learning Analytics*), яка зосереджується на зборі, вимірюванні, аналізі та інтерпретації даних про студентів і навчальне середовище з метою покращення процесу навчання. Її поява ознаменувала перехід від суто адміністративного використання освітніх даних до їх інтеграції в освітню практику для підтримки прийняття рішень, адаптації курсів до потреб студентів, вдосконалення викладання й, загалом, підвищення якості освіти.

Перші серйозні концептуальні підходи до освітньої аналітики з'явилися приблизно у 2010-х роках. Науковці почали впроваджувати методи обробки великих даних, інтелектуального аналізу даних (*Data Mining*), штучного інтелекту та машинного навчання в освітній сфері [66,110]. Таким чином, розвиток освітньої аналітики можна розглядати як наступний етап еволюції цифрової освіти: від простого використання технологій для організації навчання – до свідомого аналізу цифрової поведінки студентів з метою удосконалення методик викладання, індивідуалізації навчальних траєкторій, раннього виявлення труднощів і підвищення результативності освітнього процесу.

Поворотним моментом у становленні освітньої аналітики як самостійного напрямку досліджень та практик в освіті вважається 2011 рік. Саме тоді було організовано першу міжнародну конференцію *Learning Analytics and Knowledge*

(LAK), яка об'єднала науковців, аналітиків, освітян та технічних спеціалістів, зацікавлених у застосуванні даних для вдосконалення навчального процесу. Ця конференція стала не лише майданчиком для обміну ідеями, але й подією, що започаткувала формалізацію освітньої аналітики як міждисциплінарної галузі знань на перетині освіти, інформаційних технологій, статистики та когнітивної науки.

Того ж року вийшла друком одна з концептуально значущих публікацій у цій сфері – стаття “Penetrating the Fog: Analytics in Learning and Education”, написана Джорджем Сіменсом і Філом Лонгом. У своїй роботі автори запропонували бачення навчальної аналітики як інструменту, що дозволяє здійснювати глибший аналіз поведінки та прогресу студентів на основі даних, які генеруються у цифровому освітньому середовищі. Вони визначили навчальну аналітику як процес збирання, вимірювання, аналізу та представлення даних про студентів і їхню діяльність з метою покращення навчання, його персоналізації та підтримки прийняття педагогічних рішень [98]. Особливу увагу в публікації приділено потенціалу аналітики для підвищення прозорості освітнього процесу та прийняття обґрунтованих рішень усіма учасниками освітнього середовища – від викладачів до керівників навчальних закладів.

Важливу роль у формуванні та розвитку навчальної аналітики як міждисциплінарного наукового поля відіграла міжнародна організація SoLAR (Society for Learning Analytics Research). Ця спільнота була створена для об'єднання дослідників, практиків і розробників цифрових освітніх технологій з метою формування єдиної платформи для обговорення підходів, методів і стандартів аналізу освітніх даних. SoLAR активно підтримує наукові публікації, проводить щорічні конференції, а також сприяє поширенню найкращих практик аналітичного використання даних у навчанні. Паралельно з розвитком освітньої аналітики як галузі формувалась і суміжна дисципліна – Educational Data Mining (EDM). Хоча обидва напрями працюють із подібним джерелом (цифровими слідами освітньої діяльності), між ними існують певні відмінності у філософії,

методах та цільовому застосуванні результатів. Як зазначають Siemens і Baker (2012), EDM здебільшого орієнтується на автоматизоване виявлення закономірностей, шаблонів або аномалій у даних, часто з використанням алгоритмів машинного навчання та методів штучного інтелекту [97]. Водночас Learning Analytics більше зосереджена на людиноорієнтованому підході, який передбачає інтерпретацію даних у педагогічному контексті, адаптацію навчального середовища та підтримку прийняття освітніх рішень на основі зрозумілих і верифікованих висновків.

Таким чином, хоча обидва наукових напрямки мають спільне підґрунтя, їхні цілі та практичне застосування можуть суттєво відрізнятись. Освітня аналітика тяжіє до інтеграції аналітичних результатів у навчальний процес, залучаючи педагогів до спільної інтерпретації даних, тоді як EDM, як правило, фокусується на розвитку автоматизованих систем виявлення знань із великих освітніх масивів. Порівняльну характеристику цих двох напрямків наведено в таблиці 1.1.

Таблиця 1.1

Порівняльна характеристика Learning Analytics та Educational Data Mining

Критерій	Learning Analytics (LA)	Educational Data Mining (EDM)
Основна мета	Оптимізація навчання через інтерпретацію даних	Виявлення закономірностей у великих освітніх даних
Фокус	Педагогічна інтерпретація, підтримка прийняття рішень	Автоматизований аналіз, побудова моделей навчальної поведінки
Методи	Візуалізація, дашборди, прогностична аналітика	Машинне навчання, класифікація, кластеризація
Ключові користувачі	Викладачі, адміністратори, дизайнери курсів	Дослідники даних, розробники освітніх систем
Типовий результат	Інструменти підтримки викладання (наприклад, попередження про ризики, зворотний зв'язок)	Аналітичні моделі або алгоритми для подальшої інтеграції
Контекст застосування	Освітнє середовище, педагогічна практика	Технічні системи, адаптивне навчання
Прикладова орієнтація	Пояснення, інтерпретація, розвиток педагогічних стратегій	Виявлення закономірностей, побудова точних моделей

У сучасних дослідженнях можна виділити кілька інтерпретацій терміна «Learning Analytics» «освітня аналітика». Зокрема, класично описують її як

процес вимірювання, збору, аналізу й подання даних про студентів і навчальний контекст з метою кращого розуміння та вдосконалення освітнього середовища [98]. На думку E. Duval та K. Verbert, навчальна аналітика це аналіз цифрових слідів, які залишають студенти під час взаємодії з електронними платформами, що дозволяє вдосконалити освітній процес [29].

Узагальнюючи різні наукові підходи, освітню аналітику можна розглядати як автоматизований процес збору й аналізу цифрових слідів студентів у віртуальному освітньому середовищі для отримання нової інформації, яка допомагає ухвалювати ефективні управлінські рішення у ЗВО.

1.2. Основні задачі Learning Analytics у контексті розвитку Data Mining

Learning Analytics розглядається як міждисциплінарний напрям, що поєднує методи статистики, машинного навчання, Data Mining і візуалізації для збору, обробки та аналізу даних про навчальну діяльність здобувачів освіти з метою підвищення якості освітнього процесу та підтримки прийняття управлінських рішень. Основними задачами Learning Analytics є моніторинг навчальної активності, прогнозування успішності, виявлення ризиків академічної неуспішності, персоналізація навчання, оптимізація освітніх траєкторій. До ключових методів належать класифікація, кластеризація, регресійний аналіз, асоціативні правила, аналіз часових послідовностей, соціальний мережевий аналіз та методи пояснюваного машинного навчання. Таким чином, інтеграція методів Data Mining у Learning Analytics стала основою формування сучасних інформаційних технологій освітньої аналітики. Основні напрями досліджень та внески вчених можна згрупувати за кількома основними темами (табл. 1.2).

Таблиця 1.2

Основні напрями досліджень у сфері освітньої аналітики

Напрямки досліджень	Автори	Внески
Формування теоретичних основ та класифікація аналітики	Van Barneveld (2012), Greller (2014), Campbell (2007) [108,44,19].	запропонували модель розподілу аналітики на описову, прогностичну та пропонуючу, що широко використовується для класифікації аналітичних інструментів.
	G. Siemens, P. Long (2011) [98].	суть досліджень присвячена побудові основ та розмежуванні навчальної аналітики з інтелектуальним аналізом освітніх даних та академічною аналітикою
	Macfadyen, Dawson (2012) [65]	підкресливали важливість зв'язку між аналітикою та стратегічним управлінням у вищій освіті
Методи аналізу та технологічна база	Romero C., Ventura S. (2013) та Baker S., Inventado P. (2016) [90,11]	глибоко досліджували методи Data Mining в освітніх даних, акцентуючись на автоматизації аналізу освітніх процесів.
	Greller W., Drachsler H. (2012)[27]	вивчали очікування зацікавлених сторін і підходи до інтерпретації даних в освітній аналітиці.
	E.Lauria, J.Baron, M. Devireddy, V.Sundararaju, S. Jayaprakash (2012) S.Tseng, Y. Tsao, L.Yu, C.Chan, K.Lai (2016) [62, 103]	об'єднує системні дослідження, засновані на використанні цифрових записів про діяльність студентів у процесі їх навчання у системах електронного навчання, соціальних мережах, блогах, інших інформаційних системах
	S. Nam, S. Lonn, T. Brown, C. Davis, D. Koch (2014), M. Pistilli, J. Campbell (2014) [74,84]	розглянуто необхідність побудови моделей для вимірювання прогресу студентів та прогнозування їх успішності, удосконалення технологій навчання
Соціальні та когнітивні аспекти	Dawson S., Gasevic D.(2015) [41]	розвивали напрям соціальної навчальної аналітики, яка досліджує соціальні зв'язки та взаємодії студентів у віртуальних середовищах.
	Clow D. (2013) [20]	акцентував увагу на створенні інформаційних панелей (дашбордів) як засобів зворотного зв'язку для студентів
Етика, конфіденційність і політика	Ferguson R., Hoel T., Drachsler H. (2016) [36]	обґрунтували основи етики та конфіденційності в освітній аналітиці, запропонували таксономію ризиків і рішень для захисту даних.
	Tsai Y., Gasevic D. (2017) [102]	аналізували політику використання освітньої аналітики в різних країнах, вказуючи на необхідність чітких нормативів.

	Lang C., Macfadyen L., Slade S., Prinsloo P.(2018) [60] Rebecca Ferguson (2012) [35]	критичний аналіз відповідального використання LA, етика, політика і практика аналітики в європейському контексті.
Розробка і впровадження практичних рішень	Pistilli, Willis, Campbell (2014) [84]	досліджували, як освітня аналітика може бути інтегрована в інституційні стратегії для підвищення якості навчання.
	J. Campbell, D. Oblinger (2007), [19]	обґрунтовували підтримку викладача під час налаштування середовища навчання.
	Leitner P., Ebner M. (2017, 2019) [63]	узагальнили бар'єри та успішні практики впровадження аналітики у різних університетах Європи.
	H. Drachsler, W. Greller (2012), L. Macfadyen, S. Dawson, A. Pardo, D. Gasevic.[27,80, 60]	Розглядали питаннях реалізації та практичного застосування освітньої аналітики у закладах вищої освіти.
Побудова аналітичних моделей для прогнозування результатів навчання	Ryan S. Baker (2016) [11]	розробка моделей прогнозування успішності на основі поведінкових даних у навчальному середовищі.
	D. Gašević (2015) [41]	обґрунтовано інтеграцію моделей прогнозування з педагогічним дизайном.
Візуалізація даних та інформаційна підтримка викладачів	E. Duval (2012) [29]	один із піонерів візуалізації освітніх даних і концепції learner dashboards, розробка інтерфейсів для адаптивного навчання.
Персоналізація навчання та адаптивні системи	Bearman, M., & Ajjawi,R. (2023) [13]	Аналіз цифрових слідів для персоналізації контенту, створення адаптивного освітнього середовища, рекомендаційні системи.

Отже, сьогодні освітня аналітика розглядається як потужний інструмент для підвищення якості освіти завдяки можливостям прогнозування, персоналізації та підтримки прийняття рішень. Її розвиток відбувався під впливом численних наукових досліджень і практичних ініціатив, що заклали основи методів, технологій та етичних підходів.

Data Mining, освітній аналітичний інструмент, з'явився в 1978 році і здобув широке визнання в сучасному розумінні на початку 1990-х років. Традиційно обробка та аналіз даних здійснювалися в контексті прикладної статистики, головним чином для вирішення проблем, пов'язаних з обробкою невеликих баз даних. "Інтелектуальний аналіз даних" походить від двох концепцій: Data Mining (який виник з пошуку значущих даних у великих базах даних) і Mining (що

означає видобуток або копання) [32, 70]. Ці два процеси або потребують очищення величезних обсягів сирих даних, або інтелектуального пошуку значень у такому великому просторі. Data Mining також може означати вилучення даних, отримання інформації, копання даних, видобуток знань, аналіз шаблонів і видобуток знань у базах даних. Значний внесок у подальший розвиток теорії баз даних зробив американський математик Е.Ф. Кодд, який створив реляційну модель даних. [22] та Мічі створив метод, використовуючи алгоритм машинного навчання для розуміння як медичних даних [18].

Протягом 1990-х років Data Mining став популярним вибором як корисний та ефективний інструмент для аналізу великих і складних наборів даних. Інтернет і цифрові дані стали основною рушійною силою для Data Mining для бізнесу, урядів та академічних дослідників. [86, 111, 57, 101]. Інтелектуальний аналіз даних був ще більше вдосконалений на початку 2000-х років завдяки розвитку алгоритмів машинного навчання, зокрема нейронних мереж і глибокого навчання. Завдяки вдосконаленню технологій запису та зберігання даних, великі потоки "інформаційного видобутку" розвинулися в багатьох галузях, і з цим початкова обробка таких даних вимагає, зокрема: дані є гетерогенними (кількісними, якісними, текстовими); дані не обмежуються, результат повинен бути конкретним і зрозумілим; інструмент обробки сирих даних повинен бути простим. [101, 45, 137]. Багато вчених зробили внесок у галузь Data Mining, а саме в області інтелектуального аналізу асоціативних правил, кластеризації, паралельного та розподіленого інтелектуального аналізу даних, ансамблевого навчання, аналізу графів та ймовірнісних графічних моделей [45, 32, 57, 3]. Багато дослідників визначали цю концепцію, як показано в таблиці 1.3.

Таблиця 1.3

Трактування різних підходів поняття «Data Mining»

Parsaye K.A. Characterization of Data Mining Technologies and	Автором визначено декілька ключових характеристик інтелектуального аналізу даних, включаючи використання статистичних методів і методів машинного навчання, потребу в
---------------------------------------------------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------

Processes [83]	підготовці та очищенні даних, а також важливість знань предметної області. Автор визначає інтелектуальний аналіз даних як процес виявлення прихованих закономірностей і зв'язків у великих наборах даних.
Introduction to Data Mining and Knowledge Discovery: Third Edition. – Two Crows Corp. [105]	Data Mining - це процес, мета якого виявити нові значущі кореляції, зразки і тенденції в результаті просіювання великого об'єму даних з використанням методик розпізнавання зразків, застосування статистичних і математичних методів.
SAS Institute [93].	Data Mining - це процес виділення, дослідження і моделювання великих об'ємів даних для виявлення невідомих до цього структур (patterns) з метою досягнення переваг в бізнесі.
Data Science for Business: What you need to know about data mining and data-analytic thinking. Foster Provost & Tom Fawcett [38]	Інтелектуальний аналіз даних визначається як процес вилучення раніше невідомої та потенційно корисної інформації з великих наборів даних. Автори підкреслюють, що метою інтелектуального аналізу даних є не просто пошук цікавих шаблонів, а пошук шаблонів, корисних для прийняття рішень і вирішення бізнес-проблем.
IBM [18]	За визначенням IBM, інтелектуальний аналіз даних є критично важливим компонентом сфери бізнес-аналітики, оскільки він дозволяє організаціям отримувати цінну інформацію з великих і складних наборів даних. Підхід IBM до інтелектуального аналізу даних передбачає використання передових алгоритмів і аналітичних методів для виявлення закономірностей і тенденцій у даних, а також для створення прогнозів і рекомендацій на основі цього аналізу.

Джерело: узагальнено автором

Data Mining становлять велику цінність для керівників та аналітиків у повсякденній діяльності, за допомогою методів Data Mining вони можуть отримати відчутні переваги в конкурентній боротьбі. Data Mining є мультидисциплінарною галуззю, яка виникла і розвивається на базі досягнень прикладної статистики, розпізнавання образів, методів штучного інтелекту, теорії баз даних тощо (рис. 1.1).



Рис. 1.1. Схема Data Mining як мультидисциплінарної системи

Джерело: запропоновано автором

Data Mining має досить суттєві відмінності від інших методів аналізу даних. Традиційні методи аналізу даних (статистичні методи) і OLAP в основному орієнтовані на перевірку наперед сформульованих гіпотез і на "грубий" розвідувальний аналіз, що становить основу оперативної аналітичної обробки даних (Online Analytical Processing, OLAP), тоді як одне з основних положень Data Mining – пошук неочевидних закономірностей. OLAP більше підходить для розуміння ретроспективних даних, Data Mining спирається на ретроспективні дані для отримання відповідей на питання про майбутньому [86, 37]. У науковій літературі також підкреслюється важливість якості та підготовки даних у процесі інтелектуального аналізу даних [18, 45, 3, 38]. Це включає очищення даних, перетворення їх у формат, який можна легко проаналізувати, і вибір відповідних алгоритмів і методів для конкретних даних і аналізу.

Як правило, Data Mining включає наступні етапи [3,101]:

Збір даних: зібрана інформація може мати кілька форм, включаючи текст, аудіо, відео, категорії, числові дані та зображення.

Попередня обробка даних: щоб підготувати дані для аналізу, на цьому кроці їх очищають, конвертують та інтегрують. Основна мета попередньої обробки даних – переконатися, що дані точні, надійні та послідовні, щоб їх

можна було ефективно використовувати в аналітичних моделях.

Пошуковий аналіз: Метою дослідження даних є пошук залежностей, кореляцій, шаблонів і зв'язків, які можуть спрямовувати додатковий аналіз і моделювання.

Побудова моделі: Для створення моделі можна використовувати різні методи машинного навчання, дерева рішень, нейронні мережі або регресійний аналіз. У процесі розробки моделі необхідно вибрати відповідний алгоритм або модель для оцінки даних і створення прогнозів або класифікацій на їх основі.

Тестування моделі на наборі перевірки або через перехресну перевірку використовується для оцінки продуктивності моделі на цьому етапі.

Розвиток індустрії інтелектуального аналізу даних зумовлене декількома факторами, зокрема збільшенням доступності даних, зростаючим попитом на прогнозну аналітику та розробкою передових методів і інструментів інтелектуального аналізу даних [68]. Ще одна тенденція в індустрії інтелектуального аналізу даних полягає у використанні хмарних рішень [40, 23].

1.3. Обґрунтування доцільності використання інтелектуального аналізу даних у цифровому освітньому середовищі

Інтенсивний розвиток освітніх інформаційних технологій супроводжується накопиченням великих обсягів цифрової інформації щодо навчальних процесів. Застосування методів комп'ютерної аналітики для аналізу електронних форм освітньої інформації з метою виявлення прихованих знань дозволяє розширити можливості покращення вищої професійної освіти в країні. Як вже згадувалося, дослідження в галузі видобування освітніх даних проводяться в межах двох наукових напрямків: Educational Data Mining (EDM) та Learning Analytics (LA). Численні дослідження зарубіжних вчених стосуються специфіки застосування методів видобування даних (DM) в освітній сфері для допомоги учням, викладачам і менеджерам, моделювання поведінки студентів та проведення педагогічного прогнозування. (R. Baker [9,10], G. Cobo [21],

S. Gowda [43], C. Romero [90], M. Stewart [99], T. Feng [33]. Серед вітчизняних науковців, які проводять дослідження у сфері інтелектуального аналізу освітніх даних, можна відмітити роботи щодо дослідження сучасного стану та перспектив розвитку Educational Data Mining (Ю. Ковальчук [127]), оптимізації й аналізу результатів використання LMS Moodle засобами комп'ютерної аналітики (С. Петренко [131]), використання систем Data Mining для прогнозування педагогічних явищ та процесів (М. Коляда [128]). EDM є провідним напрямком досліджень, що має на меті вивчення внутрішньої структури даних, вилучення прихованих прогностичних факторів з них, а також розуміння освітніх наборів даних. [113].

Останнім часом передові технології, такі як Інтернет речей, датчики, штучний інтелект, та соціальні мережі інтегруються з освітньою системою для ефективного навчання [12, 15, 98]. Існує різноманіття методів аналізу даних, які практики можуть використовувати в освіті [14, 87].

Дослідження, проведені В. Tulasi [104] і іншими були спрямовані на реформи вищої освіти та дослідження рішень, отриманих з використанням обробки великих даних. Дослідження [91 та 104], аналізують та прогнозують продуктивність студента з використанням методів аналізу даних. У дослідженні [2] проаналізували різні аналітичні методи: освітній аналіз даних, Learning Analytics і Academic Analytics, і дійшли висновку, що застосування методів інтелектуального аналізу даних дають позитивні результати. Численні дослідники стверджують, що персоналізація в академічному середовищі дозволяє впровадження більш прогресивних і життєздатних форм навчання. Для досягнення цього використовуються різноманітні інструменти, такі як штучний інтелект, добування даних, машинне навчання, моделювання та статистика.

1.4. Структура наукового ландшафту освітньої аналітики на основі наукометричних баз даних

Для узагальнення важливості проведення бібліографічних досліджень здійснено опрацювання наукових розвідок, що дозволить зрозуміти, наскільки це

питання є актуальним у науковому дискурсі, з'ясувати тенденції у дослідженні вищезазначеного питання. Враховуючи останні тенденції аналізу наукових досліджень, досить популярним методом опрацювання літературних джерел стало використання багаторівневого бібліометричного аналізу в галузі EDM. Такий огляд літератури за допомогою структурного, компаративного та кластерного методів забезпечить формування методологічного та теоретичного підґрунтя для визначення основних трендів розвитку досліджуваних понять. Застосування аналітичних інструментів ScopusTools, та VOSViewer дасть можливість виявити основні зв'язки між наявними концепціями інтелектуального аналізу даних в сфері освіти та комп'ютерних наук та ідентифікувати нові й малодосліджені аспекти. Для проведення бібліометричного аналізу наукових досліджень поширеності поняття інтелектуального аналізу даних було використано наукометричну базу даних Scopus. Відбір релевантних публікацій здійснювався шляхом поєднання ключових слів «educational data mining», В наукометричній базі Scopus за пошуком «educational data mining» в галузі Computer Science було відібрано 2542 публікації, що складає понад 48% від загальної кількості публікацій за цим пошуковим запитом в розрізі 2010-2023 рр., що демонструє рис. 1.2.

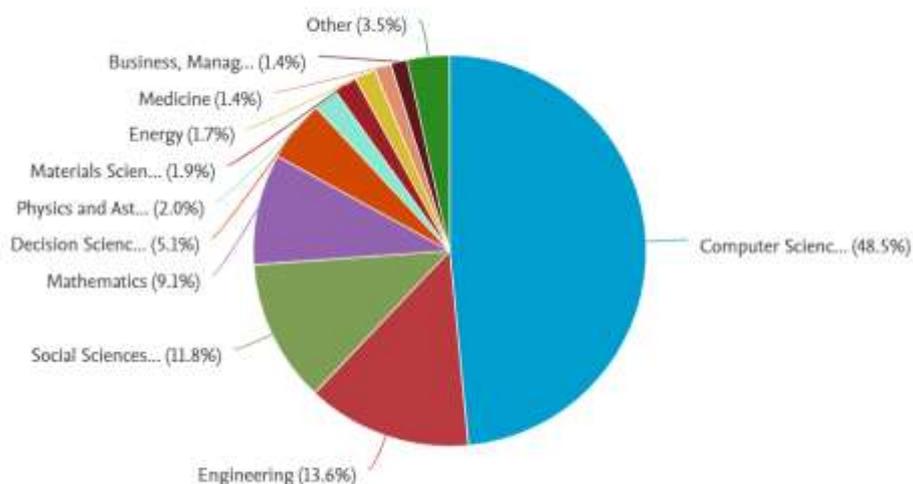


Рис.1.2. Структура публікацій за запитом «educational data mining» за науковими сферами (на основі бази даних Scopus)

Якщо аналізувати розподіл публікаційглі активності за роками, то рис.1.3. показує суттєвий щорічний приріст статей з використанням поняття «educational data mining», починаючи з 2015-2016 років.

Поєднання запропонованих ключових слів дозволило продемонструвати еволюцію досліджень. Зважаючи на кількість наукових публікацій та періоди досліджень, можна стверджувати, що поняття «educational data mining» почало більше використовуватись що пов'язаноз розвитком технологій великих даних та алгоритмів інтелектуального аналізу саме для освітніх даних.

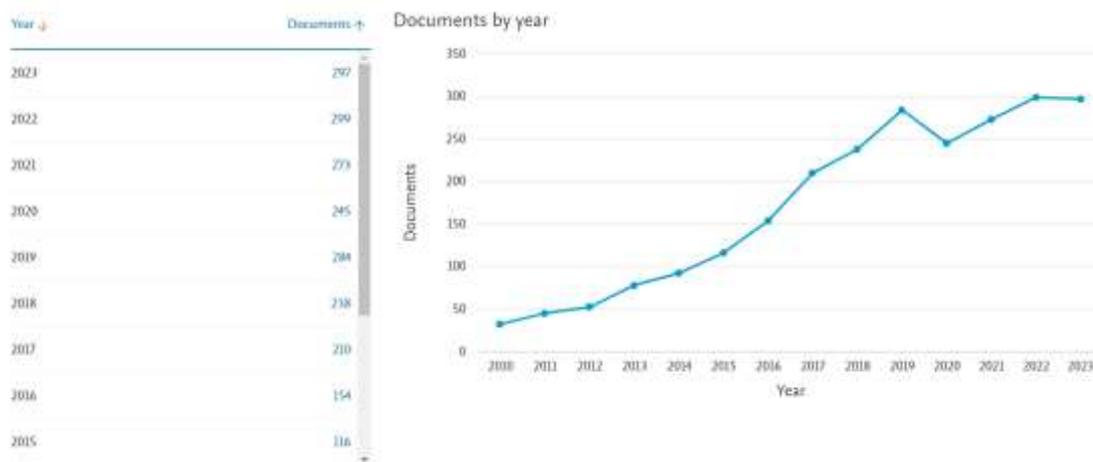


Рис.1.3. Динаміка кількості наукових публікацій, присвячених поняттю «educational data mining» (на основі бази даних Scopus)

Щодо типу публікацій то Scopus (Analyze search results), то майже 33% припадає на статті в журналах і, відповідно 67% - на публікації в рамках наукових конференцій.

Доцільним є дослідження географічної структури наукових досліджень інтелектуального аналізу даних (рис.1.4). Слід зазначити, що найбільшу кількість робіт було опубліковано вченими зі Сполучених Штатів Америки, Індії та Китаю.

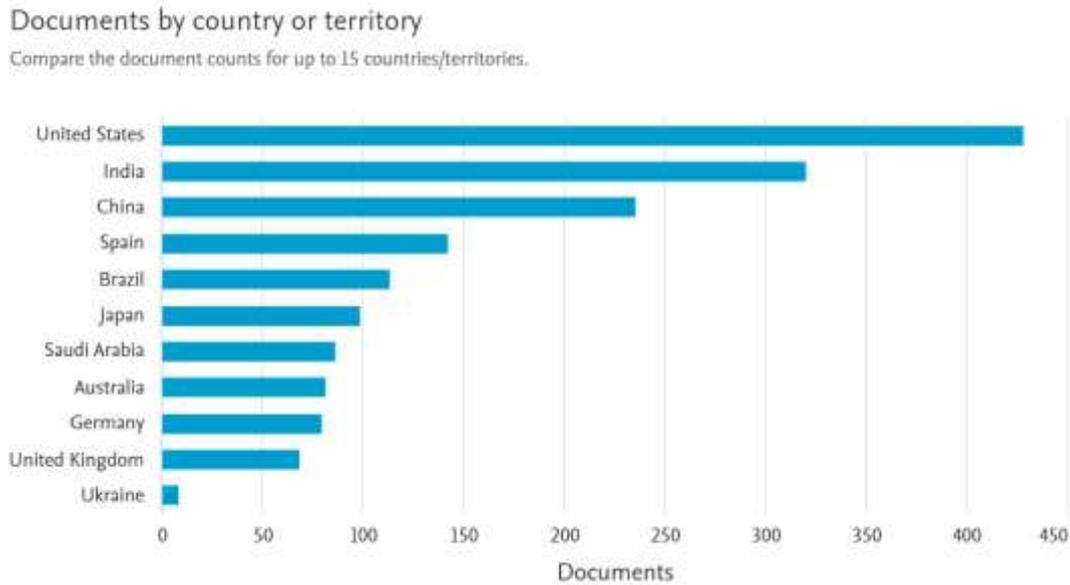


Рис. 1.4. Географічна структура наукових досліджень за запитом «educational data mining» (на основі бази даних Scopus)

Наступним кроком для узагальнення аналізу бібліографічних даних став кластерний аналіз публікацій, проіндексованих наукометричною базою даних Scopus та системою дослідницьких знань Dimensions.

Для цього ми використали аналітичну платформу VOSviewer. До переваг інструментарію програми належить ефективна візуалізація експортованих даних на основі ключових слів, цитувань та співцитувань. Для аналізу було взято 2542 наукових публікацій, проіндексованих наукометричною базою Scopus протягом 2010–2023 рр. Після програмного опрацювання даних було побудовано мапу наукових досліджень із даних за нашим запитом, яка демонструє взаємозв'язки між авторами, виданнями, цитуваннями й іншими категоріями. Проведено кластерізацію на основі кількості посилань та ресурсів, (рис.1.5).

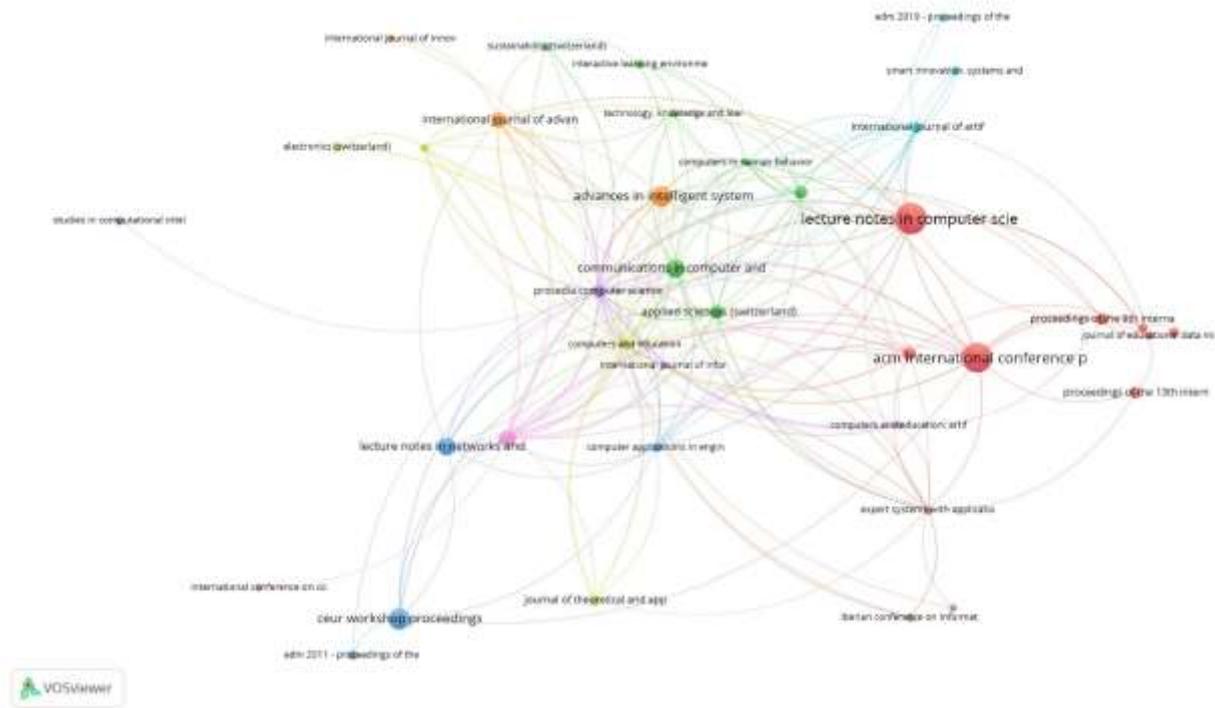


Рис. 1.5. Мапа наукових досліджень за запитом «educational data mining» (на основі бази даних Scopus)

Залежно від кількості праць у виданнях змінюється радіус кіл, що позначають самі видання. Кольором виділено кластери, на які було згруповано ресурси за (наукові журнали та апаці наукових конференцій).

Найбільші вузли відповідають виданням, в яких опублікована найбільша кількість документів та які найбільш цитовані.

Проведемо аналогічне дослідження з іншої бази – Dimensions за аналогічним запитом. Звузимо вибір і дослідимо тільки ті публікації, які були за двома сферами: Information and Computing Sciences та Education та Data Management and Data Science. Отримано 171 публікацію, що в хронологічному порядку підтверджуються стрімким приростом з 2015 року.

Щодо років публікації, бачимо, що потужні вузли мають недавні публікації. На побудованій карті чітко простежується перехід від ранніх публікацій до сучасних. Основні спостереження: Найпотужніші і найбільш цитовані вузли належать новішим публікаціям, переважно після 2017 року. Це свідчить про активну актуалізацію напрямку educational data mining останніми роками, зростання інтересу наукової спільноти та збільшення кількості практичних застосувань у сфері освітньої аналітики.

Таким чином, еволюційно-часовий аналіз підтверджує, що educational data mining є активно зростаючим і перспективним напрямом досліджень, який демонструє чіткий перехід від концептуальних робіт початку 2010-х до інтенсивних прикладних досліджень 2020-х років.

Також було побудовано мапу наукових досліджень за нашим запитом, яка демонструє взаємозв'язки між ключовими поняттями. Цікавими є дослідження карт ключових слів. Виділено чітко 2 кластери: один – це ключові слова, пов'язані з ключовим словом «education» та другий кластер – стійке утворення – пов'язане з поняттям «algorithm» (рис. 1.8.), що ще раз підкреслює мультидисциплінарність «educational data mining» та поєднання в цьому технічної та освітньої складової.

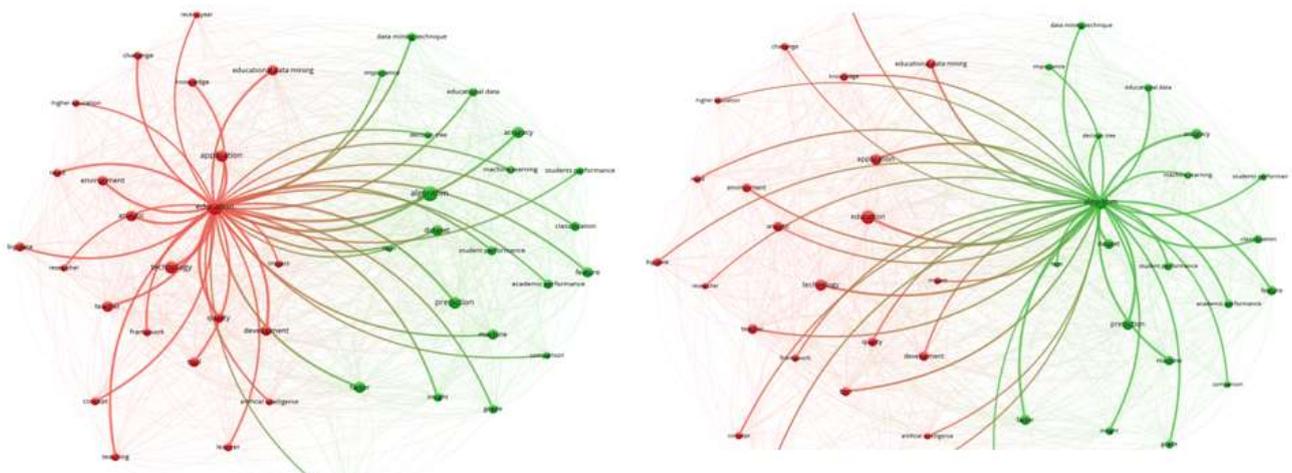


Рис.1.8. Міжкластерні взаємозв'язки ключових слів за запитом «educational data mining - цент кластеру -«education» та цент кластеру -«algorithm» (побудовано в програмі VOSviewer)

Отже, кластеризація ключових слів засвідчила існування двох домінантних

наукових напрямів: освітнього («education») та алгоритмічного («algorithm»), які взаємодіють і формують ядро міждисциплінарності цього поля. Така структура вказує на баланс між педагогічними задачами та застосуванням інтелектуального аналізу даних, що характеризує EDM як синергійний науковий напрям. Отримані результати підтверджують актуальність та складність проблематики, а також обґрунтовують необхідність подальшого розвитку інформаційних технологій для аналізу освітніх даних.

1.5 Інструментарій та завдання інтелектуального аналізу даних в освіті

Швидке зростання освітніх даних вказує на те, що дистилляція величезних обсягів даних потребує більш складного набору алгоритмів. Це питання призвело до появи сфери інтелектуального аналізу даних (EDM). Традиційні алгоритми інтелектуального аналізу даних не можна безпосередньо застосовувати до освітніх проблем, оскільки вони можуть мати конкретну мету та функцію.

Методи інтелектуального аналізу даних успішності студентів розглядалися багатьма дослідниками, зокрема, досліджували фактори, які були пов'язані з оцінками студентів наприкінці семестру, використовували можливості штучних нейронних мереж у моделюванні продуктивності студентів [17, 26, 81].

Існують й інші завдання, які можна вирішити за допомогою інтелектуального аналізу даних в освіті, такі як: прогнозування успішності студентів, аналіз використання студентами соціальних мереж та аналіз даних про онлайн-навчання [46]. На основі моделі відкритого навчання дослідники довели, що фонові характеристики не є добрими предикторами кінцевих результатів, чи досягне студент мети навчання чи ні, оскільки вони мають дані в початковій точці, тоді як інші фактори можуть сприяти труднощам, з якими студенти стикаються під час навчання [52, 25, 66].

Впровадження систем електронного навчання в освітній процес

супроводжується накопиченням великих обсягів інформації про освітній процес та цифровий слід викладачів і студентів. Застосування методів інтелектуального аналізу освітніх даних для аналізу цієї інформації та її візуалізації у вигляді інтерактивних звітів дозволяє виявляти приховані знання та закономірності, що значно покращують професійну підготовку майбутніх фахівців. Сьогодні навчальні заклади майже без винятку застосовують численні програмні комплекси для автоматизації поточних процесів у всіх основних сферах (прийом студентів, навчання, супровід студентів, забезпечення якості, управління тощо).

Чисельні дослідження зарубіжних науковців стосуються особливостей застосування методів DM в сфері надання освітніх послуг з метою надання рекомендацій студентам та викладачам, моделювання поведінки та профілю студентів, здійснення прогнозування освітніх траєкторій студентів та їх успішності [110, 80, 59, 121, 120]. Всебічного дослідження інтелектуального аналізу освітніх даних набули ці проблеми і в працях українських науковців як з педагогіки, так і з технічних наук, досліджуючи застосування методів Data Mining для підтримки прийняття рішень в освітній сфері для менеджменту якості освіти та адаптивного навчання, для підтримки інтерактивної діяльності всіх суб'єктів освітнього процесу. Вітчизняні дослідження у сфері інтелектуального аналізу освітніх даних, зокрема зосереджені на дослідженнях сучасного стану та перспектив розвитку Educational Data Mining, підготовки майбутніх фахівців з ІТ до здійснення освітньої аналітики, оптимізації й аналізу використання Big Data LMS Moodle, використання систем Data Mining для прогнозування освітніх траєкторій [125, 127, 77,74 , 131, 123]. Однак певні аспекти такого аналізу вивчені недостатньо та потребують подальшого дослідження.

Впродовж останніх років досвід експлуатації систем електронного навчання в вищій школі виявив певні суттєві недоліки, що негативно впливають на ефективність електронного та змішаного навчання. До них можна віднести погіршення та ослаблення зв'язку між студентом та викладачем, обумовлене суттєвим зменшенням їх безпосереднього спілкування [7,135]. Разом з тим,

дослідження у сфері інтелектуального аналізу освітніх даних є недостатньо системні та розрізнені. Не вирішена проблема якісних та кількісних освітніх вимірів, не недостатньо дослідженими є питання впровадження інтелектуального аналізу в освітню практику закладів вищої освіти. Все це, безумовно, посилює необхідність проведення подальших наукових досліджень щодо застосування методів та розв'язання задач Data Mining у вищій освіті в системах електронного навчання [28, 125]. Розробка інформаційних систем та технологій освітньої аналітики стає важливою науково-практичною задачею з обов'язковим залученням DM (рис 1.9).

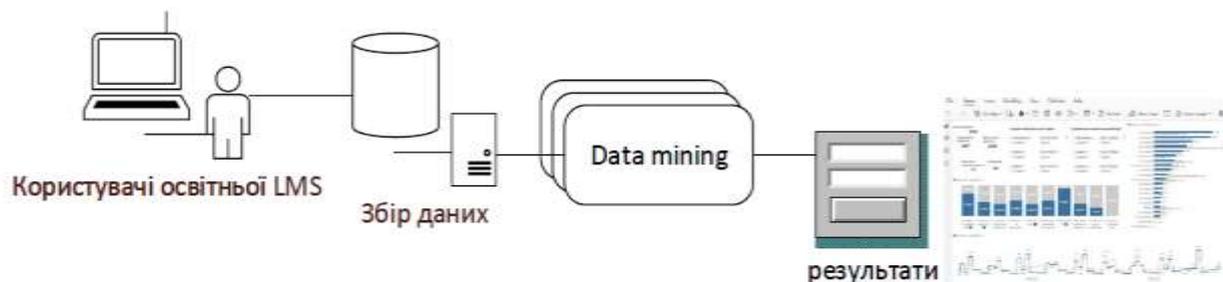


Рис. 1.9. Принципова схема процесу Educational Data Mining

Застосування методів і технологій Data Mining дає змогу розв'язати такі задачі [134, 48]: класифікація (Classification); кластеризація (Clustering); асоціація (Associations); послідовність (Sequence), або послідовна асоціація (sequential association); прогнозування (Forecasting); визначення відхилень (Deviation Detection), аналіз відхилень або викидів; оцінювання (Estimation); аналіз зв'язків (Link Analysis); візуалізація (Visualization, Graph Mining); підбивання підсумків (Summarization) – опис конкретних груп об'єктів за допомогою аналізованого набору даних.

Технології Data Mining використовують велике число методів, частина з яких запозичена з інструментарію штучного інтелекту, іншу частину складають або класичні статистичні методи, або інноваційні методи, пров'язані з використанням інформаційних технологій та систем. Перший рівень методів Data Mining базується на тому, чи зберігаються дані після опрацювання, чи вони трансформуються для подальшого використання. На рис. 1.10. показано ієрархію

методів Data Mining, де відображені тільки основні напрямки методів, причому розгалуження можна продовжувати, через те, що низка наведених методів, включають багато різновидів.

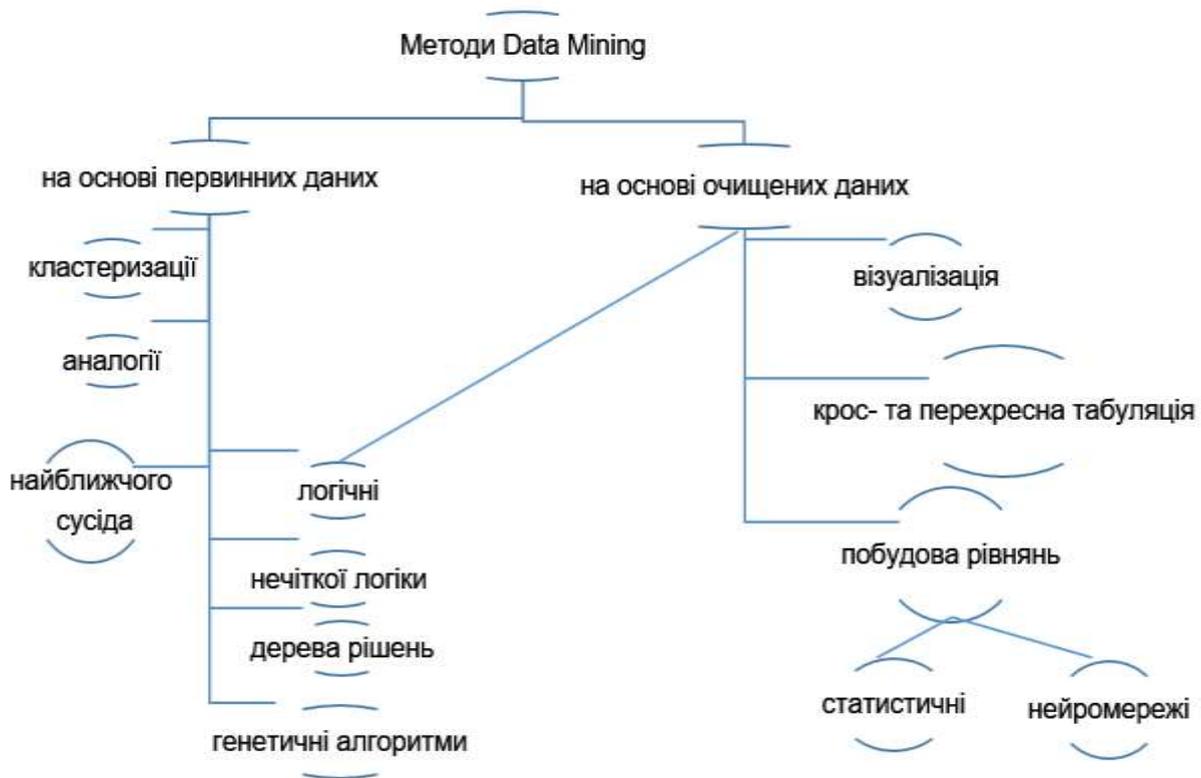


Рис. 1.10. Ієрархія методів Data Mining

У результаті проведеного аналізу було встановлено, що інтелектуальний аналіз освітніх даних є синтезом методів та засобів для розуміння й прогнозування освітніх ситуацій та розробки й використання програмного забезпечення для їх реалізації. Однак реалізація цих методів в освітній сфері для аналізу цифрових даних стосовно процесів навчання має свої особливості, обумовлені цілями аналізу та специфікою даних, які аналізуються. Тут потенційним джерелом знань для дослідника все частіше є як адміністративні бази освітніх даних рівня навчального закладу, регіону чи держави, так і Web, і бази даних і лог-файли різноманітних систем комп'ютерної підтримки навчання – CMS, LMS, ITS, системи комп'ютерного адаптивного навчання, тестування рівня навчальних досягнень тощо. Інформація про діяльність студента у середовищі електронного навчання представлена у вигляді цифрових слідів, які

містять: 1) дані про дії у системі: ідентифікація користувача, час доступу, дія та засіб з навчальним контентом, який використовувався; 2) академічні дані: підсумкова оцінка за курс, поточні оцінки; 3) час сеансу; 4) рівень активності студента.

Освітні дані, що підлягають аналізу, зазвичай мають складну структуру чи слабоструктуровані, представлені в різних системах навчання та є не завжди зрозумілими для працівників сфери освіти, оскільки вони є цифровими слідами, залишеними у логах та базах даних і стосуються різних активностей студентів у середовищі електронного навчання. Процес інтелектуального аналізу освітньої інформації в системах електронного навчання містить наступні етапи:

- етапу відбору даних: на цьому етапі відбувається ретельний відбір даних за обраним критерієм включаючи доступність даних, їх якість, тип і формат, а також семантику;
- етапу попередньої обробки: на цьому етапі здійснюється вибір підходящих стратегій по масштабуванню й нормалізації характеристик даних, а також вибір стратегії для обробки відсутніх значень атрибутів;
- етапу трансформації даних: на цьому етапі використовуються методики по зменшенню розмірності даних;
- власне етапу Data Mining - видобутку знань: на цьому етапі здійснюється застосування алгоритмів інтелектуального аналізу освітніх даних;
- етапу інтерпретації й оцінки.

Основні зусилля при проведенні інтелектуального аналізу освітніх даних направлені на адекватну підготовку та обробку даних перед тим, як до них будуть застосовані певні алгоритми аналізу [7, 16].

Застосування окремих методів Data Mining для аналізу такого роду інформації дозволяє виявляти приховані закономірності та знання, які традиційними методами аналізу отримані бути не можуть [16]. Кожен з методів призначений для розв'язання певної задачі, серед яких можна виділити основні задачі, результати та методи Data Mining (табл.1.4)

Таблиця 1.4

Задачі освітньої аналітики та методи Data Mining

Задача аналізу	Результати	Методи, що забезпечують вирішення задачі
Класифікація	встановлення чітких кількісних, статистично значимих залежностей між вхідними і дискретними вихідними змінними, які характеризують процес навчання, що дає можливість провести класифікацію об'єктів до одного зі заздалегідь відомих класів. Це дозволяє здійснювати класифікацію студентів залежно від їх попередньої чи поточної успішності чи активності у системі LMS та класифікацію ресурсів навчання	дискримінантний аналізу, Naive Bayes, k-ближніх сусідів, дерева рішень, нейронні мережі.
Прогнозування	задача передбачення значення досліджуваної величини, на основі відомих попередніх значень, що характеризують процес навчання та суб'єктів навчання. Це дозволяє моделювати та прогнозувати поведінку студентів у процесі навчання та встановлювати залежність між іншими величинами, що стосуються електронного навчання.	методи математичної статистики, нейронні мережі, часові ряди.

Кластеризація	групування об'єктів на основі різноманітних даних, що описують їх сутність. Результатом кластеризації є поділ об'єктів, які стосуються навчання, на групи споріднених, схожих об'єктів – кластери. Задача кластеризації є логічним продовженням ідеї класифікації, однак при проведенні кластеризації кількість кластерів заздалегідь невідома і визначається у процесі аналізу.	методи ієрархічного кластерного аналізу, методи k-середніх та c-середніх.
Пошук асоціативних правил та послідовностей та отримання нових знань за допомогою моделей (Discovery with Models)	дозволяє виявляти взаємозв'язки між пов'язаними подіями у наборі освітніх даних за прямими та непрямими ознаками. Розв'язання цієї задачі дозволяє виявляти правила виду «якщо умова, то наслідок», де «умова» та «наслідок» є подіями, які відбуваються у середовищі електронного навчання й мають високу ймовірність одночасної та послідовної появи.	методи машинного навчання
Візуалізація, Visual Mining	створення візуального образу аналізованих даних в режимі реального часу шляхом перетворення великих масивів цифрових даних	методи відображення складної, багатомірної інформації, спеціальні засоби аналітики (Power BI)

Кластеризація в системах електронного навчання застосовується, зазвичай, для розбивки студентів на групи, які характеризуються близькими значеннями деяких числових або якісних показників. Це дозволяє структурувати дані у випадку, коли їх структура невідома. Наприклад, студенти можуть бути розбиті на групи по подібності освітніх програм, кваліфікації, спільності цілей або інтересів, мережевій активності тощо. Для цього використовуються методи кластеризації, розроблені в прикладній статистиці, кластерному аналізі й обчислювальній математиці: ієрархічні алгоритми, алгоритм k-середніх, нечіткі алгоритми кластеризації, нейронні мережі.

