

**НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ БІОРЕСУРСІВ  
І ПРИРОДОКОРИСТУВАННЯ УКРАЇНИ**

**Факультет інформаційних технологій**

**ПОГОДЖЕНО**

Декан факультету

**ДОПУСКАЄТЬСЯ ДО ЗАХИСТУ**

Завідувач кафедри

Інформаційних технологій

(назва факультету)

комп'ютерних наук

(назва кафедри)

\_\_\_\_\_ Ігор Болбот  
(підпис) (ім'я та прізвище)

\_\_\_\_\_ Белла Голуб  
(підпис) (ім'я та прізвище)

“ \_\_\_ ” \_\_\_\_\_ 2025 р.

“ \_\_\_ ” \_\_\_\_\_ 2025 р.

**МАГІСТЕРСЬКА КВАЛІФІКАЦІЙНА РОБОТА**

на тему Аналітична система оцінювання якості навчального процесу

Спеціальність \_\_\_\_\_ 122 «Комп'ютерні науки»  
(Код і найменування)

Освітня програма Інформаційні управляючі системи та технології  
(Назва)

Орієнтація освітньої програми \_\_\_\_\_ Освітньо-професійна  
(освітньо-професійна або освітньо-наукова)

**Гарант освітньої програми**

\_\_\_\_\_ К.Т.Н., доцент  
(науковий ступінь та вчене звання)

\_\_\_\_\_ (підпис)

\_\_\_\_\_ Белла Голуб  
(ім'я та прізвище)

**Керівник магістерської кваліфікаційної роботи**

\_\_\_\_\_ К.Т.Н., доцент  
(науковий ступінь та вчене звання)

\_\_\_\_\_ (підпис)

\_\_\_\_\_ Белла Голуб  
(ім'я та прізвище)

**Виконав**

\_\_\_\_\_ (підпис)

\_\_\_\_\_ Богдан Кочубей  
(ім'я та прізвище студента)

**НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ БІОРЕСУРСІВ  
І ПРИРОДОКОРИСТУВАННЯ УКРАЇНИ**

Факультет (ННІ) інформаційних технологій

**ЗАТВЕРДЖУЮ**

**Завідувач кафедри комп'ютерних наук**  
**доцент, к.т.н. Белла Голуб**  
(науковий ступінь, вчене звання) (підпис) (ім'я та прізвище)  
**"01" листопада 2024 року**

**З А В Д А Н Н Я**

**ДО ВИКОНАННЯ МАГІСТЕРСЬКОЇ КВАЛІФІКАЦІЙНОЇ РОБОТИ СТУДЕНТУ**

Кочубею Богдану Богдановичу

(прізвище, ім'я, по батькові)

Спеціальність 122 «Комп'ютерні науки»

(код і назва)

Освітня програма Інформаційні управляючі системи та технології

(назва)

Орієнтація освітньої програми освітньо-професійна

Тема магістерської кваліфікаційної роботи Аналітична система оцінювання якості навчального процесу

затверджена наказом ректора НУБіП України від "01" листопада 2024р. №1964 «С»

Термін подання завершеної роботи на кафедру 01.12.2025

(рік, місяць, число)

Вихідні дані до магістерської кваліфікаційної роботи: синтетично згенеровані освітні дані, що моделюють структуру навчальних груп, результати тестувань студентів, рівень відвідуваності та відгуки, а також метрики якості курсів і викладацької діяльності.

Перелік питань, що підлягають дослідженню:

1. Оцінювання динаміки та якості навчального процесу із застосуванням OLAP-технологій.

2. Розробка та використання ключових показників ефективності (KPI) для моніторингу курсів та рівня задоволеності студентів.

3. Дослідження можливостей застосування методів Data Mining для освітніх даних (класифікація, пошук асоціативних правил, кластеризація).

4. Виявлення типових освітніх траєкторій та груп ризику серед студентів на основі результатів кластерного аналізу.

Перелік графічного матеріалу (за потреби)

Дата видачі завдання "01" листопада 2024 р.

**Керівник магістерської кваліфікаційної роботи**

Белла Голуб

( підпис )

(ім'я та прізвище)

**Завдання прийняв до виконання**

Богдан Кочубей

( підпис )

(ім'я та прізвище)

## Календарний план

№ з/п	Назва етапів виконання магістерської кваліфікаційної роботи	Строк виконання етапів магістерської кваліфікаційної роботи	Примітка
1	Видача завдання	01.11.2024	
2	Аналіз предметної області	01.11.2024 – 10.12.2025	
3	Моделювання предметної області	11.12.2024 – 23.01.2025	
4	Розробка системи	24.01.2025 – 18.05.2025	
5	Аналіз результатів	19.05.2025 – 31.08.2025	
6	Оформлення записки	01.09.2025 – 01.11.2025	
7	Постерна сесія	28.10.2025 – 29.10.2025	
8	Перевірка на плагіат	13.11.2025	
9	Попередній захист	01.12.2025	
10	Захист	16.12.2025	

Студент \_\_\_\_\_ Богдан Кочубей  
(підпис) (ім'я та прізвище)

Керівник магістерської кваліфікаційної роботи \_\_\_\_\_ Белла Голуб  
(підпис) (ім'я та прізвище)

## РЕФЕРАТ

Робота присвячена розробці системи для оцінювання якості навчального процесу у закладах вищої освіти.

**Об'єкт дослідження.** Оцінювання якості навчального процесу у закладах вищої освіти.

**Предмет дослідження.** Аналітична система для моніторингу та аналізу ключових показників, що характеризують якість навчання.

**Використані методи.** Використовувалися технології сховищ даних для консолідації джерел, де OLAP – для багатовимірного аналізу показників, а Data Mining – для виявлення закономірностей і побудови моделей прогнозування.

**Мета роботи.** В ході виконання досліджень обґрунтувати та реалізувати застосування OLAP і Data Mining для аналізу зібраних даних і отримання практично корисних знань, що підтримують прийняття рішень.

**Наукова складова.** Запропоновано узгоджений підхід до інтеграції OLAP–аналізу з процедурами Data Mining у єдиний аналітичний цикл, що дає змогу знаходити важливі патерни та одразу перевіряти їх на різних групах і періодах.

**Рекомендації щодо впровадження результатів.** Отримані результати можуть бути використані для раннього попередження ризику неуспішності, персоналізації підтримки студентів, коригування програм й навантаження, планування розкладу з урахуванням прогнозованих піків складності.

**Прикладна значущість роботи.** Значимість роботи полягає в підвищенні ознайомленості викладачів з якістю навчального процесу, швидше реагування на проблемні тенденції, забезпечити прозоріше й обґрунтованіше оцінювати.

Кількість сторінок – 65

Кількість ілюстрацій – 35

Кількість таблиць – 1

Кількість додатків – 2

Кількість джерел – 27

## ABSTRACT

The work is devoted to the development of a system for evaluating the quality of the educational process in higher education institutions.

**Object of research.** Evaluation of the quality of the learning process in higher education institutions.

**Subject of research.** An analytical system for monitoring and analyzing key performance indicators that characterize learning quality.

**Methods used.** Data–warehouse technologies for source consolidation; OLAP for multidimensional analysis of indicators; Data Mining for discovering patterns and building predictive and segmentation models.

**Purpose of the work.** To justify and implement the application of OLAP and Data Mining for analyzing collected data and obtaining practically useful knowledge that supports decision–making.

**Scientific novelty.** A coordinated approach to integrating OLAP analysis with Data Mining procedures into a single analytical cycle is proposed, which makes it possible to find important patterns and immediately test them on different groups and periods.

**Recommendations for implementing the results.** Use the findings for early warning of academic risk, personalization of student support, adjustment of curricula and teaching load, and schedule planning that accounts for predicted peaks of difficulty.

**Applied significance of the work.** Increases instructors’ awareness of learning quality, enables faster response to adverse trends, improves resource allocation, and makes assessment more transparent and evidence–based.

Number of pages – 65

Number of illustrations – 35

Number of tables – 1

Number of appendices – 2

Number of sources – 27

## ЗМІСТ

Оглавление	4
<b>ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ</b> .....	<b>4</b>
<b>ВСТУП</b> .....	<b>5</b>
<b>1 АНАЛІЗ ПРОЦЕСІВ ОЦІНЮВАННЯ ЯКОСТІ НАВЧАЛЬНОГО ПРОЦЕСУ ЯК ПРЕДМЕТ ДОСЛІДЖЕННЯ</b> .....	<b>7</b>
1.1 Контекст і межі предметної області .....	7
1.2 Огляд наявних рішень і аналогів .....	9
1.3 Постановка завдання .....	12
<b>2 ПРОЕКТУВАННЯ СИСТЕМИ ОЦІНЮВАННЯ ЯКОСТІ НАВЧАЛЬНОГО ПРОЦЕСУ</b> .....	<b>14</b>
2.1 Загальні положення моделювання .....	14
2.2 Діаграма прецедентів .....	16
2.3 Діаграма послідовності .....	19
<b>3 РОЗРОБКА СИСТЕМИ</b> .....	<b>25</b>
3.1 Опис основних понять .....	25
3.2 Архітектура системи .....	27
3.3 Опис джерела даних .....	30
3.4 Опис сховища даних.....	35
3.5 Побудова розгорнутого куба в середовищі проекту служби SSAS ..	37
3.6 Наповнення кубу даними .....	43
4.1 OLAP: ключові показники ефективності та звіти.....	47
4.2 Методи Data Mining для інтелектуального аналізу .....	50
<b>ВИСНОВКИ</b> .....	<b>59</b>
<b>ДЖЕРЕЛА</b> .....	<b>62</b>

## ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ

ОБД – оперативна база даних.

БД – база даних.

СД – сховище даних.

АС – аналітична система.

СУБД – система управління базами даних.

SSAS – SQL Server Analysis Services.

SSIS – SQL Server Integration Services.

SSRS – SQL Server Reporting Services.

OLAP – On-Line Analytical Processing.

DSV – Data Source View.

DF – Data Flow.

SSMS – SQL Server Management Studio.

PBI – Power BI.

DM – Data Mining (Інтелектуальний аналіз даних).

KPI – Key Performance Indicator (ключовий показник ефективності).

PCA – Principal Component Analysis (метод головних компонент).

1R – алгоритм 1–Rule (однопризнаковий класифікатор).

NB – Naive Bayes (наївний байєсівський класифікатор).

Apriori – алгоритм пошуку асоціативних правил.

K–Means – алгоритм кластеризації методом k–середніх.

MLXTend – бібліотека Python для Data Mining та машинного навчання.

Scikit–learn – бібліотека Python для реалізації алгоритмів машинного навчання.

LMS (Learning Management System) – система керування навчанням.

## ВСТУП

**Актуальність.** Підвищення якості навчального процесу на сьогоднішній день є, або ж має бути пріоритетом закладів вищої освіти, однак традиційні інструменти оцінювання (середні бали, разові зрізи успішності) дають фрагментарну картину та запізнюються з сигналами ризику. Цифровізація освітнього середовища: журнали успішності, тести, опитування, дані відвідуваності та активності формують великі, різномірні й динамічні масиви даних, що потребують не лише описового, а й причинно–прогностичного аналізу. Поєднання багатовимірної аналітики (OLAP) з методами інтелектуального аналізу даних (Data Mining) дає змогу переходити від звітності що робиться по наявним вже фактам до своєчасного виявлення закономірностей і прогнозів, персоналізованих інтервенцій для студентів, коригування змісту курсів та планування підтримки у найбільш напружені періоди.

**Об’єктом** дослідження є процес оцінювання якості навчального процесу у закладах вищої освіти.

**Предметом** дослідження виступає аналітична система для моніторингу та аналізу ключових показників, що характеризують якість навчання.

**Метою** в даній роботі є підвищення ефективності освітньої діяльності шляхом обґрунтування та практичної реалізації застосування OLAP і Data Mining для аналізу накопичених даних і отримання знань, що підтримують управлінські рішення для покращення якості освітнього процесу.

**Методи та засоби** що були використані у ході виконання магістерської роботи: Сховище даних і ETL для консолідації джерел; OLAP–аналіз (drill–down/roll–up, slice/dice) для багатовимірних зрізів та KPI; Data Mining для класифікації ризиків, кластеризації підгруп та виявлення асоціативних правил; інструментарій SQL/SSAS/Power BI та Python (pandas, scikit–learn) для побудови моделей і візуалізацій. Джерело даних у цьому проєкті – це

синтетично згенерований набір даних, що був створений на основі попередньо проведеного аналізу, ітеративного узгодження вимог та моделювання реалістичних сценаріїв.

**Наукова складова** полягає у тому, що впроваджується узгоджений підхід до інтеграції OLAP-аналізу з процедурами Data Mining у єдиний аналітичний цикл, що дає змогу швидко знаходити важливі патерни та одразу перевіряти їх на різних групах і періодах.

### **Апробація**

1. Б. Б. Кочубей, Б. Л. Голуб: АНАЛІТИЧНА СИСТЕМА ОЦІНЮВАННЯ ЯКОСТІ НАВЧАЛЬНОГО ПРОЦЕСУ. Програма конференції VII Всеукраїнської науково-практичної інтернет конференції студентів і аспірантів "Теоретичні та прикладні аспекти розробки комп'ютерних систем 2025". 24 квітня 2025 року, НУБіП України, Київ – С. 60.
2. Б. Б. Кочубей, Б. Л. Голуб: АНАЛІТИЧНІ OLAP-ТЕХНОЛОГІЇ У СИСТЕМАХ ОЦІНЮВАННЯ ЯКОСТІ НАВЧАЛЬНОГО ПРОЦЕСУ Збірник матеріалів II Міжнародно науково-практичної конференції «АКТУАЛЬНІ ПИТАННЯ РОЗВИТКУ НАУКИ ТА ТЕХНІКИ В УМОВАХ ГЛОБАЛІЗАЦІЇ», 14.05.2025. НУБіП України, ВСП «Боярський фаховий коледж НУБіП України», Боярка – С. 151–153.
3. XVI Міжнародна науково-практична конференція молодих вчених "Інформаційні технології: економіка, техніка, освіта" (м. Київ, 2025р.)

**Структура роботи** містить в собі 4 розділи. У першому розділі сформульовано мету й завдання, наведено огляд актуальних підходів та окреслено предметну область. Другий розділ присвячено моделюванню системи. У третьому розділі викладено методичну основу. Четвертий розділ містить аналіз отриманих результатів і підсумкові висновки.

Робота містить 65 сторінок, 35 малюнків та 27 джерел.

# 1 АНАЛІЗ ПРОЦЕСІВ ОЦІНЮВАННЯ ЯКОСТІ НАВЧАЛЬНОГО ПРОЦЕСУ ЯК ПРЕДМЕТ ДОСЛІДЖЕННЯ

## 1.1 Контекст і межі предметної області

Предметна область дослідження – якість навчального процесу у закладах вищої освіти. Сучасне освітнє середовище продукує великі, різномірні та часово залежні дані: оцінювання і повторні спроби, активність у LMS, результати тестувань, відвідуваність, опитування задоволеності, а також довідники курсів, груп, викладачів і навчальних періодів. Традиційні підходи – агреговані звіти та разові зрізи – обмежено відтворюють причинно–наслідкові зв’язки та запізно сигналізують про ризики.

Оцінювання якості у закладах вищої освіти стандартно організовується як повторюваний цикл удосконалення. Тобто планування цілей, потім збирання відомостей, аналітичне опрацювання, ухвалення та реалізація заходів, і перевірка досягнутих змін. На початку визначаються цілі (результати навчання, задоволеність, доступність підтримки, навчальне навантаження, тощо), погоджуються показники, періодичність вимірювань і відповідальні підрозділи (служба внутрішнього забезпечення якості, деканати/кафедри, інформаційні служби, етичні комісії). [1]

Джерела даних поєднують кількісні та якісні свідчення. До кількісних належать результати поточного й підсумкового контролю, частки складання та перескладання, відвідуваність, своєчасність завершення курсів і програм, показники вибіркового дисциплін, а також сліди навчальної активності у електронних середовищах. До якісних – анонімні опитування здобувачів і випускників, відгуки викладачів, зовнішні рецензії, спостереження за проведенням занять, портфоліо курсів та відкриті коментарі. Висновки

роблять шляхом зіставлення кількох джерел, уникаючи інтерпретацій за одним сигналом.

Агрегування здійснюють на узгоджених рівнях в такому порядку: курс – викладач – освітня програма – факультет – університет. Щоб порівняння були коректними, попередньо узгоджують визначення показників (що саме вважається “відвідуваністю”, як обчислюється “успішність”, межі семестру), установлюють порогові значення та правила тлумачення. Внутрішні результати зіставляють зі зовнішніми вимогами й критеріями акредитації. [2, 32с.].

Аналітичне опрацювання передбачає нормування та врахування контексту: вирівнювання за обсягом вибірок, складністю дисциплін, типами контролю, складом студентських груп. Окремо позначаються суттєві зміни (оновлення програм, зміна викладача, перехід на дистанційні формати), щоб уникнути хибних порівнянь до/після. Установлюють мінімальні обсяги вибірок для висновків, використовують довірчі інтервали та визначені правила роботи з аномальними значеннями.

Управлінські рішення оформлюються планами поліпшення: коригуванням силабусів і методик оцінювання, перерозподілом навантаження, наставницькими та підтримувальними заходами для окремих груп, підвищенням кваліфікації викладачів, а також організаційними змінами (розклад, формати занять). Для кожного заходу фіксують відповідальну особу, строк виконання, очікуваний ефект і спосіб перевірки. Наступний цикл оцінювання перевіряє, чи досягнуто запланованих змін.

Типові ризики такого процесу поділяються на методичні, організаційні та етично–правові. До методичних належать нерепрезентативні відповіді в опитуваннях, упередженість оцінок (поблажливість/строгість, “ефект привабливості курсу”), різночитання термінів, непорівнюваність груп. До організаційних – “втома від опитувань”, фрагментованість даних між системами, затримки в отриманні відомостей, конфлікти інтересів. Етичні та правові обмеження стосуються захисту персональних даних, мінімізації

доступу, прозорості мети збору й допустимого рівня деталізації. У зрілих практиках ці ризики ведуться у реєстрі з визначеними заходами пом'якшення (підвищення частки відповідей, стандарти тлумачення, анонімізація, незалежний експертний перегляд).[3]

У підсумку оцінювання якості – це не разова перевірка курсів і викладачів, а керований цикл що складається з узгодження цілей і мови вимірювання, збирання різних свідчень, обговорення результатів із носіями контексту, впровадження змін і перевірка їхнього впливу. Тож ця логіка і визначає реальні межі предметної області та вимоги до будь-якої аналітичної підтримки, що працює з освітніми даними.

## **1.2 Огляд наявних рішень і аналогів**

Оцінювання якості навчального процесу в сучасних закладах вищої освіти спирається на нормативні рамки зовнішнього та внутрішнього забезпечення якості, а також на стандартизовані опитувальні та доказові підходи до вимірювання результатів навчання. У європейському просторі вищої освіти орієнтиром виступають узгоджені стандарти та настанови щодо внутрішніх процедур якості (політики, формулювання результатів навчання, достовірність оцінювання, залучення студентів, регулярна самооцінка), які задають контур того, що і як має вимірюватися у вишах та освітніх програмах. Паралельно широко застосовуються респондентські інструменти із багаторазовими зрізами – зокрема національні й міжнародні опитувальники студентської залученості, що дозволяють відстежувати сталі показники середовища та навчальної взаємодії у динаміці когорти, програми й закладу загалом.

