

**НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ БІОРЕСУРСІВ І  
ПРИРОДОКОРИСТУВАННЯ УКРАЇНИ**

**ФАКУЛЬТЕТ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ**

**ДОПУСКАЄТЬСЯ ДО ЗАХИСТУ**

Завідувач кафедри  
економічної кібернетики

к.е.н., доц. \_\_\_\_\_ Володимир ХАРЧЕНКО  
(підпис)

“ \_\_\_\_ ” \_\_\_\_\_ 2025 р.

**БАКАЛАВРСЬКА КВАЛІФІКАЦІЙНА РОБОТА**

на тему

**«Прогнозування успішності студентів на основі  
освітньої аналітики»**

Спеціальність 051 – «Економіка»  
Освітня програма «Цифрова економіка»  
Програма підготовки освітньо-професійна

**Гарант освітньої програми**

К.е.н., доцент \_\_\_\_\_ Володимир ХАРЧЕНКО

**Керівник бакалаврської**

кваліфікаційної роботи, к.е.н., доц. \_\_\_\_\_ Наталія КЛИМЕНКО

**Виконала:**

\_\_\_\_\_ Аріна ДУДНІЧЕНКО

**КИЇВ – 2025**

**НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ БІОРЕСУРСІВ І  
ПРИРОДОКОРИСТУВАННЯ УКРАЇНИ  
ФАКУЛЬТЕТ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ**

Завідувач кафедри  
економічної кібернетики  
к.е.н., доц. Володимир ХАРЧЕНКО  
“20” грудня 2024 р.

**ЗАВДАННЯ**  
**до виконання бакалаврської кваліфікаційної роботи**  
**студентки Дудніченко Аріни Олександрівни**

Спеціальність 051            «Економіка»  
Освітня програма            «Цифрова економіка»  
Програма підготовки            освітньо-професійна

1. Тема роботи: «Прогнозування успішності студентів на основі освітньої аналітики» затверджена наказом ректора НУБіП України від 16.12.2024р. № 2252 «С».

2. Термін подання завершеної роботи на кафедру: 06.06.2025 р.

3. Вихідні дані до бакалаврської кваліфікаційної роботи: відкриті дані з аналітичної бази EDEBO, Інформаційної системи «Деканат» та LMS Moodle НУБіП України, матеріали періодичних видань, науково-практичних конференцій.

4. Перелік графічного матеріалу: 11 таблиці, 21 рисунки

5. Зміст пояснювальної записки (перелік питань, що підлягають дослідженню в роботі):

- Теоретичне обґрунтування застосування освітньої аналітики для прогнозування успішності студентів.
- Аналіз платформи Moodle як джерела навчальних даних.
- Огляд інструментів Excel та R для аналізу й прогнозування успішності.
- Розробка економіко-математичної моделі успішності студентів з урахуванням академічної заборгованості.
- Застосування методів машинного навчання для класифікації та прогнозування рівня успішності.
- Сегментація студентів за поведінковими та академічними характеристиками.
- Візуалізація освітніх даних і розробка інтерпретованих моделей прогнозування.

6. Дата отримання завдання 20.12.2024 р.

**Керівник бакалаврської кваліфікаційної роботи, к.е.н., доц.**

\_\_\_\_\_ Наталія КЛИМЕНКО

**Завдання прийняла до виконання** \_\_\_\_\_ Аріна ДУДНІЧЕНКО

## **ЗМІСТ**

<b>ВСТУП.....</b>	<b>5</b>
<b>РОЗДІЛ 1. ТЕОРЕТИЧНІ ОСНОВИ АНАЛІЗУ ТА ПРОГНОЗУВАННЯ УСПІШНОСТІ СТУДЕНТІВ ЗАСОБАМИ ОСВІТНЬОЇ АНАЛІТИКИ ..</b>	<b>8</b>
1.1 Освітня аналітика як інструмент дослідження навчального процесу .....	8
1.2 Характеристика платформи Moodle як джерела освітніх даних .....	13
1.3. Методи аналізу та прогнозування успішності студентів: огляд інструментів Excel та R .....	17
<b>РОЗДІЛ 2. ПОБУДОВА ЕКОНОМІКО-МАТЕМАТИЧНОЇ МОДЕЛІ УСПІШНОСТІ СТУДЕНТІВ ТА ЇЇ АНАЛІЗ .....</b>	<b>22</b>
2.1 Визначення складу моделі та кореляційний аналіз змінних.....	22
2.2 Аналіз успішності студентів з урахуванням академічної заборгованості.....	29
2.3 Динаміка успішності та рівень боргів студентів групи вступу 2021 року за шість семестрів .....	36
<b>РОЗДІЛ 3. АНАЛІЗ ТА ПРОГНОЗУВАННЯ УСПІШНОСТІ СТУДЕНТІВ НА ОСНОВІ ДАНИХ LMS І МЕТОДІВ МАШИННОГО НАВЧАННЯ В R.....</b>	<b>40</b>
3.1 Аналіз факторів навчального процесу, що впливають на успішність студентів.....	40
3.2 Сегментація студентів за профілями активності та успішності .....	48
3.3 Візуалізація та аналіз розподілу академічної успішності за допомогою боксплотів .....	53
3.4 Побудова та інтерпретація моделей прогнозування академічної успішності .....	56
<b>ВИСНОВКИ .....</b>	<b>64</b>
<b>СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ .....</b>	<b>68</b>
<b>ДОДАТКИ.....</b>	<b>73</b>

## ВСТУП

У сучасній системі вищої освіти дедалі важливішим стає використання цифрових інструментів для моніторингу та підвищення якості навчального процесу. Значне поширення систем управління навчанням (Learning Management Systems, LMS) створило нові можливості для збору, зберігання та аналізу великих обсягів освітніх даних. Інформація, яка накопичується в LMS, охоплює ключові аспекти навчальної активності студентів - від відвідуваності до оцінок, що дає змогу здійснювати глибокий аналіз академічної успішності та виявляти поведінкові патерни.

### **Актуальність роботи:**

Упровадження цифрових інструментів у вищій освіті створює нові можливості для об'єктивного оцінювання навчального процесу. Аналіз даних з LMS дозволяє своєчасно виявляти проблеми, визначати сильні та слабкі сторони навчання, а також будувати прогностичні моделі, які можуть стати основою для прийняття ефективних управлінських рішень. Результати цієї роботи мають практичну цінність для адміністрацій закладів освіти, викладачів та фахівців з освітньої аналітики.

Предметом даного дослідження є аналіз та прогнозування академічної успішності студентів на основі освітньої аналітики, сформованої з даних LMS. Об'єкт дослідження охоплює успішність студентів однієї спеціальності за період з 2021 по 2023 рік, а також ширший набір даних за перший семестр 2024-2025 навчального року, що включає всі спеціальності та курси. Такий підхід дозволяє як детально вивчити динаміку змін у невеликій вибірці, так і провести моделювання та узагальнення висновків на основі великої репрезентативної сукупності.

У першому розділі узагальнено роль освітньої аналітики як інструменту дослідження навчального процесу. Також запропонована типова структура системи освітньої аналітики та встановлені канали надходження даних для аналізу освітніх траєкторій здобувачів освіти

У другому розділі роботи реалізовано розгорнутий аналіз у середовищі Microsoft Excel, де виконано систематизацію результатів студентів у трьох академічних групах, що вступили у 2021, 2022 та 2023 роках. Досліджено середній бал, кількість студентів з академічною заборгованістю, динаміку змін показників упродовж трьох років та здійснено нормалізацію даних для забезпечення їх співставності.

У третьому розділі проведено поглиблений аналіз та прогнозування успішності студентів із використанням мови програмування R. Застосовано статистичні методи, такі як регресійний аналіз, кластеризація методом k-середніх, побудова моделей прогнозу на основі поведінкових факторів (зокрема відвідуваності). У роботі використано великий набір даних за перший семестр 2024-2025 року, що дозволяє сформувати більш об'єктивну картину та надати обґрунтовані рекомендації щодо покращення навчального процесу.

**Предмет дослідження:**

Сукупність теоретичних, методичних та практичних підходів щодо аналізу та прогнозування академічної успішності студентів на основі даних освітньої аналітики, отриманих з LMS.

**Об'єкт дослідження:**

Особливості академічної успішності студентів в розрізі семестрів та академічних груп

**Мета дослідження:**

Виявлення факторів, що впливають на академічні результати студентів, побудова аналітичних моделей і формування прогнозів успішності для підтримки якості освітнього процесу.

**У зв'язку з цим у роботі пропонуються та вирішуються наступні завдання:**

1. Узагальнення поняття освітньої аналітики та встановлення джерел даних про освітній процес

2. Формування вибірки даних про успішність студентів та аналіз академічної успішності студентів за допомогою функцій Excel.
3. Аналіз динаміки показників, що характеризують успішність студентів: середній бал, частка боржників, відмінності між роками.
4. Застосування аналітичних методів у середовищі RStudio для дослідження зв'язків між факторами та побудови прогностичних моделей.
5. Кластеризація студентів за академічними та поведінковими характеристиками та надання практичних рекомендацій щодо покращення успішності на основі результатів аналізу.

**Методи дослідження:** Теоретичною та методологічною основою даної роботи є наукові праці вітчизняних та зарубіжних вчених у галузі освітньої аналітики та education data mining, Інформаційну базу дослідження складають безпосередні дані з аналітичної бази EDEBO, Інформаційної системи «Деканат» та LMS moodle, матеріали журналів, науково-практичних конференцій, нормативні та довідкові матеріали.

Для реалізації поставлених у роботі завдань були використані такі методи та методики: монографічний – при вивченні та аналізі наукової, науково-технічної літератури; економетричні методи – при аналізі та прогнозуванні успішності студентів на основі даних LMS, методи машинного навчання в R – при розробці практичних рекомендацій.

# РОЗДІЛ 1. ТЕОРЕТИЧНІ ОСНОВИ АНАЛІЗУ ТА ПРОГНОЗУВАННЯ УСПІШНОСТІ СТУДЕНТІВ ЗАСОБАМИ ОСВІТНЬОЇ АНАЛІТИКИ

## 1.1 Освітня аналітика як інструмент дослідження навчального процесу

Освітня аналітика є відносно новим, але стрімко зростаючим напрямом у сфері освіти, який поєднує в собі аналіз великих масивів даних, педагогіку та освітні технології. Вона набуває дедалі більшого значення в навчальних закладах, платформах дистанційного навчання та серед дослідників освіти. У найзагальнішому вигляді освітня аналітика - це процес збору, аналізу, інтерпретації та використання даних про навчальний процес з метою покращення навчання й прийняття обґрунтованих рішень.

Типова структура будь-якої системи освітньої аналітики включає чотири ключові елементи: дані, аналіз, звітність та дії. По-перше, дані є вихідним ресурсом: у контексті освіти це інформація про студентів, навчальне середовище, взаємодії та результати. По-друге, дані обробляються й аналізуються за допомогою математичних і статистичних методів - у такий спосіб з'являється змістовна аналітична інформація. Третім етапом є формування звітів: результати аналізу візуалізуються у вигляді таблиць, графіків і звітів, що дозволяє виявити закономірності у поведінці студентів. Нарешті, на основі отриманих висновків ухвалюються управлінські або педагогічні рішення, тобто реалізується перехід від знання до дії [1].

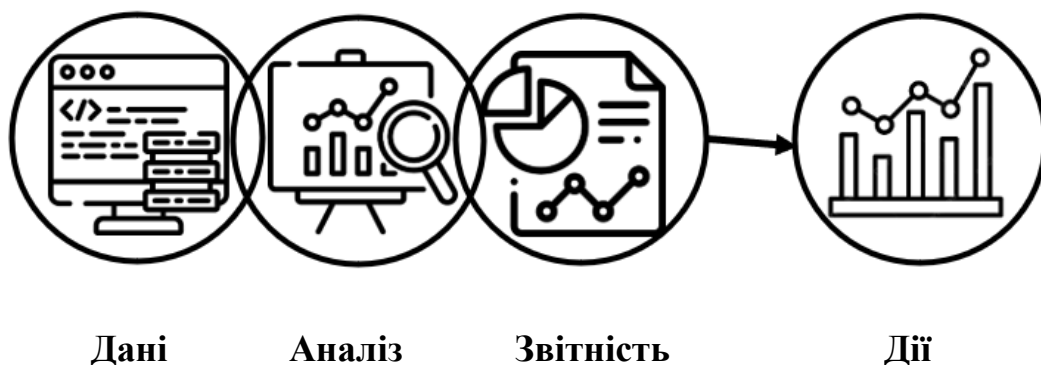


Рисунок 1.1.1 - Основні елементи аналітики навчання

Однак аналітика в освіті - це не лише про алгоритми чи звіти. Важливо також розуміти, що саме ми вимірюємо, чому це важливо та для кого це робиться. Для усвідомленого використання аналітики пропонується концепція аналітичної триади ЕПО - епістемологія, педагогіка та оцінювання. Вона допомагає осмислити, які знання ми вважаємо значущими, які методи використовуємо для їх оцінювання та яку цінність ці дані несуть для студентів, викладачів чи адміністрації.

Зокрема, епістемологія звертає увагу на те, яку природу мають знання, які ми оцінюємо, та які докази вважаються обґрунтованими. Педагогіка дає змогу зрозуміти, чому саме ці знання є важливими у певному освітньому контексті, і хто є кінцевим користувачем аналітики. Оцінювання, у свою чергу, фокусується на тому, в яких умовах, коли та як саме здійснюється фіксація навчальних результатів, і як це впливає на достовірність висновків.

Таким чином, поєднання технологічного підходу до обробки даних із критичним педагогічним осмисленням дозволяє будувати етично обґрунтовану та освітньо ефективну систему освітньої аналітики, орієнтовану не лише на контроль, а й на підтримку студента в освітньому процесі [2].

Для реалізації такої системи необхідне відповідне технологічне середовище, яке забезпечує збирання, зберігання, обробку та інтерпретацію навчальних даних у реальному часі. Саме таку роль виконують системи управління навчанням, які сьогодні є невід'ємною частиною цифрової освіти.

Системи управління навчанням (Learning Management Systems, LMS) відіграють ключову роль у сучасному освітньому середовищі як платформи для організації, реалізації й моніторингу навчального процесу. LMS не лише підтримують навчальну діяльність, а й виступають як джерело освітніх даних для аналізу. Вони забезпечують збір інформації про поведінку користувачів, активність, прогрес у навчанні, взаємодію з матеріалами та іншими учасниками курсу, що робить їх фундаментальними елементами систем освітньої аналітики.

Сучасні LMS трансформуються у напрямку відкритості, персоналізації та соціалізації. Відкритість проявляється у двох вимірах: зовнішньому (можливість

ділитися навчальними ресурсами через інтерфейси або API) та внутрішньому (інтеграція зовнішніх ресурсів і застосунків у LMS). Завдяки підтримці стандартів (наприклад, SCORM, IMS), платформи на кшталт Moodle або Sakai можуть працювати з відкритими освітніми ресурсами, сторонніми сервісами чи освітніми порталами.

Крім того, LMS поступово переходять від орієнтації на викладача до персоналізованого навчального середовища. Студенти отримують змогу керувати власними ресурсами, налаштовувати цілі навчання, ініціювати активності (наприклад, форуми чи групи), а також публікувати власний контент - блоги, відгуки, нотатки. Це сприяє розвитку автономного та самостійного навчання.

Не менш важливою є й соціальна складова LMS. Сучасні платформи підтримують створення спільнот, обмін досвідом, групову роботу, профілі користувачів, форуми між курсами, інструменти для рецензування контенту тощо. Завдяки інтеграції із соціальними медіа та розвиненим функціям комунікації, LMS сприяють формуванню навчальних мереж і залученню користувачів до взаємодії, що є важливим елементом побудови ефективного цифрового освітнього середовища.

Таким чином, LMS є не лише платформою для розміщення навчального контенту, але й багатофункціональним середовищем, що підтримує аналітику, персоналізацію, відкритість та соціальну взаємодію - критично важливі компоненти сучасної освіти. [5]

Освітня аналітика є потужним інструментом для дослідження процесів навчання, що дозволяє не лише фіксувати факти, але й розкривати приховані закономірності, взаємозв'язки та тенденції в освітньому середовищі. Її застосування охоплює як стратегічний рівень - планування розвитку освітніх програм чи удосконалення методик навчання, так і тактичний - оперативну підтримку викладачів і студентів у щоденному освітньому процесі.

Одним із ключових аспектів освітньої аналітики є її здатність забезпечувати доказову базу для прийняття рішень. Завдяки системному збору

та аналізу даних про навчальну активність (від перегляду матеріалів до проходження тестів), аналітика дозволяє викладачам вчасно виявляти труднощі студентів, прогнозувати академічні ризики та адаптувати навчальний процес до реальних потреб аудиторії. Наприклад, зниження активності студента у курсі може сигналізувати про втрату мотивації або складність матеріалу, що потребує негайної педагогічної реакції.

Крім оперативного моніторингу, аналітика дозволяє проводити ретроспективний аналіз ефективності курсів, порівнювати результати різних навчальних груп або викладацьких стратегій, виявляти сильні й слабкі сторони навчального дизайну. Це особливо важливо у великих освітніх проєктах, де ручний аналіз даних є складним або неможливим.

Окремим напрямом розвитку є предиктивна аналітика - використання історичних даних для прогнозування майбутніх результатів. Такі підходи дозволяють виявити, які саме фактори (наприклад, частота відвідування курсу, участь у форумах, кількість спроб у тестах) найбільше впливають на успішність студентів, і будувати математичні моделі, що прогнозують імовірність завершення курсу або досягнення певного рівня знань. На основі таких моделей можливо реалізувати інтелектуальні системи підтримки навчання або автоматизовані сповіщення, які орієнтовані на запобігання академічній неуспішності.

Не менш важливою функцією освітньої аналітики є забезпечення прозорості та підзвітності в освіті. Адміністрація навчального закладу отримує змогу вивчати динаміку залученості студентів, порівнювати ефективність викладачів, оптимізувати навчальне навантаження та підвищувати якість освітніх послуг на основі об'єктивних даних.

Однак застосування освітньої аналітики також супроводжується низкою етичних та правових викликів. Йдеться насамперед про питання конфіденційності, інформованої згоди на збір та обробку персональних даних, а також уникнення алгоритмічної упередженості. Для вирішення цих питань

важливо формувати політику етичного використання аналітики, яка базується на принципах прозорості, відповідальності та орієнтації на благо студента.

У підсумку, освітня аналітика перетворюється на інструмент освітніх трансформацій, що дозволяє підвищувати якість навчання, робити процеси прийняття рішень обґрунтованішими, а взаємодію між учасниками освітнього процесу - більш усвідомленою та ефективною. Її впровадження потребує не лише технічних ресурсів, а й нової культури роботи з даними в освіті, що передбачає інтеграцію педагогічного мислення, аналітичних навичок і цифрових технологій.

Освітня аналітика, незважаючи на численні переваги і перспективи, стикається із значними викликами та обмеженнями, що зумовлюють необхідність обережного й відповідального підходу до її впровадження та використання. Найважливішими з них є етичні питання, захист персональних даних, а також ризики неправильного інтерпретування даних.

Етичні питання у освітній аналітиці пов'язані з гарантією поваги до приватності студентів і викладачів. Збір, зберігання і аналіз персональних і навчальних даних мають здійснюватися лише з дотриманням принципів конфіденційності та інформованої згоди учасників освітнього процесу. Відсутність чітких правил або невідповідальне ставлення може призвести до порушення прав особи, втрати довіри і навіть юридичних наслідків для навчального закладу. Додатково, важливо уникати дискримінації, що може виникнути через упередження, закладені в алгоритми або в процес вибору даних для аналізу. Необхідно контролювати, щоб моделі не посилювали соціальні нерівності, а сприяли інклюзивності і підтримці всіх студентів.

Захист персональних даних є не менш вагомим викликом. У світі діють різноманітні нормативні акти, що регулюють обробку особистої інформації, зокрема Європейський регламент про захист даних (GDPR), який встановлює суворі вимоги до безпеки та прозорості обробки даних. Освітні організації, що впроваджують аналітичні системи, мають забезпечувати надійний захист інформації - через використання сучасних технологій шифрування, контролю доступу, а також чітких процедур щодо зберігання і видалення даних.

Відсутність таких заходів підвищує ризики витоку даних і можливих зловживань. [10]

Ризики неправильного інтерпретування даних є серйозною проблемою, що може мати негативні наслідки для навчального процесу і кар'єри студентів. Дані можуть бути неповними, неточними або мати приховані упередження, що без належного контексту та кваліфікованого аналізу призведе до хибних висновків. Наприклад, низька активність студента у системі не завжди свідчить про неуспішність - це може бути пов'язано з зовнішніми обставинами чи альтернативними формами навчання. Тому дуже важливо, щоб освітня аналітика базувалася на міждисциплінарному підході, залучала педагогів, аналітиків і психологів для комплексної інтерпретації результатів.

