



АНАЛІЗ ДАНИХ

Робоча програма навчальної дисципліни (Силабус)

Реквізити навчальної дисципліни

Рівень вищої освіти	<i>Перший (бакалаврський)</i>
Галузь знань	<i>12 Інформаційні технології</i>
Спеціальність	<i>121 Інженерія програмного забезпечення</i>
Освітня програма	<i>Інженерія програмного забезпечення мультимедійних та інформаційно-пошукових систем</i>
Статус дисципліни	<i>Вибіркова</i>
Форма навчання	<i>Очна (денна)</i>
Рік підготовки, семестр	<i>3 рік навчання, 6 семестр</i>
Обсяг дисципліни	<i>Лекції: 36 год., лабораторні заняття: 18 год, самостійна робота: 66 год.</i>
Семестровий контроль/ контрольні заходи	<i>Залік, модульна контрольна робота, календарний контроль</i>
Розклад занять	<i>Згідно розкладу на весняний семестр поточного навчального року (rozklad.kpi.ua)</i>
Мова викладання	<i>Українська</i>
Інформація про керівника курсу / викладачів	<i>Лектор: к.т.н., доцент, Онай Микола Володимирович, onay@pzks.fpm.kpi.ua Комп'ютерний практикум: к.т.н., доцент, Онай Микола Володимирович, onay@pzks.fpm.kpi.ua</i>

Програма навчальної дисципліни

1. Опис навчальної дисципліни, її мета, предмет вивчення та результати навчання

Вивчення дисципліни «Аналіз даних» дозволяє сформувати у здобувачів вищої освіти компетенції, необхідні для розв'язання складних задач професійної діяльності, пов'язаної із розробленням програмних систем для розв'язання типових задач, що виникають при аналізі великих об'ємів накопичених експериментальних даних.

Метою вивчення дисципліни «Аналіз даних» є формування у здобувачів освіти здатності проводити інноваційну діяльність, пов'язану із розробленням програмних систем для виконання аналізу даних різної структури.

Предметом дисципліни «Аналіз даних» є методи виконання програмного аналізу даних.

Вивчення дисципліни «Аналіз даних» підсилює формування у здобувачів освіти **фахових компетентностей (ФК)**, необхідних для розв'язання практичних задач професійної діяльності:

ФК08 Здатність застосовувати фундаментальні і міждисциплінарні знання для успішного розв'язання завдань інженерії програмного забезпечення.

ФК15 Здатність застосовувати фундаментальні і міждисциплінарні знання для побудови удосконалених алгоритмів пошуку.

ФК16 Здатність розробляти алгоритми реалізації методів статистичного аналізу даних.

Вивчення дисципліни «Аналіз даних» сприяє формуванню у студентів наступних **програмних результатів навчання (ПРН)** за освітньою програмою:

ПРН01 Аналізувати, цілеспрямовано шукати і вибирати необхідні для вирішення професійних завдань інформаційно-довідникові ресурси і знання з урахуванням сучасних досягнень науки і техніки.

ПРН13 Знати і застосовувати методи розроблення алгоритмів, конструювання програмного забезпечення та структур даних і знань.

ПРН25 Знати і вміти використовувати фундаментальний математичний інструментарій при побудові алгоритмів та розробленні сучасного програмного забезпечення.

ПРН27 Вміти використовувати методи статистичного аналізу даних.

ПРН39 Знати типи пошукових систем, принципи їх побудови, методи та алгоритми виконання різних типів пошуку інформації в них.

ПРН40 Знати та вміти застосовувати на практиці методи та критерії оцінювання ефективності інформаційного пошуку.

2. Пререквізити та постреквізити дисципліни (місце в структурно-логічній схемі навчання за відповідною освітньою програмою)

Успішному вивченню дисципліни «Аналіз даних» передують вивчення дисциплін «Алгоритми та структури даних» та «Алгоритмічне забезпечення мультимедійних та інформаційно-пошукових систем» навчального плану підготовки бакалаврів за спеціальністю 121 Інженерія програмного забезпечення.

Отримані в результаті засвоєння дисципліни «Аналіз даних» теоретичні знання та практичні уміння можуть бути корисні для проведення наукових досліджень та при виконанні бакалаврської кваліфікаційної роботи.

3. Зміст навчальної дисципліни

Дисципліна «Аналіз даних» передбачає вивчення тем:

Тема 1. Основні положення аналізу даних

Тема 2. Перевірка статистичних гіпотез

Тема 3. Дисперсійний аналіз

Тема 4. Кореляційний аналіз

Тема 5. Факторний аналіз

Тема 6. Кластерний аналіз

Модульна контрольна робота

Залік

4. Навчальні матеріали та ресурси

Базова література:

1. Методи аналізу даних : навчальний посібник для студентів / В.Є. Бахрушин. – Запоріжжя : КПУ, 2011. – 268 с.