У відкритій екосистемі платформ окремого розгляду потребує Open edX (модуль Aspects). Тут оцінювання базується на подієвих журналах навчальної активності, які збираються з платформи й узагальнюються у відокремленому

аналітичному середовищі. Реалізовано фіксований набір інформаційних панелей: зарахування та активні слухачі, відвідування розділів курсу, перегляд і програвання відео, темпи проходження, своєчасність здачі завдань. Дані обробляються регулярними пакетами, події нормалізуються у спеціалізованій базі, а оглядові та діагностичні індикатори подаються у вигляді періодичних зрізів із можливістю деталізації до курсу і тижня навчання. Такий підхід демонструє методологію “слідів навчання” де виконується цілеспрямоване збирання цифрових слідів і перетворення їх на узгоджені показники участі, навантаження та своєчасності.[4]

Рішення Blackboard “Analytics for Learn” реалізує інституційний контур оцінювання якості через окремий аналітичний шар. Дані з транзакційної системи навчання та з інформаційних систем закладу витягуються у власне сховище й узгоджуються у структурі, зручній для порівнянь між курсами, підрозділами й семестрами. Далі формуються стандартні панелі з показниками активності у курсі, залученості, статусів виконання завдань, розподілів оцінок, наповненості й використання курсових елементів. Важливий момент – технологічне відокремлення аналітики від поточної експлуатації платформи навчання, що дозволяє підтримувати сталу періодичність оновлення індикаторів, прозорі визначення показників і порівнюваність між підрозділами.[5]

Платформа Canvas забезпечує модуль “New Analytics”, який орієнтовано на оперативні потреби викладача й керівника програми. Основу становлять тижневі й курсові підсумки: перегляди сторінок і участь у подіях курсу, розподіли оцінок, своєчасність здачі робіт, порівняння активності та успішності у часі. Передбачено фільтри для груп, часові вікна, експорт підсумків у табличний формат і можливість адресного повідомлення студентів за виявленими критеріями, такими як низька активність чи прострочення. Методично це поєднання описових індикаторів подій що відбулись і простих діагностичних зрізів (де саме змінилася участь/якість), які придатні для регулярного педагогічного реагування.[6]

У Moodle механізм “Аналітика навчання” ґрунтується на формальному описі моделей через цілі (цільові події), індикатори (похідні показники з журналів активності й успішності), правила розбиття часу та навчальні алгоритми. Типові приклади – виявлення ризику відрахування або невиконання курсу, оцінка залученості за комбінацією дій у межах тижня чи модуля. Окремо підтримуються перевірка придатності моделей на локальних даних, порогові значення для сповіщень, а також інтерпретація прогнозів через переліки внеску індикаторів. Таким чином, базовий цикл оцінювання охоплює визначення освітньо значущої події, добір індикаторів, навчання й перевірку моделі, оповіщення викладача та фіксацію подальших дій.[7]

Поряд із платформами, у практиці забезпечення якості часто застосовуються опитувальники студентської залученості, що вимірюють сталі конструкти середовища й навчальної взаємодії (час на навчання, частоту взаємодій, характер завдань, відгук викладача). Регулярні звіти таких опитувань подають розрізи за курсами й програмами, працюють як температурні карти сильних і слабких місць та слугують відправною точкою для коригувальних дій у наступному навчальному циклі. Методологічно це стандартизована респондентська оцінка з багаторазовими хвилями спостережень і порівняннями між роками та підрозділами.

Як приклад науково обґрунтованого процесу оцінювання, що застосовується у програмах і закладах, розглядаємо цикл “доказове вдосконалення навчання”: явне формулювання результатів навчання, добір індикаторів і джерел доказів (роботи студентів, рубрики, підсумкові роботи, зовнішні іспити), розроблення процедур надійного оцінювання, зіставлення фактичних досягнень із запланованими орієнтирами, колегіальний розбір причин відхилень і фіксація конкретних дій (зміни у змісті, навантаженні, методах викладання). Ключовим є “замикання циклу”, де наступна ітерація навчання перевіряє, чи дали ці дії покращення показників. Підхід детально описано в сучасній літературі з внутрішнього забезпечення якості і

використовується як каркас для програмних самооцінок і акредитаційних процедур.[8]

Загалом розглянуті аналоги демонструють взаємодоповнювані траєкторії, де опитувальники – для вимірювання залученості та досвіду навчання у стабільних зрізах, платформні засоби – для регулярних описових і діагностичних показників діяльності у курсах, інституційні аналітичні контури – для зіставних порівнянь між програмами та періодами, а науково вивірені підходи – для доказового планування змін і перевірки їхнього ефекту в наступних ітераціях. У сукупності це задає методологічне поле, у межах якого доцільно проектувати та оцінювати власні рішення.

### **1.3 Постановка завдання**

Необхідно спроектувати й реалізувати аналітичну систему оцінювання якості навчального процесу на базі сховища даних із підтримкою OLAP та Data Mining. Система має дати відповіді на чотири ключові питання, визначені у виданому завданні:

- забезпечити оцінювання динаміки та якості навчального процесу засобами OLAP.
- розробити й застосувати систему ключових показників ефективності (KPI) для моніторингу курсів і рівня задоволеності студентів.
- дослідити можливості використання методів Data Mining для освітніх даних (класифікація, асоціативні правила, кластеризація).
- виявити типові освітні траєкторії та групи ризику за результатами кластерного аналізу.

Для реалізації цього завдання треба підготувати дані навчального процесу з урахуванням відомостей про студентів/слухачів, курси, викладачів, академічні групи, результати тестувань і відповіді на запитання тестів, а також

показники відвідуваності та результати опитувань щодо задоволеності. На основі консолідованих даних потрібно виконати багатовимірний аналіз у розрізах періодів, курсів, груп і викладачів, а також погодити набір КРІ, які відображають успішність, відвідуваність і показники задоволеності. Далі передбачається застосування методів Data Mining:

- побудова інтерпретованих моделей класифікації для раннього виявлення ризику неуспішності.
- пошук асоціативних правил між патернами навчальної активності та підсумками.
- проведення кластеризації для виділення стійких профілів навчальної поведінки й освітніх траєкторій.

Отримані результати інтегруються у звітні матеріали та OLAP–зрізи для перевірки висновків на різних групах і часових відрізках та подальшого використання у прийнятті управлінських рішень.

## 2 ПРОЕКТУВАННЯ СИСТЕМИ ОЦІНЮВАННЯ ЯКОСТІ НАВЧАЛЬНОГО ПРОЦЕСУ

### 2.1 Загальні положення моделювання

Моделювання системи – це спрощення складної структури, коли ми описуємо об’єкт через узгоджені абстракції, щоб краще його зрозуміти і керувати ним. Такий опис фіксує призначення і межі системи, показує її внутрішню будову та очікувану поведінку, а ще дає спільну мову всім, хто над нею працює. Ідея проста: поки все лишається в голові або у вигляді розрізнених тез, легко щось пропустити або потрактувати по–своєму, модель змушує домовитися про терміни, уточнити припущення і перевести бажане у перевірювані рішення.

Щоб керувати складністю, зручно тримати в голові три рівні абстракції. На концептуальному рівні ми говоримо про головні сутності й межі без прив’язки до технологій. Логічний рівень уже прояснює, які елементи взаємодіють між собою, які між ними інтерфейси, які правила і винятки. Фізичний рівень – це про розміщення артефактів у конкретному середовищі: на яких вузлах що працює, як вони з’єднані, які обмеження накладає інфраструктура. Важливо, щоб між цими рівнями була прозора відповідність: якщо змінюється термін або правило на концептуальному рівні, це має відбитись у логічній структурі й, у разі потреби, у фізичному розгортанні.[9]

Корисно розділяти два роди подань – структурні та поведінкові. Структурні відповідають на запитання “з чого складається система”: які у нас поняття, які їхні властивості, як вони пов’язані, що є частиною чого, що узагальнюється, а що – конкретизується; сюди ж відноситься топологія розміщення компонентів. Поведінкові подання пояснюють що система робить: сценарії використання, послідовності взаємодій, умови старту і завершення дій, альтернативні та помилкові гілки. Окремо кожен погляд неповний: поведінка без структури швидко повисає в повітрі, а структура без поведінки

не показує, навіщо все це потрібно. Тому в роботі вони постійно зіставляються між собою.

Далі – про самі підходи. Функціональний підхід дивиться на систему ззовні. У центрі – актори (користувачі), їхні цілі й очікувані результати. Ми описуємо послідовності кроків, які приводять до цих результатів, фіксуємо передумови і післяумови, продумуємо, що робити з альтернативами та помилками. Такий опис допомагає чесно визначити межі відповідальності: що справді входить у систему, а що – ні; що вона гарантує, а де потрібна співпраця з іншими сервісами. Важливий нюанс – на цьому етапі ми свідомо не «провалюємося» у внутрішній устрій, щоб не зацементувати рішення передчасно.[10]

Об'єктно–орієнтований підхід дивиться зсередини. Тут ми формалізуємо сталі поняття предметної області: які сутності існують, які мають атрибути, як пов'язані між собою (через асоціації, композиції, узагальнення), які кардинальності допустимі. Саме тут з'являються інваріанти – правила цілісності, що не повинні порушуватися незалежно від сценарію. Такий рівень дає стабільну термінологію і стає опорою для подальшого поділу на компоненти, визначення інтерфейсів, схем даних і політик доступу. Практично це означає, що кожен крок зі сценарію має хазяїна у структурі, а кожне правило зі структури має сенс у якомусь сценарії.[11]

Щоб моделі були не для краси, а працювали на проєкт, важлива базова дисципліна. Домовляємося про терміни і послідовно їх уживаємо, не змішуємо рівні деталізації, не ховаємо суттєві припущення й обмеження «між рядків». Корисно тримати в полі зору нефункціональні характеристики — продуктивність, надійність, безпеку, супровідність, масштабованість і сумісність. Вони впливають на форму моделі не менше, ніж власне функціональність: інколи саме вимога до затримок або ізоляції даних визначає, як розрізати компоненти і де проводити межі.

Для графічного подання ми спираємося на усталені нотації. UML версії 2.x – це, по суті, алфавіт і граматики для схематичного опису і структури, і

поведінки: він визначає типи елементів, допустимі зв'язки між ними і набір діаграм для різних поглядів на одну систему. Архітектурні матеріали зручно організовувати як набір узгоджених подань і точок зору: кожне подання висвітлює окремий інтерес (функціональний, інформаційний, інтеграційний, розгортання), але всі разом мають складати цілісну картину без суперечностей.

Нарешті, варто коротко зафіксувати інструментарій, на який спиратимемося в цьому розділі: діаграма прецедентів, діаграма послідовності, діаграма класів, діаграма розгортання. Вони покривають зовнішній і внутрішній погляди, часову логіку взаємодій і фізичний контур виконання – рівно настільки, щоб перейти від узгоджених понять і сценаріїв до предметних рішень без втрати змісту.

## **2.2 Діаграма прецедентів**

Діаграма прецедентів – це опис системи ззовні, мовою цілей і очікуваних результатів для зовнішніх учасників. Вона показує межу системи, акторів по той бік цієї межі та прецеденти всередині, з якими вони взаємодіють. Такий тип діаграми не розкриває внутрішній устрій, зате чітко відповідає на запитання хто і навіщо використовує систему, допомагає погодити очікування та уникнути змішування ролей.[12]

Діаграму прецедентів показано на рис. 2.1.

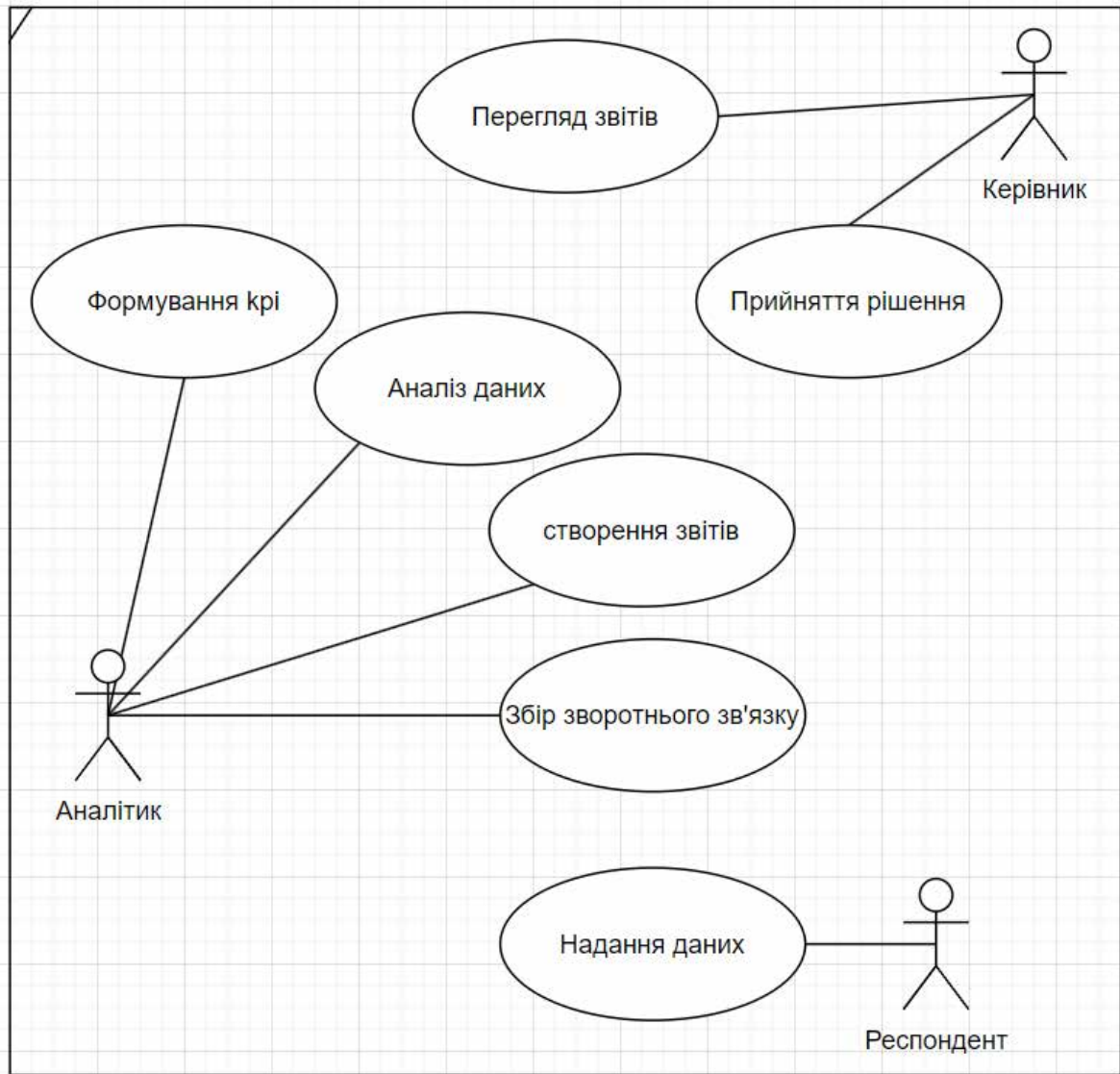


Рис. 2.1 Діаграма прецедентів до розроблюваної АС

У межах цієї системи актори та їхні обов'язки наведені в таблиці 1. Вони взаємодіють із мінімальним, але достатнім набором прецедентів, який відтворює повний цикл роботи з даними: від появи первинної інформації до управлінського рішення. На діаграмі показано такі прецеденти: «Надання даних», «Збір зворотного зв'язку», «Аналіз даних», «Формування КРІ», «Створення звітів», «Перегляд звітів», «Прийняття рішень». Респондент пов'язаний із «Наданням даних», Аналітик – зі збіркою та обробленням («Збір зворотного зв'язку», «Аналіз даних», «Формування КРІ», «Створення звітів»), Керівник – із споживанням результатів («Перегляд звітів», «Прийняття рішень»).

Таблиця 1

### Актори, які діють у системі

Актор	Опис
Керівник	Використовує систему для перегляду звітів і ухвалення управлінських рішень. Основний споживач підсумкової аналітичної інформації.
Аналітик	Відповідає за збір, обробку даних та створення аналітичних звітів. Виконує основну роботу, пов'язану з аналізом і оцінкою результатів.
Респондент	Постачає дані, необхідні для аналізу якості навчального процесу, через заповнення анкет, тестів або інших інструментів оцінювання.

Спочатку дані потрапляють у систему через “Надання даних”, далі вони акумулюються та приводяться до узгодженого вигляду в межах “Збору зворотного зв’язку”. Після цього відбувається “Аналіз даних”, у результаті якого виділяються суттєві показники; на цій основі викристалізуються “Формування КРІ” та їхні пороги/періоди. Коли зміст стає достатньо зрозумілим і структурованим, Аналітик виконує “Створення звітів”, і вже ці підсумкові матеріали стають підставою для “Перегляду звітів” та “Прийняття рішень” керівництвом. Така послідовність не є технічною залежністю, а радше відображенням реального процесу: спочатку зібрати й впорядкувати, потім проаналізувати й узагальнити, і лише далі – ухвалювати дії.

Передумови й післяумови для груп прецедентів формулюються стисло та спостережно. Для Респондента передумова – чинний період опитування та, за потреби, автентифікація; післяумова – збережений валідний запис. Для Аналітика передумови – доступність вхідних даних і права на оброблення; післяумови – очищені набори, розраховані показники та сформовані звіти. Для Керівника передумова – наявність актуальних підсумкових матеріалів;

післяумова – зафіксовані перегляди й рішення (наприклад, затвердження дій або запит на корекцію).

Межа системи проходить по контуру, де всередині залишаються операції збору, узгодження та аналізу даних, розрахунок KPI і формування звітів, а зовні – ролі користувачів і допоміжні сервіси на кшталт LMS чи засобів автентифікації. Такий поділ не тягне до діаграми внутрішні технічні деталі, але чітко фіксує відповідальність: хто подає дані, хто перетворює їх на знання, і хто на основі цих знань ухвалює рішення.

## 2.3 Діаграма послідовності

Діаграма послідовності описує сценарій роботи системи у часовому вимірі: хто з ким обмінюється повідомленнями, в якій послідовності це відбувається і який результат отримує користувач. Вона зручна там, де важливо побачити не лише що робить система, а й коли саме це стається та за яких умов сценарій переходить в альтернативну гілку. У нашому випадку така діаграма слугує мостом між прецедентами й технічними контрактами взаємодії, бо чітко розкладає кроки від ініціювання дії до отримання підтвердженого результату.[13]

У побудованих сценаріях беруть участь три об'єкти:

- Аналітик ініціює процеси, обирає метрики та перевіряє підсумкові матеріали(саме він запускає дії, що згодом призводять до формування результатів).
- Оперативне джерело даних координує увесь процес: надсилає форми респондентам, приймає та перевіряє відповіді, а також готує підсумкові дані і звіти.
- Респондент надає відповіді у формі опитування(його участь епізодична, але саме з неї починається цикл оброблення даних).

Сценарій “проведення оцінювання” у вигляді діаграми послідовності наведено на рис. 2.2.