Технічні обмеження також мають значення. Зібрані дані мають бути якісними, повними та актуальними, що часто є складним завданням через технічні несправності, різнорідність систем, недостатній рівень цифрової грамотності користувачів чи обмежені ресурси навчальних закладів. Крім того, аналітичні моделі потребують постійного оновлення і адаптації до змін в освітньому середовищі.

Таким чином, для ефективного впровадження освітньої аналітики необхідно поєднувати технологічні можливості з етичними нормами, високою компетентністю користувачів та відповідними організаційними заходами. Формування чіткої політики захисту даних, підвищення кваліфікації педагогів і аналітиків, а також інтеграція освітньої аналітики у загальну стратегію розвитку навчального закладу дозволять мінімізувати ризики і максимально використати потенціал даних для покращення навчання.

## **1.2 Характеристика платформи Moodle як джерела освітніх даних**

Moodle (Modular Object-Oriented Dynamic Learning Environment, вимовляється «Мудл») - це модульне об'єктно-орієнтоване динамічне навчальне середовище, що належить до класу систем управління навчанням (LMS). У літературі цю платформу також описують як систему управління курсами,

віртуальне навчальне середовище або інструмент для організації комп'ютеризованого, зокрема дистанційного, навчання. Moodle призначена для підтримки освітньої діяльності різного рівня - від шкільної та університетської освіти до професійного розвитку та корпоративного навчання. Платформа надає розгалужений інструментарій для викладачів, студентів і адміністраторів, включаючи засоби створення й проведення курсів, оцінювання, комунікації, управління контентом і моніторингу прогресу.

Однією з ключових переваг Moodle є її відкритий код та безкоштовна модель розповсюдження. Платформа не потребує ліцензійних чи комерційних компонентів, що робить її доступною для широкого кола освітніх закладів, зокрема в умовах обмеженого бюджету. Це дає змогу впроваджувати повноцінну систему управління навчанням без фінансових витрат, з можливістю модифікації програмного коду відповідно до індивідуальних потреб установи.

Moodle є однією з найпопулярніших освітніх платформ у світі. Станом на сьогодні нею користуються понад 400 мільйонів осіб у різних країнах. У 2018 році було зафіксовано, що кількість впроваджень Moodle перевищила сумарну кількість впроваджень усіх інших подібних платформ. У країнах Європейського Союзу дві третини закладів вищої освіти використовують саме Moodle, а в Україні, де застосування платних систем є значно обмеженим, ця платформа фактично не має конкурентів. Вибір Moodle як базової платформи для цифрового навчання в умовах української освітньої системи є цілком виправданим і порівнюється за значенням із вибором англійської мови як основної іноземної мови для вивчення.

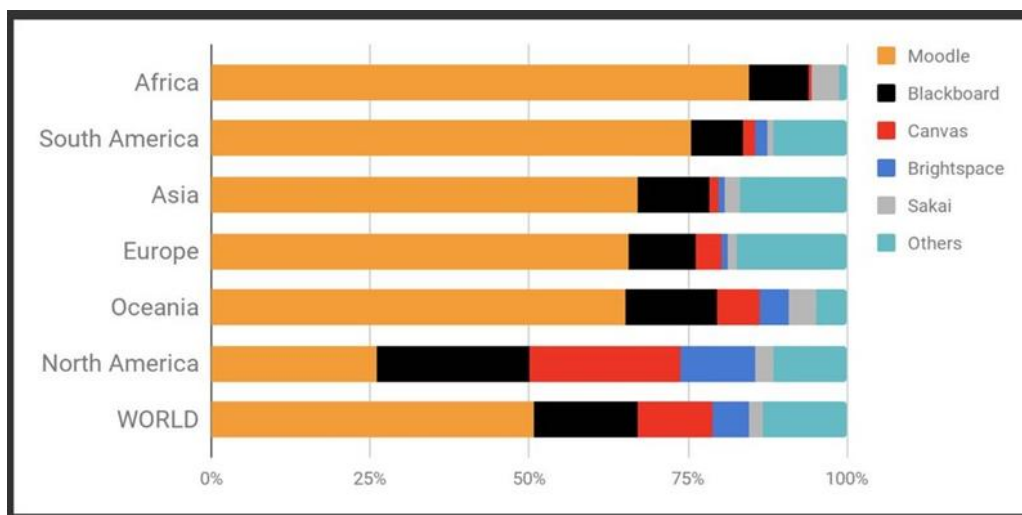


Рисунок 1.2.1 – Статистика додатків в світі за 2018 рік

Крім технічних характеристик, Moodle має потужну підтримку з боку міжнародної спільноти користувачів та розробників. Існують численні ресурси, форуми, посібники та обговорення на офіційному сайті moodle.org, де користувачі можуть отримати фахові поради, інструкції щодо інсталяції та налаштування, а також підтримку в процесі адаптації системи до конкретних умов. Moodle може бути розгорнута на домашньому комп'ютері, у локальній мережі навчального закладу або у глобальній мережі Інтернет, що робить її універсальним рішенням для організації ефективного освітнього процесу. [6]

Таким чином, Moodle виступає не лише як інструмент для розміщення навчального контенту, а як повноцінне освітнє середовище, яке забезпечує гнучкість, масштабованість і відповідність сучасним потребам цифрової освіти.

Однією з ключових переваг Moodle як системи управління навчанням є її багатий функціонал для збору, обробки та експорту освітніх даних, що дозволяє використовувати платформу як джерело для аналітичного дослідження освітнього процесу.

Серед основних функцій, які забезпечують фіксацію навчальної активності користувачів, варто виокремити: тести, завантаження завдань, журнали відвідуваності та оцінювання, а також логування активності. Наприклад, під час проходження тесту Moodle зберігає інформацію про час початку, тривалість виконання, кількість спроб, надані відповіді та отримані бали. Завдання, у свою

чергу, фіксують факт подання студентом відповіді, час завантаження файлу, коментарі викладача та оцінку. Усі ці елементи можуть бути зведені у загальний журнал оцінювання, який підтримує подальший експорт для аналізу.

Moodle дозволяє експортувати різні типи даних, які можуть використовуватись в освітній аналітиці: результати тестів, оцінки, статистику активності, журнали входу та логування дій користувачів. Ці дані є критично важливими для побудови моделей навчальної аналітики, виявлення ризиків академічної неуспішності, аналізу залученості студентів та прогнозування результатів навчання.

Що стосується форматів експорту, Moodle підтримує вивантаження даних у форматах CSV, Excel (XLS/XLSX) та JSON. Ці формати зручні для подальшої обробки в інструментах аналітики, таких як Python, R, Excel Power Query, Power BI або Google Data Studio. У межах даного дослідження дані були отримані з таких сторінок платформи Moodle, як журнал активності та журнал оцінок. [7]

Крім того, Moodle дозволяє адаптувати навчальні курси до індивідуальних потреб студентів завдяки підтримці адаптивного навчання. Це означає, що викладач може налаштувати доступ до певних матеріалів, тестів чи завдань залежно від попередніх результатів студента або виконаних дій у курсі. Такий підхід сприяє персоналізації освітнього процесу, підвищенню мотивації та ефективності навчання.

Ще однією важливою особливістю є можливість інтеграції Moodle з іншими системами та сервісами. Наприклад, через API або плагіни можна підключити сервіси відеоконференцій, бібліотечні каталоги, електронні щоденники, платформи з відкритими освітніми ресурсами тощо. Це розширює функціонал Moodle як центральної освітньої платформи навчального закладу.

З точки зору адміністрування, Moodle забезпечує гнучкі засоби управління користувачами, правами доступу, ролями, групами й курсами. Це дозволяє ефективно організувати освітній процес у великих навчальних установах з багатьма підрозділами, викладачами та студентами.

У результаті, Moodle є не лише зручним середовищем для розміщення навчального контенту, а й потужним інструментом для організації, підтримки й розвитку освітніх програм у цифровому форматі. Її широкі функціональні можливості, гнучкість та доступність роблять цю платформу оптимальним вибором для впровадження систем дистанційного та змішаного навчання в умовах сучасної освіти.

### **1.3. Методи аналізу та прогнозування успішності студентів: огляд інструментів Excel та R**

Останнім часом дедалі більшого поширення набувають аналітичні підходи до вивчення факторів, що впливають на академічну успішність студентів. Одним із таких підходів є використання методів машинного навчання, які, незважаючи на свою складність, можуть бути реалізовані навіть у середовищі Microsoft Excel. Завдяки доступності та інтуїтивному інтерфейсу Excel можна ефективно здійснювати аналіз навчальних даних без потреби у складному програмному забезпеченні.

Основна ідея такого підходу полягає у побудові прогностичних моделей на основі історичних освітніх даних. Це дозволяє виявити ключові змінні, що впливають на результати студентів, та прогнозувати їхню подальшу успішність. Використовуючи стандартні інструменти Excel, зокрема модуль «Аналіз даних» (Data Analysis Toolpak), можна здійснювати регресійний аналіз, оцінювати кореляційні зв'язки та будувати прості моделі машинного навчання.

Перед початком моделювання важливим етапом є підготовка даних, яка включає їх очищення та перетворення. Зокрема, категоріальні змінні, що містять текстові значення, мають бути закодовані у числовому вигляді, оскільки більшість алгоритмів машинного навчання оперують лише числовими даними. Це дозволяє забезпечити коректність розрахунків та інтерпретацію результатів.

У межах аналізу можуть бути розглянуті різноманітні змінні: вік, стать, освіта батьків, соціальні умови, академічні результати, кількість пропусків занять, обсяг часу, витраченого на навчання, тощо. Регресійні моделі в Excel дають змогу визначити, які саме чинники мають статистично значущий вплив на

академічні досягнення. Наприклад, може виявитися, що деякі соціальні або мотиваційні фактори (як-от причини вибору навчального закладу чи рівень підтримки з боку родини) є більш важливими для успішності студентів, ніж очевидні академічні змінні на кшталт кількості годин навчання.

Крім того, Excel дозволяє перевіряти точність моделей через порівняння прогнозованих і фактичних значень. Також можливе введення випадкових змін до вхідних параметрів, що дає змогу оцінити стійкість моделі до варіацій даних - тобто її здатність узагальнювати закономірності і не залежати від шуму чи незначних змін у вхідній інформації. [8]

Загалом, застосування методів машинного навчання в Excel є прикладом того, як навіть базові інструменти можуть бути використані для розв'язання складних освітніх завдань. Такий підхід відкриває нові можливості для педагогічної аналітики, а також сприяє прийняттю обґрунтованих рішень у сфері навчання та підтримки студентів. Водночас, у більш складних і гнучких аналітичних задачах дедалі частіше застосовується мова програмування R - потужне середовище для статистичних обчислень і візуалізації даних, яке широко використовується в наукових дослідженнях, зокрема в галузі освітньої аналітики. Завдяки широкому спектру бібліотек і пакетів R забезпечує можливість реалізації як базових статистичних методів, так і сучасних алгоритмів машинного навчання для глибокого аналізу та прогнозування академічної успішності студентів.

Однією з головних переваг R є його відкритість та гнучкість у роботі з великими обсягами освітніх даних. У процесі аналізу можна використовувати різноманітні джерела інформації, зокрема CSV- та Excel-файли, бази даних тощо. Після імпорту даних здійснюється попередня обробка, яка включає очищення даних від пропусків, нормалізацію значень, перетворення категоріальних змінних у числові (через one-hot encoding або факторизацію), а також створення нових змінних на основі наявних (feature engineering).

У контексті аналізу успішності студентів мова R дозволяє реалізувати:

Описову статистику (з використанням пакетів `dplyr`, `summarytools`), що дозволяє виявити загальні тенденції в освітніх даних;

Візуалізацію зв'язків і розподілів за допомогою `ggplot2`, що сприяє виявленню потенційних залежностей між змінними (наприклад, між кількістю пропусків занять та підсумковими оцінками);

Побудову прогностичних моделей, зокрема регресійних (лінійна, логістична регресія), дерев рішень, випадкових лісів, методів кластеризації та нейронних мереж (пакети `caret`, `randomForest`, `rpart`, `nnet`, `xgboost` тощо);

Оцінку точності моделей за допомогою метрик, таких як середньоквадратична помилка (RMSE), коефіцієнт детермінації ( $R^2$ ), точність, повнота, F-міра тощо.

Прогнозування оцінок студентів часто здійснюється з використанням фінальної оцінки (наприклад, G3) як цільової змінної, а решта параметрів - як предиктори. Аналіз дає змогу виявити, які саме змінні (наприклад, рівень освіти батьків, доступ до інтернету, наявність додаткової підтримки, соціальні фактори) мають найбільший вплив на навчальні досягнення. У процесі аналізу можна використовувати крос-валідацію, що дозволяє уникнути перенавчання моделі та підвищити її узагальнюючу здатність.

Крім прогнозування, R дозволяє також сегментувати студентів за певними ознаками (кластеризація), що є корисним для виявлення груп із подібними освітніми характеристиками - наприклад, студентів, які потребують додаткової підтримки або демонструють високу мотивацію. [9]

Загалом, застосування мови R для аналізу академічної успішності відкриває широкі можливості для глибокого розуміння освітніх процесів. Завдяки інтеграції статистики, програмування та візуалізації R забезпечує потужний інструментарій для прийняття обґрунтованих рішень у сфері освіти.

Незважаючи на значні можливості, як Excel, так і R мають певні обмеження, які необхідно враховувати при виборі інструменту для дослідження академічної успішності.

Excel має беззаперечні переваги у вигляді зручного графічного інтерфейсу та доступності для широкого кола користувачів. Проте його функціональні можливості обмежені в контексті обробки великих обсягів даних, складної логіки моделювання та автоматизації аналізу. Крім того, в Excel важко реалізувати повторювані сценарії аналізу або інтеграцію з іншими джерелами даних без використання додаткового програмування (наприклад, через VBA).

R, навпаки, забезпечує гнучкість, масштабованість та високу точність аналітики, однак потребує спеціальної підготовки. Для ефективної роботи з цією мовою необхідне розуміння принципів програмування, статистики та структури даних. Це створює певний бар'єр для викладачів або освітніх адміністраторів, які не мають достатнього досвіду в аналітичному програмуванні.

Водночас R дозволяє створювати автоматизовані та повторювані аналітичні процеси, зберігати сценарії для повторного використання, створювати інтерактивні звіти (наприклад, за допомогою R Markdown або Shiny), що робить його ідеальним інструментом для системної освітньої аналітики в навчальних закладах.

Інтеграція Excel і R - це ще один перспективний напрям. Наприклад, дані можуть збиратися й попередньо оброблятися в Excel, а потім експортуватися у форматі CSV або XLSX для подальшого аналізу в R. Такий гібридний підхід поєднує зручність Excel із аналітичною потужністю R.

У перспективі, впровадження інструментів аналізу та прогнозування успішності студентів має бути частиною ширшої стратегії освітньої аналітики, яка охоплює:

- створення центрів аналітичної підтримки у ЗВО;
- навчання персоналу основам статистичного аналізу та роботи з даними;
- розробку етичних протоколів з використання освітніх даних;
- використання результатів аналізу для побудови індивідуальних освітніх траєкторій студентів.

Таким чином, поєднання доступних і зрозумілих інструментів, як-от Excel, з більш потужними та гнучкими, такими як R, дозволяє створити надійне

підґрунтя для аналізу та прогнозування успішності студентів. Це не лише підвищує ефективність освітнього процесу, а й сприяє формуванню культури прийняття рішень, що ґрунтуються на даних.

## РОЗДІЛ 2. ПОБУДОВА ЕКОНОМІКО-МАТЕМАТИЧНОЇ МОДЕЛІ УСПІШНОСТІ СТУДЕНТІВ ТА ЇЇ АНАЛІЗ

### 2.1 Визначення складу моделі та кореляційний аналіз змінних

Успішність студентів є складною та багатогранною характеристикою, що залежить від численних змінних, які умовно можна поділити на контрольовані, неконтрольовані, випадкові та зовнішні фактори. До контрольованих змінних належать ті, що перебувають у межах впливу навчального процесу і можуть бути змінені або скориговані для покращення результатів студентів. Це, зокрема, відвідуваність лекцій і семінарів, час, який студенти витрачають на виконання завдань у системі управління навчанням (LMS), а також кількість додаткових консультацій, отриманих від викладачів. Неконтрольовані змінні - це фактори, які безпосередньо не піддаються управлінню під час навчання, але мають значний вплив на академічні результати. Серед них варто виділити психологічний стан студента, його особисті обставини, такі як робота чи сімейні проблеми, що можуть відволікати або підтримувати в навчанні.

Випадкові змінні характеризуються непередбачуваністю впливу і не піддаються жодному контролю. До них можна віднести рівень складності екзаменаційних завдань, а також зовнішні обставини, що впливають на мотивацію студентів, наприклад, карантинні обмеження, воєнні дії або зміни у законодавстві щодо освіти. Зовнішні змінні - це фактори, які мають вплив на навчальний процес з боку, але їх неможливо врахувати як сталі величини в моделі. Це, зокрема, політика університету щодо організації навчання, доступність навчальних матеріалів у LMS, а також зміни у викладацькому складі та методах викладання.

Враховуючи мету роботи - аналіз і прогнозування успішності студентів - доцільно застосувати економетричні моделі часових рядів з елементами імітаційного моделювання. Такий підхід дозволяє не лише проаналізувати динаміку змін успішності у часовому розрізі, але й зробити обґрунтовані прогнози на основі накопичених історичних даних. Цільовою функцією у цьому

випадку виступає максимізація успішності студентів, що виражається середнім балом, усередненим по групі за певний період часу.

Структура моделі складається з трьох взаємопов'язаних підмоделей. Перша - підмодель статистичного аналізу - оцінює вплив ключових змінних, таких як середній бал, кількість академічних боргів та кількість оцінюваних завдань, на загальну успішність студентської групи. Друга - підмодель регресії - за допомогою методів множинної регресії визначає, як базові параметри, зокрема конкурсний бал при вступі, кількість боржників і чисельність групи, впливають на середній бал. Третя - підмодель прогнозування - використовує часові ряди та метод екстраполяції для формування прогнозу успішності групи на наступний навчальний рік.

Для практичного аналізу було обрано вибірку даних із платформи LMS, що охоплює показники навчальної діяльності студентів протягом трьох академічних років. Вибірка включає дані восьми студентських груп, сформованих за роками вступу до університету, а також інформацію про 61 навчальний предмет (див. рис. 2.1.1), який вивчали студенти протягом цього періоду. Окрім цього, у вибірку були включені показники конкурсу при вступі, зокрема бал за вступні екзамени та інші параметри, пов'язані з процесом набору. Такий комплексний підхід до збору та аналізу даних забезпечує надійну основу для побудови моделей, що допомагають зрозуміти закономірності академічної успішності та сприяють прийняттю ефективних управлінських рішень у сфері освіти.

Рік	Група	Конкурсний бал вступ	Інформаційні технології				Програмування	Фізичне виховання	Вища математика	Філософія	Діловий протокол та етика	основи комп'ютерної інженерії	Основи програмної інженерії	практика з інформаційних технологій
			Іноземна мова (Залін)	технології (Залін)	вання (Залін)	технології (Залін)								
2022	ІПЗ-220076	190,6	90	92	100	85	95	90	90	90	94	76	99	
2022	ІПЗ-220076	154,9	80	84	95	90	90	86	83	75	90	93		
2022	ІПЗ-220076	168,3	79	84	100	78	75	75	75	75	74	83		
2022	ІПЗ-220076	129,556	95	84	87	90	85	84	75	64	74	61		
2022	ІПЗ-220076	145,1	83	74	82	82	75	70	83	80	77	60		
2022	ІПЗ-220076	158,2	75	66	68	90	70	74	86	66	75	86		
2022	ІПЗ-220076	146,222	83	80	100	74	70	74	80	69	68	82		
2022	ІПЗ-220076	189,2	95	95	100	95	90	75	90	94	92	98		
2022	ІПЗ-220076	144,5	82	74	65	60	65	72	77	76	69	79		
2022	ІПЗ-220076	170,8	92	91	96	90	90	90	94	91	90	100		
2022	ІПЗ-220076	145,5	60	60	60	60	60	71	60	61	0	0		
2022	ІПЗ-220076	183,4	95	85	98	90	90	91	95	94	90	98		
2022	ІПЗ-220076	166,7	91	86	100	90	90	90	83	90	92	97		
2022	ІПЗ-220076	189,4	93	91	95	98	90	90	97	91	90	98		
2022	ІПЗ-220076	152,8	92	91	80	90	80	75	93	90	90	80		
2022	ІПЗ-220076	194,6	95	86	100	90	90	90	96	95	92	98		
2022	ІПЗ-220076	187	95	86	100	90	90	81	90	99	91	97		
2022	ІПЗ-220076	148,6	85	85	90	90	80	74	81	90	75	74		
2022	ІПЗ-220076	125,6	65	60	60	90	60	71	72	0	60	0		
2022	ІПЗ-220076	189,2	93	84	100	90	90	75	85	91	90	96		
2022	ІПЗ-220076	142,1	74	60	60	90	60	71	66	61	60	61		
2022	ІПЗ-220076	160,8	82	90	84	90	70	77	79	79	71	69		
2022	ІПЗ-220076	142,7	82	75	60	78	70	70	77	66	64	60		
2022	ІПЗ-220076	145,5	63	61	60	60	65	72	61	0	62	60		

Рисунок 2.1.1 - Таблиця успішності студентів

До цієї вибірки додано такі додаткові параметри, як бали, отримані студентами за участь у додаткових навчальних чи позанавчальних заходах, а також інші фактори, що можуть впливати на успішність студентів, як-от наявність або відсутність академічних досягнень поза межами основної навчальної програми.

