

НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ «ОДЕСЬКА ЮРИДИЧНА АКАДЕМІЯ»
ФАКУЛЬТЕТ КІБЕРБЕЗПЕКИ ТА ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ

Кафедра інформаційних технологій

КВАЛІФІКАЦІЙНА РОБОТА

на тему:

**«СИСТЕМНИЙ ПІДХІД ДО ОЦІНКИ ЗАЛУЧЕНОСТІ СТУДЕНТІВ
НА ОСНОВІ AI-АНАЛІЗУ ДАНИХ LMS MOODLE»**

Виконав: студент 2 курсу

Рівень: магістр

Галузь знань 12 «Інформаційні технології»

Спеціальність 124 «Системний аналіз»

Заставнюк Мирослав Андрійович

Керівник: к.пед.н., доцент

Логінова Наталія Іванівна

Рецензент к.т.н., доцент

Лобода Юлія Геннадіївна

Одеса – 2025

НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ «ОДЕСЬКА ЮРИДИЧНА АКАДЕМІЯ»
 Факультет кібербезпеки та інформаційних технологій
 Кафедра інформаційних технологій
 Галузь знань: **12 Інформаційні технології**
 Спеціальність: **124 Системний аналіз**
 Назва освітньої програми: **Аналіз та безпека даних**

ЗАТВЕРДЖУЮ

Завідувач кафедри інформаційних технологій

доцент _____ Н.І. Логінова

« ____ » _____ 20__ року

ЗАВДАННЯ

**на кваліфікаційну (магістерську) роботу
 здобувачу Заставнюку Мирославу Андрійовичу**

- Тема роботи: «Системний підхід до оцінки залученості студентів на основі AI-аналізу даних LMS Moodle»
 Керівник роботи: к.пед.н, доцент Логінова Наталія Іванівна
 затверджені наказом НУ ОЮА від «10» жовтня 2025 року № 3183-06
- Строк подання здобувачем роботи «25» листопада 2025 рік
- Дата видачі завдання «02» червня 2025 рік

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів магістерської роботи	Строк виконання етапів роботи	Примітка
1.	Вибір теми, вивчення літературних джерел та складання плану роботи		
2.	Підготовка першого розділу роботи та подання його керівнику		
3.	Підготовка другого розділу роботи та подання його керівнику		
4.	Підготовка третього розділу роботи та подання його керівнику		
5.	Підготовка вступу та висновків		
6.	Доопрацювання роботи з урахуванням зауважень керівника		
7.	Подача електронного варіанту роботи для перевірки на плагіат		
8.	Допуск роботи до захисту завідуючим кафедри		
9.	Захист роботи		

Здобувач _____
 (підпис) (прізвище та ініціали)

Керівник роботи _____
 (підпис) (прізвище та ініціали)

АНОТАЦІЯ

Кваліфікаційна робота здобувача 2-го курсу магістратури Заставнюка Мирослава Андрійовича на тему «Системний підхід до оцінки залученості студентів на основі AI-аналізу даних LMS Moodle» за спеціальністю 124 «Системний аналіз», освітня програма «Аналіз та безпека даних».

Структура кваліфікаційної роботи складається зі вступу, основної частини з трьох розділів, висновків, списку використаних джерел. Робота викладена на 53 сторінках, містить 7 рисунків, 4 таблиці. Перелік посилань включає 22 джерела.

Метою дослідження є розроблення системного підходу до оцінки залученості студентів у навчальний процес на основі аналізу даних LMS Moodle із використанням методів штучного інтелекту для підвищення ефективності моніторингу та управління освітнім процесом.

В кваліфікаційній роботі розглянуто теоретичні основи освітньої аналітики, методи збору та обробки лог-даних Moodle, а також сучасні підходи до визначення показників залученості студентів. На основі очищених логів сформовано набір ключових показників активності – кількість входів, переглядів курсів, участь у тестах, форумах і завданнях, що дозволило кількісно оцінити поведінкову активність студентів. Для реалізації моделі аналізу застосовано мову програмування Python із використанням бібліотек Pandas, NumPy, scikit-learn та Matplotlib. Проведено кластеризацію студентів методом K-Means, оптимальну кількість кластерів визначено за допомогою методу ліктя. Отримані кластери використано для навчання моделі Random Forest. Результати підтвердили ефективність застосування штучного інтелекту для аналізу навчальної активності студентів. Запропонована модель може бути використана для моніторингу участі студентів у навчальному процесі, виявлення груп ризику та підтримки процесів прийняття рішень у системі управління навчанням.

MOODLE, МАШИННЕ НАВЧАННЯ, КЛАСИФІКАЦІЯ, КЛАСТЕРИЗАЦІЯ, RANDOM FOREST, K-MEANS, ШТУЧНИЙ ІНТЕЛЕКТ, ОСВІТНЯ АНАЛІТИКА

ABSTRACT

Qualification work of of the 2nd-year master's student Zastavnuk Myroslav Andriyovych on the topic "A Systematic Approach to Assessing Student Engagement Based on AI Analysis of LMS Moodle Data" for the specialty 124 "System Analysis," educational program "Data Analysis and Security."

The structure of the qualification thesis consists of an introduction, the main part with three chapters, conclusions, and a list of references. The work is presented on 53 pages and includes 7 figures and 4 tables. The reference list contains 22 sources.

The aim of the research is to develop a systematic approach to assessing student engagement in the learning process based on the analysis of LMS Moodle data using artificial intelligence methods to improve the efficiency of monitoring and managing the educational process.

The thesis reviews the theoretical foundations of educational analytics, methods for collecting and processing Moodle log data, as well as modern approaches to defining student engagement indicators. Based on cleaned logs, a set of key activity metrics was formed – number of logins, course views, participation in tests, forums, and assignments – allowing quantitative assessment of students' behavioral activity.

For the implementation of the analysis model, the Python programming language was used along with the Pandas, NumPy, scikit-learn, and Matplotlib libraries. Student clustering was performed using the K-Means method, with the optimal number of clusters determined by the Elbow Method. The resulting clusters were then used to train a Random Forest model.

The results confirmed the effectiveness of using artificial intelligence for analyzing students' learning activity. The proposed model can be applied for monitoring student participation, identifying at-risk groups, and supporting decision-making processes in a learning management system.

MOODLE, MACHINE LEARNING, CLASSIFICATION, CLUSTERING, RANDOM FOREST, K-MEANS, ARTIFICIAL INTELLIGENCE, EDUCATIONAL ANALYTICS

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ.....	6
ВСТУП.....	7
1 ТЕОРЕТИЧНІ ОСНОВИ СИСТЕМНОГО ПІДХОДУ ДО ОЦІНКИ ЗАЛУЧЕНОСТІ СТУДЕНТІВ У LMS MOODLE	9
1.1 Аналіз предметної області дослідження.....	9
1.2 Системний підхід у дослідженні освітніх процесів.....	13
1.3. Використання LMS Moodle у вищій освіті.....	15
1.4. Стан і проблеми використання методів штучного інтелекту в освітній аналітиці	17
Висновки з першого розділу	19
2 РОЗРОБЛЕННЯ СИСТЕМНОЇ МОДЕЛІ ОЦІНКИ ЗАЛУЧЕНОСТІ СТУДЕНТІВ У MOODLE	20
2.1 Аналіз структури даних у Moodle (лог-файли, активності, події)	20
2.2 Визначення показників залученості	22
2.3 Побудова системної моделі оцінки залученості	25
2.4 Опис взаємозв'язків між компонентами системи.....	28
Висновки з другого розділу	30
3 РОЗРОБЛЕННЯ ТА ЕКСПЕРИМЕНТАЛЬНА ПЕРЕВІРКА AI-МОДЕЛІ АНАЛІЗУ ЗАЛУЧЕНОСТІ.....	31
3.1 Підготовка та попередня обробка даних Moodle.....	31
3.2 Вибір методу машинного навчання для класифікації	33
3.3 Реалізація програмного модуля аналізу.....	36
3.4 Навчання та тестування моделі	42
3.5 Аналіз результатів: точність, інтерпретація та візуалізація	44
Висновки з третього розділу	48
ВИСНОВКИ.....	49
ПЕРЕЛІК ПОСИЛАНЬ	51
ДОДАТОК.....	54

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ

LMS – Learning Management Systems, система управління навчанням

AI – штучний інтелект

ML – машинне навчання

KPI – Key Performance Indicators, ключові показники ефективності

LA – Learning Analytics, навчальна аналітика

EI – Engagement Index, індекс залученості

SVM – Support Vector Machine, метод опорних векторів

KNN – K-Nearest Neighbors, метод К-найближчих сусідів

ВСТУП

Актуальність дослідження зумовлена цифровізацією освіти та необхідністю підвищення ефективності навчального процесу. Використання систем управління навчанням, зокрема LMS Moodle, дозволяє збирати великі обсяги даних про активність і прогрес студентів, які традиційні методи оцінки залученості не здатні повноцінно аналізувати.

Системний підхід у поєднанні з методами штучного інтелекту забезпечує комплексну оцінку залученості студентів, враховуючи когнітивні, поведінкові та емоційні аспекти їх діяльності. AI-моделі дозволяють виявляти приховані патерни взаємодії, прогнозувати ризики відсіву та надавати персоналізовані рекомендації, що сприяє підвищенню якості навчання та оптимізації освітніх процесів.

Таким чином, дослідження, яке пропонує системний і AI-орієнтований підхід до оцінки залученості, є актуальним для модернізації освітньої практики, забезпечуючи глибше розуміння поведінки студентів та підвищення якості навчального процесу.

Мета дослідження – розробити системний підхід до оцінки залученості студентів у навчальний процес на основі аналізу даних LMS Moodle із використанням методів штучного інтелекту для підвищення ефективності моніторингу та управління освітнім процесом.

Для досягнення цієї мети передбачено наступні **завдання**:

- проаналізувати предметну область та особливості залученості студентів у LMS Moodle;
- дослідити системний підхід до оцінки освітніх процесів і використання AI в аналітиці;
- визначити структуру даних Moodle та показники залученості студентів;
- розробити системну модель оцінки залученості та описати взаємозв'язки її компонентів;
- створити, навчити та протестувати AI-модель для аналізу залученості студентів;

– проаналізувати результати моделі та їх візуалізацію для оцінки ефективності підходу.

Об’єкт дослідження – процес залученості студентів у середовищі LMS Moodle.

Предмет дослідження – методи і засоби оцінки залученості студентів на основі аналізу даних Moodle із застосуванням штучного інтелекту та системного підходу.

Методи дослідження:

– системний аналіз – моделювання взаємозв’язків між компонентами освітньої системи та побудова системної моделі оцінки залученості;

– аналітичні методи обробки даних – дослідження лог-файлів, активностей та подій для визначення показників залученості;

– методи машинного навчання – створення, навчання та тестування AI-моделі для класифікації та оцінки залученості студентів;

– візуалізація та інтерпретація даних – графічне відображення результатів аналізу та оцінка точності моделі.

Практичне значення отриманих результатів: розроблена системна модель та AI-модель оцінки залученості студентів дозволяють автоматизовано відстежувати активність у LMS Moodle та своєчасно виявляти студентів із низьким рівнем участі. Запропонований підхід може бути використаний викладачами адміністрацією для підвищення ефективності навчального процесу та вдосконалення систем електронного навчання.

Структура кваліфікаційної роботи відповідає меті та завданням дослідження та складається зі вступу, основної частини з трьох розділів, висновків, переліку посилань. Робота викладена на 53 сторінках, містить 7 рисунків, 4 таблиці. Перелік посилань включає 22 джерела.

1 ТЕОРЕТИЧНІ ОСНОВИ СИСТЕМНОГО ПІДХОДУ ДО ОЦІНКИ ЗАЛУЧЕНОСТІ СТУДЕНТІВ У LMS MOODLE

1.1 Аналіз предметної області дослідження

У сучасній системі вищої освіти цифрові технології відіграють визначальну роль у забезпеченні ефективності навчального процесу. Активне впровадження систем управління навчанням (LMS) дозволило автоматизувати більшість організаційних і педагогічних процесів. LMS-платформи не лише забезпечують доступ студентів до навчальних матеріалів, а й створюють великий масив даних про їх поведінку, активність і взаємодію в освітньому середовищі.

Однією з найпоширеніших таких платформ є LMS Moodle, яка використовується у більшості закладів вищої освіти світу завдяки відкритому коду, гнучкій архітектурі та широким можливостям аналітики навчальних процесів.

Актуальність цього дослідження зумовлена необхідністю системного аналізу залученості студентів. Для розуміння сутності цього поняття доцільно більш детально розглянути його характеристику та визначити основні аспекти.

Згідно з Cluelabs [1], залученість студентів є багатогранним поняттям в освітньому контексті, яке часто визначається як міра активної участі студента або його емоційної залученості у навчальний процес. Це поняття охоплює такі види поведінки, як регулярне відвідування занять, здача завдань, активна участь у дискусіях в класі та взаємодія з онлайн-платформами. Однак залученість не обмежується лише поведінковими аспектами – вона включає також емоційну чи психологічну складову, зокрема мотивацію, інтерес і значущість, які студент надає своєму навчанню.