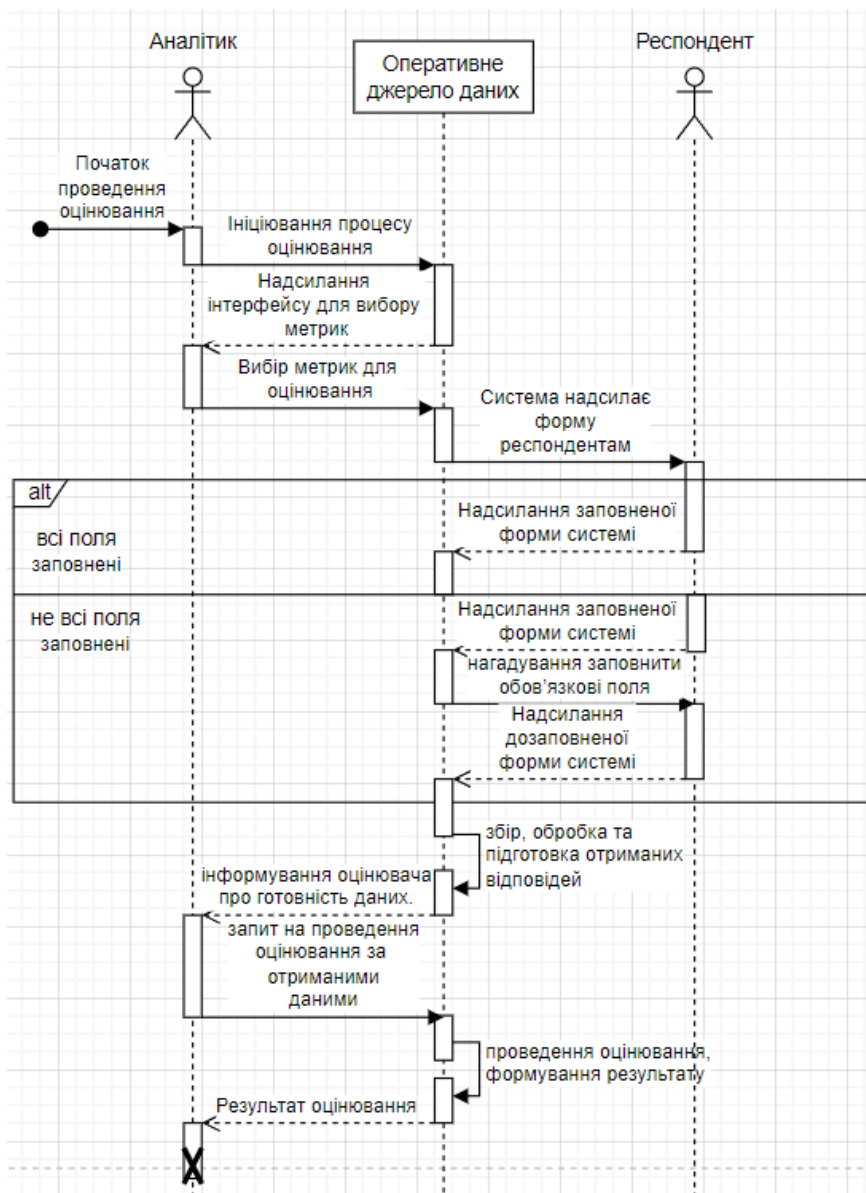


Рис. 2.2 Діаграма послідовності: проведення оцінювання.

Сценарій “проведення оцінювання” розгортається так: Аналітик запускає процес, отримує від оперативного джерела даних інтерфейс для вибору метрик і визначає перелік показників, за якими проводитиметься оцінювання. Оперативне джерело даних розсилає форму респондентам. І коли відповіді повертаються, то спочатку виконується перевірка обов’язкових полів, і якщо все заповнено – форма приймається й фіксується, якщо ні – джерело надсилає нагадування і дозволяє дозаповнити пропущені значення. Після успішного прийому відповіді збираються, приводяться до узгодженого

вигляду та готуються до аналізу. Аналітик отримує сповіщення про готовність даних, а процес завершується сформованим результатом оцінювання.

Візуалізацію послідовності подій для сценарію “збір зворотного зв’язку” подано на рис. 2.3.

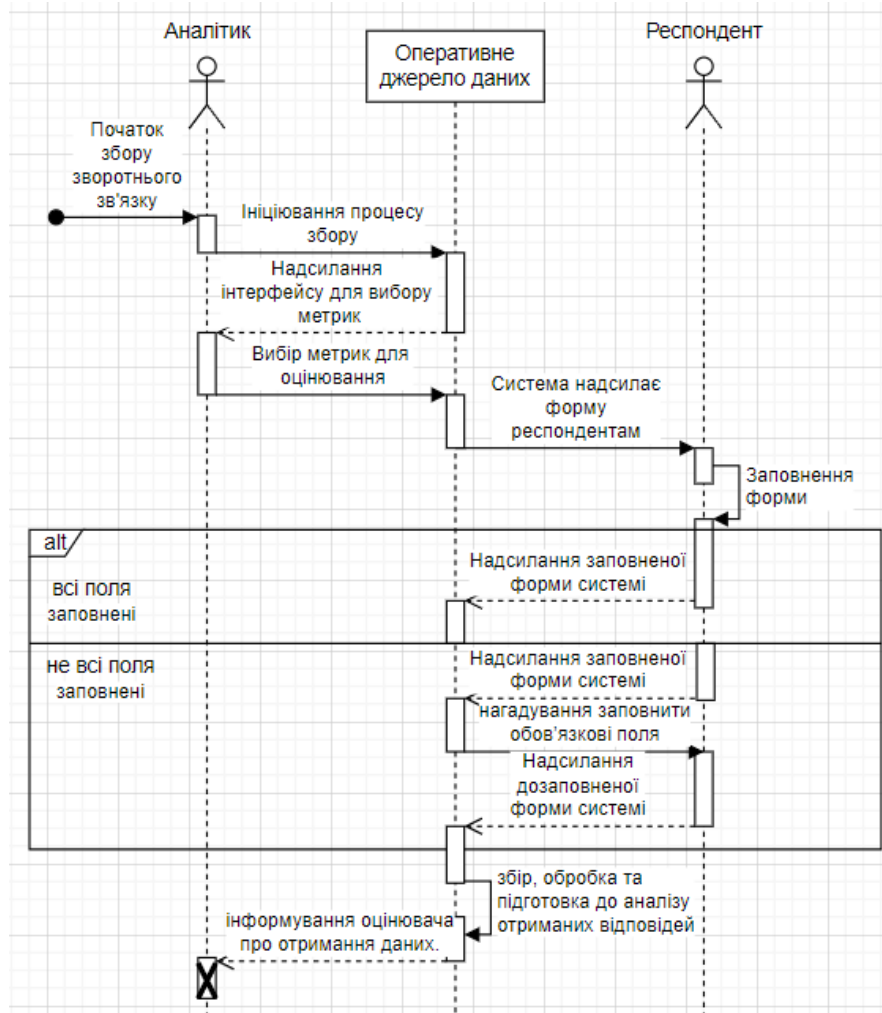


Рис. 2.3 Діаграма послідовності: збір зворотного зв’язку.

Сценарій збір зворотного зв’язку за структурою подібний, але акцент зміщено на приймання коментарів та якісних відповідей. Аналітик ініціює збір і через інтерфейс обирає метрики, за якими потрібні відгуки. Оперативне джерело даних надсилає форму респондентам, після чого знову спрацьовує перевірка обов’язкових полів із можливістю дозаповнення у випадку пропусків. Після цього відповіді збираються та попередньо опрацьовуються, щоб їх можна було використати в подальшому аналізі, і аналітик отримує повідомлення про готовність пакета даних.

Діаграму для сценарію “формування звіту” подано на рис. 2.4

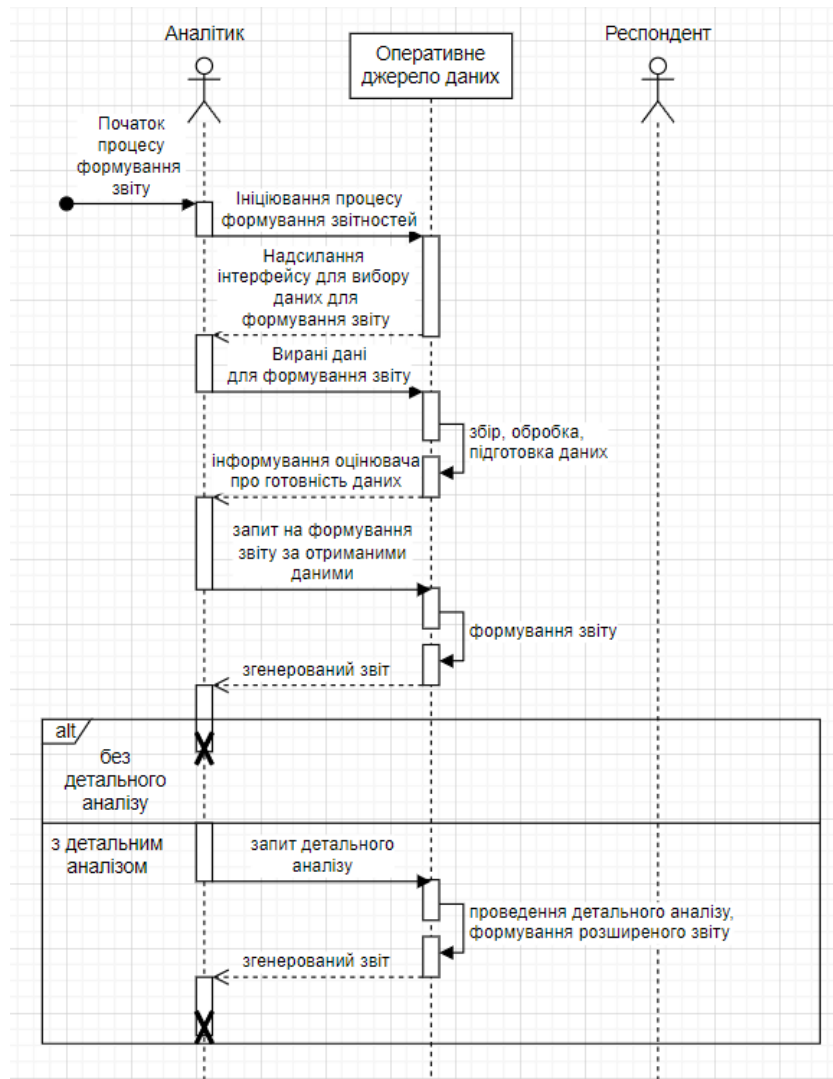


Рис. 2.4 Діаграма послідовності: формування звіту.

Сценарій формування звіту починається з ініціювання процесу аналітиком та отримання від оперативного джерела даних інтерфейсу для вибору даних. Аналітик визначає, які зрізи і періоди включати, після чого система збирає й готує обрані дані та генерує базовий звіт. Далі можливі два шляхи: якщо достатньо оглядового рівня, на цьому все завершується; якщо потрібна глибина, аналітик відправляє запит на детальний аналіз, оперативне джерело даних виконує додаткову обробку і формує розширений звіт. У діаграмі це показано окремим альтернативним фрагментом.

## 2.4 Діаграма класів

Діаграма класів показує скелет системи – сталі поняття предметної області, їхні властивості та те, як вони пов’язані між собою. На відміну від діаграм послідовності, тут не відтворюється хронологія подій; фіксується структура, яка лишається сталою незалежно від сценарію. Саме цю діаграму зручно читати як словник термінів із вбудованими правилами взаємодії: хто з ким пов’язаний, хто за що відповідає, що може існувати самостійно, а що – лише як частина більшої сутності.[14]

Структурну модель (діаграму класів) наведено на рис. 2.5.

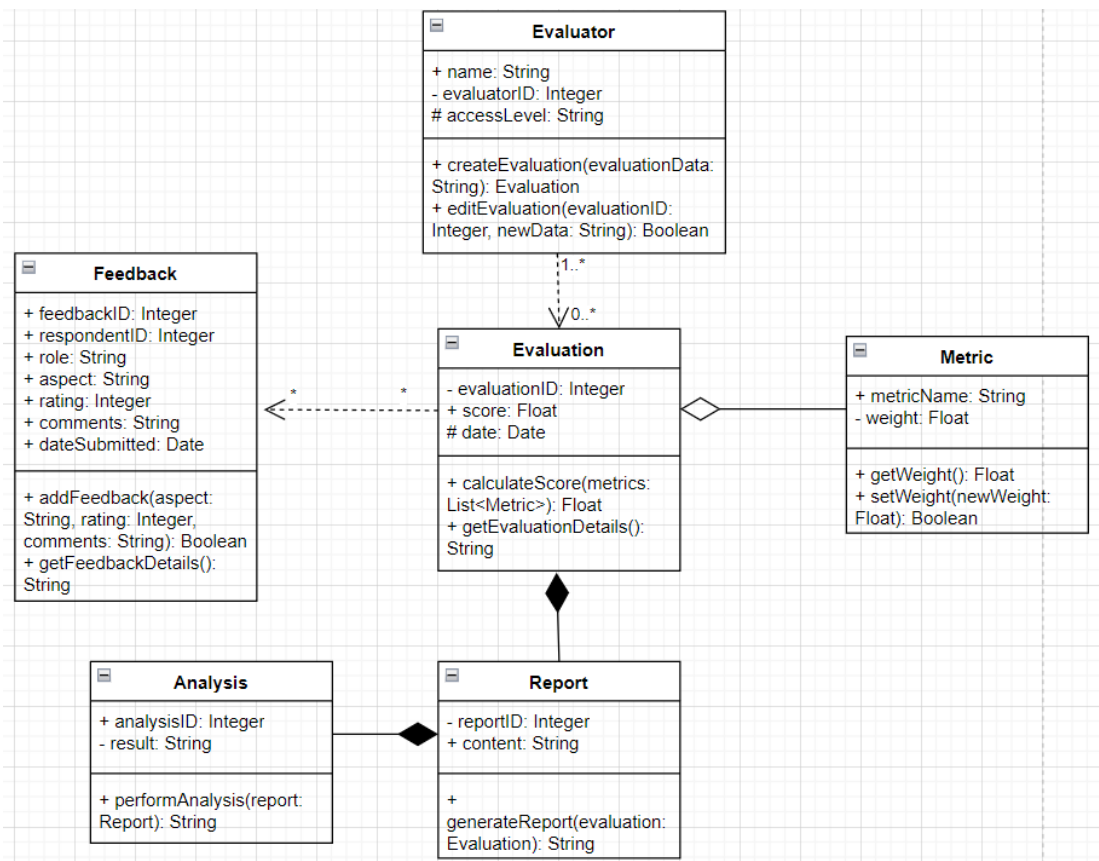


Рис. 2.5 Діаграма класів системи.

У нашій моделі головний клас – клас Evaluation (Оцінка). Він акумулює результат вимірювання якості: має ідентифікатор, підсумковий бал та дату формування; через методи підраховує оцінку на основі пов’язаних показників і повертає деталі у зручному вигляді. До цієї сутності підходять дві лінії: з одного боку – Evaluator (Оцінювач), з іншого – Metric (Метрика). Перший відображає користувача з правами створювати й коригувати оцінки. Атрибути

ідентифікації й рівня доступу пояснюють, хто саме ініціює зміни, а операції створення/редагування фіксують керівну роль цієї особи. Зв'язок із Evaluation читається просто: один оцінювач може мати багато створених ним оцінок, а у кожній оцінці є свій автор.

Metric описує критерій, за яким рахується оцінка, і його вагу. На діаграмі між Evaluation і Metric проставлено агрегування (порожній ромб біля Evaluation): оцінка оперує набором метрик, але самі метрики не «належать» їй і можуть повторно використовуватися в інших оцінюваннях. У практичному сенсі це означає, що для розрахунку беруться значення метрик, актуальні для цієї оцінки, а внесок кожної визначається її вагою.

Поруч зі вхідною частиною моделі знаходиться Feedback (Зворотний зв'язок). Клас зберігає первинні відповіді респондентів: хто саме надавав оцінку (ідентифікатор і роль), який аспект коментував, яку поставив оцінку, які залишив коментарі і коли це відбулося. На схемі між Feedback і Evaluation показано пунктирний зв'язок залежності: оцінка використовує відгуки як вхідні дані під час формування підсумку, але не володіє цими записами безпосередньо.

Результат роботи з оцінкою формалізується у Report (Звіт). Він має ідентифікатор і вміст та створюється для певної оцінки через відповідну операцію. На діаграмі зв'язок Report–Evaluation показано композицією (заповнений ромб біля Report): звіт розглядається як частина більшого цілого – конкретної оцінки, і живе стільки, скільки живе її результат.

Ще один шар – Analysis (Аналіз). Він працює вже зі сформованим звітом: бере його на вхід, виконує поглиблене опрацювання і видає результат у вигляді, визначеному реалізацією. На схемі Analysis пов'язаний із Report також композицією, що підкреслює їхній спільний життєвий цикл у межах одного пакета матеріалів. У деяких впровадженнях аналіз можуть зберігати окремо, але в цій моделі зроблено акцент на тісному логічному зв'язку від звіту до аналітичного висновку.

## 3 РОЗРОБКА СИСТЕМИ

### 3.1 Опис основних понять

У процесі розроблення системи ми спиралися на узгоджений термінологічний апарат. Далі розглянемо робочі визначення чотирьох базових понять – сховище даних, OLAP, ETL та інтелектуальний аналіз даних (Data Mining).

Сховище даних – це така спеціалізована інформаційна база, куди інтегруються відомості з різних джерел в узгодженому вигляді, де зберігається історія змін і забезпечується виконання швидких аналітичних запитів. Сховище орієнтоване на читання та аналітику, в той час як операційні бази даних оптимізовані під часті транзакції. Типові логічні моделі – “зірка” (fact + dimensions) і “сніжинка” (нормалізовані виміри). Центральна таблиця фактів містить події/вимірювання та числові міри (те, що агрегуємо), а таблиці–виміри описують контекст – хто/що/коли/де. Важливі поняття: grain (найменша одиниця обліку факту), сурогатні ключі, політика історизації довідників (SCD), цілісність посилань і узгоджені довідники між системами.[15]

OLAP (Online Analytical Processing) – підхід до багатовимірного аналізу, у якому дані подано як куб: по осях – виміри, на перетинах – міри. Користувач взаємодіє з даними з використанням типових операцій: slice/dice (вибір підмножини), drill-down/roll-up (провалювання в деталі та підняття до підсумків), pivot (перестановка осей). За способом зберігання і обчислень розрізняють MOLAP (міри та агрегати в багатовимірних структурах), ROLAP (обчислення проводиться безпосередньо по реляційних таблицях SQL) і HOLAP (гібрид: деталь у реляційній БД, агрегати — у багатовимірних структурах). Вибір режиму – це завжди компроміс між швидкістю, гнучкістю моделювання та витратами на ресурси. У сучасних інструментах також уживаються табулярні моделі з колонковим зберіганням, які зберігають

багатовимірну семантику (виміри, ієрархії, міри) і добре працюють із великими обсягами даних.[16]

ETL (Extract–Transform–Load) описує керований шлях даних від джерел до сховища. На етапі Extract дані відбираються з систем–джерел, на Transform – очищуються, узгоджуються з еталонними довідниками, приводяться до потрібних типів/форматів, розв’язуються дублікати, формуються ключі, а на Load – завантажуються у факти й виміри з дотриманням цілісності та політики історизації (SCD). Важливі аспекти ETL–процесу: контроль якості (валідаційні правила, журнали помилок), інкрементальні завантаження (щоб не перетягувати все щоразу), і оркестрація – планування задач, залежності, сповіщення. Реалізувати ETL можна по–різному: візуальними засобами на кшталт SSIS, а також SQL–скриптами чи пайплайнами на Python. Підхід обирається залежно від вимог до продуктивності, підтримованості й середовища виконання. [17]

Data Mining (інтелектуальний аналіз даних) – це методи пошуку нетривіальних, нових і корисних закономірностей у даних. Прийнято виділяти прогностичні задачі (класифікація, регресія) та описові (кластеризація, асоціативні правила, виявлення аномалій). У першому випадку моделі навчають на розмічених прикладах і оцінюють за метриками якості прогнозу (наприклад, точність/recall/F1), у другому – шукають структуру або зв’язки без цільової ознаки (метрики на кшталт support/confidence/lift для правил чи Silhouette для кластерів). Ключові практичні питання – узагальнюваність (перевірка на відкладених даних або крос–валідація), стійкість до дисбалансу класів і інтерпретованість (наскільки легко пояснити отримані результати нефахівцю).[18]

## 3.2 Архітектура системи

Архітектура системи подає інфраструктурну картину системи: які вузли залучено (робочі станції, сервери, служби), які програмні артефакти на них розміщено та якими каналами вони обмінюються даними. Це стислий опис середовища виконання й розподілу обов'язків між його частинами: де відбувається введення та зберігання, де — аналітична обробка й звітність, які з'єднання забезпечують обмін. Така схема потрібна, щоб узгодити межі доступу, спланувати навантаження і переконатися, що операційні та аналітичні процеси ізольовані та можуть масштабуватися незалежно.[19]

Архітектуру системи наведено на рис. 3.1.