Класифікація також дозволяє розбивати дані, які аналізуються, на групи споріднених об'єктів, однак кількість таких груп відома наперед. Для здійснення класифікації використовують алгоритми k-ближніх сусідів, Байєса, покриття, дерева рішень, метод опорних векторів, нейронні мережі.

Асоціативні правила застосовуються для формалізації шаблонів поведінки студента в електронному навчальному середовищі. З їх допомогою створюються типові траєкторії навчання й структури курсу, орієнтовані на цільову аудиторію або окремих споживачів освітніх послуг. Для розв'язання таких задач застосовується апарат нечіткої математики, алгоритм Аргіогі.

Розв'язання визначених задач із використанням методів інтелектуального аналізу освітніх даних складається з наступних етапів: очистка, фільтрація попередня обробка даних; виявлення закономірностей у даних на основі математичних методів; перевірка (валідація) виявлених закономірностей та моделей; прогнозування майбутніх подій у середовищі навчання на основі прогностичних моделей; використання результатів аналізу для підтримки прийняття рішень і вироблення освітньої політики.

Візуалізація дозволяє у зрозумілому для сприйняття вигляді відображати інтегровану інформацію стосовно процесів, пов'язаних з навчанням у середовищах електронного навчання. Виявлення зв'язків дозволяє моделювати освітні процеси та використовувати побудовані моделі для прогнозування

майбутньої поведінки студентів при вивченні курсу, виявляти студентів, які мають ризики неуспішності.

Основні функції, які надає програмне забезпечення інтелектуального аналізу даних (data mining), включають можливості підготовки даних, вбудовані алгоритми, підтримку прогнозного моделювання, середовище розробки на основі графічного інтерфейсу користувача та інструменти для розгортання моделей і оцінки їх ефективності. Огляд платформ які пропонують інструменти для інтелектуального аналізу наведено на рис. 1.11.

Додаткові високорівневі послуги можуть бути отримані від постачальників хмарних сервісів, які сприяють передачі великих даних, включаючи такі, як Amazon EMR (раніше Elastic MapReduce), Microsoft Azure HDInsight або Google Cloud Dataproc, через керовані сервіси. Великі дані можуть бути розміщені в хмарній системі.

Це програмне забезпечення допомагає збирати, зберігати, аналізувати, звітувати та візуалізувати результати досліджень роботи з великими даними. Як результат, загальнодоступні хмарні обчислення зараз є основним засобом розміщення систем великих даних. На додаток до складності обробки обсягу та вартості, проектування архітектури великих даних є ще однією великою проблемою, з якою стикаються користувачі.

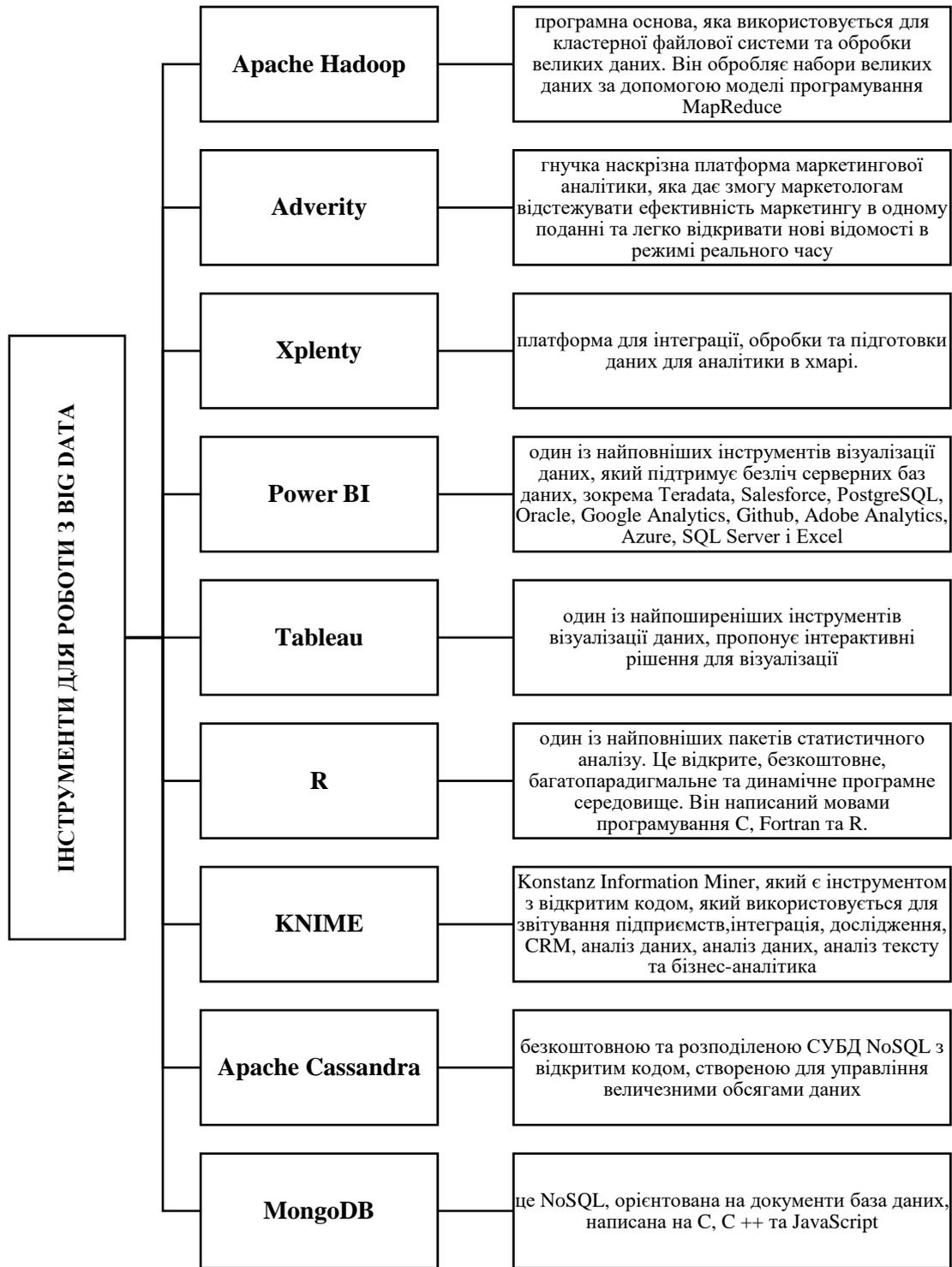


Рис. 1.11. Огляд інструментів роботи з BIG DATA

Тому ці системи великих даних потрібно налаштовувати відповідно до вимог освітніх закладів, що ставить перед ІТ-командами та розробниками додатків завдання не лише знайти "інструментарій" для всіх доступних технологій, але й розробити інформаційні технології, які дозволяють

отримувати, зберігати, обробляти та поширювати такі дані для всіх рівнів і учасників освітнього циклу будь-якого освітнього процесу.

Висновки до розділу 1

У розділі 1 здійснено комплексний теоретико-аналітичний огляд предметної області освітньої аналітики та інтелектуального аналізу освітніх даних, що дозволило сформулювати цілісне бачення сучасних підходів, понять, методів та інструментів, які формують наукове підґрунтя дослідження. Це дало можливість узагальнити практичні підходи до цифрової трансформації освітнього процесу через впровадження інструментів освітньої аналітики, які сприяють прийняттю обґрунтованих управлінських рішень, персоналізації навчання та підвищенню якості освіти в українських закладах вищої освіти.

На основі аналізу понятійного апарату встановлено, що Learning Analytics та Education Data Mining є взаємопов'язаними, але не тотожними напрямками: LA орієнтується на підтримку прийняття рішень людиною за рахунок візуалізації й інтерпретації даних, тоді як EDM акцентує на автоматизованих алгоритмічних методах виявлення закономірностей. Отже, був уточнений понятійно-термінологічний апарат Learning Analytics та EDM, адаптований до українського освітнього середовища.

Проведений огляд задач і методів Learning Analytics підтвердив, що сучасні освітні системи активно застосовують моделі інтелектуального аналізу даних, зокрема, класифікації, кластеризації, регресійного прогнозування, пошуку асоціацій, виявлення поведінкових патернів, аналізу соціальних мереж, та мультимодальної аналітики.

Проведений кластерний аналіз публікацій, проіндексованих наукометричною базою даних Scopus та системою дослідницьких знань Dimensions, за запитом «educational data mining» з використанням аналітичної платформи VOSviewer підтвердив стрімке зростання кількості наукових публікацій у сфері LA та EDM, а також виділив ключові дослідницькі кластери з центрами в поняттях «education» та «algorithm»

У ході дослідження доведено необхідність застосування методів інтелектуального аналізу даних в освіті, зумовлену зростанням обсягів цифрової активності студентів, переходом ЗВО до LMS та інших освітніх платформ, потребою у прийнятті швидких та обґрунтованих управлінських рішень. Внаслідок порівняльного аналізу визначено набір принципів (модульність, масштабованість, конфіденційність, адаптивність) та конкретних функцій (моніторинг, діагностика, прогнозування, рекомендація), які повинні реалізовуватися в аналітичних системах ЗВО що поєднують методологію EDM/LA з архітектурними вимогами LMS-інтеграції.

Узагальнення результатів, отриманих у розділі 1 дозволило сформулювати теоретичні, методологічні та концептуальні передумови для розроблення інформаційної технології освітньої аналітики, визначити вимоги до її архітектури, критерії якості, необхідні моделі та алгоритми.

Таким чином, у розділі 1 виконано всебічний аналіз предметної області та визначено методологічні, математичні й інструментальні передумови для наступних етапів дослідження. А також набули подальшого розвитку теоретичні засади освітньої аналітики, зокрема визначення структури, функцій та принципів побудови інформаційних аналітичних систем у сфері вищої освіти, що базуються на інтеграції методів інтелектуального аналізу даних із системами управління навчанням.

РОЗДІЛ 2

ОБГРУНТУВАННЯ ТА СКЛАДОВІ ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ ОСВІТНЬОЇ АНАЛІТИКИ

2.1. Джерела даних і функціональні компоненти інформаційної технології освітньої аналітики

У сучасних умовах цифрової трансформації освіти інформаційна аналітика виступає ключовим інструментом для підтримки прийняття рішень на всіх рівнях управління освітнім процесом. Необхідність розробки та впровадження інформаційної технології освітньої аналітики зумовлюється зростаючим обсягом освітніх даних, потребою в персоналізованому підході до навчання, підвищенням вимог до прозорості та підзвітності освітніх установ, а також необхідністю оперативного моніторингу якості освіти.

Загальнонаукове поняття інформаційної технологія (ІТ) включає сукупність методів, засобів і процесів, що забезпечують збір, зберігання, обробку, передачу, захист і відображення інформації з метою підтримки прийняття рішень, оптимізації діяльності та автоматизації процесів у різних сферах людської діяльності. З прикладної точки зору інформаційні технології включають інтеграцію комп'ютерного обладнання, програмного забезпечення, хмарних сервісів, інтерфейсів користувача та алгоритмів інтелектуального аналізу даних, що застосовуються для вирішення конкретних задач управління, моніторингу та оптимізації діяльності організацій. Для узагальнення цього поняття наведемо ще декілька тлумачень. Керуючись *ISO/IEC 38500:2015*. «**Information Technology (IT)** refers to the use of computers, networking, storage, and other physical devices, infrastructure and processes to create, process, store, secure, and exchange all forms of electronic data» (це використання комп'ютерів, мереж, систем зберігання даних та інших фізичних пристроїв, інфраструктури та процесів для створення, обробки, зберігання, захисту та обміну всіма видами електронних даних.) [51]. За визначенням з *Encyclopaedia Britannica* [50] "Information Technology." «**Information technology** is the use of computers to store,

retrieve, transmit, and manipulate data or information, often in the context of a business or other enterprise» (це використання комп'ютерів для зберігання, пошуку, передачі та обробки даних або інформації, часто в контексті бізнесу або інших підприємств). Визначення з відомого підручника по ІТ трактує це «IT comprises hardware, software, data, people, and processes that organizations use to produce information and support operations and decision-making» (ІТ включає в себе апаратне забезпечення, програмне забезпечення, дані, людей та процеси, які організації використовують для створення інформації та підтримки операцій і прийняття рішень.) [61]. Також наведемо визначення в контексті освіти ЄС: «Information and Communication Technologies (ICTs) in education refer to the integration of digital tools, such as computers, tablets, learning platforms, and the internet, to improve teaching, learning, and educational management» (ІКТ в освіті означають інтеграцію цифрових інструментів, таких як комп'ютери, планшети, навчальні платформи та Інтернет, з метою поліпшення викладання, навчання та управління освітою) [30]. З освітньо-управлінської точки зору «Інформаційні технології в освіті це комплекс апаратно-програмних і комунікаційних засобів, що забезпечують підтримку навчальної, наукової та адміністративної діяльності в закладах освіти [138,119].

Основною метою побудови інформаційної технології освітньої аналітики є створення інструментарію для системного збору, зберігання, обробки, аналізу та візуалізації освітніх даних з метою виявлення закономірностей, прогнозування результатів навчання та підвищення ефективності управлінських і педагогічних рішень. Вона дозволяє інтегрувати різноманітні джерела інформації: системи управління навчанням (LMS), електронні журнали, платформи електронного документообігу, опитувальники, результати ЗНО та інші державні інформаційні системи (ЄДЕБО, ДІСО, ІСУО тощо).

Інформаційне забезпечення процесів трансформації освіти є одним із ключових інструментів ефективного державного управління та модернізації освітньої системи [129, 132]. З огляду на необхідність забезпечення рівного

доступу до якісної освіти, справедливого розподілу ресурсів і підвищення ефективності управлінських рішень, важливою складовою виступає побудова й впровадження системного моніторингу освітньої сфери. Цю функцію виконують сучасні інформаційні системи управління освітою (Education Management Information System – EMIS), які забезпечують збір, обробку, зберігання, аналіз і розповсюдження освітньої статистики на базі інтегрованого цифрового середовища.

EMIS не лише автоматизує формальні адміністративні процеси, але й формує інституційну культуру управління на основі достовірних, актуальних і доступних освітніх даних. Інформація, що генерується в таких системах, має стратегічне значення для формування освітньої політики, планування бюджету, розподілу людських ресурсів і впровадження реформ. У контексті реформування освіти, особливо в умовах децентралізації, цифрової трансформації та впливу глобальних викликів, таких як пандемія COVID-19 та російська агресія, роль EMIS істотно зростає.

Суттєвим підтвердженням важливості інформаційного виміру сучасного суспільства став Всесвітній саміт з питань інформаційного суспільства (World Summit on the Information Society, WSIS), що відбувся в Женеві у березні 2018 року. Центральною темою обговорень на цьому заході стало представлення найуспішніших національних стратегій розвитку інформаційного суспільства в контексті досягнення Цілей сталого розвитку. Зокрема, обговорювалися виклики цифрової нерівності, необхідність зміцнення довіри в цифровому середовищі, проблеми кібербезпеки, а також шляхи інтеграції цифрових технологій у ключові сектори, включно з економікою, торгівлею та освітою.

Міжнародна практика демонструє різні підходи до впровадження та функціонування EMIS, зокрема, Австралії департаменти освіти штатів акумулюють статистичні й аналітичні дані в шестимісячному циклі, освітні дані згруповані на єдиному державному порталі, що створює умови для оперативного аналізу та підтримки управлінських рішень [88]. У Франції EMIS функціонує в

умовах високої децентралізації. Велика увага приділяється аналізу швидкості поширення інформації між адміністративними рівнями, достовірності звітності та ефективності процедур збирання статистики. Литовська система ŠVIS/EMIS – приклад сучасної освітньої інформаційно-аналітичної системи, що відповідає вимогам гнучкості, масштабованості та захисту персональних даних. [6]. У країнах, що розвиваються, розвиток EMIS часто підтримується міжнародними фінансовими організаціями. Світовий банк профінансував понад 230 проєктів з впровадження EMIS у 89 країнах, приділяючи особливу увагу секторам професійної освіти. [69, 100].

В Україні сформовано декілька інформаційних систем, що виконують функції, подібні до EMIS. Найбільш відомими є ІСУО, ДІСО, ЄДЕБО та АС «Школа». ЄДЕБО функціонує як національна база даних про вищу освіту, охоплюючи інформацію про дипломи, результати ЗНО, вступну кампанію, студентські квитки тощо. Єдина державна електронна база з питань освіти (ЄДЕБО) забезпечує органи державної влади, органи місцевого самоврядування, фізичні і юридичні особи інформацією про заклади вищої освіти, забезпечує збереження інформації про документи про освіту, наукові ступені, результати зовнішнього незалежного оцінювання, перебіг вступної кампанії до закладів освіти, студентські квитки державного зразка, іншої освітньої інформації. Власником ЄДЕБО є держава, розпорядником та власником інформації ЄДЕБО є Міністерство освіти і науки України. Головна цінність інформаційної технології освітньої аналітики – це практична спрямованість даної технології, шлях від сирих даних до конкретного знання, від постановки завдання до готового додатку, за підтримки якого можна приймати рішення. Інформаційна технологія повинна зв'язати в одне ціле задачі, методи, дії, закономірності, додатки, дані, інформацію, рішення.

Розглянемо потік « Дані – інформація – рішення» результатом якого має бути знання і прийняття рішення. На рис. 2.1 показано зв'язок понять «дані», «інформація» і «рішення», яка виникає в процесі прийняття рішень.



Рис. 2.1. Схема перетворення даних в інформацію та рішення

Як видно з рисунку, цей процес є циклічним. Прийняття рішень потребує інформації, яка заснована на даних. Дані формують інформацію, яка підтримує рішення і т.д. Саме ефективне використання освітніх даних через інформаційні системи та технології здатне створити підґрунтя для справедливого, інклюзивного та якісного освітнього середовища в Україні та світі.

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.

Величезне збільшення обсягів даних, що накопичуються в сфері освіти, з одного боку, і зростання потужності доступних програмних засобів і машинних алгоритмів для аналізу даних, з іншого, призвели до розвитку напрямків інтелектуального аналізу даних та аналітики навчання. Заклади освіти повинні використовувати інтелектуальний аналіз даних для вилучення корисної інформації зі своїх баз даних та інших зовнішніх платформ. Вищі навчальні заклади часто використовують системи управління навчанням (LMS) для підтримки процесу викладання та навчання. Найбільш широко використовуваною LMS в Україні, зокрема і в НУБІП України, є Moodle

(*Modular Object-Oriented Dynamic Learning Environment*). Діяльність кожного студента в Moodle фіксується у файлі журналу, коли він користується своїм особистим кабінетом. Такі дані є цінним джерелом для аналітичних процедур [47, 122].

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, зокрема, дані освітніх сервісів та платформ управління навчанням або дані пошукових систем.

Зокрема, джерела надходження даних про освітній процес узагальнено на рис 2.2.



Рис.2.2. Канали надходження даних про освітній процес

Освітні дані, які генеруються різними джерелами, як внутрішніми, так і зовнішніми щодо закладів освіти можна розділити на:

- дані про студента, такі як демографічні дані та попередня успішність;
- дані про викладача, такі як компетенції та професійний досвід.
- дані, створені під час процесів викладання навчання та оцінювання як у фізичних приміщеннях, так і за його межами;
- інфраструктурні та соціальні дані та фінансовий план, включаючи освітній та ненавчальний персонал, апаратне/програмне забезпечення, витрати.

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 (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 здобувачів освіти.

Метою закладів освіти в рамках надання освітніх послуг, зокрема і НУБП 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. EDM використовує алгоритми для різних типів освітніх даних з метою аналізу цих даних і пошуку рішень освітніх проблем [85, 132]. Іншими словами, EDM в основному займається розробкою методів і технік для дослідження зв'язку між різними типами даних, зібраних з навчальних закладів, щоб отримати краще розуміння студентів та їх навчального контексту. На рис. 2.3 показано місце застосування інтелектуального аналізу даних в інформаційній технології.



Рис.2.3. Складові інформаційної технології освітньої аналітики

Загальна архітектура технології базується на модульному принципі, що дозволяє незалежно розробляти, впроваджувати та оновлювати її компоненти без порушення цілісності системи.

Інформаційна технологія освітньої аналітики включає наступні основні модулі та процедури роботи з даними (рис.2.4)

- 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.
- 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
- Модуль інтелектуального аналізу реалізує алгоритми машинного навчання (класифікація, регресія, кластеризація, асоціативний аналіз) та інтерпретаційні моделі (SHAP) для виявлення закономірностей та прогнозування освітніх результатів.
- 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.
- Модуль управління доступом і безпеки відповідає за автентифікацію користувачів, управління ролями, контроль доступу до окремих показників та персоналізованих даних.

Основні потоки даних в системі реалізуються у вигляді послідовної ETL-процедури (Extract – Transform – Load), яка охоплює:

- вилучення даних з LMS, ЄДЕБО та інших джерел;
- трансформацію до єдиного формату (CSV);
- завантаження до централізованого сховища;

- обробку та подання результатів через аналітичні панелі(Power BI).

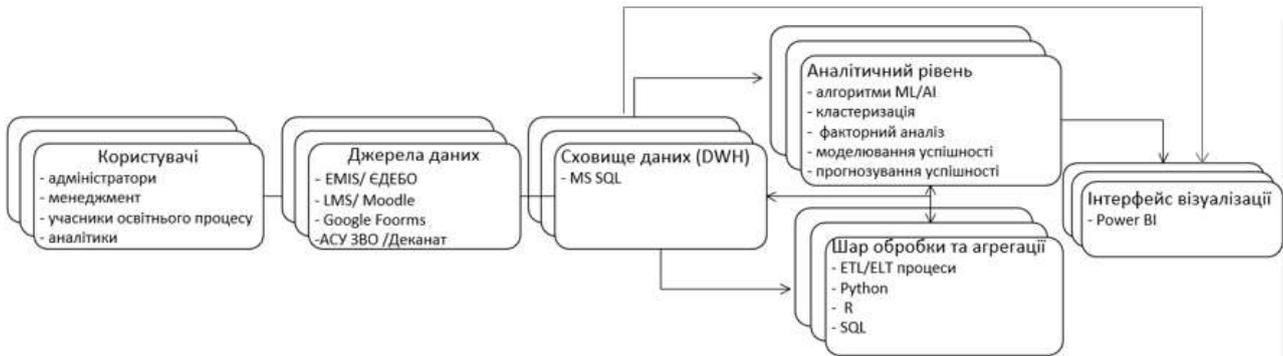


Рис. 2.4. Компоненти інформаційної технології освітньої аналітики

Варто зауважити, що на першому етапі необхідно зібрати дані з LMS, ЄДЕБО та університетської ІС «Деканат»; на другому етапі потрібно забезпечити середовище для зберігання та обробки даних, що передбачає побудову окремої бази даних або компонування моделі даних, автоматизувавши щоденне оновлення даних; на третьому етапі потрібно забезпечити використання моделей ML/AI з впровадженням аналізу ризиків неуспішності здобувачів освіти; на заключному етапі необхідно запровадити можливість формування та публікації аналітичних панелей (дашбордів)– відкритих даних в спільному цифровому середовищі НУБІП України та про успішність, динаміку відвідуваності – на рівні доступності керівників структурних підрозділів НУБІП України.

Центральним елементом технології 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. За метою цього дослідження ми проаналізували різні інструменти обробки великих наборів даних. Завдяки логічній обробці великих обсягів структурованих і неструктурованих даних вони дають менеджерам різних рівнів управління освітнім процесом та здобувачам освіти можливість легко отримувати цінну інформацію про освітній процес. Крім того, вони відіграють ключову роль у створенні аналітики в режимі реального часу, що робить їх незамінними для підтримки прийняття управлінських рішень.

Співставний аналіз цих інструментів наведено в таблиці 2.1.

Таблиця 2.1

Аналіз інструментів аналітики освітніх даних

	Опис інструментів	Переваги	Недоліки	Вартість
Zoho Analytics	<ul style="list-style-type: none"> - аналітика великих даних - візуальний аналіз і панель приладів - різноманітні компоненти - вбудовування звіту - спільний аналіз звітності розширювана платформа розробки - висока безпека 	<ul style="list-style-type: none"> - вбудований інструмент ВІ з низьким кодом; - низька загальна вартість; - простий процес розгортання; - стандарти безпеки 	ускладнена крива навчання	15-денна безкоштовна пробна версія (\$14.29 на місяць)
Apache Hadoop	<ul style="list-style-type: none"> - розподілене зберігання даних - відмовостійкість - гнучка обробка - підтримка спільноти з відкритим кодом - локальність даних - підтримка кількох NameNodes 	<ul style="list-style-type: none"> - вдосконалення автентифікації за допомогою проксі-серверів http - специфікація Hadoop Compatible Filesystem - надійна екосистема технологій та інструментів великих даних 	початкове налаштування для використання функцій сумісної файлової системи hadoop складне	безкоштовна і має відкритий код
Apache Storm	<ul style="list-style-type: none"> - розподілена система - відмовостійкість - мовна підтримка - висока пропускна здатність - висока масштабованість 	<ul style="list-style-type: none"> - аналітика в реальному часі - вбудована відмовостійкість - підтримка різних джерел даних та інтеграція з різними системами 	налаштувати і оптимізувати затримку може бути складно порівняно з деякими іншими платформами	безкоштовна і має відкритий код
Apache Cassandra	<ul style="list-style-type: none"> - висока доступність і відмовостійкість - масштабованість без простоїв - гнучка модель даних - оптимізація продуктивності 	<ul style="list-style-type: none"> - забезпечує продуктивність і низьку затримку для додатків - призначений для роботи з великими даними з лінійною масштабованістю в міру зростання даних і трафіку 	складність у налаштуванні та обслуговуванні	безкоштовна і має відкритий код

	надійність відкритого коду підтримка кількох центрів обробки даних вбудоване стиснення даних			
Cloudera	<ul style="list-style-type: none"> - об'єднує широкий спектр інструментів для обробки великих даних, машинного навчання та аналітики - надає підтримку та послуги, підвищуючи надійність і продуктивність платформи за підтримки спільноти Cloudera та експертів 	<ul style="list-style-type: none"> - пробна комплексна платформа управління даними - ефективні можливості машинного навчання - уніфікована безпека даних і управління - хмарна та локальна гнучкість - візуалізація даних у реальному часі версія 	для менших організацій варто враховувати вартість, оскільки функції та підтримка корпоративного рівня можуть мати відповідну ціну	5-денна безкоштовна, від 4000\$ в рік
Power BI	<ul style="list-style-type: none"> - платформа бізнес-аналітики, яка дозволяє створювати інтерактивні звіти, візуалізації та дашборди на основі різноманітних джерел даних. Power Query, DAX, Power Pivot-мови та інструменти для трансформації, моделювання та аналізу даних. - 	<ul style="list-style-type: none"> - Інтуїтивний інтерфейс. - Широкі можливості інтеграції підтримка Excel, SQL, Azure, Google Analytics, SharePoint, API тощо. - Потужні аналітичні функції DAX, AI-візуалізації, прогнозування, кластеризація. - Хмарна доступність. - Безпека та контроль 	Обмеження безкоштовної версії Залежність від екосистеми Microsoft	Power BI Desktop-безкоштовна Power BI Premium - від ~\$20
CouchDB	<ul style="list-style-type: none"> - одновузлова база даних Cluster Support - сумісність з http/json - надійна обробка даних - універсальна екосистема - вбудовані функції безпеки MapReduce для запитів даних 	<ul style="list-style-type: none"> - забезпечує надійні можливості реплікації та синхронізації - легка реплікація бази даних на кількох серверах - має зрозумілий http/rest API, який спрощує взаємодію з базою даних за допомогою 	обмежена підтримка складних запитів порівняно з деякими іншими базами даних NoSQL	Безкоштовна і має відкритий код

		стандартних методів http		
Apache Flink	<ul style="list-style-type: none"> - масштабована обробка даних - низька затримка - гнучкі API - потокова обробка стану - підтримка Advanced Analytics - уніфікований потік і пакетна обробка 	<ul style="list-style-type: none"> - стійкий до збоїв - хороша пропускна здатність і затримка - підтримка потокової обробки та вікон 	складність налаштування семантики часу події	безкоштовна має відкритий код за ліцензією Apache
Altair RapidMiner	<ul style="list-style-type: none"> - комплексні можливості машинного навчання - безпроблемна підготовка даних - надійні параметри візуалізації - масштабована інтеграція - автоматизований робочий процес - розширені інструменти аналітики 	<ul style="list-style-type: none"> - надійні можливості прогнозу аналітики BigData - графічний інтерфейс і параметри пакетної обробки 	велике споживання ресурсів під час великомасштабної обробки даних	30-денна безкоштовна пробна версія
Openrefine	<ul style="list-style-type: none"> - деталізація великих наборів даних за допомогою фасетів Clustering - безпека - перетворення даних - спеціальні скрипти 	<ul style="list-style-type: none"> - керування клітинками, що містять кілька значень - миттєві зв'язки між наборами даних - зручний інтерфейс 	немає функцій обробки даних у реальному часі, оскільки в основному працює з пакетними даними	безкоштовна і має відкритий код
Apache Hive	<ul style="list-style-type: none"> - SQL-подібна мова запитів - компіляція карти та редуکتора - визначення завдання в Java чи Python 	<ul style="list-style-type: none"> - горизонтальна масштабованість Apache Hive - розподілене сховище Hadoop 	більша затримка із запитами Hive	безкоштовна і має відкритий код

Узагальнено автором використовуючи платформу освітнього контенту Guru99 <https://www.guru99.com/uk/>

Як зазначалося раніше, важливий етап побудови інформаційної технології – розробити хорошу структуру для аналізу великих даних. Згідно з деякими дослідженнями [115], процес попередньої обробки даних вважається найважливішою фазою в усьому процесі аналізу даних і це може зайняти більше половини загального часу, витраченого на вирішення проблем інтелектуального аналізу даних. Інформаційна технологія освітньої аналітики заглиблюється в аналіз, інтерпритацію застосування освітніх даних, є трансформаційним підходом до покращення результатів навчання.

2.2. Порівняльний аналіз цифрових платформ управління навчанням

Орієнтуючись на дослідження, найбільш поширеними та функціонально адаптованими платформами електронного й дистанційного навчання, що використовуються у закладах вищої освіти, є такі: Moodle є однією з найпопулярніших відкритих систем управління навчанням (LMS) завдяки своїй модульній конструкції, інтерактивному інтерфейсу користувача, системі тестування, інтегрованій можливості відеоконференцій та глобальній базі користувачів. MS Learning Gateway – інтегрована система з інформацією, комунікацією та даними (ICT), побудована на Microsoft SharePoint, корпоративній та академічній екосистемі навчання. Вона надає єдиний доступ до навчального контенту для співпраці, а також інтегрує інструменти аналітики навчання. Як відкрита LMS, ATutor була розроблена з акцентом на інклюзивність та доступність, базується на SCORM та є простою системою управління контентом для адаптації навчальних закладів. Prometheus – це українська навчальна платформа, що пропонує безкоштовні онлайн-курси (переважно для підвищення кваліфікації державних службовців/освітян); вона має вбудовану сертифікацію, моніторинг та інші інструменти. IBM Lotus LMS – система навчання бізнес-класу з розширеними можливостями адміністрування, аналітики та створення контенту для великих організацій та університетів. WebCT, одна з перших платформ, що використовувалася для дистанційного

навчання, також пізніше була інтегрована з Blackboard для забезпечення багатофункціонального середовища, яке включало оцінювання, зворотний зв'язок, календарі та обговорення. SharePoint LMS базується на Microsoft SharePoint і сприяє створенню структурованого освітнього середовища для корпоративного та вищого навчання, для управління документами, співпраці, управління курсами тощо. Canvas LMS – це хмарна платформа, що забезпечує простий інтерфейс користувача, безшовну інтеграцію з Google Workspace, підтримку аналітики та мобільну доступність. Blackboard Learn – це корпоративна LMS, яка підтримує управління курсами на вимогу, забезпечує повну інтеграцію аналітики, а також інструменти зворотного зв'язку, розроблені для підвищення загального навчання з гнучкістю. Google Classroom пропонує простий спосіб налаштування всього навчального процесу, інтегрованого в сервіси Google. Schoology – це можливість організації соціального навчання, гнучкі інструменти тестування та оцінювання. Open edX – це відкрита система, що підходить для масових онлайн-програм курсів у всьому світі. Claroline – це відкрита платформа, яка дозволяє використовувати кілька мов та командам розробляти контент у навчанні. Chamilo – це зручна платформа, яка обслуговує країни, що розвиваються. ILIAS це універсальна LMS, сертифікована SCORM для використання в європейських університетах.

Кожна з цих платформ має свої переваги та обмеження, тому вибір конкретного рішення повинен ґрунтуватися на потребах освітнього закладу, доступних ресурсах, рівні цифрової компетентності користувачів та стратегічних цілях освітньої політики. Нубіп України вже майже 20 років використовує в забезпеченні освітнього процесу Moodle, тому проведемо порівняльний аналіз цієї платформи з існуючими (таблиця 2.2). Порівняльний аналіз різних LMS проведено з позицій їх аналітичного потенціалу, а саме: можливостей збору цифрових слідів, доступності лог-файлів, підтримки плагінів навчальної аналітики, відкритості API та інтеграції з зовнішніми інформаційними системами. Таким чином, розгляд альтернативних платформ

має методичний характер і спрямований на підтвердження того, що Moodle є технічно та функціонально придатним середовищем для реалізації запропонованої інформаційної технології освітньої аналітики.

Таблиця 2.2

Порівняльний аналіз платформ електронного й дистанційного навчання

Назва LMS	Тип / Сфера використання	Модель розгортання	Відкритий код (Open Source)	Підтримка стандартів (SCORM, xAPI)	Основні особливості та недоліки
Moodle	Академічна / Універсальна	Локальна / Хмара	Так	Так	Широко налаштовувана, великий ком'юніті, багато плагінів
ATutor	Академічна / Універсальна	Локальна	Так	Так	Простий інтерфейс, зручний для інклюзивної освіти
Prometheus	Академічна / MOOC	Хмара	Частково	Обмежена	Локальна платформа, фокус на український ринок
MS Learning Gateway	Корпоративна / Корпоративне навчання	Локальна / Хмара	Ні	Часткова	Інтеграція з продуктами Microsoft, зручна для Office-середовища
IBM Lotus LMS	Корпоративна	Локальна	Ні	Часткова	Старіше рішення, зосереджене на внутрішньому навчанні
WebCT	Академічна (історично)	Локальна / Хмара	Ні	Так	Об'єднана з Blackboard, застаріла
SharePoint LMS	Корпоративна / Корпоративне навчання	Локальна / Хмара	Ні	Обмежена	Інтеграція з Microsoft SharePoint, орієнтована на корпоративний сектор
Blackboard	Академічна / Комерційна	Локальна / Хмара	Ні	Так	Потужна підтримка, багатофункціональна, комерційна
Canvas	Академічна / Комерційна	Локальна / Хмара	Частково	Так	Сучасний UI, хмарне рішення, підтримка мобільних пристроїв
Google Classroom	Академічна / Спрощене	Хмара	Ні	Обмежена	Безкоштовна, інтеграція з Google-сервісами
Sakai	Академічна / Відкритий код	Локальна / Хмара	Так	Так	Модульна архітектура, підтримка співпраці
TalentLMS	Корпоративна / Комерційна	Локальна / Хмара	Ні	Так	Швидке розгортання, зручність для користувача

Можемо зробити висновок, що Moodle є однією з найпоширеніших та найбільш рекомендованих LMS у світі, особливо для закладів вищої освіти, завдяки низці

вагомих переваг, які обумовлюють її конкурентоспроможність та широке застосування.

По-перше, це система з відкритим кодом, яка дозволяє адаптувати Moodle до вимог освітніх установ без додаткових ліцензійних зборів. По-друге, вона не тільки дуже корисна, але й гнучка та масштабована, оскільки Moodle підтримує багато форматів освітнього контенту, інтерактивні формати (тести, форуми, чати, завдання), а також функції, які використовуються для персоналізації навчального процесу. По-третє, Moodle має стандарти електронного навчання, такі як SCORM та xAPI, які були встановлені, що дозволяє інтегрувати його в інші освітні ресурси та системи. Щодо формату експорту, Moodle дозволяє експорт даних у форматах CSV, Excel (XLS/XLSX) та JSON. Ці формати зручні для подальшої обробки за допомогою наших аналітичних інструментів, таких як Python, R, Excel Power Query та Power BI. Нарешті, Moodle забезпечує безпеку даних, підтримує локалізацію, зокрема українською мовою, тому місцеві громади знаходять платформу зручною та ефективною для впровадження у своїх вищих навчальних закладах

Для того, щоб проаналізувати платформу для організації освітнього процесу на рівні реалізації програмно-апаратного комплексу, доцільно звернутися до міжнародних стандартів якості програмного забезпечення, зокрема стандартів сімейства ISO 250xxx. Ці стандарти забезпечують системний підхід до оцінювання функціональних і нефункціональних характеристик інформаційних систем, таких як продуктивність, надійність, безпека, зручність використання та інші критично важливі параметри. Вони також враховують потреби безпосередніх користувачів – отримувачів освітніх послуг, що є ключовим фактором при виборі LMS. З огляду на це, платформа Moodle є оптимальним вибором, оскільки її архітектура та функціонал дозволяють реалізувати вимоги ISO 250xxx на високому рівні.

Стандарти сімейства **ISO/IEC 25000** (загально відомі як **SQuaRE** — Software Quality Requirements and Evaluation) – це міжнародні стандарти, що

призначені для допомоги розробникам, замовникам і користувачам у визначенні, вимірюванні та оцінюванні якості програмних продуктів. Підходять для будь-яких видів програмного забезпечення, включно з системами управління навчанням (LMS), які є складними інформаційними системами. Сімейство включає кілька стандартів, які розділені на групи, наприклад:

- **ISO/IEC 25010** — модель якості програмного забезпечення (визначає характеристики якості, такі як функціональність, продуктивність, безпека, зручність використання, сумісність тощо);
- **ISO/IEC 25012** — модель якості даних;
- **ISO/IEC 25020-25024** — методи вимірювання якості;
- **ISO/IEC 25040** — процес оцінки якості;

Moodle відповідає стандартам щодо гнучкості, масштабованості, безпеки даних і зручності користування, а також підтримує відкритість коду, що дає змогу адаптувати систему під індивідуальні потреби закладу. Таким чином, використання Moodle з урахуванням рекомендацій стандартів ISO 250xxx забезпечує створення надійного, ефективного та адаптивного середовища електронного навчання, яке відповідає сучасним викликам освітньої трансформації.

2.3. Аналітична та прогностична обробка освітніх даних з використанням LMS Moodle

Цифрове освітнє середовище в закладах вищої освіти поступово розвивається в повноцінний простір, що сприяє формуванню професійної ідентичності студентів та розвитку ефективних моделей поведінки. [135, 42, 114, 115]. Таке середовище зазвичай охоплює як структуровані (заздалегідь підготовлені), так і неструктуровані (спонтанно виникаючі) джерела знань, включаючи електронні навчальні матеріали, авторські курси викладачів та комунікацію на форумах і в особистих повідомленнях.

Виникнення великих даних у сфері освіти обумовлено щонайменше двома основними тенденціями цифрової епохи: по-перше, дедалі більше інституційної інформації оцифровується, що призводить до значного накопичення стандартизованих даних про студентів; по-друге, аспекти поведінки у навчанні, які раніше важко було фіксувати під час традиційного навчання в аудиторії, тепер активно реєструються через LMS. Це відкриває нові можливості для аналітики освітнього процесу.

На цій основі запропоновано підхід для використання цих даних в аналітиці навчання шляхом перетворення та інтеграції даних Moodle LMS через Microsoft Power BI у динамічні та інтерактивні візуалізовані звіти для прийняття рішень усіма учасниками освітнього процесу. Завдяки доступу до великого масиву навчальних даних від простого збору результатів до прогнозування поведінки студентів на електронних курсах постає питання вибору інструментів для застосування різноманітних методів аналітики цих даних [115]. LMS Moodle стала досить популярною і у світі серед викладачів як інструмент для створення динамічних курсів для студентів. За статистикою станом на січень 2024 р. в світі є 162,045 активних сайтів із системою Moodle, що містять 46,572,992 курси та зареєстровані у 242 країнах. Платформу використовують більше 401 млн. користувачів [72]. Найбільше користувачів в Іспанії, Сполучених Штатах Америки та Німеччині. В Україні активно функціонують сайти (1177) на базі LMS Moodle. Виокремимо основні напрями, за якими вже досить тривало виконуються наукові дослідження: розробка електронних освітніх середовищ [42]; виявлення поведінкових стратегій в онлайн-навчанні, їх зв'язок із характеристиками студентів [5]; візуалізація даних, видобутих та опрацьованих з Moodle [39, 85]; розробка та використання інструментів для аналізу даних Moodle [73]; встановлення зв'язків між активностями студентів в онлайн середовищі та їхніми успіхами в навчанні [34, 82], аналіз цифрових слідів, пов'язаних з поведінкою користувачів університетських інформаційно-освітніх систем [47,59,78]. Запровадження систем управління навчанням та розробка

інформаційних технологій формування ІТ-компетентностей у закладах освіти зробило доступними великі обсяги даних про поведінку студентів онлайн для оптимізації цифрового освітнього середовища [130, 31, 59]. Можна проаналізувати, які матеріали на курсах мають найбільшу популярність, як вчасно студенти завантажують свої роботи, який коефіцієнт залучення студентів на електронному курсі, або ж які студенти потенційно не зможуть вчасно закінчити навчання на курсі. Цей розвиток супроводжується генерацією великих обсягів інформації високої освітньої цінності й потребує автоматизованого аналізу даних із можливістю оперативного використання результатів. Ефективна візуалізація покращує сприйняття і обробку даних і полегшує прийняття обґрунтованих рішень через трансформацію великих обсягів складної інформації у зрозумілі графічні форми. Крім того це допомагає краще усвідомити приховані зв'язки і структури в інформаційному середовищі.

Стандартні процедури формування аналізів, характерні для баз даних, займають, як правило, багато часу. Щоб полегшити цей процес, використовуються інструменти обробки великих даних та виникає потреба у залученні educational data mining чи засобів інтелектуального аналізу Moodle Big Data, послідовність проведення якого запропоновано на рис.2.5.



Рис.2.5. Схема реалізації освітньої аналітики на основі Moodle Big Data

Щоб аналізувати поведінку учасників освітнього процесу, багато закладів використовують алгоритми машинного навчання з даними LMS для певного курсу або типу курсу, але лише деякі намагаються робити прогнози для великої підмножини курсів. Відповідні докази ефективності аналітики навчання для підтримки успішності студентів надаються в різних дослідженнях [49].

Необхідно відзначити два способи проведення такої навчальної аналітики: на основі наявних оперативних даних в базі системі Moodle на основі

трансформованих та підготовлених даних на іншому сервері.

Починаючи з версії Moodle 4.0 для отримання такої інформації створено спеціальний модуль “конструктор звітів” (рис. 2.6). Цей модуль дозволяє будувати власні звіти з наданої для цього поточної інформації. З метою не перевантажувати систему під час перегляду звітів, кількість показників та звітів обмежується.

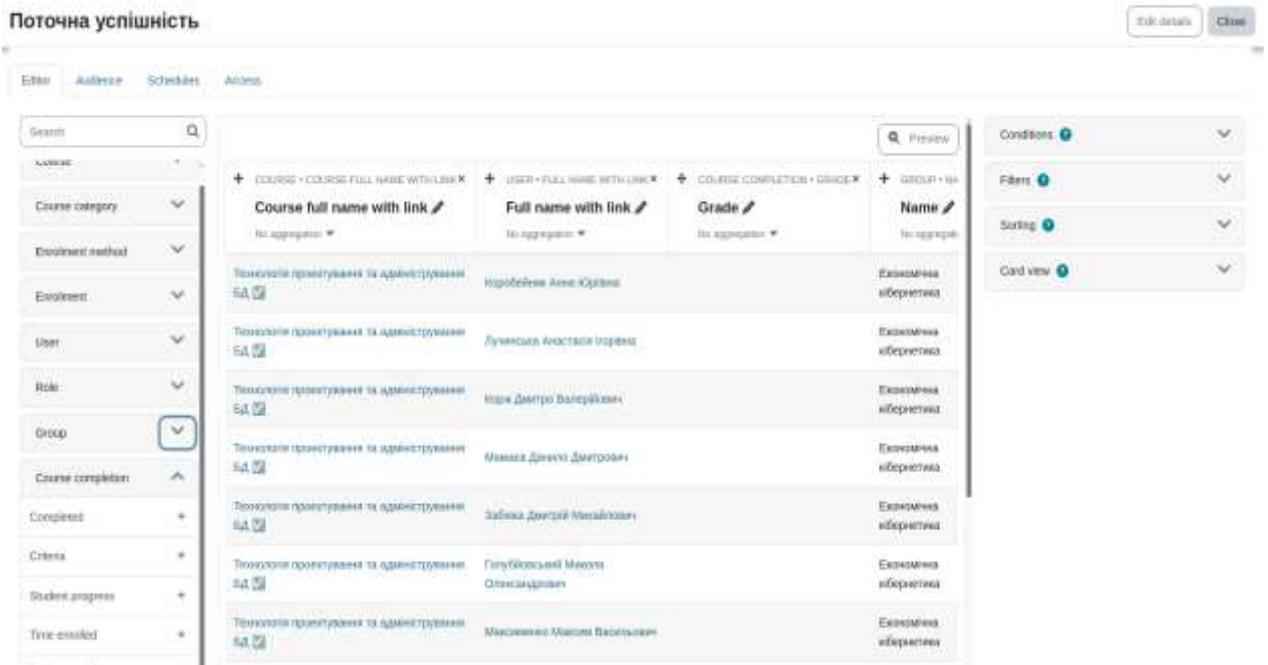


Рис. 2.6. Конструктор звітів Moodle

Наразі критичне обмеження вбудованого конструктора звітів полягає у обмеженій кількості показників, з яких у менеджера є можливість будувати аналітику. І хоча в кожній новій версії Moodle число показників збільшують, все одно воно залишається обмеженим. Проте, є вихід, якщо нам потрібно отримати певні специфічні показники для нестандартних звітів – встановити додатковий модуль Configuration Report. За допомогою цього модуля можна писати прямі запити до бази даних, з використанням мови SQL, та отримувати будь-яку інформацію, стосовно якої в базі даних залишається цифровий слід. Складність таких звітів обмежується лише вмінням та фантазією менеджера, який їх створює. Приклад такого звіту показано на рис. 2.7.

Наявні можливості для аналізу даних, що накопичуються в LMS Moodle, можна розділити на чотири категорії

- стандартні засоби системи;
- додаткові плагіни для системи;
- зовнішні засоби для аналізу даних безпосередньо з Moodle;
- різні багатофункціональні програми аналізу даних.

Стандартні засоби LMS Moodle – низка стандартних звітів, які доступні в розділі "Звіти" і кілька аналітичних моделей, які належать до базового функціоналу системи. Інформація звіту "Журнал подій" для нетехнічних користувачів виявляється мало зрозумілою і погано придатною для будь-яких висновків. Але саме ці дані часто використовують для подальшої аналітики сторонніми засобами аналізу даних. На основі журналу подій можна побудувати зведені звіти. Ці звіти формують загальне уявлення про інтенсивність використання студентами різних елементів курсу.

Додаткові плагіни для системи. Плагіни є розширенням функціональних можливостей Moodle, яких немає в базовій версії. На даний момент офіційний каталог Moodle Plugins Directory містить понад 2000 додатків [71]. Прикладами плагінів, які призначені для аналізу даних є SmartKlass, Heatmap, IntelliBoard та інші. Суттєвими недоліками цієї групи засобів є необхідність їх встановлення адміністратором, та оновлення плагінів при оновленні системи, що може викликати деякі труднощі. При роботі з аналітикою безпосередньо на "живих" даних маємо два великих мінуси: по-перше, ми отримуємо інформацію лише на конкретний час і не можемо проаналізувати стан в динаміці, по-друге, це суттєво навантажує систему, зокрема сервер бази даних (особливо на високонавантажених серверах з великою кількістю студентів та навчальних курсів).

Щоб обійти перший з мінусів розробники Moodle вже давно додали найпростіший модуль статистики, який діє за базовими принципами сховища даних: провести первинний обробіток, додати часові мітки зберегти в окремому

місці (в даному випадку, це окремі таблиці в базі Moodle). Система Moodle періодично запускає збір статистики і результати записує у спеціальні таблиці. З метою зменшення навантаження при читанні даних розробники розділили таблиці за базовими часовими характеристиками:

stats_daily: для накопичення щоденної статистики;

stats_monthly: для накопичення щомісячної статистики;

stats_weekly: для накопичення тижневої статистики;

stats_user_daily: для накопичення щоденної статистики за курс/користувача;

stats_user_weekly: для накопичення щотижневої статистики за курс/користувача;

stats_user_monthly: для накопичення щомісячної статистики за курс/користувача. Обраховані результати такого обробітку вкладаються в окремі таблиці реляційної бази даних: денні, тижневі, місячні, річні. На рис. 2.9 представлено частину схеми бази даних для фіксації аналітики використання ЕНК у місячному інтервалі. Зі схеми видно, що це простий аналіз фіксує курси та ролі користувачів на курсі. Для різних типів аналізу (найактивніші курси; зважений показник найактивніших курсів; курси з найбільшою кількістю учасників; показник участі на курсі) зберігаються два розрахованих показника, в таблиці колонки *stat1* та *stat2*.