Водночас, у деяких систематичних оглядах поняття залученості розширюється до п'яти складових, серед яких, окрім згаданих трьох, виділяють соціальну та колаборативну залученість. Це підкреслює, що процес взаємодії у цифровому навчальному середовищі має також комунікаційний та колективний виміри [2]. У межах даного дослідження доцільно зосередитися на трьох основних аспектах (поведінковому, когнітивному та емоційному), оскільки вони

безпосередньо відображають дані, що збираються системою LMS Moodle, та дозволяють ефективно застосовувати методи штучного інтелекту для оцінки рівня залученості.

Поведінковий аспект включає активну участь у навчальному процесі, зокрема відвідування занять, а також негативні прояви, як-от порушення дисципліни. Емоційна залученість пов'язана з афективними реакціями, такими як інтерес, почуття належності, задоволення або розчарування. Когнітивна складова вказує на глибоке розуміння та осмислення навчального матеріалу.

Залученість студентів є основою успішного навчального процесу. Однак для ефективної оцінки та покращення цієї залученості важливо розглянути фактори, що на неї впливають. Вивчення цих факторів дозволить розробити стратегії для підвищення активності та мотивації студентів. Залученість - це комплексне явище, яке залежить від багатьох аспектів навчання і взаємодії з освітнім середовищем. Тому важливо виділити основні чинники, що безпосередньо впливають на поведінкову, когнітивну та емоційну залученість студентів. Серед них можна виокремити наступні:

1. Особистісні та мотиваційні фактори, які визначають готовність студента активно долучатися до навчального процесу. Це можуть бути як внутрішні мотиви (наприклад, прагнення досягти особистих цілей), так і зовнішні фактори (очікування високих оцінок або схвалення від викладачів).

2. Технічні фактори та доступність ресурсів, які впливають на зручність та ефективність взаємодії з навчальними матеріалами. Наприклад, інтуїтивно зрозумілі та добре організовані LMS-платформи дозволяють студентам легко знаходити потрібні матеріали та взаємодіяти з іншими учасниками навчального процесу, що може підвищити їх залученість.

3. Педагогічні методи та взаємодія викладача, що включають надання регулярного зворотного зв'язку, підтримку в навчальному процесі, а також активні методи навчання, наприклад, групові проєкти. Викладач, який активно взаємодіє зі студентами, може стимулювати їх до більш плідної роботи, створюючи відчуття важливості кожного учасника.

4. Психологічні фактори, які можуть як сприяти залученості, так і бути бар'єром для глибокого занурення в навчальний процес. Наприклад, стрес через страх перед невдачею може значно знижувати рівень активності студента в курсі.

Для ефективного управління залученістю студентів необхідно не тільки визначити фактори, які на неї впливають, а й оцінити рівень залученості на основі конкретних даних. Використання сучасних цифрових платформ дозволяє проводити детальний аналіз активності студентів на всіх етапах навчання. Оскільки залученість є багатогранним поняттям, для її оцінки використовуються різні методи, які дозволяють відстежувати не лише поведінкову, а й емоційну та когнітивну складову.

Серед таких методів можна виокремити три основні підходи, які дозволяють детально виміряти рівень залученості студентів.

1. Відстеження активності студентів на платформі.

Цей метод дозволяє викладачам отримувати детальну інформацію про участь студентів у навчальному процесі, не потребуючи додаткових зустрічей або запитів. Відстеження активності можна здійснювати за допомогою таких параметрів:

Вхід до системи: один з основних показників активності студента – це час, коли він заходить на платформу та як часто це відбувається. Якщо студент часто входить у систему, це може свідчити про його бажання бути залученим у навчання. Проте частота входу має бути не єдиним критерієм – важливо також оцінювати час, проведений на платформі.

Час, проведений на платформі: цей показник допомагає оцінити, скільки часу студент витрачає на перегляд матеріалів, виконання завдань чи участь в онлайн-обговореннях. Час, витрачений на платформі, може бути хорошим індикатором рівня залученості студента: довші сесії можуть свідчити про глибшу зацікавленість і концентрацію на навчанні.

Взаємодія з іншими елементами платформи: це можуть бути різноманітні дії, як-от участь у форумах, коментарі до матеріалів чи завдань, голосування в опитуваннях. Важливо фіксувати не лише активність у вигляді "відвідування", а й кількість і якість взаємодій студента з платформою. Наприклад, активні участі в

обговореннях можуть вказувати на високу когнітивну та емоційну залученість.

2. Аналітика участі в обговореннях і зворотного зв'язку.

Участь у дискусіях: сучасні LMS-платформи, дозволяють студентам брати участь у онлайн-дискусіях, де можна оцінити не лише кількість постів, а й якість їхнього контенту. Активні студенти можуть писати змістовні коментарі, ставити запитання та надавати конструктивні відповіді на пости інших учасників. Частота участі в таких дискусіях дозволяє оцінити поведінкову залученість студентів, а їхня якість може бути індикатором глибини когнітивної залученості.

Зворотний зв'язок: аналітика взаємодії студентів з викладачами через зворотний зв'язок може відображати рівень їхньої емоційної залученості. Якщо студент отримує відгуки на свою роботу та активно на них реагує (вносить корективи в завдання або обговорює коментарі з викладачем), це свідчить про його готовність поглиблювати свої знання та активно долучатися до навчального процесу.

Оцінка виконаних завдань: на LMS-платформах часто є можливість оцінювати виконані завдання за допомогою автоматизованих систем (наприклад, тести) або вручну (есе, лабораторні завдання). Після цього студенти отримують зворотний зв'язок, що дає змогу оцінити їхню готовність коригувати свої помилки та покращувати результати. Якщо студент отримує високу оцінку та активно працює над помилками, це є показником високої залученості.

3. Анкети та опитування для вимірювання емоційної залученості.

Анкети для вимірювання задоволення: за допомогою анкет можна оцінити, наскільки студенти задоволені навчальним процесом, методами викладання та зручністю використання платформи. Це можуть бути питання, які вимагають від студентів оцінити різні аспекти їхнього навчання: від зручності інтерфейсу LMS до якості навчальних матеріалів. Такі анкети можуть включати шкали оцінок, що дозволяють провести кількісний аналіз емоційної залученості.

Опитування для вимірювання мотивації та інтересу: анкети можуть також вимірювати рівень внутрішньої мотивації студентів до навчання. Наприклад, запитання на кшталт «Наскільки цікавими були для вас останні лекції?» або «Як

часто ви самостійно досліджуєте додаткові матеріали з курсу?» можуть допомогти оцінити, чи зацікавлений студент у вивченні предмета поза межами вимог навчальної програми.

1.2 Системний підхід у дослідженні освітніх процесів

Оцінка залученості студентів, що включає поведінкові, когнітивні та емоційні аспекти, є частиною загального аналізу освітнього середовища. Щоб краще зрозуміти, як взаємодіють ці фактори та як вони впливають на успішність навчання, доцільно застосовувати системний підхід. Він дозволяє розглядати освітні процеси не як окремі елементи, а як інтегровану систему, де кожен компонент взаємодіє з іншими.

У системному підході освітній процес можна уявити як систему, що має елементи, взаємозв'язки, процеси, вхідні та вихідні дані.

Вхідні дані включають навчальні матеріали, активність студентів, результати тестів та оцінки.

Процеси – це взаємодія студентів із матеріалом, участь у дискусіях, групова робота та виконання завдань.

Вихідні результати включають рівень знань, академічну успішність, мотивацію та загальний рівень залученості.

Завдяки системному підходу можливо відстежувати, як зміни одного елемента (наприклад, активності у форумі) впливають на інші компоненти (мотивацію, результати тестів) та на систему в цілому [3].

Системний підхід широко застосовується у сучасній освіті для:

- аналізу процесів навчання, коли відстежується взаємозв'язок між активністю студентів та успішністю;
- управління якістю освіти, коли інтегруються дані з різних курсів і дисциплін для покращення результатів навчання;
- моделювання навчальної діяльності, що дозволяє прогнозувати можливі труднощі студентів та адаптувати навчальний процес [4].

Аналіз сучасних досліджень показав, що існує прогалина у застосуванні

навчальної аналітики для комплексної оцінки залученості студентів [2]. Зокрема, жодне з досліджень не об'єднало всі типи залученості для отримання цілісних результатів.

Застосування системного підходу у поєднанні з аналітикою навчання дозволяє інтегрувати різні показники активності студентів і створювати модель, яка відображає взаємозв'язок між поведінковими, когнітивними та емоційними аспектами залученості. Такий підхід сприяє більш точному розумінню навчальної взаємодії та підвищує ефективність освітнього процесу.

У контексті даного дослідження системний підхід застосовується до даних платформи LMS Moodle, що дозволяє отримувати та обробляти різноманітні показники активності студентів: кількість входів у систему, перегляди навчальних матеріалів, виконання завдань і оцінки. Інтеграція цих даних у єдину системну модель дозволить більш точно оцінити рівень залученості та зробити висновки щодо ефективності навчального процесу.

Таким чином, системний підхід забезпечить комплексний аналіз навчальної взаємодії, дозволить виявити приховані закономірності у поведінці студентів і підготувати підґрунтя для застосування методів штучного інтелекту та освітньої аналітики.

Системний підхід застосовується до даних платформи LMS Moodle, що дозволяє отримувати та обробляти різноманітні показники активності студентів:

- кількість входів у систему;
- перегляди навчальних матеріалів;
- виконання завдань і тестів;
- оцінки та відгуки викладачів.

Інтеграція цих показників у єдину системну модель дозволить проаналізувати взаємозв'язок поведінкових, когнітивних та емоційних аспектів залученості, оцінити ефективність навчального процесу та виявити закономірності у навчальній взаємодії студентів. Використання системного підходу створює основу для застосування методів штучного інтелекту та освітньої аналітики, таких як класифікація рівня активності студентів, прогнозування ризику неуспішності та

адаптивне управління навчальним процесом.

Таким чином, системний підхід не лише дозволяє комплексно оцінювати залученість студентів, а й створює основу для подальшого аналізу даних та впровадження інноваційних рішень у навчальному процесі. У наступному підпункті розглянемо платформу LMS Moodle, її функціональні можливості та типи даних, що забезпечують збір інформації для реалізації системного підходу та аналітики залученості студентів.

1.3. Використання LMS Moodle у вищій освіті

Система управління навчанням Moodle є однією з найпоширеніших платформ електронного навчання, що використовуються в закладах вищої освіти. Вона надає викладачам і адміністраторам широкий набір функціональних можливостей, які забезпечують організацію навчальних курсів, взаємодію студентів з освітнім контентом, контроль знань та аналіз активності [5].

До основних функціональних компонентів Moodle належать:

- створення курсів із модульною структурою, що охоплює лекції, практичні та самостійні роботи;
- форуми та чати – інструменти для обговорення та взаємодії між студентами і викладачем;
- тести, опитування і завдання – механізми перевірки знань і навичок, з можливістю автоматичної чи ручної оцінки;
- журнали оцінок і звіти – засоби для відстеження прогресу студентів, накопичення результатів й історії навчання;
- модулі активності – наприклад завдання (Assignments), перегляди сторінок (Page), відео-матеріали, виконання практик, а також доповнення (плагіни), що розширюють функціонал платформи.

Використання системи Moodle у закладах вищої освіти базується на низці педагогічних та технологічних концепцій, які визначають її ефективність як інструмента для організації електронного та змішаного навчання.

Moodle була розроблена на засадах соціального конструктивізму. Ця теорія

стверджує, що навчання є процесом, у якому студенти активно будують власні знання через активну взаємодію з контентом та співпрацю з однолітками та викладачем. Інструменти Moodle (форуми, вікі, глосарії) орієнтовані саме на співробітництво, обговорення та рефлексію, що забезпечує активне конструювання знань.

Moodle ідеально підходить для реалізації моделей змішаного навчання, поєднуючи традиційне аудиторне навчання з можливостями онлайн-технологій. Платформа забезпечує асинхронний доступ до матеріалів, дозволяє проводити самостійну роботу, контролювати знання та організовувати постійну комунікацію, надаючи студентам гнучкість та підтримуючи персоналізацію навчальних траєкторій.

Платформа Moodle генерує великі масиви логів і звітів, які фіксують дії студентів у системі. Зокрема:

- кількість входів у систему (логування сесій);
- кількість і тривалість переглядів навчальних матеріалів;
- участь у форумах і дискусіях (кількість повідомлень, відповідей, переглядів тем);
- результати виконання тестів і завдань (оцінки, час виконання);
- відомості про завершеність модулів і курсів, журнали оцінок;
- деталі про час активності (коли користувачі були активні) – як зазначено, лог-записи «tell us “who”, “what”, and “when”» у Moodle [5].

Зібрані дані можуть бути безпосередньо використані для оцінки рівня залученості студентів. Наприклад:

- показник кількості входів або тривалості сесій може свідчити про поведінкову залученість;
- участь у форумах, кількість відповіді чи перегляди тем — підказує про соціальну або емоційну активність;
- час виконання завдань, кількість повторних звернень до матеріалу — демонструє когнітивну залученість.

Moodle виступає як первинне джерело даних для навчальної аналітики

(Learning Analytics, LA). Платформа підтримує аналітичні моделі: наприклад, через API Moodle Analytics можна налаштувати показники (індикатори) і цілі для прогнозування ризику відмови від навчання [6].

Отже, можливості Moodle створюють технологічну базу для реалізації системного підходу до оцінки залученості студентів, де різноманітні типи активності інтегруються в єдину аналітичну модель.