Загальний рейтинг студентів, що є кінцевим показником ефективності їхнього навчання, включає не лише результати іспитів, а й ураховує додаткові складові, що можуть відображати рівень залученості студентів у навчальний процес. Вся вибірка включає 188 студентів, що дозволяє забезпечити статистичну значущість аналізу та дослідження різних чинників, що можуть впливати на їхню успішність.

Для першого етапу дослідження було обрано чотири основні змінні: конкурсний бал вступу, навчальний рейтинг студента, додатковий рейтинг студента та тип фінансування (бюджет/контракт). На основі цих показників проведено кореляційний аналіз, результати якого представлені на рис. 2.1.2. Отримані значення кореляційних коефіцієнтів допомагають виявити силу та напрямок зв'язку між змінними.

	Конкурсний бал вступу	Навчальний рейтинг студента	Додатковий рейтинг студента	Бюджет/контракт
Конкурсний бал вступу	1			
Навчальний рейтинг студента	0,6028102	18	1	
Додатковий рейтинг студента	0,3802335	0,3559	08	1
	82	0,4525	0,2718	
Бюджет/контракт	0,4924449	11	93	15
				1

Рисунок 2.1.2 - Кореляційна залежність атрибутів

Сильний позитивний кореляційний зв'язок спостерігається між навчальним рейтингом студента та конкурсним балом вступу (коефіцієнт кореляції = 0,6028). Це свідчить про те, що вищий вступний бал тісно пов'язаний із вищим академічним рейтингом під час навчання. Така залежність є очікуваною, оскільки високі вступні бали можуть відображати сильну початкову підготовку студента, що надалі позитивно позначається на його академічних досягненнях.

Регрессионная статистика								
Множественный R	0,602810218							
R-квадрат	0,363380159							
Нормированный R-квадрат	0,359823624							
Стандартная ошибка	13,5847118							
Наблюдения	181							
Дисперсионный анализ								
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значимость F</i>			
Регрессия	1	18855,36444	18855,36444	102,1725124	2,7409E-19			
Остаток	179	33033,44664	184,5443946					
Итого	180	51888,81108						
	Коэффициенты	Стандартная ошибка	t-статистика	P-значение	Нижние 95%	Верхние 95%	Нижние 95,0%	Верхние 95,0%
Y-пересечение	72,57058329	8,369988144	8,670332865	2,54043E-15	56,0540402	89,08712637	56,0540402	89,08712637
Навчальний рейтинг студента	1,224043403	0,121095995	10,10804197	2,7409E-19	0,985084017	1,463002789	0,985084017	1,463002789

Рисунок 2.1.3 - Регресія

В аналізі регресії, де змінними виступають вступний бал і навчальний рейтинг студента, модель використовується для передбачення їхньої успішності в навчанні. В цьому контексті основні показники регресії допомагають оцінити якість моделі та її значимість для прогнозування навчальних досягнень студентів. Перший ключовий показник - множинний коефіцієнт кореляції

(Multiple R), що дорівнює 0.6028. Цей показник характеризує ступінь лінійного зв'язку між змінними, де 1.0 вказує на ідеальний лінійний зв'язок, а значення, близькі до нуля, свідчать про слабкий зв'язок. У даному випадку, значення 0.6028 свідчить про помірний позитивний зв'язок між незалежною змінною (навчальний рейтинг студента) та залежною змінною (успішність). Це означає, що навчальний рейтинг має суттєвий, хоча й не абсолютний, вплив на успішність студента, що робить цей фактор важливим, але недостатньо вирішальним.

Далі розглянемо коефіцієнт детермінації (R-квадрат), значення якого становить 0.3634. Цей показник показує частку варіації в успішності, яка може бути пояснена моделлю, що враховує вступний бал і навчальний рейтинг. Іншими словами, модель пояснює приблизно 36.34% варіації в навчальних досягненнях студентів, що є помірним результатом. Хоча 36.34% - це не дуже високий показник, він свідчить про те, що навчальний рейтинг відіграє важливу, але не виключну роль у формуванні успішності. Решта 63.66% варіації залишаються непоясненими і можуть залежати від інших, не включених у модель факторів, таких як мотивація, якість викладання, особисті обставини тощо.

Нормований R-квадрат (Adjusted R-squared), значення якого становить 0.3598, також є важливим показником, особливо для моделей із кількома змінними. Цей показник враховує кількість змінних і кількість спостережень, що робить його більш точним у моделях із кількома предикторами. Нормоване значення R-квадрата, яке близьке до звичайного R-квадрата, вказує на стабільність моделі. Це означає, що навіть за додаванням змінних чи корекцією розміру вибірки, модель зберігає свою пояснювальну здатність. Різниця між значенням звичайного R-квадрата і нормованого невелика, що свідчить про те, що модель оптимально адаптована для аналізу обраних даних.

Стандартна помилка моделі, яка дорівнює 13.585, відображає середнє відхилення прогнозованих значень від фактичних результатів. Ця помилка вказує на рівень невизначеності або середню похибку прогнозу, і значення 13.585 є доволі значним. Це може означати, що хоча модель і надає корисні прогнози, вони супроводжуються певною мірою невизначеності, яка може бути важливою

для інтерпретації результатів. Висока стандартна помилка сигналізує про те, що для підвищення точності прогнозів необхідно досліджувати й додаткові змінні або розширити модель іншими предикторами.

Результати дисперсійного аналізу (ANOVA) також свідчать про статистичну значимість моделі. Значення F-статистики дорівнює 102.1725, а значущість F (p-значення) становить 2.74E-19, що практично дорівнює нулю. Таке високе значення F-статистики та дуже мале значення p-значення (близьке до нуля) дозволяють зробити висновок, що між незалежними змінними та успішністю студента існує статистично значущий зв'язок. Це підтверджує, що обрана модель адекватно описує вплив навчального рейтингу та вступного балу на навчальні досягнення і що ймовірність випадкової появи такої кореляції надзвичайно мала.

Коефіцієнти регресії є наступним важливим аспектом. Перетин із віссю Y (Y-пересечение), значення якого становить 72.57, показує передбачувану успішність студентів, якщо навчальний рейтинг дорівнює нулю. Хоча ця точка перетину має обмежену інтерпретаційну цінність у реальному контексті, вона все ж таки необхідна для побудови регресійного рівняння. Основний коефіцієнт для навчального рейтингу студента дорівнює 1.224, що означає, що з підвищенням навчального рейтингу на одну одиницю успішність студента зростає в середньому на 1.224 бали. Це позитивне значення свідчить про те, що навчальний рейтинг має прямий вплив на підсумковий результат студента, що підтверджує його значимість у моделі.

Нарешті, результати для T-статистики та значущості (p-значення) для коефіцієнтів свідчать про статистичну значущість кожного коефіцієнта в моделі. Для обох коефіцієнтів (перетин із віссю Y і навчальний рейтинг) значення p-значення дуже малі (приблизно 0 для навчального рейтингу). Це означає, що обидва коефіцієнти є статистично значущими і мають вагомий внесок у модель. Таким чином, навчальний рейтинг виступає важливим фактором, який допомагає передбачити успішність студента.

Отже, результати регресійного аналізу вказують на те, що модель є статистично значущою і може пояснити певну частину успішності студентів на основі їхнього навчального рейтингу та вступного балу. Хоча коефіцієнт детермінації вказує на те, що значна частина варіації залишається неврахованою, наявні результати підтверджують, що навчальний рейтинг є вагомим предиктором успішності. Це вказує на доцільність приділення більшої уваги навчальному рейтингу як одному з факторів, що сприяють успішному навчанню.

Також звернемося до візуалізації факторів на графіку (рис. 2.1.4), який наочно ілюструє виявлені кореляційні залежності. Згідно з графіком, підтверджується тісний зв'язок між конкурсним балом вступу та навчальним рейтингом студента, який був виражений кореляційним коефіцієнтом 0,6028. Візуально це проявляється у схожих тенденціях зміни значень для цих показників: студенти з високими вступними балами, як правило, демонструють високий рівень успішності протягом навчання, що відображається у високих навчальних рейтингах.

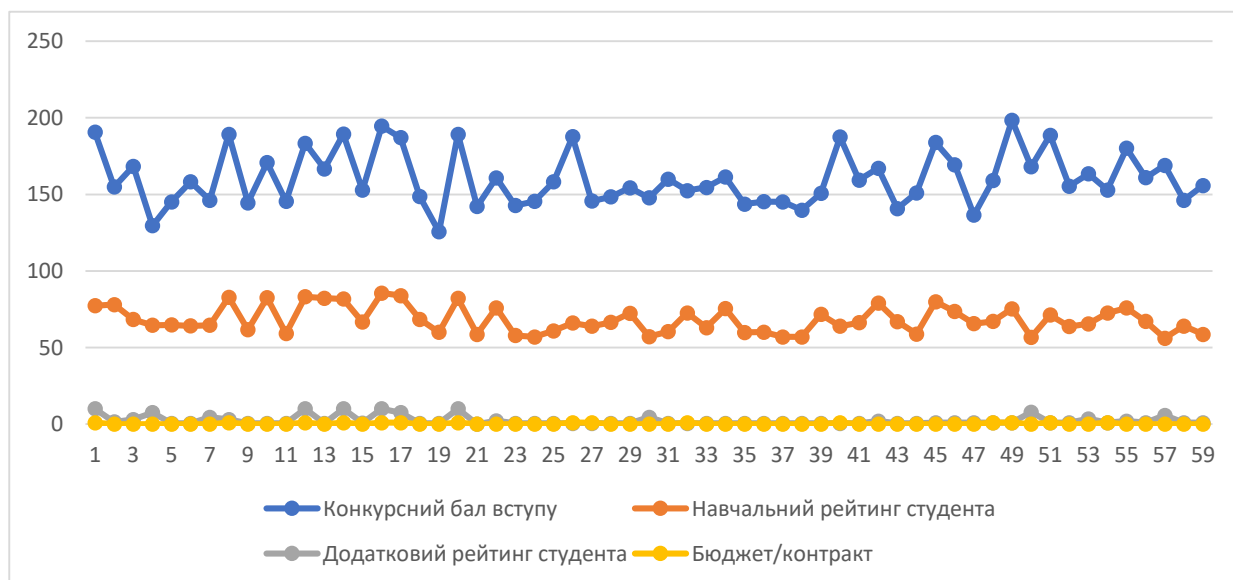


Рисунок 2.1.4 - Графік факторів

Також існує помітний зв'язок між конкурсним балом вступу та типом фінансування (бюджет або контракт) із коефіцієнтом кореляції 0,4925. Це означає, що студенти з вищими вступними балами частіше потрапляють на

бюджетні місця, що свідчить про важливість конкурсу при розподілі фінансування.

Крім того, існує кореляція між навчальним рейтингом студента та типом фінансування (0,4526). Це також логічно, адже студенти на бюджеті можуть мати кращу академічну мотивацію чи додаткову підтримку, що впливає на їхні результати під час навчання.

Інші взаємозв'язки, як-от між додатковим рейтингом студента та іншими показниками, також вказують на певну залежність, але є менш значущими, що відображається в нижчих значеннях кореляції (до 0,38). Ці кореляційні зв'язки свідчать про логічний взаємозв'язок між показниками, обраними для дослідження, та дозволяють глибше розуміти фактори, що впливають на академічну успішність студентів.

На основі даних про конкурсний бал, тип фінансування (бюджет/контракт), навчальний та додатковий рейтинги можливо створити прогнозу модель для оцінки майбутньої академічної успішності студентів. Лінійна регресія дозволяє передбачати навчальний рейтинг, оскільки конкурсний бал і тип фінансування показали значну кореляцію з результатами навчання. Ця модель допоможе університету оптимізувати підтримку для студентів, ефективно розподіляти ресурси та покращувати освітній процес, забезпечуючи прогнозування на основі ключових показників вступу та навчання.

## **2.2 Аналіз успішності студентів з урахуванням академічної заборгованості**

Для детальнішого аналізу успішності студентів було розраховано середні бали по кожному предмету для всіх груп, використовуючи два підходи: з урахуванням боргів (борг оцінюється як 0 балів) та без урахування боргів. Додатково було підраховано кількість боржників для кожного предмету, що дозволяє оцінити вплив кількості боргів на середній бал. У таблиці 2.2.1 представлено отримані дані, де середні значення відображено кольоровим градієнтом від зеленого (вищі значення) до червоного (нижчі значення), що візуально демонструє ситуацію із середніми балами по предметах.

Рік	Група	Конкурсний бал вступу	Іноземна мова (Залі)	Інформаційні технології (Залі)	Програмування (Залі)	Фізичне виховання (Залі)	Вища математика (Залі)	Філософія (Ісп)	Діловий протокол та етика спілкування (Ісп)	основи комп'ютерної інженерії (Ісп)	Основи програмування (Ісп)	практика з інформаційних технологій (Залі)	Навчальна практика з групової динаміки (Ісп)	Правова культура особистості (Ісп)	алгебра та аналітика (Ісп)	Програмування технологій (Ісп)	Інформаційні технології (Ісп)	
2023	ІПЗ-230076	159	64	90	79	74	67	74	79	63	74	0	62	83	60	60	64	
2023	ІПЗ-230076	168,545	90	82	90	90	65	68	79	78	84	0	80	94	78	90	61	
2023	ІПЗ-230076	132	72	81	60	82	60	63	75	63	75	0	64	65	60	0	60	
2023	ІПЗ-230076	123,4	62	63	60	80	60	74	75	60	64	0	0	62	60	0	60	
2023	ІПЗ-230076	140,3	64	75	60	90	60	68	77	61	74	0	65	65	60	0	60	
2023	ІПЗ-230076	148,545	67	84	64	90	65	83	86	63	85	0	78	80	60	60	60	
2023	ІПЗ-230076	173,909	82	83	76	90	86	75	80	85	84	0	92	97	92	95	76	
2023	ІПЗ-230076	146,25	83	90	96	94	85	85	84	96	90	0	94	95	92	98	82	
2023	ІПЗ-230076	155,5	90	90	82	90	79	86	91	74	90	0	83	99	74	85	80	
2023	ІПЗ-230076	142,222	60	75	60	90	60	78	86	0	70	0	62	92	60	60	65	
2023	ІПЗ-230076	122,1	80	86	66	74	65	76	70	64	65	0	62	90	60	60	64	
2023	ІПЗ-230076	165,182	64	75	65	80	65	74	81	65	75	0	64	90	60	60	63	
2023	ІПЗ-230076	179	88	96	97	90	80	90	90	88	81	0	95	90	91	100	82	
2023	ІПЗ-230076	143,3	72	62	69	82	60	71	87	63	74	0	64	75	60	75	60	
2023	ІПЗ-230076	169,545	80	81	88	90	60	71	97	74	76	0	79	96	76	85	75	
2023	ІПЗ-230076	147,364	60	74	60	60	60	66	72	60	74	0	62	74	60	60	60	
2023	ІПЗ-230076	172,909	92	85	82	90	75	69	93	84	90	0	82	97	74	97	60	
2023	ІПЗ-230076	155	74	79	94	85	70	69	86	62	77	0	60	85	64	80	64	
2023	ІПЗ-230076	137,8	60	63	60	90	60	60	73	60	60	0	60	79	60	60	0	
2023	ІПЗ-230076	143,273	73	64	60	60	60	62	74	60	62	0	61	74	60	60	63	
середнє з боргами			81,58011	80,071823	78,801105	83,569061	72,961326	79,237569	81,337017	71,696133	75,563536	49,712707	51,099448	70,61326	77,292818	71,248619	74,828729	71,116022
середнє не враховуючи боржників			81,58011	80,071823	80,129213	83,569061	72,961326	79,237569	81,337017	74,154286	76,407821	74,983333	75,811475	74,74269	78,595506	73,272727	80,142012	73,554286
боржники кількість			0	0	3	0	0	0	0	6	2	10	8	10	3	5	12	6
боржники			0	0	-1,3281085	0	0	0	0	-2,4581531	-0,8442853	-25,270626	-24,712028	-4,1294304	-1,3026879	-2,0241085	-5,3132826	-2,4382636

Рисунок 2.2.1 - Таблиця з середніми значеннями

З результатів видно, що існує певна залежність між кількістю боржників та середнім балом, обчисленим без урахування боргів: зокрема, більша кількість боржників призводить до зниження середнього балу при їхньому виключенні з розрахунків. Проведення кореляційного аналізу підтвердило наявність зв'язку, хоча цей зв'язок не є достатньо сильним.

	середнє не враховуючи боржників	боржники кількість
середнє не враховуючи боржників	1	
боржники кількість	0,12863399	1

Рисунок 2.2.2 - Кореляційна залежність двох факторів

На основі цього можна припустити, що для частини студентів ці предмети є нормальними за рівнем складності, тоді як інша частина групи, яка має заборгованості, може стикатися з труднощами через фактори, що не обов'язково пов'язані із самим предметом. Серед можливих причин - особисте зацікавлення, мотивація та рівень відвідуваності, що відрізняються між студентами.

Наприклад, на графіку було побудовано дві порогові лінії: одну на рівні 75 балів для середнього результату (без урахування боржників) та іншу на позначці 10 для кількості боржників. Завдяки цим лініям можна легко визначити предмети, де середній бал нижчий за 75, а кількість боржників перевищує 10. Ці порогові значення можуть змінюватися залежно від цілей аналізу, дозволяючи

адаптувати підхід до конкретних потреб. Таким чином, усі точки (предмети), що розташовані між двома червоними лініями, вважаються проблемними, оскільки демонструють низькі середні бали та високу кількість боржностей, що вказує на потенційні труднощі для студентів у цих дисциплінах.

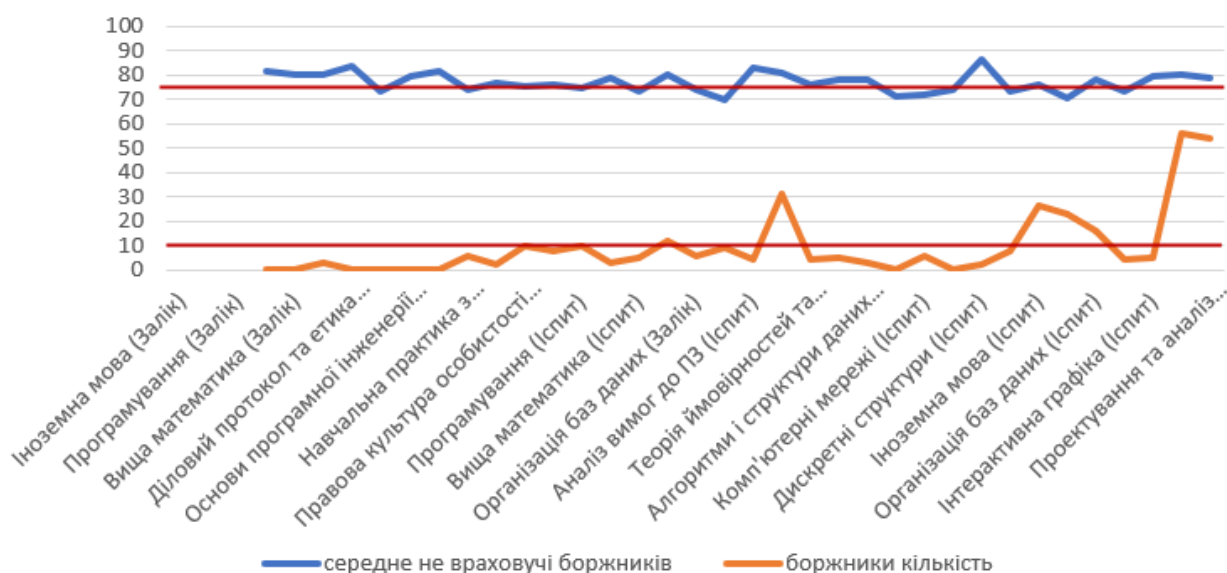


Рисунок 2.2.3 - Графік середніх балів та кількості боржників з пороговими лініями

Аналіз успішності показав, що низка предметів має низькі середні бали та велику кількість боржників. Для покращення результатів рекомендується організувати додаткові консультації та заняття для студентів, яким складно засвоювати матеріал. Також доцільно переглянути та адаптувати навчальні матеріали складних тем, щоб зробити їх більш структурованими та доступними для розуміння. Запровадження заохочень за відвідування та активну участь допоможе підвищити зацікавленість і залученість студентів. Важливо також здійснювати моніторинг успішності на ранніх етапах курсу, щоб своєчасно виявляти студентів, які потребують підтримки. Додатково можна переглянути критерії оцінювання та розглянути можливість запровадження більш гнучкої системи оцінювання, щоб підтримати студентів у процесі навчання. Такі кроки дозволять підвищити середній бал, зменшити кількість боржників та загалом покращити академічну успішність студентів.