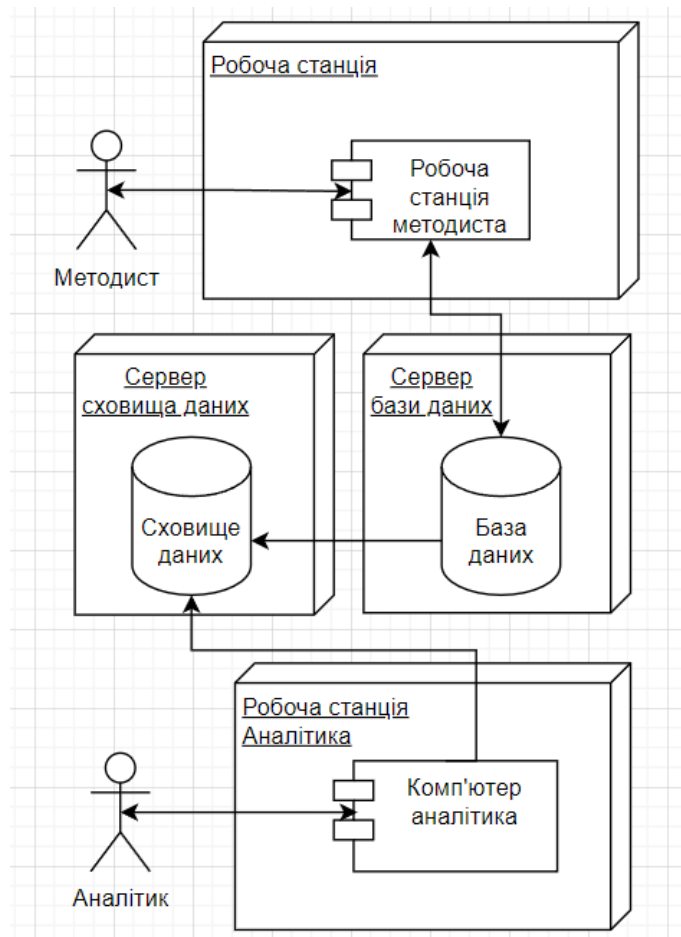


Рис. 3.1 Архітектура системи

Архітектурна конфігурація системи розподіляє функції між операційним і аналітичним контурами, фіксуючи, де саме виконуються процеси вилучення, трансформації та завантаження даних (ETL), побудова

багатовимірних подань, перегляди в режимі аналітичної обробки онлайн (OLAP) та дослідницькі процедури інтелектуального аналізу даних.

Робоча станція методиста зосереджує предметне налаштування, де визначаються показники та порогові значення, узгоджуються довідники, фіксуються правила агрегації й специфікація даних (опис структур і правил оброблення) для подальшого завантаження у сховище. Методист задає, які метрики потраплятимуть до куба та які виміри будуть зафіксовані (час, курс, група, метрика), як обробляються пропуски та хибні значення, а також коли і що саме має оновлюватися у звітності. Масові перетворення на цьому вузлі не виконуються – тобто станція лише конфігурує параметри, за якими працюють серверні процедури вилучення, перетворення і завантаження; деталі реалізації потоків та розкладів подано далі у розділі 3.

Сервер бази даних виступає оперативним джерелом подій і показників, бо на ньому відбувається узгоджене приймання транзакцій та їх фіксація. Щоб не впливати на продуктивність операційного контуру, застосовуються лише мінімальні підготовчі дії – перевірка обов’язкових полів, уніфікація часових позначок, контроль дублювання. З цього вузла ініціюється вилучення у межах ETL: формуються вибірки сирих записів для подальшого очищення, нормалізації й поєднання з довідниками в аналітичному контурі. Далі в розділі 3 показано, як ці дані за допомогою засобів Integration Services (SSIS) переносяться і перетворюються перед завантаженням у сховище.

Центром аналітичного контуру є сервер сховища даних EducationQualityDW. Тут відбувається трансформація та завантаження у зіркову модель із таблицею фактів QualityFact та узгодженими вимірами DateDim, CourseDim, GroupDim, MetricDim; на основі цієї моделі будується та актуалізується багатовимірний куб у службах аналізу (SSAS). Куб – упорядковане подання даних за вимірами з попередньо обчисленими підсумками для швидкого доступу. На рівні сховища реалізовано OLAP-обробку, пов’язану з практичними потребами звітності: підтримуються ключові показники ефективності (KPI), такі як, Quality і Satisfaction, з

відстеженням динаміки, а також керовані перегляди на кшталт “Курси/метрики” та “Семестровий звіт”. Технічно застосовуються класичні OLAP-операції: деталізація (drill-down), узагальнення (roll-up), зріз (slice) та комбінований зріз (dice). Це дає змогу переходити від підсумкового перегляду КРІ до переліку конкретних курсів у вибраному періоді, а далі – до груп, де відхилення найбільш виражені. Конкретні схеми, регламенти завантаження SSIS і параметри оброблення куба SSAS подано далі у розділі 3, тут же в свою чергу, зафіксовано, що саме сервер сховища є місцем виконання OLAP і зберігання багатовимірних подань.

Робоча станція аналітика поєднує навігацію за зрізами куба та процедури інтелектуального аналізу даних. Аналітик використовує багатовимірні подання для формування навчальних і тестових вибірок та застосовує методи, що доповнюють звітну аналітику. 1-Rule (1R) слугує базовою лінією: одноатрибутний класифікатор, який для кожного значення вибраної ознаки прив’язує найімовірніший клас, забезпечує прозору інтерпретацію та швидку оцінку роздільної здатності ознак. Наївний Байєс реалізує ймовірнісну класифікацію, бо комбінуються апіорні ймовірності класів з умовними ймовірностями значень ознак (за припущенням незалежності), модель повертає інтерпретовані ймовірнісні оцінки. У реалізації застосовується згладжування, щоб уникати нульових частот. Для виявлення сталих поєднань ознак застосовуються асоціативні правила, де дані представляються у вигляді індикаторів наявності ознак, далі виконується пошук частих наборів і побудова правил умова  $\>$  наслідок із оцінюванням підтримки (support) та довіри (confidence). Додаткові показники підсилення (lift) і conviction використовують для відсіювання випадкових або крихких зв’язків. Для сегментації без попередніх міток використовується кластеризація K-Means: ознаки нормують, категоріальні перетворюють на індикатори, центроїди ініціалізують (k-means++), після чого алгоритм по чергово призначає об’єкти найближчим центрам і оновлює їх до збіжності. Кількість кластерів добирають емпірично за індексом силуета (Silhouette Score), а для

інтерпретації багатовимірних даних використовують аналіз головних компонент (PCA). Типовий цикл роботи аналітика при цьому такий: у багатовимірному поданні обирається релевантний зріз, на його основі формується навчальна і тестова підвибірки, послідовно навчаються/оцінюються прості та складніші моделі (1–Rule, наївний Байєс, K–Means, правила асоціацій), результати перевіряються на узгодженість із агрегованими переглядами (OLAP, KPI) і за потреби уточнюються параметри або вихідні зрізи; підсумком стають інтерпретовані висновки та узгоджені звітні матеріали.

Отже у підсумку бачимо таку картину.

1. Методист задає семантику та регламенти оновлення, оперативний
2. сервер надає достовірне джерело
3. сервер сховища EducationQualityDW реалізує ETL–консолідацію і обслуговує SSAS–куб із KPI та OLAP–звітами
4. робоча станція аналітика поєднує ці подання з методами IR, наївного Байєса, асоціативних правил і K–Means (з оцінюванням індексом силуета та підтримкою PCA).

Технічні кроки побудови й наповнення куба за SSIS/SSAS, а також приклади звітів і метрик подано далі у розділі 3. Тут визначено, де саме в системі виконуються відповідні процеси і який зміст мають застосовані аналітичні методи.

### **3.3 Опис джерела даних**

Джерелом для аналізу є синтетично згенерований набір освітніх даних, сформований спеціально під завдання оцінювання якості навчального процесу. Від відкритих наборів відмовилися через конфіденційність і брак потрібної деталізації. Замість цього структура й обсяги даних були узгоджені ітеративно через серії уточнювальних запитань, щоб зафіксувати предметну область, масштаби, метрики та сценарії подальшого аналізу. Орієнтовний

обсяг – близько 15 тис. записів подій оцінювання з розбивкою за роками, семестрами та місяцями.

Перший блок запитань визначав рамки та правила: перелік спеціальностей і кафедр, типи курсів (базові, вибіркові, практичні), обрану шкалу оцінювання (100–бальну), потребу у додаткових атрибутах для студентів і викладачів, наявність спеціальних показників у відгуках та приклади типових аналітичних питань (порівняння між періодами, вплив зовнішніх подій). Також погоджено урахування пропусків/нездач, семестрову розбивку та можливість змін складу студентів і викладачів упродовж років.

Узгоджувальний перелік запитань, що визначив рамки предметної області та вимоги до синтетичного датасету, подано на рис. 3.2–3.4.

- 2 Які саме спеціальності є на факультеті Інформаційних технологій у вашому ВНЗ? (Будь ласка, наведіть перелік хоча б 3-5 спеціальностей.)
- 3 Які типи курсів (наприклад, базові, вибіркові, практичні) мають бути присутні у датасеті? Чи є приклади назв курсів?
- 4 Яка система оцінювання використовується для курсів (наприклад, 100-бальна, 5-бальна, ECTS тощо)?
- 5 Чи потрібно враховувати додаткові атрибути для студентів та викладачів (наприклад, стать, вік, стаж, науковий ступінь)?
- 6 Чи є якісь специфічні показники або метрики, які мають бути у відгуках/оцінках курсів (наприклад, окремо оцінка змісту, організації, актуальності, або лише загальна оцінка)?
- 7 Чи є у вас приклади типових аналітичних запитів, які ви плануєте виконувати над цими даними (наприклад, порівняння оцінок до/після пандемії, вплив війни на якість освіти тощо)?
- 8 Чи потрібно моделювати пропуски/нездачі курсів студентами (наприклад, не всі студенти склали всі курси)?
- 9 Чи потрібно враховувати семестри (осінній/весняний) окремо, чи достатньо простої розбивки по роках?
- 10 Чи потрібно моделювати зміни у складі студентів/викладачів протягом 4 років (наприклад, випуск/відрахування студентів, зміна викладачів)?

Будь ласка, надайте відповіді, щоб я міг максимально якісно підготувати скрипт генерації даних!

рис. 3.2 Блок запитань 1 (спеціальності/кафедри, типи курсів, шкала оцінювання, додаткові атрибути, приклади аналітики).

Другий блок стосувався масштабів та часових правил: визначено орієнтовні кількості студентів і викладачів, типовий ритм (скільки курсів проходить студент на рік і скільки веде викладач за семестр), правило оцінювання лише за реально відвіданими курсами, ознаку очного/дистанційного формату для різних років, можливість коротких текстових коментарів, очікувані розподіли оцінок у складні періоди, частку відрахувань і рівень деталізації груп/потоків.

2. Яка оптимальна кількість студентів і викладачів для одного факультету (наприклад, 50, 100, 200 студентів; 10, 20, 30 викладачів)? Чи є побажання щодо співвідношення студентів до викладачів?
3. Чи правильно, що кожен студент проходить 6-8 курсів на рік, а кожен викладач веде 1-3 курси на семестр?
4. Для оцінювання курсів: чи залишає кожен студент відгук лише по тим курсам, які реально відвідував у відповідному семестрі? (Тобто не всі студенти оцінюють всі курси.)
5. Чи потрібно додавати до даних ознаку дистанційного/очного навчання (наприклад, у 2022-2023 роках більшість курсів дистанційні через пандемію/війну)?
6. Чи потрібно додавати до оцінок/метрик додаткові текстові коментарі (наприклад, короткий фідбек)?
7. Які оптимальні межі/розподіли для оцінок студентів і викладачів? Наприклад, середній бал 7-8/10, але з пониженням у роки пандемії/війни, більша дисперсія, поява низьких оцінок.
8. Чи потрібно моделювати відрахування студентів (наприклад, 2-3% студентів не завершують навчання)?
9. Чи потрібно враховувати академічні групи/потоки, чи достатньо лише спеціальності?

Після отримання відповідей я запропоную оптимальні обсяги даних (кількість студентів, викладачів, курсів, метрик, оцінок) для погодження.

рис. 3.3 Блок запитань 2 (обсяги, навантаження, очне/дистанційне, коментарі, розподіли, відрахування, деталізація груп).

Третій блок фіксував набір метрик і класи аналітичних задач. Для курсів і викладання узгоджено показники якості (зокрема: Clarity, Relevance, Engagement, Practical Value, Workload, Assessment Fairness, Support), мінімальні обсяги вибірок для задач Data Mining (класифікація, асоціативні правила, кластеризація), доцільність групування за спеціальністю та профілем

викладача, моделювання різної складності курсів і академічної неуспішності, а також часові вікна оцінювання (кожен семестр чи раз на рік).

- 1 Які саме метрики ви вважаєте найбільш релевантними для оцінки якості курсів у ВНЗ IT-напрямку? (Наприклад: clarity, relevance, engagement, practical value, workload, assessment fairness, support, resources, etc.) Чи достатньо 5-7 метрик для студентів і 3-5 для викладачів, чи потрібно більше/менше?
- 2 Для аналітики (1-Rule, асоціативні правила, кластеризація, Naive Bayes) оптимально мати хоча б 100-150 студентів і 15-20 викладачів, чи погоджуєтесь ви з такими обсягами?
- 3 Чи потрібно для кожної спеціальності створювати окремі групи, чи достатньо 1-2 груп на спеціальність (наприклад, ПЗ-21, КС-21, КБ-21 тощо)?
- 4 Чи потрібно, щоб викладачі також були розподілені по групах (наприклад, гуманітарні/технічні), чи достатньо просто вказати їх профіль (спеціалізацію)?
- 5 Чи потрібно моделювати різні рівні складності курсів (наприклад, базові, просунуті), які можуть впливати на оцінки?
- 6 Чи потрібно додавати до оцінок викладачів/студентів якісь додаткові ознаки (наприклад, чи був курс обов'язковим чи вибірковим)?
- 7 Чи погоджуєтесь ви з тим, що 10 курсів (3-4 спільних для всіх спеціальностей, решта – унікальні для окремих спеціальностей) – це оптимально для цього датасету?
- 8 Чи потрібно додавати випадки академічної неуспішності (наприклад, оцінка <4/10, не склав курс) як окремі записи?
- 9 Які часові межі для оцінювання: наприкінці кожного семестру (тобто 2 рази на рік), чи достатньо одного разу на рік?

рис. 3.4 Блок запитань 3 (перелік метрик якості, розміри вибірок для DM, групування, складність курсів, неуспішність, часові вікна).

За підсумком узгоджень сформовано склад і структуру даних. Варто підкреслити, що наведені нижче приклади відображають лише окремі фрагменти повного синтетичного набору, згенерованого спеціально для цього проєкту з урахуванням вимог до структури, змісту та реалістичності. Таблиця груп/кафедр містить стабільний ідентифікатор, назву (шифр), підрозділ/спеціальність і роль групи (student / lecturer), що дозволяє будувати зрізи успішності та чітко відділяти студентські й викладацькі записи.

Репрезентативні зрізи базових таблиць джерела даних показано на рис. 3.5–3.7.

	id_group	name	department	groupname
1	1	ПЗ-21-1	Інженерія програмного забезпечення	student
2	2	ПЗ-21-2	Інженерія програмного забезпечення	student
3	3	КС-21-1	Комп'ютерні науки	student
4	4	КС-21-2	Комп'ютерні науки	student
5	5	КБ-21-1	Кібербезпека	student
6	6	КБ-21-2	Кібербезпека	student
7	7	ІСТ-21-1	Інформаційні системи та технології	student
8	8	ІСТ-21-2	Інформаційні системи та технології	student
9	9	КІ-21-1	Комп'ютерна інженерія	student
10	10	КІ-21-2	Комп'ютерна інженерія	student
11	11	Викладачі ПЗ	Інженерія програмного забезпечення	lecturer
12	12	Викладачі КС	Комп'ютерні науки	lecturer
13	13	Викладачі КБ	Кібербезпека	lecturer
14	14	Викладачі ІСТ	Інформаційні системи та технології	lecturer
15	15	Викладачі КІ	Комп'ютерна інженерія	lecturer

Рис. 3.5 Таблиця груп (id, назва/шифр, кафедра/спеціальність, role).

Таблиця слухачів (студентів) містить службовий ідентифікатор, ім'я та прізвище, контактну адресу, роль і посилання на групу; це дає змогу простежувати успішність у межах груп, курсів і семестрів та пов'язувати дані з подіями оцінювання.

	id_listeners	Name_listeners	Surname_listeners	Contact_information	Role_listeners	id_group
1	1	Андрій	Коваль	андрій.коваль@email.ua	student	1
2	2	Марія	Шевченко	марія.шевченко@email.ua	student	1
3	3	Іван	Мельник	іван.мельник@email.ua	student	1
4	4	Олена	Бондар	олена.бондар@email.ua	student	1
5	5	Юрій	Кравчук	юрій.кравчук@email.ua	student	1
6	6	Тетяна	Сидоренко	тетяна.сидоренко@email.ua	student	1
7	7	Віктор	Гнатюк	віктор.гнатюк@email.ua	student	1
8	8	Оксана	Петренко	оксана.петренко@email.ua	student	1
9	9	Дмитро	Лисенко	дмитро.лисенко@email.ua	student	1
10	10	Наталія	Романюк	наталія.романюк@email.ua	student	1
11	11	Павел	Вишневський	павел.вишневський@email.ua	student	2
12	12	Ева	Ковальська	ева.ковальська@email.ua	student	2
13	13	Адам	Зелінський	адам.зелінський@email.ua	student	2
14	14	Анна	Яблонська	анна.яблонська@email.ua	student	2
15	15	Міхал	Камінський	міхал.камінський@email.ua	student	2
16	16	Лаура	Калініна	лаура.калініна@email.ua	student	2
17	17	Едгар	Біріньш	едгар.бірзінш@email.ua	student	2

Рис. 3.6 Таблиця слухачів (id, ПІБ, контакт, group\_id, role).

Центральним елементом є таблиця результатів/подій оцінювання: посилання на тест/подію та слухача, дата у форматі YYYY-MM-DD, назва метрики (Selective\_metric) і числові значення за кількома критеріями

(наприклад, Rating\_results, Answers\_results, Interest\_results). Саме з цієї таблиці будуться агрегати та КРІ.

	id_results	id_tests	Rating_results	Answers_results	Interest_results	Selective_metric	id_metric	id_listeners	test_date
1	1	1	6	6	66	Clarity	1	1	2022-06-07
2	2	1	7	7	78	Relevance	2	1	2022-06-05
3	3	1	7	7	73	Engagement	3	1	2022-06-15
4	4	1	7	7	74	Practical Value	4	1	2022-06-18
5	5	1	8	8	70	Workload	5	1	2022-06-05
6	6	1	7	7	80	Assessment Fairness	6	1	2022-06-15
7	7	1	7	7	80	Support	7	1	2022-06-18
8	8	2	6	6	54	Clarity	1	1	2022-12-08
9	9	2	6	6	70	Relevance	2	1	2022-12-07
10	10	2	8	8	73	Engagement	3	1	2022-12-12
11	11	2	3	3	40	Practical Value	4	1	2022-12-07
12	12	2	7	7	67	Workload	5	1	2022-12-10
13	13	2	8	8	73	Assessment Fairness	6	1	2022-12-16
14	14	2	8	8	78	Support	7	1	2022-12-12
15	15	3	7	7	64	Clarity	1	1	2023-06-08
16	16	3	6	6	70	Relevance	2	1	2023-06-11
17	17	3	8	8	71	Engagement	3	1	2023-06-13

Рис. 3.7 Таблиця результатів (id, id\_tests, Selective\_metric, значення показників, дата, id\_listeners).