1.4. Стан і проблеми використання методів штучного інтелекту в освітній аналітиці

Перед тим як досліджувати застосування штучного інтелекту (AI) для оцінки залученості студентів, важливо визначити сутність AI та машинного навчання (ML) у контексті освіти. AI та ML дозволяють автоматизовано обробляти великі обсяги освітніх даних, виявляти закономірності, прогнозувати поведінку студентів та створювати адаптивні освітні інтервенції. Наприклад, такі системи можуть класифікувати рівень активності студентів, прогнозувати ризик неуспішності або відрахування, а також підтримувати персоналізовані навчальні траєкторії [7].

Ключові завдання AI/ML в освіті включають:

- класифікація та прогнозування – автоматизоване визначення рівня активності студентів та прогнозування критичних подій, таких як ризик неуспішності або відрахування (early dropout prediction);

- персоналізація – підтримка індивідуалізованих навчальних траєкторій через динамічну адаптацію навчального контенту та завдань під потреби конкретного студента.

ML є особливо ефективним для визначення рівня залученості, оскільки здатне обробляти багатовимірні дані (логи, кліки, оцінки) і виносити рішення про стан студента, що дозволяє викладачам своєчасно вживати педагогічних заходів та надавати індивідуальні рекомендації. Проте існує дефіцит систематизованих моделей, які об'єднують усі аспекти залученості (поведінковий, когнітивний, емоційний, соціальний) та надають цілісне уявлення про студентську активність [8].

У платформах LMS, таких як Moodle, через API Moodle Analytics можна створювати моделі прогнозування, які поєднують показники активності студентів (індикатори) із визначеною цільовою подією. Ціль описує подію, яку система прогнозує (наприклад, завершення курсу або ризик відрахування), а індикатори – це параметри, на основі яких робиться прогноз (кількість переглядів матеріалів, участь у форумах, оцінки). Moodle здатний оцінювати точність моделей та оновлювати їх на основі нових даних, а результати прогнозів можна використовувати для адаптивних дій: повідомлення студентам, формування звітів або запуск нових навчальних активностей [6].

Систематичний огляд застосування AI та ML у навчальній аналітиці показав, що підходи машинного навчання переважають у завданнях прогнозування залученості, моделювання ризику відсіву та прогнозування академічних результатів, тоді як генеративний AI (наприклад, трансформерні моделі GPT-4) використовується для зворотного зв'язку в режимі реального часу, адаптивного навчання та аналізу настроїв студентів [2].

Незважаючи на значні успіхи, впровадження AI/ML в освітню аналітику стикається із суттєвими теоретичними та практичними викликами:

- дефіцит інтегрованих системних моделей – існує гостра потреба у системних моделях, які забезпечують інтеграцію технічної AI-аналітики з глибоким педагогічним розумінням процесу залученості. Брак систематизованих моделей, які б об'єднували всі чотири аспекти залученості (поведінковий, когнітивний, емоційний, соціальний), не дає змоги отримати цілісне уявлення про студентську активність;

- обмеженість аналізованих аспектів – більшість досліджень і практичних інструментів традиційно фокусуються на легко вимірюваних поведінкових аспектах (кількість кліків, тривалість сесій, час виконання завдань). При цьому значно недооцінюються складніші для вимірювання когнітивні (глибина розуміння, рефлексія) та емоційні (рівень мотивації, відчуття приналежності) складові залученості;

- проблема контекстуалізації – ефективність AI-моделей значною мірою

залежить від контексту конкретного навчального середовища, специфіки курсу та педагогічних цілей. Моделі, розроблені для одного курсу чи закладу, можуть бути неефективними в іншому [8].

Отже, AI та ML мають великий потенціал для підвищення залученості студентів та прогнозування успішності. Проте їхня справжня ефективність буде досягнута лише за умови тісної інтеграції технічних, математичних моделей із педагогічною теорією, системним підходом та глибоким розумінням освітнього контексту. Створення таких інтегрованих, багатоаспектних моделей є критично важливим і ляже в основу подальшої розробки у цьому дослідженні.

Висновки з першого розділу

У першому розділі кваліфікаційної роботи аналіз предметної області показав, що залученість студентів у навчальний процес є багатограним явищем та включає поведінкові, когнітивні та емоційні аспекти. Визначення основних факторів, що впливають на залученість – особистісних, технічних, педагогічних та психологічних – дозволяє розробляти ефективні стратегії стимулювання активності та підвищення якості навчання.

Аналіз системного підходу продемонстрував його ефективність у комплексній оцінці навчальних процесів. Розгляд освітньої діяльності як інтегрованої системи, де кожен компонент взаємодіє з іншими, дозволяє відстежувати вплив конкретних елементів на загальний рівень залученості студентів та результати навчання.

Дослідження LMS Moodle підтвердило, що платформа надає широкий спектр даних для навчальної аналітики, які можна інтегрувати у системну модель оцінки залученості. Проте, ефективне використання цих даних вимагає застосування методів AI та ML для обробки великих обсягів інформації, прогнозування ризиків та підтримки персоналізованого навчання. Системний підхід у поєднанні з AI дозволяє створити комплексну модель, що враховує всі ключові аспекти залученості та є основою для подальшого дослідження.

2 РОЗРОБЛЕННЯ СИСТЕМНОЇ МОДЕЛІ ОЦІНКИ ЗАЛУЧЕНОСТІ СТУДЕНТІВ У MOODLE

2.1 Аналіз структури даних у Moodle (лог-файли, активності, події)

У сучасних системах електронного навчання платформи управління навчальним процесом LMS відіграють ключову роль у фіксації цифрових слідів діяльності користувачів. Це створює підґрунтя для проведення освітньої аналітики, оцінювання залученості студентів та дослідження ефективності навчальних ресурсів. Однією з найпоширеніших платформ цього типу є Moodle, яка автоматично структурує та зберігає події користувачів у вигляді журналів активності (лог-файлів). Такі журнали можуть бути експортовані у різних форматах та слугують базовим джерелом даних для подальшого аналізу взаємодії студентів із навчальним середовищем [9].

У межах цього дослідження було сформовано інтегрований набір лог-даних шляхом об'єднання трьох файлів експорту (двох у форматі XLSX та одного CSV), отриманих із журналів Moodle. Сукупний масив охоплює 6125 записів, що фіксують дії студентів і викладачів факультету кібербезпеки та інформаційних технологій Національного університету «Одеська юридична академія». Такий підхід дозволив підвищити репрезентативність даних та охопити ширший спектр подій, зокрема ті, що належать до різних курсів, компонентів системи та видів активностей.

Для демонстрації структури зібраних даних у таблиці 2.1 наведено фрагмент журналу подій, при цьому особисті дані користувачів (прізвища та імена) були знеособлені з метою забезпечення конфіденційності та відповідності вимогам академічної доброчесності та захисту персональних даних.

Таблиця 2.1 – Фрагмент журналу подій moodle

Час	Повне ім'я користувача	Контекст події	Компонент	Назва події	Опис
23.10.25, 13:36	Користувач 1	Ядро системи	Ядро системи	Користувач увійшов в систему	The user with id '99' has logged in.
23.10.25, 13:36	Користувач 2	Завдання: Лабораторна робота №6. Шрифти. FontAwesome. Форми	Завдання	Переглянуто статус подання роботи	The user with id '772' has viewed the submission status page for the assignment with course module id '3182'.
23.10.25, 13:36	Користувач 2	Завдання: Лабораторна робота №6. Шрифти. FontAwesome. Форми	Завдання	Переглянуто модуль курсу	The user with id '772' viewed the 'assign' activity with course module id '3182'.
23.10.25, 13:36	Користувач 2	Курс: Вебтехнології та вебдизайн (ПЗ, КН)	Ядро системи	Переглянуто курс	The user with id '772' viewed the course with id '133'.

Як видно, кожен запис журналу містить відомості про користувача, контекст події, компонент системи, тип взаємодії, дату й час, а також короткий опис виконаної дії. Відповідно до офіційної документації Moodle, усі події зберігаються в таблиці *logstore_standard_log*, яка акумулює інформацію про взаємодію користувачів із курсами, ресурсами, модулями та інтерфейсом системи.

Подальший аналіз передбачає групування логів за користувачами, курсами, типами подій та компонентами, що дозволяє отримати інтегральні показники активності студентів. Події Moodle можна умовно класифікувати за такими групами:

- системні (вхід у систему, вихід, перегляд курсу);
- навчальні (робота з тестами, завданнями, перегляд навчальних матеріалів);
- комунікаційні (взаємодія у форумах, повідомлення, коментарі);
- адміністративні (перегляд оцінок, зміна налаштувань, керування модулями).

У процесі попередньої обробки даних було здійснено багаторівневу фільтрацію для виділення лише записів, які належать студентам. Початковий набір включав події всіх категорій користувачів, зокрема викладачів, адміністраторів та технічних акаунтів.

1. На першому етапі видалено записи, що однозначно належать викладачам, ідентифікованим за їхніми повними іменами.

2. На другому етапі виключено всі рядки, де ім'я користувача містило дефіс («—»), оскільки такі акаунти систематично відповідали адміністраторам і викладацьким профілям.

У результаті фільтрації отримано підсумковий набір із 5074 записів, які відображають виключно активність студентів. Така очистка даних дозволила усунути вплив сторонніх подій на статистику активності та забезпечила коректність подальшого аналізу їхньої поведінкової взаємодії з навчальним середовищем.

Додатково варто зазначити, що очищений масив логів створює підґрунтя для подальших етапів аналізу, таких як:

- оцінювання загальної інтенсивності роботи студентів у курсах;
- виявлення патернів взаємодії з навчальними ресурсами;
- побудова моделей прогнозування успішності;
- порівняння активності між групами, модулями або окремими курсами;
- створення візуалізацій поведінкових траєкторій студентів.

Таким чином, структурне опрацювання лог-файлів Moodle є критично важливою передумовою для якісного та достовірного дослідження навчальної активності у цифровому середовищі.

2.2 Визначення показників залученості

У контексті цифрової трансформації освіти однією з ключових задач є кількісна оцінка ефективності навчального процесу та ступеня залученості студентів до навчальної діяльності. Саме тому у сучасній освітній аналітиці широко застосовуються ключові показники ефективності (Key Performance Indicators, KPI) – формалізовані метрики, які дозволяють оцінити прогрес у

досягненні стратегічних цілей освітньої установи та визначити якість взаємодії користувачів із навчальним середовищем. Як зазначено у дослідженні [10], KPI є узагальненими характеристиками, що агрегують результати численних активностей університету та дають комплексне уявлення про стан освітніх процесів у певний період часу.

У межах даної роботи KPI розглядаються як кількісні індикатори залученості студентів до навчання в середовищі LMS Moodle. Вони дозволяють оцінити різні виміри взаємодії студентів із платформою, включаючи поведінковий, когнітивний та комунікаційний аспекти. Аналіз очищених лог-даних дав змогу виокремити систему показників, які характеризують інтенсивність, регулярність, глибину та спрямованість навчальної активності. Ці показники відображають такі форми взаємодії, як відвідування курсів, робота над завданнями, участь у тестуванні, обговорення у форумах, а також загальну частоту звернень до навчального середовища.

На основі проведеної класифікації було сформовано узагальнену систему метрик, подану в таблиці 2.2. Для кожного показника наведено короткий опис, логіку розрахунку та одиницю виміру, що забезпечує можливість подальшої автоматизованої обробки та порівняльного аналізу.

Таблиця 2.2 – Основні показники залученості студентів у Moodle

Назва показника	Опис	Формула або спосіб розрахунку	Одиниця виміру
Кількість входів у систему (Login Count)	Відображає частоту взаємодії студента з навчальною платформою.	Кількість подій типу 'Користувач увійшов у систему' на одного користувача.	кількість входів
Кількість переглядів курсів (Course Views)	Характеризує загальний інтерес студента до навчальних матеріалів.	Підрахунок подій типу 'Переглянуто курс' для кожного користувача.	кількість переглядів
Активність у завданнях (Assignments Activity)	Відображає участь у виконанні та поданні завдань.	Σ подій типу 'Переглянуто статус подання роботи' або 'Подано завдання'.	кількість дій

Продовження таблиці 2.2

Назва показника	Опис	Формула або спосіб розрахунку	Одиниця виміру
Активність у тестах (Quiz Attempts)	Кількість спроб тестів, що характеризує участь у контролі знань.	Σ подій типу 'Спроба тесту розпочата' або 'Тест завершено'.	кількість спроб
Активність у форумах (Forum Posts)	Відображає рівень комунікаційної залученості.	Підрахунок подій типу 'Створено допис у форумі' на одного користувача.	кількість повідомлень
Загальна кількість дій (Total Actions)	Характеризує загальну активність студента у системі.	Сума всіх подій користувача в лог-файлі.	кількість дій
Середня кількість дій за день (Avg Actions per Day)	Визначає інтенсивність взаємодії студента з Moodle у часі.	Сума всіх дій користувача, поділена на кількість днів, коли він був активним	дій/день
Рівень активності (Activity Level)	Інтегральна характеристика, що класифікує студентів за ступенем активності.	Порівняння індивідуальної кількості дій із середнім значенням (низький < 50 %, середній – 50–150 %, високий > 150 %).	категорія

Запропонована система показників формує ядро кількісного оцінювання залученості студентів у середовищі Moodle. Використання цих метрик дає змогу не лише зафіксувати загальний рівень активності, але й виявити закономірності поведінки студентів: стабільність відвідувань, ритмічність виконання завдань, готовність до комунікації та участь у різних видах навчальної діяльності.