Для детального аналізу успішності студентів та виявлення тенденцій у навчанні було взято дані по чотирьох ключових предметах, що викладаються на першому курсі: "Основи програмної інженерії", "Групова динаміка і комунікації", "Програмування" та "Вища математика". Дослідження охоплює студентів груп 2021, 2022 та 2023 років вступу, а основними показниками для аналізу стали середній бал та кількість боржників за кожен предмет (рис. 2.2.4). Такий підхід дозволяє не лише відстежити динаміку успішності, але й оцінити зовнішні чинники, які могли вплинути на результати навчання.

	Основи програмної інженерії (Іспит)	Групова динаміка і комунікації (Іспит)	Програмування (Іспит)	Вища математика (Іспит)
сер.2021	79,11111	75,56452	80,07937	70,9375
сер.2022	73,75385	73,90476	79,82813	68,29032
сер.2023	76,45098	74,78261	80,71429	70,45652
кількість боржників 2021	1	2	1	0
кількість боржників 2022	1	3	2	4
кількість боржників 2023	0	5	9	5

*Рисунок 2.2.4 - Таблиця середніх оцінок та боргів*

Аналіз середніх балів показує, що у 2021 році студенти продемонстрували відносно високий рівень успішності з усіх предметів. Однак у 2022 році спостерігається зниження середнього балу, що може бути пов'язане з кількома чинниками. Перш за все, варто відзначити, що саме у 2022 році студенти зазнали суттєвого впливу воєнних дій, які могли спричинити психологічний стрес, нестабільність та відволікання від навчання. У 2023 році середній бал частково відновлюється, що може свідчити про певну адаптацію до нових умов. Однак, попри часткове зростання середнього балу, кількість боржників у 2023 році досягає рекордного значення. Це вказує на те, що частина студентів має серйозні труднощі з навчанням і не може повністю адаптуватися до нових вимог та умов.

Для об'єктивнішого порівняння групи було нормалізовано дані з урахуванням різної кількості студентів у кожному році (рис 2.2.5).

кількість студентів 2021	64
кількість студентів 2022	66
кількість студентів 2023	51

*Рисунок 2.2.5 - Таблиця кількості студентів*

Так, розраховано зважений середній бал та зведену кількість боржників на 100 студентів, що дозволило зробити результати порівнянними між групами (рис. 2.2.6). Аналіз нормалізованих даних показав, що найбільші труднощі з навчанням виникли у студентів 2022 року, коли, окрім зниження середнього балу, значно зросла кількість боржників. Це може бути обумовлено тим, що саме у цей період навчальний процес був порушений через частковий або повний перехід на дистанційне навчання. Відсутність стабільного зв'язку, недостатня технічна підтримка та адаптація до нових форматів навчання могли негативно вплинути на засвоєння матеріалу.

	Основи програмної інженерії (Іспит)	Групова динаміка і комунікації (Іспит)	Програмування (Іспит)	Вища математика (Іспит)
зважений середній бал 2021	79,11111	75,56452	80,07937	70,9375
зважений середній бал 2022	73,7535	73,90476	79,82813	68,29032
зважений середній бал 2023	76,4508	74,78261	80,71429	70,45652
зважена кількість боржників 2021	1,5625	3,125	1,5625	0
зважена кількість боржників 2022	1,51512	4,545455	3,030303	6,060606
зважена кількість боржників 2023	0	9,803922	17,64706	9,803922

*Рисунок 2.2.6 - Таблиця зважених середніх*

Побудуємо для наглядності графіки. Перший з зображенням середніх балів успішності студентів для чотирьох ключових дисциплін - "Основи

програмної інженерії", "Групова динаміка і комунікації", "Програмування" та "Вища математика". Графік дозволяє побачити зміну середнього балу для кожного предмета серед студентів різних років вступу.

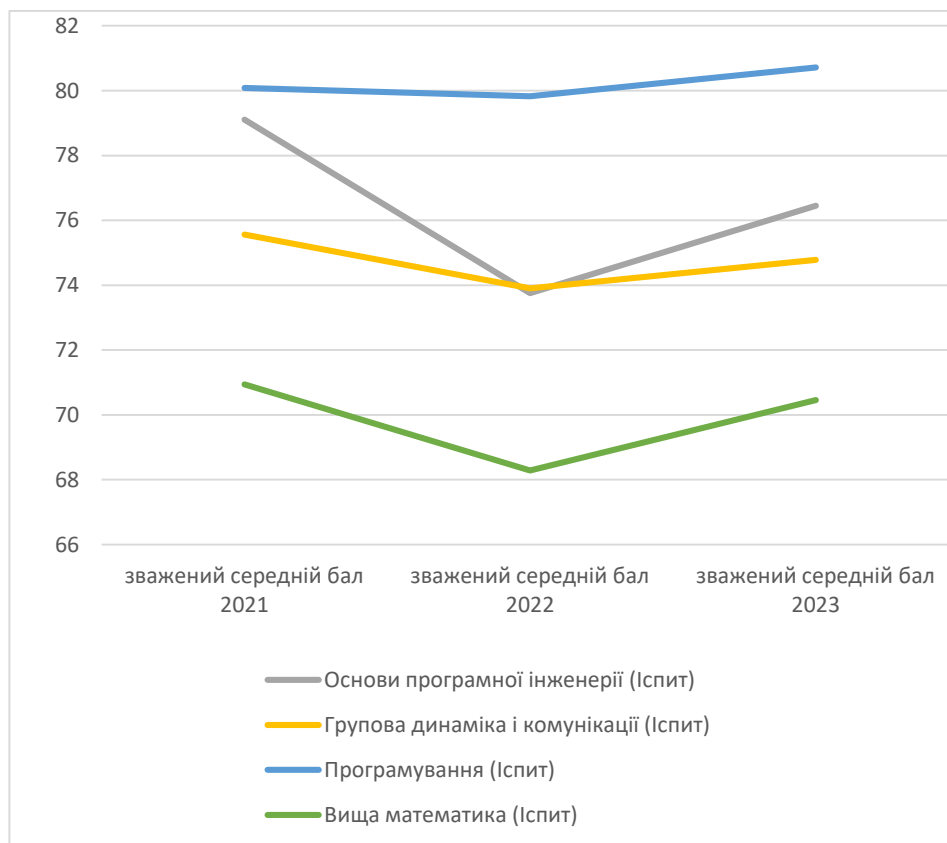


Рисунок 2.2.7 - Графік зважених середніх

Аналіз окремих дисциплін: Особливо помітні зміни в успішності з дисципліни "Групова динаміка і комунікації", де середній бал знизився у 2022 році, що може свідчити про складність організації комунікативного навчання в дистанційному форматі. Інші предмети, такі як "Програмування", демонструють відносну стабільність показників, однак "Вища математика" у всіх групах залишається предметом з найнижчим середнім балом, що вказує на її високу складність для студентів.

Другий графік демонструє зважену кількість боржників, він показує середній бал та кількість боржників, зведену до 100 студентів, для кожного року вступу. Такий підхід дозволяє уникнути впливу різної чисельності груп на результати, що робить дані об'єктивнішими для порівняння.

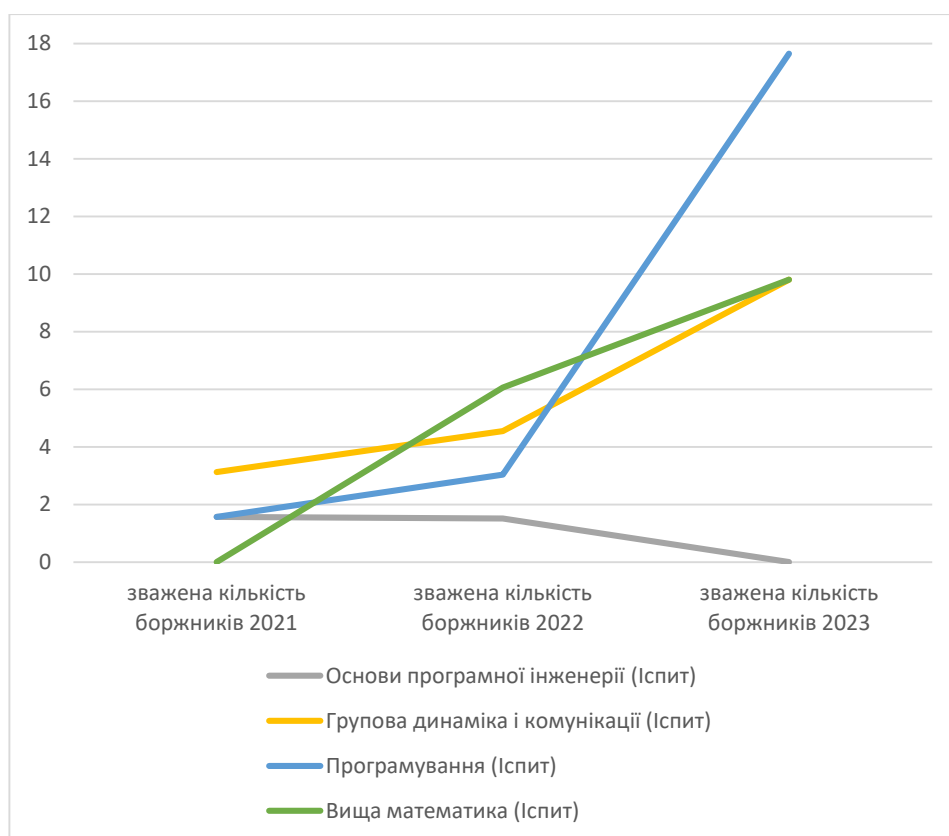


Рисунок 2.2.8 - Графік зважених середніх боржників

Можливі фактори впливу на успішність студентів:

Психологічний стрес та нестабільність через війну. Військові дії, що розпочалися у 2022 році, мали значний вплив на психологічний стан студентів. Постійний стрес, невпевненість у майбутньому, часткове або повне переселення та необхідність адаптуватися до нових умов життя могли знизити концентрацію та мотивацію до навчання. Ці фактори, безумовно, вплинули на успішність студентів.

Перехід на дистанційне навчання. У зв'язку з пандемією COVID-19 та воєнними подіями багато університетів змушені були перейти на дистанційне навчання. Хоча це рішення було необхідним, дистанційний формат має певні обмеження, зокрема, зниження рівня взаємодії з викладачами, відсутність структурованого навчального середовища та труднощі з технічною підтримкою. Студенти, особливо на першому курсі, могли відчувати нестачу індивідуального підходу та підтримки, що негативно вплинуло на їхні результати.

Недостатня технічна база та умови для навчання вдома. Не всі студенти мають доступ до належної технічної бази (наприклад, комп'ютери, стабільний інтернет), що є необхідним для дистанційного навчання. Крім того, умови вдома не завжди сприяють ефективному навчанню через відсутність спеціально обладнаного простору, відволікаючі фактори та інші побутові обмеження.

Відсутність соціальної підтримки та ізоляція. Дистанційне навчання та пандемія обмежили соціальну взаємодію студентів, що є важливим фактором для навчання, особливо на першому курсі, коли студенти тільки адаптуються до нових умов. Відсутність живого спілкування з однокурсниками та викладачами могла призвести до зниження мотивації, втрати відчуття підтримки та солідарності.

Зміни в програмі та адаптація до нових вимог. Навчальні програми могли зазнати певних змін у зв'язку з новими умовами навчання, що також могло вплинути на результати студентів. Різке зростання складності навчальних матеріалів, неадаптованих до дистанційного формату, або зменшення часу на підготовку могли спричинити зростання кількості боржників.

Проблеми з дисципліною та самоконтролем. Дистанційне навчання вимагає від студентів вищого рівня самоконтролю та дисципліни, що для молодших студентів може бути викликом. Відсутність регулярних занять в аудиторіях та прямих взаємодій з викладачами може призвести до меншої уваги до навчання та до накопичення боргів.

### **2.3 Динаміка успішності та рівень боргів студентів групи вступу 2021 року за шість семестрів**

У цьому розділі проводиться детальний аналіз успішності студентів групи, яка вступила в 2021 році, протягом шести семестрів навчання. Основною метою дослідження є виявлення тенденцій у зміні рівня успішності студентів та вивчення динаміки боргів, що дозволить зробити висновки щодо загальної мотивації та зусиль студентів у процесі навчання.

Для початку було розраховано середні бали студентів по кожному предмету за кожний семестр, причому в обчисленнях не враховувалися студенти, які мають борги за відповідний період. Це дозволило отримати більш об'єктивну картину успішності, оскільки виключено вплив студентів, які не виконали вимоги курсу та мають незадовільні оцінки. Таким чином, було отримано середнє значення балів для кожного семестру.

Подальше обчислення середніх значень по семестрах дає можливість проаналізувати загальні зміни в успішності групи без урахування конкретних предметів, а також оцінити, як змінюється рівень навчальних досягнень студентів у загальному контексті. Оскільки аналіз не враховує боржників, це дозволяє більш чітко побачити, як колективно змінюється успішність у групі, незалежно від індивідуальних проблем з навчанням.

Отримані середні бали за кожен семестр були зображені на графіку (рис. 2.3.1), де, окрім самих значень, була накладена лінія тренду. Спостереження за цією лінією дозволило виявити спад успішності студентів, що свідчить про тенденцію зниження мотивації до навчання та загального рівня успішності протягом часу. Це може бути результатом кількох факторів, таких як зниження інтересу до навчальних предметів, зростаюча складність навчальної програми або зміни в організації навчального процесу.

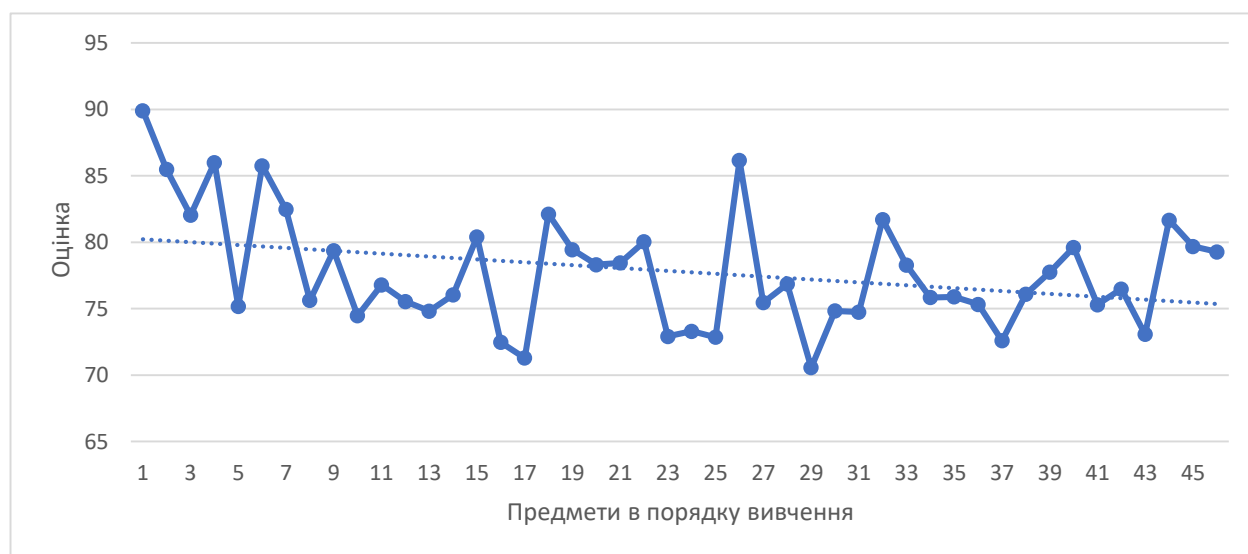


Рисунок 2.3.1 - Графік успішності групи впродовж 3 років

У подальшому було проаналізовано динаміку боргів студентів, використовуючи аналогічний метод. Для кожного семестру було підраховано кількість студентів, які не змогли здати іспити або виконати інші вимоги курсу, що призвело до наявності боргів. Дані про борги також були представлені у таблиці (рис. 2.3.2), де кожному семестру відповідають середні бали та кількість студентів з боргами.

	1 семестр	2 семестр	3 семестр	4 семестр	5 семестр	6 семестр
сер	83,8	76,0	77,8	75,3	76,3	77,8
едне	3195	9431	1319	6593	0882	5215
бор						
ги	2	4	11	20	59	89

Рисунок 2.3.2 - Зведена таблиця по успішності

На основі цих даних було побудовано ще один графік, на якому стовпцями відображено середні оцінки, а лінією - кількість боржників (рис.2.3.3).

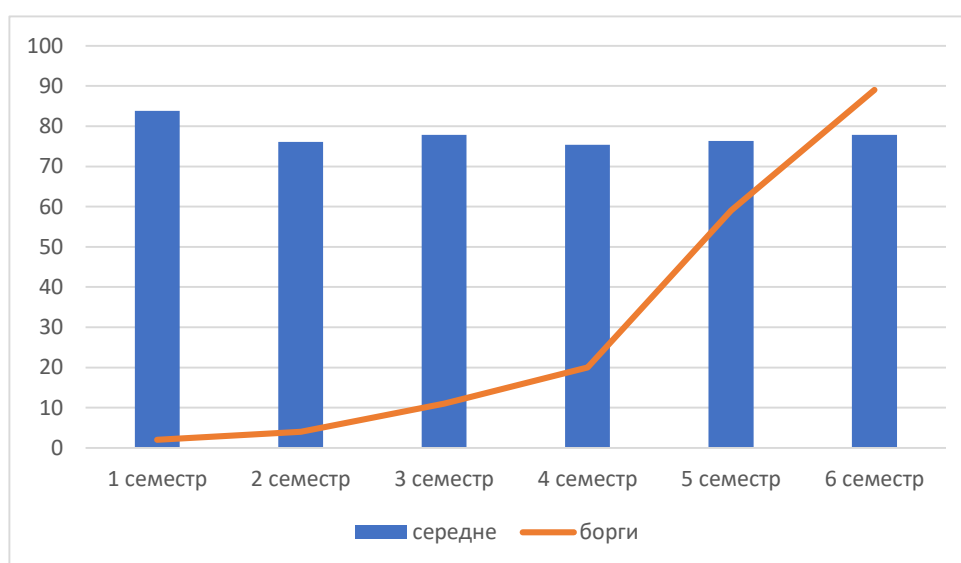


Рисунок 2.3.3 – Графік середніх оцінок та боргів в розрізі 6 семестрів

Результати показали, що кількість студентів з боргами поступово зростає протягом навчання, і до 2023 року цей показник зростає в рази. Це, ймовірно, свідчить про зростаючі труднощі студентів в успішному освоєнні матеріалу. Водночас, середній бал знижується, що підтверджує зменшення загального рівня успішності серед студентів, ймовірно, через накопичення боргів.

Таким чином, аналіз показав важливі тенденції: з одного боку, зниження середнього балу студентів по семестрах, а з іншого - значне зростання кількості боргів, що свідчить про погіршення успішності групи і, ймовірно, про зниження рівня мотивації студентів до навчання. Ці результати можуть бути корисними для подальших досліджень і для коригування навчальних планів, підходів до викладання та підтримки студентів у процесі навчання.

Для вирішення проблеми зниження успішності та зростання кількості боргів серед студентів групи вступу 2021 року необхідно впровадити комплексний підхід, що включає індивідуалізовану підтримку студентів, організацію додаткових консультацій та занять, мотиваційні програми, психологічну допомогу, а також перегляд навчальних планів та адаптацію курсів під потреби студентів. Важливо створити умови для активної взаємодії студентів через групові заняття та менторство, використання новітніх технологій для навчання та забезпечення регулярного зворотного зв'язку для корекції навчального процесу. Такий підхід дозволить зменшити кількість боргів і покращити загальний рівень успішності студентів.

## **РОЗДІЛ 3. АНАЛІЗ ТА ПРОГНОЗУВАННЯ УСПІШНОСТІ СТУДЕНТІВ НА ОСНОВІ ДАНИХ LMS І МЕТОДІВ МАШИННОГО НАВЧАННЯ В R**

### **3.1 Аналіз факторів навчального процесу, що впливають на успішність студентів**

Для проведення аналізу та побудови моделі прогнозування успішності студентів було використано набір даних, отриманий з системи управління навчанням (LMS) факультету інформаційних технологій. Вибірка охоплює дані про студентів з першого по четвертий курс, що навчаються на різних спеціальностях цього факультету. Загальний обсяг даних становить 74 427 записів. Інформація включає широкий спектр характеристик, зокрема: навчальну групу студента, унікальний ідентифікатор особи, ПБ, код курсу, назву навчальної дисципліни, дату та номер заняття, відмітку про присутність, нормалізований ідентифікатор, стать студента (чоловіча або жіноча), джерело фінансування навчання (бюджет або контракт), курс навчання, конкурсний бал при вступі, кількість входів на курс у LMS, а також поточні академічні бали, отримані студентом. Такий обсяг та структура даних дозволяють здійснити як описовий аналіз освітньої діяльності, так і побудову моделей прогнозування успішності з використанням сучасних методів машинного навчання.