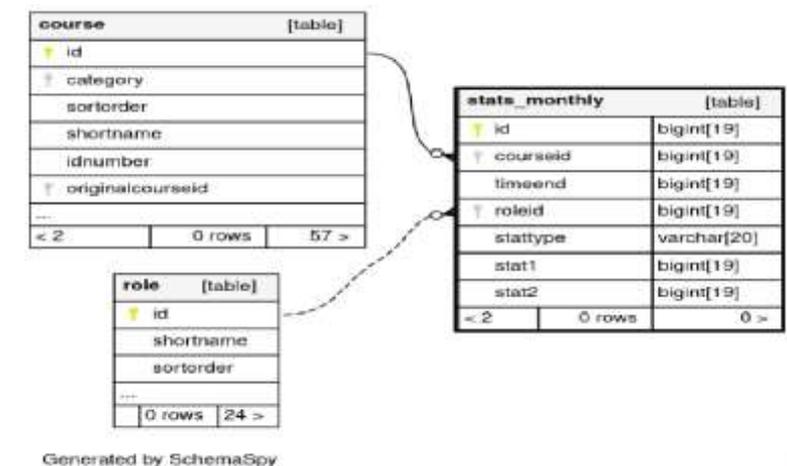


Рис. 2.9. Модель даних порівняння електронних курсів

Але для реальної роботи з аналізом навчальної інформації необхідно створювати окреме сховище даних, налаштовувати систему перетворення

первинних даних (з “живої” бази Moodle) до сховища, підключати до сховища даних систему бізнес аналітики для побудови різноманітних звітів, графіків та панелей даних.

Зовнішні засоби для аналізу даних Moodle. До цієї групи належать автономні рішення для аналізу даних Moodle. Багато наявних зовнішніх рішень для аналізу даних Moodle засновані на додатковій аналітиці даних журналу подій. Такі рішення маю інтуїтивно зрозумілий англomовний інтерфейс, не містять у собі складних, спеціалізованих методів інтелектуального аналізу даних і можуть використовуватися широкою академічною аудиторією. Дані рішення пропонують додатковий аналіз і візуалізацію даних журналу подій.

Різні багатофункціональні програми аналізу даних. Варто виділити декілька груп таких інструментів, які можуть бути використані для аналізу даних, отриманих з Moodle:

1. Статистичні та математичні програми (Statistica, SPSS Statistics тощо; Matlab, Octave тощо; Excel тощо);
2. Програми для інтелектуального аналізу даних та бізнес аналітики та візуалізації (Orange, Knime, RapidMiner, Weka, Power BI, Deductor та ін.);
3. Платформи та мови програмування (R, Python (як найбільш використовувані для аналізу даних), Java та ін.);
4. СУБД і мови запитів (Oracle, MySQL, Microsoft SQL Server та ін.). Цю групу засобів виділено у зв'язку з тим, що LSM Moodle використовує для зберігання інформації реляційну базу даних;
5. Інструменти аналізу веб-сайтів (використовують лог-аналізатори: Webalizer, AWStats та ін.). Ці засоби можуть бути використані, оскільки Moodle також являє собою освітній веб-сайт. Застосування цих засобів, як правило, вимагає від користувача певних навиків використання статистичних методів, алгоритмів інтелектуального аналізу даних, мов програмування, процесів попереднього оброблення даних, практики роботи в різних програмах, навичок програмування тощо.

Порівняльний аналіз груп засобів аналітики Moodle Big Data наведений в таб.2.11.

Таблиця 2.11

Порівняльний аналіз засобів Moodle Big Data Analytics

Інструменти аналітики великих даних Moodle	Переваги	Недоліки
стандартні (внутрішні) інструменти системи (облікові записи)	<ul style="list-style-type: none"> – доступність викладачам і адміністраторам – наявні в Moodle за замовчуванням 	<ul style="list-style-type: none"> – недостатня функціональність і наочність для прийняття рішень
додаткові плагіни для системи	<ul style="list-style-type: none"> – реалізують додаткові функції 	<ul style="list-style-type: none"> – вузькоспрямовані – необхідне встановлення адміністратором системи
зовнішні інструменти для аналізу даних безпосередньо з Moodle	<ul style="list-style-type: none"> – орієнтація на навчальну аналітику та роботу саме з даними системи – простота використання 	<ul style="list-style-type: none"> – мало у вільному доступі – англомовний інтерфейс
багатофункціональне програмне забезпечення для аналізу даних	<ul style="list-style-type: none"> – широкий функціонал – можливості для якісного та багатоаспектного аналізу 	<ul style="list-style-type: none"> – вимагають фахових цифрових і математичних компетентностей

У даному дослідженні інформаційна технологія освітньої аналітики використовує дані LSM Moodle для реалізації моделей прогнозування успішності здобувачів освіти з аналізом та візуалізацією через Power BI.

2.4. Методичні, алгоритмічні та модельні засади інформаційної технології освітньої аналітики

Технологія добування даних базується на шаблонах, які представляють приховані закономірності у великому обсязі інформації. Виявлення таких закономірностей, які зазвичай не видно неозброєним оком, призводить до вирішення багатьох загальних завдань, пов'язаних з глибоким аналізом даних. Саме тому завдання добування даних часто прирівнюють до ідентифікації шаблонів або методів їх ідентифікації.

У науковій спільноті немає єдиного підходу до класифікації завдань, які слід віднести до сфери добування даних [1, 14, 18, 19, 26, 45]. Однак багато з цих дослідників добування даних вказують на такі основні теми: класифікація, кластеризація, прогнозування, виявлення асоціативних правил, візуалізація даних, виявлення аномалій, оцінка, аналіз зв'язків та узагальнення даних. Кожна з цих галузей має свої методи, алгоритми та сфери застосування, тому інструменти добування даних можуть бути адаптовані до різних завдань у різних галузях (освіта, медицина, фінанси, маркетинг тощо) [24, 30, 57, 78, 86].

Зосереджуючись на характеристиках освітніх даних та їх джерелах окреслимо основні класи методів добування даних, які використовуються освітньою аналітикою, що застосовуються як аналітичний шар запропонованої технології:

1. Методи класифікації– дозволяють віднести студента або подію до певної категорії на основі набору вхідних характеристик. Використані алгоритми:

- дерева рішень (Decision Trees),
- логістична регресія,
- ансамблеві моделі (Random Forest, Gradient Boosting).

У практиці аналітики освіти класифікація дозволяє передбачити ймовірність відрахування студента, успішність за підсумками курсу чи потребу в додатковій підтримці.

2. Методи кластеризації – використовуються для виокремлення груп студентів із подібними навчальними траєкторіями або стилями навчання без попереднього маркування даних. Використані алгоритми:

- k-середніх (K-Means),
- ієрархічна кластеризація (Hierarchical Clustering),

Кластеризація допоможе у формуванні адаптивних навчальних маршрутів, групуванні освітніх програм або визначенні "ризикових" категорій студентів.

3. Аналіз асоціацій – виявлення часто повторюваних патернів або правил співіснування ознак.

4. Регресійний аналіз – дасть можливість довести чи спростувати залежності між освітніми предикторами (наприклад, відвідуваність, активність у LMS, попередня успішність, активність) та результатами навчання.
5. Методи часових рядів і прогнозування –дозволять проаналізувати поведінку студентів у динаміці, передбачати зміни рівня успішності, потребу в інтервенціях тощо.
6. Аналіз соціальних мереж (SNA) –можуть бути використані для дослідження взаємодії учасників освітнього процесу в цифровому середовищі (наприклад, форумах, спільних проєктах), виявлення лідерів думок, слабких зв'язків і структур комунікацій.
7. Нейронні мережі та глибоке навчання –методи глибокого навчання (deep learning), включаючи (CNN) і (RNN) нейронні мережі, можуть бути використані для автоматичної обробки складних даних: текстів, зображень, відео.
8. Семантичний аналіз і обробка природної мови (NLP) –застосовується для аналізу письмових відгуків, обговорень у чатах, відкритих відповідей у тестах, що дозволяє оцінити рівень залучення, емоційний стан і змістовну якість комунікацій.

Застосування цих методів залежить від типу та обсягу даних, рівня деталізації, наявної інфраструктури та цілей аналітики. Зведена інформація стосовно призначення описаних методів наведена в таблиці 2.3.

Таблиця 2.3

Порівняння основних класів методів добування даних, які використовуються освітньою аналітикою

Група методів	Приклади алгоритмів	Призначення в освітній аналітиці
Класифікація	Decision Trees, Naive Bayes, SVM, Logistic Regression, Random Forest	Передбачення результатів навчання, ризику відрахування, оцінювання потреб у підтримці
Кластеризація	K-Means, DBSCAN, Ієрархічна кластеризація	Виявлення груп студентів за стилем навчання, поведінкою, ризиками
Аналіз асоціацій	Apriori, FP-Growth	Виявлення закономірностей у діях студентів (наприклад, які ресурси переглядають перед успішним тестом)
Регресійний аналіз	Лінійна, множинна,	Моделювання залежностей між

	поліноміальна регресія	змінними (активність – бали, мотивація – успішність тощо)
Прогнозування / часові ряди	ARIMA, Prophet, LSTM	Аналіз динаміки навчальної активності, прогноз змін успішності або потреб в інтервенції
Аналіз соціальних мереж (SNA)	Центральність, щільність, когезія мереж	Аналіз комунікацій у форумах, проектних групах, виявлення ізольованих або ключових учасників
Нейронні мережі та глибоке навчання	CNN, RNN, LSTM, Autoencoders	Обробка складних даних (зображення, відео, послідовності дій), розпізнавання емоцій, поведінкових патернів
Обробка природної мови (NLP)	BERT, TF-IDF, Word2Vec, Sentiment Analysis	Аналіз письмових відповідей, чатів, емоцій, тем обговорення

Комплексне застосування описаних методів у рамках інформаційної технології освітньої аналітики забезпечить розв’язання поставлених задач, зокрема: дасть можливість раннього виявлення ризиків (низька успішність, відрахування); забезпечить підтримку управлінських рішень через глибоку аналітику поведінки студентів; забезпечить оптимізацію змісту курсів через аналіз ефективності навчального контенту.

Ці методи лежать в основі побудови інформаційної технології освітньої аналітики, де використання різних алгоритмів забезпечує адаптивність, гнучкість та можливість оцінювання як індивідуальних, так і колективних освітніх траєкторій. Характеристика основних методів для освітньої аналітики наведено в таблиці 2.4.

Таблиця 2.4

Характеристика основних моделей та методів освітньої аналітики

Метод	Формалізація математична модель / формула	Логіка роботи	Приклад застосування в освіті
Класифікація (Decision Tree)	Рекурсивне розбиття простору ознак: $Gini(D) = 1 - \sum_{i=1}^k p_i^2$ або $Entropy(D) = - \sum_{i=1}^k p_i \log_2 p_i$	Дані розділяються за ознаками на підмножини, які мінімізують ентропію або індекс Джіні. Для	Визначення груп ризику серед студентів (відрахування, низька успішність).

		передбачення категорій (наприклад, ризик академічної неспішності).	
Кластеризація (K-Means)	$\min_C \sum_{i=1}^k \sum_{x_j \in C_i} \ x_j - \mu_i\ ^2$	Алгоритм ітеративно обчислює центри кластерів та переоб'єднує точки до найближчого кластера. Для виявлення груп студентів зі схожою поведінкою.	Групування студентів за стилями навчання, рівнем участі, цифровою активністю.
Асоціативні правила (Apriori)	$A \Rightarrow B;$ $support = P(A \cup B), confidence = \frac{P(A \cup B)}{P(A)}$	Генерує часті набори елементів та виводить залежності між подіями (наприклад, студенти, які проходили тему А, часто успішно складають тему В).	Встановлення залежностей між темами курсу та успішністю їх засвоєння.
Регресійний аналіз	Лінійна модель: $y = \beta_0 + \beta_1 x_1 + \dots + \beta_n x_n + \epsilon$	Прогнозування змінної на основі інших факторів. Встановлює залежність цільової змінної (наприклад, підсумковий бал) від предикторів (кількість входів, КБ при вступі тощо).	Прогноз успішності за показниками активності, початковими знаннями.
ARIMA (часові ряди)	$y_t = c + \sum_{i=1}^p \phi_i y_{t-i} + \sum_{j=1}^q \theta_j \epsilon_{t-j} + \epsilon_t$	Проводиться аналіз і прогнозування серій освітньої	Аналіз змін активності студентів протягом

		активності або змін в оцінках на основі минулих значень.	семестру.
Аналіз соціальних мереж (SNA)	Міра центральності (напр. degree): $C_D(v) = deg(v)$ або міжцентральність: $C_B(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$	Оцінюється взаємодія студентів у навчальному середовищі.	Виявлення ізольованих студентів або лідерів у командній роботі.
Глибоке навчання (DNN)	Звичайна модель: $y = f(Wx + b)$, де f — активаційна функція (ReLU, sigmoid)	Нейронна мережа формує складну нелінійну функцію, що апроксимує залежності у даних.	Індивідуальне прогнозування освітньої траєкторії студента.
Обробка природної мови (NLP)	TF-IDF: $tfidf_{t,d} = tf_{t,d} \cdot \log \frac{N}{df_t}$ Word2Vec: векторне представлення слів	Аналіз текстових даних – відповідей студентів, повідомлень, форумів – для виявлення смислів, тональностей, тем.	Автоматичний аналіз відповідей студентів, виявлення емоцій у відгуках.
Метод опорних векторів (SVM)	$f(x) = \text{sign}(w^T x + b)$	Побудова гіперплощини – розділяє класи з максимальним зазором.	Класифікація студентів за рівнем успішності або за ризиком академічної неуспішності.
Метод головних компонент (PCA)	$Z = XW$, де W — матриця власних векторів	Зменшення розмірності даних при збереженні найбільшої варіації.	Побудова візуалізацій освітніх характеристик або кластерів студентів.

Висновки до розділу 2.

У межах другого розділу нами було проведено комплексне обґрунтування інформаційної технології освітньої аналітики та визначено її ключові структурні, методичні й алгоритмічні засади, визначено вимоги до інформаційних систем та характеристик даних, що використовуються для аналітичної підтримки управлінських рішень.

Також ми здійснили систематизацію джерел освітніх даних і сформували компоненти інформаційної технології освітньої аналітики, яка охоплює підсистеми збору, обробки, інтеграції, зберігання та аналізу даних. Це дало можливість запропонувати методи інтеграції та підготовки освітніх даних із різних джерел – Єдиної державної електронної бази з питань освіти (ЄДЕБО), системи управління навчальним процесом (LMS moodle) та ІС «Деканат» для формування повноцінного багатовимірного профілю здобувача освіти

На основі порівняння найбільш поширених систем управління навчанням (LMS), визначено їх переваги й обмеження з точки зору подальшого застосування в аналітичних задачах. Це дозволило сформуванати вимоги до архітектури інтегрованої інформаційної платформи, орієнтованої на підтримку освітнього процесу закладу вищої освіти. Також запропоновано технологічний підхід до стандартизації, нормалізації та очищення даних, а також алгоритми формування аналітичних вибірок, що забезпечують коректність і повноту їх подальшого опрацювання методами інтелектуального аналізу.

Особливу увагу у розділі приділено вибору моделей Data Mining для задач освітньої аналітики. На основі порівняльного аналізу сучасних методів обґрунтовано доцільність застосування моделей (Дерева рішень, Випадковий ліс, LightGBM (легка градієнтна модель підсилення і Логістична регресія) для прогнозування успішності здобувачів освіти, моделей кластеризації для групування студентів за поведінковими патернами та регресійних моделей для виявлення впливових показників на успішність.

Таким чином, у розділі сформовано цілісне методичне та алгоритмічне підґрунтя інформаційної технології освітньої аналітики, що забезпечує можливість її подальшої реалізації, програмної розробки та застосування в задачах підтримки прийняття рішень у закладах вищої освіти.

РОЗДІЛ 3

РОЗРОБКА ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ ОСВІТНЬОЇ АНАЛІТИКИ

3.1. Архітектура інформаційної технології освітньої аналітики

Для системного підходу до опису складних інформаційних процесів, що включають взаємодію різномірних даних, аналітичних процедур та управлінських рішень використаємо побудову функціональної моделі інформаційної технології освітньої аналітики з використанням методології SADT (Structured Analysis and Design Technique). У контексті дослідження це дозволило формалізувати логіку перетворення освітніх даних у знання та підтримку прийняття рішень у закладах вищої освіти. Застосування SADT-діаграм дає можливість структурувати процес функціонування інформаційної технології у вигляді ієрархічної системи взаємопов'язаних функцій. Представимо на рис.3.1. SADT- модель для опису функціональної структури нашої технології у вигляді ієрархії функцій та потоків даних, що демонструє, які входи система обробляє, які виходи вона продукує, які механізми використовує та які керуючі впливи визначають її поведінку.

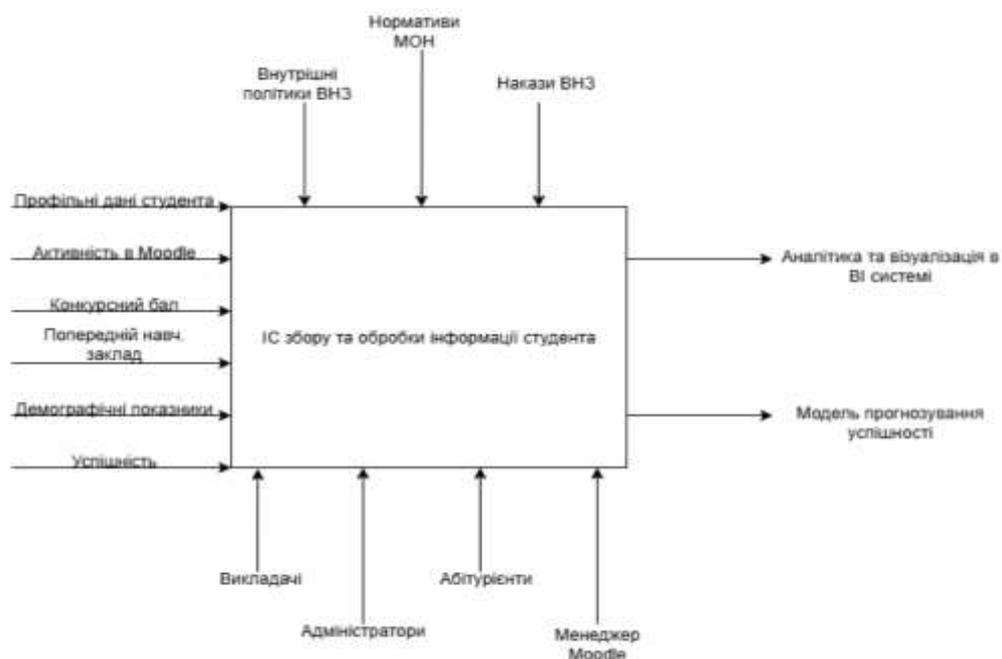


Рис. 3.1. Функціональна модель модуля збору та обробки інформації студента за методологією SADT (IDEF0)

Саме вона становлює рамки та призначення інформаційної технології освітньої аналітики.

Перший рівень декомпозиції А-0 і побудова А0 означає, що основна функція розбивається на кілька підпроцесів, які описують внутрішню структуру системи. Для інформаційної технології освітньої аналітики А0 включає:

1. Збір та валідацію освітніх даних.
2. Структуризацію даних.
3. Аналітичне моделювання та прогнозування.
4. Інтерпретацію та візуалізація результатів.

Деталізуємо А-0 по відповідних блоках роботи з даними та отримаємо перший рівень декомпозиції –діаграму А0, що деталізує внутрішні процеси та структуру інформаційної технології. (рис.3.2)

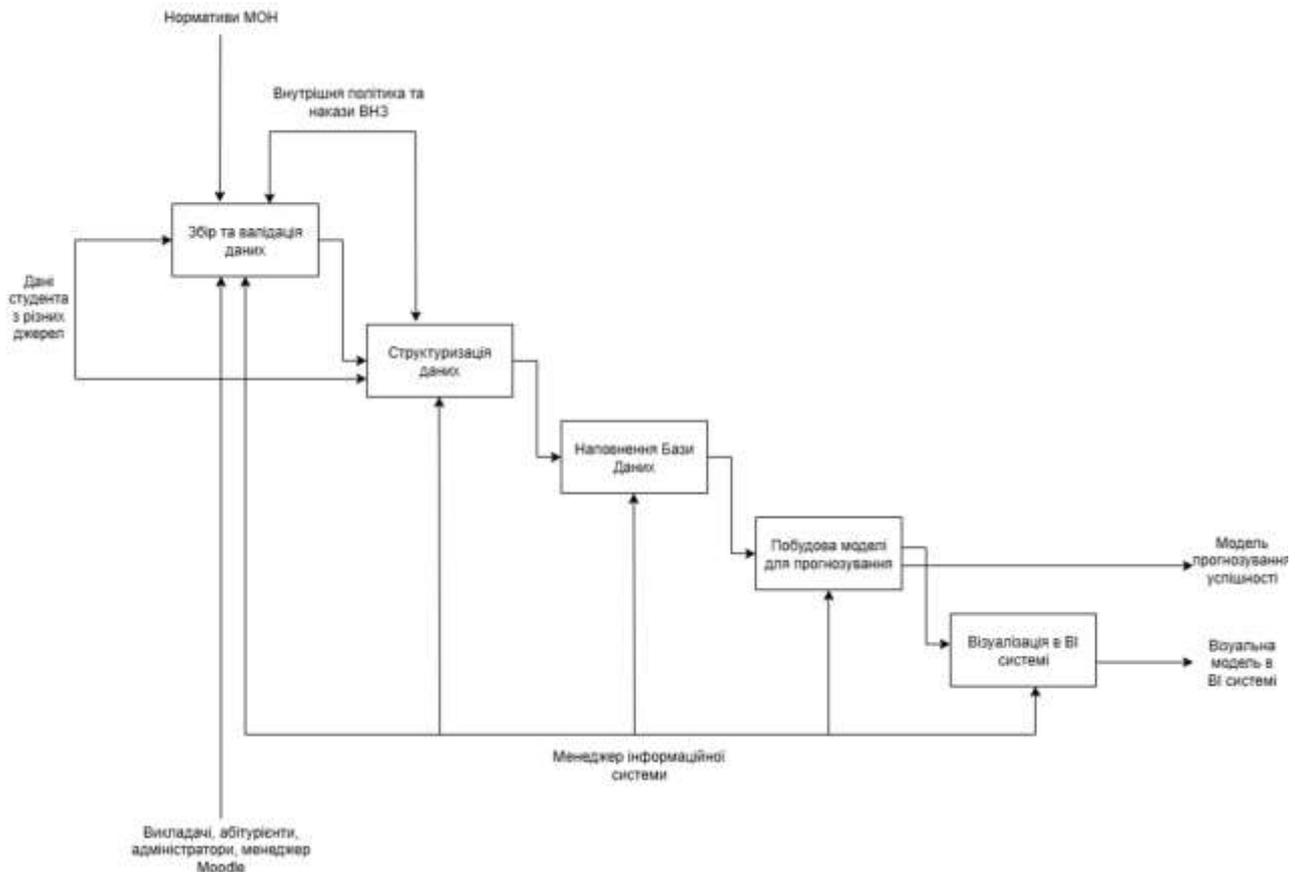


Рис. 3.2. Діаграма А0 рівня модуля «збору та обробки інформації студента»

Проведемо також декомпозицію, що розкладає головну функцію (із А-0) на підфункції. Вона допомагає зрозуміти, з яких складових етапів складається

базова функція «модель прогнозування успішності» і як потоки даних проходять між ними (рис.3.3).

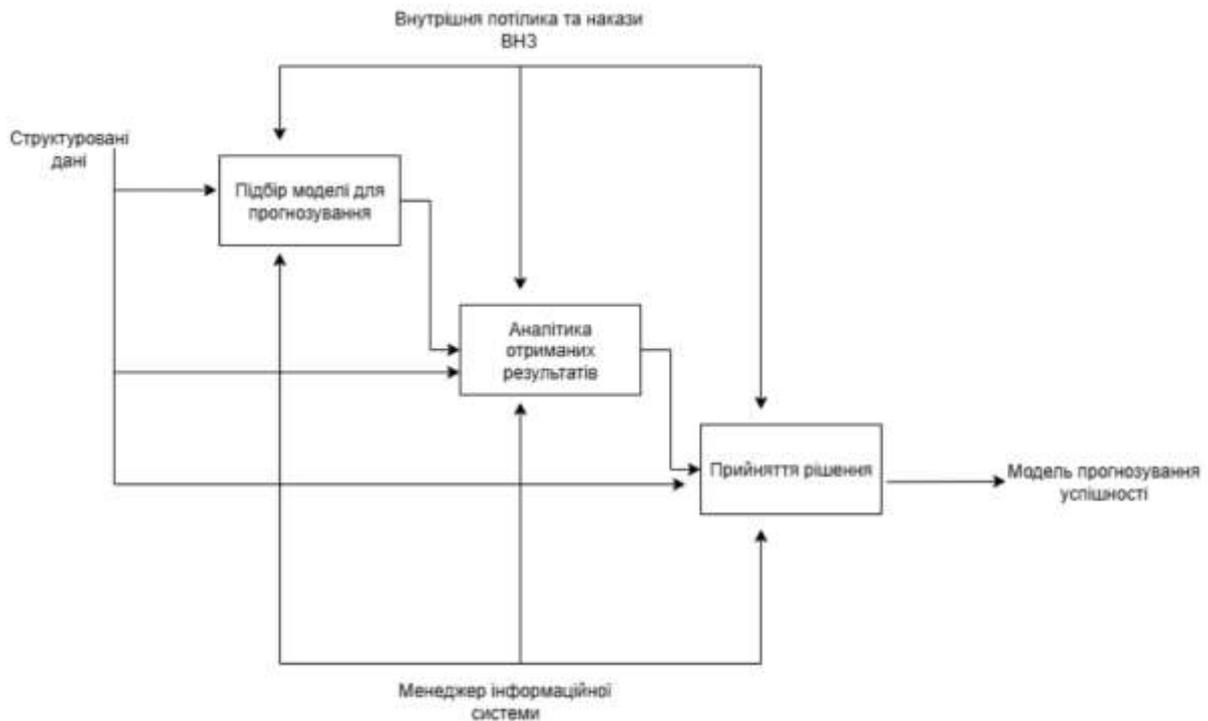


Рис.3.3. Деталізація базової функції «модель прогнозування успішності»

Крім того, SADT-моделі забезпечують формалізований опис бізнес-процесів і функціональних потоків даних, що є необхідним для подальшої розробки архітектури системи, UML-моделей, BPMN-діаграм і програмної реалізації.

Архітектура інформаційної технології повинна відповідати принципам відкритості, масштабованості, гнучкості, безпеки та сумісності з чинними державними стандартами, зокрема ISO/IEC 25010 (вимоги до якості програмного забезпечення) і ISO/IEC 23988 (достовірність комп'ютеризованого тестування).

Розробка інформаційної технології освітньої аналітики потребувала створення цілісної архітектури, яка б забезпечувала ефективне збирання, обробку, інтеграцію, аналіз та візуалізацію освітніх даних з різних джерел. Запропонована архітектура ґрунтується на модульному принципі, що дозволяє забезпечити гнучкість, масштабованість, адаптивність до різних технічних і

організаційних умов функціонування закладів вищої освіти [117,123,124,78]. Вона також орієнтована на підтримку інтеграції з наявними інформаційними системами (зокрема, ЄДЕБО, ІСУ «Деканат», LMS Moodle тощо) та реалізацію методів інтелектуального аналізу даних.

У процесі розробки інформаційної технології освітньої аналітики важливим етапом є формалізований опис її архітектури, що забезпечує цілісне уявлення про структуру системи, взаємодію компонентів і логіку обробки освітніх даних. З цією метою в роботі використано нотацію UML (*Unified Modeling Language*) – рис. 3.4. Для повного представлення архітектури інформаційної технології освітньої аналітики використано декілька типів UML-діаграм, кожна з яких виконує окрему функцію: Use Case Diagram застосована для формалізації взаємодії між системою та користувачами (аналітик, викладач, адміністрація, студент). Component Diagram використана для відображення логічної структури технології. Sequence Diagram застосована для моделювання сценаріїв прогнозування успішності. Deployment Diagram використана для відображення фізичного розміщення компонентів (сховище даних, ML-сервер, ВІ-платформа).

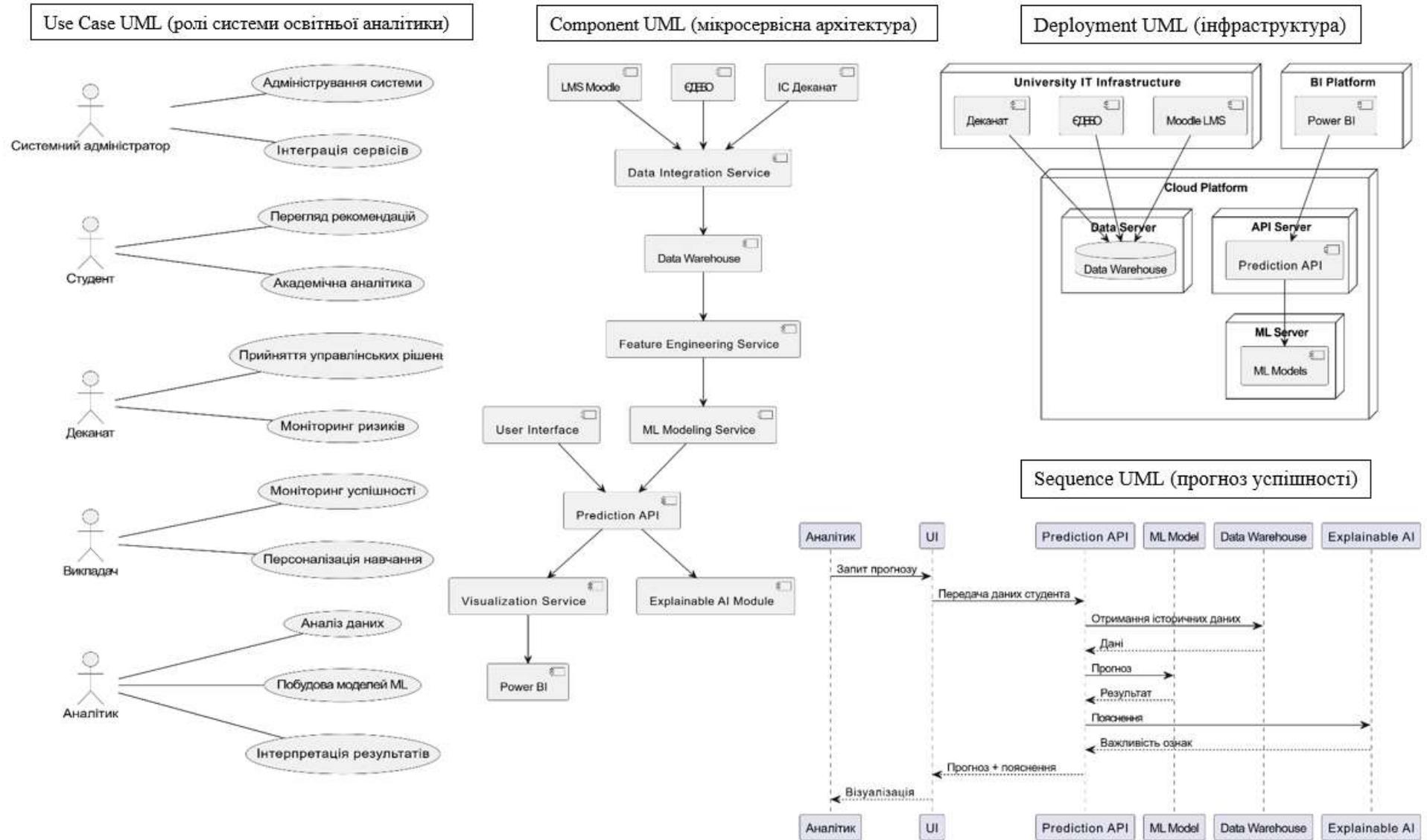


Рис. 3.4. Інформаційна технологія освітньої аналітики в нотатції UML моделей

Застосування різних типів UML-моделей забезпечило структуроване, системне та відтворюване представлення компонентів, інтерфейсів і сценаріїв функціонування розробленої технології.

Розробка інформаційної технології освітньої аналітики передбачає формалізацію бізнес-процесів збору, обробки та використання освітніх даних у діяльності закладу вищої освіти. У цьому контексті доцільним є застосування нотації BPMN (*Business Process Model and Notation*) для моделювання процесів функціонування системи. Розробка BPMN діграми для моделювання бізнес-процесу освітньої аналітики наведена на рис. 3.5. На відміну від UML, що переважно орієнтована на програмну архітектуру, BPMN дозволяє моделювати управлінські та аналітичні процеси, показати взаємодію між людьми та інформаційними системами, продемонструвати прикладний характер технології.

Це є важливим, оскільки інформаційна технологія освітньої аналітики використовується не лише в ІТ-середовищі, а й у системі управління якістю освіти.

Ця модель відображає основні етапи роботи системи: збір даних, їх обробка та подальше використання у бізнес-аналітичній платформі (ВІ-системі). Діаграма розділена на три логічні блоки, які репрезентують відповідні функціональні напрями.

На першому етапі «збір даних моделі» здійснюється аналіз наявної інформації з трьох основних джерел: системи управління навчанням (Moodle), Єдиної державної електронної бази з питань освіти (ЄДЕБО), а також інформаційної системи управління навчальним процесом (ІС «Деканат»). Додатково система може підключатися до відкритих державних реєстрів для збагачення даних про здобувачів освіти чи викладачів. Усі ці дані можуть бути представлені в різних форматах (SQL, XML, CSV, API-відповіді), тому тут важливо було забезпечити підтримку гнучких засобів витягування, перетворення і завантаження (ETL – Extract – Transform – Load).

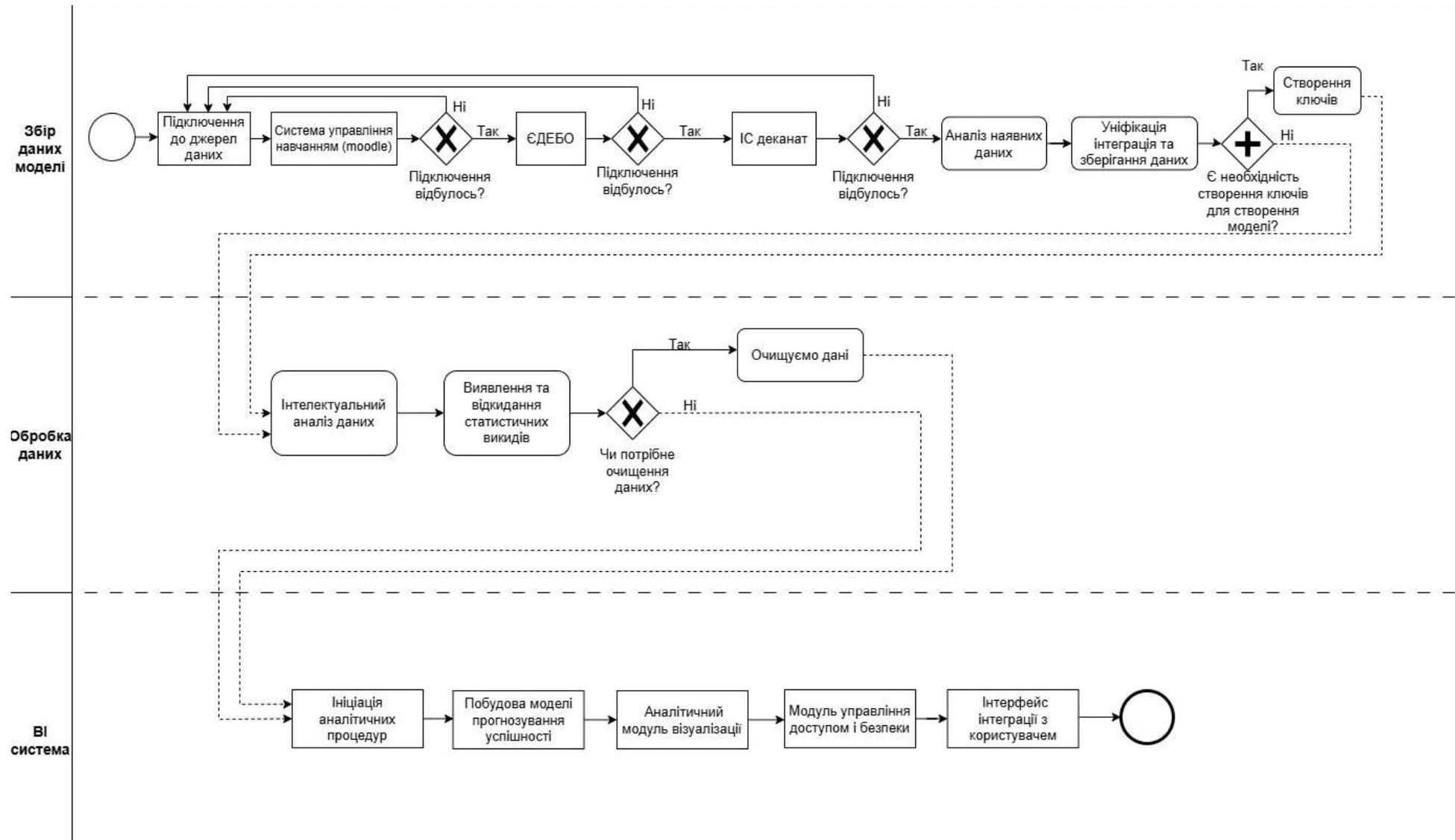


Рис. 3.5. BPMN діаграма бізнес-процесу освітньої аналітики

Після збору дані проходять етап уніфікації, тобто узгодження форматів і структур, що дозволяє інтегрувати їх у єдине аналітичне сховище. На етапі цього етапі відбувається стандартизація інформації за єдиною структурою освітніх подій. Тобто, поєднання записів з LMS (вхід, активність у курсі, результати оцінювання) з демографічною та навчальною інформацією з ЄДЕБО (спеціальність, освітня програма, форма навчання, джерело фінансування) дозволяє сформувати багатовимірний профіль студента. У структурі такого профілю враховуються як кількісні ознаки конкурсний бал, кількість входів, середній бал, тривалість активності), так і категоріальні (стать, курс, статус зарахування тощо). У випадку, якщо для моделювання необхідно створити спеціальні ідентифікаційні ключі або агреговані змінні, система виконує цю операцію автоматично.

Другий блок охоплює «інтелектуальну обробку даних». На цьому етапі здійснюється виявлення статистичних аномалій, перевірка на наявність викидів, дублювань та пропущених значень. При потребі дані проходять процедуру очищення. Такий підхід забезпечує якість і достовірність інформаційної бази, що надалі використовується для побудови моделей. На цьому етапі реалізовано інструментарій статистичного та інтелектуального аналізу. Далі реалізується формування предиктивних і діагностичних моделей. Це дозволяє формувати персоналізовані рекомендації, адаптувати вміст курсів або запропонувати гнучкі моделі навчання.

Завершальний блок системи представлений Ві-компонентами, які забезпечують візуалізацію, керування доступом і формування результатів аналізу. Аналітичний модуль візуалізації генерує звіти, діаграми та дашборди для прийняття управлінських рішень. На цьому етапі в системі Power Ві була побудована модель даних, схема якої зображена на рис.3.6.

Для візуалізації звітів в середовищі інтерфейс користувача дозволяє взаємодіяти з даними у зручний спосіб, зокрема через браузерні додатки. На паралелі модуль побудови моделей прогнозування успішності реалізує

алгоритми машинного навчання для виявлення студентів групи ризику або аналізу впливу факторів на результати навчання.

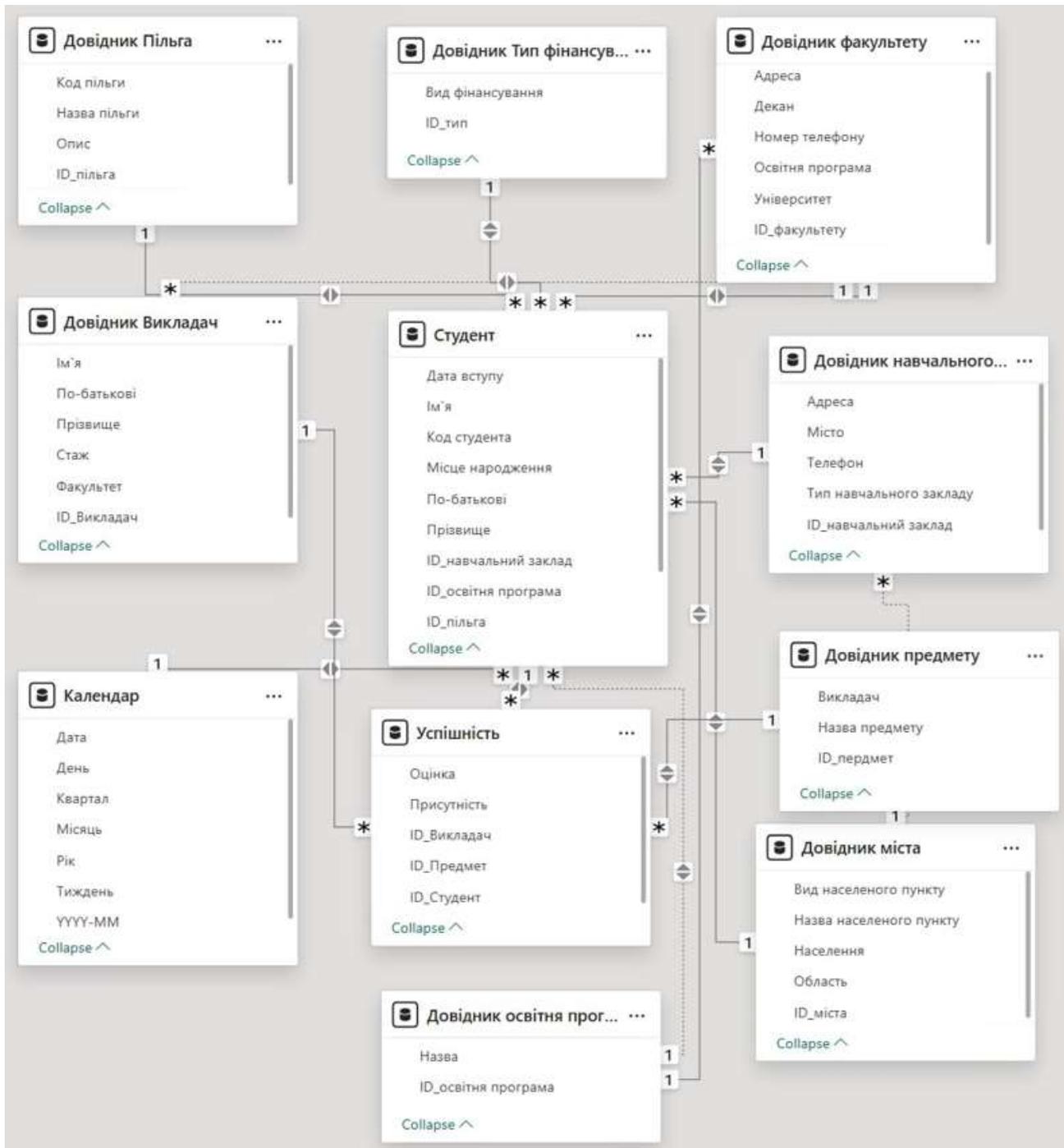


Рис. 3.6. Модель даних з Power BI

Отримані результати аналітики передаються до модуля візуалізації, який реалізує побудову дашбордів, графіків, інтерактивних звітів та теплових карт.

Для цього інтегрується платформа Power BI, яка забезпечує публікацію звітів як у внутрішньому освітньому середовищі, так і у відкритому доступі. Публічні візуалізації дозволяють користувачам аналізувати як індивідуальні освітні траєкторії, так і системні показники якості освіти, зокрема у порівнянні між підрозділами, факультетами або освітніми програмами.

Таким чином, побудована архітектура інформаційної технології реалізує повний цикл освітньої аналітики: від отримання сирих даних до прийняття управлінських рішень на основі аналітичних індикаторів. Вона дозволяє забезпечити науково обґрунтований моніторинг якості освіти, виявлення ризиків академічної неуспішності, оцінку ефективності освітніх програм та персоналізовану підтримку учасників освітнього процесу. А реалізована аналітична модель надає можливість здійснювати не лише ретроспективний аналіз освітніх процесів, а й впроваджувати елементи передбачувального управління, що базується на надійних алгоритмах аналізу даних, адаптивних до динаміки освітнього середовища та потреб ЗВО.

3.2.Формалізація атрибутів та ознак успішності здобувачів освіти

Успішність здобувачів освіти є багатовимірною характеристикою, що охоплює не лише академічні результати, але й рівень сформованості професійних компетентностей, мотиваційні установки, здатність до самостійного вирішення прикладних задач, а також навички командної роботи та комунікації в технічному середовищі. Її не можна обмежити лише кількісними показниками, такими як оцінки чи кількість виконаних лабораторних і проєктних завдань. Сучасні наукові підходи до оцінювання успішності передбачають інтеграцію академічних, поведінкових і технологічних параметрів, що дозволяє сформуванати комплексне уявлення про якість освітнього процесу та ефективність професійного становлення здобувача освіти. Насамперед, ключовими атрибутами успішності виступають:

- рівень засвоєння знань і сформованість професійних компетентностей, які відповідають програмним результатам навчання, передбаченим освітньою програмою;
- інтегровані ключові компетентності, зокрема здатність до критичного та системного мислення, комунікації в різних контекстах, цифрова та інформаційна грамотність, здатність до самоосвіти, прийняття обґрунтованих рішень у складних ситуаціях;
- застосування знань на практиці, що виявляється через виконання прикладних кейсів, участь у проєктній діяльності, стажування, конкурсах або стартап-активностях.

Окреме значення мають поведінкові індикатори, що свідчать про глибину залучення здобувача освіти в навчальний процес: відвідуваність занять (в офлайн або онлайн форматі), своєчасність виконання завдань (дотримання дедлайнів), активність у дискусіях, форумах участь у груповій роботі, а також ініціативність у виборі індивідуальної освітньої траєкторії. Ці чинники, у поєднанні з високим рівнем академічної мотивації, є передумовами для стабільного досягнення результатів. У сучасному освітньому середовищі, що більше орієнтується на кількісні дані, важливу роль у вимірюванні успішності починають відігравати аналітичні показники навчальної діяльності, які акумулюються в електронних системах управління навчанням (LMS), освітніх аналітичних платформах або EMIS. До таких показників належать: кількість входів у систему, тривалість активного навчання, індекси залучення, час на виконання завдань, рівень освоєння матеріалів, участь у форумах чи вебінарах тощо.

Оцінка успішності, таким чином, потребує комплексного підходу, що поєднує кількісні та якісні методи, інтегрує результати освітнього моніторингу, цифрових комунікацій, а також враховує зворотний зв'язок з боку викладачів, роботодавців та самих здобувачів освіти. Тільки на основі таких багатовимірних даних можливе формування об'єктивного уявлення про успішність як одного здобувача, так і ефективність освітньої програми загалом.

Загально визнано, що якість засвоєних знань та рівень академічної успішності студентів значною мірою визначаються ефективністю взаємодії між основними учасниками освітнього процесу. У рамках стрімкого розвитку комп'ютерних технологій, інженерії програмного забезпечення, аналізу великих даних і систем штучного інтелекту, а також в умовах глобалізації, розширення географічної мобільності й підвищення цифрової обізнаності учасників освітнього процесу, цифрові платформи стають ключовими інструментами підготовки висококваліфікованих фахівців у різних галузях.

Разом із тим, в освітньому середовищі часто фіксуються проблеми, пов'язані з недостатнім рівнем мотивації викладачів, обмеженою цифровою компетентністю, а також труднощами, які виникають під час впровадження та повноцінного використання навчальних систем. На рис.3.7 узагальнено фактори, які впливають на успішність навчання студентів при використанні платформ електронного навчання, зокрема LMS Moodle.

Рівень навчальних досягнень здобувачів освіти розглядається як складова інтегрального індикатора якості сформованих знань, умінь і навичок, що значною мірою залежить від ефективності програмно-апаратного забезпечення електронного навчального середовища, змістовного наповнення та логіко-структурної організації навчальних матеріалів, обраних педагогічних підходів, форм організації комунікації між викладачем і здобувачем освіти, наявності інтерактивних засобів комунікації між учасниками освітнього процесу, а також дієвості систем моніторингу та оцінювання результатів навчання.



Рис.3.7. Графічна модель успішності здобувача освіти

3.3. Математичні моделі оцінювання успішності здобувачів освіти

Враховуючи суттєвий вплив якості навчального контенту, якості кадрового забезпечення та якості платформ навчання на академічні результати здобувачів освіти, особливої актуальності набуває впровадження формалізованих методик їх оцінювання. Зокрема, ефективними інструментами виступають методи статистичного аналізу, які забезпечують можливість

дослідження інтенсивності використання освітніх ресурсів, характеру взаємодії з ними, результатів успішності, а також отримання якісного зворотного зв'язку від усіх учасників освітнього процесу. Додатково, застосування алгоритмів штучного інтелекту відкриває можливості для виявлення прихованих закономірностей у навчальній поведінці здобувачів та здійснення прогнозування їхньої успішності на основі аналізу багатовимірних освітніх даних. З іншого боку і професійна та цифрова компетентність викладачів, як ключової ланки у реалізації змісту освітніх програм, потребує власної формалізації. З цією метою доцільним є використання апарату теорії множин у поєднанні зі статистичними показниками, що відображають відповідність кваліфікаційного рівня вимогам ліцензійних умов, а також результативність наукової, методичної та освітньої діяльності викладацького складу. Такий підхід забезпечує цілісну та обґрунтовану оцінку кадрового потенціалу в контексті цифрового управління якістю освіти. Зазначені фактори успішності є складними інтегральними характеристиками, які доцільно розглядати через призму їх структурних компонентів. Це передбачає розроблення системи базових атрибутів та відповідних кількісних метрик. Для представлення концептуальної математичної моделі, яка формалізує формування успішності здобувача освіти та враховуючи, що критерії впливу на успішність за весь термін навчання є комплексними, доцільно використати систему моделей, які на верхньому рівні описує концептуальну сутність, на середньому – атрибути, які формують характеристики вищого рівня, а на нижньому – метрики та методи які дозволяють кількісно виражати та обчислювати значення атрибутів.