Додаткова література:

1. Benjamin S. Duran, Patrick L. Odell Cluster Analysis a Survey. – Springer-Verlag. – Berlin-Heidelberg New York. – 1974.

Використати для опанування практичних навичок з дисципліни.

2. Jared Dean Big Data Mining, and Machine Learning [Електронний ресурс], 2014. Режим доступу: https://www.booksfree.org/wp-content/uploads/2022/06/Big-Data-Data-Mining-and-Machine-Learning-by-Jared-Dean-pdf-free-download-booksfree.org_.pdf

Використати для опанування теоретичним матеріалом з дисципліни.

3. Mohammed J. Zaki, Wagner Meira Jr. *Data Mining and Analysis. Fundamental Concepts and Algorithms* [Електронний ресурс], 2014. Режим доступу:

http://pzs.dstu.dp.ua/DataMining/bibl/mohammed_j_zaki_wagner_meira_jr_data_mining_and_analysis_fun.pdf

Використати для опанування практичних навичок з дисципліни.

5. Jiawei Han, Micheline Kamber, Jian Pei *Data Mining: Concepts and Techniques. Third Edition* [Електронний ресурс], 2012. Режим доступу:

<http://myweb.sabanciuniv.edu/rdehkharghani/files/2016/02/The-Morgan-Kaufmann-Series-in-Data-Management-Systems-Jiawei-Han-Micheline-Kamber-Jian-Pei-Data-Mining.-Concepts-and-Techniques-3rd-Edition-Morgan-Kaufmann-2011.pdf>

Використати для вивчення принципів розв'язання нелінійних рівнянь. Матеріали знаходяться у вільному доступі в Інтернеті.

Навчальний контент

5. Методика опанування навчальної дисципліни (освітнього компонента)

№ з/п	Тип навчального заняття	Опис навчального заняття
<i>Тема 1. Основні положення аналізу даних</i>		
1	<i>Лекція 1. Класифікація методів аналізу даних</i>	<i>Кореляційний аналіз. Дисперсійний аналіз. Регресійний аналіз. Коваріаційний аналіз. Дискримінантний аналіз. Кластерний аналіз. Аналіз часових рядів</i>
2	<i>Лекція 2. Варіаційна статистика</i>	<i>Теоретичні положення варіаційної статистики та приклади побудови</i>
3	<i>Комп'ютерний практикум 1</i>	<i>Завдання: Розробити програму для побудови типових графіків варіаційної статистики</i>
<i>Тема 2. Перевірка статистичних гіпотез</i>		
4	<i>Лекція 3. Основні положення статистичних гіпотез</i>	<i>Основні поняття. Параметричні тести. Непараметричні тести. Визначення моделей розподілу емпіричних даних.</i>
5	<i>Лекція 4. Ідентифікація функції</i>	<i>Ідентифікація функції розподілу однорідної вибірки. Ідентифікація функції розподілу неоднорідної вибірки.</i>
6	<i>Комп'ютерний практикум 2</i>	<i>Завдання: Розробити програму для графічної інтерпретації моделей розподілу емпіричних даних</i>
<i>Тема 3. Дисперсійний аналіз</i>		
7	<i>Лекція 5. Однофакторний аналіз</i>	<i>Факторна або міжгрупова варіація. Залишкова або внутрішньогрупова варіація. Ранговий</i>