Наведені фрагменти й ілюстрації демонструють робочий зріз набору даних на момент підготовки розділу. Для кожної сутності визначено первинні й зовнішні ключі для цілісності посилань, а назви в довідниках нормалізовано, щоб уникати дублікатів і спростити перевірки якості. Дані є синтетичними та не містять персональних відомостей реальних осіб. Висновки за результатами аналізу розглядаються як методологічні та в разі впровадження потребують підтвердження на продуктивних даних. Логічну модель сховища, політику історизації та правила завантаження наведено окремо в п. 3.4.

### 3.4 Опис сховища даних

Сховище організовано за схемою зірка з однією центральною таблицею фактів і набором узгоджених вимірів навколо неї. Така модель спрощує з'єднання, забезпечує передбачувані агрегування та слугує єдиною семантичною основою для побудови КРІ й OLAP-зрізів. На рисунку подано логічну структуру сховища з основними сутностями і зв'язками.[20]

Логічну структуру сховища даних (зіркова модель) подано на рис. 3.8.

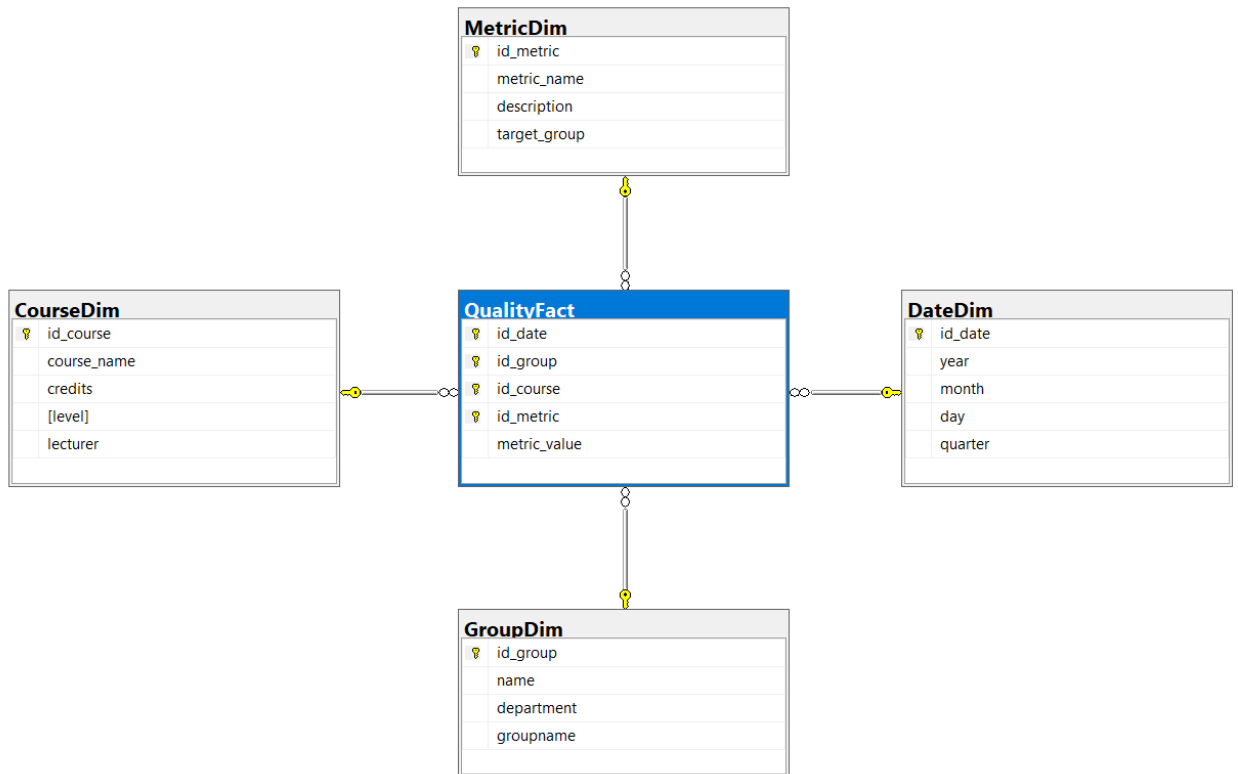


Рис. 3.8 Структура сховища даних

Таблиця фактів і вимірів:

**QualityFact**: основна таблиця фактів, яка об'єднує дані з усіх вимірів та містить

значення метрик, пов'язаних із якістю навчального процесу.

**DateDim**: таблиця часу для аналізу в часових розрізах.

**CourseDim**: таблиця, яка містить деталі про курси, їх складність і викладачів.

**GroupDim**: таблиця, яка надає інформацію про студентські групи.

**MetricDim**: таблиця вимірів, яка описує параметри та метрики для аналізу якості.

Рівень деталізації таблиці фактів визначено як одне агреговане значення конкретної метрики якості для певної академічної групи на певному курсі у конкретну дату. Тобто кожен запис у фактовій таблиці відповідає комбінації  $id\_date \times id\_group \times id\_course \times id\_metric$ . Такий рівень деталізації зменшує обсяг куба, пришвидшує агрегації й водночас залишає можливість розширення моделі (за потреби – до рівня окремого студента).

Усі виміри мають сурогатні числові ключі, на які посилається фактова таблиця зовнішніми ключами. Для запобігання дублюванню вводиться унікальність комбінації ключів зерна у факті. Значення `metric_value` трактується як попередньо агрегована міра (наприклад, середній бал або інший підсумок), тому по часовій осі застосовуються відповідні правила згортання (середнє/медіана/зважене середнє – згідно з визначеннями KPI). Це забезпечує від некоректних сумувань у звітах.

Ієрархії у вимірах задаються явно. Для часу використовується послідовність рік, квартал/семестр, місяць, день. Для курсів – атрибути на кшталт дисципліни, кредитності, рівня та викладача. Для груп – спеціальність/підрозділ і шифр групи. Для метрик – словник назв і описів, який стандартизує семантику показників у всіх звітах і моделях.

### **3.5 Побудова розгорнутого куба в середовищі проекту служби SSAS**

Служби аналізу SQL Server (SSAS) є одним із ключових інструментів для роботи з OLAP (Online Analytical Processing), завдяки якому можна створювати багатовимірні моделі даних для аналізу великих масивів інформації. Цей механізм підтримує побудову аналітичних кубів, що дають швидкий доступ до даних і спрощують процес їхнього аналізу. Використовуючи SSAS у середовищі Visual Studio можливо створювати гнучкі рішення, які можливо адаптувати та використовувати в багатьох випадках. SSAS дозволяє працювати з багатовимірними (MOLAP) та табличними (Tabular) моделями, підтримуючи різноманітні вимоги до аналітичної обробки даних.[21]

Процес побудови OLAP-куба у середовищі SSAS включає визначення джерела даних, налаштування вимірів і створення таблиці фактів. Куб дозволяє аналізувати дані у різних аспектах, таких як час, курси, групи чи оцінювальні метрики, в результаті чого він може бути використаний як ефективний

інструмент для прийняття рішень. У цьому розділі буде детально розглянуто основні етапи створення розгорнутого куба для аналітичної системи оцінювання якості навчального процесу.

Першим етапом є налаштування підключення до джерела даних. У нашому випадку джерелом даних виступає оперативна база даних (ОБД), створена раніше для системи. За допомогою Data Source Wizard було додано підключення до бази даних, як показано на рис.3.9.

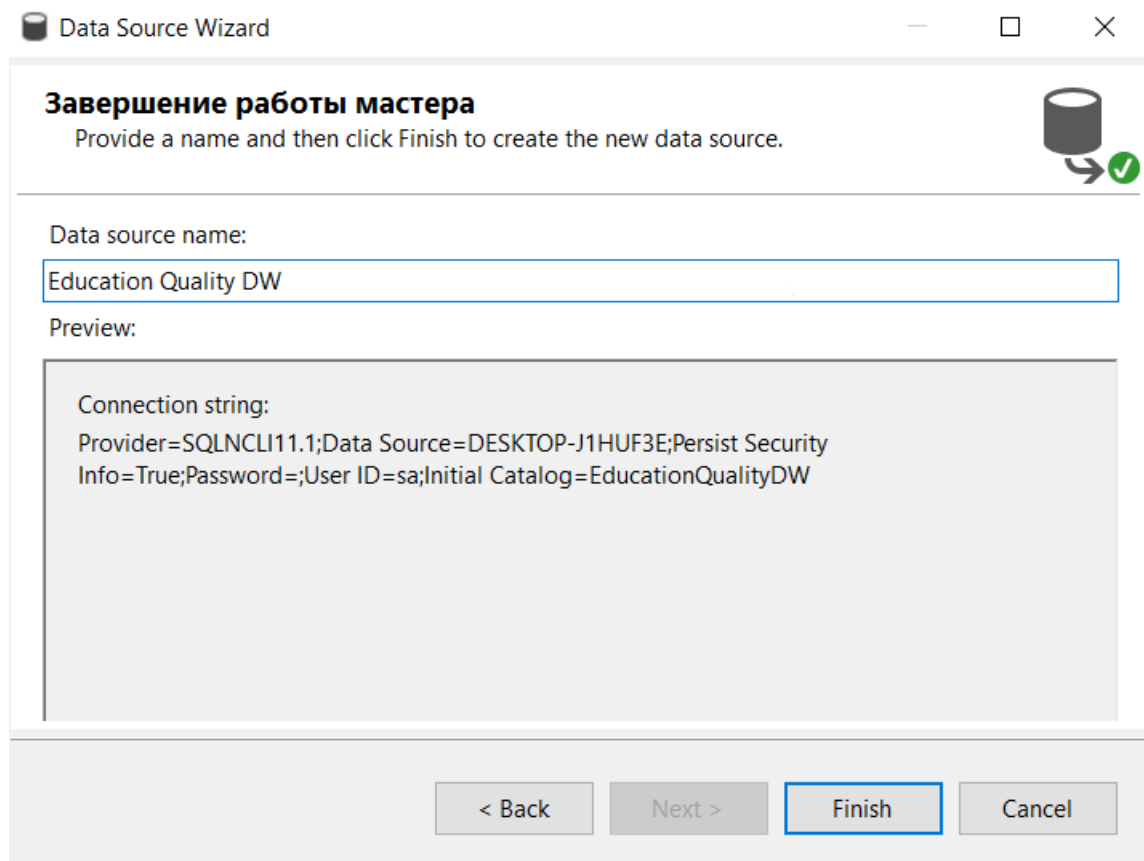


Рис. 3.9 Додавання підключення

Наступним кроком стало створення уявлення джерела даних за допомогою Data Source View Wizard. DSV виступає як логічна модель, яка абстрагується від фізичного джерела даних і дозволяє створювати зв'язки між таблицями, об'єднувати поля або змінювати їх назви. На рис. 3.10 зображено процес створення DSV для використання таблиць з оперативної бази даних.

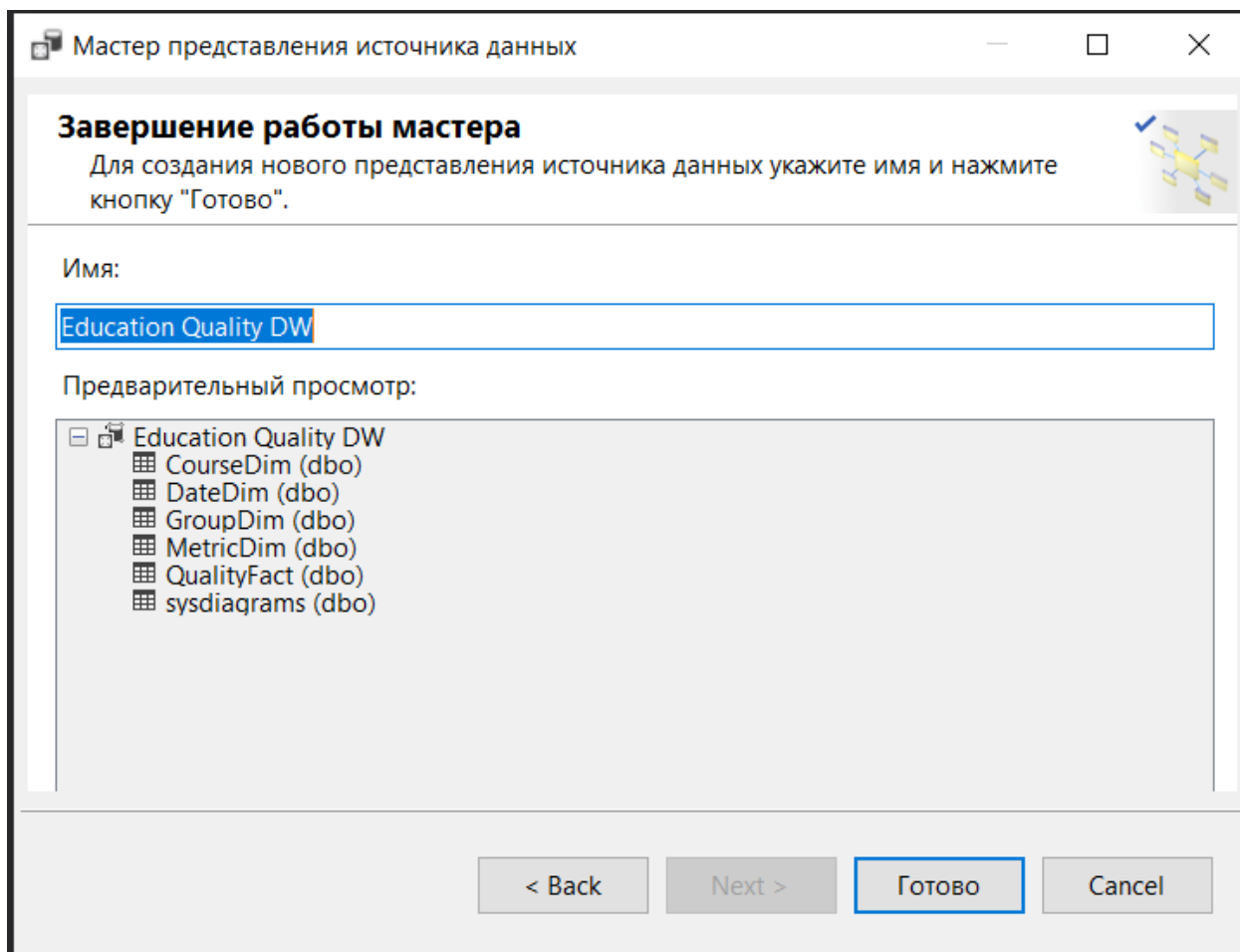


Рис. 3.10 Створення уявлення джерела даних

За допомогою Cube Wizard було створено багатовимірний куб. Під час налаштування було виконано наступні дії (рис. 3.11–3.13):

1. Вибір таблиці фактів QualityFact як основи для куба.
2. Додавання створених вимірів до куба.
3. Вибір атрибутів, які будуть використовуватися у вимірах та зв'язках.

Мастер кубов

### Выбор таблиц групп мер

Выберите представление или диаграмму источника данных, а затем выберите таблицы, которые будут использованы для групп мер.

Представление источника данных:  
Education Quality DW

Таблицы групп мер: Предложить

- CourseDim
- DateDim
- GroupDim
- MetricDim
- QualityFact
- sysdiagrams

< Back   Next >   Finish >>|   Cancel

Мастер кубов

### Выберите меры

Выберите меры, которые необходимо включить в куб.

- Мера
  - Course Dim
    - Credits
    - Число Course Dim
  - Date Dim
    - Year
    - Month
    - Day
    - Quarter
    - Число Date Dim
  - Quality Fact
    - Metric Value
    - Число Quality Fact
  - Metric Dim
    - Число Metric Dim
  - Group Dim
    - Число Group Dim

< Back   Next >   Finish >>|   Cancel

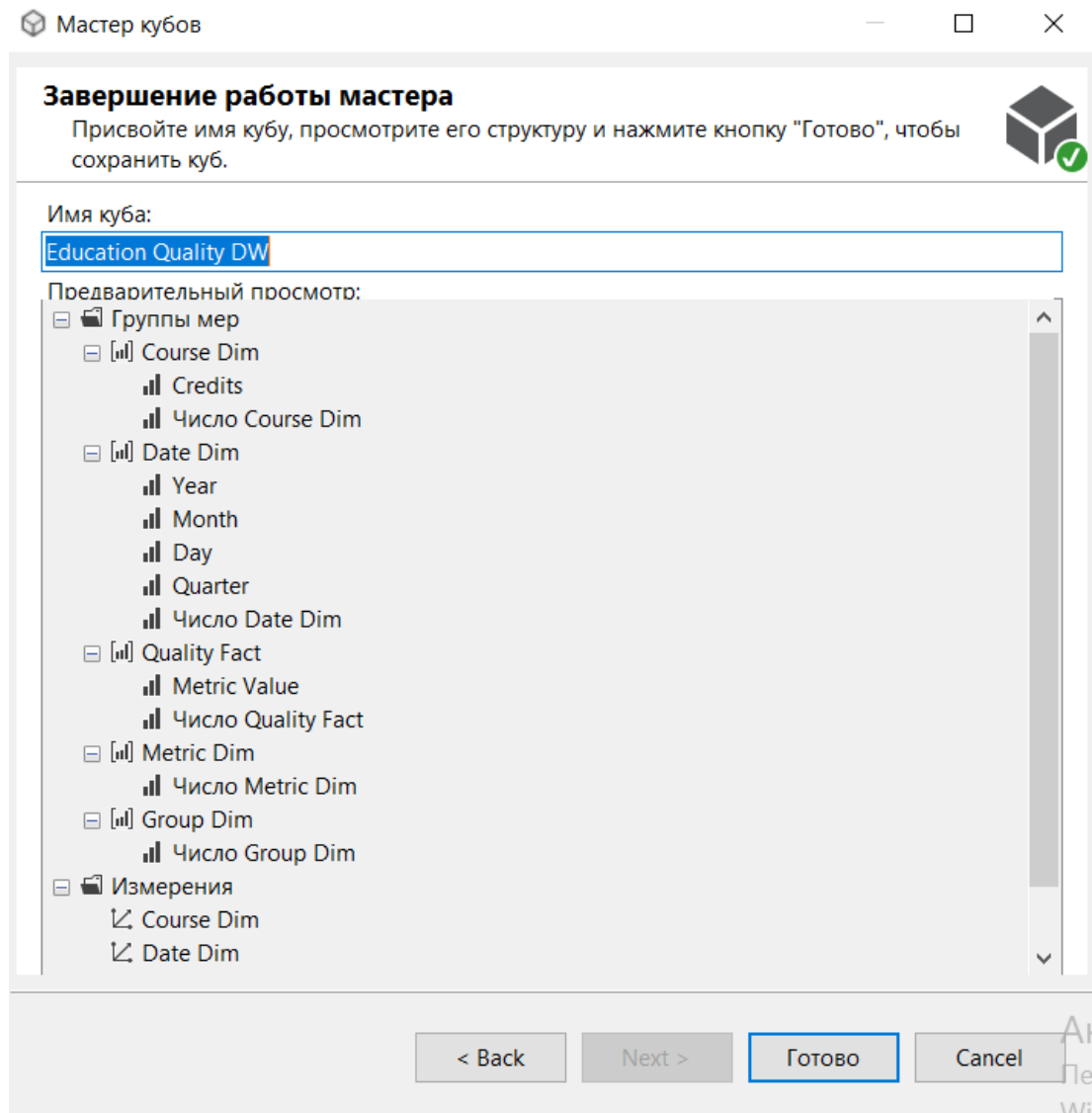


Рис. 3.11–3.13 Формування кубу.