Концептуально, успішність студентів можна розглядати у вигляді максимізації функції, що залежить від системи факторів впливу, кожен з яких також є функціями

$$Y = \{Q_m, Q_k, Q_t, Q_s\} \rightarrow \max, \quad (3.1)$$

де:

Y – результативний показник успішності студента є інтегральним

індикатором, який може включати як академічні результати, так і поведінкові.

Q_m – інтенсивність та регулярність використання системи управління навчанням (LMS Moodle) важливий параметр, що відображає технічну взаємодію студента з освітнім середовищем.

Q_k –якість, структурованість і доступність освітніх ресурсів у цифровому середовищі визначає змістовну складову навчання.

Q_t – рівень професійної компетентності викладача та ефективність його педагогічних впливів впливає на якість викладання, мотивацію та залучення студентів.

Q_s – навчальна активність здобувача освіти, включаючи участь в електронних курсах, виконання завдань та комунікацію, характеризує активність, самостійність і відповідальність у навчальному процесі.

Запропонована система змінних для оцінювання успішності студента є логічно обґрунтованою та охоплює ключові чинники, що впливають на результативність навчання в цифровому середовищі. Вона враховує технічний аспект (Q_m – використання LMS), змістовий компонент (Q_k –якість ресурсів), педагогічний вплив (Q_t –професійність викладача) та особисту активність студента (Q_s – ефективність роботи). Це дозволяє розглядати успішність не як ізольований результат, а як функцію взаємодії між різними елементами освітнього середовища. Для формалізації показників якості та ефективності функціонування платформи дистанційного навчання доцільно використовувати модель, що описує вплив технічних факторів на академічну успішність здобувачів освіти та якість сприйняття ними навчального контенту.

У цьому контексті модель Q_m (якість технічної складової платформи) подається у вигляді системи:

$$Q_m = \{Q_q(Q_{use}, Q_{pm}, Q_{pc}, Q_{la}, Q_{err}, Q_{ui}, Q_{ac}, Q_{sc}, Q_{adapt})Q_A\}, \quad (3.2)$$

Модель Q_m подається як складова, що включає:

Q_q –якість програмного продукту, яка базується на стандарті ISO/IEC 25010, що

охоплює:

Q_{use} –якість використання (user experience, зручність, ефективність);

Q_{pm} –якість програмних модулів (функціональність, надійність, сумісність);

Q_{pc} –якість програмного коду (читабельність, підтримуваність, ефективність виконання).

Q_{ia} –затримки при завантаженні ресурсів (latency)

Q_{err} –частота технічних збоїв або помилок (error rate).

Q_{ui} –зручність інтерфейсу користувача (user interface usability)

Q_{ac} –доступність для користувачів з особливими потребами (accessibility compliance, наприклад WCAG)

Q_{sc} –здатність платформи витримувати зростання кількості користувачів.

Q_{adap} –адаптивність до різних пристроїв (мобільна версія, планшети).

QA –якість апаратного забезпечення (сервери, мережеві ресурси, доступність, стабільність роботи).

Ця модель дозволяє комплексно оцінити технічну складову платформи, яка прямо впливає на сприйняття навчального контенту та успішність студентів.

Формалізоване представлення розміщених у системі онлайн навчання освітніх матеріалів (ресурсів та діяльностей) Q^k можна відобразити набором відповідних атрибутів чи властивостей. Однак, при їх визначенні, потрібно також обґрунтувати показник для оцінювання міри реалізації того чи іншого атрибуту. У зв'язку з цим формальний вигляд якості та ефективності навчальних матеріалів пропонується представити так (3.3)

$$Q^k = \left\{ \begin{array}{l} \underbrace{Q_i^k}_{A_{ij}^k} \left| \begin{array}{l} M_{ij}^k \\ \widetilde{A}_{ij}^k \end{array} \right. \end{array} \right\} \quad (3.3)$$

Q_i^k –деяка комплексна властивість, що описує якість наявного матеріалу в електронному курсі, $i = 1..n$ - кількість комплексних властивостей;

A_{ij}^k – атрибут, який відображає якість опублікованого навчального матеріалу, $j = 1..k$ – кількість характеристик, що описують властивості;

M_{ij}^k – кількісна метрика представлення j -го атрибуту i -ї властивості якості навчального матеріалу.

Наведемо перелік комплексних властивостей навчального контенту, кожна з яких описано через вимірні атрибути та кількісні метрики (табл 3.1). Запропонована модель виступає базисом для побудови системи моніторингу та оцінювання якості матеріалів у цифровому середовищі.

Таблиця 3.1

Перелік комплексних властивостей навчального контенту

Комплексна властивість Q_i^k	Вимірні атрибути A_{ij}^k	Кількісні метрики M_{ij}^k
Змістова повнота Q_1^k	Обсяг тем робочої програми, що охоплюються матеріалом Відповідність програмним результатам навчання	% тем зі списку програми, що представлені в матеріалі
Актуальність Q_2^k	Дата останнього оновлення Посилання на сучасні джерела	термін матеріалу (у місяцях/роках); кількість зовнішніх бібліографічних посилань за останні 5 років
Якість викладу Q_3^k	Число логічних блоків Наявність висновків	середній бал студентських відгуків; % позитивних коментарів
Складність (рівень) Q_4^k	Кількість ключових термінів Глибина пояснень	середній результат тестів за матеріалом (%); середня тривалість проходження модуля (хвилин)
Інтерактивність Q_5^k	Число інтерактивних елементів (тести, симуляції) Кількість елементів медіа (відео, анімації)	% студентів, які взаємодіяли з інтерактивом; середня тривалість взаємодії з кожним елементом (сек)
Доступність Q_6^k	Наявність субтитрів/альтернативних форматів Відповідність WCAG	% контенту з субтитрами/транскриптом; Рейтинг доступності за перевіркою WCAG (%)
Зрозумілість Q_7^k	Кількість прикладів/ілюстрацій; Рівень читаємості тексту (Flesch–Kincaid)	Індекс читаємості (F–K score); % повторних звернень до матеріалу (skips)
Залученість Q_8^k	Кількість коментарів/запитань на форумі Активність у дискусіях	середня кількість коментарів на студента; % студентів, що взяли участь у дискусії, форумах
Технічна стабільність Q_9^k	Час завантаження сторінки Кількість технічних помилок	Середній час завантаження (сек); кількість HTTP-помилки типу 404 (Not Found) та 500 (Internal Server Error), що виникають у межах однієї сесії користувача

Аналогічно до (3.3) можна описати професійність викладачів при

провадженні дистанційної освіти. Формула (3.4) дуже подібна до попередньої, єдине, що відрізняється складом та інтерпретацією властивостей.

$$Q^t = \left\{ \begin{array}{l} \underbrace{Q_i^t}_{A_{ij}^t} \left| \begin{array}{l} M_{ij}^t \\ \widetilde{A}_{ij}^t \end{array} \right. \end{array} \right\} \quad (3.4)$$

Q_i^t – комплексна властивість показника професійності викладача (освіта, вчене звання, кваліфікація, відповідність ліцензійним умовам, практичний досвід роботи), $i = 1..n$ - кількість комплексних властивостей;

A_{ij}^t – атрибут, який відображає ефективність інструкторів курсу,

$j = 1..k$ - кількість властивостей;

M_{ij}^t – кількісна метрика для представлення оцінки професійності викладача.

Розроблена структурована пропозиція комплексних властивостей професійної компетентності викладача, яка враховує сучасні вимоги освітнього процесу, цифровізації, інноваційної діяльності та відповідності стандартам якості освіти (табл. 3.2).

Таблиця 3.2

Перелік комплексних властивостей професійної компетентності викладача

Комплексна властивість Q_i^t	Вимірні атрибути A_{ij}^t	Кількісні метрики M_{ij}^t
Науково-кваліфікаційний рівень Q_1^t	Освітній та науковий ступінь, вчене звання Спеціальність за дипломом	-PhD, доктор наук, професор, доцент (бінарна чи шкала балів) Стаж у науково-педагогічній роботі (роки)
Досвід педагогічної діяльності Q_2^t	Загальний стаж роботи, Педагогічне навантаження Кількість дисциплін, що викладає	Роки/місяці Години/рік N дисциплін
Дослідницька активність Q_3^t	Публікації у виданнях категорії А, В, Н-індекс Цитування у Scopus/Web of Science/Google Scholar	- Кількість публікацій за 5 років Н-індекс Кількість цитувань
Методична	Розроблені навчально-методичні	-N методичних

активність Q_4^t	матеріали Участь в розробці освітніх програм	матеріалів N навчальних курсів, оновлених згідно з ОП
Цифрова компетентність Q_5^t	Володіння LMS, цифровими платформами Рівень цифрової грамотності	Сертифікати (Coursera, Prometheus, Google) Кількість інтерактивних елементів у курсі
Інноваційна педагогічна активність Q_6^t	Застосування змішаного/онлайн навчання Інтерактивні технології, проєктне навчання	- % дисциплін із застосуванням ІКТ- Кількість новацій у курсах
Професійний розвиток і підвищення кваліфікації Q_7^t	Кількість курсів/тренінгів/стажувань за останні 3 роки	N сертифікатів Загальна тривалість навчання (годин)
Міжнародна активність Q_8^t	Участь у міжнародних конференціях/проєктах Публікації у міжнародних журналах	N конференцій N проєктів Кількість спільних публікацій
Якість освітньої діяльності (викладання) Q_9^t	Відгуки та оцінки студентів Аналітика курсу (доступність, актуальність, зворотний зв'язок, інтерактивність)	Середній бал студентського опитування % зворотного зв'язку на повідомлення в LMS
Організаційна активність Q_{10}^t	Участь у роботі вчених/методичних рад, кафедральних комісіях	Кількість засідань, у яких взяв участь Роль (учасник, керівник, координатор)

Аналогічно до (3.3) опишемо активність та наполегливість роботи здобувача освіти в системі електронного навчання. Формула (3.5) має спільну концепцію з попередніми, але відрізняється складом та інтерпретацією властивостей.

$$Q^s = \left\{ \begin{array}{c} Q_i^s \\ A_{ij}^s \end{array} \middle| \begin{array}{c} M_{ij}^s \\ \widetilde{A}_{ij}^s \end{array} \right\} \quad (3.5)$$

Q_i^s – показники зацікавленості та наполегливості роботи студента),

$i = 1..n$ - кількість комплексних властивостей;

A_{ij}^s – атрибут, що описує ефективність роботи в ЕНК,

$j = 1..k$ - кількість властивостей;

M_{ij}^s – кількісна метрика для представлення ефективності та

вмотивованості роботи студента.

Розроблений перелік для представлення властивостей зацікавленості та наполегливості студента, відповідних атрибутів та метрик їх оцінювання наведено в таблиці 3.3

Таблиця 3.3

Перелік комплексних властивостей зацікавленості та наполегливості здобувача освіти

Комплексна властивість Q_i^S	Вимірні атрибути A_{ij}^S	Кількісні метрики M_{ij}^S
Залученість до навчання Q_1^S	Кількість входів до LMS Середня тривалість сесії Частота активності протягом тижня	Вхід/тиждень Години за сеанс або тиждень Активні дні / 7
Регулярність навчання Q_2^S	Розподіл активностей по календарю курсу	% тижнів з активністю
Цілеспрямованість Q_3^S	Відсоток виконаних обов'язкових завдань Середній час виконання завдань	% виконання Годин/завдання
Ініціативність Q_4^S	Кількість повідомлень у форумах	Повідомлень/курс
Наполегливість у навчанні Q_5^S	Кількість повторних переглядів/проходжень тестів Рівень завершеності курсу	Повторів/тема % завершених модулів
Організованість / вчасність Q_6^S	Відсоток завдань, зданих до дедлайну	% вчасно зданих
Рефлексивність / самооцінка Q_7^S	Участь в опитуваннях про власний прогрес Середня оцінка саморефлексії	% заповнених форм Бал за шкалою

Ще одним з важливих показників і критеріїв, який необхідно врахувати визначаючи якість освітнього процесу, є відгук про курс здобувачів освіти. Для цього у платформах онлайн навчання передбачають модулі зворотного зв'язку, або функціонал, що дає змогу під час або після проходження курсів мати відгук у результаті опрацювання анкет про якість та практичність контенту, простоту та доступність матеріалу, про задоволеність організацією та структурою навчання на відповідній платформі, тощо. Для формалізації показника відгуків з позиції прогнозування успішності студентів можна представити в наступному вигляді:

$$F = \{N(M), Z\} \quad (3.6)$$

де $N(M)$, – множина моделей представлених у (3.1);

Z – одне із значень множини відносно типу відгуку, що характеризує якість ЕНК, якість матеріалу ЕНК, методів викладання викладача, та якість особистої наполегливості при роботі з ЕНК, що описується категорійними змінними. Z можна представити у вигляді множини $Z = \{p, n, n_0\}$ (3.7)

де p – значення, що фіксує позитивний відгук про вплив на успішність;

n – значення, що відповідає негативному відгуку і впливу на успішність;

n_0 – значення, що характеризує нейтральний вплив курсу на задоволеність здобувача, тобто курс жодним чином не вплине на успішність здобувача.

У результаті сформовано концептуальну систему математичних моделей, яка забезпечує можливість структурованого аналізу та ідентифікації ключових чинників, що впливають на успішність здобувачів освіти в умовах використання систем дистанційного навчання, зокрема LMS Moodle (рис.3.8).

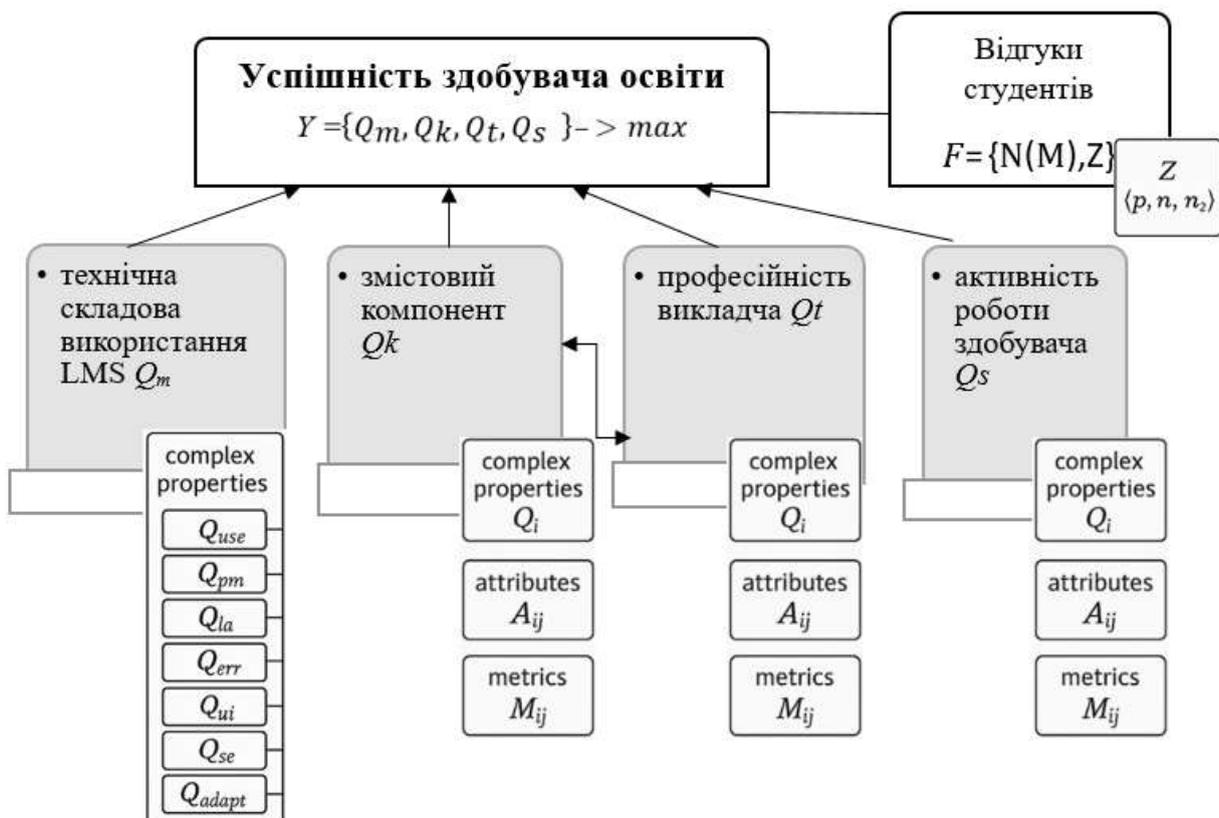


Рис. 3.8. Система математичних моделей, для ідентифікації ключових чинників, що впливають на успішність здобувачів освіти

Розроблена узагальнена система математичних моделей формування успішності здобувачів освіти відображає повний спектр факторів, що потенційно впливають на результати навчання, зокрема характеристики освітнього контенту, професійність викладача, технічна стабільність освітнього середовища та поведінкова активність студента. Зазначена модель має концептуальний характер і слугує теоретичною основою для формування інформаційної технології освітньої аналітики. Разом із тим, під час експериментальної апробації буде використано звужену множину змінних, які можуть бути об'єктивно, кількісно та автоматизовано виміряні на основі наявних цифрових слідів у освітніх інформаційних системах (LMS Moodle, ЄДЕБО, ІС «Деканат»). Саме тому до експериментальної моделі включено такі атрибути, як демографічні характеристики здобувачів освіти, результати вступної кампанії та показники активності в LMS. Показники якості освітнього контенту та професійності викладача, попри їх важливість з педагогічної точки зору, у досліджуваному освітньому середовищі не мали уніфікованих числових індикаторів, придатних для використання в алгоритмах машинного навчання без додаткових суб'єктивних оцінок або експертного анкетування. Включення таких параметрів у експериментальну модель могло б знизити достовірність результатів та суперечити принципам доказової освітньої аналітики.

Таким чином, відмінність між теоретичною та експериментальною моделями зумовлена переходом від концептуального рівня опису до практично реалізованої моделі, адаптованої до реальних умов функціонування цифрового освітнього середовища. Запропонована інформаційна технологія передбачає можливість подальшого розширення набору ознак за умови появи формалізованих і стандартизованих показників якості контенту та викладацької діяльності.

3.4. Статистичний аналіз поведінкових та академічних предикторів успішності здобувачів освіти

З огляду на залучені джерела даних доцільно виділити такі блоки предикторів: демографічні, академічні та поведінкові. Аналіз предикторів проводиться з метою визначення суттєвих факторів, що впливають на якість навчання та успішність здобувачів освіти. З огляду на наявні інформаційні дані відкритих джерел та розраховані аналітичні показники побудовано лінійну регресійну модель, яка відображає залежність результатів освітньої діяльності від факторних ознак. Лінійна регресія залишається одним із найчастіших досліджень при виявленні стійких статистично-значимих закономірностей. Лінійна багатофакторна регресійна модель розглядає дві групи змінних:

- *незалежні змінні* (предиктори, x_i) – змінні, що впливають на інші показники та змінюється самостійно.

- *залежна змінна* (цільова змінна, y) – змінна, значення якої прогнозується на основі незалежних змінних.

Модель визначає оптимальну лінійну функцію, що описує цей зв'язок

$$y(x) = a_0 + a_1x_1 + a_2x_2 + \dots + a_nx_n \quad (3.8)$$

де: y – результативна ознака (якість підготовки випускників як успішність за весь час навчання); x_i – факторні ознаки, які впливають на успішність; n – кількість факторних ознак; Як факторні ознаки використаємо атрибути, які отримали з LMS, ІС Деканат та ЄДБО:

- *ID здобувача освіти*
- *Відмітка присутності (present/absent)* – дозволяє аналізувати рівень участі студента в навчальному процесі;
- *Стать студента* – додатковий соціальний параметр, який може впливати на освітні траєкторії;
- *Форма фінансування* – бюджет або контракт (після кодування: 0 – бюджет, 1 – контракт);
- *Курс навчання* – актуальний рівень студента в межах освітньої програми;
- *Конкурсний бал при вступі* – значення, з яким студент зарахований до ЗВО;

- *Кількість входів на курс* –цифрова активність у межах дистанційного курсу (кількість авторизованих входів);

Підсумковий бал з дисципліни –результат навчання, який розглядається як основний індикатор успішності та результативна ознака.

Вищезазначені показники були отримані чи розраховані на основі використання даних LMS Moodle НУБП України. Для з'ясування, наскільки зібрані показники можуть бути прогностичними для майбутньої успішності, було побудовано діаграму розсіювання (рис.3.9) для перевірки зв'язку між конкурсним балом та підсумковим балом.

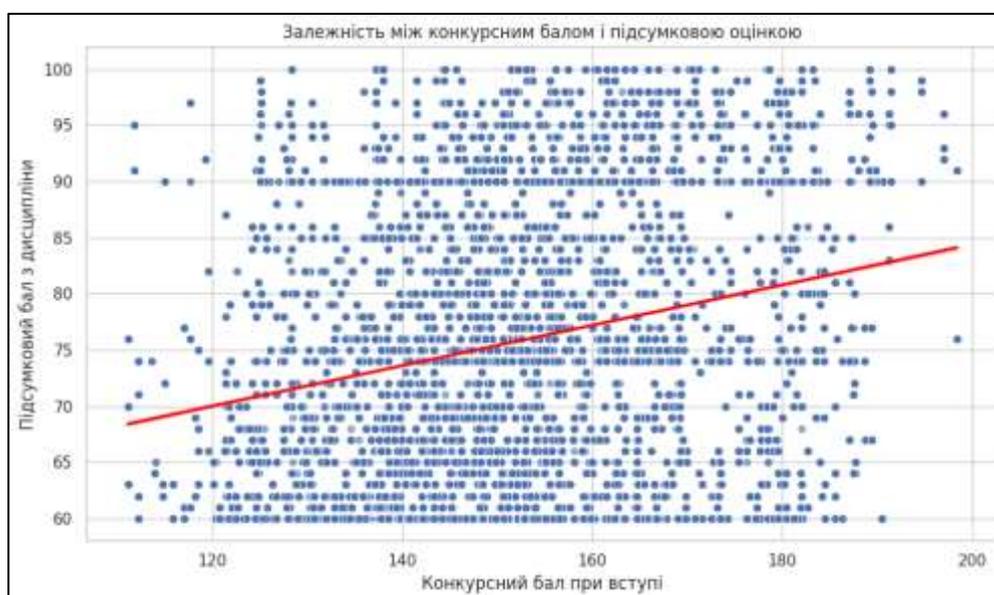


Рис. 3.9. Діаграма розсіювання між конкурсним балом при вступі та підсумковим балом навчання

Аналіз вихідних даних свідчить, що здобувачі освіти з вищими конкурсними балами, як правило, демонструють кращі академічні результати. Водночас серед студентів із однаковими вступними показниками спостерігається варіативність у рівні підсумкової успішності, що вказує на багатофакторну природу цього показника та підтверджує, що конкурсний бал не є визначальним або єдиним предиктором академічних досягнень.

На рис. 3.10 представлено лінійний графік, що ілюструє зв'язок між кількістю відвідувань ЕНК (лекцій, практичних занять) та підсумковими балами студента з відповідного освітнього компонента. Представлений точковий графік

ілюструє взаємозв'язок між показниками активності здобувачів освіти, зокрема, кількістю відвідувань (участь у лекціях, практичних заняттях або взаємодію з віртуальними ресурсами курсу) та їх підсумковими оцінками. Нанесена також лінія лінійного тренду та довірчого інтервалу.

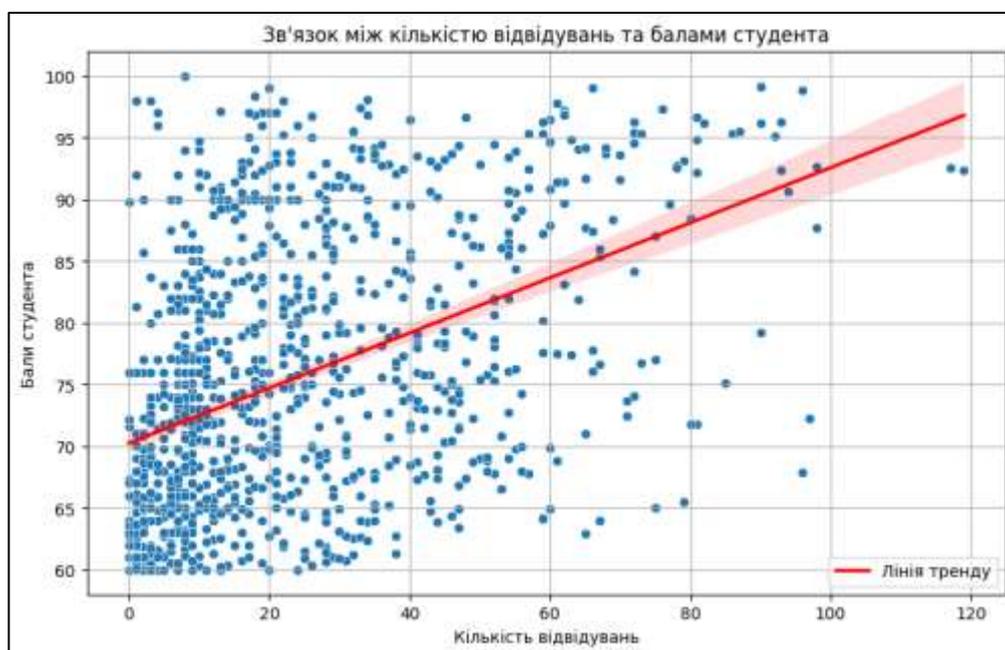


Рис. 3.10. Зв'язок між відвідуванням занять та підсумковими балами

Аналіз графіка свідчить про наявність позитивної кореляції між кількістю відвідувань та рівнем академічної успішності: із зростанням кількості відвідувань, як правило, підвищуються й підсумкові бали. Водночас, наявний розкид емпіричних значень відносно трендової прямої демонструє, що існують окремі випадки, коли висока відвідуваність не корелює з високими результатами, а також ситуації, коли студенти із мінімальним рівнем участі демонструють задовільну або високу успішність. Це вказує на потребу врахування додаткових чинників при аналізі успішності освітнього процесу.

Перед проведенням детального кореляційного аналізу, для попередньої візуальної оцінки зв'язків між кількісними змінними успішності, рис.3.11 побудовано матрицю розсіювання що ілюструє розподіли та парні взаємозв'язки між "Балами студента", "Конкурсним балом при вступі (КБ при вступі)" та

"Кількістю входів на курс".

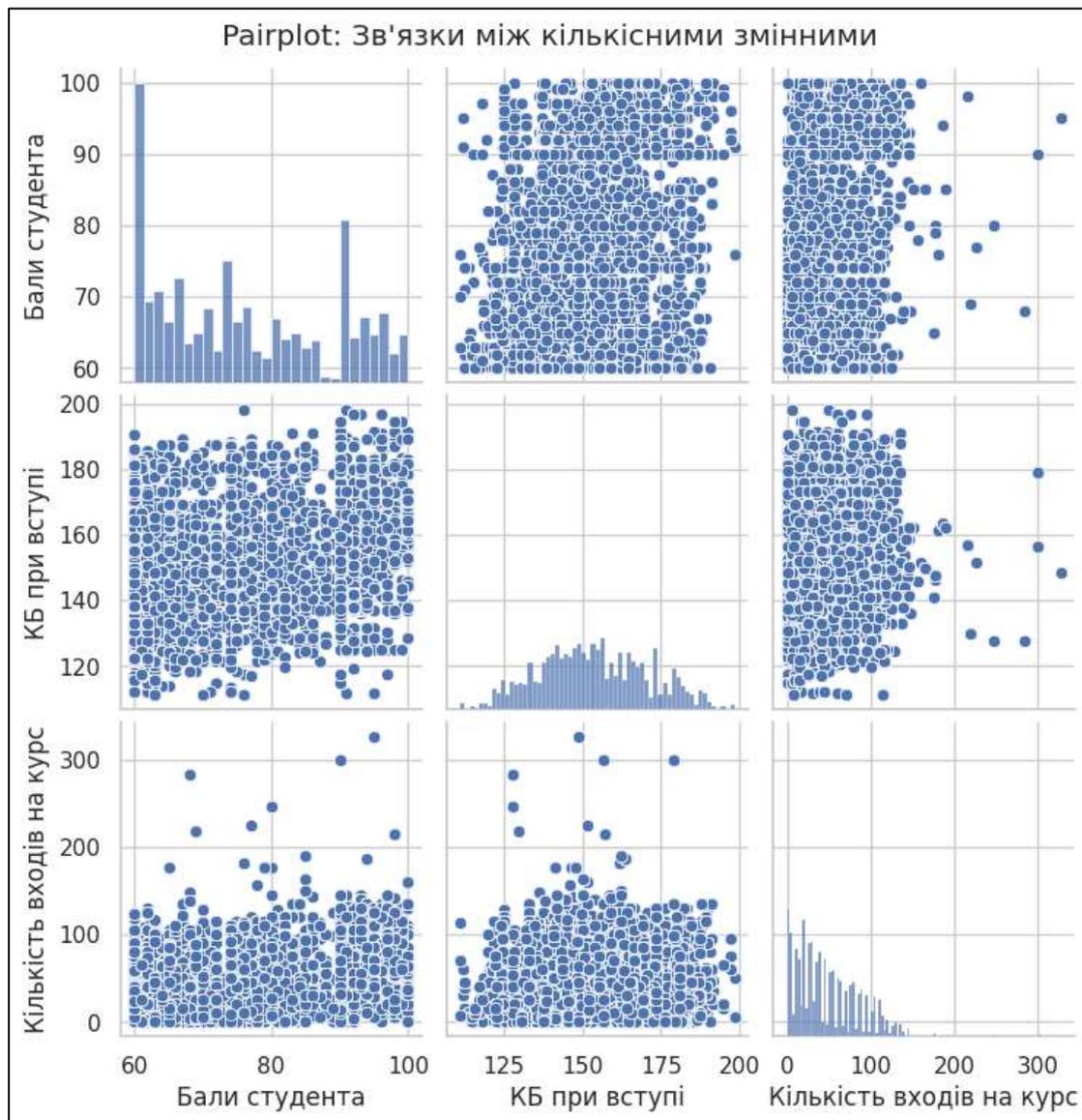


Рис. 3.11. Матриця розсіювання (Pairplot) зв'язків між кількісними змінними успішності

Аналіз розподілів показав, що бали студента мають кілька піків, конкурсний бал при вступі – майже нормальний розподіл (пік 150-160), а кількість входів на курс сильно скошений вправо з переважанням низьких значень (0-20).

Результати аналізу парних взаємозв'язків між змінними не виявили наявності сильних лінійних кореляцій. Зокрема, високий конкурсний бал при вступі не обов'язково корелює з високими академічними результатами студента. Так само, кількість входів до навчального курсу не демонструє вираженої лінійної залежності з підсумковими оцінками, а також слабо пов'язана з

конкурсним балом. Отже, зазначені три змінні: конкурсний бал, кількість входів до курсу та підсумковий бал не є вичерпними предикторами академічної успішності або рівня залученості здобувачів освіти, що свідчить про необхідність розширення моделі за рахунок додаткових факторів або використання в інформаційній технології методів машинного навчання.

Побудуємо кореляційну матрицю, щоб встановити міру взаємозв'язку між змінними. Дані наведено в таблиці 3.4.

Таблиця 3.4

Кореляційна матриця атрибутів

	Стать	Джерело фінансування	Конкурсний бал при вступі	Активність в LMS	Успішність студента
Стать X_1	1				
Джерело фінансування X_2	-0,068253	1			
Конкурсний бал при вступі X_3	0,119743	-0,567698124	1		
Активність в LMS X_4	0,073592	-0,107342992	0,103279	1	
Успішність студента Y	0,21538	-0,252690386	0,233815	0,357614	1

Кореляційний аналіз виявив статистично значиму залежність між змінними «Джерело фінансування» та «Конкурсний бал при вступі», для яких встановлено середній негативний кореляційний зв'язок ($r = -0,568$). Це свідчить про те, що здобувачі освіти з вищими конкурсними балами частіше навчаються за кошти державного бюджету, тоді як студенти з нижчими вступними балами переважно фінансують навчання самостійно. Також зафіксовано слабку негативну кореляцію між «Джерелом фінансування» та «Успішність студента» ($r = -0,253$), що демонструє загальну тенденцію до вищої академічної результативності серед студентів-бюджетників. Проте помірна сила зв'язку не дозволяє зробити остаточні висновки щодо наявності причинно-наслідкових залежностей.

Водночас встановлено позитивну кореляцію між «Активність в LMS» та «Успішність студента» ($r = 0,358$), яка вказує на наявність зв'язку між

цифровою активністю та навчальною успішністю: чим частіше студент взаємодіє з освітнім середовищем, тим вищими є його оцінки. Щодо змінної «Стать студента», виявлено слабкий позитивний зв'язок з успішністю ($r = 0,215$), який може відображати незначні гендерні відмінності в результатах навчання, але не має визначального характеру.

<i>Регресійна статистика</i>					
Множинний R	0,459784343				
R-квадрат	0,511401642				
Нормований R-квадрат	0,210265333				
Стандартна помилка	10,90085838				
Спостереження	2781				

Дисперсійний аналіз						
	<i>Df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значимість F</i>	
Регресія	4	88428,72113	22107,18028	186,04241	2,018E-141	
Залишок	2776	329868,5086	118,8287135			
Всього	2780	418297,2298				

	<i>Коефіцієнти</i>	<i>Стандартна помилка</i>	<i>t-статистика</i>	<i>P-Значення</i>	<i>Нижні 95%</i>	<i>Верхні 95%</i>
Y-перетин	62,61444685	2,438859192	25,67366212	1,3E-130	57,83228561	67,396608
X1	5,581224524	0,557603255	10,00931123	3,413E-23	4,487865514	6,6745835
X2	-3,848840303	0,51258311	-7,508714647	7,999E-14	-4,85392296	2,8437576
X3	0,066838242	0,014791984	4,518544688	6,487E-06	0,037833839	0,0958426
X4	0,108619433	0,005792297	18,75239467	5,708E-74	0,097261788	0,1199771

Результати множинної регресії підтвердили статистично значущий вплив обраних предикторів на навчальну успішність здобувачів освіти. Отримане значення коефіцієнта детермінації ($R^2 = 0,51$) свідчить, що модель пояснює понад 51% варіації успішності, а це достатній рівень пояснювальної здатності моделі, що є типовим для освітніх досліджень із складною мультифакторною природою результатів навчання. Аналіз коефіцієнтів показав, що найбільш вагомими факторами виступають показники цифрової активності студентів у середовищі LMS. Водночас відмінність між коефіцієнтами R^2 та нормованим R^2 свідчить про можливу наявність шуму або слабших предикторів та підтверджує доцільність застосування нелінійних методів інтелектуального аналізу даних, які дозволяють

врахувати складні взаємозв'язки між освітніми показниками. Це обґрунтовує використання нелінійних ансамблевих моделей у подальших експериментах.

Отримані результати підтвердили мультифакторний характер освітньої успішності та були використані для оптимізації структури ознак у прогностичних моделях. Це свідчить про можливу наявність інших, більш вагомих чинників, не включених до моделі, зокрема таких, як рівень навчальної мотивації, якість викладацької діяльності, інтенсивність академічної підтримки, соціально-економічне становище студентів, регулярність відвідування занять тощо, які включені до розробленої концептуальної моделі (3.1).

Висновки до розділу 3.

У третьому розділі було розроблено та обґрунтовано інформаційну технологію освітньої аналітики як інтегроване програмно-алгоритмічне рішення, що забезпечує повний цикл аналітичної обробки освітніх даних у закладі вищої освіти. Запропонована технологія дозволяє об'єднувати різноманітні дані з LMS, ІС «Деканат» та інших цифрових сервісів, здійснювати їх очищення, нормалізацію, інтеграцію, побудову багатовимірних наборів ознак і подальше моделювання навчальної успішності здобувачів освіти.

На основі проведених досліджень вдалося сформувати архітектуру технології освітньої аналітики, яка включає модулі збору даних, формування ознак, моделювання, оцінювання результатів та візуалізації. Використання BPMN є логічним доповненням UML-моделювання: UML описує програмну та системну архітектуру, а BPMN відображає бізнес-процеси та організаційний контекст. Поєднання цих підходів дозволило забезпечити комплексне представлення інформаційної технології освітньої аналітики на концептуальному, функціональному, організаційному та технічному рівнях. Було розроблено формальний опис атрибутів, що характеризують демографічні, освітні та поведінкові характеристики студентів, а також алгоритмічні

процедури їх перетворення у структурований формат, придатний для інтелектуального аналізу.

Запропонована узагальнена система математичних моделей формування навчальної успішності здобувачів освіти охоплює комплекс факторів, що потенційно впливають на результати навчання, зокрема характеристики освітнього контенту, професійну компетентність викладача, технічну надійність освітнього середовища та поведінкову активність студентів. Дана модель має концептуальний характер і використовується як теоретична основа для розроблення інформаційної технології освітньої аналітики.

Таким чином, у розділі повністю реалізували поставлене завдання – розроблено інформаційну технологію освітньої аналітики як комплексне інтегроване рішення, що поєднує методи інтелектуального аналізу даних із наявними освітніми інформаційними системами НУБП України та забезпечує можливість подальшого прогнозування успішності здобувачів освіти.

РОЗДІЛ 4

ПРОГНОЗУВАННЯ УСПІШНОСТІ ЗДОБУВАЧІВ ОСВІТИ З ВИКОРИСТАННЯМ АЛГОРИТМІВ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ

4.1. **Методологія та етапи побудови прогностичної моделі успішності здобувачів освіти**

Прогнозування успішності здобувачів освіти є ключовим елементом освітньої аналітики, що сприяє покращенню якості навчального процесу, своєчасному виявленню ризиків академічної неуспішності та наданню індивідуальної підтримки студентам. Методологія формування таких прогнозів базується на інтеграції статистичних, машинних та аналітичних підходів, які забезпечують обробку освітніх даних і виявлення закономірностей у поведінці та результатах учнів.

Основою цієї методології є етап попередньої обробки даних (data preprocessing), що включає очищення, нормалізацію та кодування вхідних змінних. До них належать соціально-демографічні характеристики студентів (вік, стать, джерело фінансування), навчальні показники (конкурсний бал, оцінки, відвідуваність, цифрова активність) а також параметри освітнього середовища. Ці дані формують набір предикторів для використання в математичних моделях прогнозування.

Перед початком інтелектуального аналізу даних необхідно провести попередній аналіз (Exploratory Data Analysis). Цей етап передбачає дослідження основних характеристик набору даних з метою отримання уявлення про ключові закономірності та взаємозв'язки. Головне завдання полягає у виявленні найзначніших характеристик і змінних даних, а також виявленні потенційних аномалій чи викидів у неоднорідних даних. Це полегшує вибір статистичних методів і алгоритмів машинного навчання для ефективного аналізу.

Для етапу EDA використовувались різноманітні інструменти, включаючи статистичні методи (кореляційний і регресійний аналіз) та методи візуалізації (гістограми, секторні діаграми, діаграми розсіювання і теплові карти). Ці

інструменти допомагають знаходити закономірності у даних, визначати зв'язки між змінними й відкривати унікальні або несподівані патерни для подальшого дослідження.

Прогнозування академічної успішності студентів є критично важливим завданням для ухвалення обґрунтованих рішень щодо їхнього навчального прогресу. Алгоритми інтелектуального аналізу можуть бути застосовані для прогнозування успіхів студентів і визначення чинників, які впливають на точність цих прогнозів. На рисунку 4.1. показано методологію реалізації запропонованої роботи.

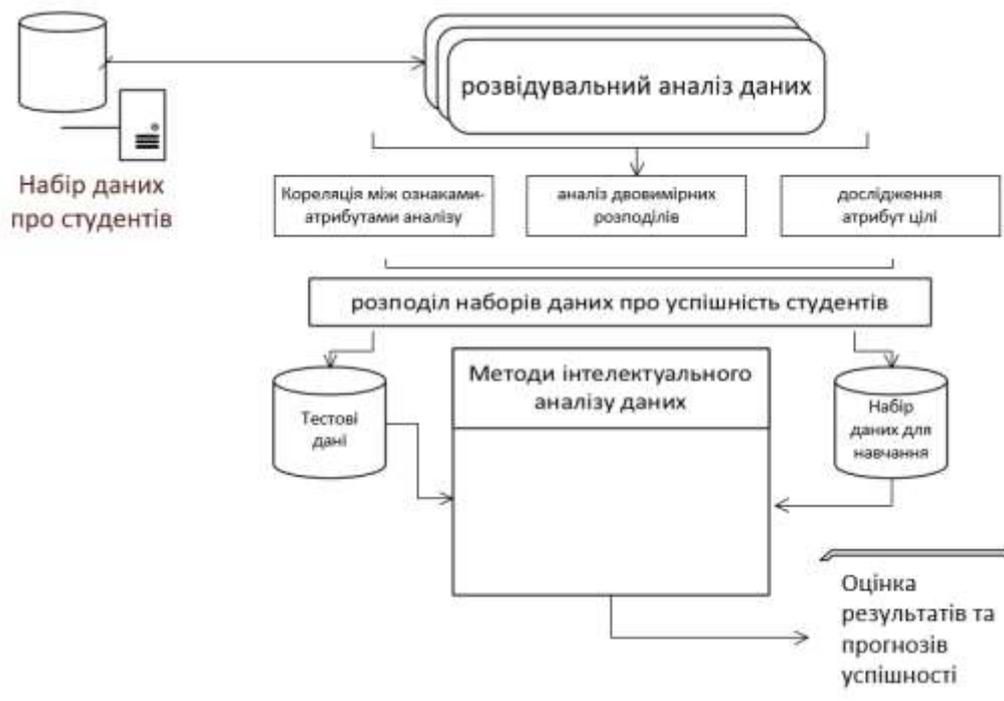


Рис. 4.1. Візуалізація методології прогнозування академічної успішності здобувачів освіти

Оцінка результатів супроводжується етапом валідації (train/test split, крос-валідація), що дозволяє оцінити точність та узагальнювальну здатність моделей. При використанні ML-моделей використовуються метрики точності класифікації (Accuracy, Precision, Recall, F1-score) та середньоквадратичні помилки для регресії (RMSE, MAE). Інтерпретація результатів передбачає урахування педагогічного контексту, адже прості регресійні моделі не дають можливості зробити чіткі висновки, а при використанні складних моделей

машинного навчання (Дерева рішень, Випадковий ліс, LightGBM і логістична регресія), виникає проблема дисбалансу класів та «чорного ящика» – складність або неможливість пояснити, як саме модель прийняла певне рішення.

4.2. Алгоритмічна та програмна реалізація моделей прогнозування успішності здобувачів освіти

У попередньому розділі було обґрунтовано концептуальні та методичні засади побудови прогностичних моделей успішності здобувачів освіти. Наступним етапом є їх алгоритмічна формалізація та програмна реалізація, що забезпечує практичне застосування запропонованих підходів у межах інформаційної технології освітньої аналітики. Зазначимо, що програмна підтримка реалізації збору та підготовки даних, а також подальше моделювання задачі прогнозування успішності студентів здійснюється за допомогою мови програмування Python за розробленим алгоритмом (рис.4.2.) Окреслимо множину атрибутів, отриманих з обраних джерел даних.

Перше джерело даних (дані LMS Інформаційний портал НУБІП України) містить інформацію про відвідування й успішність студентів підрозділу, маючи наступні атрибути:

- Група – шифр групи студентів,
- КодОсоби – код студента,
- Студент – прізвище й ініціали студента,
- КодКурсу – код навчального курсу (дисципліни),
- Дисципліна – назва дисципліни, яку вивчає студент,
- Дата – дата проведення заняття за дисципліною відносно студента,
- Пара – час проведення заняття у форматі HH:MM:SS,
- Відмітка – відмітка про присутність студента на занятті (присутній або відсутній),
- pozm id – ідентифікаційний код студента,
- Стать – стать студента,
- Джерело фінансування – позначка про тип фінансування: бюджетне або контрактне,
- Курс – рік навчання студента,
- КБ при вступі – кількість балів при вступі в ВНЗ,
- Кількість входів на курс – загальна кількість входів студента на курс через платформу електронного навчання,
- Бали студента – фінальний бал у кінці вивчення дисципліни.

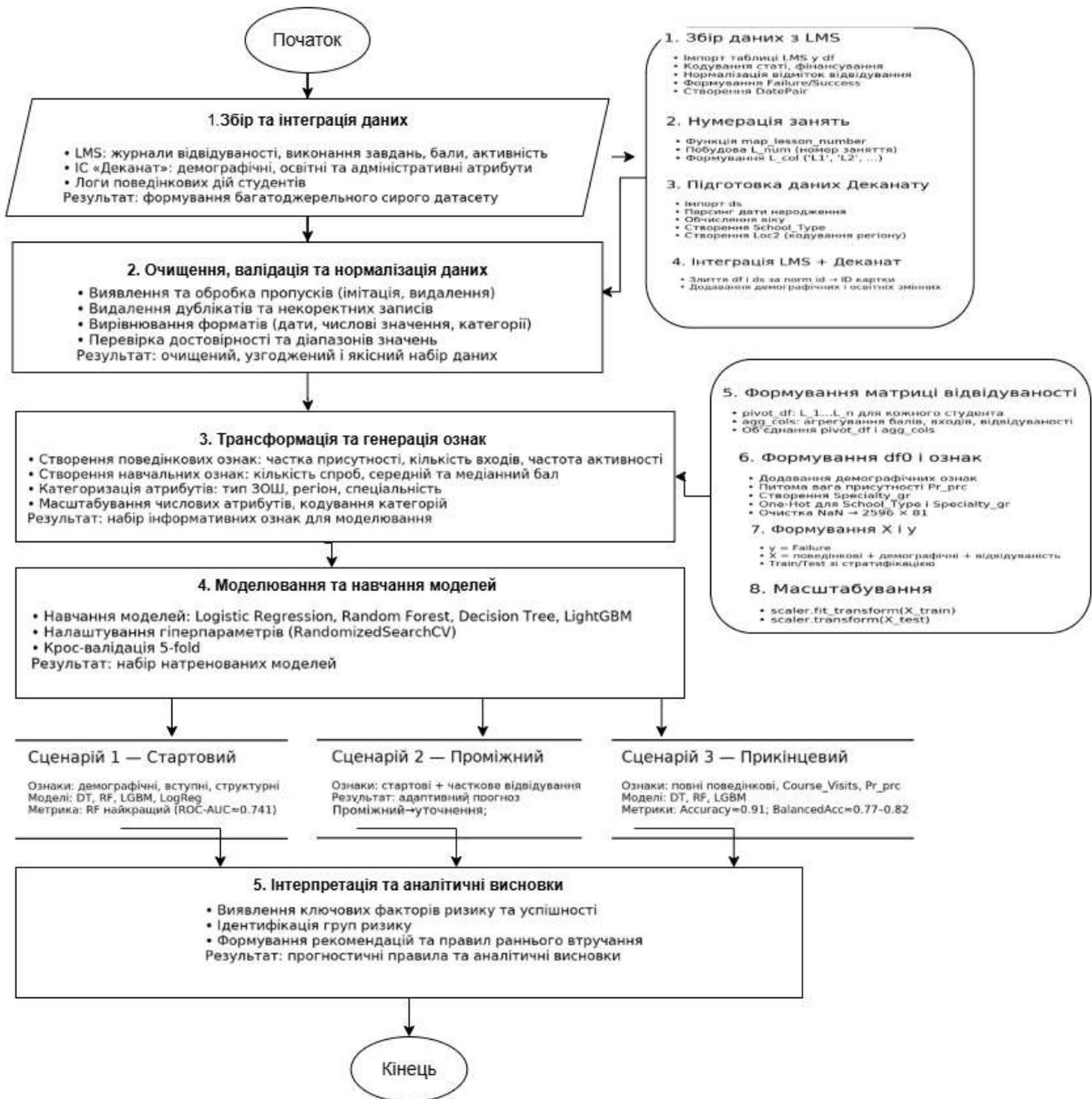


Рис. 4.2. Послідовність реалізації прогностичного моделювання успішності здобувачів освіти

Завантажені дані зберігаються в датафреймі під назвою df. Підготовка даних для подальшого моделювання передбачає визначення наступних змінних (Feature Engineering):

Gender – цілочисельна змінна, яка приймає значення 0, якщо стать жіноча, або 1, якщо – чоловіча:

$$Gender = \begin{cases} 0, & \text{якщо стать жіноча} \\ 1, & \text{якщо стать чоловіча} \end{cases}$$

Funding – цілочисельна змінна, яка також приймає два значення відповідно до джерела фінансування:

$$Funding = \begin{cases} 0, & \text{якщо бюджет} \\ 1, & \text{якщо контракт} \end{cases}$$

Check – цілочисельна змінна, що характеризує присутність студента на парі:

$$Check = \begin{cases} 0, & \text{відсутній} \\ 1, & \text{присутній} \end{cases}$$

Failure – булева змінна, яка сигналізує про неуспішне завершення навчальної дисципліни:

$$Failure = \begin{cases} 0, & \text{якщо} \geq 60 \text{ балів} \\ 1, & \text{якщо} < 60 \text{ балів} \end{cases}$$

Success – булева змінна, яка сигналізує про успішне завершення навчальної дисципліни:

$$Success = \begin{cases} 0, & \text{якщо} < 60 \text{ балів} \\ 1, & \text{якщо} \geq 60 \text{ балів} \end{cases}$$

В скрипті на мові програмування Python наведені вище змінні визначаються наступним чином:

```
df['Gender'] = LabelEncoder().fit_transform(df['Стать'])
df['Funding'] = LabelEncoder().fit_transform(df['Джерело фінансування'])
df["Відмітка"].unique()
    array(['absent', 'present', 'Present'], dtype=object)

df['Check'] = df['Відмітка'].map({'present': 1, 'Present':1, 'absent': 0})
df["Failure"] = (df["Бали студента"] < 60).astype(int)
df["Success"] = (df["Бали студента"] >= 60).astype(int)
```

Далі, вводимо нову змінну DatePair формату «дата-час» (YYYY-MM-DD HH:MM:SS), яка відображає дату й час запису про присутність студента на занятті, тобто об'єднуємо «Дата» і «Пара»:

```
df["DatePair"] = pd.to_datetime(df["Дата"].astype(str) + ' ' +
                                df["Папа"].astype(str))
```

Зважаючи на принцип постійного відстеження успішності студентів, прогнозування має здійснюватися в будь-який момент часу протягом навчального семестру і враховувати актуальні дані про поточний стан роботи студента. При такому підході необхідно мати унікальний номер кожного заняття в межах дисципліни для окремої групи:

```
def map_lesson_number(group):
    unique_dates = sorted(group['DatePair'].unique())
    mapping = {v: i+1 for i, v in enumerate(unique_dates)}
    return group['DatePair'].map(mapping)

# Застосовуємо .apply замість .transform
df['L_num'] = df.groupby(['Дисципліна', 'Група'],
                        group_keys=False).apply(map_lesson_number)
```

```
df['L_num'] = df['L_num'].astype(int)
```

Виконання цієї частини коду дасть дві змінні щодо номеру заняття:

- `L_num` – порядковий номер;
- `L_col` – текстова змінна, що складається з літери «L» і порядкового номеру.