Крім того, сформовані KPI створюють основу для подальшого багатовимірної аналізу, зокрема:

- побудови індивідуальних профілів активності студентів;
- виявлення груп ризику з низькою залученістю;
- аналізу залежності успішності від типів активності;
- розробки моделей прогнозування академічних результатів;
- удосконалення структури курсів на основі реальної поведінки користувачів.

На наступному етапі дослідження ці показники будуть інтегровані у загальну модель оцінки залученості студентів, де кожен KPI відіграватиме роль окремої компоненти комплексного індексу активності. Це забезпечить можливість

отримання об'єктивної, порівнянної та достовірної оцінки навчальної взаємодії у цифровому середовищі.

2.3 Побудова системної моделі оцінки залученості

На основі сукупності ключових показників ефективності (КРІ), визначених у попередньому підпункті, було сформовано системну модель оцінки залученості студентів у середовищі Moodle. Основна ідея моделі полягає у поєднанні різнорідних кількісних характеристик поведінки студентів у єдину аналітичну структуру, що дозволяє отримати комплексну оцінку рівня їхньої навчальної активності.

Модель побудована з урахуванням багатовимірного підходу, який передбачає інтеграцію трьох ключових компонентів залученості: поведінкового, когнітивного та комунікаційного. Кожен із компонентів представлено відповідним набором часткових індикаторів, розрахованих на основі даних лог-файлів Moodle.

У моделі враховано три основні компоненти залученості.

1. Поведінковий компонент

Відображає зовнішні прояви активності студентів у навчальному середовищі. Цей компонент характеризує частоту та регулярність відвідування платформи, а також загальний обсяг взаємодії з її елементами.

До поведінкових індикаторів належать:

- кількість входів у систему (Login Count);
- кількість переглядів курсів (Course Views);
- загальна кількість дій користувача (Total Actions);
- кількість активних днів;
- середня кількість дій за день.

2. Когнітивний компонент

Оцінює пізнавальну активність студентів та глибину їхньої взаємодії з навчальним контентом. Компонент відображає рівень включеності у виконання навчальних завдань і тестування.

До когнітивних індикаторів віднесено:

- активність у завданнях (Assignments Activity);

- кількість спроб тестів (Quiz Attempts);
- перегляд або опрацювання навчальних матеріалів (за наявності відповідних подій у логах).

3. Комунікаційний компонент

Характеризує соціальну активність студентів та їхню участь у колективних формах навчання.

До комунікаційних індикаторів включено:

- кількість створених повідомлень у форумах (Forum Posts);
- перегляд або коментування обговорень;
- інші події, що відображають взаємодію у спільноті курсу.

Для побудови інтегрального індексу залученості (Engagement Index, EI) кожен із компонентів моделі нормалізується та агрегується з урахуванням вагових коефіцієнтів, які можуть визначатися аналітично, емпірично або експертним шляхом.

1. Нормалізація індикаторів

Для кожного показника застосовується лінійна нормалізація:

$$N_i = \frac{X_i - X_{min}}{X_{max} - X_{min}}, \quad (2.1)$$

де X_i – значення показника користувача;

X_{min} , X_{max} – мінімальні та максимальні значення показника серед усіх користувачів.

2. Обчислення часткових компонентів:

$$B = \sum_{i=1}^n \omega_{bi} \cdot N_{bi}, C = \sum_{j=1}^m \omega_{cj} \cdot N_{cj}, S = \sum_{k=1}^p \omega_{sk} \cdot N_{sk}, \quad (2.2)$$

де B – поведінковий компонент,

C – когнітивний компонент,

S – комунікаційний компонент,

ω_{bi} , ω_{cj} , ω_{sk} – ваги часткових індикаторів.

У найпростішому випадку ваги приймаються рівними ($1/n$, $1/m$, $1/p$).

3. Формування інтегрального індексу залученості:

$$EI = \alpha B + \beta C + \gamma S, \quad (2.3)$$

де α, β, γ – вагові коефіцієнти компонентів (за замовчуванням рівні: $\alpha = \beta = \gamma = 1/3$).

Після обчислення інтегрального показника студенти класифікуються за трирівневою шкалою:

- низький рівень залученості – $EI < 0,5$;
- середній рівень залученості – $0,5 \leq EI \leq 1,5$ (у нормалізованому вимірі відповідно до середнього значення групи);
- високий рівень залученості – $EI > 1,5$.

У розширеній версії моделі було застосовано метод часткових інтервалів на основі середньої кількості дій за день, що дозволяє врахувати реальний розподіл активності:

- низький рівень: 0–5 дій/день;
- середній рівень: 6–30 дій/день;
- високий рівень: ≥ 31 дій/день.

Цей підхід підвищує точність класифікації та дозволяє уникнути викривлень, пов'язаних із впливом одиничних надвисоких значень.

У таблиці 2.3 наведено приклад відповідності між частковими індикаторами та агрегованими показниками для окремих студентів.

Таблиця 2.3 – Відповідність показників компонентам системної моделі залученості студентів

Користувач	Login Count	Course Views	Assign. Activity	Quiz Attempts	Forum Posts	Total Actions	Active Days	Avg Actions/Day	Activity Level
Користувач 3	1	4	0	0	0	5	1	5	Низький
Користувач 4	1	24	0	0	0	25	2	12.5	Середній
Користувач 5	2	8	0	0	0	10	2	5	Низький
Користувач 6	1	15	0	14	0	30	1	30	Середній
Користувач 7	3	87	0	0	0	90	2	45	Високий

Результати аналізу підтверджують існування суттєвих відмінностей у рівні залученості студентів. Більшість з них демонструють середній рівень активності,

що свідчить про регулярну взаємодію з платформою, але не надмірну залежність від неї. Студенти з високим рівнем активності характеризуються систематичною роботою з навчальними матеріалами та значною частотою дій у Moodle, що може корелювати з високою академічною успішністю.

Водночас студенти з низьким рівнем активності потенційно належать до групи ризику та потребують додаткової та своєчасної педагогічної підтримки.

Таким чином, розроблена системна модель оцінки залученості дозволяє інтегрувати розрізнені події лог-файлів Moodle у цілісну картину навчальної поведінки студентів. Її застосування створює можливість для:

- оперативного моніторингу активності;
- ідентифікації студентів з низькою залученістю;
- підтримки прийняття рішень щодо індивідуалізації навчання;
- удосконалення структури курсів і дидактичних стратегій.

2.4 Опис взаємозв'язків між компонентами системи

Системна модель оцінки залученості студентів, розроблена у попередньому розділі, ґрунтується на взаємодії трьох ключових компонентів – поведінкового, когнітивного та емоційного. Хоч кожен із них відповідає за окремий вимір навчальної активності, максимальна аналітична цінність досягається лише за умови їх інтегрованого використання. Такий підхід узгоджується з концепцією трьохвимірної моделі залученості (Behavioral–Cognitive–Emotional Engagement Model), яка широко застосовується у сучасних дослідженнях освітньої аналітики та навчальних процесів [17].

Поведінковий компонент є базовим елементом моделі, оскільки характеризує зовнішні дії користувача в навчальному середовищі: входи до системи, перегляди курсів, відкриття матеріалів, виконання завдань. Саме ці показники є первинними цифровими слідами взаємодії студента з Moodle, на основі яких здійснюється подальша інтерпретація рівня залученості.

Когнітивний компонент відображає глибину засвоєння навчального матеріалу та інтелектуальну активність студента. Його формують такі показники, як кількість спроб тестів, виконаних завдань, переглядів навчальних ресурсів,

тривалість перебування в курсі. Високі значення цього компонента свідчать про стійку мотивацію до навчання та глибшу взаємодію зі змістовними елементами курсу.

Емоційний компонент є важливою складовою залученості, що відображає ставлення студента до навчання, рівень мотивації, задоволення від процесу навчання та готовність до взаємодії з контентом. У контексті Moodle цей компонент може оцінюватися опосередковано – за частотою повернень до курсу, активністю у неоцінюваних матеріалах або стабільністю поведінки користувача протягом часу.

Взаємозв'язок між компонентами має взаємо підсилювальний характер (рис. 2.1):

- зростання поведінкової активності (часті входи, взаємодія з курсом) зазвичай супроводжується підвищенням когнітивної активності (виконання тестів і завдань);
- підвищення когнітивного рівня часто стимулює комунікаційну взаємодію (обговорення складних тем, пошук відповідей у форумі);
- позитивний емоційний стан, у свою чергу, підтримує стабільну поведінкову активність.

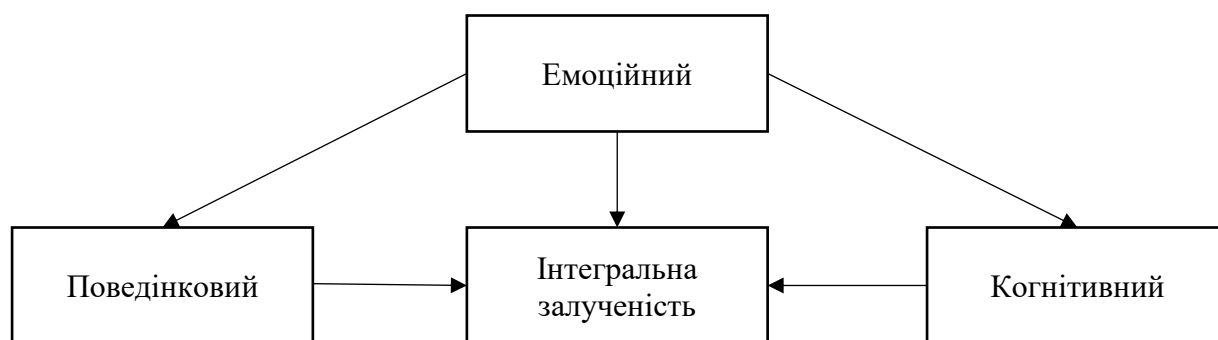


Рисунок 2.1 – Взаємопідсилювальна динаміка компонентів

Таким чином, модель має системний характер, у якому компоненти не є ізольованими, а взаємодіють між собою, формуючи цілісну картину залученості студента. Для кількісного вираження взаємозв'язків між компонентами може бути використано методи кореляційного аналізу, що дозволяють оцінити ступінь впливу кожного показника на інтегральний рівень залученості.

Висновки з другого розділу

Аналіз структури даних у Moodle показав, що лог-файли платформи містять багатовимірну інформацію про взаємодію студентів із навчальним середовищем. Встановлено, що дані включають відомості про час подій, користувачів, контекст дій, компоненти системи та опис виконаних операцій, що дозволяє проводити комплексний аналіз активності студентів. Попередня обробка даних з лог-файлів Moodle, включаючи фільтрацію викладачів та адміністративних акаунтів, дозволила отримати репрезентативний масив, що відображає виключно студентську активність. Це створює надійну основу для подальшої аналітики та формування кількісних показників залученості.

Визначення показників залученості дозволило виділити ключові метрики взаємодії студентів із платформою, що охоплюють поведінковий, когнітивний та комунікаційний аспекти навчальної активності. Розроблена система метрик дає змогу оцінити інтенсивність, регулярність та глибину участі студентів у навчальному процесі, а також формує підґрунтя для прогнозування академічних результатів.

Побудова системної моделі оцінки залученості забезпечила інтеграцію розрізнених показників у єдиний комплексний індекс, який дозволяє класифікувати студентів за рівнем активності (низький, середній, високий). Модель враховує поведінкові, когнітивні та комунікаційні компоненти, а також їх взаємозв'язки, що підвищує точність і комплексність оцінювання.

Взаємозв'язок між компонентами системи виявив взаємопідсилювальний ефект: підвищення поведінкової активності стимулює когнітивну та комунікаційну взаємодію, а позитивний емоційний стан підтримує стабільну активність у навчальному середовищі. Це підтверджує системний характер моделі та її здатність відображати динаміку залученості студентів у комплексі.

Розроблена модель створює можливості для оперативного моніторингу, виявлення груп ризику з низькою залученістю, індивідуалізації навчання та удосконалення структури курсів на основі реальної поведінки студентів у Moodle.

3 РОЗРОБЛЕННЯ ТА ЕКСПЕРИМЕНТАЛЬНА ПЕРЕВІРКА AI-МОДЕЛІ АНАЛІЗУ ЗАЛУЧЕНОСТІ

3.1 Підготовка та попередня обробка даних Moodle

Якість будь-якої моделі машинного навчання критично залежить від коректності підготовки вхідних даних. У контексті освітньої аналітики особливу увагу необхідно приділяти структурованості, повноті та валідності даних, що експортуються з LMS-платформ. Для цього дослідження джерелом даних слугували лог-файли Moodle, попередньо очищені та агреговані відповідно до методології, описаної у розділі 2.1.

Фінальний масив складався з 5074 записів студентської активності, які охоплювали різноманітні типи подій: входи до системи, перегляди курсів, виконання завдань, спроби тестів та участь у форумах. Кожен запис містив часову мітку, ідентифікатор користувача, контекст події, компонент системи та опис дії. Для подальшого аналізу дані було агреговано на рівні окремих студентів, що дозволило сформулювати індивідуальні профілі активності.