На фінальному етапі було проведено обробку куба(рис. 3.14). Результатом є створений куб QualityCube, який забезпечує аналіз ефективності навчального процесу за різними вимірами, як показано на рис. 3.15.

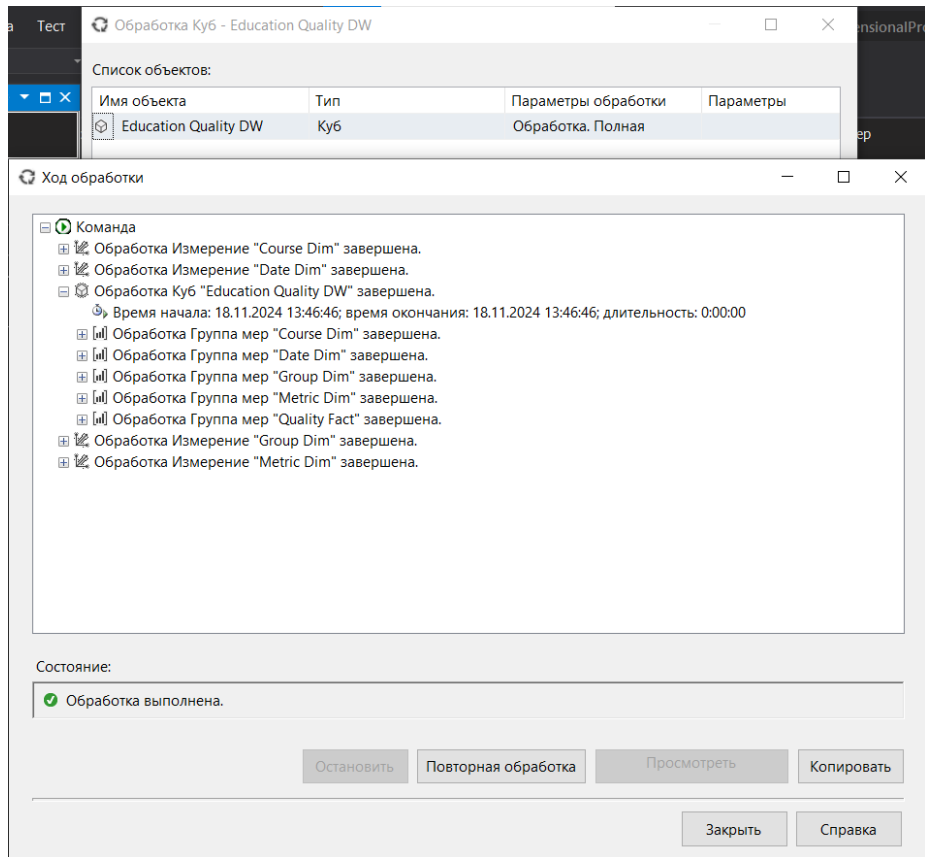


Рис. 3.14 Обработка куба.

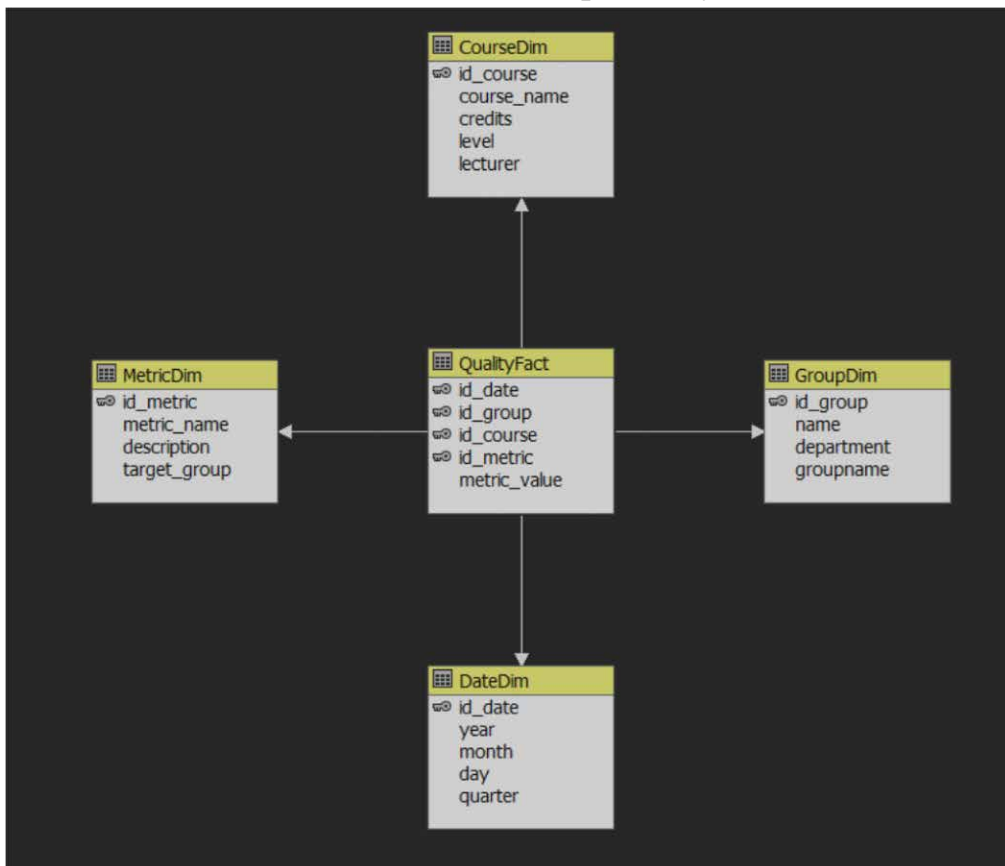


Рис. 3.15 Створений куб.

### 3.6 Наповнення кубу даними

Процес передачі даних до сховища є одним із ключових етапів побудови аналітичної системи, оскільки забезпечує інтеграцію інформації з різних джерел у єдину базу для подальшої обробки. Для реалізації цього процесу використовується SQL Server Integration Services (SSIS) – відповідний ETL–інструмент, що дозволяє здійснювати вилучення, трансформацію та завантаження (Extract, Transform, Load) даних з оперативної бази до сховища. Завдяки SSIS забезпечується створення потоків даних, які адаптують вихідну інформацію до потреб аналітичної системи, що підвищує її ефективність і продуктивність.[22]

Інтеграція даних за допомогою SSIS дозволяє:

- Об'єднувати дані з різних джерел у централізоване сховище.
- Оптимізувати структуру даних для виконання аналітичних запитів.
- Забезпечити швидке наповнення сховища за допомогою автоматизованих потоків даних.

Процес реалізації передачі даних, що зображений на рис. 3.16, до сховища даних у нашій аналітичній системі базується на таких етапах:

1. Заповнення таблиць вимірів.
2. Заповнення таблиці фактів.

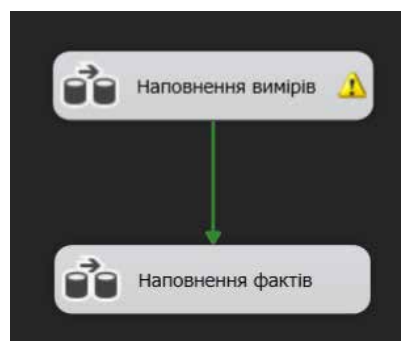


Рис. 3.16 Потоки даних для наповнення СД

На першому етапі було реалізовано перенесення даних з оперативної бази даних (ОБД) до таблиць вимірів у сховищі даних (СД). За допомогою Data Flow Task у середовищі SSIS були налаштовані потоки для кожної таблиці вимірів. У процесі реалізації використовувалися наступні компоненти:

1. Source Assistant: для вибору джерела даних.
2. Data Conversion: для приведення типів даних до вимог сховища.
3. OLE DB Destination: для завантаження даних у таблиці вимірів.

На рис. 3.17 зображено процес заповнення даних для таблиць виміру.

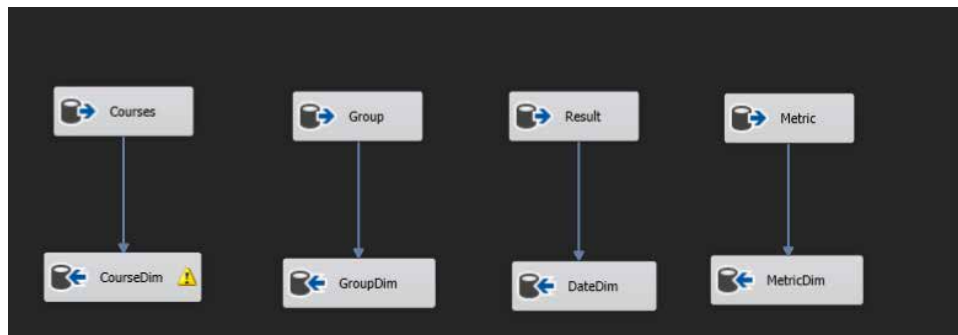


Рис. 3.17 Наповнення вимірів

На рис. 3.18 зображено процес налаштування потоку даних для таблиці CourseDim. Потік формує дані про курси (ідентифікатори, назви, типи – базовий/вибірковий/практичний, кредити, рівень/складність, викладача) на основі оперативної бази.

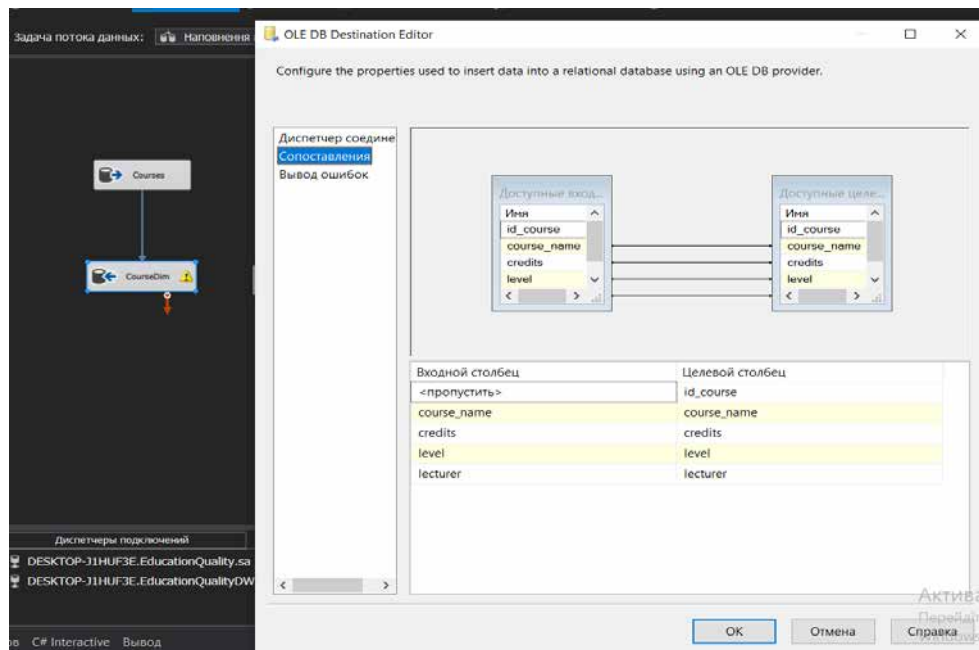


Рис. 3.18 Заповнення таблиці виміру CourseDim

На другому етапі відбувається наповнення таблиці фактів QualityFact, що містить основні метрики для аналізу якості навчального процесу. Перед записом у QualityFact дані агрегуються до визначеного рівня деталізації. Процес включав об'єднання даних з таблиць вимірів і фактів оперативної бази, а також трансформацію даних перед їх завантаженням до СД. Використовувались такі елементи SSIS:

- Merge Join: для об'єднання даних з різних джерел.
- Derived Column: для розрахунку додаткових полів.
- Conditional Split: для фільтрації даних відповідно до критеріїв.

На рис. 3.19–3.20 зображено приклад створення запиту для наповнення таблиці фактів. Запит формує дані про курси, групи, дати, метрики та значення оцінок, об'єднуючи їх з відповідними вимірами.

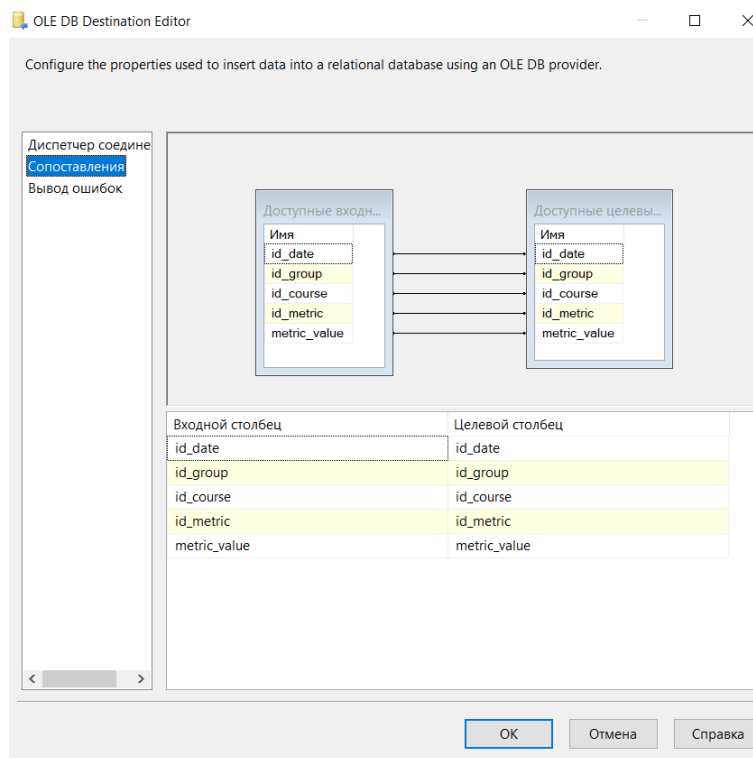
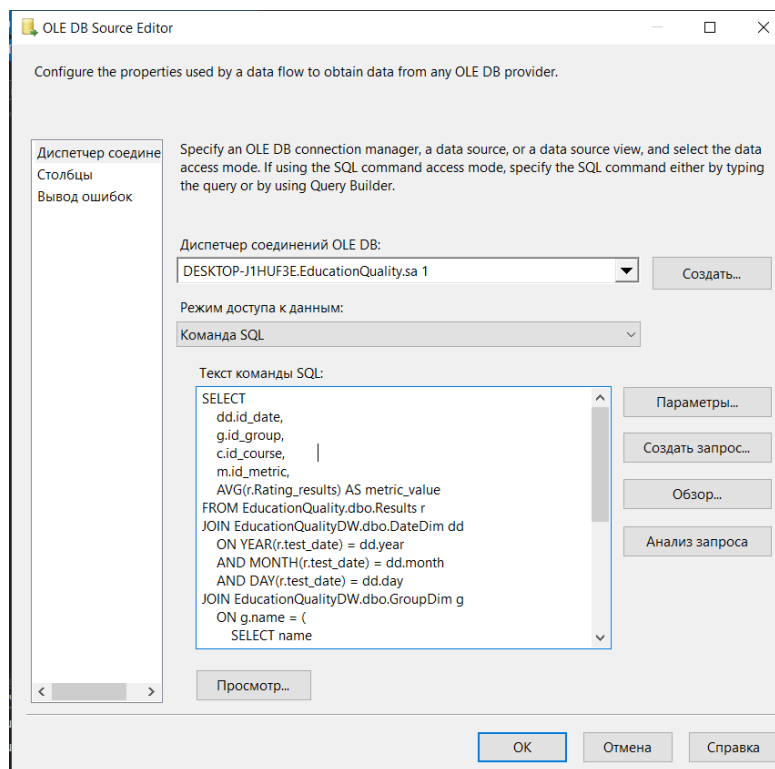


Рис. 3.19–3.20 Процесс формування вибірки та передачі даних на основі сформованої вибірки

Після завершення процесу передачі даних у сховищі даних було отримано заповнені таблиці вимірів і фактів, які дозволяють здійснювати багатовимірний аналіз.

## 4 АНАЛІЗ РЕЗУЛЬТАТІВ ДОСЛІДЖЕННЯ

### 4.1 OLAP: ключові показники ефективності та звіти

Ключові показники ефективності (Key Performance Indicators) в більшості випадків розглядаються як кількісні метрики, що забезпечують регулярний моніторинг досягнення визначених цілей у навчальному процесі. Використовуючи їх можна зосередитися на суттєвих аспектах якості, своєчасно виявляти слабкі місця та фіксувати динаміку змін у стандартизованому форматі, придатному для порівнянь між періодами, курсами та академічними групами.[23]

В ході виконання дослідження використовуються два узгоджені показники, які продемонстровані на рис. 4.1: середня якість курсу (Quality) та рівень задоволеності студентів (Satisfaction). Обидвоє обчислюються на агрегованих даних за 2023 рік із порівнянням до попереднього періоду. В результаті чого можна фіксувати поточний стан і відстежувати тенденції: які курси демонструють покращення, а які – погіршення динаміки.

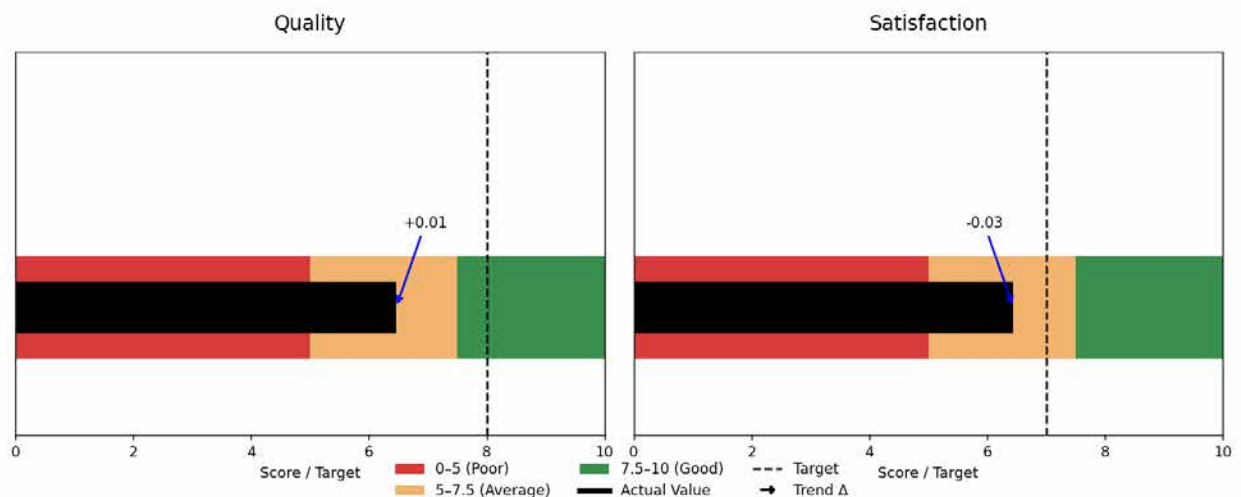


рис. 4.1 Підсумковий перегляд КРІ.

За підсумковим переглядом обидва показники утримуються переважно в межах середньої зони (близько 6 – 7,5 бала за 10-бальною шкалою), не досягаючи встановлених цільових орієнтирів. При цьому Quality демонструє

незначне покращення відносно попереднього року (+0,01), що свідчить про стабільність або поступове вдосконалення, тоді як Satisfaction має невеликий негативний зсув (-0,03), що вказує на окремі організаційні чи методичні чинники, які потребують подальшої перевірки. У такій конфігурації KPI дозволяє оперативно локалізувати курси та періоди з відхиленнями і задати пріоритети для поглибленого аналізу у звітних формах OLAP і на етапі моделювання.

У межах OLAP-блоку звіти виконують прикладну роль: вони дають швидкі, відтворювані відповіді на конкретні запитання й допомагають перейти від загальної картини KPI до точкових управлінських дій. Ми зосереджуємося на двох формах, які покривають базові потреби: порівняння курсів у фіксованому періоді та відстеження динаміки показників упродовж навчального року.