Фрагмент сформованого масиву даних `df` зображено на рис.4.3.

Група	Кодробот	Студент	Найменування	Дисципліна	Дата	Пара	Відвітка	Семінар	Стать	Джерело фінансування	Курс	КІ при вступі	Кількість академічних років	Бали студента	leader	funding	Check	failure	Success	DatePair	L_num	L_col
0	04-210010	0892301	Ривален Д.	5104	2024-10-15	10:30:00	absent	0939361	Чоловік	Контракт	4	177.6	0	60	1	1	0	0	1	2024-10-15 10:30:00	1	L_1
1	04-210010	0892301	Ривален Д.	5104	2024-10-17	10:30:00	absent	0939361	Чоловік	Контракт	4	177.6	0	60	1	1	0	0	1	2024-10-17 10:30:00	2	L_2
2	04-210010	0892301	Ривален Д.	5104	2024-10-22	10:30:00	absent	0939361	Чоловік	Контракт	4	177.6	0	60	1	1	0	0	1	2024-10-22 10:30:00	3	L_3
3	04-210010	0892301	Ривален Д.	5104	2024-10-24	10:30:00	absent	0939361	Чоловік	Контракт	4	177.6	0	60	1	1	0	0	1	2024-10-24 10:30:00	4	L_4
4	04-210010	0892301	Ривален Д.	5104	2024-10-25	10:30:00	absent	0939361	Чоловік	Контракт	4	177.6	0	60	1	1	0	0	1	2024-10-25 10:30:00	5	L_5
40068	04-230020	0475001	Явор Р.	770	2024-10-10	12:10:00	absent	10939182	Чоловік	Бюджет	2	109.2	10	76	1	0	0	0	1	2024-10-10 12:10:00	10	L_10
40069	04-230020	0475001	Явор Р.	770	2024-11-01	12:10:00	absent	10939182	Чоловік	Бюджет	2	109.2	10	76	1	0	0	0	1	2024-11-01 12:10:00	10	L_10
40070	04-230020	0475001	Явор Р.	770	2024-11-05	12:10:00	absent	10939182	Чоловік	Бюджет	2	109.2	10	76	1	0	0	0	1	2024-11-05 12:10:00	10	L_10
40071	04-230020	0475001	Явор Р.	770	2024-09-28	15:30:00	absent	10939182	Чоловік	Бюджет	2	109.2	10	76	1	0	0	0	1	2024-09-28 15:30:00	9	L_9
40072	04-230020	0475001	Явор Р.	770	2024-10-12	12:10:00	absent	10939182	Чоловік	Бюджет	2	109.2	10	76	1	0	0	0	1	2024-10-12 12:10:00	10	L_10

Рис. 4.3. Фрагмент масиву даних `df`

Друге джерело (дані ІС Деканат) містить загальну інформацію про студентів підрозділу. Завантажені дані з цього джерела зберігаються в датафреймі

ds. Їх підготовка також включає процедуру визначення змінних (Feature Engineering):

1. Шифр спеціальності (Specialty). Перші три символи у стовпці «Спеціальність», і є цим шифром:

```
ds['Specialty'] = ds.Спеціальність.str[:3]
```

2. Дата народження (Birth_date) і вік (Age) студента:

```
ukr_months = {'січня': 1, 'лютого': 2, 'березня': 3, 'квітня': 4, 'травня': 5, 'червня': 6, 'липня': 7, 'серпня': 8, 'вересня': 9, 'жовтня': 10, 'листопада': 11, 'грудня': 12}
```

```
def parse_ukr_date(date_str):
    try:
        day, month, year = date_str.replace(' п.', '').split()
        return pd.Timestamp(year=int(year), month=ukr_months[month],
day=int(day))
    except:
        return pd.NaT
```

```
ds['Birth_date'] = ds['Дата народження'].apply(parse_ukr_date)
```

```
ds['Age'] = (pd.to_datetime("2024-09-01").year - ds['Birth_date'].dt.year)
```

В останньому рядку 2024-09-01 – це дата початку семестру, хоча замість нього може бути дата прогнозування в межах семестру.

3. Тип освітнього закладу (School_Type). Стовпець «Школа» у вихідних даних містить повну назву попереднього освітнього закладу, в якому навчався студент.

Нехай, S – множина всіх можливих текстових значень у стовпці “Школа”. Тоді, для визначення змінної School_Type застосовується функція перетворення категоріальної змінної в числову:

$$F: S \rightarrow \mathbb{N} \cup \text{NaN}, \quad (4.1)$$

де \mathbb{N} – множина натуральних чисел;

NaN – значення “не число” (для відсутніх даних).

Ця функція F є композитною, що складається з процедур екстракції, категоризації й перетворення на числові значення:

$$F(s) = f_{\text{encode}} \left(f_{\text{replace}} \left(f_{\text{extract}}(s) \right) \right), s \in S. \quad (4.2)$$

Функція $f_{\text{extract}}(s)$ використовує перелік ключових слів, за якими можна ідентифікувати коротку назву освітнього закладу, наприклад, центр, школа, ліцей, гімназія, коледж тощо. Вона витягує з тексту перший вираз, який відповідає одному з визначених шаблонів. Функція $f_{\text{replace}}(\cdot)$ замінює вилучений текстовий фрагмент на значення зі словника “mapping” (“ЗОШ”, “ПЗОШ”, “ПТО” тощо), створеного на думку автора / деканату. Нарешті, $f_{\text{encode}}(\cdot)$ замінює категоріальні значення на числові значення відповідно до заданого словника.

Нижче наведено відповідний Python-код:

```
ds['School_Type'] = ds['Школа'].str.extract(
r"(центр|школа|ліцей| ... )
, flags=re.IGNORECASE
)
mapping = {
    'загальноосвітня школа': 'ЗОШ',
    'зош i-iii': 'ЗОШ',
    'школа': 'ЗОШ',
    ... # приховано
}

ds['School_Type'] = ds['School_Type'].replace(mapping)
ds['School_Type'] = ds['School_Type'].map({'ЗОШ':1, 'ПЗОШ':2, 'ПТО':3,
    'ФПВО':4, 'ВО':5, 'Центр':6, 'Інше':7, 'Зарубіжжя':8})
```

4. Регіон (Loc). На підставі того ж стовпця «Школа» вилучаємо регіон, в якому ця школа розташована: Київ, Донецька область, Запорізька область, Київська область, Полтавська область, Чернігівська область тощо. Перетворення категоріальних змінних в числові значення дає нову змінну Loc2:

```
ds['Loc2'] = LabelEncoder().fit_transform(ds['Loc'])
```

Виконання зазначених вище кроків дозволило сформувати масив даних, фрагмент якого проілюстровано на рис. 4.4.

Спеціальність	Explanation	Освітня програма	Profesio	Курс	Група	Попередній заклад освіти	Документ про здобуття освіти	Інформація про попереднє навчання	КБ при вступі	КФ без отриманого КБ	Семестр	Discipline	Mark	Вказ	Specialty	Birth_date	Age	School_Type	Loc	Loc2
122 Кваліфікаційні науки	NaN	Кваліфікаційні науки	NaN	1.0	NaN	NaN	Скорочено про здобуття повної загальної освіти...	NaN	147.615	16	1.0	Дисципліна проносил на етапі отримання	81.8	План №227 м. Київ	122	2007-07-18	17	2	Київ	48
122 Кваліфікаційні науки	NaN	Кваліфікаційні науки	NaN	1.0	NaN	NaN	Скорочено про здобуття повної загальної освіти...	NaN	147.615	16	1.0	Філософія	79.8	План №227 м. Київ	122	2007-07-18	17	2	Київ	48
122 Кваліфікаційні науки	NaN	Кваліфікаційні науки	NaN	2.0	КН-236626	NaN	Скорочено про здобуття повної загальної освіти...	NaN	139.182	16	1.0	Дисципліна проносил на етапі отримання	65.8	Тристоронній план Житомирської області	122	2006-08-03	18	2	Житомирська область	34
122 Кваліфікаційні науки	NaN	Кваліфікаційні науки	NaN	2.0	КН-236626	NaN	Скорочено про здобуття повної загальної освіти...	NaN	139.182	16	1.0	Філософія	68.8	Тристоронній план Житомирської області	122	2006-08-03	18	2	Житомирська область	34
122 Кваліфікаційні науки	NaN	Кваліфікаційні науки	NaN	2.0	КН-236626	NaN	Скорочено про здобуття повної загальної освіти...	NaN	139.182	16	2.0	Виклад матеріалу	3.8	Тристоронній план Житомирської області	122	2006-08-03	18	2	Житомирська область	34

Рис. 4.4. Фрагмент масиву даних ds

Наступним кроком є об'єднання отриманих даних шляхом додавання до датафрейму df обраних колонок з датафрейму ds:

```
df = df.merge(ds[['ID картки', 'ID ФО', 'Здобувач', 'Birth_date', 'Age',
                'Вступ на основі', 'Форма навчання', 'Чи здобувався ступень за іншою спеціальністю', 'Чи скорочений термін навчання', 'Specialty', 'Освітня програма', 'Попередній заклад освіти', 'Школа', 'School_Type', 'Loc', 'Loc2']],
             how='left', left_on='norm id', right_on='ID картки')
```

Отже, отримали загальний датафрейм df, який зберігає всю інформацію про студента, що була на момент вступу до ЗВО та накопичена протягом поточного навчального семестру.

Після цього здійснюється групування даних по дисциплінах і студентах з відомостями про відвідування:

```
pivot_df = df.drop_duplicates().pivot_table(
    index=['КодКурсу', 'КодОсоби'],
    columns='L_col',
    values='Check',
    fill_value=0,
    aggfunc='max' # якщо випадково дублювання – беремо максимум
).reset_index()
```

Рис. 4.5 демонструє фрагмент датафрейму pivot_df, який містить відомості про відвідування студентом кожної дисципліни. Упорядковані заняття рознесені

по колонках L_1, L_2, L_3 і так далі. Цифра 1 в цих колонках говорить про присутність студента на занятті, тоді як 0 свідчить про його відсутність.

id	Назва курсу	Назва	L_1	L_2	L_3	L_4	L_5	L_6	L_7	L_8	L_9	L_10	L_11	L_12	L_13	L_14	L_15	L_16	L_17	L_18	L_19	L_20	L_21	L_22	L_23	L_24	L_25	L_26	L_27
0	210	888888	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	210	730034	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	210	730091	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
3	210	730490	1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	210	730114	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
...
2905	5475	730204	0	1	0	0	1	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2906	5475	730034	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2907	5475	740571	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2908	5475	7467107	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2909	5475	740888	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Рис. 4.5.Фрагмент датафрейму pivot_df

Код нижче дозволяє отримати агреговані дані про навчання студента, уникаючи дублюючих рядків, зокрема, загальний бал за вивчення дисципліни, успішне або неуспішне проходження дисципліни, кількість входів на електронну версію дисципліни, кількість відвіданих занять.

```
agg_cols = df.drop_duplicates().groupby(['КодКурсу', 'КодОсоби',
    'Дисципліна', 'Студент', 'Gender', 'Курс', 'Funding', 'КБ при
    вступі']).agg({
    'Бали студента' : 'max',
    'Failure' : 'max',
    'Success' : 'max',
    'Кількість входів на курс' : 'max',
    'Check' : 'sum',
    'Відмітка' : 'count'
}).reset_index()
```

Фрагмент датафрейму agg_cols показано на рис. 4.6. Він містить відомості про дисципліни, які вивчали студенту підрозділу. Врахована фінальна кількість балів і відповідний результат вивчення дисциплін – успіх (Success) або неуспіх (Failure). Вказано кількість входів на електронний навчальний курс, кількість відвіданих занять «Check». При цьому колонка «Відмітка» тепер вказує на те, скільки всього було перевірок присутності студента за кожною з дисциплін. Крім того, враховані особисті дані такі, як рік навчання, стать, джерело фінансування, кількість балів при вступі.

КодКурсу	КодОсоби	Discipline	Студент	Gender	Курс	Passing	№ раз вступ	Важ студента	Failure	Success	Кількість проходів на курс	Check	Відмітка	
0	210	5300145	Прогнозування соціально-економічних процесів (...)	Левинський О.	1	4	1	173.500	98	0	1	90	18	20
1	210	7230534	Прогнозування соціально-економічних процесів (...)	Дядько Д.	1	4	1	138.700	60	0	1	10	2	20
2	210	7330051	Прогнозування соціально-економічних процесів (...)	Овчаренко М.	1	4	1	195.000	92	0	1	90	18	20
3	210	7354892	Прогнозування соціально-економічних процесів (...)	Перчик Д.	1	4	0	150.500	64	0	1	30	8	20
4	210	7369174	Прогнозування соціально-економічних процесів (...)	Павел Р.	1	4	1	130.702	64	0	1	35	7	20
2965	5475	7362084	Методи прогнозування	Науменко С.	0	4	0	145.700	100	0	1	50	10	18
2966	5475	7368344	Методи прогнозування	Слободанко М.	1	4	1	122.600	82	0	1	0	0	2
2967	5475	7435271	Методи прогнозування	Торшак М.	1	4	1	131.800	82	0	1	0	0	2
2968	5475	7457187	Методи прогнозування	Костун В.	1	4	1	122.400	0	1	0	0	0	2
2969	5475	7524850	Методи прогнозування	Мусиков Є.	1	4	1	114.800	0	1	0	0	0	2

Рис. 4.6. Фрагмент датафрейму `agg_cols`

На підставі злиття двох масивів `pivot_df` і `agg_cols` отримуємо `df0`:

```
df0 = pd.merge(agg_cols, pivot_df, on=['КодКурсу', 'КодОсоби'])
```

Додатково розраховано нову змінну – питому вагу присутності студента на заняттях з дисципліни (співвідношення кількості відвіданих занять до кількості перевірок (всього занять):

```
df0['Pr_prc'] = df0.Check/df0.Відмітка
```

Далі до `df0` додається низка колонок з `df`:

```
df0 = pd.merge(df0, df[['ID ФО', 'Specialty', 'Вступ на основі', 'Форма навчання', 'Чи здобувався ступень за іншою спеціальністю', 'Чи скорочений термін навчання', 'Birth_date', 'Age', 'School_Type', 'Loc', 'Loc2']], left_on='КодОсоби', right_on='ID ФО', how='left')
```

В процесі обробки даних варто передбачити виокремлення спеціальностей, які не відносяться до підрозділу або галузі, наприклад, для факультету інформаційних технологій всі спеціальності крім 051, 122, 123, 125 і 126 позначаються 'no_FIT', а далі за потреби – фіктивним шифром 999:

```
df0['Specialty_gr'] = np.where(~np.isin(df0['Specialty'], [51, 121, 122, 123, 125, 126]), 'no_FIT', df0['Specialty'])
df0.loc[df0.Specialty_gr=='no_FIT', 'Specialty_gr'] = 999
df0.Specialty_gr = df0.Specialty_gr.astype(int)
```

Рівень попередньої освіти (Pre_Degree) є категоріальною змінною. Для перевірки його впливу на успішність навчання студента під час розробки моделі прогнозування слід конвертувати у числовий тип, де 1 позначає повну загальну середню освіту, 2 – молодший спеціаліст, 3 – бакалавр:

```
df0['Pre_Degree'] = df0['Pre_Degree'].replace({
    'Повна загальна середня освіта': '1',
    'Бакалавр': '3',
    'Молодший спеціаліст': '2'
}).astype(int)
```

Одним з можливих шляхів підвищення точності моделі прогнозування успішності навчання студентів може бути One-Hot кодування таких змінних як School_Type і Specialty_gr:

```
df0_encoded = df0.copy()

# Створюємо OneHot колонки окремо
school_dummies = pd.get_dummies(df0_encoded['School_Type'], prefix='School',
                                dtype=int)
spec_dummies = pd.get_dummies(df0['Specialty_gr'], prefix='Spec', dtype=int)

df0 = pd.concat([df0_encoded, school_dummies, spec_dummies], axis=1)
```

```
del df0_encoded
```

Задля спрощення процесу моделювання деякі колонки було перейменовано:

```
df0.rename(columns={
    'Курс': 'Year',
    'КБ при вступі': 'Admission_Points',
    'Бали студента': 'Score',
    'Кількість входів на курс': 'Course_Visits',
    'Відмітка': 'N_Classes',
    'Вступ на основі': 'Pre_Degree',
    'Форма навчання': 'Form',
    'Чи здобувався ступень за іншою спеціальністю': 'Another_Spec',
    'Чи скорочений термін навчання': 'Short'},
    inplace=True)
```

Отже, запропоновані процедури збору та попередньої обробки даних за 1-й семестр 2024-2025 року навчання на факультеті інформаційних технологій формують масив даних df0 розмірністю 2790 рядків і 81 стовпців:

```
Index(['КодКурсу', 'КодОсоби', 'Дисципліна', 'Студент', 'Gender', 'Year',
      'Funding', 'Admission_Points', 'Score', 'Failure', 'Success',
      'Course_Visits', 'Check', 'N_Classes', 'Pr_prc', 'L_1', 'L_2', 'L_3',
      'L_4', 'L_5', 'L_6', 'L_7', 'L_8', 'L_9', 'L_10', 'L_11', 'L_12',
      'L_13', 'L_14', 'L_15', 'L_16', 'L_17', 'L_18', 'L_19', 'L_20',
      'L_21', 'L_22', 'L_23', 'L_24', 'L_25', 'L_26', 'L_27', 'L_28',
      'L_29', 'L_30', 'L_31', 'L_32', 'L_33', 'L_34', 'L_35', 'L_36',
      'L_37', 'L_38', 'L_39', 'L_40', 'ID ФО', 'Specialty', 'Pre_Degree',
      'Form', 'Another_Spec', 'Short', 'Birth_date', 'Age', 'School_Type',
      'Loc', 'Loc2', 'Specialty_gr', 'School_1', 'School_2', 'School_3',
      'School_4', 'School_5', 'School_6', 'School_7', 'Spec_121',
      'Spec_122', 'Spec_123', 'Spec_125', 'Spec_126', 'Spec_51',
      'Spec_no_FIT'],
      dtype='object')
```

Остаточним кроком формування df0 є фільтрація від рядків з пропусками, в результаті якої кількість рядків скоротилась на 374 і склала 2596.

Описова статистика. Частка студентів, які мають хоча б один борг, тобто принаймні за 1 дисципліною набрали менше 60 балів, складає 21,5% (табл. 4.1).

Таблиця 4.1

Розподіл студентів за успішністю

Успіх / неуспіх	Кількість осіб	Частка, %
Мають хоча б 1 борг (< 60 б.)	215	21,5
Не мають боргів	785	78,5
УСЬОГО	1000	100

Кількість дисциплін, за якими хоча б один студент має менше 60 балів складає 44 з 79 дисциплін, тобто 56%. В табл. 4.2. наведено розподіл 1 тис. студентів по спеціальностям, які вивчали дисципліни на факультеті інформаційних технологій.

Таблиця 4.2

Розподіл студентів за спеціальностями

Шифр спеціальності	Кількість осіб	Частка, %
121	318	31.8
122	279	27.9
125	121	12.1
123	120	12.0
051	84	8.4
126	67	6.7
174	4	0.4

73	2	0.2
201	2	0.2
61	1	0.1
71	1	0.1
181	1	0.1
Разом	1000	100

Прогнозування успішності навчання здобувачів освіти за дисциплінами протягом семестру може здійснюватися на підставі трьох підходів, що відрізняються за часовою ознакою реалізації моделі прогнозування:

«Стартовий» – момент прогнозування t^{Pred} стоїть раніше початку вивчення дисципліни t^{Start} на $\Delta t'_1$ або на початку, або після нього протягом $\Delta t'_2$:
 $t^{Pred} \in [t^{Start} - \Delta t'_1, t^{Start} + \Delta t'_2], \forall 0 \leq \Delta t'_1 \leq T'1 \wedge 0 \leq \Delta t'_2 \leq T'2.$ (4.3)

«Проміжний» – момент прогнозу t^{Pred} обирається в будь який момент протягом семестру, але до його фінальної частини:

$$t^{Pred} \in (t^{Start} + \Delta t'_2, t^{Fin} - \Delta t'_3], \forall 0 \leq \Delta t'_2 \leq T'2 \wedge 0 \leq \Delta t'_3 \leq T'3. \quad (4.4)$$

«Прикінцевий» – прогноз здійснюється в момент t^{Pred} , близький до підсумкового контролю:

$$t^{Pred} \in (t^{Fin} - \Delta t'_3, t^{Fin}], \forall 0 \leq \Delta t'_3 \leq T'3. \quad (4.5)$$

Розробка моделі прогнозу ґрунтується на засадах класифікації, як одного з розділів статистичного аналізу та машинного навчання.

Цільовою змінною в моделі прогнозування на основі класифікації обрано ознаку неуспішного завершення курсу (Failure). Звідси витікає започаткування бінарної задачі класифікації щодо завчасного виявлення неуспіху серед студентів з вивчення дисциплін:

```
y = df0['Failure']
```

Після вибору ознак класифікації здійснюється розподіл на тренувальний і тестовий набори даних, 70% і 30% відповідно:

```
X_train, X_test, y_train, y_test = train_test_split(X_df, y, test_size=0.3,
random_state=1234, stratify=y)
```

Такий розподіл дає наступні набори даних в розрізі успішності навчання студентів:

```
Number of unsuccesed in train set: 192
Number of   succeeded in train set: 1625
Number of unsuccesed in test  set: 82
Number of   succeeded in test  set: 697
```

При цьому застосування аргументу `stratify=y` гарантує, що співвідношення класів у тренувальній і тестовій вибірках буде таким самим, як і у вихідних даних. Це доречно для досліджуваної вибірки з малою часткою позитивних результатів, де співвідношення класів “1” і “0” складає майже 1:5.

Обробка цих наборів відбувається завдяки наступному трансформеру `ColumnTransformer()`, де `scaler` – це спосіб масштабування даних, зокрема метод нормалізації “minmax” або стандартизації “standard”:

```
preprocessor = ColumnTransformer(
    transformers=[
        ('num', scaler, colnames)
    ],
    remainder='drop')
```

В набір моделей включені Древа рішень, Випадковий ліс, `LightGBM` (легка градієтна модель підсилення) і Логістична регресія:

```
models =
{
    "DecisionTree": DecisionTreeClassifier(),
    "RandomForest": RandomForestClassifier(),
    "LightGBM":     LGBMClassifier(),
    "LogReg":       LogisticRegression()
}
```

Перетворення даних і реалізація алгоритму класифікації містяться в `Pipeline`:

```
for name, clf in models.items():
    pipe = Pipeline([('prep', preprocessor),
                    ('clf',  clf)])
    pipe.fit(X_train, y_train)
```

Разом з тим можна скористатись методом автоматичного підбору гіперпараметрів для моделей машинного навчання `RandomizedSearchCV`. На

відміну від повного перебору (GridSearchCV), він випадково вибирає комбінації параметрів із заданого розподілу, що значно прискорює пошук (додаток В).

4.3. Апробація підходів прогнозування успішності студентів (на базі вибірки здобувачів освіти НУБІП України)

4.3.1. Сценарійний підхід 1 «Стартовий»

Виявлення можливого неуспіху студента за дисципліною на початку її вивчення. Вибрано наступні ознаки:

- 'Gender',
- 'Year',
- 'Funding',
- 'Admission_Points',
- 'Pre_Degree'
- 'Age',
- 'School_Type',
- ' Loc2',
- 'Specialty_gr'

В якості способу масштабування обрано MinMaxScaler(). Результати реалізації методу автоматичного підбору гіперпараметрів для моделей машинного навчання RandomizedSearchCV з використанням метрики 'roc_auc' наведено в додатку В.

Порівняльний аналіз ефективності моделей машинного навчання за результатами, отриманими в рамках задачі бінарної класифікації задля виявлення неуспіху по вивченню дисциплін студентами здійснено з використанням методу 5-кратної перехресної перевірки (5-fold Cross-Validation) та метрики ROC-AUC ґрунтується на знайдених найкращих значеннях гіперпараметрів (базовий варіант).

Застосований метод крос-валідації передбачає розбиття набору тренувальних даних (x_{train} , y_{train}) на п'ять рівних частин (фолдів); модель послідовно навчається на чотирьох частинах і тестується на п'ятій. Процедура повторюється п'ять разів. В скрипті втілена стратифікована вибірка, яка

забезпечує збереження пропорцій класів у кожному фолді, що особливо важливо при роботі з незбалансованими наборами даних.

Результати порівняльного аналізу за CV метриками представлені у форматі: *Середнє значення (Mean) ± Стандартне відхилення (SD)*, де середнє значення характеризує загальну ефективність моделі, а стандартне відхилення – варіативність (стабільність) її продуктивності на різних підмножинах даних.

Крос-валідаційні (CV) метрики якості моделей на тренувальному наборі (X_{train} , y_{train}):

```
LightGBM 5-fold CV Results:
  ROC_AUC: 0.7373 ± 0.0291
    F1: 0.3233 ± 0.0197
PRECISION: 0.2131 ± 0.0131
  RECALL: 0.6726 ± 0.0586
ACCURACY: 0.7028 ± 0.0168

RandomForest 5-fold CV Results:
  ROC_AUC: 0.7407 ± 0.0258
    F1: 0.3267 ± 0.0482
PRECISION: 0.2960 ± 0.0395
  RECALL: 0.3650 ± 0.0621
ACCURACY: 0.8415 ± 0.0109

DecisionTree 5-fold CV Results:
  ROC_AUC: 0.6956 ± 0.0378
    F1: 0.3033 ± 0.0183
PRECISION: 0.1982 ± 0.0054
  RECALL: 0.6571 ± 0.1075
ACCURACY: 0.6841 ± 0.0266

LogReg 5-fold CV Results:
  ROC_AUC: 0.6433 ± 0.0501
    F1: 0.2445 ± 0.0295
PRECISION: 0.1502 ± 0.0173
  RECALL: 0.6617 ± 0.1020
ACCURACY: 0.5702 ± 0.0241
```

Ці значення метрик показують, наскільки добре кожна з моделей узагальнює дані, які вона "бачила", але у різних комбінаціях:

ROC-AUC (Area Under the Receiver Operating Characteristic Curve) є показником якості бінарної класифікації, що відображає здатність моделі розрізняти класи. Значення 0.5 відповідає випадковому вгадуванню, 1.0 – ідеальній класифікації.

Усі протестовані моделі демонструють ROC-AUC в діапазоні 0.643-0.741, що свідчить про їх помірну дискримінативну здатність: помітно кращі за випадкове вгадування, але далекі від високої точності. Найкращий результат показав Random Forest (0.7407 ± 0.0258), який також продемонстрував найвищу стабільність (найменше стандартне відхилення 0.0258), що свідчить про добру узагальнюючу здатність. Проте низький загальний рівень ROC-AUC сигналізує про необхідність подальшого вдосконалення моделей.

З практичного погляду моделі з ROC-AUC $\approx 0.64-0.74$ можуть бути прийнятним для задач попереднього скринінгу, коли головне – відсіяти очевидні випадки з подальшою експертною перевіркою, тобто знайти більшість ризикованих студентів без точного прогнозування, оскільки моделі дають багато хибних тривог. При цьому значення ROC-AUC > 0.7 вважаються задовільними для практичного застосування, особливо в гуманітарних науках, де поведінкові фактори мають значну варіабельність.

F1-Score – є компромісом між точністю (precision) та повнотою (recall), особливо важливим для незбалансованих даних. Низькі значення метрики для всіх моделей (< 0.33) підкреслюють складність прогнозування меншого класу «неуспіх». Висока стандартна відхилення для RandomForest вказує на нестабільність результатів між фолдами.

Precision (влучність) – оцінює частку правильно класифікованих позитивних прогнозів. Низькі значення метрики для всіх моделей свідчать про ризик надмірної кількості хибних спрацювань при прогнозуванні неуспіху. Це може бути прийнятним для скринінгу, коли важливіше не пропустити ризикованих студентів.

Recall (повнота) – визначає частку справжніх позитивних екземплярів, які модель правильно ідентифікувала.

Accuracy (точність) – частка правильно класифікованих прикладів. З огляду на інші метрики бачимо, що високі значення цієї метрики спричинені переважанням класу «успіх» у даних.

Метрики точності моделей машинного навчання щодо виявлення неуспіху вивчення дисципліни з використанням визначених гіперпараметрів для 1-го підходу «стартовий» на тестовій вибірці вказані в зведеній табл. 4.3.

Таблиця 4.3

**Тестові метрики реалізації моделей машинного навчання
для 1-го підходу («стартовий»)**

Модель	Accuracy	ROC_AUC	Balanced Accuracy	PR-AUC
LightGBM	0.6470	0.6862	0.6090	0.2002
RandomForest	0.8062	0.6857	0.6334	0.2343
DecisionTree	0.5995	0.6708	0.6309	0.2710
LogReg	0.5546	0.6169	0.5951	0.1566

Проаналізуємо метрики кожної моделі згідно з базовим варіантом:

- *Accuracy*: RandomForest демонструє найвищу точність – правильно класифікує 80% студентів, але це пояснюється переважанням класу «успіх» у даних;

- *ROC_AUC*: середня здатність моделей розрізнити класи “успіх” і “неуспіх” є нижчою за 0.7, що свідчить про їх помірну або слабку узагальнюючу здатність;

- *Balanced Accuracy*: Light GBM лідирує за середнім відсотком правильних прогнозів для кожного класу окремо (68,6%), що є більш надійною метрикою при дисбалансі класів;

- *PR-AUC* (площа під Precision-Recall кривою): для незбалансованих даних часто є більш інформативною метрикою, ніж ROC-AUC, а її значення ≤ 0.271 говорять про проблеми ідентифікації рідкісного класу “неуспіх”.

Доповнимо отриману інформацію класифікаційними звітами, враховуючи, що всі моделі працюють зі 697 екземплярами класу 0 (“успіх”) і 82 екземплярами класу 1 (“неуспіх”), тобо у співвідношенні $\sim 8.5:1$. Ці звіти містять метрики точності моделей машинного навчання з прогнозування неуспіху вивчення дисципліни, зокрема, precision, recall, f1-score і accuracy:

1. Decision Tree:

	precision	recall	f1-score	support
0	0.9385	0.5911	0.7254	697

1	0.1618	0.6707	0.2607	82
accuracy			0.5995	779
macro avg	0.5501	0.6309	0.4930	779
weighted avg	0.8567	0.5995	0.6764	779

Classification Report for Imbalanced Labels:

	pre	rec	spe	f1	geo	iba	sup
0	0.9385	0.5911	0.6707	0.7254	0.6297	0.3933	697
1	0.1618	0.6707	0.5911	0.2607	0.6297	0.3996	82
avg/total	0.8567	0.5995	0.6623	0.6764	0.6297	0.3940	779

Отже, для класу 1 маємо багато помилок першого роду: Recall=67%, що означає, що модель знаходить 2/3 студентів з “неуспіхом”, а Precision=16%, тобто 84% прогнозів «неуспіх» є хибними. Для класу 0 – багато помилок другого роду.

2. Random Forest:

	precision	recall	f1-score	support
0	0.9252	0.8522	0.8872	697
1	0.2482	0.4146	0.3105	82
accuracy			0.8062	779
macro avg	0.5867	0.6334	0.5989	779
weighted avg	0.8540	0.8062	0.8265	779

Classification Report for Imbalanced Labels:

	pre	rec	spe	f1	geo	iba	sup
0	0.9252	0.8522	0.4146	0.8872	0.5944	0.3688	697
1	0.2482	0.4146	0.8522	0.3105	0.5944	0.3379	82
avg/total	0.8540	0.8062	0.4607	0.8265	0.5944	0.3656	779

Отже, модель характеризується найвищими F1-score і balanced accuracy, але пропускає 58% справжніх неуспішних студентів (Recall).

3. Light GBM:

	precision	recall	f1-score	support
0	0.9271	0.6571	0.7691	697
1	0.1614	0.5610	0.2507	82
accuracy			0.6470	779
macro avg	0.5443	0.6090	0.5099	779
weighted avg	0.8465	0.6470	0.7145	779

Classification Report for Imbalanced Labels:

	pre	rec	spe	f1	geo	iba	sup
0	0.9271	0.6571	0.5610	0.7691	0.6071	0.3722	697
1	0.1614	0.5610	0.6571	0.2507	0.6071	0.3651	82
avg/total	0.8465	0.6470	0.5711	0.7145	0.6071	0.3714	779

Отже, модель має найкращу дискримінаційна здатність і потенціал для впізнавання класу 1, але – низький Precision і PR-AUC.

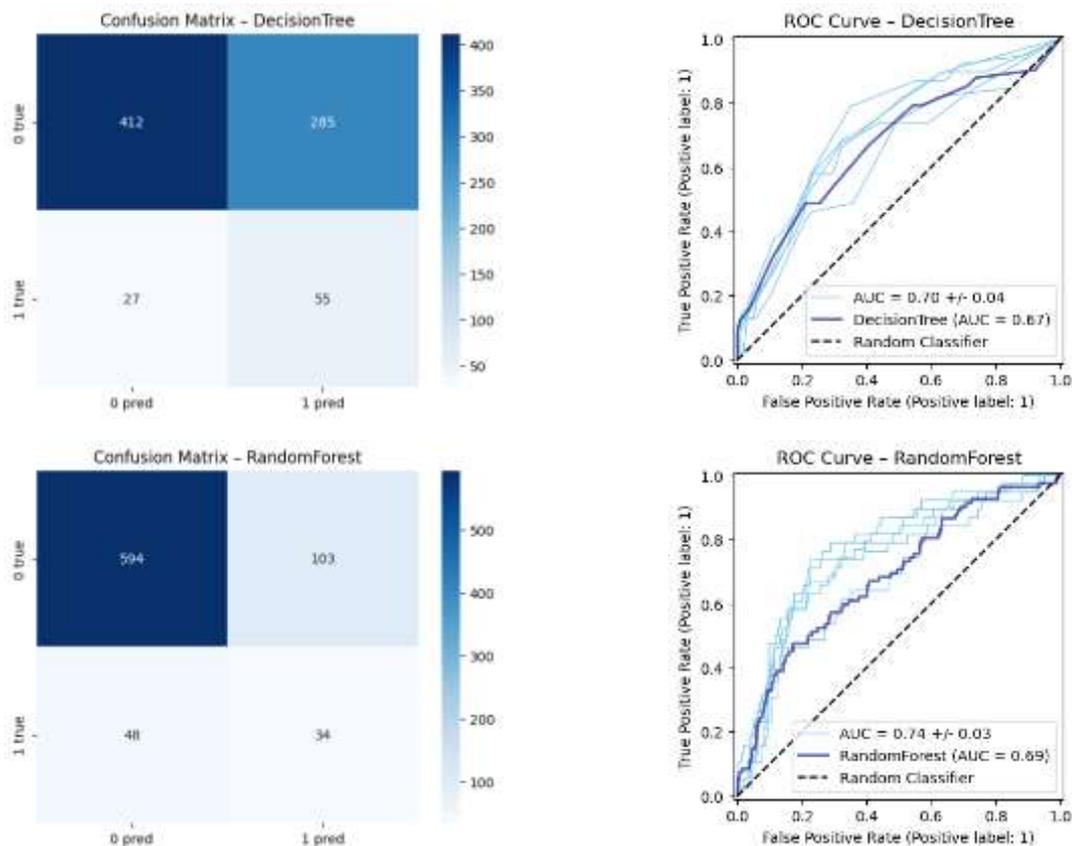
4. Логістична регресія (LogReg):

	precision	recall	f1-score	support
0	0.9289	0.5438	0.6860	697
1	0.1429	0.6463	0.2340	82
accuracy			0.5546	779
macro avg	0.5359	0.5951	0.4600	779
weighted avg	0.8462	0.5546	0.6384	779

Classification Report for Imbalanced Labels:

	pre	rec	spe	f1	geo	iba	sup
0	0.9289	0.5438	0.6463	0.6860	0.5928	0.3478	697
1	0.1429	0.6463	0.5438	0.2340	0.5928	0.3551	82
avg/total	0.8462	0.5546	0.6355	0.6384	0.5928	0.3486	779

Отже, можна зробити висновок про непридатність моделі для задачі класифікації через нездатність вловити нелінійну залежність. Матриця плутанини (помилки) для кожної моделі класифікації, спрямованої на виявлення неуспіху проходження дисциплін на його початку, а також відповідний графік ROC показані на рис. 4.7.



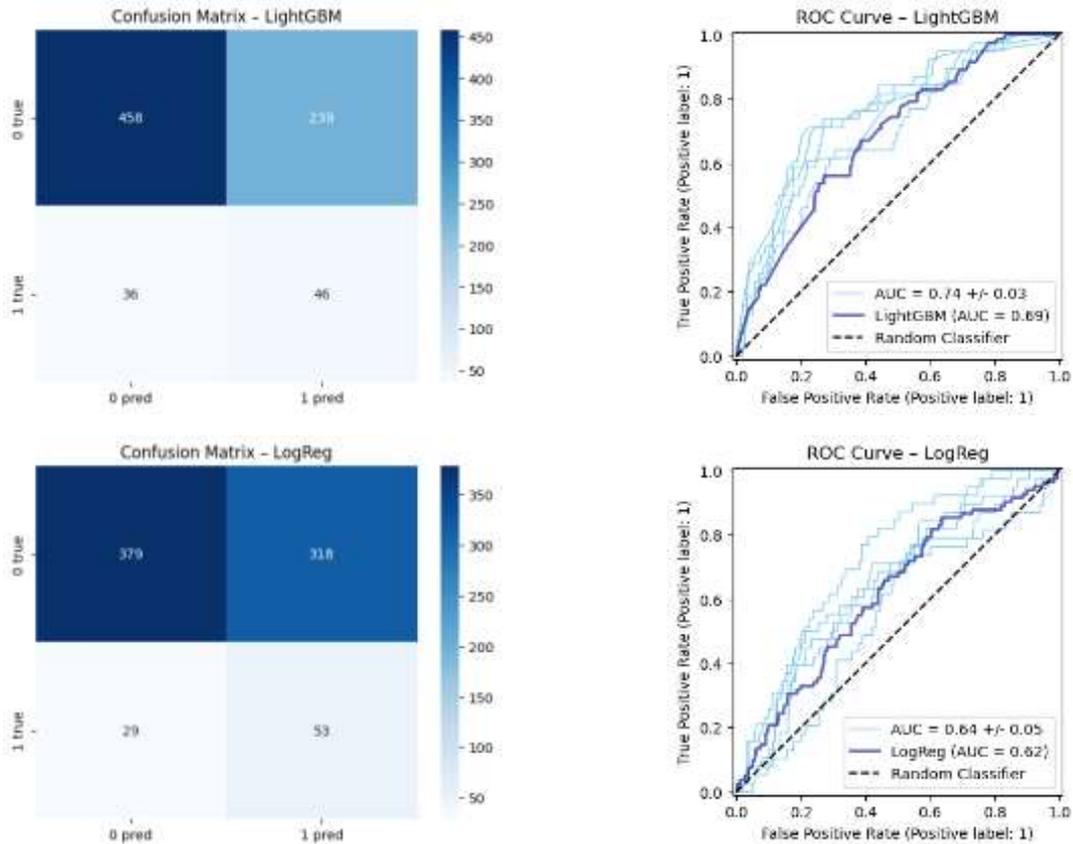


Рис. 4.7. Матриці плутанини (confusion matrix) для 1-го підходу («стартовий»)
 На рис. 4.8. наведені графіки «Precision-Recall» для всіх моделей.

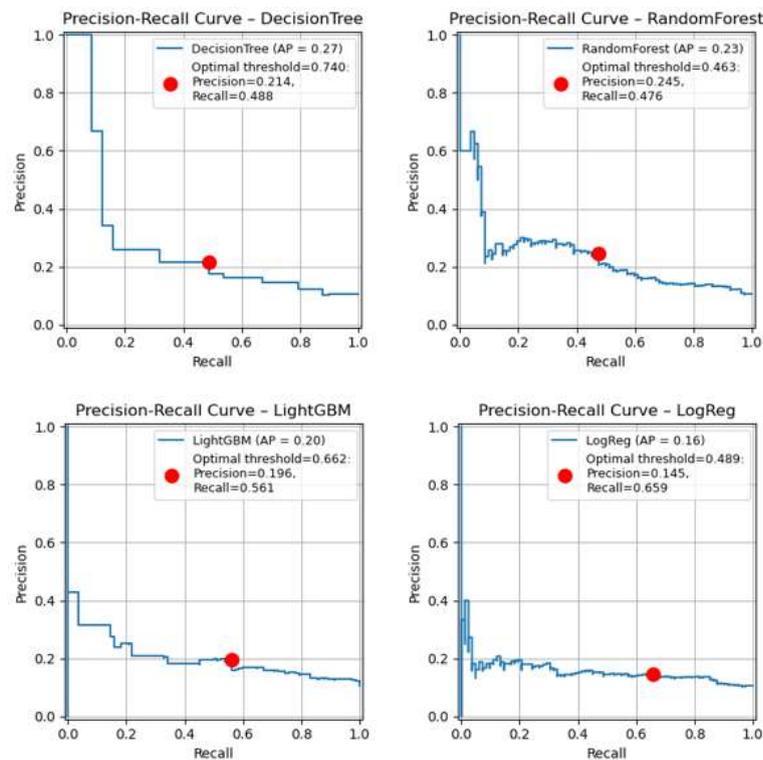


Рис. 4.8. Графіки «Precision-Recall» (підхід 1 – «стартовий»)

Про важливість ознак згідно отриманих результатів реалізації моделей свідчать діаграми на рис. 4.9. Цей метод показує, які ознаки модель найчастіше використовувала для побудови розщеплень під час тренування.

Як видно з рис. 4.9, найбільший внесок в результати прогнозування надає з 9 ознак «Admission_Points» (кількість балів при вступі). Далі до групи найважливіших ознак відносяться «Specialty_gr» (спеціальність, на якій навчається студент), його вік «Age» і регіон «Loc2», де попередньо навчався.

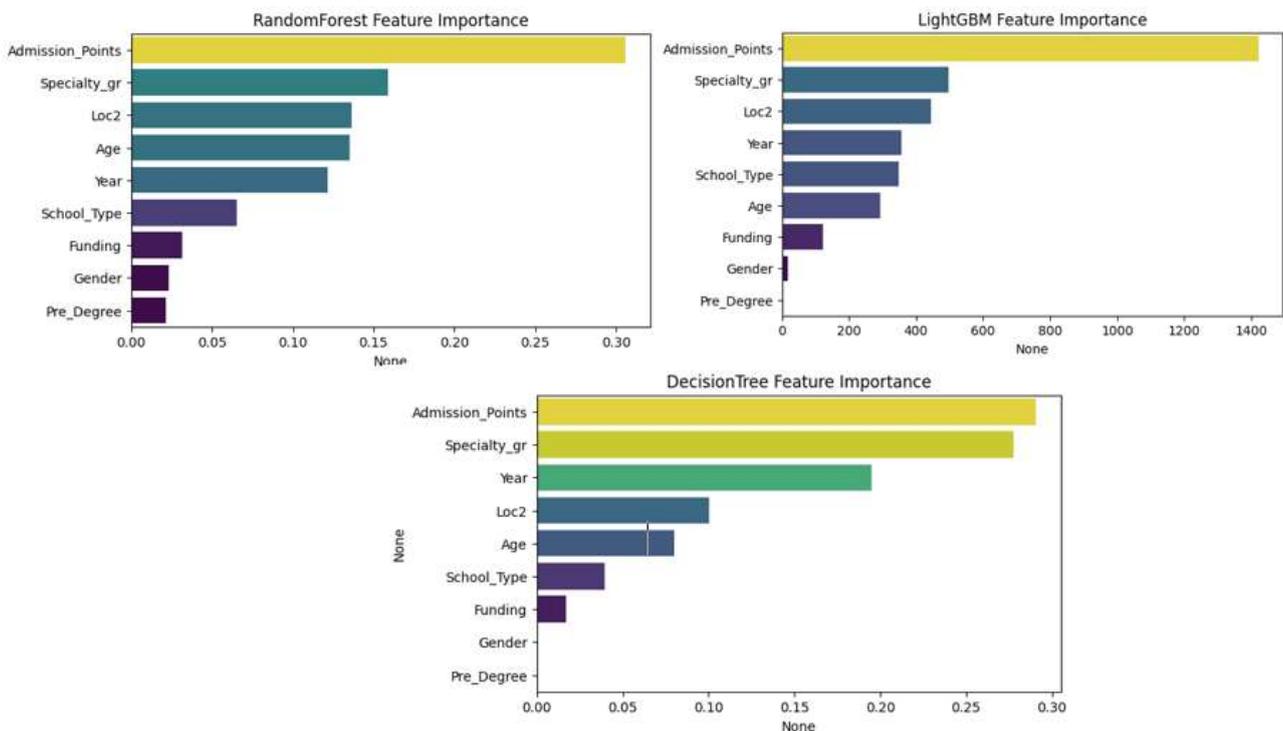


Рис. 4.9. Важливість ознак за моделями машинного навчання для 1-го підходу («стартовий»)

Доповнимо інформацію про важливість ознак за кожною моделлю пермутаційною важливістю (рис. 4.10). На графіку зображено "Mean AUC Decrease" (середнє зниження AUC) – на скільки в середньому погіршилася якість моделі (ROC-AUC) після "зіпсування" конкретної ознаки. Чорні лінії навколо стовпчиків – це довірчі інтервали (варіативність за результатами 10 повторних ермутацій). Метод пермутаційної важливості ознак оцінює, наскільки сильно просідає якість моделі, якщо «перемішати» (random shuffle) значення певної ознаки у тестовому наборі:

- якщо після перемішування модель втрачає точність → ознака була важливою для прогнозування;
- якщо ж зміна майже не впливає на якість → ознака не критично впливає на рішення моделі;
- в окремих випадках з'являються і негативні значення: це означає, що "зашумлена" ознака навіть випадково покращувала роботу моделі (ознака скоріше заважає, ніж допомагає).

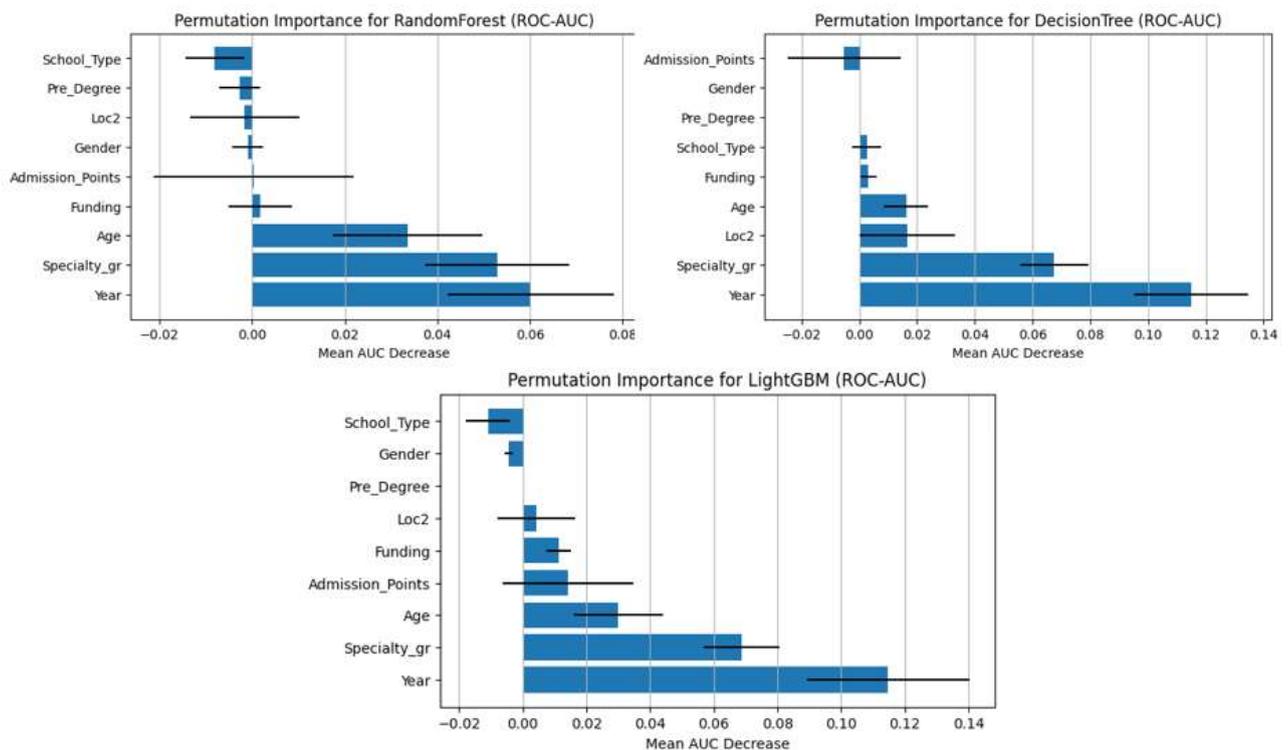


Рис. 4.10. Пермутаційна важливість ознак

При цьому процедура отримання результатів була наступною:

1. Базова оцінка: модель оцінюється на тестовому наборі даних (X_{test} , y_{test}) з використанням метрики `roc_auc`.
2. Пермутація: Для кожної ознаки окремо її значення у тестовому наборі випадковим чином перемішуються. Це руйнує зв'язок між цією ознакою та цільовою змінною.
3. Повторна оцінка: Модель знову оцінюється на модифікованих даних.