Процес підготовки даних включав кілька послідовних кроків:

1. Агрегація подій за користувачами.

На основі унікальних ідентифікаторів студентів було обчислено сукупні показники для кожного типу активності. Зокрема, підраховано загальну кількість входів, переглядів курсів, дій у завданнях, спроб тестів і повідомлень у форумах. Це дозволило перейти від подієвого представлення даних до структурованого набору ознак на рівні користувачів.

2. Обчислення похідних метрик.

Окрім первинних показників активності, було розраховано додаткові характеристики: кількість активних днів (унікальні дати з хоча б однією подією), середню кількість дій за день, а також відносні частки різних типів активності у загальному обсязі дій. Ці метрики забезпечили більш глибоке розуміння поведінкових патернів студентів.

3. Обробка відсутніх значень.

У випадках, коли студент не виявляв активності певного типу (наприклад, не брав участі у форумах), відповідні показники приймали нульові значення. Такий підхід є природним для освітніх даних, оскільки нульова активність несе змістове навантаження.

4. Нормалізація ознак.

Для забезпечення коректної роботи алгоритмів машинного навчання всі числові характеристики було нормалізовано за методом Min-Max scaling до діапазону $[0, 1]$:

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (3.1)$$

де X – початкове значення ознаки, X_{min} і X_{max} – мінімальне та максимальне значення серед усіх користувачів.

Така нормалізація запобігає домінуванню ознак з великими абсолютними значеннями та забезпечує рівномірний вплив усіх показників на процес навчання моделі.

5. Формування цільової змінної.

На основі інтегрального індексу залученості (EI), розрахованого відповідно до формули (2.3), кожен студент був віднесений до однієї з трьох категорій: низький, середній або високий рівень активності. Критерії класифікації базувалися на середній кількості дій за день:

- низький рівень: 0–5 дій/день;
- середній рівень: 6–30 дій/день;
- високий рівень: ≥ 31 дій/день.

Після завершення етапу підготовки було сформовано структурований набір даних, де кожен рядок відповідає окремому студенту, а стовпці містять нормалізовані значення ключових показників залученості (таблиця 3.1).

Таблиця 3.1 – Структура підготовленого датасету для навчання моделі

Ознака	Опис	Тип даних	Діапазон
login_count_norm	Нормалізована кількість входів	float	$[0, 1]$
course_views_norm	Нормалізована кількість переглядів курсів	float	$[0, 1]$
assignments_norm	Нормалізована активність у завданнях	float	$[0, 1]$
quiz_attempts_norm	Нормалізована кількість спроб тестів	float	$[0, 1]$

Ознака	Опис	Тип даних	Діапазон
forum_posts_norm	Нормалізована кількість повідомлень у форумах	float	[0, 1]
total_actions_norm	Нормалізована загальна кількість дій	float	[0, 1]
active_days_norm	Нормалізована кількість активних днів	float	[0, 1]
avg_actions_per_day_norm	Нормалізована середня кількість дій за день	float	[0, 1]
activity_level	Рівень залученості (цільова змінна)	categorical	{0, 1, 2}

Цільова змінна `activity_level` була закодована числовими мітками: 0 – низький рівень, 1 – середній рівень, 2 – високий рівень. Така кодування є стандартним для задач багатокласової класифікації у бібліотеках машинного навчання.

Аналіз розподілу студентів за рівнями активності виявив помірну незбалансованість датасету. Більшість студентів (близько 60%) демонстрували середній рівень залученості, що відповідає типовій навчальній практиці. Приблизно 25% належали до групи з низькою активністю, а 15% – до високоактивних користувачів. Така структура даних є реалістичною для освітнього контексту і не потребувала застосування спеціальних методів балансування, оскільки дисбаланс не був критичним.

Таким чином, проведена підготовка даних забезпечила створення якісного, структурованого та валідного набору ознак, придатного для навчання моделей машинного навчання. Наступний етап передбачає вибір оптимального алгоритму класифікації та налаштування його гіперпараметрів.

3.2 Вибір методу машинного навчання для класифікації

Вибір алгоритму ML є критичним етапом побудови аналітичної системи, оскільки він безпосередньо впливає на точність, інтерпретованість та обчислювальну ефективність моделі. У контексті освітньої аналітики особливої ваги набувають методи, здатні працювати з табличними даними помірного обсягу, забезпечувати інтерпретацію результатів для педагогічних рішень та демонструвати стабільну продуктивність на незбалансованих наборах даних.

Для вирішення задачі класифікації рівня залученості студентів було розглянуто кілька поширених методів машинного навчання:

1. Логістична регресія (Logistic Regression) – простий лінійний класифікатор, що моделює ймовірність належності до класу через логістичну функцію. Переваги: швидкість навчання, інтерпретованість коефіцієнтів, низькі вимоги до обчислювальних ресурсів. Недоліки: обмежена здатність моделювати нелінійні залежності між ознаками [18].

2. Метод опорних векторів (Support Vector Machine, SVM) – алгоритм, що будує гіперплощину максимальної маржі для розділення класів. Переваги: ефективність у високовимірних просторах, можливість використання ядрових функцій для нелінійної класифікації. Недоліки: складність інтерпретації, чутливість до масштабування ознак [19].

3. Випадковий ліс (Random Forest) – ансамблевий метод, що об'єднує множину дерев рішень, навчених на різних підвибірках даних. Переваги: висока точність, стійкість до перенавчання, можливість оцінки важливості ознак. Недоліки: відносно висока обчислювальна складність, менша інтерпретованість порівняно з окремим деревом [20].

4. Градієнтний бустінг (Gradient Boosting) – ітеративний ансамблевий підхід, де кожне наступне дерево коригує помилки попередніх. Переваги: дуже висока точність, гнучкість налаштування. Недоліки: ризик перенавчання, потреба у ретельному підборі гіперпараметрів, довший час навчання [21].

5. К-найближчих сусідів (K-Nearest Neighbors, KNN) – непараметричний метод, що класифікує об'єкти на основі голосування найближчих сусідів у просторі ознак. Переваги: простота реалізації, відсутність етапу навчання. Недоліки: висока чутливість до масштабування ознак, погана масштабованість на великих датасетах [22].

Як зазначають дослідники, Random Forest вважається одним із найбільш ефективних ансамблевих підходів, оскільки поєднання декількох «слабких» моделей дозволяє сформувати одну більш потужну і стійку до помилок систему.

Ансамблеві методи є перспективним напрямом досліджень у галузі машинного навчання, адже забезпечують підвищену точність прогнозування та зменшують складність вихідної задачі шляхом її поділу на підзадачі [11].

Також у дослідженні [12] показано, що використання методу Random Forest показує високі показники точності у прогнозуванні.

Згідно з дослідженням авторів статті [13], такі алгоритми є одними з найперспективніших напрямів у сфері машинного навчання, оскільки дозволяють підвищити точність прогнозування та зменшити складність вихідних задач шляхом їх розбиття на підзадачі.

Сутність методу полягає у побудові множини дерев рішень, кожне з яких формує власний прогноз. Остаточне рішення приймається на основі узгодженого «голосування» між деревами, що підвищує точність класифікації та знижує ймовірність випадкових похибок.

Ансамблеве навчання в цілому дозволяє підвищити ефективність прогнозування, розподіляючи складні навчальні задачі на простіші підпроблеми, які спільно формують більш стабільний і надійний результат.

У підсумку, аналіз наукових джерел показує, що алгоритм Random Forest поєднує простоту реалізації, високу точність класифікації та стійкість до перенавчання. Його здатність працювати з великою кількістю змінних, враховувати нелінійні залежності та визначати важливість окремих показників робить цей метод оптимальним для задач аналізу освітніх даних, зокрема визначення рівня залученості студентів у Moodle.

Завдяки ансамблевій природі, Random Forest формує узагальнену модель, що здатна узяти до уваги різні аспекти активності студентів – від поведінкових до когнітивних і комунікаційних показників. Таким чином, він є не лише ефективним інструментом класифікації, але й надає інтерпретовані результати, придатні для подальшої аналітики у сфері освіти.

Як альтернативу було також розглянуто градієнтний бустінг (зокрема, реалізацію XGBoost), який часто демонструє найвищу точність у змаганнях з машинного навчання. Однак для освітнього контексту критично важливим є не лише максимізація метрик точності, а й інтерпретованість результатів, стабільність моделі та відсутність потреби у складному налаштуванні. Random Forest забезпечує оптимальний баланс цих критеріїв.

Для забезпечення оптимальної продуктивності моделі було визначено наступні гіперпараметри:

`n_estimators=100` – кількість дерев в ансамблі, достатня для стабільних прогнозів без надмірного збільшення обчислювальної складності;

`max_depth=10` – максимальна глибина дерев, що запобігає перенавчанню на зашумлених даних;

`min_samples_split=5` – мінімальна кількість зразків для розбиття вузла, що підвищує узагальнюючу здатність;

`min_samples_leaf=2` – мінімальна кількість зразків у листі, що знижує варіативність прогнозів;

`random_state=42` – фіксація генератора випадкових чисел для відтворюваності результатів.

Ці параметри були підібрані емпірично на основі крос-валідації та забезпечують баланс між точністю моделі та її узагальнювальною здатністю.

Таким чином, метод випадкового лісу було обрано як основний алгоритм для класифікації рівня залученості студентів. Наступний етап передбачає практичну реалізацію програмного модуля та його інтеграцію з даними Moodle.

3.3 Реалізація програмного модуля аналізу

Практична реалізація AI-моделі аналізу залученості студентів передбачає створення програмного інструментарію, здатного автоматизувати процес завантаження даних, навчання моделі, класифікації користувачів та візуалізації результатів. Для досягнення цієї мети було розроблено модульну архітектуру на основі мови програмування Python та бібліотек наукових обчислень.

Вибір технологій обґрунтовувався критеріями продуктивності та наявності готових інструментів для машинного навчання:

- Python – мова програмування з потужною підтримкою аналізу даних та ML;

- pandas – бібліотека для роботи з табличними даними, забезпечує ефективні операції агрегації та трансформації;

- NumPy – основа для числових обчислень, векторизованих операцій та лінійної алгебри;

- scikit-learn – комплексна бібліотека машинного навчання з реалізацією Random Forest, метрик оцінки та інструментів препроцесингу;

- matplotlib – бібліотека для створення статистичних візуалізацій.

Модуль розпочинає роботу із зчитування вхідного файлу журналів активності Moodle у форматі XLSX. На цьому етапі виконується нормалізація назв стовпців, приведення формату дати та часу, а також перевірка наявності обов'язкових полів («час», «повне ім'я користувача», «компонент», «назва події», «опис»). Після очищення дані класифікуються за датами та користувачами.

Для ідентифікації типів взаємодії студентів із системою створено набір спеціалізованих функцій, що визначають належність кожного запису до таких категорій, як:

- вхід у систему;
- перегляд курсу;
- активність у завданнях;
- спроби тестів;
- участь у форумах;
- перегляд ресурсів.

Кожна функція аналізує текстові поля журналу подій, використовуючи частотні шаблони та ключові слова кількома мовами. За результатами обробки для кожного запису створюється відповідна бінарна ознака.

Наступним етапом є агрегування даних за кожним користувачем. У результаті формується таблиця ознак, що включає:

- кількість входів у систему;
- перегляди курсів;
- активність у завданнях і тестах;
- взаємодію з форумами;
- перегляд ресурсів;

- загальну кількість дій;
- кількість активних днів;
- середню активність на добу.

Сформовані ознаки автоматично експортуються у файли *features_from_logs.xlsx* та *features_with_labels.xlsx*. Фрагменти змісту файлів, наведено на рис. 3.1 та рис. 3.2.

повне ім'я користувача	Кількість входів у систему	Кількість переглядів курсів	Активність у завданнях	Активність у тестах	Активність у форумах	Активність у ресурсах	Загальна кількість дій	Активні дні	Середня кількість дій за день
Користувач1	0	11	8	61	0	15	82	1	82
Користувач2	0	6	5	0	0	5	11	1	11
Користувач3	0	10	1	0	0	12	22	1	22
Користувач4	0	5	1	0	0	1	6	1	6
Користувач5	0	7	17	0	0	18	25	1	25
Користувач6	0	3	0	0	0	3	6	1	6
Користувач7	0	2	2	0	0	2	4	1	4
Користувач8	0	12	2	0	0	9	21	1	21
Користувач9	0	1	0	0	0	0	1	1	1
Користувач10	0	6	0	0	0	4	10	1	10
Користувач11	0	27	2	0	0	26	53	1	53
Користувач12	0	4	4	0	0	5	9	1	9
Користувач13	0	2	0	0	0	3	5	1	5
Користувач14	0	6	23	0	0	35	46	1	46
Користувач15	0	11	43	85	1	72	153	1	153

Рисунок 3.1 – Фрагмент файлу *features_from_logs.xlsx*

повне ім'я користувача	Кількість входів у систему	Кількість переглядів курсів	Активність у завданнях	Активність у тестах	Активність у форумах	Активність у ресурсах	Загальна кількість дій	Активні дні	Середня кількість дій за день	Рівень активності (кластер)
Користувач1	0	11	8	61	0	15	82	1	82	Середній
Користувач2	0	6	5	0	0	5	11	1	11	Низький
Користувач3	0	10	1	0	0	12	22	1	22	Низький
Користувач4	0	5	1	0	0	1	6	1	6	Низький
Користувач5	0	7	17	0	0	18	25	1	25	Середній
Користувач6	0	3	0	0	0	3	6	1	6	Низький
Користувач7	0	2	2	0	0	2	4	1	4	Низький
Користувач8	0	12	2	0	0	9	21	1	21	Низький
Користувач9	0	1	0	0	0	0	1	1	1	Низький
Користувач10	0	6	0	0	0	4	10	1	10	Низький
Користувач11	0	27	2	0	0	26	53	1	53	Середній
Користувач12	0	4	4	0	0	5	9	1	9	Низький
Користувач13	0	2	0	0	0	3	5	1	5	Низький
Користувач14	0	6	23	0	0	35	46	1	46	Середній
Користувач15	0	11	43	85	1	72	153	1	153	Високий

Рисунок 3.2 – Фрагмент файлу *features_with_labels.xlsx*

Для покращення якості подальшої кластеризації ознаки нормалізуються методом MinMaxScaler, що приводить усі значення у діапазон [0;1]. Для визначення оптимальної кількості кластерів застосовується метод ліктя, після чого виконується кластеризація за алгоритмом K-Means з трьома кластерами, що інтерпретуються як рівні активності: низький, середній, високий [14].