		<i>однофакторний аналіз Краскела-Уолліса. Критерій Джонкхієра-Терпстра</i>
8	<i>Лекція 6. Двофакторний аналіз</i>	<i>Дисперсійний аналіз за двома ознаками. Ранговий критерій Фрідмана. Критерій Пейджа. Приклад виконання дисперсійного аналізу</i>
9	<i>Комп'ютерний практикум 3</i>	<i>Завдання: Розробити програму для проведення однофакторного та двофакторного дисперсійного аналізу</i>
10	<i>Модульна контрольна робота. Частина 1</i>	
<i>Тема 4. Кореляційний аналіз</i>		
11	<i>Лекція 7. Кореляційний аналіз кількісних ознак</i>	<i>Вибірковий коефіцієнт детермінації. Коефіцієнт кореляції Пірсона. Коефіцієнт кореляції Фехнера. Коваріаційна матриця</i>
12	<i>Комп'ютерний практикум 4</i>	<i>Завдання: Розробити програму для обчислення коефіцієнтів кореляції</i>
13	<i>Лекція 8. Кореляційний аналіз порядкових ознак</i>	<i>Рангова кореляція. Коефіцієнт рангової кореляції Спірмена. Коефіцієнт рангової кореляції Кендалла.</i>
14	<i>Лекція 9. Кореляційний аналіз номінальних ознак</i>	<i>Ф-коефіцієнт Пірсона. Середньоквадратична спряженість. Показник подібності Жаккара</i>
15	<i>Комп'ютерний практикум 5</i>	<i>Завдання: Розробити програму для аналізу рангової подібності</i>
16	<i>Лекція 10. Кореляційний аналіз змішаних ознак</i>	<i>Коефіцієнт Гауера. Бісеріальний коефіцієнт кореляції. Бісеріальний коефіцієнт кореляції. За таблицею Келлі-Вуда.</i>
17	<i>Лекція 11. Множинна кореляція</i>	<i>Частинний коефіцієнт кореляції. Коефіцієнт множинної кореляції. Канонічний кореляційний аналіз. Коефіцієнт конкордації.</i>
18	<i>Комп'ютерний практикум 6</i>	<i>Завдання: Розробити програму для обчислення коефіцієнту множинної кореляції та конкордації</i>
<i>Тема 5. Факторний аналіз</i>		
19	<i>Лекція 12. Метод головних компонент</i>	<i>Матриця факторного відображення. Критерій Кайзера. Критерій відсіювання.</i>
20	<i>Лекція 13. Метод головних факторів</i>	<i>Повна факторна матриця. Факторне відображення. Фундаментальна теорема факторного аналізу</i>
21	<i>Комп'ютерний практикум 7</i>	<i>Завдання: Розробити програму для проведення факторного аналізу</i>
<i>Тема 6. Кластерний аналіз</i>		
22	<i>Лекція 14. Метод найближчого сусіда та метод дальнього сусіда</i>	<i>Міри подібності (відмінності). Інформаційна статистика. Відстань Спірмена. Відстань Кендалла. Евклідова відстань. Зважена евклідова відстань. Ієрархічні та неієрархічні методи. Метод найближчого та дальнього сусіда</i>
23	<i>Лекція 15. Метод середнього зв'язку та метод центрів тяжіння</i>	<i>Узагальнена К-відстань. Відстань Геммінга. Метод середнього зв'язку та метод центрів</i>

		<i>тяжіння. Приклади роботи методу середнього зв'язку та методу центрів тяжіння</i>
24	<i>Комп'ютерний практикум 8</i>	<i>Завдання: Розробити програму для проведення кластерного аналізу</i>
25	<i>Лекція 16. Метод k-середніх</i>	<i>Узагальнений метод k-середніх та його модифікації. Приклади роботи алгоритму, що реалізує метод k-середніх</i>
26	<i>Комп'ютерний практикум 9</i>	<i>Підсумки</i>
27	<i>Модульна контрольна робота. Частина 2</i>	

6. Самостійна робота студента/аспіранта

Дисципліна «Аналіз даних» ґрунтується на самостійній підготовці до аудиторних занять на теоретичні та практичні теми.

<i>№ з/п</i>	<i>Назва теми, що виноситься на самостійне опрацювання</i>	<i>Кількість годин</i>	<i>Література</i>
1	<i>Підготовка до лекцій</i>	16	1-5
2	<i>Підготовка до комп'ютерного практикуму</i>	27	1-5
3	<i>Підготовка до модульної контрольної роботи. Частина 1</i>	9	1-5
4	<i>Підготовка до модульної контрольної роботи. Частина 2</i>	9	1-5
5	<i>Підготовка до заліку</i>	5	1-5

Політика та контроль

7. Політика навчальної дисципліни (освітнього компонента)

Відвідування занять. Відсутність на аудиторному занятті не передбачає нарахування штрафних балів, оскільки фінальний рейтинговий бал студента формується виключно на основі оцінювання результатів навчання. Разом з тим, обговорення результатів виконання тематичних завдань, а також презентація / публічний виступ та участь у обговореннях та доповнення на семінарах оцінюватимуться під час аудиторних занять. Для активної участі у роботі семінару студент готується за рекомендованою викладачем до певного семінарського заняття літературою. Участь у роботі семінару також передбачає підготування доповідей та співдоповідей у межах усіх занять.

Пропущені контрольні заходи оцінювання. Кожен студент має право відпрацювати пропущені з поважної причини (лікарняний, мобільність тощо) заняття за рахунок самостійної роботи. Детальніше за посиланням: <https://kpi.ua/files/n3277.pdf>.