4. Обчислення важливості: Важливість ознаки визначається як різниця між базовою ROC-AUC та ROC-AUC після перемішування.

5. Усереднення: Процес повторюється 10 разів ($n_repeats=10$) для кожної ознаки, щоб отримати стабільну оцінку (середнє значення та стандартне відхилення).

Проведемо інтерпретацію графіку на прикладі Випадкового лісу. Year, Specialty_gr, Age – найбільш значущі ознаки, адже їх перемішування знижує ROC-AUC у середньому на 0.04-0.06. Це говорить, що модель в основному спирається саме на ці ознаки. Вузькі "вуса" (довірчі інтервали) для Year та Specialty_gr свідчать про стабільність цього висновку. Admission_Points та Funding мають середнє зниження AUC близьке до нуля. Широкий довірчий інтервал для Admission_Points (від'ємні та додатні значення) вказує на те, що вплив цієї ознаки є нестабільним або вона сильно корелює з іншими важливими ознаками. Решта ознак ("Funding", "Gender", "School_Type" тощо), судячи з усього, не несуть значущої прогностичної інформації для моделі. Розбіжність між методами оцінки важливості, зазвичай, пов'язана з тим, що Feature Importance схильний завищувати значущість ознак з більшою кількістю унікальних значень ("Admission_Points"), оскільки такі ознаки статистично частіше стають "вигідними для спліту" на тренувальному наборі даних, тоді як Permutation Importance показує, наскільки спотворюється модель на нових даних після "зашумлення" ознаки на тестовому наборі даних. Усунення неважливих ознак призводить до модифікації базового варіанту, реалізація якої характеризується приблизно такими ж значеннями метрик моделей в табл. 4.4.

Таблиця 4.4

Тестові метрики реалізації моделей машинного навчання для 1-го підходу («стартовий»)

Модель	Accuracy	ROC_AUC	Balanced Accuracy	PR-AUC
LightGBM	0.6701	0.6866	0.6381	0.2056
RandomForest	0.7458	0.6864	0.6589	0.2315
DecisionTree	0.6829	0.6754	0.6291	0.2705
LogReg	0.6149	0.6632	0.6126	0.2023

Зазначимо, що використання моделей можливе у випадку, коли потрібно встановити, чи зможе конкретний студент пройти певну дисципліну, отримуючи відповідні прогнози (додаток В).

Отже, в рамках 1-го підходу “стартовий” найкраще використовувати RandomForest і DecisionTree задля попереднього скринінгу з подальшою експертною перевіркою, а LightGBM – для задач ранжування студентів за ризиком. При цьому при оцінці моделей фокус варто робити на F1-score, balanced accuracy та PR-AUC, а також дивитись на confusion matrix для розуміння характеру помилок. Для створення більш ефективної системи прогнозування необхідно збільшити обсяг даних і покращити якість ознак.

4.3.2. Сценарійний підхід 2 «Проміжний».

До низки ознак, що надаються моделі за цим підходом, можуть належати ті ж ознаки, що використовуються перед початком навчального семестру, а також відвідування певної кількості занять. Для досліджуваного періоду обрано 10 занять. Тому *1-й варіант* переліку ознак містить:

- 'Gender'
- 'Year'
- 'Funding'
- 'Admission_Points'
- 'Pre_Degree'
- 'Age'
- 'School_Type'
- 'Loc2'
- 'Specialty_gr'
- 'L_1', 'L_2', 'L_3', 'L_4', 'L_5', 'L_6', 'L_7', 'L_8', 'L_9', 'L_10'

Гіперпараметри моделей згідно підходу 2 «проміжний» і 1-го варіанту ознак визначені наступним чином:

```
"DecisionTree": DecisionTreeClassifier(
    criterion = "gini"
    , max_depth = 10
    , splitter="random"
    , min_samples_split = 2
    , min_samples_leaf = 6
    , min_weight_fraction_leaf = 0.0
    , max_features = None
    , class_weight = "balanced"
```

```

    , random_state = 1234
    ),

"RandomForest": RandomForestClassifier(
    criterion = "gini"
    , n_estimators = 2000
    , class_weight = "balanced"
    , min_samples_split = 4
    , min_samples_leaf = 6
    , max_features = "sqrt"
    , bootstrap = False
    , random_state=1234
    , n_jobs=-1
    ),

"LightGBM": LGBMClassifier(
    n_estimators = 10000
    , learning_rate = 0.007
    , max_depth = 10
    , class_weight = 'balanced'
    , num_leaves = 31
    , random_state=1234
    , n_jobs=-1
    , boosting_type = 'dart'
    ),

"LogReg": LogisticRegression(
    max_iter = 1000
    , solver = 'liblinear'
    , class_weight='balanced'
    , random_state = 1234
    , n_jobs=-1
    , penalty='l2'
    , C = np.float64(0.1)
    )
}

```

Судячи з метрик точності моделей за 2-м підходом “проміжний” і 1-м варіантом переліку, наведених в табл. 4.5, можна сказати, що на тестовому наборі даних найкраще відпрацювали Light GBM і Random Forest для базового типу прогнозування («Basic»). В таблиці також наведено метрики за модифікованим типом прогнозування – з оптимізованим порогом Threshold на основі Precision-Recall кривої, що максимізує F1-score для рідкісного класу 1 – «неуспіх» («Opt-F1»).

Таблиця 4.5

Метрики точності моделей за 2-м підходом “проміжний” і 1-м варіантом переліку ознак

Характеристики	Моделі							
	Decision Tree		Light GBM		Log Reg		Random Forest	
Тип прогнозу	Basic	Opt-F1	Basic	Opt-F1	Basic	Opt-F1	Basic	Opt-F1
Threshold	0.5000	0.9351	0.5000	0.3423	0.5000	0.7821	0.5000	0.6015
Accuracy	0.7805	0.9140	0.9101	0.9101	0.7510	0.8703	0.9012	0.9217
Balanced Accuracy	0.7751	0.7744	0.7722	0.7991	0.7747	0.7608	0.7887	0.8002
F1_Score	0.4242	0.5939	0.5833	0.6067	0.4049	0.5025	0.5792	0.6347
Precision	0.2930	0.5904	0.5698	0.5625	0.2705	0.4215	0.5248	0.6235
Recall	0.7683	0.5976	0.5976	0.6585	0.8049	0.6220	0.6463	0.6463
ROC_AUC	0.8273		0.9048		0.8351		0.8945	
PR_AUC	0.4682		0.6548		0.4184		0.6334	
CV_ROC_AUC_mean	0.8179		0.8588		0.8336		0.8842	
CV_ROC_AUC_std	0.0273		0.0151		0.0293		0.0156	
CV_F1_mean	0.4308		0.5821		0.3676		0.5974	
CV_F1_std	0.0264		0.0446		0.0179		0.0307	

В табл. 4.6. знаходяться метрики продуктивності моделей за другим модифікованим типом прогнозування «Opt-TR» – з порогом для класу 1, що забезпечує цільовий рівень Recall у розмірі 0.75.

Таблиця 4.6

Метрики продуктивності моделей з визначеним під цільовий рівень Recall порогом для класу 1 (тип «Opt-TR») згідно з підходом 2 і варіантом 1 переліку ознак

Модель	RandomForest	LightGBM	LogReg	DecisionTree
Threshold	0.1071	0.4017	0.5503	0.5905
Accuracy	0.8703	0.8511	0.7908	0.7715
Balanced Accuracy	0.8199	0.8092	0.7755	0.7755
F1_Score	0.5511	0.5167	0.4321	0.4183
Precision	0.4336	0.3924	0.3024	0.2857
Recall	0.7561	0.7561	0.7561	0.7805

Далі проаналізуємо Light GBM за базовим типом («Basic»).

Класифікаційний звіт на тестовому наборі щодо реалізації моделі Light GBM має вигляд:

```
precision    recall  f1-score   support
0           0.9524    0.9469    0.9496     697
1           0.5698    0.5976    0.5833     82
accuracy                0.9101     779
```

macro avg	0.7611	0.7722	0.7665	779
weighted avg	0.9121	0.9101	0.9111	779

Модель LightGBM показує загальну високу точність (Accuracy = 0.9101), але ключове питання – наскільки добре вона впізнає рідкісний клас неуспіху. Balanced Accuracy говорить, що модель у середньому правильно впізнає 77% прикладів кожного класу. ROC-AUC означає, що модель добре розділяє класи. Низьке стандартне відхилення (± 0.0151) для ROC-AUC свідчить про високу стабільність моделі між фолдами, а в такому разі й про сильну узагальнюючу здатність, надійність результатів і відсутність перенавчання. PR-AUC є ключовою метрикою для незбалансованих класів, а її значення 0.65 означає, що модель краще, ніж випадкове вгадування, але ще вартує покращення.

Модель відмінно впізнає успішних студентів, що значною мірою забезпечується великою кількістю прикладів цього класу.

Помірна влучність (Precision) для класу "1", коли модель правильно прогнозує "неуспіх" у 57% випадків, означає, що 43% хибнопозитивних прогнозів – ризик для задач, де помилки коштують дорого. Крім того, стандартне відхилення цієї метрики за крос-валідацією (0.0579) означає, що точність для класу "1" коливається між фолдами.

Помірна повнота для класу "1" (Recall) говорить про те, що модель впізнає 60% усіх справжніх неуспішних студентів, тоді як 40% з них залишаються невпізнаними (хибнонегативні). Результати крос-валідації мають аналогічну влучності інтерпретацію.

Графічна ілюстрація результатів реалізації моделі Light GBM, а також важливість ознак наведена на рис. 4.11. Отже, модель Light GBM стабільна за загальними метриками (Accuracy, ROC-AUC), але нестабільна за метриками рідкісного класу (F1, Precision, Recall), що підтверджує складність його виявлення. Тому модель Light GBM для 2-го підходу з 1-м варіантом ознак може бути корисним інструментом для скринінгу з обов'язковою експертною перевіркою.

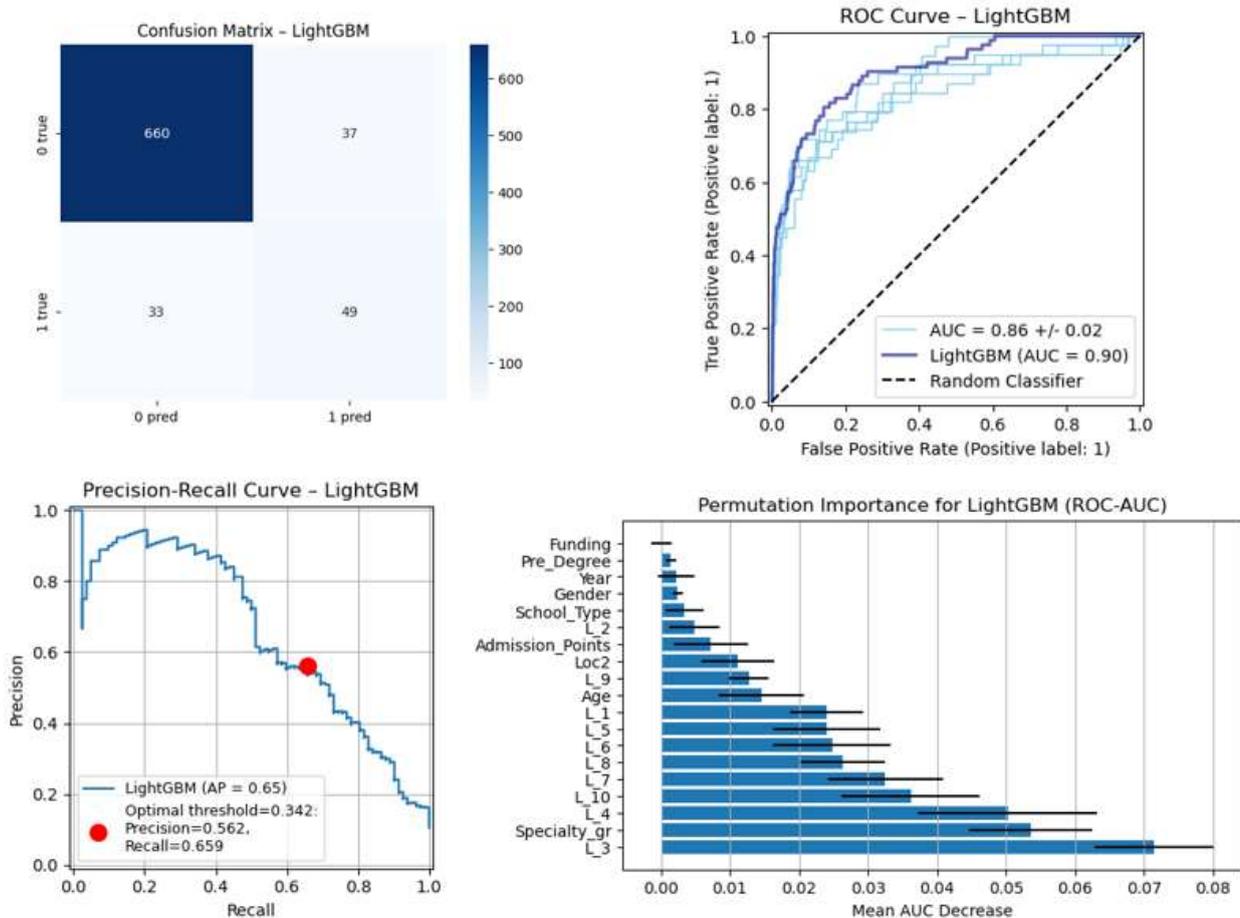


Рис. 4.11. Інфографіка результатів реалізації моделі Light GBM для 2-го підходу з 1-м варіантом ознак і відображенням їх важливості

Подібні висновки стосуються й моделі Random Forest для 2-го підходу з 1-м варіантом ознак, виходячи з результатів крос-валідації та класифікаційного звіту, а також графіків на рис. 4.12-4.13 (матриці плутанини, ROC-кривих з крос-валідації та прогнозу на тестовому наборі даних, PR-кривої разом з точкою порогу та діаграми пермутаційної важливості):

```

--- RandomForest (Basic) ---
      precision    recall  f1-score   support
0         0.9572     0.9311     0.9440         697
1         0.5248     0.6463     0.5792          82
 accuracy         0.9012         779
 macro avg         0.7410     0.7887     0.7616         779
 weighted avg         0.9117     0.9012     0.9056         779

```

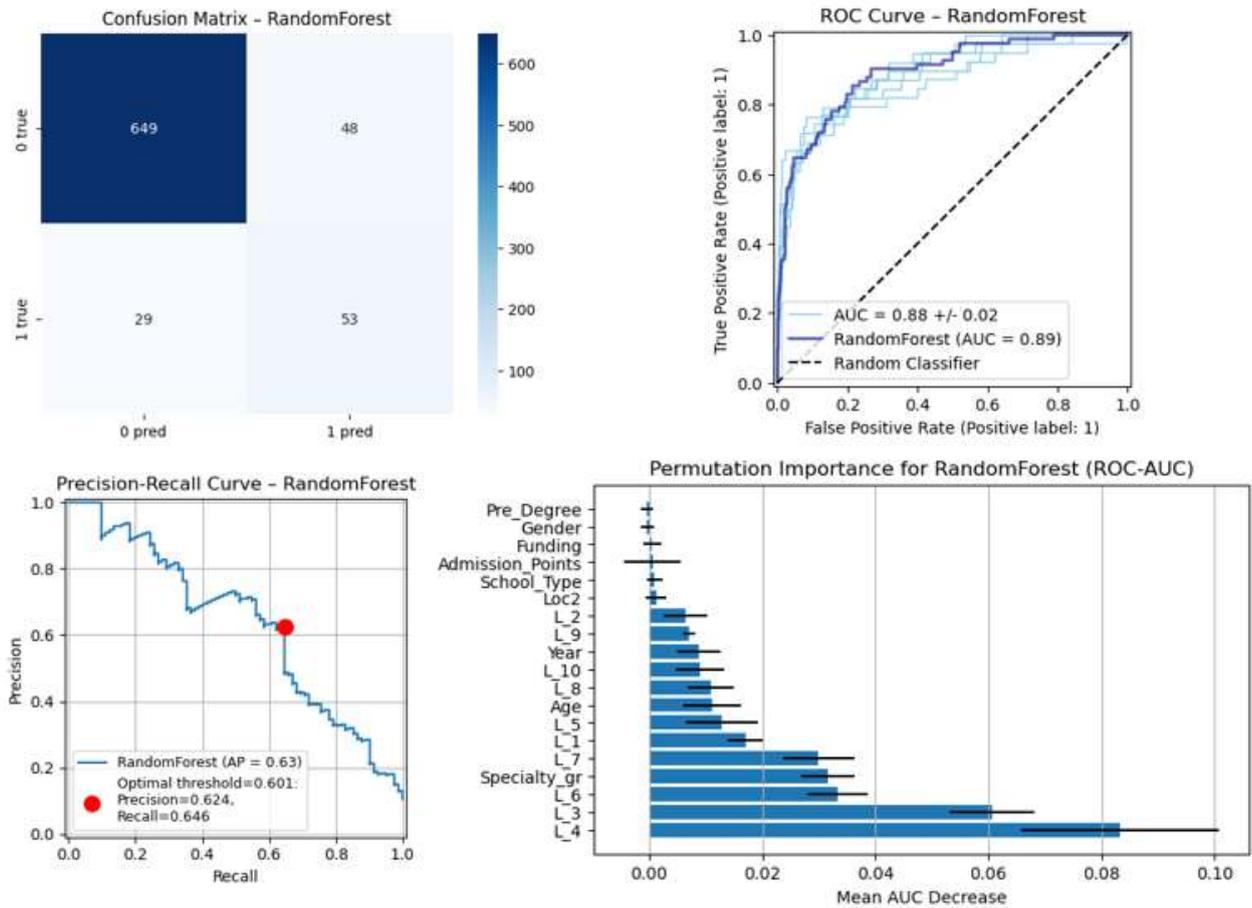


Рис. 4.12. Інфографіка результатів реалізації моделі Random Forest для 2-го підходу з 1-м варіантом ознак і відображенням їх важливості

Матриці плутанини для прогнозів з базовим і оптимізованими порогами «Opt-F1» і «Opt_TR» в рамках 2-го підходу та 1-го варіанту ознак представлені на рис. 4.13.



Рис. 4.13. Матриці плутанини для прогнозів з базовим і оптимізованими порогами «Opt-F1» і «Opt_TR» (підхід 2, варіант 1)

Класифікаційні звіти для кожного з модифікованих типів прогнозу «Opt-F1» і «Opt-TR»:

```

--- RandomForest (Opt-F1) ---
      precision    recall  f1-score   support
0         0.9582     0.9541    0.9561     697
1         0.6235     0.6463    0.6347      82
 accuracy                    0.9217     779
 macro avg                    0.7909     0.8002    0.7954     779
 weighted avg                  0.9230     0.9217    0.9223     779

```

```

--- RandomForest (Opt-TR) ---
      precision    recall  f1-score   support
0         0.9678     0.8623    0.9120     697
1         0.3924     0.7561    0.5167      82
 accuracy                    0.8511     779
 macro avg                    0.6801     0.8092    0.7143     779
 weighted avg                  0.9072     0.8511    0.8704     779

```

```

--- LightGBM (Opt-F1) ---
      precision    recall  f1-score   support
0         0.9590     0.9397    0.9493     697
1         0.5625     0.6585    0.6067      82
 accuracy                    0.9101     779
 macro avg                    0.7608     0.7991    0.7780     779
 weighted avg                  0.9173     0.9101    0.9132     779

```

```

--- LightGBM (Opt-TR) ---
      precision    recall  f1-score   support
0         0.9686     0.8838    0.9242     697
1         0.4336     0.7561    0.5511      82
 accuracy                    0.8703     779
 macro avg                    0.7011     0.8199    0.7377     779
 weighted avg                  0.9122     0.8703    0.8850     779

```

Отже, можна стверджувати, що для підходу 2 «проміжний» з 1-м варіантом ознак потенційно підходять наступні стратегії прогнозування академічної неуспішності з метою мінімізації пропусків здобувачів освіти:

- використовувати LightGBM «Opt-TR» або LogReg “Basic” для початкового виявлення (скринінгу), орієнтовно 75-80%, ризикових випадків, далі використовувати LightGBM «Opt-F1» або/i RandomForest «Opt-F1» для підтвердження – фільтрації хибних позитивних (FP)

- використовувати RandomForest «Opt-F1», якщо потрібно мати мінімальну кількість хибно позитивних (FP)

- використовувати RandomForest «Opt-TR» або LightGBM «Opt-TR» як основні моделі з метою отримання мінімуму пропусків «неуспіху»

– використовувати LogReg «Basic» або «Opt-TR» в разі максимального пріоритету на мінімум пропусків «неуспіху»

Далі викладемо **2-й варіант** переліку ознак в рамках **2-го підходу «проміжний»**. До переліку ознак у цьому випадку належать:

```
'Gender',
'Year',
'Funding',
'Admission_Points',
'Age',
'School_Type',
'Loc2',
'Specialty_gr',
'L_1', 'L_2', 'L_3', 'L_4', 'L_5', 'L_6', 'L_7', 'L_8',
'L_9', 'L_10',
'Course_Visits',
'Pr_prc'.
```

Слід звернути увагу на дві остані ознаки 'Course_Visits' і 'Pr_prc' – кількість входів на платформу електронного навчання і питому вагу присутності студента на заняттях в рамках окремої дисципліни, відповідно. Під час тренування моделей надходять фактичні значення за весь семестр. При практичному застосуванні таких моделей задля завчасної ідентифікації «неуспіху» на вхід моделей подаються очікувані значення цих ознак.

Гіперпараметри моделей згідно підходу 2 «проміжний» і 2-го варіанту ознак визначені наступним чином:

```
models = {
  "DecisionTree": DecisionTreeClassifier(
    criterion = "entropy"
    , max_depth = None
    , splitter="random"
    , min_samples_split = 4
    , min_samples_leaf = 2
    , min_weight_fraction_leaf = 0.0
    , max_features = None
    , class_weight = "balanced"
    , random_state = 1234
  ),
  "RandomForest": RandomForestClassifier(
    criterion = "gini"
    , n_estimators = 2000
    , class_weight = "balanced"
```

```

, min_samples_split = 8
, min_samples_leaf = 4
, max_features = "sqrt"
, bootstrap = True
, random_state=1234
, n_jobs=-1
),

"LightGBM": LGBMClassifier(
    n_estimators = 1000
, learning_rate = 0.015
, max_depth = -1
, class_weight = 'balanced'
, num_leaves = 63
, max_features = 'sqrt'
, random_state=1234
, n_jobs=-1
, boosting_type = 'gbdt'
, verbose=-1
),

"LogReg": LogisticRegression(
    max_iter = 1000
, solver = 'liblinear'
, class_weight='balanced'
, random_state = 1234
, n_jobs=-1
, penalty='l2'
, C = np.float64(10)
)
}

```

Оцінка точності моделей виявлення неуспішного вивчення дисципліни студентами протягом одного семестру спирається на метрики в табл. 4.7. в розрізі двох типів прогнозування – базового («Basic») та з оптимізованим порогом Threshold на основі Precision-Recall кривої, що максимізує F1-score для рідкісного класу 1 – «неуспіх» («Opt-F1»).

Таблиця 4.7

**Метрики точності моделей
за 2-м підходом “проміжний” і 2-м варіантом переліку ознак**

Характеристики	Моделі							
	Decision Tree		Light GBM		Log Reg		Random Forest	
Тип прогнозу	Basic	Opt-F1	Basic	Opt-F1	Basic	Opt-F1	Basic	Opt-F1
Threshold	0.5000	1.0000	0.5000	0.9857	0.5000	0.8355	0.5000	0.5564
Accuracy	0.9345	0.9525	0.9564	0.9641	0.7818	0.9384	0.9551	0.9589
Balanced Accuracy	0.8397	0.8174	0.8303	0.8346	0.8081	0.8311	0.8081	0.8103

F1_Score	0.6982	0.7413	0.7639	0.7971	0.4481	0.7037	0.7445	0.7612
Precision	0.6782	0.8689	0.8871	0.9821	0.3053	0.7125	0.9273	0.9808
Recall	0.7195	0.6463	0.6707	0.6707	0.8415	0.6951	0.6220	0.6220
ROC_AUC	0.8485		0.9220		0.8991		0.9209	
PR_AUC	0.6444		0.7993		0.6567		0.7894	
CV_ROC_AUC_mean	0.8195		0.9136		0.8666		0.9258	
CV_ROC_AUC_std	0.0176		0.0107		0.0144		0.0083	
CV_F1_mean	0.6068		0.7040		0.4195		0.7055	
CV_F1_std	0.0417		0.0152		0.0219		0.0217	

Зокрема, «Проміжний» сценарій не позиціонується як безпосередньо готовий до застосування в реальному часі, а розглядається як аналітичний еталон, що дозволяє оцінити потенціал поведінкових показників за умови їх повної спостережуваності. Отримані у таблиці 4.7 високі значення Ассурасу (0,93–0,96) відображають саме цю граничну оцінку і не інтерпретуються як гарантована точність у реальних умовах середини семестру. У практичному режимі використання запропонованої інформаційної технології передбачається подання на вхід моделей часткових або прогнозованих значень поведінкових показників, сформованих на основі накопичених даних на конкретний момент часу. З точки зору мінімізації пропусків (FN) пороги слід змістити у бік нижчих значень. Таким чином забезпечується зростання Recall до цільового рівня, наприклад, в діапазоні 0.75–0.85. При цьому також забезпечується прийнятний рівень Precision. Таблиця 4.8 містить метрики продуктивності моделей з пошуком порогів для класу 1, які дозволяють досягти цільовий рівень Recall. Позначимо такий тип прогнозування через «Opt_TR».

Таблиця 4.8

Метрики продуктивності моделей з визначенням під цільовий рівень Recall порогом для класу 1 (тип «Opt_TR») згідно з підходом 2 і варіантом ознак 2

Модель	RandomForest	LightGBM	LogReg	DecisionTree
Threshold	0.3629	0.0181	0.7084	0.0000
Accuracy	0.9255	0.9024	0.8755	0.1053
Balanced Accuracy	0.8508	0.8433	0.8228	0.5000
F1_Score	0.6813	0.6238	0.5611	0.1905
Precision	0.6200	0.5250	0.4460	0.1053
Recall	0.7561	0.7683	0.7561	1.0000

Маючи сильний дисбаланс класів, тобто успішних студентів значно більше, ніж неуспішних – 78,5% проти 21,5% , у таблиці 4.8 (DecisionTree, Opt-TR) ми отримали Accuracy 0,1053 при Recall 1,0. Так, при дисбалансі класів Accuracy вводить в оману, модель відносить усі спостереження до класу «неуспішний». З метою уникнення подібних тривіальних рішень у всіх експериментах застосовано методи оптимізації моделей та використано збалансовані метрики оцінювання. Доповнимо метрики з табл. 4.7 класифікаційними звітами (додаток В):

Наведені результати реалізації моделей з базовим порогом «Basic» показують, що моделі LightGBM та RandomForest знову демонструють найкращу продуктивність. Вони мають високу PR-площу ($PR-AUC \approx 0.80$), що свідчить про помірну продуктивність у виявленні неуспіху. Обидві моделі мають малі стандартні відхилення, що підтверджує стабільність у крос-валідації.

Модель LightGBM демонструє високу загальну точність та гарну дискримінативність ($ROC-AUC > 0.9$). Вона добре впізнає обидва класи за рахунок збалансованої точності. Частка хибнопозитивних попереджень низька, оскільки частка правильних позитивних передбачень (Precision) ближче до 0.89. При цьому $Recall \approx 0.67$. Графічна ілюстрація результатів реалізації моделі представлена на рис. 4.14, де зображено матрицю плутанини, ROC-криві з крос-валідації, ROC-криву з прогнозу на тестовому наборі даних, PR-криву та точку порогу на ній, а також діаграму пермутаційної важливості ознак.

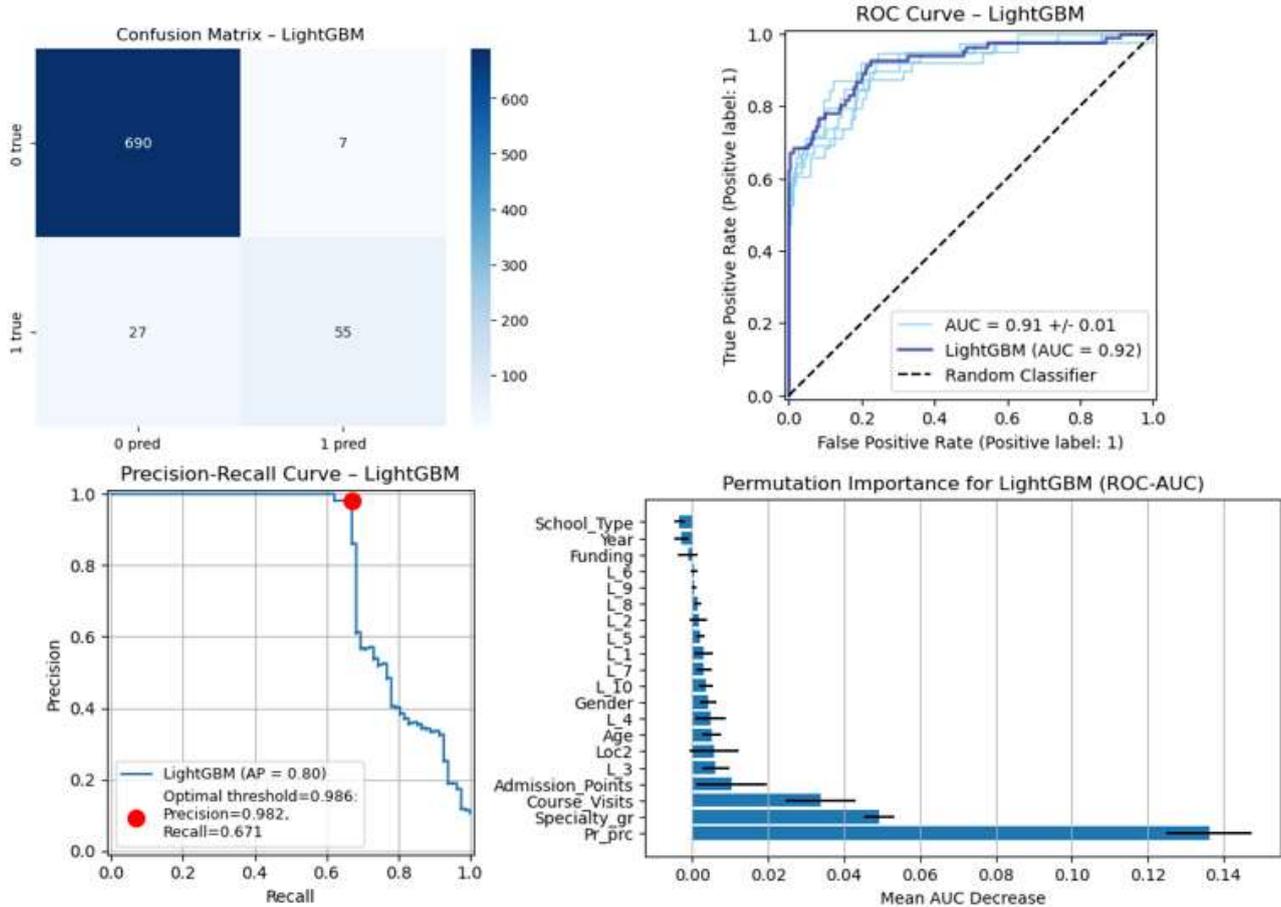


Рис. 4.14. Інфографіка результатів реалізації моделі Light GBM і важливості ознак згідно 2-го підходу («проміжний») і 2-го варіанту переліку ознак (тип «Basic»)

Для моделі Random Forest точність (Accuracy) та ROC-AUC схожі з аналогічними метриками LightGBM, але Recall нижчий за LightGBM, тоді як влучність (Precision) дуже висока. F1 на рівні 0.7445 вказує на добре збалансований компроміс між Precision і Recall для класу 1 («неуспіх») на тестових даних, але все ж таки існує ризик пропускати частину дійсних випадків через більш низьку повноту (Recall). Матриця плутанини, ROC-крива, PR-крива та пермутаційна важливість ознак для моделі з Random Forest проілюстровані на рис. 4.15.

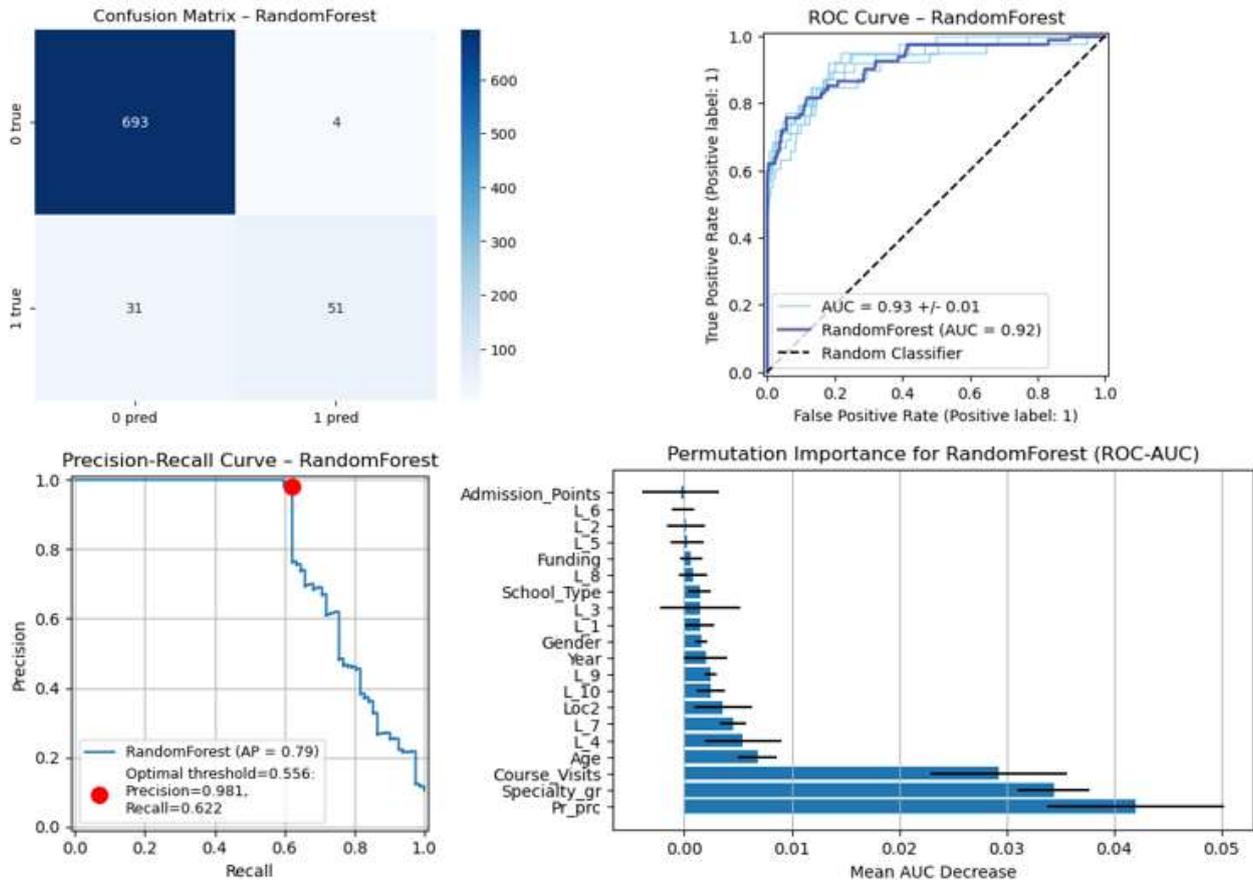


Рис. 4.15. Інфографіка результатів реалізації моделі Random Forest і важливості ознак (підхід 2 і варіант 2 переліку ознак, тип «Basic»)

Модель, побудована на основі дерева рішень (DecisionTree), демонструє найвищу збалансовану точність (BalancedAccuracy) на тестових даних (0.84), а також має достатній рівень F1-score для класу 1. Добре впізнає успішних студентів, але має середню здатність виявляти неуспіх (67.82% прогнозів "неуспіху" правильні, 71.95% справжніх "неуспіхів" виявлені).

Візуалізація результатів реалізації моделі представлена на рис. 4.16.

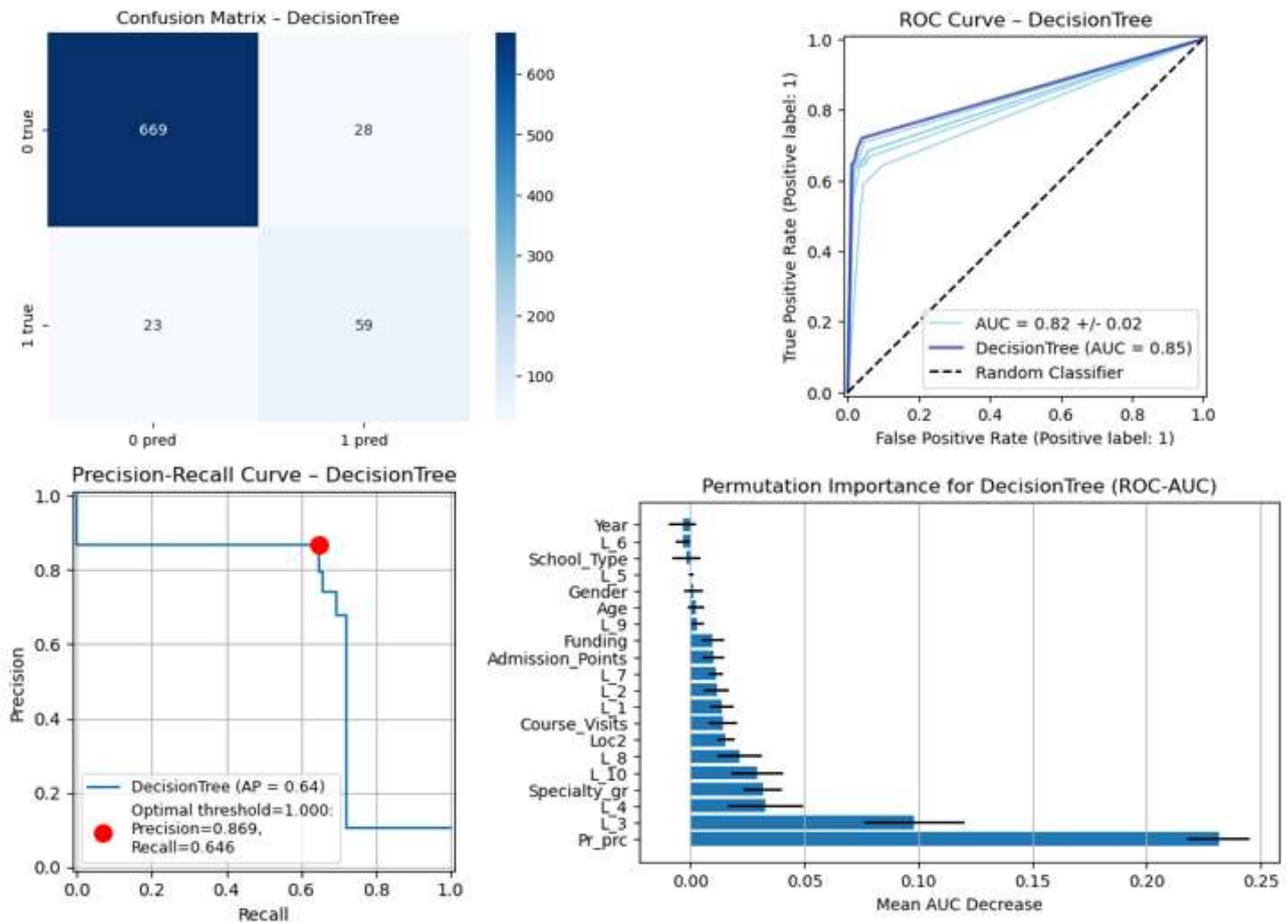


Рис. 4.16. Інфографіка результатів реалізації моделі DecisionTree і важливості ознак (підхід 2 і варіант 2 переліку ознак, тип «Basic»)

Нарешті, рис. 4.17 відображає результати застосування логістичної регресії для прогнозування академічної неуспішності студентів. LogReg має високу повноту, але низьку точність. Може бути корисною для попереднього скринінгу, але нерідко з хибнопозитивними прогнозами. Тому потребує додаткової оптимізації.

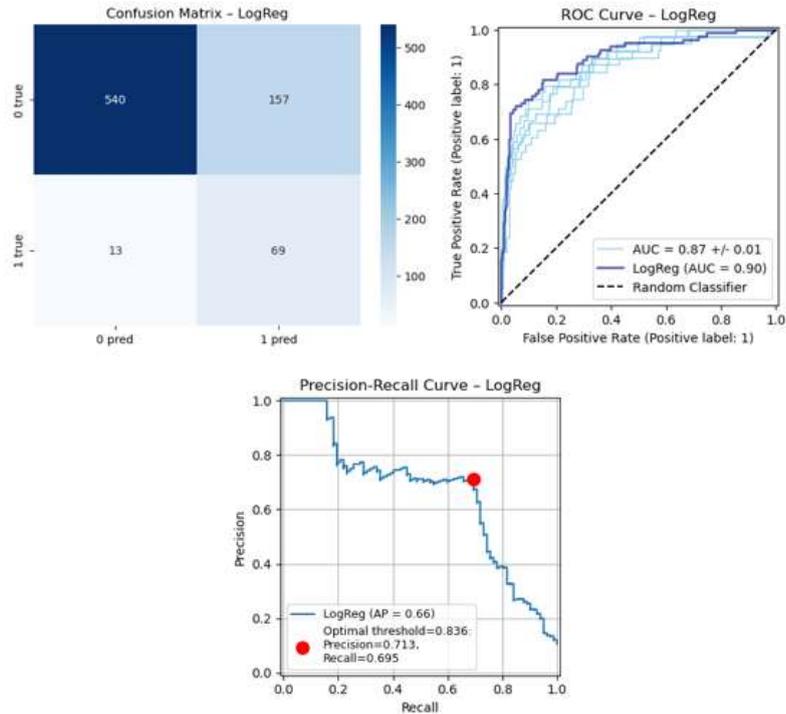


Рис. 4.17. Інфографіка результатів реалізації моделі LogReg і важливості ознак (підхід 2 і варіант 2 переліку ознак, тип «Basic»)

Приклад ідентифікації неуспіху на основі підходу 2 і варіанту 2 (базовий поріг):

```

=== DecisionTree ===
Ймовірності для 30-го тестового зразка (факт. клас 1):
  Клас '0' (індекс 0): 0.0000
  Клас '1' (індекс 1): 1.0000
Передбачений клас: 1
Впевненість у прогнозі: 1.0000

=== RandomForest ===
Ймовірності для 30-го тестового зразка (факт. клас 1):
  Клас '0' (індекс 0): 0.0264
  Клас '1' (індекс 1): 0.9736
Передбачений клас: 1
Впевненість у прогнозі: 0.9736

=== LightGBM ===
Ймовірності для 30-го тестового зразка (факт. клас 1):
  Клас '0' (індекс 0): 0.0000
  Клас '1' (індекс 1): 1.0000
Передбачений клас: 1
Впевненість у прогнозі: 1.0000

=== LogReg ===
Ймовірності для 30-го тестового зразка (факт. клас 1):
  Клас '0' (індекс 0): 0.0005
  Клас '1' (індекс 1): 0.9995
Передбачений клас: 1
Впевненість у прогнозі: 0.9995

```

На рис. 4.18 наряду з матрицями плутанини для моделей з базовим порогом («Basic») наведені матриці плутанини для прогнозів з оптимізованим порогом «Opt-F1», який максимізує F1-score, і порогом «Opt_TR», котрий забезпечує цільовий рівень 0.75 повноти Recall. Матриці за базовим порогом розташовані ліворуч, за першим типом оптимізованих порогів – в центрі, а за другим – праворуч.

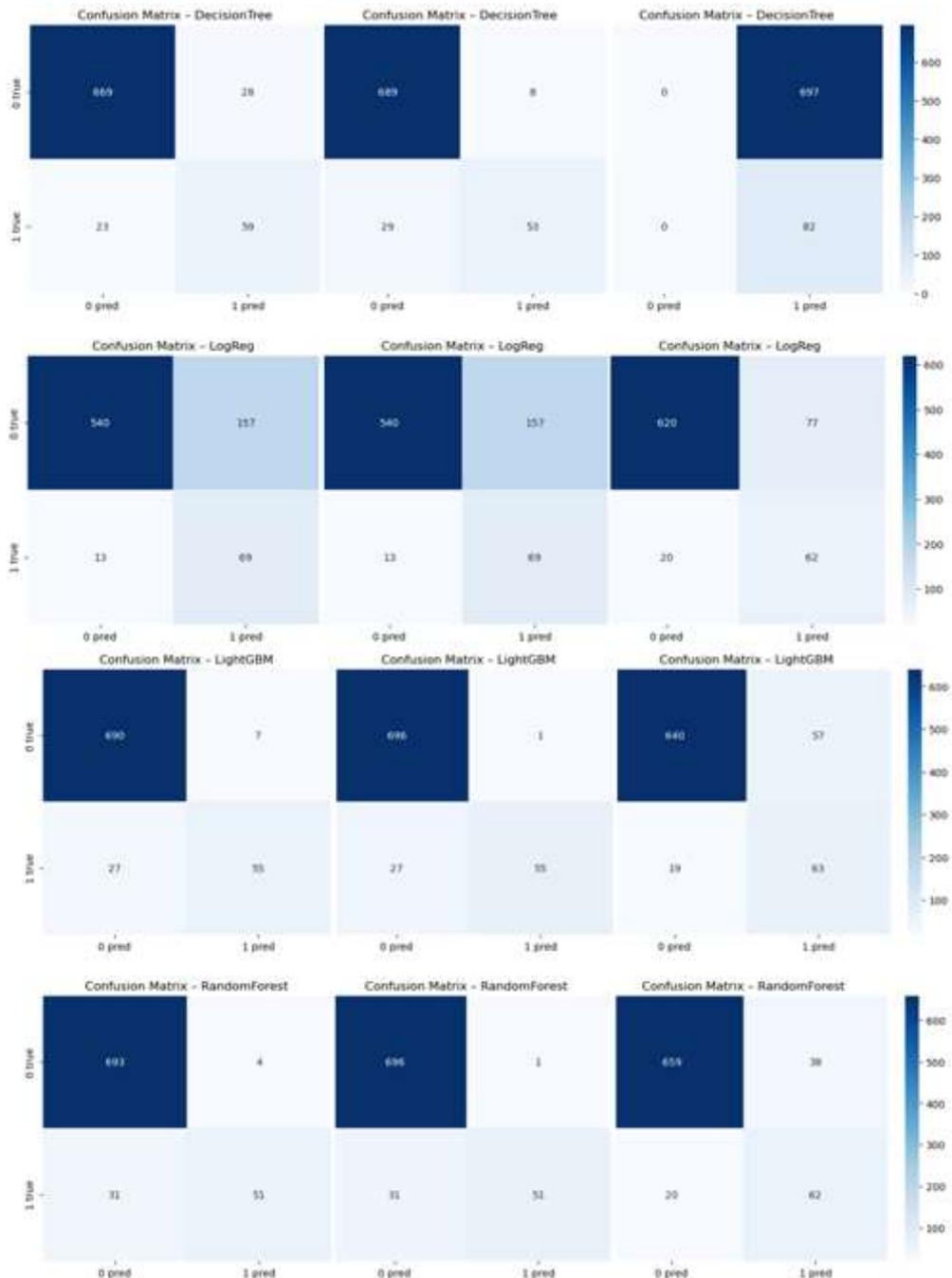


Рис. 4.18. Матриці плутанини для прогнозів з базовим і оптимізованими порогоми «Opt-F1» і «Opt_TR» (підхід 2, варіант 2)

Отже, можна стверджувати, що для підходу 2 «проміжний» з 2-м варіантом ознак потенційно підходять наступні стратегії прогнозування академічної неспішності з метою мінімізації пропусків неуспіхів:

- використовувати Random Forest або/i Light GBM за типом «Opt-TR» для початкового виявлення ризикових студентів (скринінгу), далі використовувати RandomForest або/i Light GBM за типом «Opt-F1» для підтвердження ризику

- використовувати модель Random Forest або/i Light за типом «Opt-F1», якщо важливо мати меншу кількість хибних позитивів, тобто помилково позначених успішних студентів як неуспішних

- використовувати модель Random Forest або/i Light за типом «Opt-TR», якщо більш важливим є мінімум пропусків «неуспіхів»

- використовувати LogReg за типом «Opt-TR» для скринінгу, далі використовувати RandomForest або/i Light GBM за типом «Opt-F1» для підтвердження ризику.

4.3.3. Сценійний підхід 3 «Прикінцевий»

Передбачається, що прогнозування відбувається перед підсумковим контролем.