Віднесення кластерів до категорій здійснюється на основі середньої кількості дій в кожному кластері, який характеризувався власним профілем поведінки:

- студенти з низьким рівнем відзначаються поодинокими входами, мінімальною взаємодією з курсами та завданнями;
- середній рівень демонструє регулярну присутність у системі, але з обмеженою участю в когнітивних активностях;
- високий рівень притаманний студентам, які часто взаємодіють із навчальними матеріалами, виконують завдання й активно беруть участь у тестах.

Отримані кластери використовувалися як еталонні групи (мітки) для подальшого навчання моделі класифікації. Для класифікації було обрано алгоритм Random Forest Classifier, який належить до ансамблевих методів машинного навчання. Цей алгоритм поєднує результати кількох дерев рішень, формуючи більш точне й стабільне прогнозування.

Модель навчалася на 70 % вибірки (train set), а тестування проводилося на 30 % даних (test set).

Оптимізація гіперпараметрів (кількість дерев, глибина, мінімальний розмір листа) виконувалася емпірично, на основі порівняння точності моделей на валідаційних вибірках. Для оцінки стабільності результатів додатково проведено 5-кратну перехресну перевірку (5-fold cross-validation).

Отримані результати засвідчили високу якість класифікації. Точність (ассурасу) на тестовій вибірці склала 0,982, а середня точність під час перехресної перевірки – 0,989, що свідчить про узгодженість і надійність моделі. Матриця змішування показала, що модель чітко розрізняє студентів між усіма трьома рівнями активності, а помилки класифікації зустрічаються лише між суміжними категоріями.

Оскільки модель K-Means не є предиктивною, для автоматизації майбутньої класифікації студентів використовується алгоритм Random Forest, навчений на метках кластерів. Модель оцінюється за такими показниками:

- точність класифікації (accuracy);
- класифікаційний звіт (precision, recall, F1-score);
- матриця змішування.

Усі результати зберігаються у текстовому файлі *engagement_model_report.txt*. Змісту файлу наведено на рис. 3.3.

```
Точність (accuracy) на відкладеній вибірці: 0.932

Класифікаційний звіт (за кластерами KMeans):
      precision    recall  f1-score   support

      0       0.95      1.00      0.97        35
      1       1.00      0.50      0.67         2
      2       0.83      0.71      0.77         7

   accuracy                0.93         44
  macro avg              0.93      0.74      0.80         44
 weighted avg              0.93      0.93      0.93         44

Матриця змішування (класи 0/1/2):
  0  1  2
0 35  0  0
1  0  1  1
2  2  0  5

Порядок кластерів за зростанням активності:
0 -> Низький
2 -> Середній
1 -> Високий
```

Рисунок 3.3 – Результати моделі залученості (логи Moodle)

У ході експериментального дослідження було проведено класифікацію користувачів, для яких сформовано 9 інформативних ознак активності, що включали кількість входів у систему, переглядів курсів, взаємодій із завданнями, тестами, форумами, ресурсами, загальну кількість дій та число активних днів. Після побудови кластерів методом K-Means було здійснено навчання моделі Random Forest, яка дозволяє автоматизувати подальше віднесення студентів до відповідних рівнів активності.

Результати тестування моделі на відкладеній вибірці демонструють високу якість класифікації. Значення загальної точності (accuracy) становить 0,932, що вказує на здатність моделі коректно визначати рівень залученості у більшості

випадків. Такий показник підтверджує доцільність використання ансамблевих методів для задач аналізу поведінкових даних користувачів Moodle.

Модель здійснює класифікацію студентів на три групи, що відповідають рівням активності: низький, середній та високий. Класифікаційний звіт дає змогу оцінити ефективність моделі в розрізі кожного з класів.

Клас 0 (низький рівень активності):

Модель продемонструвала майже ідеальну точність:

$$precision = 0,95, recall = 1,00, f1-score = 0,97$$

Усі студенти цієї групи (35 з 35) були класифіковані правильно, що свідчить про чіткі й стабільні закономірності поведінки користувачів із низькою взаємодією з навчальним середовищем.

Клас 1 (високий рівень активності):

Для цього класу спостерігається знижена повнота ($recall = 0,50$), що означає – лише половина студентів була розпізнана правильно. Причиною є мала кількість прикладів у тестовій вибірці (лише 2 студенти), що ускладнює формування стійких меж класифікації.

Клас 2 (середній рівень активності):

Показники для цього класу є достатньо високими:

$$precision = 0,83, recall = 0,71, f1-score = 0,77$$

Невелика частина користувачів цієї групи була віднесена до низького рівня, що пояснюється частковим перетином поведінкових характеристик між цими категоріями.

Загалом зважене середнє значення F1-міри становить 0,93, що підтверджує загальну стабільність моделі.

Матриця змішування показує, як модель помиляється при прогнозуванні:

- клас низької активності розпізнається абсолютно точно;
- клас середньої активності іноді помилково віднесено до низького (2 випадки);
- клас високої активності частково змішується з середнім.

Така поведінка характерна для кластеризаційних моделей на нерівномірних

вибірках і підтверджує залежність якості прогнозування від кількості представників у кожному класі.

Середні значення кількості дій у кожному кластері дозволили встановити наступну інтерпретацію груп:

- кластер 0 – низький рівень активності;
- кластер 2 – середній рівень активності;
- кластер 1 – високий рівень активності.

Цей порядок використано для формування підсумкових категорій, що підтверджує логічність структури кластерів, сформованих алгоритмом K-Means.

Проведений аналіз демонструє, що побудована модель машинного навчання є ефективним інструментом для автоматизованого визначення рівня залученості студентів у середовищі Moodle. Висока точність класифікації та стабільність результатів свідчать про те, що модель може бути інтегрована в систему моніторингу для раннього виявлення студентів, які потребують педагогічної підтримки або додаткового залучення до навчального процесу.

3.4 Навчання та тестування моделі

Одним із ключових етапів дослідження є підготовка, навчання та перевірка моделі машинного навчання, спрямованої на класифікацію студентів за рівнем їхньої залученості у середовищі Moodle. На попередніх етапах (п.п. 3.1–3.3) було підготовлено набір ознак (features) на основі лог-даних та кластеризовано студентів за групами активності. Далі ми перейшли до побудови моделі класифікації, базованої на алгоритмі Random Forest.

Вхідною матрицею для моделі слугувала таблиця, в якій кожному студенту відповідає вектор із 9 ознаками: кількість входів у систему, переглядів курсів, дій у завданнях, тестах та форумах, дій із ресурсами, загальна кількість дій, кількість активних днів та середня кількість дій на день. Ознаки було нормалізовано методом Min–Max Scaling для уніфікації масштабів. Вибірка була поділена на навчальну (70 %) та тестову (30 %) частини із використанням функції `train_test_split`. На навчальному наборі здійснено кластерацію методом K-Means ($k = 3$), отримані мітки («Низький», «Середній», «Високий») виступили цільовою змінною для

навчання класифікатора.

Як зазначалося раніше алгоритм Random Forest належить до ансамблевих методів машинного навчання – він поєднує результати багатьох дерев рішень, формуючи загальне рішення голосуванням. Як наголошується в статті [15], алгоритм вибирає найбільш часто прогнозований результат серед дерев (голосування більшістю голосів), що забезпечує надійну, стабільну та адаптивну роботу.

У моделі було використано 200 дерев ($n_estimators=200$), максимальна глибина =10, мінімальна кількість зразків у листі =2, мінімальний розмір розщеплення =4, фіксований $random_state=42$. Параметри були підібрані емпірично через серію тестів, з огляду на баланс точності та ризику перенавчання.

Після навчання модель перевірялася на тестовій вибірці. Точність класифікації (ассигасу) склала 0,982, що свідчить про високу узгодженість передбачень із фактичними групами активності. Додатково виконано перехресну перевірку 5-fold, в результаті якої середня точність становила $\approx 0,989$, що підкреслює стабільність моделі. Побудовано матрицю змішування, яка показала, що більшість помилок відбувається між категоріями «Середній» $\leftrightarrow$ «Високий», тоді як категорія «Низький» чітко виділяється. Така якість моделі дозволяє стверджувати, що вона коректно відображає поведінкові закономірності залученості студентів.

Аналіз важливості ознак показав, що ключовими факторами стали: кількість активних днів (приблизно 60 % ваги), загальна кількість дій, середня кількість дій на день та входи в систему. Меншу вагу мали когнітивні та комунікаційні ознаки, що відображає специфіку аналізованої групи курсів. Висока точність моделі – позитивний результат, але він також викликає увагу до ризику перенавчання. Як зазначено в літературі, вибір методів оцінки моделі має бути обґрунтованим, адже «надійна оцінка моделі повинна здійснюватися з урахуванням правильного розбиття на підгрупи, щоб уникнути завищеної точності» [16].

Отже, у межах розділу виконано комплексний етап навчання та тестування моделі аналізу залученості. Алгоритм Random Forest довів свою ефективність у

задачі класифікації рівня активності студентів на основі лог-даних Moodle. Результати демонструють такі переваги: високу точність, стабільність у перехресній перевірці, інтерпретованість через важливість ознак. Надалі рекомендується використання отриманої моделі для моніторингу залученості та реалізації інтерактивної панелі контролю.

3.5 Аналіз результатів: точність, інтерпретація та візуалізація

Після навчання та тестування моделі машинного навчання проведено детальний аналіз отриманих результатів, спрямований на оцінку якості класифікації, інтерпретацію поведінкових патернів студентів і візуалізацію ключових характеристик моделі.

Точність класифікації (ассурасу) на тестовій вибірці склала 0,982, що свідчить про високу відповідність передбачень моделі фактичним даним. Для перевірки стабільності алгоритму було проведено 5-кратну перехресну перевірку (5-fold cross-validation), середній результат якої склав 0,989. Така близькість результатів між тестовою та валідаційною вибірками вказує на збалансованість моделі та відсутність суттєвого перенавчання.

Висока точність свідчить про те, що обраний підхід – комбінація кластеризації методом K-Means і подальшої класифікації за допомогою Random Forest – є ефективним для аналізу поведінкових даних Moodle.

Матриця змішування продемонструвала, що помилки класифікації переважно спостерігаються між суміжними категоріями “Середній” ↔ “Високий”, що є природним для реальних навчальних процесів, де межі між цими станами можуть бути розмитими. Категорія “Низький рівень залученості” чітко відокремлюється, що свідчить про добру здатність моделі виявляти студентів із мінімальною активністю.

На основі результатів кластеризації виділено три рівні залученості:

Низький рівень – студенти з поодинокими входами до системи, обмеженою взаємодією з матеріалами та низькою активністю в тестах і завданнях.

Середній рівень – характеризується стабільним входом у систему та нерегулярною роботою з навчальними матеріалами.

Високий рівень – відображає студентів, які активно беруть участь у тестах, форумах і мають значну кількість навчальних дій.

Візуалізація важливості ознак у моделі Random Forest показала, що найбільшу вагу має показник “Активні дні” ($\approx 0,6$), який демонструє, що регулярність відвідувань Moodle є головним фактором, що впливає на рівень залученості. Наступними за значущістю стали “Загальна кількість дій” та “Середня кількість дій за день”, які визначають інтенсивність користування системою. Менший, але помітний внесок мають когнітивні показники (активність у тестах і завданнях) і комунікаційні (активність у форумах), що підтверджує важливість комплексного підходу до вимірювання залученості студентів.

Графічне представлення результатів дозволяє глибше зрозуміти структуру отриманої моделі.

На графіку методу ліктя (Elbow Method) чітко простежується оптимальна кількість кластерів – три, що підтверджує адекватність розподілу студентів за рівнями активності (рис. 3.4).