Звіт Курси × метрики, що продемонстрований на рис. 4.2 відповідає на запитання: які дисципліни за підсумком року демонструють кращі/гірші значення показників і де спостерігається підвищена варіативність між академічними групами. Результат подано як впорядковану вітрину з атрибутами курсу (назва, кредити, рівень, викладач) і агрегованим значенням метрики. Лідери — кандидати на поширення практик; аутсайдери — на предметну перевірку причин (навантаження, формат оцінювання, темп курсу).

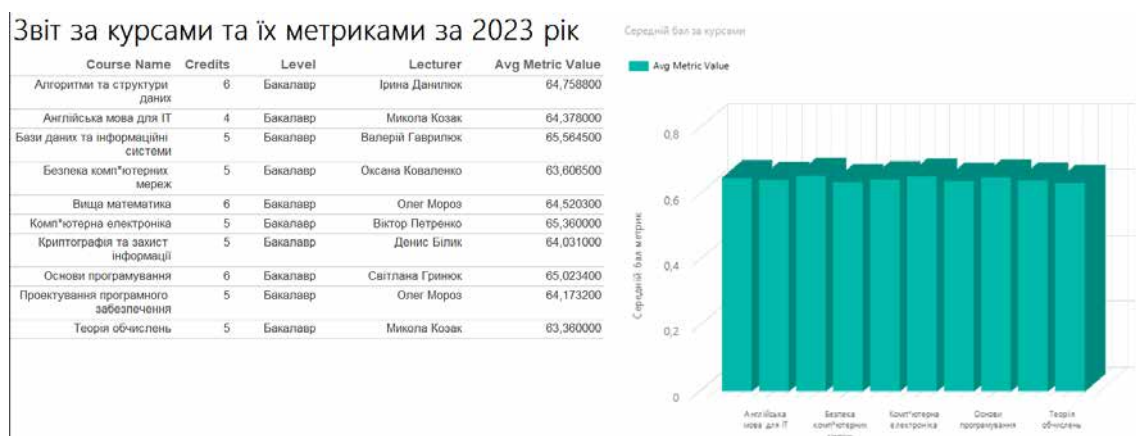


Рис. 4.2 Результати формування звіту за курсами та метриками за визначений рік.

Семестровий звіт (рис. 4.3) показує, як змінюються ті самі показники в часі та чи збігаються коливання між Quality і Satisfaction. Він дає відповідь на запитання: у яких періодах виникають локальні спади/покращення, чи є сезонність, і чи пов'язана вона з організаційними факторами навчального процесу. Легко помічаємо розбіжності, коли якість утримується стабільною, а задоволеність просідає, — це сигнал для окремого розбору формату взаємодії зі студентами.

## Семестровий звіт за 2023 рік

	2 квартал	4 квартал
Алгоритми та структури дан	65,451600	64,070500
Англійська мова для ІТ	64,918500	63,846100
Бази даних та інформаційні системи	64,615300	66,610100
Безпека комп'ютерних мер	62,280700	64,769200
Вища математика	64,318100	64,723100
Комп'ютерна електроніка	65,483800	65,238000
Криптографія та захист інформації	63,281200	64,769200
Основи програмування	65,261500	64,778400
Проектування програмного забезпечення	64,218700	64,126900
Теорія обчислень	65,172400	61,791000

Рис. 4.3 Результати формування семестрового звіту за 2023 рік

У парі з підсумковими КРІ ці два звіти дають відповіді на практичні питання: які курси варто переглянути першими та в яких семестрах змінюється картина показників. Після їх перегляду формується короткий перелік дій — від корекції методики оцінювання до перевірки розкладу й навантаження на потоки.

## 4.2 Методи Data Mining для інтелектуального аналізу

**Алгоритм 1–Rule (One–rule, 1R)** належить до найпростіших методів класифікації: класи утворюються за єдиним правилом, побудованим на одному атрибуті. Попри мінімалізм, підхід корисний для первинної оцінки роздільної здатності ознак і як базова лінія для подальшого порівняння зі складнішими моделями.[24]

У межах дослідження 1–Rule застосовано для класифікації якості навчального процесу на основі даних сховища EducationQualityDW (Microsoft SQL Server). Цільовою змінною слугує категорія QualityCategory з трьома класами (Low, Medium, High), що утворюється на основі числового показника metric\_value з таблиці QualityFact. Перетворення виконано обчислюваним полем: значення metric\_value до 6 включно віднесено до Low, у (6; 8] – до Medium, понад 8 – до High. Така дискретизація переводить числову оцінку в узгоджену категоріальну шкалу, придатну для класифікації.

Незалежні змінні сформовано з атрибутів предметної області: CourseName, GroupName, Year, Month. Дані завантажуються додатком на C# / Windows Forms, який об'єднує таблицю фактів (QualityFact) з вимірами (DateDim, GroupDim, CourseDim, MetricDim) і для кожної кандидатної ознаки виконує стандартну процедуру 1–Rule: групування значень, визначення найбільш поширеного класу в кожній групі (MajorityClass), підрахунок кількості коректних класифікацій і помилок, а також розрахунок точності (Accuracy). Ознака з найменшою кількістю помилок вважається найкращою для побудови правила.

Аналіз

Course	Total	MajorityClass	CorrectCount	Accuracy (%)
Вища математи...	2489	Medium	1282	51,50662916834...
Основи програ...	2543	Medium	1324	52,06449075894...
Англійська мов...	2527	Medium	1314	51,99841709537
Алгоритми та ст...	2495	Medium	1292	51,78356713426...
Проектування п...	514	Medium	249	48,44357976653...
Теорія обчислень	511	Medium	260	50,88062622309...

GroupName	Total	MajorityClass	CorrectCount	Accuracy (%)
ПЗ-21-1	2144	Medium	1118	52,14552238805...
ПЗ-21-2	2134	Medium	1094	51,26522961574...
КС-21-1	2575	Medium	1324	51,41747572815...
КС-21-2	2121	Medium	1130	53,27675624705...
КБ-21-1	2115	Medium	1084	51,25295508274...
Викладачі ПЗ	384	Medium	202	52,60416666666...

Year	Total	MajorityClass	CorrectCount	Accuracy (%)
2023	3249	Low	1696	52,20067713142...
2022	3278	Low	1670	50,94569859670...
2025	3259	Medium	2025	62,13562442467...
2024	3313	Medium	2044	61,69634772109...

Правила для кожного атрибута:  
 Аналіз за CourseName:  
 Якщо CourseName = "Вища математика", то клас = "Medium" (імовірність: 51,5%)  
 Якщо CourseName = "Основи програмування", то клас = "Medium" (імовірність: 52,1%)  
 Якщо CourseName = "Англійська мова для ІТ", то клас = "Medium" (імовірність: 52,0%)  
 Якщо CourseName = "Алгоритми та структури даних", то клас = "Medium" (імовірність: 51,8%)  
 Якщо CourseName = "Проектування програмного забезпечення", то клас = "Medium" (імовірність: 48,4%)  
 Якщо CourseName = "Теорія обчислень", то клас = "Medium" (імовірність: 50,9%)  
 Якщо CourseName = "Курсова робота з аналізу даних", то клас = "Medium" (імовірність: 51,5%)

Опрацювати

Рис. 4.4 Результати роботи 1–R.

На ілюстрації для кожної ознаки наведено сформовані правила разом із кількістю правильних спрацювань та точністю. Зокрема, для окремих груп (ПЗ–21–1) домінує клас Medium із точністю близько 52,7%; для інших груп показники коливаються біля 50–53%. За ознакою Year спостерігається невеликий зсув: у 2023 році домінує Medium ( $\approx 50,8\%$ ), у 2022 — Low ( $\approx 50,7\%$ ), тобто відмінності між роками є, але вони несуттєві. Узагальнюючи, жодна з розглянутих одиничних ознак не забезпечує високої точності: класи розподілені відносно рівномірно, і 1–Rule фіксує лише слабкі, але інтерпретовані зв'язки. Це узаконює використання 1–Rule як базової лінії та водночас мотивує перехід до багатовимірних методів.

**Метод наївного Байєса (Naive Bayes Classifier)** є статистичним алгоритмом класифікації, що широко використовується в галузі Data Mining для аналізу великих обсягів даних. Його популярність зумовлена поєднанням математичної простоти з доволі високою точністю результатів у задачах, пов'язаних із категоріальними змінними. Метод базується на теоремі Байєса, яка дозволяє визначити ймовірність приналежності об'єкта до певного класу за відомих значень його ознак, за умови незалежності цих ознак одна від одної. [25]

В ході виконання дослідження класифікатор наївного Байєса застосовано для визначення категорії якості освітнього процесу на даних сховища EducationQualityDW (Microsoft SQL Server). Ключовим показником, що підлягав класифікації, був `metric_value` – числове значення, що характеризує конкретну освітню метрику. На основі цього значення було сформовано допоміжну змінну `QualityCategory`, яка поділяє записи на три категорії:

- Low – якщо `metric_value`  $\leq 6$ ;
- Medium – якщо `metric_value`  $> 6$  та  $\leq 8$ ;
- High – якщо `metric_value`  $> 8$ .

Категорію якості (`QualityCategory`) було обрано як цільову змінну, а незалежними ознаками виступали: назва курсу (`CourseName`), назва групи (`GroupName`), рік (`Year`) та місяць (`Month`).

Реалізація класифікатора наївного Байєса у середовищі C# була виконана у вигляді десктопного додатка Windows Forms. Програма з'єднується з базою даних, виконує запит до кількох таблиць сховища (`QualityFact`, `CourseDim`, `GroupDim`, `DateDim`, `MetricDim`), об'єднує дані у відповідні структури і формує навчальну вибірку. Після попередньої обробки дані використовуються для побудови ймовірнісної моделі.

У процесі навчання програма:

- обчислює апіорні ймовірності для кожного класу на основі частоти їх зустрічання;

- розраховує умовні ймовірності для кожного значення атрибутів у контексті кожного класу;
- застосовує згладжування Лапласа, щоб уникнути нульових значень;
- класифікує кожен запис на основі добутку ймовірностей (у логарифмічній формі).

Результати роботи алгоритма продемонстровані на рис. 4.5.

FullDate	GroupName	CourseName	MetricValue	ActualCategory	PredictedCategory
11.12.2023	ПЗ-21-1	Вища математи...	7,00	Medium	Low
11.12.2023	ПЗ-21-1	Вища математи...	4,00	Low	Low
11.12.2023	ПЗ-21-1	Вища математи...	6,00	Low	Low
11.12.2023	ПЗ-21-1	Вища математи...	5,00	Low	Low
11.12.2023	ПЗ-21-1	Вища математи...	7,00	Medium	Low
11.12.2023	ПЗ-21-1	Основи програ...	8,00	Medium	Low
11.12.2023	ПЗ-21-1	Основи програ...	7,00	Medium	Low
11.12.2023	ПЗ-21-1	Основи програ...	5,00	Low	Low
11.12.2023	ПЗ-21-1	Основи програ...	6,00	Low	Low
11.12.2023	ПЗ-21-1	Англійська мов...	6,00	Low	Low
11.12.2023	ПЗ-21-1	Англійська мов...	5,00	Low	Low
11.12.2023	ПЗ-21-1	Алгоритми та ст...	6,00	Low	Low
11.12.2023	ПЗ-21-1	Алгоритми та ст...	6,00	Low	Low

Naive Bayes Classification Report:  
 Total instances: 13099  
 Correctly classified: 7435  
 Overall Accuracy: 56,76%

Prior Probabilities:  
 Medium : 51,61%  
 Low : 36,20%  
 High : 12,19%

Рис. 4.5 Результат роботи Наївного Байеса

У таблиці результатів, представлений на рисунку, наведено приклади класифікації окремих записів: для кожної дати, групи та курсу вказано фактичне значення метрики якості (MetricValue), реальну категорію якості (ActualCategory) та категорію, яку передбачила модель (PredictedCategory).

Загалом, класифікатор наївного Байеса опрацював 13 099 записів, з яких 7 345 були віднесені до правильного класу з загальною точністю 56,7%, що є

суттєвим покращенням порівняно з попередніми результатами. Такий рівень точності свідчить про те, що модель здатна враховувати залежності між ознаками та класом якості, навіть попри припущення про незалежність ознак.

У нижній частині вікна наведено розподіл ймовірностей для кожного класу: найбільшу частку складає клас "Medium" (61,61%), далі йде "Low" (26,18%) та "High" (12,19%).

**Асоціативні правила.** Описані на рис. 4.6 асоціативні правила пов'язують поєднання ознак (наприклад, «курс × група») з типовими наслідками. Їхню надійність оцінювали за чотирма показниками: підтримкою (частка записів, де правило спрацьовує), довірою (ймовірність наслідку за умови антецеденту), ліфтом/підсиленням (наскільки зв'язок сильніший за випадковий) і конвікцією (стійкість правила, якщо наслідок не відбувається). [26]

Наприклад, одне з правил має вигляд: якщо студент вивчає курс "Бази даних та інформаційні системи" у групі "ПЗ–21–2", то з великою ймовірністю він також пов'язаний із певною іншою освітньою подією або метрикою. Для цього правила довіра становить понад 80%, а ліфт перевищує 1,5, що свідчить про сильну асоціацію між цими ознаками. Подібні закономірності повторюються і для інших курсів та груп, наприклад, для "Безпеки комп'ютерних мереж" у тій же групі або для "Проектування програмного забезпечення".

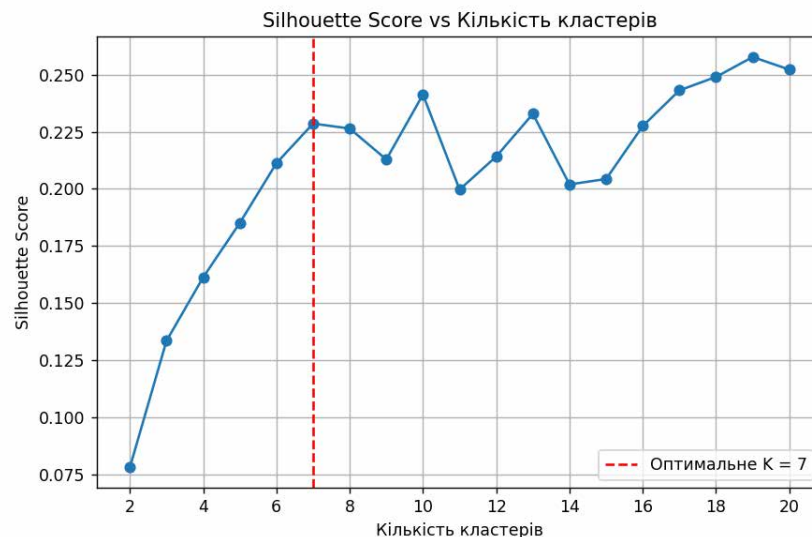
Особливої уваги заслуговують правила з максимальною довірою (наближеною до 1) та високою конвікцією, адже вони вказують на практично детермінований зв'язок між антецедентом і консеквентом. Це означає, що в межах наявних даних наслідок майже завжди настає за умови виконання відповідної умови. Водночас у таблиці є й правила з довірою у межах 0,5–0,7, які хоч і не є абсолютними, але все ж вказують на значну ймовірність спільної появи ознак.

Асоціативні правила:						
	Антеcedенти (умови)	Консеквенції (результат)	Підтримка антеcedентів (частота)	Підтримка консеквенцій (частота)	Підтримка (частота)	Довіра (ймовірн)
0	(Course=Бази даних та інформаційні системи)	(Group=КС-21-2)	0.038324	0.161921	0.032293	0.8
1	(Course=Безпека комп'ютерних мереж)	(Group=КС-21-1)	0.038171	0.196580	0.032369	0.8
2	(Course=Комп'ютерна електроніка)	(Group=КС-21-1)	0.038400	0.161463	0.032445	0.8
3	(Course=Криптографія та захист інформації)	(Group=КС-21-1)	0.039316	0.196680	0.033285	0.8
4	(Course=Проектування програмного забезпечення)	(Group=ПЗ-21-1)	0.039240	0.163677	0.033514	0.8
5	(Course=Теорія обчислень)	(Group=ПЗ-21-2)	0.039011	0.162913	0.032980	0.8

Рис. 4.6 Результат роботи алгоритму пошуку асоціативних правил

**Кластерний аналіз** застосовано для виділення однорідних груп спостережень без заздалегідь заданих міток. У нашому контексті це дає змогу побачити типові портрети навчальних ситуацій і відмежувати масові сценарії від нішевих. Працюємо з числовим вектором ознак, сформованим із агрегованих показників у розрізі курсів/груп і календаря. Перед моделюванням виконано стандартизацію, категоріальні атрибути перетворено на індикатори. Така підготовка забезпечує коректну роботу алгоритму на різномасштабних змінних.

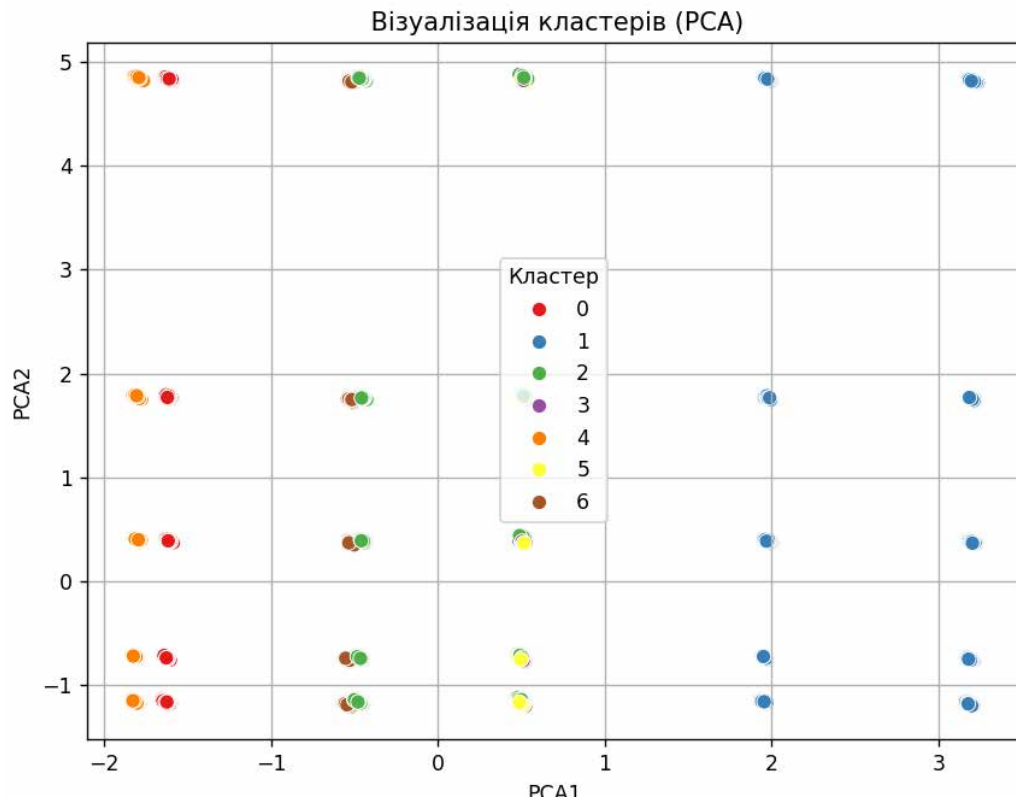
Для сегментації використано K–Means. Алгоритм ініціалізує k центрів (центроїдів), відносить кожний запис до найближчого центру, оновлює центри як середні значення точок кластера і повторює кроки до збіжності. Вибір кількості кластерів здійснювався емпірично за показником Silhouette Score, як бачимо з рисунків 4.7 та 4.8: порівнювалися значення для кількох k, після чого обирали конфігурацію з найкращим балансом: щільність всередині – роздільність між кластерами. За цим критерієм оптимальним виявився поділ на сім кластерів, локальний максимум силуета спостерігається коли  $k = 7$ . [27]



```
Прирости Silhouette Score:  
K = 3: приріст = 0.0554  
K = 4: приріст = 0.0278  
K = 5: приріст = 0.0238  
K = 6: приріст = 0.0262  
K = 7: приріст = 0.0173  
K = 8: приріст = -0.0022  
K = 9: приріст = -0.0135  
K = 10: приріст = 0.0284  
K = 11: приріст = -0.0416  
K = 12: приріст = 0.0145  
K = 13: приріст = 0.0187  
K = 14: приріст = -0.0309  
K = 15: приріст = 0.0024  
K = 16: приріст = 0.0232  
K = 17: приріст = 0.0156  
K = 18: приріст = 0.0059  
K = 19: приріст = 0.0087  
K = 20: приріст = -0.0055  
  
Обрано оптимальну кількість кластерів: K = 7
```

Рис. 4.7–4.8 Визначення оптимальної кількості кластерів за значенням Silhouette Score

Візуалізація у площині двох головних компонент (PCA), яка зображена на рис. 4.9, підтверджує коректність розбиття: хмари точок компактні, межі між більшістю кластерів виразні, перетинів небагато. Кожен кластер має власний центр ваги, що вказує на наявність природної структури даних, а не випадкового групування. Така картина дозволяє інтерпретувати сегменти як різні сценарії організації навчального процесу.