Під час обробки даних у середовищі програмування R одним із перших кроків є приведення стовпця з датами до відповідного формату типу Date. Далі з цієї дати виділяється місяць, який записується в окрему колонку для подальшого аналізу сезонних особливостей. Стовпець із часом занять також перетворюється у відповідний формат типу Time. Крім того, змінна «Відмітка» (наявність студента на занятті) переводиться у факторний тип даних, що дозволяє зручно працювати з категоріальними значеннями під час аналізу та побудови моделей (код наведений в додатках).

На цьому етапі дослідження була побудована перша аналітична модель, метою якої є вивчення зв'язку між активністю студента в системі навчання (кількістю входів на курс) та отриманими балами. Для коректнішої оцінки моделі

з аналізу були виключені студенти з нульовими балами, оскільки такі результати зазвичай є наслідком неявки на підсумковий контроль або нездачі іспиту чи заліку. Нульові бали суттєво спотворюють статистичну картину, адже не відображають реального рівня знань, а радше технічний факт академічної заборгованості.

Для візуалізації залежності була побудована діаграма розсіювання (scatter plot), де по осі X відображено кількість входів на курс, а по осі Y - підсумкові бали студента за дисципліну. Для кращого розуміння тенденцій у даних додано лінію лінійної регресії.

### Активність vs. Бали (без нульових балів)

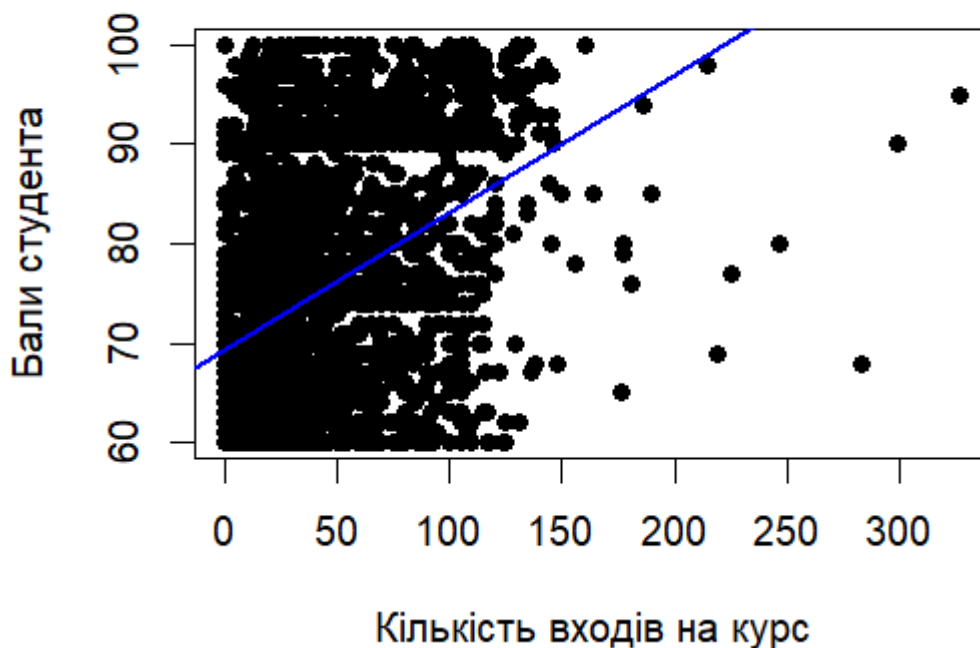


Рисунок 3.1.1 Графік впливу навчальної активності на успішність студентів

Графік, побудований на основі відібраних даних, демонструє закономірність між активністю студентів у навчальній системі (виміряною кількістю входів на курс) та їхніми підсумковими балами. Більшість точок на графіку зосереджені у верхній частині діаграми, що відповідає діапазону приблизно від 60 до 100 балів. Це свідчить про те, що переважна частина студентів, включених до вибірки, показали позитивні результати в навчанні. У

той же час кількість входів на курс у них варіюється досить значно - від майже нульової до понад 300. Помітно також певне скупчення точок у проміжку від 0 до 150 входів, що свідчить про те, що значна частина студентів має помірний рівень активності в LMS.

На графіку також чітко видно синю лінію регресії, яка після виключення нульових балів набуває позитивного нахилу. Це означає, що серед студентів із позитивними балами простежується тенденція зростання успішності зі збільшенням активності в курсі. Лінія регресії починається приблизно з рівня 68 балів при нульових входах і поступово піднімається до понад 90 балів при кількості входів, що наближається до 250-300. Такий вигляд регресійної лінії підтверджує наявність прямої залежності між частотою взаємодії з навчальними матеріалами та підсумковими оцінками, принаймні серед студентів, які вже продемонстрували базовий рівень результативності.

Отримані результати свідчать про наявність позитивної кореляції між навчальною активністю та успішністю студентів. Виключення записів з нульовими балами дозволило побачити більш реалістичну картину: серед тих студентів, які продемонстрували хоча б мінімальний рівень успішності, підвищена активність у системі LMS загалом супроводжується вищими результатами. Проте варто зауважити, що навколо лінії регресії спостерігається значний розкид точок. Це означає, що лінійна модель, хоча й демонструє загальну тенденцію, не пояснює всіх варіацій у балах. Існують студенти з малою кількістю входів, які отримали високі бали, та навпаки - активні студенти з нижчими результатами. Це може свідчити про вплив додаткових факторів, таких як якість викладання, індивідуальні здібності, самостійна підготовка, рівень початкових знань тощо. Також ймовірно, що після досягнення певного порогу активності подальше її зростання вже не має суттєвого впливу на успішність.

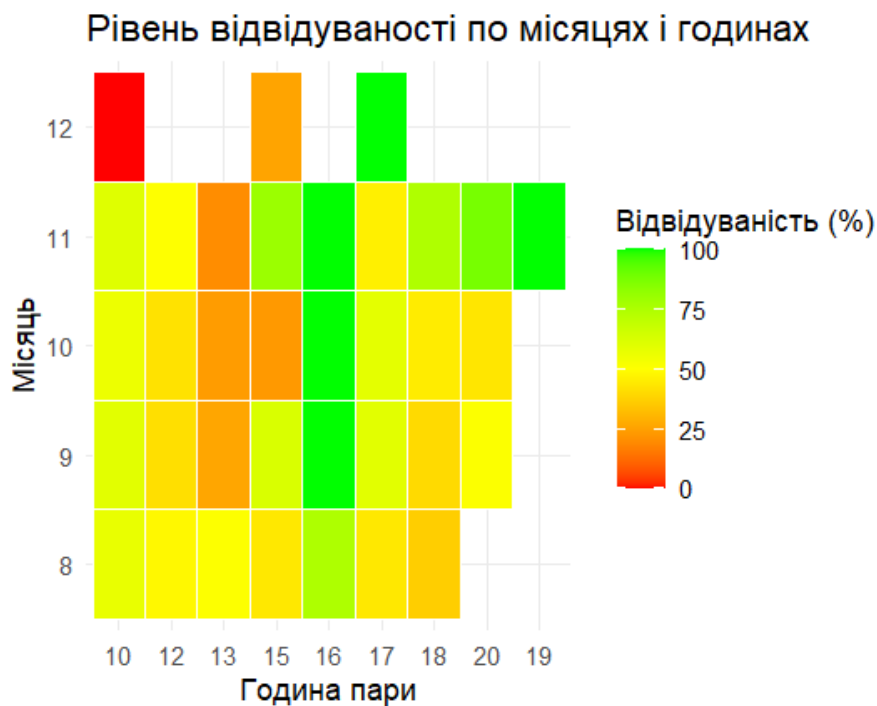
Таким чином, побудована модель дозволяє зробити висновок про те, що активність у навчальній системі загалом корелює з успішністю, але для глибшого розуміння необхідно враховувати додаткові змінні та використовувати складніші моделі прогнозування.

У другій моделі розглядається залежність між відвідуваністю студентів та тим, коли саме відбувається навчальне заняття - зокрема, аналізується вплив місяця та години на ймовірність присутності. Для цього було проведено агрегацію даних із розрахунком середнього рівня відвідуваності для кожної унікальної комбінації місяця та години проведення заняття. Це дозволило виявити часові закономірності, які можуть бути важливими при плануванні навчального процесу.

	Місяць	Година	СерВідвідуваність
1	8	02:00	45.652174
2	8	10:30	57.158530
3	8	12:00	52.941176
4	8	12:10	43.350743
5	8	13:00	0.000000
6	8	13:15	100.000000
7	8	13:50	52.342286
8	8	15:10	0.000000
9	8	15:13	75.000000
10	8	15:30	55.991875
11	8	16:50	75.000000
12	8	17:10	43.543544

*Рисунок 3.1.2 Агрегована таблиця даних по даним проведення занять*

Для зручності аналізу було побудовано теплову карту, на якій візуалізовано середній відсоток відвідуваності студентів. Кожна клітинка графіка відповідає певному місяцю та годині (без урахування хвилин), а колірна шкала відображає рівень відвідуваності. З аналізу були виключені записи, що містять час 02:00, оскільки це некоректні дані - навчальні заняття не можуть проводитися в другій годині ночі.



*Рисунок 3.3.3 Теплова карта середньої відвідуваності за місяцями та годинами*

Отримана візуалізація дозволила виявити кілька характерних закономірностей. Найвища відвідуваність спостерігається в денні години, переважно з 12:00 до 17:00. Це узгоджується з загальною тенденцією, за якої студенти демонструють найбільшу активність у середині дня. Ранкові години (до 12:00) зазвичай характеризуються нижчим рівнем відвідуваності, так само як і вечірні (після 17:00), що може бути пов'язано як із фізіологічними особливостями, так і з розкладом інших занять або особистими обставинами студентів.

Щодо місячної динаміки, то у вересні відвідуваність залишається на відносно високому рівні протягом усього дня, що може пояснюватися мотиваційним підйомом на початку навчального року. У грудні ж спостерігається зниження відвідуваності, особливо в ранкові години. Це може бути пов'язано як з втомою студентів наприкінці семестру, так і з підвищеним навантаженням у період перед сесією, коли частина занять може переноситися або скасовуватися.

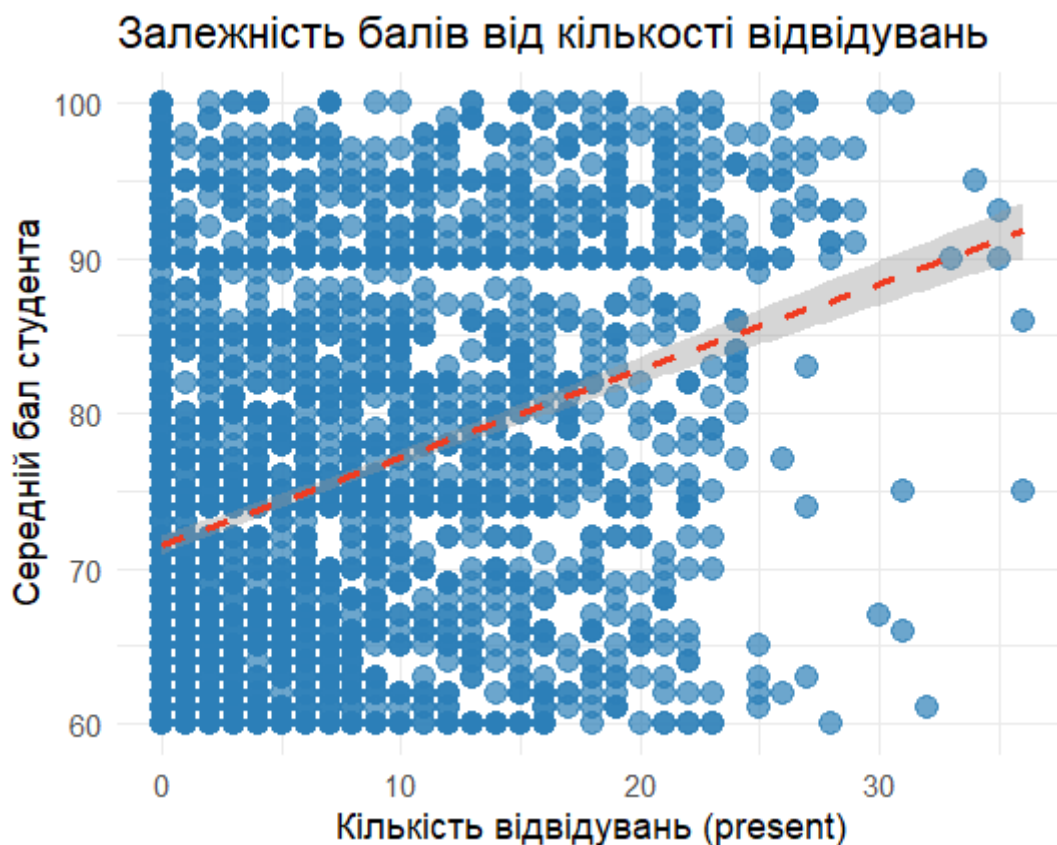
Години з 12:00 до 17:00 демонструють найбільш стабільну та високу відвідуваність упродовж більшості місяців. У той час як вечірні години,

особливо з 18:00 до 20:00, характеризуються зниженням рівня присутності студентів, що може свідчити про втому, додаткову зайнятість або зниження мотивації наприкінці дня.

Загалом, результати цього аналізу підтверджують наявність стійких часових патернів у поведінці студентів. Після виключення аномальних значень (зокрема часу 02:00), основні закономірності залишилися незмінними, що підтверджує їхню стабільність. Ці висновки можуть бути практично застосовані під час формування розкладу занять, зокрема - при визначенні оптимального часу для проведення важливих навчальних активностей або заходів, спрямованих на підвищення відвідуваності.

У рамках третьої моделі було досліджено, чи існує взаємозв'язок між фактичною кількістю відвідувань занять студентом та його підсумковим балом за курс. З метою підвищення точності аналізу з вибірки було виключено записи з нульовими балами, оскільки вони здебільшого відображають випадки академічної заборгованості або неспаних заліків та іспитів, і таким чином можуть спотворювати загальну статистичну картину.

Для побудови моделі де після попередньої обробки даних збудовано лінійну регресійну модель. Візуалізація результатів представлена у вигляді точкової діаграми, на якій по осі X відображено кількість відвідувань занять студентом, а по осі Y - його середній бал за курс. Кожна точка на графіку відповідає одному студенту в контексті конкретної дисципліни.



*Рисунок 3.1.4 Залежність балів від кількості відвідувань (без нульових оцінок)*

На графіку чітко видно, що зростання кількості відвідувань загалом супроводжується підвищенням підсумкових балів. Лінія тренду (червона пунктирна), побудована методом лінійної регресії, демонструє помірно позитивний нахил, що свідчить про наявність прямої залежності між відвідуваністю та результатами навчання. Сіра зона навколо лінії позначає довірчий інтервал на рівні 95%, що дозволяє оцінити надійність прогнозу.

Особливу увагу привертає скупчення точок у діапазоні до 10 відвідувань, де значення балів варіюються досить широко - від 60 до 100. Це свідчить про те, що хоча деякі студенти можуть досягати високих результатів навіть за низької відвідуваності, така ситуація є менш типовою. Натомість у діапазоні 20–35 відвідувань спостерігається концентрація балів у межах 85–100, що вказує на стабільну тенденцію: чим більше студент відвідує заняття, тим вищою є його успішність.

Residuals:

Min	1Q	Median	3Q	Max
-28.552	-10.041	-1.537	9.224	28.463

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	71.53732	0.32264	221.73	<2e-16
Кількість_відвідувань	0.56295	0.03097	18.18	<2e-16

Multiple R-squared: 0.1063, Adjusted R-squared: 0.106  
F-statistic: 330.5 on 1 and 2778 DF, p-value: < 2.2e-16

*Рисунок 3.1.5 Виведені результати лінійної моделі*

Побудована модель регресії має наступний вигляд:

$$\text{Бали студента} = 71.54 + 0.56 \times \text{Кількість відвідувань}$$

Це означає, що кожне додаткове відвідування заняття асоціюється зі збільшенням середнього балу приблизно на 0.56 бала, за інших рівних умов. Коефіцієнт при змінній “Кількість\_відвідувань” є статистично значущим ( $p < 2e-16$ ), що підтверджує достовірність виявленої залежності.

Коефіцієнт детермінації ( $R^2$ ) становить 0.106, тобто модель пояснює приблизно 10.6% варіації у балах студентів. Це свідчить про наявність помірного, але стабільного зв'язку між відвідуваністю та академічними результатами. Стандартна похибка залишків дорівнює 11.6 бала, що вказує на розкиданість значень навколо регресійної прямої - типове явище для поведінкових та освітніх даних.

Таким чином, результати моделювання дозволяють зробити висновок про існування статистично значущої, хоча й не абсолютно сильної, позитивної залежності між кількістю відвідувань занять та середнім балом студента. Цей фактор може бути врахований у системах прогнозування успішності студентів на основі освітньої аналітики, проте його варто аналізувати в сукупності з іншими важливими показниками.

### 3.2 Сегментація студентів за профілями активності та успішності

З метою поглибленого аналізу поведінкових моделей студентів у контексті навчальної активності та академічної успішності було застосовано метод кластеризації k-середніх (k-means). Для побудови моделі кластеризації було використано два ключові параметри: кількість відвідувань занять студентом та його середній бал за курс. Попередньо з вибірки було виключено спостереження з нульовими середніми балами, оскільки вони, як правило, пов'язані з академічною заборгованістю або відсутністю оцінювання, що могло б суттєво вплинути на точність розподілу за кластерами.

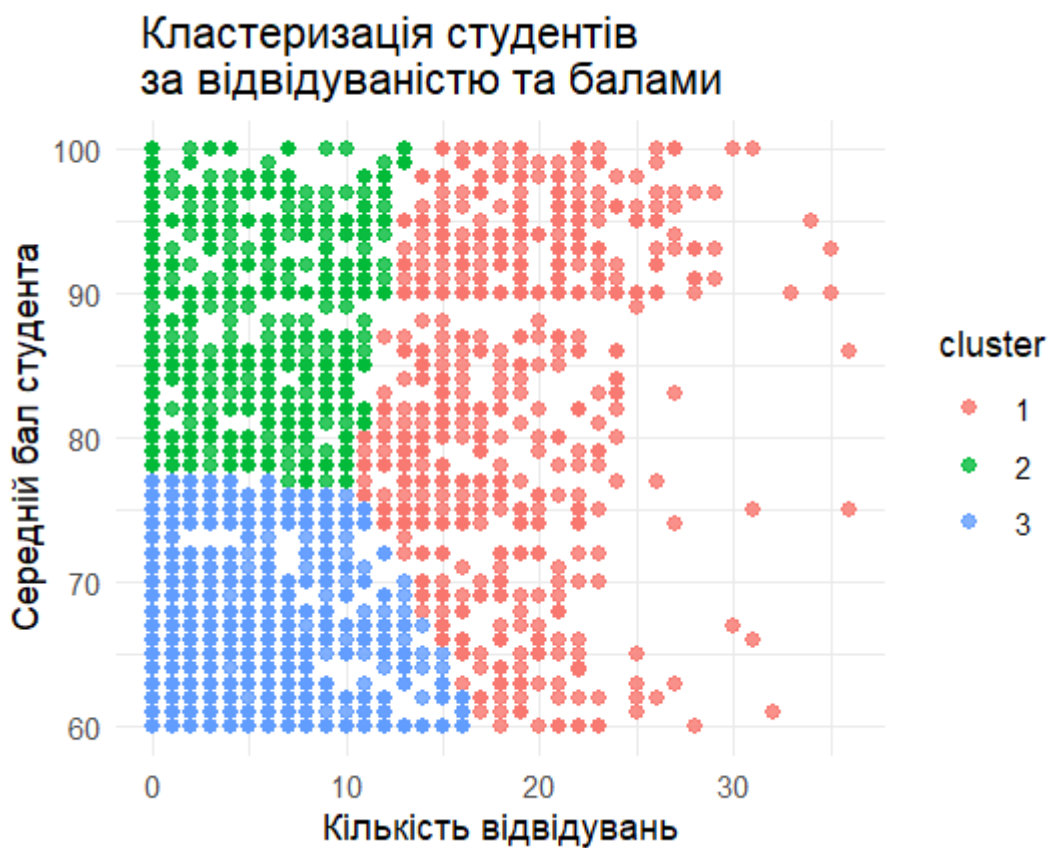


Рисунок 3.2.1 Графік кластеризації студентів за відвідуваністю та балами

Результати кластерного аналізу дозволили виділити три основні групи студентів, які демонструють характерні відмінності у своїй навчальній поведінці. Перший кластер позначений червоним кольором включає переважну більшість студентів з високим рівнем відвідуваності та середніми або високими

результатами навчання. Така конфігурація свідчить про те, що регулярна присутність на заняттях зазвичай корелює з кращими академічними показниками. У цій групі також трапляються студенти, які, незважаючи на високу відвідуваність, мають нижчі бали, що може вказувати на наявність додаткових впливових чинників, таких як складність предмету, особливості засвоєння матеріалу або рівень особистої мотивації.