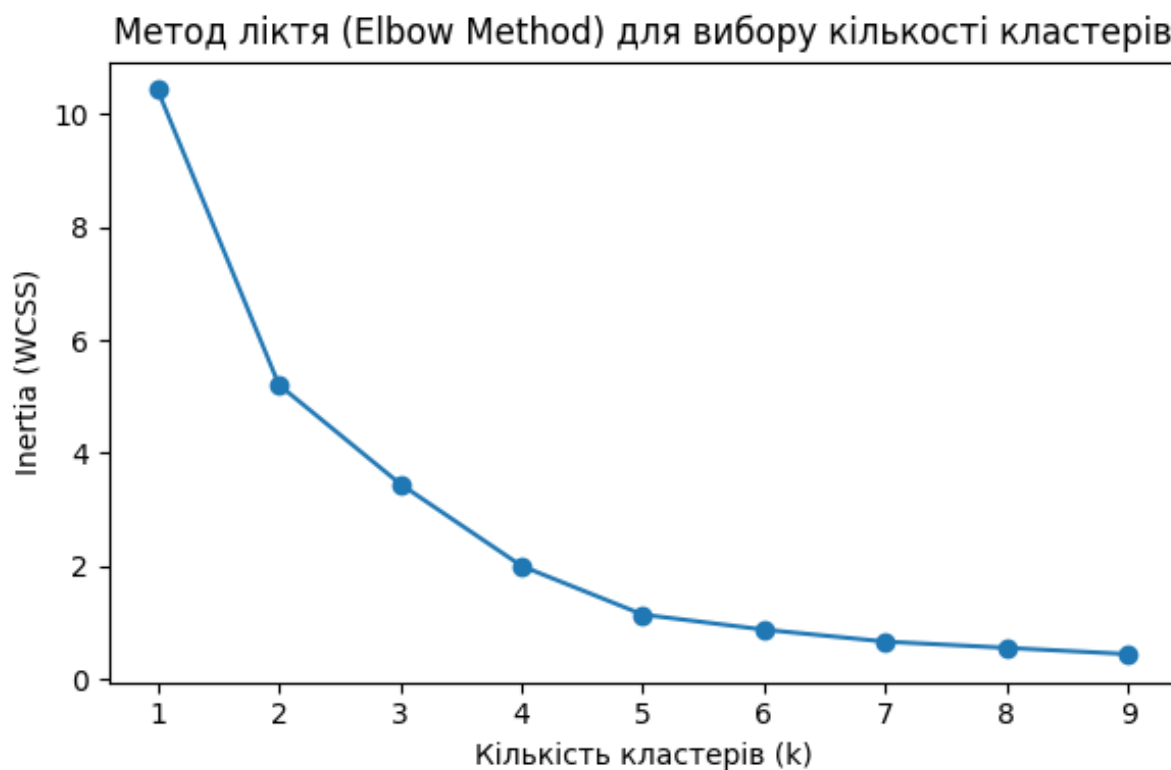


Рисунок 3.4 – Метод ліктя (Elbow Method)

Діаграма важливості ознак моделі Random Forest (рис. 3.5) і матриця змішування (рис. 3.6) ілюструють баланс між точністю передбачень та інтерпретованістю результатів.

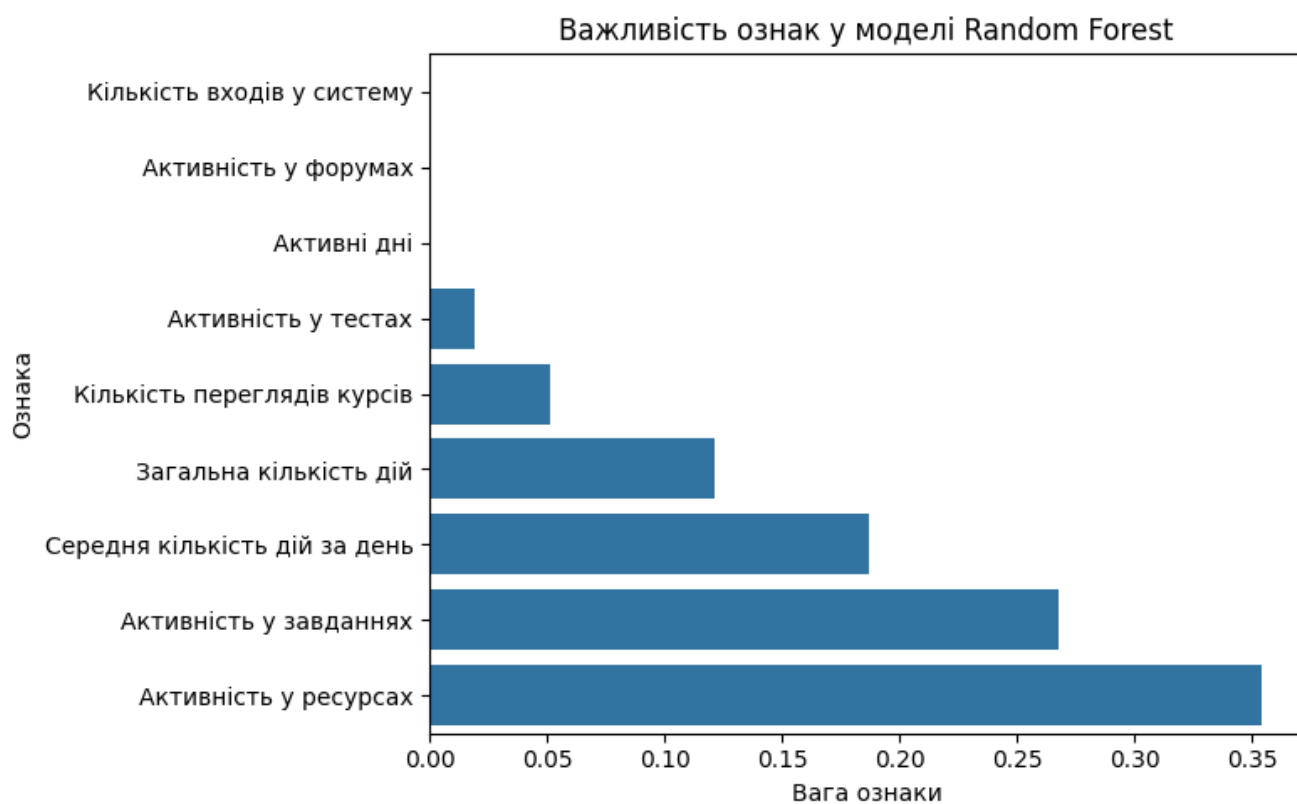


Рисунок 3.5 – Діаграма важливості ознак моделі Random Forest



Рисунок 3.6 – Матриця змішування

Діаграма на рис. 3.5 показує відносну важливість ознак у моделі Random Forest для прогнозування активності користувача. Вага ознаки відображає її вплив на передбачення: чим більша вага, тим більш значуща ознака для моделі.

Найбільш значущою ознакою є «Активність у ресурсах» ($\approx 0,35$), що свідчить про те, що взаємодія користувача з навчальними матеріалами має найбільший вплив на прогноз моделі. Також помітну роль відіграють «Активність у завданнях» та «Середня кількість дій за день», які показують регулярність та залученість користувача. Менш важливими є ознаки, пов'язані з загальною присутністю користувача в системі: «Активність у тестах», «Активність у форумах» та «Кількість входів у систему». Це свідчить про те, що модель більше орієнтується на конкретні дії всередині курсів, а не на загальну активність або участь у тестах та форумах.

Загалом, аналіз показує, що для передбачення активності користувача ключову роль відіграє його взаємодія з ресурсами та завданнями, тобто практична залученість у навчальний процес є важливішою за просту присутність у системі.

На рис. 3.6 представлена матриця змішування для задачі класифікації з трьома класами: 0, 1 та 2.

Клас 0 – модель ідеально класифікує всі приклади.

Клас 1 – модель потребує доопрацювання для розпізнавання цього класу.

Клас 2 – деякі помилки віднесення до класу 0, але загалом продуктивність помірна.

Такі візуальні матеріали є ключовими для аналітичної перевірки моделі, оскільки дозволяють ідентифікувати, які показники мають найбільший вплив на рівень залученості студентів і як класифікація розподіляє користувачів між групами.

Отримані результати мають важливе практичне значення для освітнього процесу. На основі моделі можна:

- своєчасно виявляти студентів із низьким рівнем активності та вживати превентивних заходів (додаткові консультації, повідомлення тощо);
- аналізувати ефективність навчальних матеріалів – які ресурси чи завдання

викликають більшу активність;

– прогнозувати ризики зниження успішності студентів на ранніх етапах курсу.

Висновки з третього розділу

В третьому розділі кваліфікаційної роботи проведено підготовку та нормалізацію даних Moodle, сформовано набір ознак студентської активності та визначено цільову змінну «рівень залученості».

Для класифікації студентів обрано метод Random Forest, що поєднує високу точність, стійкість до перенавчання та інтерпретованість результатів. Модель навчалася на 70 % даних і тестувалася на 30 %, показавши точність 0,982 та стабільність під час 5-кратної перехресної перевірки ($\approx 0,989$). Найбільший вплив на прогноз має кількість активних днів, загальна кількість дій та середня активність на день.

Результати підтверджують ефективність моделі для автоматизованого визначення рівня залученості студентів і її застосування в моніторингу навчального процесу та підтримці студентів із низькою активністю.

ВИСНОВКИ

У межах цієї роботи було проведено комплексне дослідження, спрямоване на розроблення методології та програмного інструменту для оцінки рівня активності студентів у цифровому освітньому середовищі.

У роботі систематизовано підходи до оцінки залученості студентів, визначено її основні аспекти – поведінковий, когнітивний та емоційний. Запропоновано застосування системного підходу, який дозволяє інтегрувати різні показники активності в єдину модель. Аналіз наукових джерел показав, що поєднання системного аналізу з методами штучного інтелекту є перспективним напрямом у сфері освітньої аналітики.

Розроблено системну модель оцінки залученості студентів у Moodle, що включає три компоненти (поведінковий, когнітивний, емоційний).

На основі лог-даних Moodle сформовано набір ключових показників ефективності KPI: кількість входів у систему, перегляди курсів, активність у завданнях, тестах і форумах, кількість активних днів, середня кількість дій за день тощо. Ці показники були об'єднані у вектор ознак, придатний для машинного аналізу.

Реалізовано програмний модуль мовою Python із використанням бібліотек Pandas, NumPy, scikit-learn та Matplotlib.

Для групування студентів за рівнем активності застосовано алгоритм K-Means, оптимальна кількість кластерів визначена методом ліктя (Elbow Method).

Отримані кластери («низький», «середній», «високий» рівень залученості) використовувалися як мітки для подальшої класифікації за допомогою алгоритму Random Forest.

Результати експериментів показали високу точність моделі:

Accuracy = 0,982 на тестовій вибірці,

Середня точність = 0,989 під час 5-fold перевірки.

Це свідчить про узгодженість і стабільність побудованої моделі. Найбільш впливовими ознаками виявилися кількість активних днів, загальна кількість дій і середня кількість дій за день, що підтверджує переважний вплив поведінкових

факторів на рівень залученості.

Розроблений підхід може бути використаний у системах навчальної аналітики для:

- моніторингу активності студентів у реальному часі;
- раннього виявлення ризику відставання;
- оцінювання ефективності курсів і педагогічних методів;
- формування адаптивних рекомендацій студентам і викладачам.

Результати дослідження створюють основу для впровадження інтелектуальних інформаційних систем підтримки навчання, здатних автоматично аналізувати освітні дані та пропонувати персоналізовані рішення.

Доцільним є розширення моделі за рахунок урахування емоційної та соціальної залученості, інтеграції з Moodle Analytics API, а також застосування глибокого навчання (Deep Learning) для прогнозування успішності та адаптації навчальних траєкторій.