Розподіл спостережень між кластерами є нерівномірним, що теж очікувано для освітніх даних. За результатами зображеними на рис. 4.10 можемо бачити що найбільший кластер (№2) містить 3497 записів (майже третина вибірки). Далі – кластери 1 (3038), 4 (2543), 0 (2489); найменші – 3 (515), 5 (503) та 6 (514). Отже, маємо кілька масових сценаріїв і кілька відносно малих, потенційно специфічних сегментів. Масові кластери доцільно розглядати як еталонні патерни, тоді як малі – як кандидати на окрему увагу (ризик/ніші).

Профілі центроїдів уточнюють зміст сегментів. У найбільшому кластері (№2) підвищені середні значення для дисциплін “Безпека комп’ютерних мереж” та “Комп’ютерна електроніка”, що може вказувати на їх стабільну успішність або на специфіку груп, які їх проходять. Кластер 1 вирізняється високим внеском “Бази даних та інформаційні системи” і ознаками підвищеної академічної насиченості (нижчі середні за кредитами), що корелює з інтенсивнішими програмами. Менші кластери мають виразні ніші: у №5 домінує “Криптографія та захист інформації”, у №6 – “Проектування

програмного забезпечення”. Такі профілі зручні для планування адресних дій: підтримки груп ризику, локальних змін навантаження, масштабування кращих практик.

```
Розподіл записів по кластерах:
Кластер
0 2489
1 3038
2 3497
3 515
4 2543
5 583
6 514
Name: count, dtype: int64

Центроїди кластерів (повністю):
metric_value  year  quarter  credits  groupname_student  course_name_Англійська мова для ІТ  course_name_Бази даних та інформаційні системи  course_name_Безпека комп'ютерних
0  0.019528 -0.004375  0.003194  0.784070  -0.004538  -0.488905  -0.199627
1 -0.018638  0.005654  0.010257 -1.538874  0.003810  1.619114  -0.199627
2  0.000323 -0.003183 -0.000476  0.420713  -0.004282  -0.488905  0.548131
3  0.007394  0.012796 -0.005444 -0.484053  0.000135  -0.488905  -0.199627
4 -0.000430 -0.008159 -0.014168  0.784070  0.002417  -0.488905  -0.199627
5  0.003686  0.027344 -0.009559 -0.484053  -0.004502  -0.488905  -0.199627
6  0.004516  0.010210  0.012055 -0.484053  0.020899  -0.488905  -0.199627
```

Рис. 4.10 Представлення центрів кластерів

Узагальнюючи, кластерний аналіз виявив приховану сегментацію даних і дав керовані портрети навчальних сценаріїв: від поширених, стабільних конфігурацій до невеликих спеціалізованих підгруп. Ці результати узгоджуються з якісною візуальною картиною (РСА) і кількісним критерієм (силует) та можуть слугувати основою для подальших управлінських рішень – від налаштування розкладу і навантаження до цільової методичної підтримки.

## ВИСНОВКИ

Одержані у роботі результати підтвердили доцільність інтеграції двох взаємодоповнювальних підходів: багатовимірної аналітики (OLAP) і методів інтелектуального аналізу даних (Data Mining) для задач оцінювання якості навчального процесу. Побудоване сховище даних за схемою «зірка» (QualityFact + DateDim, CourseDim, GroupDim, MetricDim), розгорнутий куб SSAS і налаштовані потоки SSIS забезпечили відтворюваний контур збору, узгодження й аналізу даних. Використання синтетичного набору (~15 тис. записів), сформованого за погодженими правилами, дозволило без ризиків для персональних даних продемонструвати повний цикл: від підготовки до багатовимірних зрізів і побудови моделей.

OLAP дав основу для першої лінії діагностики: через зрізи за часом, курсами й групами були отримані внесені у куб KPI та звіти, які фіксують поточний стан і динаміку показників. Два ключові KPI(узагальнена якість курсу та рівень задоволеності) показали, що у 2023 році обидва індикатори перебувають у середній зоні ( $\approx 6 - 7,5$  за 10-бальною шкалою), причому якість має слабо позитивну динаміку ( $\approx +0,01$  р/р), а задоволеність незначно знижується ( $\approx -0,03$  р/р). Для тлумачення змін використано сталу шкалу та однакові правила згортання; пороги ухвалення управлінських дій визначалися як інтервали відхилень від цільових значень, що забезпечує зіставність між роками й семестрами. Семестрові й курсові звіти SSRS дозволили локалізувати курси/періоди із систематичними відхиленнями та задати конкретні запити для подальшого дослідження.

Методи Data Mining перетворили описові спостереження на формалізовані перевірки. Одноатрибутний класифікатор 1–Rule, побудований для категорії якості (Low/Medium/High), продемонстрував типовий для збалансованих сукупностей рівень точності трохи вище 50%: жоден окремий атрибут (CourseName, GroupName, Year, Month) не є достатнім провісником класу. Як базову лінію слід вважати мажоритарний або випадковий

класифікатор ( $\approx 33\%$  для трьох класів або фактичну частку найбільшого класу), тож 1–Rule її перевищує, однак показує обмежену роздільну здатність одиничних ознак. Наївний Байєс, що враховує спільні розподіли ознак, опрацював 13 099 записів і досяг загальної точності 56,7%. Оцінювання проводилося на відкладеній вибірці, додатково аналізувалися матриця помилок і розподіл ймовірностей класів: найбільшу частку становить клас Medium ( $\approx 61,61\%$ ), далі Low ( $\approx 26,18\%$ ) та High ( $\approx 12,19\%$ ), що підтверджує «центральність» вибірки і пояснює складність відокремлення класів за окремими атрибутами.

Пошук асоціативних правил (Apriori) виявив закономірності зі значеннями підтримки та довіри вище робочих порогів і підсиленням  $lift > 1$ , що поєднують типові контексти курсів, груп і періодів з очікуваним рівнем якості. Для уникнення випадкових поєднань застосовувалася мінімальна підтримка, а для оцінки міцності – показник conviction; результати згорталися до неперекривних шаблонів для зручності інтерпретації у звітності. Кластеризація (K–Means із добором k за індексом силуета) виокремила сім стійких сегментів із нерівномірним наповненням; локальний максимум індексу силуета спостерігається саме за  $k=7$ , а візуалізація у просторі головних компонент (PCA) підтверджує відокремленість великих “типових” кластерів від нішевих підгруп. Виявлені правила та профілі кластерів звірено з OLAP–зрізами та KPI, що забезпечило узгодженість статистичних висновків із звітними показниками і дозволило локалізувати відхилення до рівня курсів, груп і періодів.

Практична цінність підходу полягає у побудові керованого циклу поліпшень: OLAP дає наочну картину проблемних зон, а моделі Data Mining уточнюють, які поєднання умов систематично ведуть до просідання або зростання показників. Таким чином, управлінські дії можуть плануватися не на вмання, а на основі відтворюваних критеріїв (пороги KPI, довіра/підсилення для правил, належність до кластерів), що безпосередньо інтегруються у звітні форми та панелі моніторингу.

Дослідження має притаманні для прототипів обмеження: характер і обсяг даних, вибір порогів та спрощені моделі можуть впливати на точність і узагальнюваність висновків. Отримані результати варто трактувати як робочу основу, яку слід перевіряти і доповнювати на ширших вибірках та в інших навчальних контекстах; доцільно провести перехресну перевірку для класифікаційних моделей, розглянути вагові схеми для незбалансованих класів та перевірити чутливість кластеризації до масштабування ознак. Є простір для посилення рішення: розширити джерела та деталізацію моделі, підсилити валідацію результатів і щільніше інтегрувати знахідки у регулярну звітність і панелі моніторингу для оперативного використання.

Сукупно це дало змістовні відповіді на дослідницькі питання і підтвердило, що мету досягнуто: показано, як поєднання описового й прогностичного аналізу переводить контроль якості у керований цикл “виявити – пояснити – втрутитися”. Підхід є технологічно реалістичним для впровадження і масштабування, а окреслені обмеження – прозорими та такими, що можуть бути зняті подальшою валідацією на реальних даних і розширенням методичного інструментарію.

## ДЖЕРЕЛА

1. Воробйова О., Дебич М., Луговий В., Оржель О., Слюсаренко О., Таланова Ж., Трима К.; за ред. В. Лугового, Ж. Таланової. Механізми оцінювання якості вищої освіти в умовах євроінтеграції. Київ: Ін-т вищої освіти НАПН України, 2020. 220 с. Доступно: <http://repository.ldufk.edu.ua/handle/34606048/32198> (дата звернення: 02.09.2025).
2. Стандарти і рекомендації щодо забезпечення якості в Європейському просторі вищої освіти (ESG). Київ: ТОВ «ЦС», 2015. 32 с. Режим доступу: [https://www.britishcouncil.org.ua/sites/default/files/standards-and-guidelines\\_for\\_qa\\_in\\_the\\_ehea\\_2015.pdf](https://www.britishcouncil.org.ua/sites/default/files/standards-and-guidelines_for_qa_in_the_ehea_2015.pdf) (дата звернення: 02.09.2025).
3. Merritt D. J. Bias, the Brain, and Student Evaluations of Teaching. 2012. 44 с. Режим доступу: [https://www.dickinson.edu/download/downloads/id/6296/bias\\_the\\_brain\\_and\\_student\\_evaluations\\_of\\_teaching.pdf](https://www.dickinson.edu/download/downloads/id/6296/bias_the_brain_and_student_evaluations_of_teaching.pdf) (дата звернення: 02.09.2025).
4. Open edX. Aspects: документація. 2024 – 2025. Режим доступу: <https://docs.openedx.org/projects/openedx-aspects/en/latest/index.html> (дата звернення: 05.09.2025).
5. Blackboard (Anthology). Analytics for Learn: довідка адміністратора (SaaS). Режим доступу: [https://help.blackboard.com/Learn/Administrator/Hosting/Tools\\_Management/Analytics\\_for\\_Learn](https://help.blackboard.com/Learn/Administrator/Hosting/Tools_Management/Analytics_for_Learn) (дата звернення: 05.09.2025).
6. Ohio University. Understanding Analytics in Canvas. 2024. Режим доступу: <https://help.ohio.edu/TDClient/30/Portal/KB/PrintArticle?ID=1068> (дата звернення: 05.09.2025).
7. Moodle. Using analytics. Режим доступу: [https://docs.moodle.org/en/Using\\_analytics](https://docs.moodle.org/en/Using_analytics) (дата звернення: 05.09.2025).
8. Jankowski N. Closing the Loop: Using Assessment Results to Enhance Student Learning. National Institute for Learning Outcomes Assessment

- (NILOA), 2019. Режим доступу: <https://www.sfasu.edu/sites/default/files/2019-08/closing-the-loop-natasha-jankowski-niloa.pdf> (дата звернення: 05.09.2025).
9. Бардус І. О., Лазарєв М. М., Ніценко А. В. Бази даних: у схемах. Київ: 2017. Режим доступу: [https://duikt.edu.ua/uploads/1\\_1800\\_32106091.pdf](https://duikt.edu.ua/uploads/1_1800_32106091.pdf) (дата звернення: 15.09.2025).
10. Тютюник О. О., Бринза Н. О., Удовенко С. Г. Проектування інформаційних систем: методичні рекомендації. Харків: ХНЕУ, 2024. Режим доступу: <https://repository.hneu.edu.ua/bitstream/123456789/32513/1/2024-Тютюник%20О%20О,%20Бринза%20Н%20О,%20Удовенко%20С%20Г.pdf> (дата звернення: 16.09.2025).
11. Дудзяний І. М. Об'єктно-орієнтоване моделювання програмних систем. Львів: Видавничий центр ЛНУ імені Івана Франка, 2007. 108 с. Режим доступу: [https://fis.tntu.edu.ua/data/elibrary/3/ooop\\_uml.pdf](https://fis.tntu.edu.ua/data/elibrary/3/ooop_uml.pdf) (дата звернення: 16.09.2025).
12. Sparx Systems. UML Use Case Diagram: Actors and relationships. Режим доступу: <https://www.sparxsystems.eu/languages/uml/diagrams/usecasediagram/> (дата звернення: 19.09.2025).
13. IBM Developer. Explore the UML sequence diagram. 2023. Режим доступу: <https://developer.ibm.com/articles/the-sequence-diagram/> (дата звернення: 21.09.2025).
14. Дудзяний І. М. Об'єктно-орієнтоване моделювання програмних систем. Львів: Видавничий центр ЛНУ імені Івана Франка, 2007. 108 с. Режим доступу: [https://fis.tntu.edu.ua/data/elibrary/3/ooop\\_uml.pdf](https://fis.tntu.edu.ua/data/elibrary/3/ooop_uml.pdf) (дата звернення: 23.09.2025).
15. Visual Paradigm. UML Class Diagram Tutorial. Режим доступу: <https://www.visual-paradigm.com/guide/uml-unified-modeling-language/uml-class-diagram-tutorial/> (дата звернення: 23.09.2025).
16. Chaudhuri S., Dayal U., Narasayya V. An overview of business intelligence technology. Communications of the ACM, 2011, 54(8), 88–98. Режим

- доступу: <https://dl.acm.org/doi/10.1145/1978542.1978562> (дата звернення: 23.09.2025).
17. Kimball R., Caserta J. The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data. Indianapolis: Wiley, 2004. Режим доступу: <https://sisis.rz.htw-berlin.de/inh2006/1235036.pdf> (дата звернення: 23.09.2025)
  18. Han J., Kamber M., Pei J. Data Mining: Concepts and Techniques. 3rd ed. Waltham, MA: Morgan Kaufmann, 2011. Режим доступу: [https://ia800702.us.archive.org/7/items/datamining\\_201811/DS-book%20u5.pdf](https://ia800702.us.archive.org/7/items/datamining_201811/DS-book%20u5.pdf) (дата звернення: 23.09.2025).
  19. IBM. Deployment diagrams in UML modeling. Режим доступу: <https://www.ibm.com/docs/en/rational-soft-arch/9.7.0?topic=diagrams-deployment> (дата звернення: 28.09.2025).
  20. Oracle. Data Warehousing Guide: Data Warehousing Logical Design (розділ про star schema, conformed dimensions). Режим доступу: <https://docs.oracle.com/en/database/oracle/oracle-database/19/dwhsg/data-warehouse-logical-design.html> (дата звернення: 01.10.2025).
  21. Microsoft. SQL Server Analysis Services overview. Режим доступу: <https://learn.microsoft.com/en-us/analysis-services/ssas-overview?view=sql-analysis-services-2025> (дата звернення: 03.10.2025).
  22. MSSQLTips. SQL Server Integration Services SSIS 2016 Tutorial. Режим доступу: <https://www.mssqltips.com/tutorial/sql-server-integration-services-ssis-2016-tutorial/> (дата звернення: 06.10.2025).
  23. QAA. UK Quality Code, Advice and Guidance: Monitoring and Evaluation. Режим доступу: <https://www.qaa.ac.uk/docs/qaa/quality-code/advice-and-guidance-monitoring-and-evaluation.pdf> (дата звернення: 10.10.2025).
  24. Data Mining – Concepts and Applications. Запорізький національний університет, 2022. Режим доступу: <https://files.znu.edu.ua/files/Bibliobooks/Inshi70/0051095.pdf> (дата звернення: 13.10.2025).

25. Rish I. An empirical study of the naive Bayes classifier. T. J. Watson Research Center, IBM, 2001. Режим доступа: <https://faculty.cc.gatech.edu/~isbell/reading/papers/Rish.pdf> (дата звернення: 15.10.2025).
26. Hahsler M. Measures for association rules. Режим доступа: <https://mhahsler.github.io/arules/docs/measures.pdf> (дата звернення: 18.10.2025).
27. Tan P.-N., Steinbach M., Kumar V. Cluster Analysis: Basic Concepts and Algorithms. In: Introduction to Data Mining. Minneapolis: University of Minnesota, 2006. Режим доступа: <https://www-users.cse.umn.edu/~kumar/dmbook/ch8.pdf> (дата звернення: 21.10.2025).

## ДОДАТКИ

### ДОДАТОК А

#### Обчислення KPI (Quality, Satisfaction, Trend)

```
# Обчислення KPI
import pyodbc, pandas as pd

def load_and_compute(conn_str: str):
    with pyodbc.connect(conn_str) as conn:
        df = pd.read_sql(SQL_KPI, conn)

    df["scaled"] = df["metric_value"] * 10 # нормування до 1–10

    def kpi(year, student_only=False):
        sub = df[df["year"] == year]
        if student_only:
            sub = sub[sub["target_group"] == "student"]
        return sub["scaled"].mean()
    k = {
        "Quality": {"value": kpi(2023), "goal": 8, "trend": kpi(2023) –
kpi(2022)},
        "Satisfaction": {"value": kpi(2023, True), "goal": 7, "trend": kpi(2023,
True) – kpi(2022, True)},
    }
    return k
```

#### KPI “Quality” (Value / Goal)

Value:

AVG([Date Dim].[Id Date].&[start]:[Date Dim].[Id Date].&[end],  
[Measures].[ScaledMetricValue])

Goal:

8

#### KPI “Satisfaction” (Value / Goal)

Value:

AVG([Date Dim].[Id Date].&[start]:[Date Dim].[Id Date].&[end],  
([Measures].[ScaledMetricValue], [Metric Dim].[Target Group].&[student]))

Goal:

7

**SQL–запит створення семестрового звіту:**

```

SELECT
    c.course_name AS CourseName,
    d.quarter AS Quarter,
    AVG(q.metric_value) AS AvgMetricValue
FROM
    QualityFact q
JOIN
    CourseDim c
    ON q.id_course = c.id_course
JOIN
    DateDim d
    ON q.id_date = d.id_date
— Фільтруємо тільки 2023 рік
WHERE
    d.year = 2023
GROUP BY
    c.course_name,
    d.quarter
ORDER BY
    d.quarter,
    AvgMetricValue DESC;

```

**SQL–запит створення звіту за курсами та метриками**

```

SELECT
    c.course_name AS CourseName,
    c.credits AS Credits,
    c.[level] AS Level,
    c.lecturer AS Lecturer,
    AVG(q.metric_value) AS AvgMetricValue
FROM
    QualityFact q
JOIN
    CourseDim c
    ON q.id_course = c.id_course
JOIN
    DateDim d
    ON q.id_date = d.id_date
WHERE
    d.year = 2023
GROUP BY

```

```
c.course_name,  
c.credits,  
c.[level],  
c.lecturer  
ORDER BY  
AvgMetricValue DESC;
```