Другий кластер позначений зеленим кольором представлений студентами з низькою відвідуваністю, але високими середніми балами. Така категорія демонструє потенціал до самостійного опанування матеріалу та, ймовірно, активно використовує альтернативні освітні ресурси, зокрема електронні курси, відеолекції або інші форми самонавчання. Наявність цього кластеру є показовою, оскільки він руйнує стереотип про обов'язковість тісного зв'язку між відвідуваністю та академічним успіхом, а також свідчить про потребу у диференційованому підході до оцінювання освітньої ефективності.

Третій кластер позначений синім кольором об'єднує студентів з низьким рівнем відвідуваності та відносно низькими балами. Ця група становить найбільший ризик з точки зору академічної неуспішності та може вимагати цілеспрямованого педагогічного супроводу. Причини такої поведінки можуть включати низьку мотивацію, відсутність доступу до ресурсів, зовнішні обставини або інші індивідуальні бар'єри, що перешкоджають ефективному навчанню.

Візуалізація результатів кластеризації чітко ілюструє розподіл студентів за трьома основними групами: зелений кластер сконцентрований у верхньому лівому секторі діаграми (низька відвідуваність, високі бали), червоний – праворуч (висока відвідуваність і високі/середні бали), а синій – у нижньому лівому секторі (низька відвідуваність і низькі бали). Така диференціація дозволяє краще зрозуміти різноманіття навчальних стратегій студентів та потенційні зони ризику.

На основі проведеного аналізу можна зробити висновок про наявність статистично обґрунтованих моделей поведінки у навчальному процесі.

Встановлено, що регулярна відвідуваність переважно супроводжується високими або задовільними результатами, проте існує частка студентів, які демонструють високу академічну успішність навіть за низької присутності на заняттях. Водночас, студенти з низькою відвідуваністю та низькими балами потребують окремого педагогічного втручання, зокрема індивідуального менторства, консультацій або адаптаційних заходів у межах системи підтримки навчання. Застосування кластерного підходу в освітній аналітиці дозволяє не лише виявляти типові сценарії студентської поведінки, а й оптимізувати стратегії підтримки успішності, орієнтуючись на конкретні потреби кожної групи.

З огляду на виявлені особливості поведінкових груп студентів, доцільним є впровадження низки заходів, спрямованих на підвищення ефективності освітнього процесу та покращення академічних результатів.

По-перше, студентам, які демонструють високі результати за умови низької відвідуваності, варто запропонувати індивідуалізовані освітні траєкторії з використанням гнучких форм навчання. Це може включати асинхронні формати проходження курсу, посилений акцент на самостійну роботу або онлайн-доступ до навчальних матеріалів.

По-друге, для студентів, які належать до групи з підвищеним ризиком академічної неуспішності (низька відвідуваність та низький середній бал), доцільно реалізовувати комплекс заходів підтримки. Йдеться про впровадження менторських програм, проведення роз'яснювальної роботи щодо важливості регулярного навчання, а також забезпечення доступу до додаткових ресурсів, консультацій і індивідуальних завдань.

По-третє, з метою підвищення мотивації до відвідування занять, доцільно впроваджувати стимулювальні заходи. Це можуть бути бонусні оцінки за активну участь, використання інтерактивних освітніх технологій, активний зворотний зв'язок від викладачів та інші форми залучення студентів до освітнього процесу.

Крім того, рекомендовано впровадження системи постійного моніторингу відвідуваності у реальному часі. Такий підхід дозволить оперативно виявляти

тенденції до зниження активності та своєчасно реагувати шляхом персоналізованих втручань з боку адміністрації або викладацького складу.

У рамках дослідження було проведено однофакторний дисперсійний аналіз (ANOVA) для перевірки статистичної гіпотези щодо впливу джерела фінансування на академічні результати студентів. Незалежною змінною виступав фактор «Джерело фінансування», який включає дві категорії - «Бюджет» та «Контракт». Залежною змінною було обрано показник «Середній бал студента», обчислений на основі даних LMS за один семестр навчального року 2024–2025.

#### Методика дослідження

Дисперсійний аналіз дозволяє оцінити, чи спостерігається статистично значуща різниця між середніми значеннями успішності в двох незалежних групах. Для цього порівнюється міжгрупова варіативність із внутрішньогруповою за допомогою F-критерію Фішера.

Результати проведеного аналізу наведено в таблиці 3.2.2.

Джерело варіації	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Джерело фінансування	1	1106333	1106333	747.2	< 2e-16
Залишки (residuals)	74425	110192413	1481		

Таблиця 3.2.2 - Результати дисперсійного аналізу (ANOVA) щодо впливу джерела фінансування на успішність студентів

#### Інтерпретація результатів

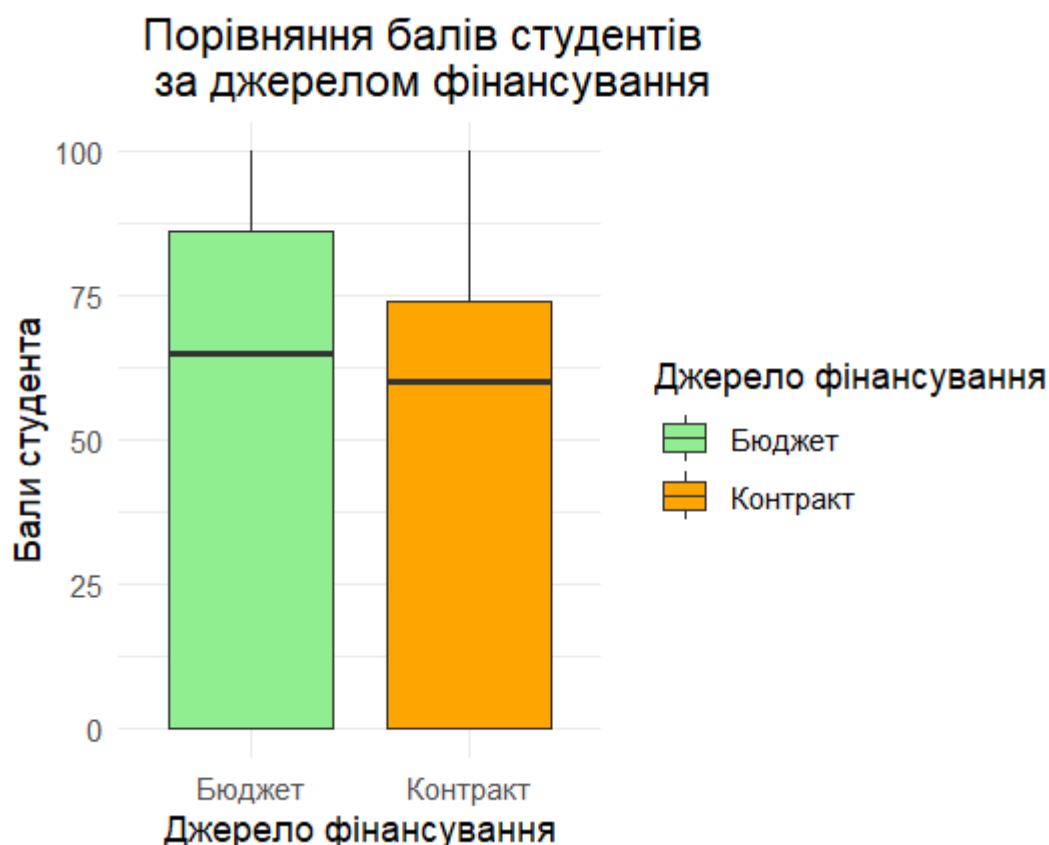
Отримане значення F-критерію (747.2) є надзвичайно високим, що вказує на суттєві відмінності між середніми балами у двох групах. Значення p-рівня < 2e-16 є статистично значущим на всіх традиційних рівнях значущості ( $\alpha = 0.05$ ;  $\alpha = 0.01$ ;  $\alpha = 0.001$ ). Це означає, що ймовірність випадкового виникнення такої різниці між групами практично дорівнює нулю.

Таким чином, нулева гіпотеза про відсутність впливу джерела фінансування на академічну успішність студентів - відхиляється. Можна з

високим ступенем упевненості стверджувати, що тип фінансування пов'язаний із результатами навчання студентів.

#### Візуалізація результатів

Для наочності побудовано стовпчикову діаграму, яка відображає середні бали студентів залежно від типу фінансування (рис. 3.2.3).



*Рисунок 3.2.3 - Порівняння середнього балу студентів за джерелом фінансування*

Графік підтверджує числові результати: студенти бюджетної форми навчання в середньому демонструють вищі академічні показники, ніж їхні колеги, які навчаються на контрактній основі. Це спостереження може мати кілька пояснень:

Виявлена різниця в успішності між студентами бюджетної та контрактної форм навчання може бути зумовлена низкою факторів. Зокрема, студенти, які навчаються за державним замовленням, зазвичай мають вищі конкурсні бали при вступі, що свідчить про початково сильнішу академічну підготовку. Крім того, бюджетна форма передбачає певні зобов'язання перед державою, що може

посилювати мотивацію до навчання та відповідальність за власні результати. Натомість студенти контрактної форми часто змушені поєднувати навчання з роботою через фінансові обставини, що, у свою чергу, зменшує час і ресурси, які вони можуть присвячувати академічній діяльності.

Отримані результати підтверджують гіпотезу про те, що джерело фінансування має статистично значущий вплив на академічну успішність студентів. Надалі доцільно поглибити аналіз із урахуванням додаткових факторів (конкурсний бал, дисципліна, стать, форма навчання), а також дослідити внутрішньогрупову варіативність для виявлення потенційних винятків або підгруп із нетиповою поведінкою.

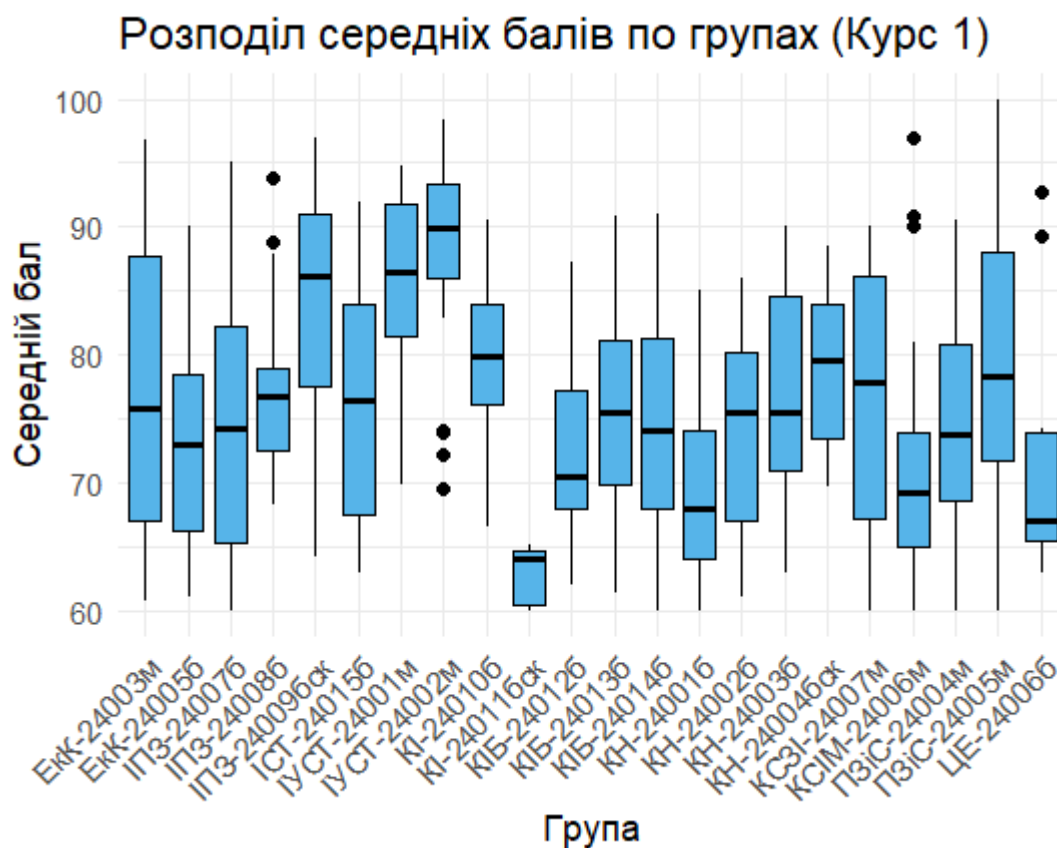
### **3.3 Візуалізація та аналіз розподілу академічної успішності за допомогою боксплотів**

У цьому підрозділі здійснено візуалізацію розподілу середніх балів студентів перших та четвертих курсів із різних спеціальностей з використанням боксплотів (boxplots) - одного з найбільш інформативних статистичних інструментів описової аналітики. Побудова боксплотів дозволяє одночасно оцінити як централізовані тенденції (медіану), так і варіативність, асиметрію розподілу та наявність викидів. Це дає змогу глибше зрозуміти структуру академічної успішності в межах кожної академічної групи та між ними.

На рисунку 3.3.1 подано боксплоти середніх балів студентів першого курсу за різними академічними групами. Горизонтальна вісь репрезентує назви груп (спеціальностей), вертикальна - середній бал. Візуалізовані п'ять основних статистичних характеристик: мінімальне значення, перший квартиль, медіана, третій квартиль та максимальне значення. Окремі точки за межами вусів відображають викиди - аномальні значення, які істотно відрізняються від основної маси спостережень.

Аналізуючи дані першого курсу, можна відзначити значну варіативність середніх балів між групами вже на початковому етапі навчання. Зокрема, окремі спеціальності, такі як ІУСТ-24002м, демонструють відносно високі медіанні бали та щільний розподіл, що свідчить про стабільну академічну підготовку

студентів. Натомість інші групи, як-от ЕкК-24003м чи ПЗ-24005б, характеризуються розлогим міжквартильним розмахом та значною кількістю викидів, що свідчить про нерівномірний рівень засвоєння матеріалу всередині груп. Загалом медіанні значення варіюються від приблизно 65 до 90 балів, що є ознакою широкого спектра академічної успішності серед студентів, які тільки розпочали навчання у вищому закладі освіти.

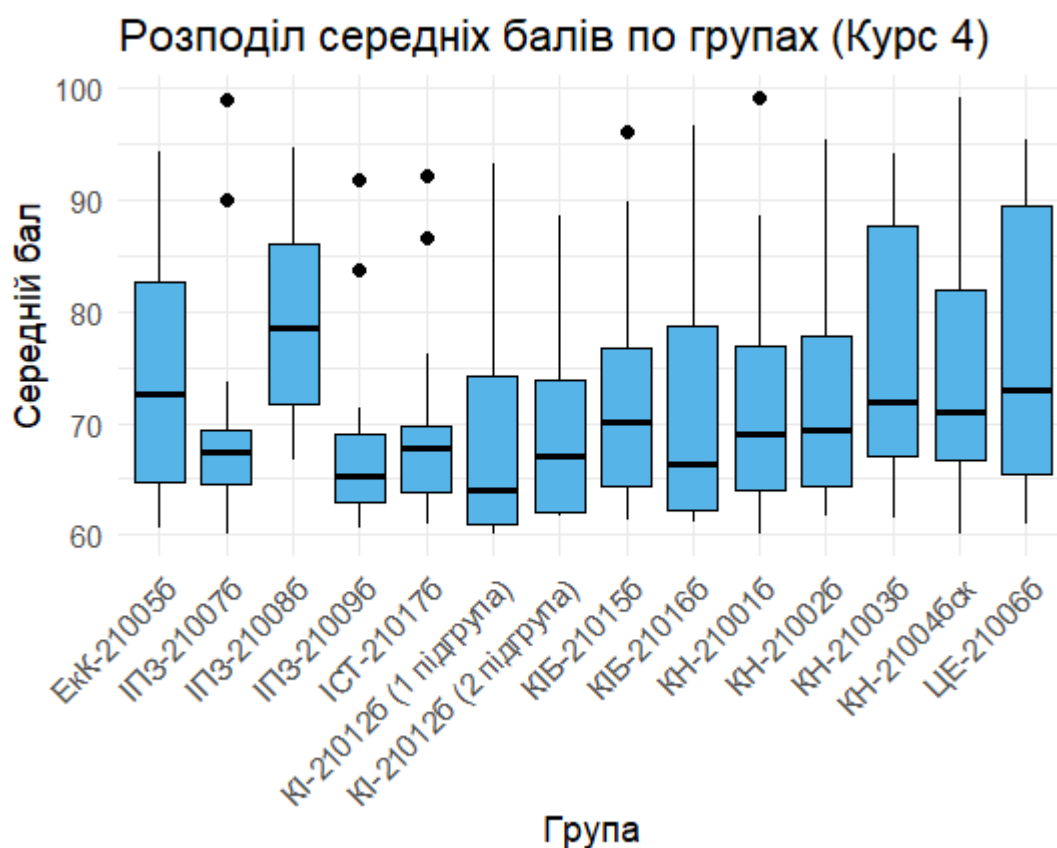


*Рисунок 3.3.1 - Розподіл середніх балів студентів 1-го курсу за академічними групами (боксплоти)*

На рисунку 3.3.2 наведено аналогічний розподіл середніх балів, але вже для студентів четвертого курсу. Візуальне порівняння дозволяє зробити низку важливих спостережень. По-перше, у частини груп спостерігається стабілізація успішності: розподіл стає компактнішим, а медіанні значення зберігають сталість. Це характерно, зокрема, для груп КН-210016 та КН-210026, де можна припустити стабільне академічне середовище та ефективну навчальну динаміку. Водночас у деяких спеціальностях, таких як ПЗ-21007б чи ІСТ-21009б,

спостерігаються коливання як у значеннях медіани, так і в розширенні міжквартильного розмаху, що може свідчити про нерівномірну академічну траєкторію студентів протягом навчання.

У студентів четвертого курсу також фіксується наявність викидів, причому вони трапляються як у верхній, так і нижній частині розподілу. Це свідчить про індивідуальні траєкторії розвитку студентів - зокрема, наявність як дуже сильних, так і тих, хто суттєво відстає від загального рівня групи. Причинами такої динаміки можуть бути як зовнішні обставини (наприклад, поєднання навчання з роботою або особисті труднощі), так і внутрішні освітні чинники (відсутність академічної підтримки, зниження мотивації, або складність фінального етапу навчання).



*Рисунок 3.3.2 - Розподіл середніх балів студентів 4-го курсу за академічними групами (боксплоти)*

Причини виявлених відмінностей в академічній успішності між спеціальностями та між курсами можуть мати багатовимірну природу. Одним із

головних чинників є рівень підготовки абітурієнтів на етапі вступу: спеціальності з вищим конкурентним балом часто залучають студентів з вищим академічним потенціалом. Крім того, істотне значення має структура навчального плану, інтенсивність дисциплін, кількість практичних занять і форма проведення оцінювання, які можуть відрізнятися між освітніми програмами. Не менш важливими є психолого-педагогічні фактори, зокрема мотивація до навчання, адаптація до освітнього середовища, а також наявність внутрішньої та зовнішньої підтримки. На четвертому курсі, як правило, спостерігається або підвищення академічної відповідальності, або зниження зацікавленості у зв'язку із фокусом на працевлаштування чи завершення навчання. Крім того, відбір студентів у процесі навчання (відрахування, переривання) формує нову "усереднену" вибірку до випуску, що також впливає на форму та характер розподілу.

З метою покращення академічної успішності доцільно проводити регулярний аналіз навчальних планів, ефективності викладацьких стратегій і запровадження інноваційних форм навчання. Важливо організувати систему раннього виявлення студентів, які демонструють ризик низької успішності, та надавати їм індивідуальну академічну підтримку. Обмін педагогічними практиками між кафедрами й освітніми програмами, а також впровадження інтерактивного та адаптивного навчання можуть сприяти підвищенню рівня засвоєння знань та зменшенню дисперсії результатів. Результати, отримані за допомогою боксплотів, підтверджують необхідність такого підходу - адже глибокий аналіз варіативності в навчальних досягненнях створює базу для прийняття стратегічних управлінських рішень у сфері якості освіти.

### **3.4 Побудова та інтерпретація моделей прогнозування академічної успішності**

З метою виявлення ключових чинників, що впливають на академічну успішність здобувачів вищої освіти, було побудовано лінійну регресійну модель, яка дозволяє прогнозувати середній бал студента залежно від низки соціально-демографічних та освітніх характеристик. До аналізу включено лише ті

спостереження, в яких студент мав позитивні оцінки за всіма дисциплінами, що дозволило уникнути спотворення результатів через відсутність або формальний характер оцінювання.