Але фактично, моделі за цим підходом можуть використовуватись будь-коли.

Тому є два варіанти застосування моделі:

варіант 1 – наприкінці курсу, перед підсумковим контролем. На вхід моделі подаються фактичні значення всіх ознак;

варіант 2 – протягом семестру. На вхід моделі подаються очікувані значення відвідування та входів на електронну платформу онлайн-навчання.

Основний набір ознак для моделей в рамках 3-го підходу містить:

```
'Gender',
'Year',
'Funding',
'Admission_Points',
'Age',
'School_Type',
'Loc2',
'Specialty_gr',
'Course_Visits',
'Pr_prc'.
```

Звісно, цей набір може бути модифікований, з одного боку, за рахунок додавання інших ознак, а з іншого боку, шляхом вилучення найменш важливих ознак. Завдяки `RandomizedSearchCV()` визначено наступні гіперпараметри моделей в рамках підходу 3 «прикінцевий»:

```
models = {
    "DecisionTree": DecisionTreeClassifier(
        criterion = "entropy"
        , max_depth = None
        , splitter="random"
        , min_samples_split = 2
        , min_samples_leaf = 1
        , min_weight_fraction_leaf = 0.0
        , max_features = None
        , class_weight = "balanced"
        , random_state = 1234
    ),

    "RandomForest": RandomForestClassifier(
        criterion = "gini"
        , n_estimators = 2000
        , class_weight = "balanced"
        , min_samples_split = 8
        , min_samples_leaf = 4
        , max_features = None
        , bootstrap = True
        , random_state=1234
        , n_jobs=-1
    ),

    "LightGBM": LGBMClassifier(
        n_estimators = 2000
        , learning_rate = 0.015
        , max_depth = 10
        , class_weight = 'balanced'
        , num_leaves = 31
        , max_features = 'log2'
        , random_state=1234
        , n_jobs=-1
        , boosting_type = 'gbdt'
        , verbose=-1
    ),

    "LogReg": LogisticRegression(
        max_iter = 1000
        , solver = 'liblinear'
        , class_weight='balanced'
        , random_state = 1234
        , n_jobs=-1
        , penalty='l2'
```

```

    , C = np.float64(0.1)
    )
}

```

Оцінка точності моделей прогнозування неуспішного вивчення дисципліни студентами протягом одного семестру згідно підходу 3 «прикінцевий» відображена в табл. 4.9. В таблиці показані метрики для двох типів прогнозування – базового («Basic») та з оптимізованим порогом Threshold на основі Precision-Recall кривої, що максимізує F1-score для рідкісного класу 1 – «неуспіх» («Opt-F1»).

Таблиця 4.9

Метрики точності моделей за 3-м підходом “прикінцевий” для двох типів прогнозування: «Basic» і «Opt-F1»

Характеристики	Моделі							
	Decision Tree		Light GBM		Log Reg		Random Forest	
Тип прогнозу	Basic	Opt-F1	Basic	Opt-F1	Basic	Opt-F1	Basic	Opt-F1
Threshold	0.5000	0.9621	0.5000	0.9986	0.5000	0.8154	0.5000	0.9040
Accuracy	0.9230	0.9294	0.9422	0.9602	0.7522	0.9589	0.9409	0.9589
Balanced Accuracy	0.8278	0.8260	0.8225	0.8110	0.7755	0.8049	0.8433	0.8049
F1_Score	0.6591	0.6746	0.7097	0.7669	0.4062	0.7576	0.7195	0.7576
Precision	0.6170	0.6552	0.7534	1.0000	0.2716	1.0000	0.7195	1.0000
Recall	0.7073	0.6951	0.6707	0.6220	0.8049	0.6098	0.7195	0.6098
ROC_AUC	0.8301		0.9111		0.8738		0.9173	
PR_AUC	0.4910		0.7896		0.7493		0.7953	
CV_ROC_AUC_mean	0.8105		0.9015		0.8574		0.9072	
CV_ROC_AUC_std	0.0203		0.0277		0.0277		0.0154	
CV_F1_mean	0.6458		0.6546		0.3925		0.6273	
CV_F1_std	0.0294		0.0318		0.0271		0.0378	

Доповнимо метрики в табл. 4.9 класифікаційними звітами в розрізі типів прогнозування (Додаток В).

Як видно зі звітів, найкращими є RandomForest і LightGBM з обома типами прогнозування Basic і Opt-F1. Вони показують високі значення ROC-AUC ($\approx 0.91-0.92$) і PR-AUC ($\approx 0.79-0.80$), що говорить про суттєву здатність розрізняти класи та пріоритезувати позитивні випадки. Високий F1-Score ($\approx 0.70-$

0.77) і Balanced Accuracy ($\approx 0.80\text{--}0.84$) вказують на здатність балансувати між влучністю та повнотою прогнозів для незбалансованого класу.

Рис. 4.19. графічно ілюструє результати реалізації LightGBM для прогнозування академічної неуспішності, зокрема: матрицю плутанини для основного типу прогнозування «Basic», ROC-криві з крос-валідації та ROC-криву з прогнозу («Basic») на тестовому наборі, PR-криву разом з точкою оптимального порогу та діаграму пермутаційної важливості ознак.

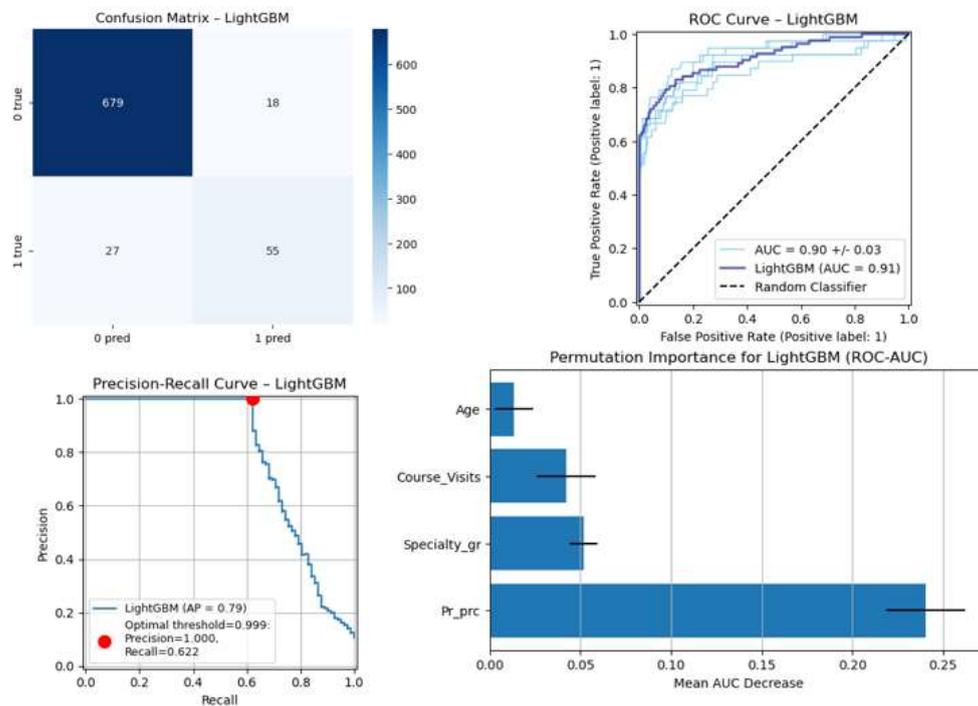


Рис. 4.19. Інфографіка результатів Light GBM і важливості ознак згідно 3-го підходу («прикінцевий») і типу прогнозування «Basic»

Візуалізація результатів перевірки продуктивності Random Forest на тестовому наборі даних за базовим типом прогнозування здійснена на рис. 4.20.

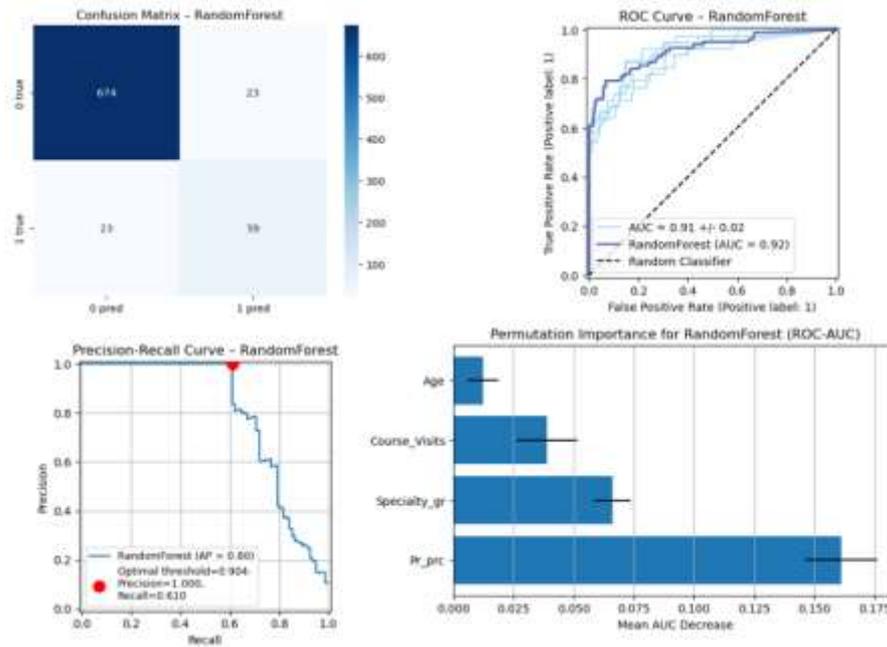


Рис. 4.20. Інфографіка результатів RandomForest і важливості ознак згідно 3-го підходу («прикінцевий») і типу прогнозування «Basic»

Результати перевірки DecisionTree на тестовому наборі даних проілюстровано на рис. 4.21. Попри непоганий Recall і Balanced Acc, низький PR-AUC свідчить про слабку здатність до пріоритизації позитивного класу.

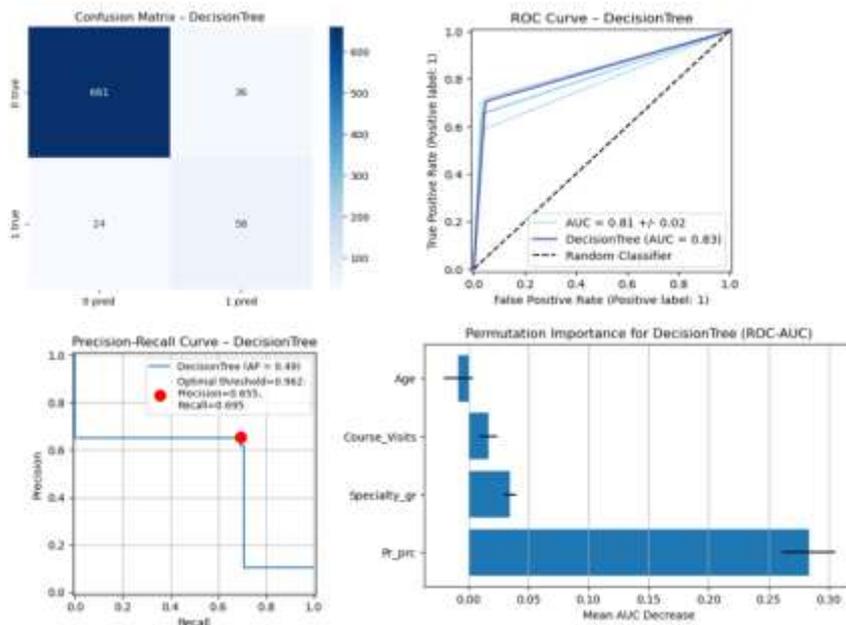


Рис. 4.21 Інфографіка результатів DecisionTree і важливості ознак згідно 3-го підходу («прикінцевий») і типу прогнозування «Basic»

Рисунок 4.22 демонструє результати застосування LogReg на тестових даних. Відповідно до крос-валідації модель має низький середній F1-score

(≈ 0.393), що відображає його схильність до високого Recall при низькій Precision. Вона добре «підбирає» клас 1 (мінімум пропусків), але ціною великої кількості хибних спрацьовувань.

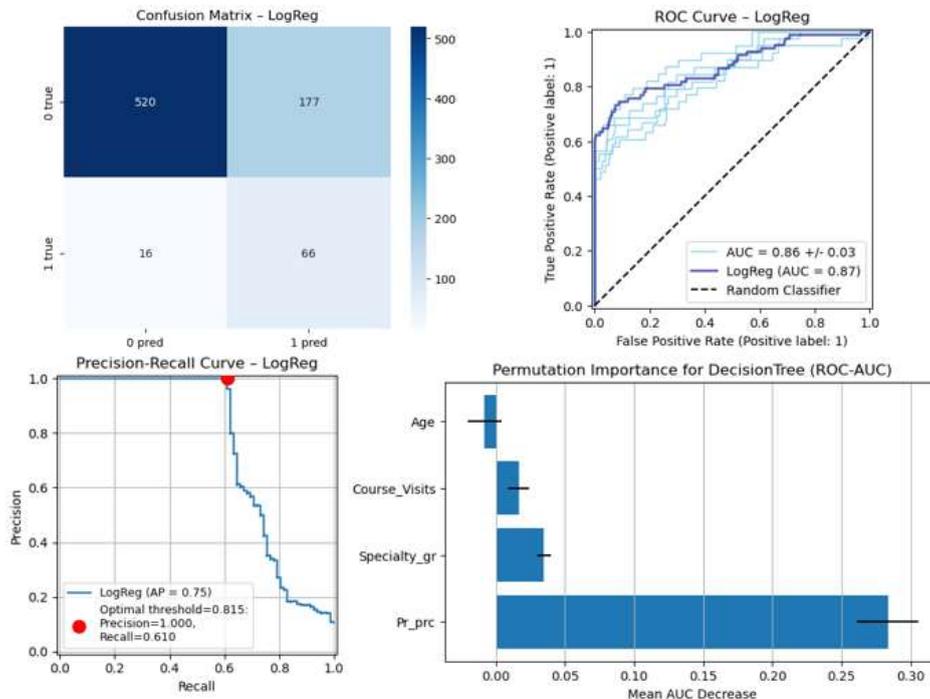


Рис. 4.22. Інфографіка результатів LogReg згідно 3-го підходу («прикінцевий») і типу прогнозування «Basic»

Вплив оптимізації порога (тип «Opt-F1») призвів до наступного:

- у RandomForest, LightGBM, LogReg, оптимізований поріг порівняно з базовим порогом привів до підвищення Precision до 1 та легкого падіння Recall до 0.61–0.62, що в цілому підвищило F1 до 0.76–0.77 й Accuracy до приблизно 0.96, але знизило Balanced Accuracy.

- У DecisionTree оптимізація дала малий зсув (Precision ≈ 0.655 , Recall ≈ 0.695), F1 зріс до ≈ 0.675 .

З точки зору мінімізації пропусків (FN) пороги слід змістити у бік нижчих значень, добираючи за PR-кривою для цільового рівня Recall (наприклад, ≥ 0.75 –0.85) із прийнятним рівнем Precision. Таблиця 4.10 містить метрики продуктивності моделей з пошуком порогів для класу 1, які дозволяють досягти цільовий рівень Recall. Позначимо такий тип прогнозування через «Opt_TR».

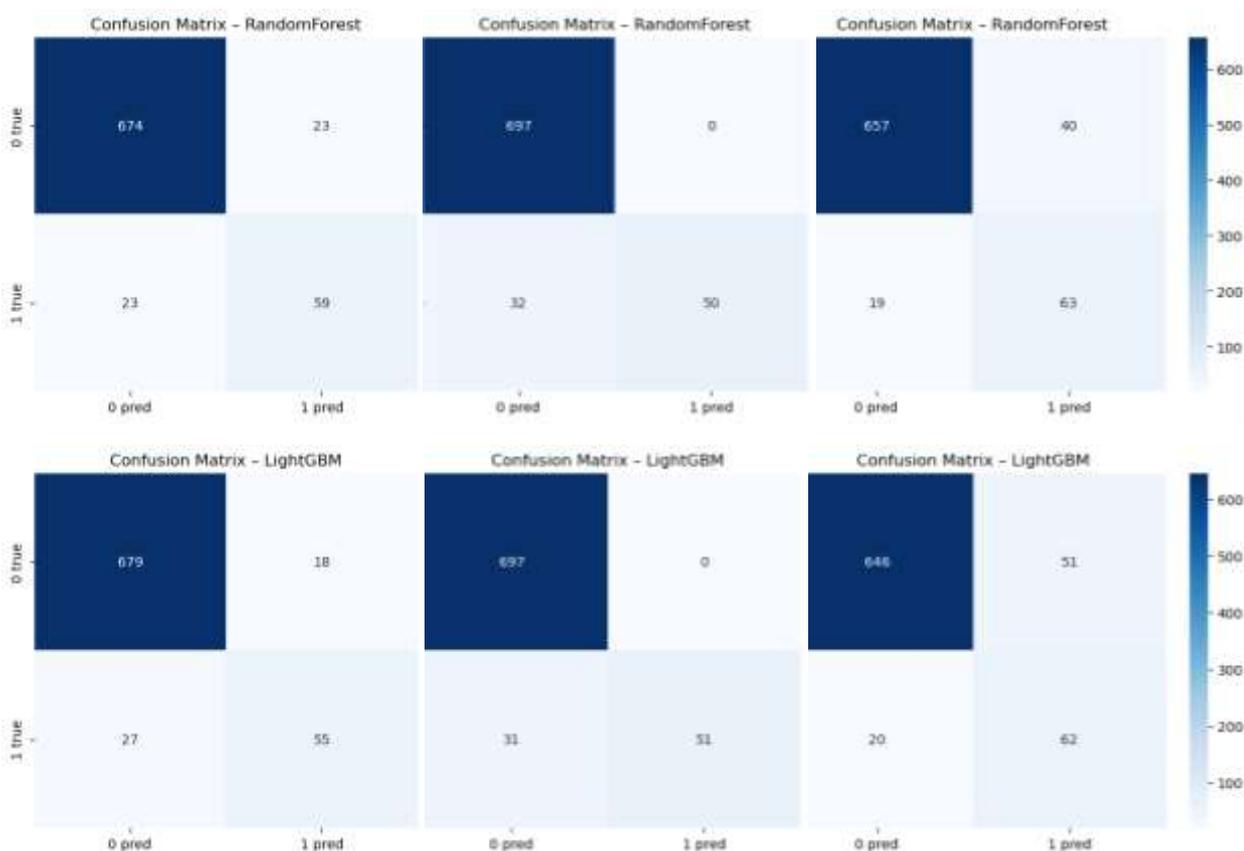
RandomForest має Precision = 1.0 – всі студенти, яких система визначила як "неуспішні", справді мають ризик неуспіху, а Recall = 0.6098 – 61% справді неуспішних студентів будуть виявлені.

Таблиця 4.10

**Метрики продуктивності моделей з визначеним під цільовий рівень
Recall порогом для класу 1 (тип «Opt_TR»)**

Модель	RandomForest	LightGBM	LogReg	DecisionTree
Threshold	0.4092	0.0668	0.6801	0.0000
Accuracy	0.9243	0.9089	0.8665	0.1053
Balanced Accuracy	0.8555	0.8415	0.8178	0.5000
F1_Score	0.6811	0.6359	0.5439	0.1905
Precision	0.6117	0.5487	0.4247	0.1053
Recall	0.7683	0.7561	0.7561	1.0000

На рис.4. 23 наряду з матрицями плутанини для моделей з базовим порогом («Basic») наведені матриці плутанини для прогнозів з оптимізованим порогом «Opt-F1», що максимізує F1-score, і порогом «Opt_TR», що забезпечує цільовий рівень 0.75 повноти Recall. Матриці за базовим порогом розташовані ліворуч, за першим типом оптимізованих порогів – в центрі, а за другим – праворуч.



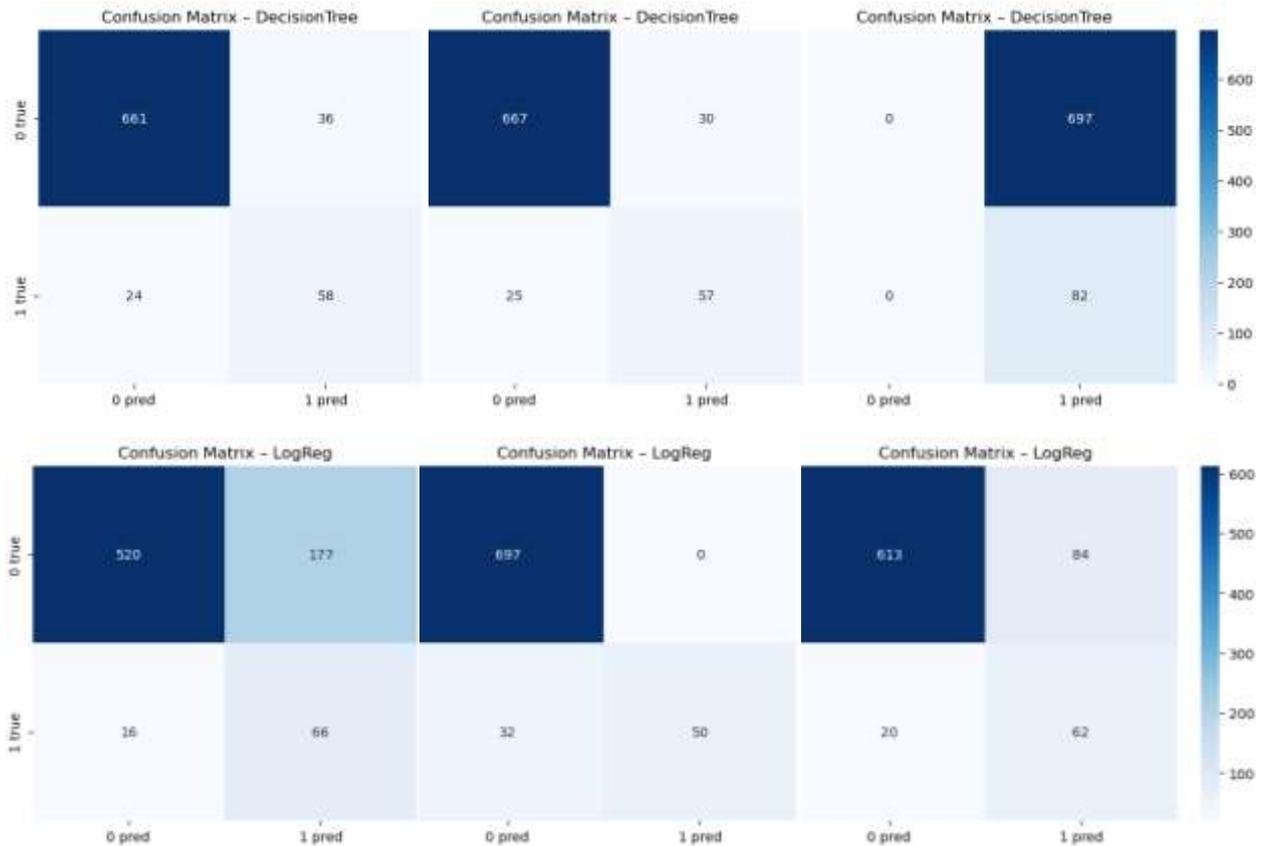


Рис. 4.23. Матриці плутанини для прогнозів з базовим і оптимізованими порогамі «Opt-F1» і «Opt_TR» (підхід 3)

Отже, сформулюємо потенційні стратегії застосування моделей для 3-го підходу «прикінцевий»:

- якщо важливо мінімізувати хибні позитиви, тобто кількість помилково позначених успішних студентів як неуспішних, використовується RandomForest (Opt-F1 з порогом 0.904)

- далі використовується LightGBM (Opt-F1) як додатковий фільтр для студентів, яких оцінила RandomForest як «можливо неуспішні», щоб мінімізувати обсяг ручної перевірки, але зберегти точність

- якщо важливо мінімізувати пропуски (FN), тобто кількість помилково позначених неуспішних студентів як успішних, здійснюється скринінг завдяки RandomForest (Opt-TR з порогом 0.409) або LightGBM (Opt-TR з порогом 0.067) або ансамблю цих двох моделей.

4.4. Візуалізація як етап аналізу даних в освітній аналітиці

Візуалізація даних в запропонованій інформаційній технології використовується не лише як інструмент представлення результатів аналізу, а як ключовий етап аналітичного процесу, що дозволяє ідентифікувати закономірності, аномалії, тренди та взаємозв'язки у великих обсягах освітніх даних. Вона виступає критично важливим мостом між початковими (сирими) даними та інтерпретованим знанням, підвищуючи ефективність прийняття рішень у сфері освіти. Основні функції візуалізації в освітній аналітиці можна подати у вигляді структури (рис.4.24):

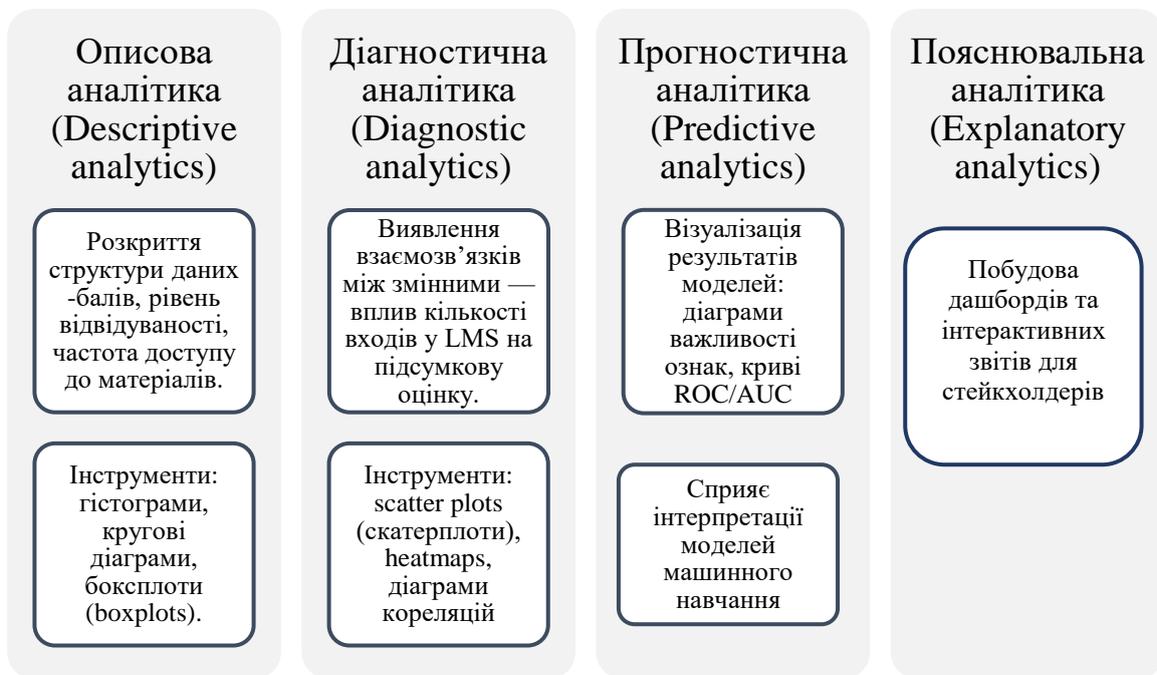


Рис. 4.24. Основні функції візуалізації в освітній аналітиці

При візуалізації освітніх даних у нашому дослідженні було дотримано принципів інформативності (діаграми передають зміст, точності (візуалізація не повинна викривлювати співвідношення чи динаміку), доступності (зрозумілість графіків для всіх споживачів цих звітів без технічної підготовки) та інтерактивності (на сучасних ВІ-платформах, зокрема Microsoft Power BI є можливість досліджувати дані глибше). При обґрунтуванні та виборі ВІ системи

враховувались такі показники, що були також зазначені в літературних джерелах:

- Інтеграція з освітніми інформаційними системами

Power BI має розвинену екосистему конекторів до поширених джерел даних, зокрема до LMS (Moodle, Canvas), Excel, SQL Server, SharePoint, Google Analytics, що спрощує імпорт даних з академічних журналів, систем управління навчанням, інших баз даних. У науковій роботі [64] підкреслено, що Power BI забезпечує безперебійну інтеграцію з різними освітніми платформами, що дозволяє здійснювати оновлення даних у режимі реального часу та створювати динамічні звіти.

- Зрозумілість інтерфейсу та доступність

Power BI має відносно простий графічний інтерфейс, який дозволяє користувачам без глибоких знань створювати інтерактивні дашборди та користуватися ними. «Зручний інтерфейс та економічна ефективність Power BI роблять його ідеальним інструментом для візуалізації даних та звітності в освітній сфері» [76].

- Функціональні можливості

Power BI підтримує потужну мову аналітичних запитів DAX (Data Analysis Expressions) і має можливість використовувати R та Python-скрипти для поглибленої аналітики, що дозволило реалізувати складні метрики освітньої успішності, прогнозування, сегментацію студентів та використовувати цю BI систему в комплексі для функціонування інформаційної технології освітньої аналітики.

- Використання в освітніх проєктах

Враховуючи проведені дослідження інтенсивності використання цих платформ для освітніх потреб, можна стверджувати, що Power BI активно використовується в аналітичних освітніх проєктах по всьому світу, зокрема для моніторингу прогресу студентів, візуалізації їх відвідуваності, виявлення ризиків відрахування, аналізу доступу до електронного контенту [56].

У межах концепції прозорості та відкритості вищої освіти в Україні важливу роль відіграє публічне представлення даних про діяльність університету у формі інтерактивних візуалізацій. Згідно з принципами відкритих даних та рекомендаціями UNESCO Open Educational Data Framework [107], університети повинні забезпечувати доступ до статистичної та аналітичної інформації про освітній процес, кадровий склад, фінансові витрати, результати наукової діяльності, індикатори якості освіти тощо. В рамках апробації інформаційної технології освітньої аналітики для НУБІП України була реалізована задача формування та публікації відповідних дашбордів відкритих даних. Для цього використовувалась запропонована інформаційна технологія, яка в рамках цього етапу передбачала:

- обробку та узгодження структури даних, створення моделі даних в ВІ системі;
- побудову інтерактивних візуалізацій за допомогою інструментів Power BI;
- публікацію звітів у вигляді вбудованих модулів в цифровому середовищі університету.

Розроблені дашборди включають інтерактивні звіти для загального користування – відкриті. Вони включають дані про: структуру університету (кількість факультетів, спеціальностей, освітніх програм, відокремлених підрозділів НУБІП України); статистику контингенту здобувачів освіти (за гендерною ознакою, за факультетами, за спеціальностями та освітніми програмами в розрізі базового закладу та відокремлених підрозділів) (рис.4.25).

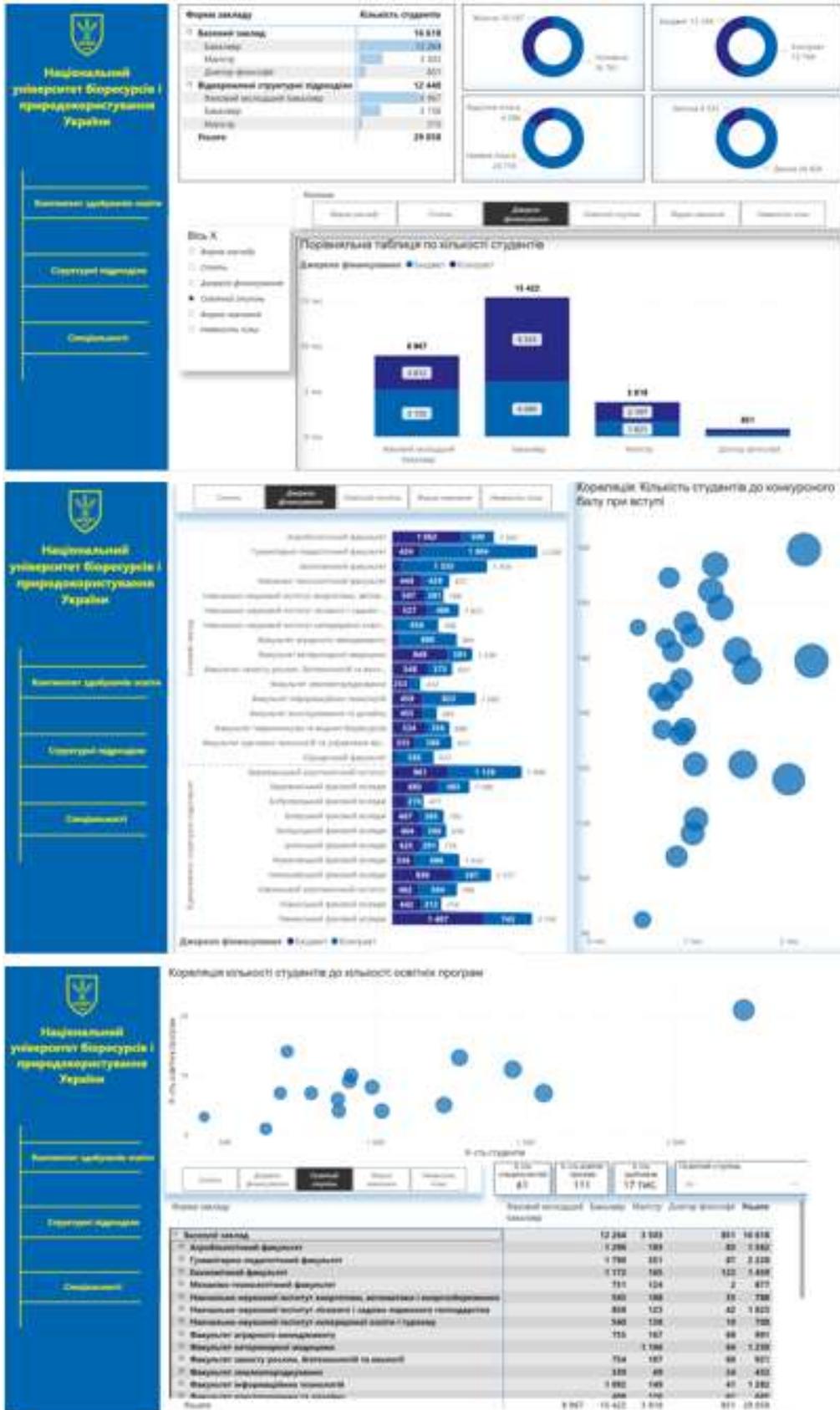


Рис.4.25. Реалізації дашбордів інформаційної технології освітньої аналітики для НУБІП України

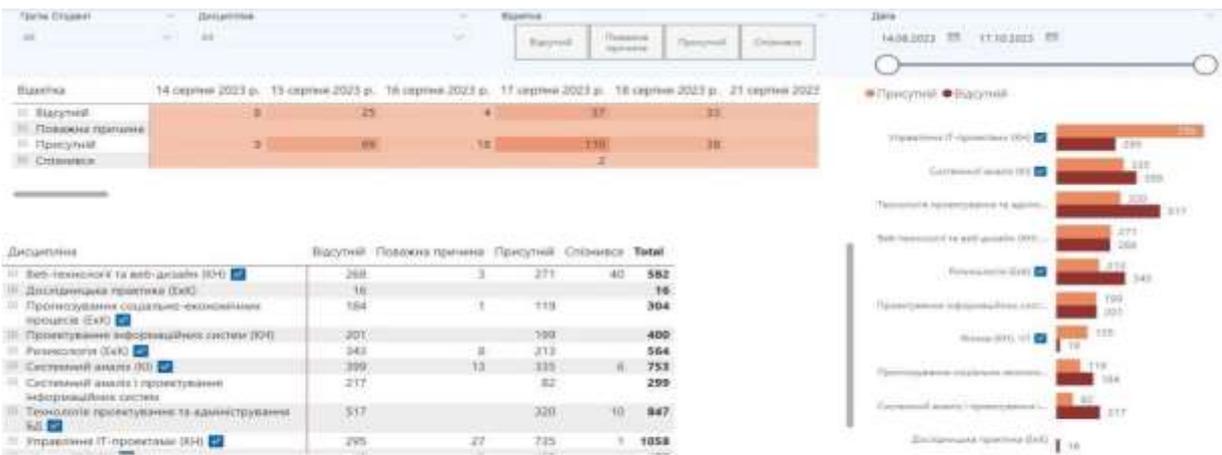
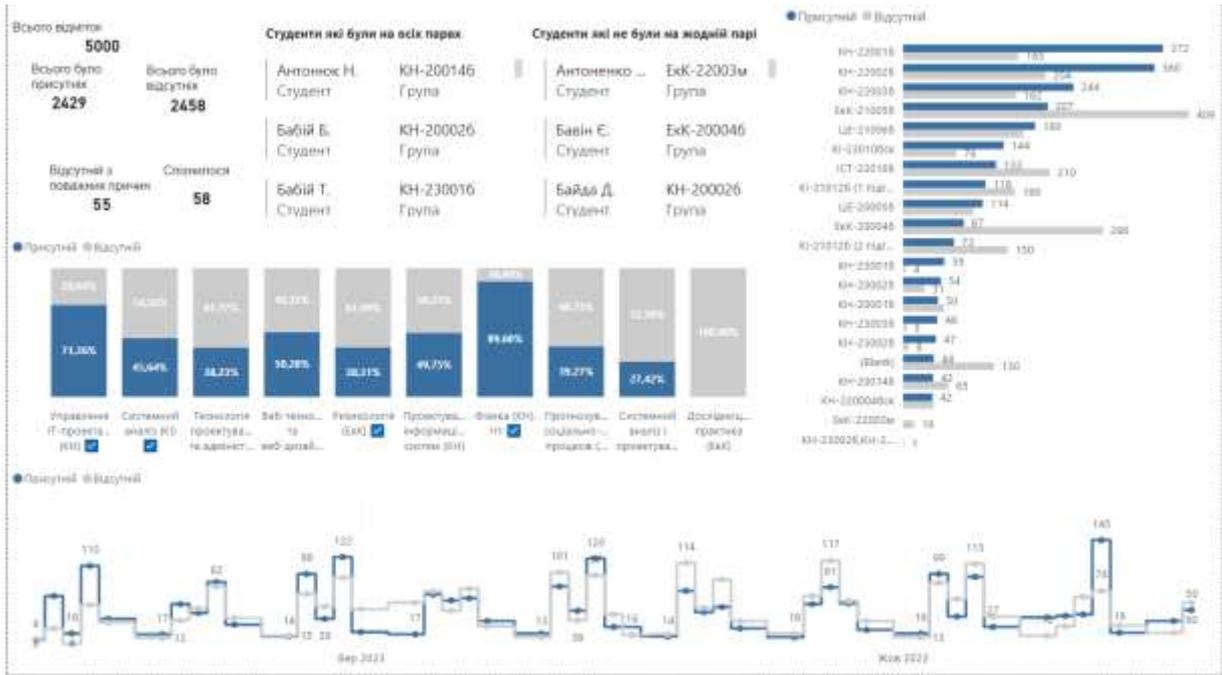
Для реалізації функції публікації інтерактивної аналітики за розробленою інформаційною технологією використовується можливість вбудованого публікування дашбордів Power BI (Publish to Web) та інтеграція з університетськими порталами через iframe або REST API. Такий підхід дозволяє динамічно оновлювати дані, не потребуючи ручного редагування сторінок сайту.

Також результатом роботи нашої технології є дашборди для службового користування деканів та керівництва відділів забезпечення якості освіти, що включають інформацію про: динаміку успішності студентів за факультетами, спеціальностями, освітніми програмами, курсами, формами навчання та фінансування, за окремими освітніми компонентами; аналітику відвідуваності курсів у LMS, доступності навчальних матеріалів; тренди присутності та залученості здобувачів освіти.

Такі результати аналітики дають можливість керівникам підрозділів оперативно реагувати на індикатори ризиковості щодо успішності здобувачів освіти. Приклади відповідної аналітики по факультету інформаційних технологій подано на рис.4.26.

Таким чином, впровадження системи візуалізації відкритих даних університету є складовою цифрової трансформації ЗВО, що відповідає принципам відкритої науки та сприяє розвитку аналітичної культури в Україні.

Конкретними доказами ефективності і реальними практиками використання такого підходу виступають дашборди МОН/Інституту освітньої аналітики, спрямовані на публічне представлення даних. Інтерактивний дашборд МОН (також відомий як Education Indicators Dashboard) охоплює ключові показники освіти (вступна кампанія, бюджет, соціальний профіль, динаміка змін). Він розміщений на сайті МОН; щодинамічно оновлюється через API системи АІКОМ. Містить наступні модулі: національний, регіональний, локальний рівень, профіль громади, бюджетний огляд, кореляційний аналіз.



Група: ЦЕ-220066 | **Студент:** All | **Дисципліна:** Кроссплатформне програмування... | **Відмітка:** All | **Дата:** All | **Присутній:** 76 | **Відсутній:** 34

Група	Дисципліна	Студент	20 вересня 2024 р.	26 вересня 2024 р.	27 вересня 2024 р.	31 вересня 2024 р.	2 вересня 2024 р.	3 вересня 2024 р.	9 вересня 2024 р.
ЦЕ-220066	Кроссплатформне програмування (Python)	Запорожець А.	Присутній	Присутній	Присутній	Присутній	Присутній	Присутній	Присутній
		Кожаря В.	Присутній	Присутній	Відсутній	Відсутній	Відсутній	Відсутній	Відсутній
		Кондратюк М.	Відсутній	Відсутній	Відсутній	Відсутній	Відсутній	Відсутній	Відсутній
		Рогов Д.	Присутній	Присутній	Відсутній	Присутній	Відсутній	Присутній	Присутній
		Стефанчук В.	Присутній	Присутній	Присутній	Присутній	Присутній	Присутній	Присутній
Терещак І.	Присутній	Присутній	Присутній	Присутній	Присутній	Присутній	Присутній	Присутній	



Рис. 4.26. Дашборди освітньої аналітики для ФІТ НУБІП України

Публікація таких звітів забезпечує реалізацію принципу прозорості діяльності ЗВО, підтримку внутрішньої системи якості освіти, підвищення довіри з боку здобувачів освіти, батьків, партнерів, органів управління, прийняття обґрунтованих управлінських рішень на основі доказових відкритих даних.

Висновки до розділу 4.

Реалізація та апробація розробленої інформаційної технології освітньої аналітики здійснювалася на реальній вибірці здобувачів освіти НУБіП України та охопила повний цикл практичного використання інформаційної технології: від збору даних, формування ознак, навчання моделей до отримання прогнозів, інтерпретації та візуалізації результатів. Ефективність технології підтверджено експериментально, зокрема, у ході апробації показано, що розроблена інформаційна технологія забезпечує високу прогностичну здатність, особливо на завершальних етапах навчання. Так, для Сценарію 3 ми отримали такі оцінки Balanced Accuracy 0.77–0.82, та високу стабільність моделей RandomForest та LightGBM. Це демонструє, що технологія здатна надійно відтворювати залежності у великих освітніх даних та підтримувати процеси академічного моніторингу й управління якістю освіти. Використання трьох сценаріїв («Стартовий», «Проміжний», «Прикінцевий») дозволило оцінити точність прогнозування на різних етапах навчання.

Порівняння результатів довело, що точність моделей закономірно зростає зі збільшенням доступних даних; моделі адекватно реагують на зміну набору ознак і алгоритми демонструють стабільні показники на незалежних підвибірках. Усі ці фактори свідчать про статистичну надійність та достовірність моделювання.

Практичну цінність підтверджено прикладним застосуванням, зокрема, під час апробації технології було доведено її реальну користь для закладу вищої освіти. Сценарій 1 дає можливість раннього виявлення студентів групи ризику,

Сценарій 2 забезпечує можливість оперативного відстеження динаміки успішності протягом семестру, сценарій 3 формує здатність точного прогнозування результатів навчання перед підсумковим контролем.

Візуалізація даних у межах розробленої інформаційної технології виступила не лише засобом представлення результатів аналізу, але й важливим аналітичним етапом, що забезпечує виявлення прихованих закономірностей, аномалій і тенденцій у великих масивах освітніх даних. Використання інтерактивних дашбордів Power BI підтвердило свою ефективність як інструменту підтримки прийняття рішень, підвищуючи прозорість, інтерпретованість і практичну цінність отриманих аналітичних результатів. Таким чином, алгоритмічна та програмна реалізація прогностичних моделей забезпечила інтеграцію методів інтелектуального аналізу даних у функціональну структуру інформаційної технології освітньої аналітики та створила основу для її практичного впровадження в діяльність НУБІП України.

ВИСНОВКИ

У дисертаційній роботі вирішено науково-прикладну задачу розроблення інформаційної технології освітньої аналітики, яка ґрунтується на методах інтелектуального аналізу даних і забезпечує підвищення ефективності прийняття управлінських рішень у закладах вищої освіти на основі комплексного аналізу освітніх даних. Проблему вирішено за рахунок інформаційної технології, яка об'єднує процеси збору, інтеграції, інтелектуального аналізу та візуалізації даних про освітню діяльність у цифровому середовищі і забезпечує підтримку управлінських рішень на основі даних у режимі реального часу. Розв'язання цієї задачі забезпечує можливість системного переходу від описової до прогностичної освітньої аналітики в українських закладах вищої освіти, сприяє розвитку національної інфраструктури освітніх даних і створює основу для впровадження цифрових механізмів управління якістю освіти.

У процесі виконання поставлених завдань отримано такі основні результати:

1. На основі аналізу сучасного стану розвитку освітньої аналітики в Україні та за кордоном визначено ключові тенденції цифровізації освітнього процесу, які полягають у зростанні ролі даних у прийнятті управлінських рішень, поширенні систем Learning Analytics та Educational Data Mining, інтеграції електронних навчальних середовищ із аналітичними інструментами, а також переході від описової звітності до прогностичної аналітики. Встановлено наявність прогалин у практиці використання методів інтелектуального аналізу даних у національних інформаційних системах освіти, що обмежує можливості своєчасного прогнозування успішності та управління якістю освіти. На основі аналізу наукового ландшафту освітньої аналітики, сформованого за публікаціями у провідних наукометричних базах, систематизовано основні науково-методичні підходи, виявлено актуальні проблеми та домінантні напрями розвитку галузі. Досліджено методи, моделі та інформаційні технології збору, обробки й аналізу освітніх даних, зокрема методи класифікації, кластеризації,

регресійного аналізу, пошуку асоціативних правил, а також використання алгоритмів машинного навчання для прогнозування навчальної успішності та виявлення прихованих закономірностей у цифрових освітніх середовищах. Оцінено їх переваги та обмеження з точки зору практичного застосування в системах управління освітою.

2. Уточнено понятійно-категоріальний апарат і структуру процесу освітньої аналітики. Визначено її місце в системі цифрового управління освітою, сформовано функціональну модель, що охоплює етапи збору, попередньої обробки, аналізу, візуалізації та інтерпретації освітніх даних. Запропоновано методи збору, підготовки та інтеграції освітніх даних із різних джерел (LMS(moodle), ІС Деканат та ЄДБО). Обґрунтовано принципові засади побудови, структуру та функції інформаційної технології освітньої аналітики з використанням методів інтелектуального аналізу даних

3. Обґрунтовано вибір методів та моделей інтелектуального аналізу даних для вирішення завдань освітньої аналітики та прогнозування успішності здобувачів освіти. Сформовано математичні моделі освітньої аналітики, які описують взаємозв'язки між параметрами навчальної активності, участі в онлайн-курсах і результатами навчання. Розроблена модель прогнозування навчальної успішності здобувачів освіти дозволила підвищити точність прогнозів ризиків неуспішності до 70-80 %.

4. Розроблено інформаційну технологію освітньої аналітики, яка реалізує запропоновані моделі, методи та алгоритми в інтегрованому цифровому середовищі ЗВО (платформа DigitalNubip). Задача проєктування реалізована через розробку архітектури моделі, структури та функціональних компонентів інформаційної технології. Обґрунтовано доцільність застосування Power BI як інструменту побудови інтерактивних аналітичних звітів і візуалізацій в освітній аналітиці, що дозволяє реалізувати відкриті дашборди для потреб закладів вищої освіти.

5. Здійснено апробацію інформаційної технології освітньої аналітики з використанням методів інтелектуального аналізу даних на реальних навчальних даних здобувачів НУБІП України, які отримані з трьох джерел, що підтвердило коректність алгоритмів, достовірність аналітичних висновків і практичну придатність розробленого рішення. Оцінено ефективність і практичну цінність інформаційної технології. Доведено, що впровадження розробленої технології дозволяє: скоротити час аналітичної обробки освітніх даних на рівні деканатів; підвищити точність прогнозів академічної успішності (неуспішності) до 0,7-0,8 за моделями машинного навчання та відповідними сценаріями; покращити якість управлінських рішень щодо підтримки здобувачів із ризиком неуспішності.

Розроблена інформаційна технологія освітньої аналітики є комплексним науково-технічним рішенням, що інтегрує сучасні методи збору та інтелектуального аналізу даних, моделі та програмні засоби для підтримки прийняття рішень у сфері управління освітнім процесом. Впровадження запропонованих підходів сприяє підвищенню ефективності цифрової трансформації закладів вищої освіти України та формуванню нової культури використання освітніх даних для управління якістю освітнього процесу.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Adebayo, A. O., & Chaubey, M. S.(2019). Data mining classification techniques on the analysis of student's performance. *GSI*, 7(4) ,45–52.
2. Agasisti, T.; Bowers, A. J.(2017). Data Analytics and Decision Making in Education: Towards the Educational Data Scientist as a Key Actor in Schools and Higher Education Institutions. *Handbook of Contemporary Education Economics*, 184.
3. Aggarwal, C. C. (2015). *Data Mining: The Textbook*. Springer.
4. Agrawal, H., & Mavani, H.(2015). Student performance prediction using machine learning. *International Journal of Engineering Research and Technology*, 4(03), 111–113.
5. Akcapinar G.(2015). *Profiling students' approaches to learning through Moodle logs*. Proceedings of the Multidisciplinary Academic Conference, Prague, 242–248.
6. Apie ŠVIS. (2023). Informacinė sistema. URL: <https://www.itc.smm.lt/en/> (дата звернення: 23.10.2025)
7. Arghir, D.-C.(2024). Implementation of learning management systems with generative artificial intelligence functions in the post-pandemic environment. *Information Technologies and Learning Tools*, 100(2), 217-232. <https://doi.org/10.33407/itlt.v100i2.5518>
8. Baker R. S., Yacef K.(2020). The state of educational data mining in 2020: A review and future visions. *Journal of Educational Data Mining*, 12(1), 1–17. DOI: 10.5281/zenodo.3555611.
9. Baker R., Siemens G.(2014). *Education Data Mining and Learning Analytics*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139519526.016>
10. Baker R., Yacef K.(2009). The state of Educational data mining in 2009 : A review and future vision . *Journal of Educational Data Mining*, 1(1), 3–17.
11. Baker S., Inventado P.(2016). *Educational Data Mining and Learning*

Analytics: Potentials and possibilities for online education. *Emergence and Innovation in Digital Learning* / G. Veletsianos, Ed., pp. 83–98.