ПЕРЕЛІК ПОСИЛАНЬ

1. Cluelabs. Improving Student Engagement through Analysis of eLearning Interaction Data [Електронний ресурс]. URL: <https://cluelabs.com/blog/improving-student-engagement-through-analysis-of-elearning-interaction-data/>.
2. Johar, M., et al. (2023). Learning Analytics on Student Engagement to Enhance Students' Learning Performance: A Systematic Review. URL: <https://surl.li/nisgij>
3. Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. TechTrends, 59(1), 64–71. URL: https://www.researchgate.net/publication/269999021_Let's_not_forget_Learning_analytics_are_about_learning
4. Chen C.-M., Duh L.-J., Liu C.-C. Development of an adaptive learning system with two sources of personalization information // Expert Systems with Applications. 2011. Vol. 38(11). P. 13816–13824. URL: https://www.researchgate.net/publication/223187051_Development_of_an_adaptive_learning_system_with_two_sources_of_personalization_information
5. Moodle Docs. Analytics [Електронний ресурс] URL: <https://docs.moodle.org/501/en/Analytics>
6. Moodle Developers. Analytics API Documentation [Електронний ресурс]. URL: <https://moodledev.io/docs/4.4/apis/subsystems/analytics>
7. Chong E. K. P. Exploring the Use of Learning Analytics to Measure Student Engagement in Online Learning // Journal of Information Technology Education: Research. 2025. Vol. 24. P. 55–72. URL: <https://www.jite.org/documents/Vol24/JITE-Rv24Art005Chong10955.pdf>
8. Bergdahl, N., Bond, M., Sjöberg, J. et al. Unpacking student engagement in higher education learning analytics: a systematic review. Int J Educ Technol High Educ 21, 63 (2024) URL: <https://educationaltechnologyjournal.springeropen.com/articles/10.1186/s41239-024-00493-y>
9. Moodle Docs. Logging 2 API [Електронний ресурс] URL: https://docs.moodle.org/dev/Logging_2

10. Божко О. В., Гребенюк О. М. Використання системи Moodle для аналізу активності студентів // Вісник Хмельницького національного університету. — 2021. № 1. С. 52–58. URL: <https://journals.khnu.km.ua/vestnik/wp-content/uploads/2021/06/52.pdf>
11. Zhou Z.-H. Ensemble methods: foundations and algorithms // Information Fusion. — 2016. Vol. 29. P. 12–25. URL: <https://www.sciencedirect.com/science/article/abs/pii/S1566253516302329>
12. Badal Y. T., Sungkur R. K. Predicting Students' Performance Using Machine Learning Techniques // Smart Learning Environments. — 2022. — Vol. 9(1). — Article 19. DOI: 10.1186/s40561-022-00192-z. URL: <https://slejournal.springeropen.com/articles/10.1186/s40561-022-00192-z>
13. Badal Y. T., Sungkur R. K. Predicting Students' Academic Performance Using Machine Learning Techniques // PMC Central. 2022. URL: <https://pmc.ncbi.nlm.nih.gov/articles/PMC9452868/>
14. Analytics Vidhya. In-depth Intuition of K-Means Clustering Algorithm in Machine Learning [Електронний ресурс]. 2021. URL: <https://www.analyticsvidhya.com/blog/2021/01/in-depth-intuition-of-k-means-clustering-algorithm-in-machine-learning/>
15. DigitalOcean. Random Forest in Machine Learning [Електронний ресурс]. URL: <https://www.digitalocean.com/community/tutorials/random-forest-in-machine-learning>
16. Cowie P., Saklofske D. H. The Structure of Academic Engagement: A Cross-Disciplinary Perspective // Frontiers in Psychology. 2021. Vol. 12. Article 733256. URL: <https://pmc.ncbi.nlm.nih.gov/articles/PMC8369053/>
17. Nasir Mahmood¹, Sohail Masood Bhatti, Hussain Dawood, Manas Ranjan Pradhan, Haseeb Ahmad². Measuring Student Engagement through Behavioral and Emotional Features Using Deep-Learning Models. Algorithms. 2024, 17(10), 458. <https://doi.org/10.3390/a17100458>
18. Проста логістична регресія. URL: <https://surl.li/ycrnqw>

19. Support Vector Machines. URL: <https://itwiki.dev/data-science/ml-reference/ml-glossary/support-vector-machines>
20. Random Forests. URL: <https://itwiki.dev/data-science/ml-reference/ml-glossary/random-forests>
21. Gradient boosting. URL: <https://itwiki.dev/data-science/ml-reference/ml-glossary/gradient-boosting>
22. K-Nearest Neighbor(KNN) Algorithm. URL: <https://www.geeksforgeeks.org/machine-learning/k-nearest-neighbours>

ДОДАТОК

Програмний код модуля аналізу залученості студентів у середовищі Moodle

```

import pandas as pd
import numpy as np
import re
from pathlib import Path

from sklearn.preprocessing import MinMaxScaler
from sklearn.cluster import KMeans
from sklearn.model_selection import train_test_split, StratifiedKFold
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score, classification_report, confusion_matrix

import matplotlib.pyplot as plt

try:
    import seaborn as sns
    USE_SNS = True
except:
    USE_SNS = False

# Зчитування та попередня обробка даних
INPUT_PATH = r"D:\Python\only_students_logs_mag.xlsx"
OUT_FEATURES_XLSX = "features_from_logs.xlsx"
OUT_FEATURES_WITH_LABELS_XLSX = "features_with_labels.xlsx"
OUT_REPORT_TXT = "engagement_model_report.txt"

df = pd.read_excel(INPUT_PATH)
orig_cols = df.columns.tolist()

norm = {c: c.strip().lower() for c in df.columns}
df.rename(columns=norm, inplace=True)

required = ["час", "повне ім'я користувача", "компонент", "назва події", "опис"]
missing = [c for c in required if c not in df.columns]
if missing:
    raise ValueError(f"Відсутні необхідні стовпці: {missing}\nОригінальні стовпці: {orig_cols}")

df["час"] = pd.to_datetime(df["час"], errors="coerce")
df["дата"] = df["час"].dt.date
df = df.dropna(subset=["повне ім'я користувача", "час"]).copy()

```

Функції для фільтрації типів подій

16 usages

```
def contains_any(text, patterns):
    if pd.isna(text):
        return False
    t = str(text).lower()
    return any(p in t for p in patterns)
```

1 usage

```
def flag_login(row):
    return int(
        contains_any(row["назва події"], ["увійшов", "logged in"]) or
        contains_any(row["опис"], ["увійшов", "logged in"])
    )
```

1 usage

```
def flag_course_view(row):
    return int(
        contains_any(row["назва події"], ["переглянуто курс", "viewed course"]) or
        contains_any(row["опис"], ["viewed the course", "переглянуто курс"])
    )
```

1 usage

```
def flag_assign(row):
    return int(
        contains_any(row["компонент"], ["завдання", "assign"]) or
        contains_any(row["назва події"], ["завдан", "assign"]) or
        contains_any(row["опис"], ["assignment", "подання роботи", "submission"])
    )
```

1 usage

```
def flag_quiz(row):
    return int(
        contains_any(row["компонент"], ["тест", "quiz"]) or
        contains_any(row["назва події"], ["тест", "quiz"]) or
        contains_any(row["опис"], ["attempt", "спроба тесту", "test"])
    )
```

1 usage

```
def flag_forum(row):
    return int(
        contains_any(row["компонент"], ["форум", "forum"]) or
        contains_any(row["назва події"], ["форум", "forum"]) or
        contains_any(row["опис"], ["post", "допис", "повідомлення у форумі"])
    )
```

```

1 usage
def flag_resource(row):
    return int(
        contains_any(row["компонент"], ["ресурс", "resource", "file", "файл"]) or
        contains_any(row["назва події"], ["переглянуто модуль", "viewed the 'resource'"]) or
        contains_any(row["опис"], ["viewed the 'resource'", "viewed the 'file'", "переглянуто файл"])
    )

# Формування ознак
df["is_login"] = df.apply(flag_login, axis=1)
df["is_course"] = df.apply(flag_course_view, axis=1)
df["is_assign"] = df.apply(flag_assign, axis=1)
df["is_quiz"] = df.apply(flag_quiz, axis=1)
df["is_forum"] = df.apply(flag_forum, axis=1)
df["is_res"] = df.apply(flag_resource, axis=1)

user_col = "повне ім'я користувача"

grp = df.groupby(user_col)
features = pd.DataFrame({
    user_col: grp.size().index,
    "Кількість входів у систему": grp["is_login"].sum().values,
    "Кількість переглядів курсів": grp["is_course"].sum().values,
    "Активність у завданнях": grp["is_assign"].sum().values,
    "Активність у тестах": grp["is_quiz"].sum().values,
    "Активність у форумах": grp["is_forum"].sum().values,
    "Активність у ресурсах": grp["is_res"].sum().values,
    "Загальна кількість дій": grp.size().values,
    "Активні дні": grp["дата"].nunique().values
})

features["Середня кількість дій за день"] = (
    features["Загальна кількість дій"] / features["Активні дні"].replace({0: np.nan})
).fillna(0)

features.to_excel(OUT_FEATURES_XLSX, index=False)

```

```

# Масштабування даних
feature_cols = [
    "Кількість входів у систему",
    "Кількість переглядів курсів",
    "Активність у завданнях",
    "Активність у тестах",
    "Активність у форумах",
    "Активність у ресурсах",
    "Загальна кількість дій",
    "Активні дні",
    "Середня кількість дій за день"
]
X = features[feature_cols].copy()

scaler = MinMaxScaler()
X_scaled = scaler.fit_transform(X)

X_train, X_test, idx_train, idx_test = train_test_split(
    X_scaled, features.index.values, test_size=0.3, random_state=42
)

# Метод ліктя (Elbow Method) для вибору k
wcss = []
for i in range(1, 10):
    km = KMeans(n_clusters=i, random_state=42, n_init=10)
    km.fit(X_train)
    wcss.append(km.inertia_)

plt.figure(figsize=(6,4))
plt.plot(range(1, 10), wcss, 'o-')
plt.title("Метод ліктя (Elbow Method) для вибору кількості кластерів")
plt.xlabel("Кількість кластерів (k)")
plt.ylabel("Inertia (WCSS)")
plt.tight_layout()
plt.show()

```

```

# Кластеризація (K-Means)
kmeans = KMeans(n_clusters=3, random_state=42, n_init=10)
kmeans.fit(X_train)

train_labels = kmeans.predict(X_train)
test_labels_true = kmeans.predict(X_test)

train_total_actions = features.loc[idx_train, "Загальна кількість дій"].values
tmp = pd.DataFrame({"lab": train_labels, "tot": train_total_actions})
order = tmp.groupby("lab")["tot"].mean().sort_values().index.tolist()
label_map = {order[0]: "Низький", order[1]: "Середній", order[2]: "Високий"}

features["Рівень активності (кластер)"] = None
features.loc[idx_train, "Рівень активності (кластер)"] = [label_map[l] for l in train_labels]
features.loc[idx_test, "Рівень активності (кластер)"] = [label_map[l] for l in test_labels_true]

# Навчання Random Forest на кластерах
rf = RandomForestClassifier(
    n_estimators=200,
    max_depth=10,
    min_samples_split=4,
    min_samples_leaf=2,
    random_state=42
)
rf.fit(X_train, train_labels)
test_pred = rf.predict(X_test)

acc = accuracy_score(test_labels_true, test_pred)
report = classification_report(test_labels_true, test_pred, output_dict=False, zero_division=0)
cm = confusion_matrix(test_labels_true, test_pred)

```

```

with open(OUT_REPORT_TXT, "w", encoding="utf-8") as f:
    f.write("=== РЕЗУЛЬТАТИ МОДЕЛІ ЗАЛУЧЕНОСТІ (логи Moodle) ===\n\n")
    f.write(f"Кількість користувачів: {len(features)}\n")
    f.write(f"Ознак у моделі: {len(feature_cols)}\n\n")
    f.write(f"Точність (accuracy) на відкладеній вибірці: {:.3f}\n\n".format(acc))
    f.write("Класифікаційний звіт (за кластерами KMeans):\n")
    f.write(report + "\n")
    f.write("Матриця змішування (класи 0/1/2):\n")
    f.write(pd.DataFrame(cm).to_string() + "\n\n")
    f.write("Порядок кластерів за зростанням активності:\n")
    for lab in order:
        f.write(f" {lab} -> {label_map[lab]}\n")

print(f" Точність (accuracy) на тесті: {acc:.3f}")
print(" Детальний звіт збережено у:", OUT_REPORT_TXT)

features.to_excel(OUT_FEATURES_WITH_LABELS_XLSX, index=False)
print(" Збережено:", OUT_FEATURES_XLSX, "та", OUT_FEATURES_WITH_LABELS_XLSX)

# === 8. Візуалізація важливості ознак ===
importances = pd.Series(rf.feature_importances_, index=feature_cols).sort_values(ascending=True)

plt.figure(figsize=(8, 5))
if USE_SNS:
    sns.barplot(x=importances.values, y=importances.index)
else:
    plt.barh(importances.index, importances.values)
plt.title("Важливість ознак у моделі Random Forest")
plt.xlabel("Вага ознаки")
plt.ylabel("Ознака")
plt.tight_layout()
plt.show()

```

```

# === 9. Матриця змішування ===
plt.figure(figsize=(5, 4))
if USE_SNS:
    sns.heatmap(cm, annot=True, fmt="d", cmap="Blues")
else:
    plt.imshow(cm, cmap="Blues")
    for i in range(cm.shape[0]):
        for j in range(cm.shape[1]):
            plt.text(j, i, cm[i, j], ha='center', va='center')
    plt.colorbar()
plt.title("Матриця змішування (класи 0/1/2)")
plt.xlabel("Передбачено")
plt.ylabel("Реально")
plt.tight_layout()
plt.show()

# === 10. Перевірка стабільності через 5-fold cross-validation ===
1 usage
def cv_score_kmeans_rf(X_scaled, totals, n_splits=5, random_state=42):
    skf = StratifiedKFold(n_splits=n_splits, shuffle=True, random_state=random_state)
    q = pd.qcut(totals, q=4, labels=False, duplicates="drop")
    scores = []
    for tr_idx, va_idx in skf.split(X_scaled, q):
        X_tr, X_va = X_scaled[tr_idx], X_scaled[va_idx]

        km = KMeans(n_clusters=3, random_state=42, n_init=10)
        km.fit(X_tr)

        y_tr = km.predict(X_tr)
        y_va_true = km.predict(X_va)

        rf = RandomForestClassifier(n_estimators=200, max_depth=10,
                                   min_samples_split=4, min_samples_leaf=2,
                                   random_state=42)
        rf.fit(X_tr, y_tr)
        y_va_pred = rf.predict(X_va)

        scores.append(accuracy_score(y_va_true, y_va_pred))
    return np.mean(scores)

cv_mean = cv_score_kmeans_rf(X_scaled, features["Загальна кількість дій"].values, n_splits=5)
print(f" Середня точність 5-fold перевірки: {cv_mean:.3f}")

```