Модель враховує такі змінні: конкурсний бал при вступі (КБ), джерело фінансування (бюджет або контракт), курс навчання (від 1 до 4), стать студента (чоловіча або жіноча), а також перехоплення - базове значення, що відповідає студентці першого курсу на бюджетній формі навчання з нульовим конкурсним балом. Таблиця 3.4.1 подає оцінки коефіцієнтів моделі разом із відповідними статистичними характеристиками.

<b>term</b>	<b>estimate</b>	<b>std.error</b>	<b>statistic</b>	<b>p.value</b>
Джерело фінансування: Бюджет	0.0000	NA	NA	NA
Джерело фінансування: Контракт	-3.7109	0.5428	-6.8365	0
КБ при вступі	0.1335	0.0165	8.0768	0
Курс: 1	0.0000	NA	NA	NA
Курс: 2	-3.0717	0.6066	-5.0641	0
Курс: 3	-4.9114	0.6155	-7.9795	0
Курс: 4	-5.1730	0.6361	-8.1325	0
Перетин (Intercept)	65.9896	2.7255	24.2118	0
Стать: Жіноча	0.0000	NA	NA	NA
Стать: Чоловіча	-6.2804	0.5828	-10.7766	0

*Таблиця 3.4.1 - Результати регресійного аналізу академічної успішності студентів (без нульових оцінок)*

Результати побудованої регресійної моделі демонструють високий рівень статистичної значущості більшості включених змінних, що підтверджується нульовими або близькими до нуля р-значеннями. Це свідчить про надійність отриманих оцінок та дозволяє зробити низку обґрунтованих висновків щодо чинників, які впливають на академічну успішність студентів.

Зокрема, одним із ключових предикторів виявився конкурсний бал при вступі: кожен додатковий бал КБ асоціюється із зростанням середнього навчального балу приблизно на 0.13 бала. Це означає, що студенти з вищими вступними результатами, як правило, демонструють кращу успішність протягом навчання, що є підтвердженням ролі початкового академічного капіталу у формуванні освітньої траєкторії.

Аналіз змінної «джерело фінансування» засвідчив, що студенти контрактної форми навчання, за інших рівних умов, отримують у середньому на 3.71 бала менше, ніж студенти, які навчаються за бюджетною формою. Це може бути зумовлено відмінностями у рівні попередньої підготовки, навчальній мотивації або академічних очікуваннях.

Важливим фактором виявився і курс навчання: студенти другого курсу мають середній бал на 3.07 бала нижчий, третього - на 4.91 бала, а четвертого - на 5.17 бала нижчий порівняно з першокурсниками. Така тенденція може бути наслідком ускладнення навчальної програми на старших курсах, появи додаткових навантажень або зниження інтересу до формальної освіти.

Значущим чинником виявилася й стать студента: чоловіки в середньому мають на 6.28 бала нижчу академічну успішність, ніж жінки. Такий розрив підтверджується численними дослідженнями, які вказують на стабільно вищі результати жінок у формальній освіті, зокрема завдяки кращій дисциплінованості, плануванню та відповідальному ставленню до навчання.

Коефіцієнт перехоплення (Intercept) становить 65.99 і є умовною базою для обчислення середніх балів студентів з різними характеристиками. Він інтерпретується як очікуване значення середнього балу для студентки першого курсу, яка навчається на бюджеті та має нульовий конкурсний бал. Попри те, що така ситуація є суто гіпотетичною (КБ = 0 практично не трапляється), цей параметр є важливим для коректного тлумачення решти змінних моделі.

Водночас наявність значень NA у стандартних помилках, t-статистиках і р-значеннях для певних змінних пояснюється тим, що ці категорії є базовими у використаному факторному кодуванні. До таких належать жіноча стать,

бюджетна форма навчання та перший курс. Їхні коефіцієнти встановлюються на рівні нуля, а всі інші категорії оцінюються як відхилення від цих референтних груп.

Отримана модель має практичне значення для управлінських рішень у сфері вищої освіти. Зокрема, вона може бути використана для ранньої діагностики потенційних академічних труднощів, ще на етапі вступу. Наприклад, студенти-чоловіки контрактної форми навчання з низьким конкурсним балом та високим курсом можуть бути віднесені до групи ризику, що потребує додаткової академічної підтримки. Такий підхід відповідає принципам адаптивної та орієнтованої на студента освіти, сприяючи реалізації індивідуальних освітніх траєкторій та забезпеченню рівного доступу до якісної освіти.

Окрім побудови та інтерпретації регресійної моделі, в рамках дослідження було реалізовано інструмент для практичного використання результатів моделі прогнозування - написано програмний код на мові R, який дає змогу розрахувати прогнозовану оцінку студента за певною дисципліною на основі його індивідуальних характеристик. Таким чином, користувач може самостійно задати значення ключових змінних - конкурсного балу при вступі, статі, джерела фінансування, курсу та конкретної навчальної дисципліни - і отримати прогноз середнього балу на основі побудованої регресійної моделі.

Наприклад, для моделювання було задано такі параметри (повний код в додатках):

```
# Прогноз на основі лінійної регресії з дисципліною
new_data_reg <- data.frame(
  КБ_при_вступі = 168,6,
  Стать = factor("чоловіча", levels = levels(agg_data_unique$Стать)),
  Джерело_фінансування = factor("Контракт", levels = levels(agg_data_unique$Джерело_фінансування)),
  Курс = factor(4, levels = levels(agg_data_unique$Курс)),
  Дисципліна = factor("Імітаційне моделювання", levels = levels(agg_data_unique$Дисципліна))
)

predicted_reg <- predict(regression_model, newdata = new_data_reg)
cat("Прогнозована оцінка (лінійна регресія):", round(predicted_reg, 2), "\n")

> cat("Прогнозована оцінка (лінійна регресія):", round(predicted_reg, 2), "\n")
Прогнозована оцінка (лінійна регресія): 73.25
```

Рисунок 3.4.2 - Фрагмент коду з прогнозуванням (1)

У цьому прикладі йдеться про студента чоловічої статі, який навчається на 4 курсі за контрактною формою з конкурсний балом 168.6, а прогноз стосується його оцінки з дисципліни «Імітаційне моделювання». Результатом виконання коду стала прогнозована оцінка 73.25 бала, що є цілком реалістичним показником для такого профілю студента відповідно до виявлених залежностей у моделі.

Для порівняння було розраховано прогноз і для кардинально протилежного випадку - студентки бюджетної форми навчання на 1 курсі з високим конкурсним балом (187) та дисципліною «Окремі розділи математичної статистики»:

```
# Прогноз на основі лінійної регресії з дисципліною
new_data_reg <- data.frame(
  КБ_при_вступі = 187,
  Стать = factor("Жіноча", levels = levels(agg_data_unique$Стать)),
  Джерело_фінансування = factor("Бюджет", levels = levels(agg_data_unique$Джерело_фінансування)),
  Курс = factor(1, levels = levels(agg_data_unique$Курс)),
  Дисципліна = factor("Окремі розділи математичної статистики", levels = levels(agg_data_unique$Дисципліна))
)

predicted_reg <- predict(regression_model, newdata = new_data_reg)
cat("Прогнозована оцінка (лінійна регресія):", round(predicted_reg, 2), "\n")

> cat("Прогнозована оцінка (лінійна регресія):", round(predicted_reg, 2), "\n")
Прогнозована оцінка (лінійна регресія): 90.95
```

*Рисунок 3.4.3 - Фрагмент коду з прогнозуванням (2)*

У цьому випадку, студентка з максимально сприятливими характеристиками отримує прогнозовану оцінку 90.95 бала, що суттєво вище за попередній приклад. Така різниця в результатах підтверджує значущість кожної з включених у модель змінних - конкурсного балу, статі, форми фінансування та курсу - і водночас свідчить про точність моделі в межах вибірки.

Таким чином, розроблений код надає зручний інструмент для персоналізованого прогнозування академічної успішності, що може бути використаний як студентами, так і адміністрацією навчального закладу для діагностики, профілактики ризиків та прийняття обґрунтованих освітніх рішень.

Далі розглянуто результати аналізу важливості змінних, отриманих за допомогою моделі Random Forest. Дана модель не лише дозволяє здійснювати

точне прогнозування, але й надає можливість оцінити внесок кожної змінної у загальну точність прогнозу. Для цього використовуються два основні показники:

**%IncMSE (Percentage Increase in Mean Squared Error)** - відсоткове зростання середньоквадратичної помилки моделі при виключенні змінної. Чим вищий цей показник, тим важливішою є змінна для точності прогнозу.

**IncNodePurity (Increase in Node Purity)** - показник, який відображає загальне покращення "чистоти" вузлів дерева при використанні конкретної змінної. Вищі значення свідчать про більшу інформативність ознаки.

	%IncMSE	IncNodePurity
КБ_при_вступі	26.168005	148559.355
Стать	8.534915	9927.379
Джерело_фінансування	25.512256	41150.632
Курс	30.909522	154617.832
Дисципліна	27.855689	602390.596

*Рисунок 3.4.3 – Результати аналізу моделі Random Forest*

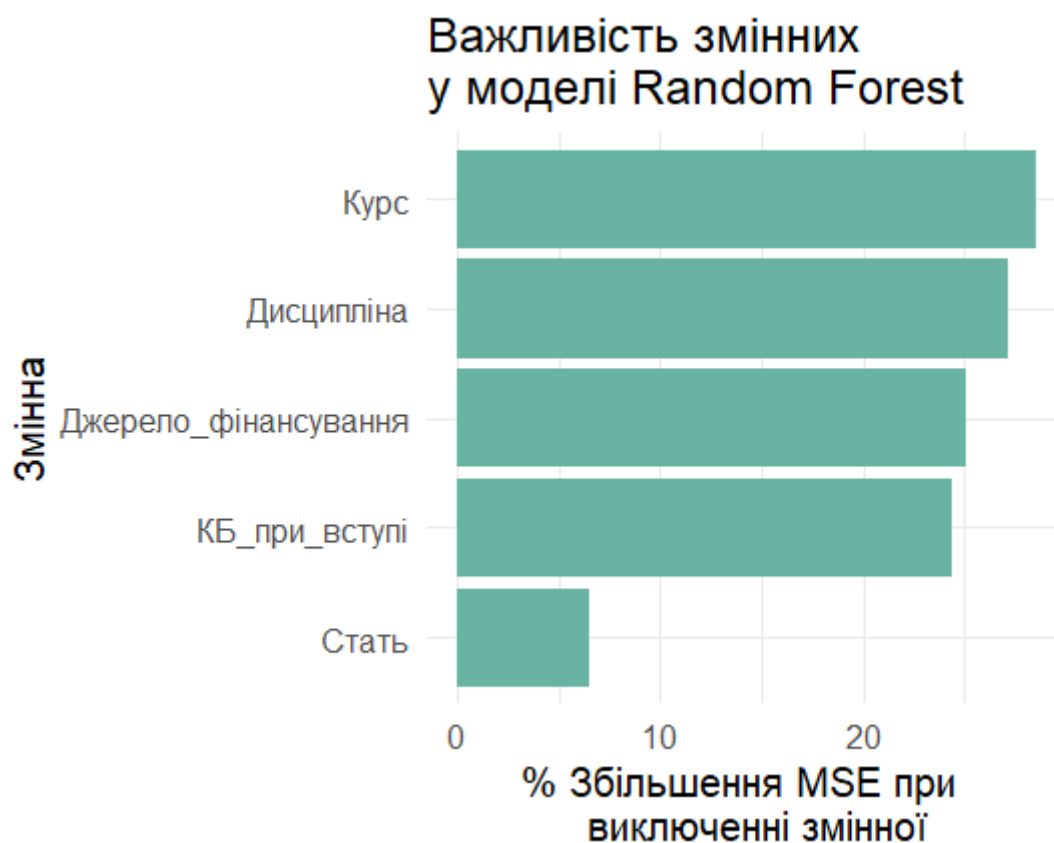
Інтерпретація результатів важливості змінних:

Курс навчання (30.91% збільшення MSE) виявився найважливішим фактором. Це свідчить про те, що рік навчання суттєво впливає на середню оцінку студента. Збільшення складності дисциплін або підвищення вимог на старших курсах може пояснити цей ефект.

Дисципліна (27.86% IncMSE) є другою за важливістю змінною. Очевидно, що специфіка предмету суттєво впливає на оцінювання: одні дисципліни можуть бути об'єктивно складнішими або мати інший стиль викладання/оцінювання.

Конкурсний бал при вступі (26.17%) та джерело фінансування (25.51%) також мають значний вплив. Перший чинник відображає академічну підготовку студента на момент вступу, що логічно корелює з подальшою успішністю. Другий може вказувати на відмінності в мотивації чи підтримці з боку освітнього середовища.

Стать (8.53%) показала найменший вплив у цій моделі, що вказує на другорядну роль гендеру у формуванні оцінок порівняно з іншими змінними.



*Рисунок 3.4.4 – Графік важливості змінних*

На графіку важливості змінних (Рис. 3.4.4), побудованому за критерієм  $\%IncMSE$ , чітко видно, що курс навчання є домінуючим чинником. Це дає змогу припустити, що з кожним наступним роком студенти стикаються з дедалі більшими труднощами, які суттєво впливають на їхню академічну успішність.

Також видно значний внесок змінної дисципліна, що зумовлено різною складністю та особливостями викладання окремих навчальних предметів. Конкурсний бал при вступі є надійним предиктором, адже вищий бал зазвичай свідчить про кращу підготовку абітурієнта. Форма фінансування вказує на певні соціальні чи психологічні аспекти, які також можуть впливати на навчальний процес.

Натомість стать студента показала найменший вплив, що свідчить про гендерну нейтральність середньої оцінки у межах цієї вибірки та моделі.

Отже, модель Random Forest підтверджує, що основними предикторами академічної успішності студентів є курс навчання, конкретна дисципліна,

конкурсний бал при вступі та форма фінансування. Аналіз важливості змінних дає змогу краще зрозуміти, на які аспекти варто звертати увагу при формуванні освітньої стратегії, підборі навчального навантаження та плануванні підтримки для студентів.

Таким чином, проведений аналіз і побудовані моделі дозволяють не лише ідентифікувати ключові фактори, що впливають на академічну успішність студентів, але й практично застосувати отримані результати для підтримки освітнього процесу. Використання регресійної моделі в поєднанні з інструментом прогнозування дає змогу персоналізувати оцінювання та вчасно виявляти студентів із ризиком низької успішності, що сприяє ефективнішому розподілу ресурсів і розвитку адаптивних освітніх стратегій.

Аналіз важливості змінних за допомогою моделі Random Forest підтвердив домінування таких чинників, як курс навчання, дисципліна, конкурсний бал і форма фінансування, в той час як гендерний аспект виявився менш суттєвим. Це вказує на необхідність фокусування уваги на академічних і організаційних параметрах, що формують умови навчання і підтримки студентів.

Отже, комплексний підхід, що поєднує статистичний аналіз із сучасними методами машинного навчання, відкриває нові можливості для підвищення якості вищої освіти через більш точне прогнозування і цілеспрямовану роботу з потенційними проблемами успішності студентів.

## ВИСНОВКИ

У межах бакалаврської роботи було побудовано комплексну економіко-математичну модель для аналізу та прогнозування академічної успішності студентів на основі освітньої аналітики з використанням даних з LMS (Learning Management System), а також додаткових параметрів, отриманих із систем управління освітнім процесом.

На першому етапі було визначено склад моделі, до якої увійшли як класичні змінні (конкурсний бал вступу, джерело фінансування, середній бал, кількість боржників), так і поведінкові (відвідуваність, активність у LMS, час навчальних занять, тощо). Аналіз середніх значень успішності та динаміки балів студентів групи вступу 2021 року протягом шести семестрів дозволив виявити стійку тенденцію до зниження академічних показників. Це може свідчити про зменшення навчальної мотивації, перевантаження програмами або зовнішній вплив (зміни в умовах навчання, карантин, стрес тощо).

Проведений кореляційний аналіз підтвердив існування статистично значущих зв'язків між рядом змінних. Зокрема, виявлено сильну позитивну кореляцію між конкурсним балом та середнім рейтингом студента. Також простежено зв'язок між джерелом фінансування та успішністю: студенти-бюджетники, як правило, демонструють вищі результати. Це може свідчити про необхідність додаткової підтримки для студентів на контрактній формі.

Крім того, було проведено кластеризацію студентів за ознаками відвідуваності та успішності, що дозволило сегментувати аудиторію на групи з різними освітніми ризиками. Подібний підхід забезпечує можливість для цільової педагогічної інтервенції - наприклад, організації додаткових занять для груп із підвищеним ризиком академічної неуспішності.

Аналіз структури успішності за окремими дисциплінами з урахуванням боржників виявив, що на деяких предметах значна кількість студентів має академічну заборгованість. Це призводить до зниження загального середнього балу та вказує на потенційні проблеми в методиці викладання, навантаженні чи доступності навчальних матеріалів.

Особливу увагу приділено вивченню факторів, які безпосередньо впливають на академічну успішність. Встановлено, що активність у системі LMS (виконання завдань, кількість спроб, своєчасність здачі), регулярна відвідуваність занять, а також час проведення занять (наприклад, ранкові пари мають нижчу відвідуваність) є контрольованими змінними з високим впливом. Отримані результати дозволяють зробити висновки щодо доцільності більш гнучкого розкладу або змішаного формату занять.

У рамках роботи побудовано дві основні моделі прогнозування: на основі лінійної регресії та Random Forest. Перша дозволила чітко простежити зв'язки між окремими змінними та академічними результатами, тоді як друга - ідентифікувати найважливіші предиктори успішності з урахуванням взаємодії змінних і нелінійних ефектів. Візуалізація даних через боксплоти допомогла виявити відхилення та особливості розподілу балів серед студентів.

Модель має трикомпонентну структуру:

Аналітична частина (описово-статистичний аналіз успішності, боргів, відвідуваності);

Моделююча частина (регресійне моделювання та кластеризація);

Прогнозна частина (передбачення академічної успішності на основі історичних даних).

У результаті було доведено, що успішність студентів є багатофакторною категорією, яка формується під впливом як контрольованих (відвідуваність, активність, форма навчання), так і неконтрольованих (психоемоційний стан, сімейні обставини), а також випадкових змінних (умови проведення іспитів, карантин тощо).

Практичне значення отриманих результатів полягає в можливості впровадження системи раннього моніторингу студентів із ризиками академічної неуспішності. На основі кластеризаційного аналізу та даних освітньої платформи LMS можна ідентифікувати окремі групи студентів, що потребують підвищеної уваги з боку викладачів та адміністрації. Така система дозволить виявляти

потенційні труднощі ще на ранніх етапах навчання та вчасно реагувати на проблеми.

Важливим напрямом є персоналізоване планування академічної підтримки для студентів, які мають борги з окремих дисциплін. Використання даних про успішність, динаміку балів та відвідуваність дозволяє формувати індивідуальні маршрути підтримки, що можуть включати додаткові консультації, наставництво або адаптовані форми контролю знань. Це сприятиме підвищенню рівня засвоєння матеріалу й зменшенню кількості академічних заборгованостей.

Також рекомендовано переглянути та оптимізувати розклад занять з урахуванням виявленого впливу часу доби на рівень відвідуваності. Результати дослідження свідчать про те, що ранкові пари характеризуються нижчими показниками присутності студентів, тому можливе зміщення деяких занять на пізніший час або впровадження більш гнучких форматів (наприклад, змішаного навчання) з урахуванням фізіологічних та поведінкових особливостей студентської аудиторії.

Окремої уваги потребує вдосконалення системи збору освітньої аналітики. Доцільним є розширення набору відстежуваних параметрів шляхом включення поведінкових і контекстних змінних, таких як тривалість роботи з матеріалами в LMS, час здачі завдань, активність у форумах, реакція на зворотний зв'язок, а також зовнішні обставини, які можуть впливати на навчальний процес. Це підвищить точність моделі та дозволить будувати більш чутливі прогнози щодо успішності.

У перспективі розроблена модель може бути інтегрована в загальну інформаційну систему університету, що дозволить автоматизувати процеси збору, аналізу й візуалізації освітніх даних. Крім того, модель є достатньо гнучкою для адаптації під різні освітні програми або спеціальності. Для подальшого вдосконалення передбачається розширення переліку змінних за рахунок включення результатів психометричних опитувань, даних про рівень цифрової грамотності та показників позааудиторної активності студентів.

Таким чином, побудована модель є не лише засобом аналізу вже наявних даних, але й ефективним інструментом для формування системи прогностичної освітньої аналітики. Її використання сприятиме підвищенню якості управлінських рішень в освітньому процесі, своєчасному реагуванню на ризики та ефективнішому плануванню академічної підтримки.

## СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Learning Analytics [Електронний ресурс] // ResearchGate. – Режим доступу: [https://www.researchgate.net/publication/365405659\\_Learning\\_Analytics](https://www.researchgate.net/publication/365405659_Learning_Analytics) (дата звернення: 15.12.2024).
2. High-Level Architecture for Learning Analytics [Електронний ресурс] // SolarResearch. – Режим доступу: <https://solaresearch.org/wp-content/uploads/2017/05/hla17.pdf> (дата звернення: 28.12.2024).
3. Susnjak T., Ramaswami G.S., Mathrani A. Learning analytics dashboard: a tool for providing actionable insights to learners // Int J Educ Technol High Educ. 2022. Vol. 19. №12. <https://doi.org/10.1186/s41239-021-00313-7> (дата звернення: 19.01.2025).
4. Learning Management Systems in a Changing Environment [Електронний ресурс] // ResearchGate. – Режим доступу: [https://www.researchgate.net/publication/267625754\\_Learning\\_Management\\_Systems\\_in\\_a\\_Changing\\_Environment](https://www.researchgate.net/publication/267625754_Learning_Management_Systems_in_a_Changing_Environment) (дата звернення: 10.02.2025).
5. Learning Management Systems in a Changing Environment [Електронний ресурс] // ResearchGate. – Режим доступу: [https://www.researchgate.net/publication/267625754\\_Learning\\_Management\\_Systems\\_in\\_a\\_Changing\\_Environment](https://www.researchgate.net/publication/267625754_Learning_Management_Systems_in_a_Changing_Environment) (дата звернення: 25.02.2025).
6. Moodle documentation: Grade book overview [Електронний ресурс] // Moodle. – Режим доступу: <https://moodle.org/mod/page/view.php?id=8174> (дата звернення: 03.03.2025).
7. Grade export in Moodle [Електронний ресурс] // Moodle Docs. – Режим доступу: [https://docs.moodle.org/500/en/Grade\\_export](https://docs.moodle.org/500/en/Grade_export) (дата звернення: 12.03.2025).
8. Danalyst J. Predicting student performance using machine learning in Excel [Електронний ресурс] // Medium. – Режим доступу: <https://medium.com/@JusticeDanalyst/predicting-student-performance-using-machine-learning-in-excel-7504fc9c1c97> (дата звернення: 24.03.2025).

9. Predicting Academic Success in Higher Education: Literature Review and Best Practices [Електронний ресурс] // ResearchGate. – Режим доступу: [https://www.researchgate.net/publication/338717237\\_Predicting\\_Academic\\_Success\\_in\\_Higher\\_Education\\_Literature\\_Review\\_and\\_Best\\_Practices](https://www.researchgate.net/publication/338717237_Predicting_Academic_Success_in_Higher_Education_Literature_Review_and_Best_Practices) (дата звернення: 02.04.2025).
10. Siemens G., Long P. Penetrating the Fog: Analytics in Learning and Education // EDUCAUSE Review. 2011. Vol. 46, № 5. С. 30–40. – Режим доступу: <https://er.educause.edu/articles/2011/9/penetrating-the-fog-analytics-in-learning-and-education> (дата звернення: 08.04.2025).
11. Галушак М. Ю. Дослідження методів прогнозування для оцінки успішності студентів університету за підсумками поточного навчання [Електронний ресурс] / М. Ю. Галушак // Чорноморський національний університет імені Петра Могили. – Режим доступу: [er.nau.edu.ua](http://er.nau.edu.ua) (дата звернення: 12.04.2025).
12. Predicting Student Performance from LMS Data: A Comparison of 17 Blended Courses Using Moodle LMS [Електронний ресурс] // IEEE Xplore Digital Library. – Режим доступу: <https://ieeexplore.ieee.org/> (дата звернення: 15.04.2025).
13. Коровайна В.С. Дослідження методів прогнозування для оцінки успішності студентів університету за підсумками поточного навчання [Електронний ресурс] / В.С. Коровайна // Харківський національний університет радіоелектроніки. – Режим доступу: <https://openarchive.nure.ua/handle/document/28737> (дата звернення: 18.04.2025).
14. Литвинова С. Г., Бідюк П. І. Використання машинного навчання для прогнозування успішності студентів [Електронний ресурс] // Інститут інформаційних технологій і засобів навчання НАПН України. – Режим доступу: <https://journal.iitta.gov.ua/> (дата звернення: 22.04.2025).
15. Yadav S. K., Pal S., Singh V. An Improved Approach for Predicting Student Performance Using Educational Data Mining [Електронний ресурс] // International Journal of Computer Applications. – Режим доступу: <https://www.ijcaonline.org/> (дата звернення: 25.04.2025).

16. Марчук О., Сафонова М. Аналітичні системи підтримки навчання на основі використання LMS Moodle [Електронний ресурс] // Наукові праці Миколаївського національного університету ім. В.О. Сухомлинського. – Режим доступу: <http://mdu.edu.ua/> (дата звернення: 28.04.2025).
17. Romero C., Ventura S. Educational Data Mining: A Review of the State of the Art [Електронний ресурс] // IEEE Transactions on Systems, Man, and Cybernetics. – Режим доступу: <https://ieeexplore.ieee.org/> (дата звернення: 30.04.2025).
18. Білоус О.В., Ткаченко М.А. Моделі прогнозування академічної успішності студентів на основі даних з LMS [Електронний ресурс] // Науковий вісник Чернівецького національного університету. – Режим доступу: <https://journals.chnu.edu.ua/> (дата звернення: 01.05.2025).
19. Aldowah H., Al-Samarraie H., Fauzy W.M. Educational data mining and learning analytics for 21st century higher education: A review and synthesis // Telematics and Informatics. 2019. Vol. 37. P. 13-49. doi:10.1016/j.tele.2019.01.007 (дата звернення: 29.04.2025).
20. Pal N., Dahiya O. Role of Learning Management System for Evaluating Students' progress in Learning Environment // Proceedings of 5th International Conference on Contemporary Computing and Informatics, IC3I 2022. P. 1800-1806 (дата звернення: 30.04.2025).
21. Ковальчук Ю. О. Пошук, отримання й аналіз даних в освіті: сучасний стан і перспективи розвитку. Інформаційні технології і засоби навчання, 2016. Том 50. № 6. С. 152–164. DOI: 10.33407/itlt.v50i6.1284 (дата звернення: 25.04.2025).
- 22.
23. Nikolaienko, S.M., Shynkaruk, V.D., Kovalchuk, V.I. & Kocharyan, A.B. Використання Big Data в освітньому процесі сучасного університету. Information Technologies and Learning Tools, 2017. №60(4), С.239. <https://doi.org/10.33407/itlt.v60i4.1681> (дата звернення: 25.04.2025).

24. Morze, N.V., Smyrnova-Trybulska, E., Glazunova, O. Design of a university learning environment for SMART education. *Smart Technology Applications in Business Environments*, pp. 221-248.
25. Петренко С. В. Оптимізація й аналіз результатів використання LMS Moodle у системі змішаного навчання в університеті. *Інформаційні технології і засоби навчання*, 2017. Т. 61, № 5. С. 140–150.
26. Drigas, A., Leliopoulos, P. The Use of Big Data in Education. *International Journal of Computer 2014. Science Issues*, 11, 5 (дата звернення: 25.04.2025).
- 27.
28. Bognár, L., Fauszt, T., Nagy, G.Z.: Analysis of Conditions for Reliable Predictions by Moodle Machine Learning Models. *International Journal of Emerging Technologies in Learning (iJET)*. 2021. №16(06), pp.106–121. doi:10.3991/ijet.v16i06.18347 (дата звернення: 17.03.2025).
29. Romero C., Ventura S. *Data mining in education // Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*. – 2013. – v. 3 (1), – P.12–27. (дата звернення: 17.03.2025).
30. Baker R., Siemens G. *Education Data Mining and Learning Analytics [Електронний ресурс]*. – Режим доступу : <http://www.columbia.edu/~rsb2162/BakerSiemensHandbook2013.pdf>. (дата звернення: 17.03.2025).
31. Baker R., Yacef K. The state of Educational data mining in 2009 : A review and future vision [Електронний ресурс] // *Journal of Educational Data Mining*. 2009. V. 1, № 1. – P. 3–17. Режим доступу : <http://www.educationaldatamining.org/JEDM/index.php/JEDM/article/download/8/2>. (дата звернення: 12.01.2025).
32. Скрипник А., Клименко Н., Костенко І. РІВЕНЬ ОСВІЧЕНОСТІ НАСЕЛЕННЯ В ГАЛУЗІ ЦИФРОВИХ ТЕХНОЛОГІЙ ТА ЗРОСТАННЯ ЕКОНОМІК КРАЇН. *Information Technologies and Learning Tools Інформаційні технології і засоби навчання 2020 Том 78№ 4 с.278-297* <https://doi.org/10.33407/itlt.v78i4.2948> (дата звернення: 28.02.2025).

33. Ye.V., Klymenko., Олена, Глазунова. (2024). Moodle big data analytics за допомогою power bi. Grail of science, 201-203. doi: 10.36074/grail-of-science.19.01.2024.035 (дата звернення: 29.12.2024).
34. Глазунова О., Клименко Є. ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ АНАЛІТИКИ ОСВІТНІХ ДАНИХ, Наука і техніка сьогодні. № 2(43) (2025). СЕРІЯ «Техніка» [https://doi.org/10.52058/2786-6025-2025-2\(43\)-1145-1155](https://doi.org/10.52058/2786-6025-2025-2(43)-1145-1155)
35. Hlazunova, O., Klymenko, N., Mokriiev, M., Nehrey, M., Klymenko, Y. (2025). Data Analysis Technologies for Enhanced Educational Processes: A Case Study Using the Moodle LMS. In: Hu, Z., Yanovsky, F., Dychka, I., He, M. (eds) Advances in Computer Science for Engineering and Education VII. ICCSEEA 2024 2024. Lecture Notes on Data Engineering and Communications Technologies, vol 242. Springer, Cham. [https://doi.org/10.1007/978-3-031-84228-3\\_58](https://doi.org/10.1007/978-3-031-84228-3_58) (дата звернення: 28.12.2024).

## ДОДАТКИ

### Код на мові R для аналізу відвідуваності та успішності студентів

```
r
Копировать Редактировать
setwd("D:/универ/4 курс/2 семестр/ДИПЛОМ/R")

install.packages("readxl")
library(readxl)
data <- read_excel("Відвідування_успішність_зведено_ФІТ_1семестр_повний.xlsx")

# Підготовка даних
data$Дата <- as.Date(data$Дата, format = "%d.%m.%Y")
data$Місяць <- as.numeric(format(data$Дата, "%m"))
data$Година <- format(data$Папа, "%H:%M")
data$Відмітка <- as.factor(data$Відмітка)

# Модель 1: Активність впливає на бали (без нульових балів)
filtered_data <- subset(data, `Бали студента` > 0)
modell_filtered <- lm(`Бали студента` ~ `Кількість входів на курс`, data =
filtered_data)
summary(modell_filtered)
plot(filtered_data$`Кількість входів на курс`, filtered_data$`Бали студента`,
xlab = "Кількість входів на курс", ylab = "Бали студента", pch = 19,
main = "Активність vs. Бали (без нульових балів)")
abline(modell_filtered, col = "blue", lwd = 2)

library(dplyr)
library(ggplot2)

# Модель 2: Вплив місяця і часу на ймовірність бути присутнім
data$Відвідуваність <- ifelse(data$Відмітка == "present", 1, 0)
agg <- data %>%
  group_by(Місяць, Година) %>%
  summarise(СерВідвідуваність = mean(Відвідуваність) * 100)

agg_filtered <- agg %>%
  mutate(Година_група = as.factor(sub(".*", "", as.character(Година)))) %>%
  filter(!Година_група %in% c("02", "11")) %>%
  group_by(Година_група, Місяць) %>%
  summarise(СерВідвідуваність = mean(СерВідвідуваність, na.rm = TRUE))

ggplot(agg_filtered, aes(x = Година_група, y = factor(Місяць), fill =
СерВідвідуваність)) +
  geom_tile(color = "white") +
  scale_fill_gradient2(low = "red", mid = "yellow", high = "green", midpoint =
50,
name = "Відвідуваність (%)") +
  labs(x = "Година пари", y = "Місяць", title = "Рівень відвідуваності по
місяцях і годинах") +
  theme_minimal()

# Модель 3: Вплив реальної відвідуваності на результат студента
agg_data <- data %>%
  group_by(КодОсоби, КодКурсу) %>%
  summarise(
    Кількість_відвідувань = sum(Відвідуваність),
    Бали_студента = mean(`Бали студента`)
  ) %>%
  filter(Бали_студента != 0)
```

```

model3 <- lm(Бали_студента ~ Кількість_відвідувань, data = agg_data)
summary(model3)

ggplot(agg_data, aes(x = Кількість_відвідувань, y = Бали_студента)) +
  geom_point(color = "#2c7fb8", size = 3, alpha = 0.7) +
  geom_smooth(method = "lm", color = "#f03b20", se = TRUE, linetype = "dashed")
+
  labs(
    title = "Залежність балів від кількості відвідувань",
    x = "Кількість відвідувань (present)",
    y = "Середній бал студента"
  ) +
  theme_minimal()

# Кластеризація студентів
agg_data_clean <- agg_data %>%
  filter(Бали_студента > 0) %>%
  select(КодОсоби, КодКурсу, Кількість_відвідувань, Бали_студента) %>%
  na.omit()

cluster_data_scaled <- scale(agg_data_clean[, c("Кількість_відвідувань",
"Бали_студента")])

set.seed(123)
kmeans_result <- kmeans(cluster_data_scaled, centers = 3)
agg_data_clean$cluster <- as.factor(kmeans_result$cluster)

ggplot(agg_data_clean, aes(x = Кількість_відвідувань, y = Бали_студента, color =
cluster)) +
  geom_point(size = 2, alpha = 0.8) +
  labs(
    title = "Кластеризація студентів за відвідуваністю та балами",
    x = "Кількість відвідувань",
    y = "Середній бал студента"
  ) +
  theme_minimal()

# ANOVA: вплив джерела фінансування на бали
data$`Джерело фінансування` <- as.factor(data$`Джерело фінансування`)
aov_result <- aov(`Бали студента` ~ `Джерело фінансування`, data = data)
summary(aov_result)

ggplot(data, aes(x = `Джерело фінансування`, y = `Бали студента`, fill =
`Джерело фінансування`)) +
  geom_boxplot() +
  labs(title = "Порівняння балів студентів за джерелом фінансування",
    x = "Джерело фінансування",
    y = "Бали студента") +
  theme_minimal() +
  scale_fill_manual(values = c("Контракт" = "orange", "Бюджет" = "lightgreen"))

library(tidyr)
library(stringr)

# Аналіз середніх балів по групах для курсу 1
data_clean <- data %>%
  filter(`Бали студента` > 0, !is.na(`Бали студента`), Курс == 1) %>%
  separate_rows(Група, sep = ",") %>%
  mutate(Група = str_trim(Група))

avg_scores <- data_clean %>%
  group_by(Група, Студент) %>%

```

```

  summarise(Середній_бал = mean(`Бали студента`, na.rm = TRUE), .groups =
"drop")

ggplot(avg_scores, aes(x = Група, y = Середній_бал)) +
  geom_boxplot(fill = "#56B4E9", color = "black") +
  labs(
    title = "Розподіл середніх балів по групах (Курс 1)",
    x = "Група",
    y = "Середній бал"
  ) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

# Аналіз середніх балів по групах для курсу 4
data_clean <- data %>%
  filter(`Бали студента` > 0, !is.na(`Бали студента`), Курс == 4) %>%
  separate_rows(Група, sep = ",") %>%
  mutate(Група = str_trim(Група))

avg_scores <- data_clean %>%
  group_by(Група, Студент) %>%
  summarise(Середній_бал = mean(`Бали студента`, na.rm = TRUE), .groups =
"drop")

ggplot(avg_scores, aes(x = Група, y = Середній_бал)) +
  geom_boxplot(fill = "#56B4E9", color = "black") +
  labs(
    title = "Розподіл середніх балів по групах (Курс 4)",
    x = "Група",
    y = "Середній бал"
  ) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

# Додатковий огляд даних
str(data)
summary(dplyr::select(data, `Бали студента`, `КБ при вступі`))
table(data$Стать)
table(data$`Джерело фінансування`)
table(data$Курс)
dim(data)

```

## **Прогнозування оцінок студентів у вищій освіті: застосування регресії та випадкових лісів**

```

# --- Лінійна регресія ---

# 1. Очищаємо дані: прибираємо нульові оцінки
data_clean <- data %>%
  filter(`Бали студента` != 0)

# 2. Агрегуємо дані по студенту і дисципліні
agg_data_unique <- data_clean %>%
  group_by(Студент, Дисципліна) %>%
  summarise(
    Бали_студента = mean(`Бали студента`, na.rm = TRUE),
    КБ_при_вступі = mean(`КБ при вступі`, na.rm = TRUE),
    Стать = first(Стать),
    Джерело_фінансування = first(`Джерело фінансування`),
    Курс = first(Курс),
    .groups = "drop"
  )

# 3. Перетворюємо фактори з правильними рівнями

```

```

agg_data_unique <- agg_data_unique %>%
  mutate(
    Стать = factor(Стать, levels = c("Жіноча", "Чоловіча")),
    Джерело_фінансування = factor(Джерело_фінансування, levels = c("Бюджет",
"Контракт")),
    Курс = factor(Курс)
  )

# 4. Створення лінійної регресії
regression_model <- lm(
  Бали_студента ~ КБ_при_вступі + Стать + Джерело_фінансування + Курс,
  data = agg_data_unique
)

# 5. Форматування виводу результатів
regression_summary <- tidy(regression_model)

reference_levels <- tibble::tibble(
  term = c("Стать: Жіноча", "Джерело фінансування: Бюджет", "Курс: 1"),
  estimate = 0,
  std.error = NA,
  statistic = NA,
  p.value = NA
)

regression_summary_clean <- regression_summary %>%
  mutate(term = case_when(
    term == "(Intercept)" ~ "Перетин (Intercept)",
    term == "КБ_при_вступі" ~ "КБ при вступі",
    term == "СтатьЧоловіча" ~ "Стать: Чоловіча",
    term == "Джерело_фінансуванняКонтракт" ~ "Джерело фінансування: Контракт",
    term == "Курс2" ~ "Курс: 2",
    term == "Курс3" ~ "Курс: 3",
    term == "Курс4" ~ "Курс: 4",
    TRUE ~ term
  ))

full_model_table <- bind_rows(regression_summary_clean, reference_levels)

# Вивід таблиці з результатами регресії
full_model_table %>%
  arrange(term) %>%
  kable(digits = 4, caption = "Результати регресії без нульових оцінок", align =
"с")

# --- Прогнозування на основі лінійної регресії ---

new_data_reg <- data.frame(
  КБ_при_вступі = 187,
  Стать = factor("Жіноча", levels = levels(agg_data_unique$Стать)),
  Джерело_фінансування = factor("Бюджет", levels =
levels(agg_data_unique$Джерело_фінансування)),
  Курс = factor(1, levels = levels(agg_data_unique$Курс)),
  Дисципліна = factor("Окремі розділи математичної статистики", levels =
levels(agg_data_unique$Дисципліна))
)

predicted_reg <- predict(regression_model, newdata = new_data_reg)
cat("Прогнозована оцінка (лінійна регресія):", round(predicted_reg, 2), "\n")

# --- Модель Random Forest ---

# Повторна агрегація (з видаленням NA)

```

```

agg_data_unique <- data %>%
  group_by(Студент, Дисципліна) %>%
  summarise(
    Бали_студента = mean(`Бали студента`, na.rm = TRUE),
    КБ_при_вступі = mean(`КБ при вступі`, na.rm = TRUE),
    Стать = first(Стать),
    Джерело_фінансування = first(`Джерело фінансування`),
    Курс = first(Курс),
    .groups = "drop"
  ) %>%
  drop_na()

# Факторні змінні для RF (якщо потрібно)
agg_data_unique <- agg_data_unique %>%
  mutate(
    Стать = factor(Стать),
    Джерело_фінансування = factor(Джерело_фінансування),
    Курс = factor(Курс),
    Дисципліна = factor(Дисципліна)
  )

# Створення моделі Random Forest
rf_model <- randomForest(
  Бали_студента ~ КБ_при_вступі + Стать + Джерело_фінансування + Курс +
  Дисципліна,
  data = agg_data_unique,
  ntree = 500,
  importance = TRUE
)

print(rf_model)

# Важливість змінних
importance_df <- as.data.frame(importance(rf_model))
importance_df$Variable <- rownames(importance_df)

# Візуалізація важливості змінних
ggplot(importance_df, aes(x = reorder(Variable, `%IncMSE`), y = `%IncMSE`)) +
  geom_bar(stat = "identity", fill = "#69b3a2") +
  coord_flip() +
  labs(
    title = "Важливість змінних у моделі Random Forest",
    x = "Змінна",
    y = "% Збільшення MSE при виключенні змінної"
  ) +
  theme_minimal(base_size = 12)

```