12. Baker R.(2010). Data Mining for Education. *International encyclopedia of education*, 7 (3), 112–118.

13. Bearman M., & Ajjawi R. (2023). Learning to work with the black box: Pedagogy for a world with artificial intelligence. *British Journal of Educational Technology*, 54, 1160–1173. <https://doi.org/10.1111/bjet.13337>

14. Bie R., Mehmood R., Ruan S., Sun Y., Dawood H. (2016). Adaptive Fuzzy Clustering by Fast Search and Find of Density Peaks. *Personal and Ubiquitous Computing*, 20 (5), 785–793.

15. Blanco T., Casas R., Manchado-Pérez E., Asensio Á., López- Pérez J. M. (2017). From the Islands of Knowledge to a Shared Understanding: Interdisciplinarity and Technology Literacy for Innovation in Smart Electronic. Product Design. *International Journal of Technology and Design Education*, 27 (2), 329–362.

16. Bogнар L., Fauszt T., Nagy G.Z. (2021). Analysis of Conditions for Reliable Predictions by Moodle Machine Learning Models. *International Journal of Emerging Technologies in Learning*, 16(06), 106–121. doi:10.3991/ijet.v16i06.18347

17. Brijesh Kumar Baradwaj, Saurabh Pal.(2020). Mining Educational Data to Analyze Students Performance. *International Journal of Advanced Computer Science and Applications*, Vol. 2, No. 6, 66-69.

18. Brown M. (2012). Data mining techniques. Identify patterns with predictive analytics. IBM. URL: <https://developer.ibm.com/technologies/analytics/articles/ba-data-mining-%20techniques/> (дата звернення 23.10.2025)

19. Campbell J., DeBlois P., Oblinger D.(2007). Academic Analytics: A New Tool for a New Era. *EDUCAUSE Review*, 42, 4, 40–57.

20. Clow D.(2013). An overview of learning analytics. *Teaching in Higher Education*, 18, 6, 683–695.

21. Cobo G., García-Solórzano D., Morán J. A.(2012). *Using agglomerative*

hierarchical clustering to model learner participation profiles in online discussion forums. Proc. 2nd International Conference on Learning Analytics and Knowledge, 311, 323–29.

22. Codd E. F.(1970). A relational model of data for large shared data banks. *Commun, ACM* 13, 6, 377–387. <https://doi.org/10.1145/362384.362685>

23. Data Analytics Market Size , Share & Growth Report, 2022-2030. URL: <https://www.grandviewresearch.com/industry-analysis/data-analytics-market-report> (дата звернення: 23.10.2025)

24. Data Mining Tools Market Size, Share, Trends and Industry Analysis. MarketsandMarkets (2023) URL: <https://www.marketsandmarkets.com/Market-Reports/data-mining-tools-market-259286296.html> (дата звернення: 23.10.2025)

25. Diaz-Choque M., Chamorro O., Ortega-Galicio O., Arévalo-Tuesta J., Cáceres-Cayllahua E., Dávila-Laguna R., Aybar-Bellido I., Siguas-Jerónimo Y.(2023). Contributions of Data Mining to University Education, in the Context of the Covid-19 Pandemic: A Systematic Review of the Literature. *International Journal of Online and Biomedical Engineering*, №19, 16-33. <https://doi.org/10.3991/ijoe.v19i12.40079> .

26. Doug A. (2015). Data Mining, URL: <https://laits.utexas.edu/~anorman/BUS.FOR/course.mat/Alex/> (дата звернення: 23.10.2025)

27. Drachsler H., Greller W. (2012). *The pulse of Learning Analytics understandings and expectations from the stakeholders*. Proceedings of the 2nd international conference on Learning Analytics and Knowledge, Vancouver, Canada, 120–129.

28. Drigas, A. S., Leliopoulos, P.(2014). The Use of Big Data in Education. *International Journal of Computer Science Issues*, 11 (5), 58.

29. Duval E., Verbert K .(2012). Learning Analytics. *ELEED: E-Learning and Education*, 8(1). 0009-5-33367.

30. European Commission.(2020). Digital Education Action Plan 2021–2027 Luxembourg: Publications Office of the European Union, URL::

<https://education.ec.europa.eu/focus-topics/digital-education/actions> (дата звернення: 30.06.2025)

31. Farias-Gaytan, S., Aguaded, I., Ramirez-Montoya, M.-S. (2023). Digital transformation and digital literacy in the context of complexity within higher education institutions: a systematic literature review. *Humanities and Social Sciences Communications*, 10 (1), 386.

32. Fayyad, U. M., G. Piatetsky-Shapiro, P. Smyth, and Ft. Uthurusamy (1996). *Advances in Knowledge Discovery and Data Mining*, (AKDDM), AAAI/MIT Press.

33. Feng T., Shibin W., Cheng Z., Qinghua Z.(2008). *Research on e-learner personality grouping based on fuzzy clustering analysis*. Proc. 12th Int. Conf. Computer Supported Cooperative Work in Design, 1035–1040.

34. Fenu G., Marras M., Meles M.(2017). A Learning Analytics Tool for Usability Assessment in Moodle Environments. *Journal of e-Learning and Knowledge Society*,13(3), 23-34. DOI: 10.20368/1971-8829/1388

35. Ferguson R. (2012). Learning Analytics: drivers, developments and challenges. *International Journal of Technology Enhanced Learning*, 4, № 5–6, 304–317.

36. Ferguson R., Hoel T., Scheffel M., Drachsler H.(2016). Guest editorial: Ethics and privacy in learning analytics. *Journal of learning analytics*, 3,1, 5–15.

37. Forbes. (2019). Big Data Analytics: A Comprehensive Overview.

38. Foster Provost, Tom Fawcett. (2013). *Data Science for Business: What you need to know about data mining and data-analytic thinking*. O'Reilly Media, 408

39. Gabroveanu M., Diaconescu I.(2008). *Extracting semantic annotations from Moodle data*. In Proceedings of the 2nd East European Conference on Rule-Based Applications, Germany, 428: 1–7.

40. Gartner. (2019). Top 10 Cloud Myths for Data Mining. <https://c.digitalisationworld.com/news/35773/gartner-highlights-the-top-10-cloud-myths>

41. Gašević D., Dawson S., Siemens G.(2015). Let's not forget: Learning analytics are about learning. *TechTrends*, 59,1, 64–71.
42. Glazunova O., Morze N., Golub B., Voloshyna T., Parhomenko O. (2020). *Learning style identification system: Design and data analysis*. CEUR Workshop Proceedings, 2732: 793–807
43. Gowda Sujith M., Baker Ryan S., Corbett Albert T., Rossi Lisa M. (2012). Towards automatically detecting whether student learning is shallow. *International Journal of Artificial Intelligence in Education*, 23 (1-4), 50-70
44. Greller W., Ebner M., Schön M.(2014). *Learning Analytics: From Theory to Practice–Data Support for Learning and Teaching*. *Computer Assisted Assessment. Research into E-Assessment* / M. Kalz, R. Marco, Eds. New York, USA, Springer, 79–87.
45. Han, J., Kamber, M., & Pei, J. (2011). *Data Mining: Concepts and Techniques*. Morgan Kaufmann.
46. Herrera, O. L.(2006). *Investigation of the role of pre- and post-admission variables in undergraduate institutional persistence, using a Markov student flow model*. PhD Dissertation, North Carolina State University, USA.
47. Hlazunova O., Klymenko N., Mokriiev M., Nehrey M., Klymenko, Ye. (2025). Data Analysis Technologies for Enhanced Educational Processes: A Case Study Using the Moodle LMS. *Lecture Notes on Data Engineering and Communications Technologies*, 242, 670 – 682.
48. Ian H. Witten (2011). *Data Mining: Practical Machine Learning Tools and Techniques* / Ian H. Witten, Eibe Frank, Mark A. Hall (3rd ed.) Morgan Kaufmann, 664.
49. Ifenthaler D., Yau J.YK. (2020). Utilising learning analytics to support study success in higher education: a systematic review. *Education Tech Research Dev*, 68: 1961–1990 <https://doi.org/10.1007/s11423-020-09788-z>
50. Information technology. Encyclopaedia Britannica
51. ISO/IEC 38500:2015. Information technology –Governance of IT for the

organization. Geneva: International Organization for Standardization, 2015. 20 p.

52. Jun J.(2005). *Understanding dropout of adult learners in e-learning*. PhD Dissertation, The University of Georgia, USA.

53. Kabakchieva D. (2013). Predicting Student Performance by Using Data Mining Methods for Classification. *Cybernetics and Information Technologies*, 13. 10.2478/cait-2013-0006.

54. Kasthuriarachchi, K.T.S., Liyanage, S.R.(2019). Predicting Students' Academic Performance Using Utility Based Educational Data Mining. *Lecture Notes in Electrical Engineering*, 542. _

55. Khojastehfar A., Abarbekouh H., Safai S. and Fatemi O.(2023). *Analysis of Teaching and Learning in Moodle with the Help of Visualization with Power BI: Case Study of e-Learning Course of University of Tehran*. ICeLeT, 1-7.

56. Kumar R., Chen W.(2020). Using Power BI in Higher Education Analytics. *Journal of Learning Analytics*, 7(3), 112–125.

57. Kumar V. (2018). *Data Mining for Business Analytics: Concepts, Techniques, and Applications*. Pearson Education Limited.

58. Kuzminska O., Morze, N., Osadchyi, V. (2023). Digitization of learning environment. *Journal of Physics*, 2611(1), 012024

59. Lakhno V., Akhmetov B., Makulov K., Tynymbayev B., Tsiutsiura S., Tsiutsiura M., & Chubaievskiy, V. (2024). Formation of Models for Registering Systemic Processes in The Digital Educational Environment of the University Based on Log File Analysis. *International Journal of Electronics and Telecommunications*, 70(2), 389-396.

60. Lang C., Macfadyen L., Slade S., Prinsloo P., Sclater N.(2018). *The complexities of developing a personal code of ethics for learning analytics practitioners: implications for institutions and the field*. Proc. 8th Int. Conf. Learning Analytics and Knowledge, Sydney, Australia, 436–440.

61. Laudon K. C., Laudon J. P.(2020). *Management Information Systems: Managing the Digital Firm* (16th ed.). Harlow: Pearson Education, 624

62. Lauria E. (2012). *Mining academic data to improve college student retention: An open source perspective*. Proceedings of the 2nd international conference on Learning Analytics and Knowledge, Vancouver, Canada,. 139–142.
63. Leitner P., Ebner M.(2019). Learning Analytics Challenges to Overcome in Higher Education Institutions. *Utilizing Learning Analytics to Support Study*, 7, 91–104
64. Lopez A., Sánchez J., Rojas L.(2022). Data-Driven Decision-Making in Universities: The Role of Power BI. *Educational Technology & Society*, 25(2), 77–89.
65. Macfadyen L., Dawson S.(2012). Numbers Are Not Enough. Why e-Learning Analytics Failed to Inform an Institutional Strategic Plan. *Educational Technology & Society*, 15, 3,149–163
66. Manhiça, R., Santos, A., Cravino, J.(2022). The use of artificial intelligence in learning management systems in the context of higher education : *Systematic literature review, 1-6*. <https://doi.org/10.23919/CISTI54924.2022.9820205>
67. Manohar A., Gupta P., Priyanka V., Uddin M. (2016). Utilizing Big Data Analytics to Improve Education, *ASEE*.
68. Market Research.(2025). Data Mining Tools Market Outlook 2025-2034.URL <https://www.marketresearch.com/OG-Analysis-v3922/Data-Mining-Tools-Outlook-Share-42770509> (дата звернення: 15.12.2025)
69. Ministry of Federal Education and Professional Training. URL: <http://www.moent.gov.pk/> . (дата звернення: 15.12.2025).
70. Mitchell, T. M. (1997). *Machine Learning*. McGraw-Hill.
71. Moodle piugins. URL: <https://moodle.org/local/plugins/?q=>
72. Moodle Statistics. URL: <https://stats.moodle.org/>
73. Morze N., Kuzminska O., Glazunova O., Korolchuk V., Mokriiev M., Varchenko-Trotsenko L., Zolotukha R.(2022). *Moodle Tools for Educational Analytics of the Use of Electronic Resources of the University's Portal*. Proceedings of the 1st Symposium on Advances in Educational Technology, 20 (2):. 444-451. ISSN 978-989-758-558-6
74. Morze, N. V., Smyrnova-Trybulska, E., & Glazunova, O. (2021). *Design*

of a University Learning Environment for SMART Education. In I. Management Association (Ed.), *Research Anthology on Preparing School Administrators to Lead Quality Education Programs* (pp. 518-545). IGI Global Scientific Publishing. <https://doi.org/10.4018/978-1-7998-3438-0.ch024>

75. Nam S., Lonn S., Brown T., Davis C., Koch D.(2014). *Customized course advising: investigating engineering student success with incoming profiles and patterns of concurrent course enrollment*. Proc. 4th Int. Conf. Learning Analytics and Knowledge, Indianapolis, USA, 6–25.

76. Nguyen H., Bui T. (2021). Empowering Data Literacy in Higher Education with Microsoft Power BI. *International Journal of Educational Technology*, 18, 29–40.

77. Nikolaienko S.M., Shynkaruk V.D., Kovalchuk V.I. & Kocharyan A.B. (2017). Використання Big Data в освітньому процесі сучасного університету. *Information Technologies and Learning Tools*, 60(4), 239. [10.33407/itlt.v60i4.1681](https://doi.org/10.33407/itlt.v60i4.1681).

78. Kryvoruchko O., Shestak Ya., Zavorodnya E., Fesenko A. (2025). *Higher education cyber resilience: Intelligent protection of educational, administrative and resource systems*. CSDP'2025, Lviv, Ukraine.

79. OECD (2021). *Artificial Intelligence and the Future of Skills*. URL: <https://www.oecd.org/en/about/projects/artificial-intelligence-and-future-of-skills.html> (дата звернення: 15.12.2025)

80. Okike E., Morogosi M. (2020). Educational Data Mining for Monitoring and Improving Academic Performance at University Levels. *International Journal of Advanced Computer Science and Applications*, 11. <https://dx.doi.org/10.14569/IJACSA.2020.0111171> .

81. Osofisan A., Adeyemo O., Oluwasusi S.(2022). Empirical Study of Decision Tree and Artificial Neural Network Algorithm for Mining Educational Database. *African Journal of Computing & ICT*, 7, No. 2, 191-193

82. Pal N., Dahiya O. (2022). *Role of Learning Management System for Evaluating Students' progress in Learning Environment*. Proceedings of 5th IC3I, 1800–1806
83. Parsaye K. A. (1998). Characterization of Data Mining Technologies and Processes. *The Journal of Data Warehousing*, 1, 12-24
84. Pistilli M., Willis J., Campbell J. (2014) Analytics through an institutional lens: Definition, theory, design, and impact. *Learning Analytics*, 79–102.
85. Poon L.K.M., Kong SC., Yau T.S.H., Wong M., Ling, M. (2017). Learning Analytics for Monitoring Students Participation Online: Visualizing Navigational Patterns on Learning Management System. Blended Learning. New Challenges and Innovative Practices. ICBL. *Lecture Notes in Computer Science*, 10309. https://doi.org/10.1007/978-3-319-59360-9_15
86. Provost F. & Fawcett T. (2013). *Data Science for Business: What you need to know about data mining and data-analytic thinking*. O'Reilly Media.
87. Qian G., Wu Y., Ferrar, D., Qiao P., Hollande F. (2016). Semisupervised Clustering by Iterative Partition and Regression with Neuroscience Applications. *Computational intelligence and neuroscience*, 1, 4037380.
88. Queensland Government. (2024). Education and training. URL: <https://www.qld.gov.au/education/further-ed/vet> (дата звернення: 15.12.2025)
89. Rienties B., Cross S., Zdrahal Z. (2018). Implementing a learning analytics intervention and evaluation framework: What works? *Educational Technology Research and Development*, 66, 103–126.
90. Romero C., Ventura S.(2023). Data mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 3. № 1, 12–27
91. Saa A. A.(2016). Educational Data Mining & Students' Performance Prediction. *International Journal of Advanced Computer Science and Applications*, 7 (5), 212–220.
92. Sajid W. A., Mirzah N., Mustafa F. M. and Shabala Y. (2024). *Educational Institution Management Information System*. 35th Conference of Open

Innovations Association (FRUCT), Tampere, Finland, 625–632 doi: 10.23919/FRUCT61870.2024.10516390.

93. SAS (2023). Leading in data and AI 2023. Annual Report and Corporate Overview

94. Sclater N.(2017). *Learning Analytics Explained*. Routledge, 240 p.

95. Siafis, V., Rangoussi M. (2021). *Visualization of educational data mined from E-Learning platforms: A comparative evaluation of tools*. ICERI–2021, Seville, Spain. <https://doi.org/10.21125/iceri.2021>

96. Siemens G., Gašević D., Dawson S. (2015). Preparing for the digital university: a review of the history and current state of distance, blended, and online learning. <https://doi.org/10.13140/RG.2.1.3515.8483> .

97. Siemens G., Baker, R. S. J. d. (2012). *Learning analytics and educational data mining: Towards communication and collaboration*. In Proceedings of the 2nd International Conference on Learning Analytics and Knowledge, 252–254.

98. Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE Review*, 46(5), 30–40.

99. Stewart M. F., Chisholm C. U.(2012). Comparative analysis of emotional competency within distinct student cohorts, *Global J. of Engng. Educ.*, V. 14, 2, 163–169.

100. TAALIM (2018) Database User Manual for Training Provider (TP). URL: <https://www.tvf.af> (дата звернення: 15.03.2025)

101. Tan, P. N., Steinbach, M., & Kumar, V. (2021). *Introduction to Data Mining*. Pearson Education Limited.

102. Tsai Y., Gasevic D.(2017). *Learning analytics in higher education — challenges and policies: a review of eight learning analytics policies*. Proc. 7th Int. Conf. Learning Analytics and Knowledge. Vancouver, Canada, 233–242.

103. Tseng S., Tsao Y., Yu L., Chan C., Lai K .(2016). Who will pass? Analyzing learner behaviors in MOOCs. *Research and Practice in Technology Enhanced Learning*, 11, 8.

104. Tulasi B.(2019). Significance of Big Data and Analytics in Higher Education. *International Journal of Computer Applications*, 68 (14).
105. Two Crows Corporation. (1999). *Introduction to Data Mining and Knowledge Discovery*(3rd ed.). TwoCrows Corporation. ISBN 1-892095-02-5
106. UNESCO (2021). Recommendation on the Ethics of Artificial Intelligence. Paris: UNESCO, 2021.
107. UNESCO Institute for Statistics.(2023). *Data for Education: A Guide for Policymakers to Leverage Education Data*. Montreal: UIS, 72 .
108. Van Barneveld A., Arnold K., Campbell J.(2012). Analytics in higher education: Establishing a common language. *Educause Learning Initiative*, 1(1), 1–11.
109. Volkova N., Rizun N., Nehrey M.V.(2019). *Data science: Opportunities to transform education*. CEUR Workshop Proceedings, 2433: 48-73.
110. Williamson B.(2017). *Introduction: Learning machines, digital data and the future of education*. SAGE Publications Ltd. <https://doi.org/10.4135/9781529714920> .
111. Witten I. H., Frank E., & Hall M. A. (2016). *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann.
112. Wong, W., Fu, A. W. (2002). Incremental Document Clustering for Web Page Classification. *In Enabling Society with Information Technology*. Springer Japan, 2002; pp 101–110.
113. Yan-lin L. L. Z. (2010). The Application of the Internet of Things in Education. *Modern Educational Technology*, 2 (005).
114. Бондарчук А., Жебка В., Корецька В., Шилкіна А. Порівняльна характеристика web-орієнтованих інструментів автоматизації освітнього процесу в умовах цифрової трансформації. *Публічно-управлінські та цифрові практики*, 1, 13-21.
115. Глазунова О., Кравченко В., Саяпіна Т., Волошина Т., & Корольчук, В.(2024). Вплив технології мікронавчання на мотивацію та академічні досягнення здобувачів вищої освіти. *Електронне наукове фахове видання*

“Відкрите освітнє е-середовище сучасного університету”, 17, 30–41.

<https://doi.org/10.28925/2414-0325.2024.173>

116. Глазунова О., Клименко, Є., Волошина, Т., Мокрієв М., Вороненко О. (2024). Освітня аналітика в університетах: інструменти для аналізу та прогнозування. *Телекомунікаційні та інформаційні технології*, 2, 49-59

117. Глазунова О., Клименко Є. (2025). Інформаційна технологія аналітики освітніх даних. *Наука і техніка сьогодні*, 2(43)

[https://doi.org/10.52058/2786-6025-2025-2\(43\)-1145-1155](https://doi.org/10.52058/2786-6025-2025-2(43)-1145-1155)

118. Горошкіна О.(2021). Цифрова трансформація вищої освіти України: аналітичний звіт. Київ: Інститут вищої освіти НАПН України, 2021,72 с.

119. Дем’яненко І. М., Семенюк С. В. (2020) *Інформаційні технології в освіті: навчальний посібник*. Київ: КНЕУ, 268 с.

120. Клименко Є., Глазунова О. (2023). *Врахування компетентісного підходу в освітніх траєкторіях здобувачів освіти*. Collection of Scientific Papers «SCIENTIA», Chicago, USA, 127–128

121. Клименко Є., Глазунова О. (2022). *Компетентісний підхід в освітніх траєкторіях випускників ІТ спеціальностей ЗВО на базі платформи Moodle*. Економіко-правовий розвиток сучасної України : матеріали XI Всеукраїнської наукової конференції студентів, аспірантів та молодих вчених, Одеса, С. 197-199. <http://dspace.onua.edu.ua/handle/11300/23228>

122. Клименко Є.(2024). Аналіз освітніх даних у системах електронного навчання.Збірник тез Міжнародної науково-практичної конференції молодих вчених «інформаційні технології: економіка, техніка, освіта ‘2024», К.:НУБіП України, с.79-81

123. Клименко Є.(2025) Розробка інформаційних систем прогнозу аналітики успішності здобувачів освіти. Матеріали Міжнародної наукової інтернет-конференції Інформаційне суспільство: технологічні, економічні та технічні аспекти становлення, 2025 випуск 98.

124. Клименко Є., Глазунова, О. (2025). Архітектура інформаційної

технології освітньої аналітики з використанням інтелектуального аналізу даних. *Інформаційні технології та суспільство*, 2 (17), 69-75. <https://doi.org/10.32689/maup.it.2025.2.10>

125. Клименко Є., Глазунова О. (2024). Moodle BIG DATA analytics за допомогою POWER BI. *Grail of Science*, №35, 201–203. <https://doi.org/10.36074/grail-of-science.19.01.2024.035>

126. Клименко Є., Глазунова, О. (2024). Методи інтелектуального аналізу освітніх даних у системах електронного навчання. *Інформаційні технології та суспільство*, 2 (13), 34-40. <https://doi.org/10.32689/maup.it.2024.2.5>

127. Ковальчук Ю. (2015) Пошук, отримання й аналіз даних в освіті: сучасний стан і перспективи розвитку, *Інформаційні технології і засоби навчання*, 50(6), 152–164.

128. Коляда М. (2014). Використання систем видобутку знань Data Mining для прогнозування педагогічних процесів та явищ. *Вісник післядипломної освіти*, 11, 80–88.

129. Кремень, В. Г., Биков, В. Ю., Ляшенко, О. І., Литвинова, С. Г., Луговий, В. І., Мальований, Ю. І., Топузов, О. М. (2022). Науково-методичне забезпечення цифровізації освіти України: стан, проблеми, перспективи. *Вісник Національної академії педагогічних наук України*, 4(2), 1–49. <https://doi.org/10.37472/v.naes.2022.4223>

130. Криворучко О., Захаров Р. (2024). Моделювання інформаційної технології формування компетентностей здобувачів освіти. *Кібербезпека: освіта, наука, техніка*, 4(24), 196–204.

131. Петренко С. В. (2017). Оптимізація й аналіз результатів використання LMS Moodle у системі змішаного навчання в університеті. *Інформаційні технології і засоби навчання*, 61(5), 140–150.

132. Погоріла, Ю. В., & Василенко, В. Ю. (2024). Data Mining в освітній сфері: особливості використання. *Прикладні аспекти сучасних міждисциплінарних досліджень*, 48-51.

133. Пригодій М.А., Гуржій А.М., Лупаренко Л.А., Гуменний О.Д., Белан В.Ю.(2025). *Проектування та застосування цифрової платформи професійної підготовки спеціалістів: інформаційно-аналітичні матеріали*. Київ: Інститут професійної освіти НАПН України, 37 с.
134. Ситник В. Ф.(2007) Інтелектуальний аналіз даних (дейтамайнінг): навч. посіб. К.: КНЕУ. 376 с.
135. Співак С.М., Білоус В.В., Горбатовський Д.В., Бондарчук, А.П. (2025) Адаптація освіти до ринку 3D - графіки за допомогою ШІ. *Телекомунікаційні та інформаційні технології*, 89 (4), 215-221
136. Староста В. І.(2021). MOODLE до, під час і після пандемії covid-19: використання студентами бакалаврату та магістратури. *Електронне наукове фахове видання "Відкрите освітнє е-середовище сучасного університету"*, 10, 216–230. <https://doi.org/10.28925/2414-0325.2021.1018>
137. Томашевський О. М., Цегелик Г. Г., Вітер М. Б., Дудук В. І. Інформаційні технології та моделювання бізнес-процесів. навч. посіб. Київ. Центр навчальної літератури (ЦУЛ), 296 с.
138. Юринець Р. С.(2018). Інформаційні системи і технології в управлінні. Львів: ЛНУ ім. Івана Франка, 312 с.

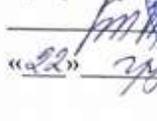
ДОДАТКИ

Додаток А

Акт впровадження результатів роботи в освітній процес

ЗАТВЕРДЖУЮ

Проректор з наукової роботи
та інноваційної діяльності
Національного університету біоресурсів
і природокористування України
доктор сільськогосподарських наук,
професор


Оксана ТОНХА
«22» грудня 2025 р.



ПОГОДЖЕНО

Керівник центру забезпечення якості
освіти Національного університету
біоресурсів і природокористування
України
кандидат педагогічних наук, доцент


Ярослав РУДИК
«22» грудня 2025 р.

А К Т

про впровадження/використання результатів
дисертації на здобуття ступеня доктора філософії
у навчальний процес

Цим актом стверджується, що результати дисертації на тему: **Інформаційна технологія освітньої аналітики на основі методів інтелектуального аналізу даних**,

яку представлено на здобуття ступеня доктора філософії
з галузі знань 12 – Інформаційні технології та спеціальності 122 – Комп'ютерні науки,
виконаної Климченко Євгенієм Олеговичем,
впроваджено у навчальну програму під час викладання дисциплін «Business intelligence системи» та «Розумні цифрові технології в освіті» на кафедрі інформаційних систем і технологій у процесі підготовки фахівців ступеня вищої освіти «Бакалавр» зі спеціальності Інформаційні системи та технології та освітнього ступеня «Магістр» освітньо-професійної програми Інформаційно-комунікаційні технології в освіті у Національному університеті біоресурсів і природокористування України.

При викладанні дисципліни «Business intelligence системи» здобувачі освіти ознайомлені з технологіями розробка та використання панелей управління у BI системах, познайомились з розробкою BI систем, опанувати моделювання даних в Power BI з використанням датасетів, зібраних в ході дисертаційного дослідження, знайомились з візуалізацією результатів аналізу освітніх даних засобами Python та Power BI, що є складовою інформаційної технології, розробленої здобувачем ступеня доктора філософії.

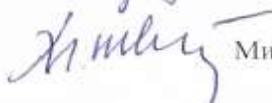
При викладанні дисципліни «Розумні цифрові технології в освіті» було доведено основи освітньої аналітики та цифрової трансформації освіти, запропоновані розроблені підходи аналізу ризиків академічної неуспішності та освітніх втрат, забезпечено практичні навички щодо збору та обробки освітніх даних з різних джерел, ознайомлено з створенням інтерактивних дашбордів для освітнього моніторингу, розроблених під час дисертаційного дослідження.

Декан факультету інформаційних технологій
д.т.н., професор

Заступник декана з наукової роботи
д.с.н., професор

Завідувач кафедри
к.е.н., доцент, почесний професор


Ігор БОЛБОТ

Володимир КРАВЧЕНКО

Михайло ШВИДЕНКО

Додаток Б

Акт впровадження результатів роботи у виробництво**ЗАТВЕРДЖУЮ**

Проректор з науково-педагогічної роботи
та адміністративно-господарської
діяльності, кандидат економічних наук,
доцент

Володимир С ВТУШЕНКО
Володимир С ВТУШЕНКО

«06» *серпня* 2026 р.

**ПОГОДЖЕНО**

Начальник інформаційно-
обчислювального центру

Віктор ТEPЛЮК
Віктор ТЕРЛЮК

«06» *серпня* 2026 р.

А К Т

**про впровадження/використання результатів
дисертації на здобуття ступеня доктора філософії
у виробничий процес**

Цим актом стверджується, що результати дисертації на тему: **Інформаційна технологія освітньої аналітики на основі методів інтелектуального аналізу даних**, яку подано на здобуття ступеня доктора філософії з галузі знань 12 – Інформаційні технології та спеціальності 122 – Комп’ютерні науки, виконаної Клименком Євгенієм Олеговичем, впроваджено як модуль освітньої аналітики в інформаційну систему управління освітнім процесом Nubip Digital.

Метою дисертаційного дослідження є обґрунтування та розробка інформаційної технології освітньої аналітики, яка забезпечує багаторівневий аналіз та інтерпретацію освітніх даних на основі методів інтелектуального аналізу даних з метою прогнозування результатів навчання здобувачів вищої освіти. Запропоновані методи та розроблену інформаційну технологію освітньої аналітики застосовано для прогнозування успішності студентів факультету інформаційних технологій НУБІП України та візуалізації звітів про освітню діяльність університету.

Декан факультету
інформаційних технологій
д.т.н., професор

Ігор БОЛБОТ
Ігор БОЛБОТ

Заступник декана
з наукової роботи
д.е.н., професор

Володимир КРАВЧЕНКО
Володимир КРАВЧЕНКО

Код реалізації моделей та обчислення метрик ефективності

1. Метод автоматичного підбору гіперпараметрів для моделей машинного навчання *RandomizedSearchCV*

```

param_grids = {
    "DecisionTree": {
        "max_depth": [None, 5, 10, 20, 30],
        "min_samples_split": [2, 4, 8, 12],
        "min_samples_leaf": [1, 2, 4, 6],
        "criterion": ["gini", "entropy"],
        "splitter": ["best", "random"]
    },
    "RandomForest": {
        "n_estimators": [100, 200, 300, 500],
        "max_features": ["sqrt", "log2", None],
        "min_samples_split": [2, 4, 8],
        "min_samples_leaf": [1, 2, 4, 6],
        "criterion": ["gini", "entropy"],
        "bootstrap": [True, False]
    },
    "LightGBM": {
        "n_estimators": [100, 250, 500, 1000],
        "learning_rate": [0.01, 0.02, 0.05, 0.1],
        "max_depth": [-1, 5, 10, 20],
        "num_leaves": [15, 31, 63],
        "boosting_type": ["gbdt", "dart", "rf"]
    },
    "LogReg": {
        "C": np.logspace(-3, 3, 7),
        "solver": ["lbfgs", "liblinear", "saga"],
        "penalty": ["l2", "none"]
    }
}

search = RandomizedSearchCV(
    estimator=model,
    param_distributions=param_grid,
    n_iter=20,
    scoring='roc_auc'
    cv=5,
    random_state=1234,
    n_jobs=-1,
    verbose=1
)

```

2. Результати реалізації методу автоматичного підбору гіперпараметрів для моделей машинного навчання *RandomizedSearchCV* з використанням метрики *'roc_auc'*

```

- 🔍 Searching best parameters for: DecisionTree
- Fitting 5 folds for each of 20 candidates, totalling 100 fits
- ✓ Best params: {'splitter': 'best', 'min_samples_split': 4,
  'min_samples_leaf': 2, 'max_depth': 5, 'criterion': 'gini'}
- 📊 Best score: 0.7207775355548635
-
- 🔍 Searching best parameters for: RandomForest
- Fitting 5 folds for each of 20 candidates, totalling 100 fits
- ✓ Best params: {'n_estimators': 200, 'min_samples_split': 4,
  'min_samples_leaf': 4, 'max_features': 'sqrt', 'criterion': 'gini',
  'bootstrap': True}
- 📊 Best score: 0.7407725526834839
-
- 🔍 Searching best parameters for: LightGBM
- Fitting 5 folds for each of 20 candidates, totalling 100 fits
- ✓ Best params: {'num_leaves': 15, 'n_estimators': 250, 'max_depth': 10,
  'learning_rate': 0.01, 'boosting_type': 'dart'}
- 📊 Best score: 0.7373181770995536
-
- 🔍 Searching best parameters for: LogReg
- Fitting 5 folds for each of 20 candidates, totalling 100 fits
- ✓ Best params: {'solver': 'liblinear', 'penalty': 'l2', 'C':
  np.float64(0.01)}
- 📊 Best score: 0.6864150316619951

```

3. Прогнози реалізації моделей машинного навчання для 1-го підходу

```

=== DecisionTree ===
Ймовірності для 2-го тестового зразка (факт. клас 0):
  Клас '0' (індекс 0): 0.4623
  Клас '1' (індекс 1): 0.5377
Передбачений клас: 1
Впевненість у прогнозі: 0.5377

=== RandomForest ===
Ймовірності для 2-го тестового зразка (факт. клас 0):
  Клас '0' (індекс 0): 0.8310
  Клас '1' (індекс 1): 0.1690
Передбачений клас: 0
Впевненість у прогнозі: 0.8310

=== LightGBM ===
Ймовірності для 2-го тестового зразка (факт. клас 0):
  Клас '0' (індекс 0): 0.6626
  Клас '1' (індекс 1): 0.3374
Передбачений клас: 0
Впевненість у прогнозі: 0.6626

```

```

=== LogReg ===
Ймовірності для 2-го тестового зразка (факт. клас 0):
  Клас '0' (індекс 0): 0.4576
  Клас '1' (індекс 1): 0.5424
Передбачений клас: 1
Впевненість у прогнозі: 0.5424
-----

=== DecisionTree ===
Ймовірності для 15-го тестового зразка (факт. клас 1):
  Клас '0' (індекс 0): 0.0000
  Клас '1' (індекс 1): 1.0000
Передбачений клас: 1
Впевненість у прогнозі: 1.0000

=== RandomForest ===
Ймовірності для 15-го тестового зразка (факт. клас 1):
  Клас '0' (індекс 0): 0.4587
  Клас '1' (індекс 1): 0.5413
Передбачений клас: 1
Впевненість у прогнозі: 0.5413

=== LightGBM ===
Ймовірності для 15-го тестового зразка (факт. клас 1):
  Клас '0' (індекс 0): 0.3781
  Клас '1' (індекс 1): 0.6219
Передбачений клас: 1
Впевненість у прогнозі: 0.6219

=== LogReg ===
Ймовірності для 15-го тестового зразка (факт. клас 1):
  Клас '0' (індекс 0): 0.5274
  Клас '1' (індекс 1): 0.4726
Передбачений клас: 0
Впевненість у прогнозі: 0.5274

```

4. Метрики точності моделей за 2-м підходом “проміжний” і 2-м варіантом переліку ознак

```

=== LightGBM (Basic) ===
      precision    recall  f1-score   support
0         0.9623     0.9900     0.9760         697
1         0.8871     0.6707     0.7639          82
 accuracy         0.9564
macro avg         0.9247     0.8303     0.8699         779
weighted avg         0.9544     0.9564     0.9536         779

=== LightGBM (Opt-F1) ===
      precision    recall  f1-score   support
0         0.9627     0.9986     0.9803         697
1         0.9821     0.6707     0.7971          82
 accuracy         0.9641
macro avg         0.9724     0.8346     0.8887         779
weighted avg         0.9647     0.9641     0.9610         779

=== LightGBM (Opt-TR) ===
      precision    recall  f1-score   support

```

0	0.9712	0.9182	0.9440	697
1	0.5250	0.7683	0.6238	82
accuracy			0.9024	779
macro avg	0.7481	0.8433	0.7839	779
weighted avg	0.9242	0.9024	0.9102	779

=== RandomForest (Basic) ===

	precision	recall	f1-score	support
0	0.9572	0.9943	0.9754	697
1	0.9273	0.6220	0.7445	82
accuracy			0.9551	779
macro avg	0.9422	0.8081	0.8599	779
weighted avg	0.9540	0.9551	0.9511	779

=== RandomForest (Opt-F1) ===

	precision	recall	f1-score	support
0	0.9574	0.9986	0.9775	697
1	0.9808	0.6220	0.7612	82
accuracy			0.9589	779
macro avg	0.9691	0.8103	0.8694	779
weighted avg	0.9598	0.9589	0.9548	779

=== RandomForest (Opt-TR) ===

	precision	recall	f1-score	support
0	0.9705	0.9455	0.9578	697
1	0.6200	0.7561	0.6813	82
accuracy			0.9255	779
macro avg	0.7953	0.8508	0.8196	779
weighted avg	0.9336	0.9255	0.9287	779

=== DecisionTree (Basic) ===

	precision	recall	f1-score	support
0	0.9668	0.9598	0.9633	697
1	0.6782	0.7195	0.6982	82
accuracy			0.9345	779
macro avg	0.8225	0.8397	0.8308	779
weighted avg	0.9364	0.9345	0.9354	779

=== DecisionTree (Opt-F1) ===

	precision	recall	f1-score	support
0	0.9596	0.9885	0.9739	697
1	0.8689	0.6463	0.7413	82
accuracy			0.9525	779
macro avg	0.9142	0.8174	0.8576	779
weighted avg	0.9501	0.9525	0.9494	779

=== DecisionTree (Opt-TR) ===

	precision	recall	f1-score	support
0	0.0000	0.0000	0.0000	697
1	0.1053	1.0000	0.1905	82
accuracy			0.1053	779
macro avg	0.0526	0.5000	0.0952	779
weighted avg	0.0111	0.1053	0.0201	779

=== LogReg (Basic) ===

	precision	recall	f1-score	support
0	0.9765	0.7747	0.8640	697
1	0.3053	0.8415	0.4481	82
accuracy		0.7818		779
macro avg	0.6409	0.8081	0.6560	779
weighted avg	0.9058	0.7818	0.8202	779

```

=== LogReg (Opt-F1) ===
      precision    recall  f1-score   support
     0       0.9642     0.9670     0.9656     697
     1       0.7125     0.6951     0.7037     82
  accuracy                   0.9384     779
 macro avg       0.8384     0.8311     0.8347     779
weighted avg       0.9377     0.9384     0.9380     779

```

```

=== LogReg (Opt-TR) ===
      precision    recall  f1-score   support
     0       0.9688     0.8895     0.9274     697
     1       0.4460     0.7561     0.5611     82
  accuracy                   0.8755     779
 macro avg       0.7074     0.8228     0.7443     779
weighted avg       0.9137     0.8755     0.8889     779

```

5. Метрики точності моделей за 3-м підходом “прикінцевий” для двох типів прогнозування: «Basic» і «Opt-F1»

```

--- DecisionTree (Basic) ---
      precision    recall  f1-score   support
     0   0.9650   0.9484   0.9566     697
     1   0.6170   0.7073   0.6591     82
  accuracy                   0.9230     779
 macro avg   0.7910   0.8278   0.8078     779
weighted avg   0.9283   0.9230   0.9253     779

```

```

--- DecisionTree (Opt-F1) ---
      precision    recall  f1-score   support
     0   0.9639   0.9570   0.9604     697
     1   0.6552   0.6951   0.6746     82
  accuracy                   0.9294     779
 macro avg   0.8095   0.8260   0.8175     779
weighted avg   0.9314   0.9294   0.9303     779

```

```

--- RandomForest (Basic) ---
      precision    recall  f1-score   support
     0   0.9670   0.9670   0.9670     697
     1   0.7195   0.7195   0.7195     82
  accuracy                   0.9409     779
 macro avg   0.8433   0.8433   0.8433     779
weighted avg   0.9409   0.9409   0.9409     779

```

```

--- RandomForest (Opt-F1) ---
      precision    recall  f1-score   support
     0   0.9561   1.0000   0.9776     697
     1   1.0000   0.6098   0.7576     82
  accuracy                   0.9589     779
 macro avg   0.9781   0.8049   0.8676     779
weighted avg   0.9607   0.9589   0.9544     779

```

```

--- LightGBM (Basic) ---
      precision    recall  f1-score   support
     0   0.9618   0.9742   0.9679     697
     1   0.7534   0.6707   0.7097     82
  accuracy                   0.9422     779
 macro avg   0.8576   0.8225   0.8388     779
weighted avg   0.9398   0.9422   0.9407     779

```

```

--- LightGBM (Opt-F1) ---

```

	precision	recall	f1-score	support
0	0.9574	1.0000	0.9782	697
1	1.0000	0.6220	0.7669	82
accuracy		0.9602		779
macro avg	0.9787	0.8110	0.8726	779
weighted avg	0.9619	0.9602	0.9560	779

--- LogReg (Basic) ---

	precision	recall	f1-score	support
0	0.9701	0.7461	0.8435	697
1	0.2716	0.8049	0.4062	82
accuracy		0.7522		779
macro avg	0.6209	0.7755	0.6248	779
weighted avg	0.8966	0.7522	0.7974	779

--- LogReg (Opt-F1) ---

	precision	recall	f1-score	support
0	0.9561	1.0000	0.9776	697
1	1.0000	0.6098	0.7576	82
accuracy		0.9589		779
macro avg	0.9781	0.8049	0.8676	779
weighted avg	0.9607	0.9589	0.9544	779

Додаток Д

СПИСОК ОПУБЛІКОВАНИХ ПРАЦЬ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

Статті у наукових виданнях,

включених до міжнародних наукометричних баз даних

Scopus та/або Web of Science Core Collection

1. Hlazunova O., Klymenko N., Mokriiev M., Nehrey M., **Klymenko Ye.** 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. P. 670–682. *(Глазуноюю О.Г. визначено загальну концепцію дослідження, наукове консультування, участь у формуванні методологічних засад, наукової новизни роботи та узагальненні результатів. Клименко Н.А. проведено літературний науковий огляд та запропоновано ідею дослідження щодо застосування технологій освітньої аналітики в системах дистанційного навчання. Мокрієвим М.В. здійснена технічна реалізація збору даних із LMS Moodle, налаштування інструментів аналітики та забезпечення інтеграції програмних засобів аналізу даних. Негрей М.В. здійснила проведення статистичної валідації та інтерпретації результатів отриманих аналітичних моделей. Клименко Є.О. представив результати технології аналізу даних, та сформував рекомендації щодо впровадження систем аналізу великих даних у закладах вищої освіти).*

Статті у наукових виданнях,

включених до Переліку наукових фахових видань України

2. 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. С. 49–59. *(Глазуноюю О.Г. сформовано концептуальні засади дослідження, визначено стратегічні напрями цифрової трансформації освітнього середовища, обґрунтовано необхідність використання цифрових інструментів освітньої аналітики для підтримки управлінських рішень у закладах вищої освіти.*

*Клименком Є.О. розроблено структуру та логіку реалізації процесу аналізу великих масивів освітніх даних та здійснено експериментальне дослідження інформаційної технології освітньої аналітики на прикладі навчального порталу Національного університету біоресурсів і природокористування України. **Волошина Т.В.** провела аналіз груп цифрових інструментів освітньої аналітики, систематизувала підходи до класифікації аналітичних засобів LMS Moodle. Мокрієвим М.В. сформовано систему освітніх індикаторів для оцінювання активності та залученості здобувачів освіти. **Вороненко О.В.** провів експертну оцінку використання результатів освітньої аналітики та аналіз практичної значущості отриманих результатів для освітніх програм.)*

3. **Клименко Є., Глазунова О.** Методи інтелектуального аналізу освітніх даних у системах електронного навчання. Інформаційні технології та суспільство. 2024. № 2 (13). С. 34–40. *(Клименко Є.О. представив теоретичне дослідження розвитку, основних задач і методів інтелектуального аналізу для застосування в інформаційних системах і технологіях, обґрунтував можливості використання методів інтелектуального аналізу для прогнозування результатів навчання та підтримки прийняття управлінських рішень Глазунова О.Г. здійснювала наукове консультування щодо методології освітньої аналітики та визначила перспективи застосування методів інтелектуального аналізу даних в інформаційних системах).*

4. Глазунова О., **Клименко Є.** Інформаційна технологія аналітики освітніх даних. Наука і техніка сьогодні. 2025. № 2 (43). С. 1147–1155. *(Глазуновою О.Г. надано наукове консультування щодо застосування методів Data Mining в освітньому середовищі, проведено рецензування та загальне керівництво роботою над статтею. Клименко Є.О. визначив та обґрунтував основні складові інформаційної технології аналітики освітніх даних із використанням методів інтелектуального аналізу даних та Big Data на основі журналів активності LMS Moodle).*

5. **Клименко Є., Глазунова О.** Архітектура інформаційної технології освітньої аналітики з використанням інтелектуального аналізу даних. Інформаційні технології та суспільство. 2025. № 2 (17). С. 69–75. *(Клименко Є.О. обґрунтував результати роботи розробленої інформаційної технології освітньої аналітики на основі використання методів інтелектуального аналізу освітніх даних. Глазуновою О.Г. здійснено методологічне обґрунтування дослідження та проведений комплексний аналіз можливостей імплементації інтелектуального аналізу в інформаційну технологію освітньої аналітики).*

Тези наукових доповідей

6. **Клименко Є. О.** Компетентнісний підхід в освітніх траєкторіях випускників ІТ спеціальностей ЗВО на базі платформи Moodle. Економіко-правовий розвиток сучасної України: XI Всеукраїнська наукова конференція студентів, аспірантів та молодих вчених, що присвячена 25-й річниці з дня заснування Національного університету «Одеська юридична академія», м. Одеса, 25 листопада 2022 року: тези доповіді. Одеса, 2022. С. 197–199.

7. **Клименко Є., Глазунова О.** Врахування компетентісного підходу в освітніх траєкторіях здобувачів освіти. Sectoral research XXI: characteristics and features: V International Scientific and Theoretical Conference, Chicago, USA, February 3, 2023: Conference Paper. Chicago, USA, 2023. P. 127–128. *(Клименко Є. Обґрунтував доцільність впровадження інформаційних технологій з використанням модульного об'єктно-орієнтованого динамічного навчального середовища для ефективної персоналізації навчання студентів у вищих навчальних закладах. Глазуновою О.Г. визначено концептуальні рамки дослідження та доведено вплив трансформацій на організацію освітнього процесу у закладах вищої освіти).*

8. **Клименко Є., Глазунова О.** Moodle Big Data analytics за допомогою Power BI. Science in Motion: Classic and modern tools and methods in scientific investigations: II International Scientific and Practical Conference, Vinnytsia–Vienna, Ukraine-Austria, January 19, 2024: Conference Paper. Vinnytsia–Vienna, Ukraine-

Austria, 2024. P. 201–203. *(Клименком Є.О. розроблені підходи до аналізу великих масивів даних LMS Moodle з інтеграцією з Power BI та надані результати аналізу великих даних з платформи електронного навчання Moodle в НУБІП України. Глазуновою О.Г сформульовано постановку задачі аналізу великих масивів даних, що зберігаються в СУБД та здійснене наукове консультування щодо застосування цих методів в освітніх інформаційних системах).*

9. **Клименко Є.** Аналіз освітніх даних у системах електронного навчання. Інформаційні технології: економіка, техніка, освіта 2024: Міжнародна науково-практична конференція молодих вчених, м. Київ, 7–8 листопада 2024 року: тези доповіді. Київ, 2024. С. 79–81.

10. **Клименко Є.** Розробка інформаційних систем прогнозування успішності здобувачів освіти. Інформаційне суспільство: технологічні, економічні та технічні аспекти становлення: Міжнародна наукова інтернет-конференція, Тернопіль–Ополе, Україна–Польща, 15–16 квітня 2025 року: тези доповіді. Тернопіль, 2025, випуск 98 С.22-25.

11. **Клименко Є.** Прогнозування успішності здобувачів освіти методами інтелектуального аналізу даних. Інформаційні технології: економіка, техніка, освіта '2025: XVI Міжнародна науково-практична конференція молодих вчених, м. Київ, 28–29 жовтня 2025 року: тези доповіді. Київ, 2025. С. 89–90.

12. Hlazunova O., Klymenko N., **Klymenko Ye.** Machine Learning in Educational Analytics Systems. Cyfryzacja Edukacji na Uczelniach Przyrodniczych: IX Międzynarodowa Konferencja, Wrocławiu, Polska, Listopad 20–21, 2025: tezy raportu. *(Глазунова О.Г. обгрунтувала доцільність використання методів машинного навчання в освітній аналітиці. Клименко Н.А. провела інтерпретацію результатів машинного навчання в контексті освітнього процесу. Клименком Є.О проведено аналіз сучасних алгоритмів машинного навчання для задач прогнозування успішності здобувачів освіти та розроблено архітектурну модель інформаційної технології освітньої аналітики з використанням ML-алгоритмів).*