Процедура оскарження результатів контрольних заходів оцінювання. Студент може підняти будь-яке питання, яке стосується процедури контрольних заходів та очікувати, що воно буде розглянуто згідно із наперед визначеними процедурами. Студенти мають право аргументовано оскаржити результати контрольних заходів, пояснивши з яким критерієм не погоджуються відповідно до оціночного. Календарний контроль проводиться з метою підвищення якості навчання студентів та моніторингу виконання студентом вимог силабусу.

Академічна доброчесність. Політика та принципи академічної доброчесності визначені у розділі 3 Кодексу честі Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського». Детальніше: <https://kpi.ua/code>.

Норми етичної поведінки. Норми етичної поведінки студентів і працівників визначені у розділі 2 Кодексу честі Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського». Детальніше: <https://kpi.ua/code>.

Інклюзивне навчання. Засвоєння знань та умінь в ході вивчення дисципліни «Науково-дослідна діяльність у комп'ютерній інженерії» може бути доступним для більшості осіб з особливими освітніми потребами, окрім здобувачів з серйозними вадами зору, які не дозволяють виконувати завдання за допомогою персональних комп'ютерів, ноутбуків та/або інших технічних засобів.

Навчання іноземною мовою. У ході виконання завдань студентам може бути рекомендовано звернутися до англomовних джерел. Призначення заохочувальних та штрафних балів Відповідно до Положення про систему оцінювання результатів навчання сума всіх заохочувальних балів не може перевищувати 10% рейтингової шкали оцінювання.

Всі студенти повинні відвідувати лекційні та практичні заняття, на яких потрібно активно працювати над засвоєнням навчального матеріалу. За об'єктивних причин (наприклад - хвороба, міжнародне стажування) навчання може відбуватись в он-лайн формі індивідуально за погодженням із керівником курсу.

Політика щодо дедлайнів та перескладання:

Роботи, які здаються із порушенням термінів без поважних причин, оцінюються на нижчу оцінку. Перескладання модулів відбувається із дозволу деканату за наявності поважних причин (наприклад, лікарняний).

Політика щодо академічної доброчесності:

Усі письмові роботи перевіряються на наявність плагіату і допускаються до захисту із коректними текстовими запозиченнями не більше 20%. Списування під час контрольних робіт заборонені (в т. ч. із використанням мобільних пристроїв).

8. Види контролю та рейтингова система оцінювання результатів навчання (PCO)

Протягом семестру студентами виконують 8 комп'ютерних практикумів. Максимальна кількість балів за кожний комп'ютерний практикум: 6 балів.

Бали нараховуються за:

- якість виконання комп'ютерного практикуму: 0-2 бали;
- відповідь на теоретичні запитання під час захисту комп'ютерного практикуму: 0-2 бали;
- своєчасне представлення роботи до захисту: 0-2 бали.

Критерії оцінювання якості виконання:

- 2 бали – робота виконана якісно, в повному обсязі;
- 1 бал – робота виконана в повному обсязі, але містить незначні помилки;
- 0 балів – робота виконана не в повному обсязі, або містить суттєві помилки.

Критерії оцінювання відповіді:

- 2 бали – відповідь повна, добре аргументована;
- 1 бал – в цілому відповідь правильна, але має недоліки або незначні помилки;
- 0 балів – немає відповіді або відповідь неправильна.

Критерії оцінювання своєчасності представлення роботи до захисту:

- 2 бали – робота представлена до захисту не пізніше вказаного терміну;
- 0 балів – робота представлена до захисту пізніше вказаного терміну.

Максимальна кількість балів за виконання та захист комп'ютерних практикумів:
6 балів × 8 комп. практ. = 48 балів.

Завдання на **модульну контрольну роботу** складається з 3 запитань – 1 теоретичного та 2 практичних. Відповідь на теоретичне запитання оцінюється 6 балами, а відповідь на практичне запитання оцінюється 10 балами.

Критерії оцінювання кожного теоретичного запитання контрольної роботи:

6 балів – відповідь правильна, повна, добре аргументована;
5 балів – відповідь правильна, розгорнута, але не дуже добре аргументована;
4 балів – в цілому відповідь правильна, але має недоліки;
3 балів – у відповіді є незначні помилки;
1-2 бали – у відповіді є суттєві помилки;
0 балів – немає відповіді або відповідь неправильна.

Критерії оцінювання практичного запитання контрольної роботи:

9-10 балів – відповідь правильна, розрахунки виконані у повному обсязі;
7-8 балів – відповідь правильна, але не дуже добре підкріплена розрахунками;
5-6 балів – в цілому відповідь правильна, але має недоліки;
3-4 балів – у відповіді є незначні помилки;
1-2 бали – у відповіді є суттєві помилки;
0 балів – немає відповіді або відповідь неправильна.

Максимальна кількість балів за модульну контрольну роботу:

2 роботи * (6 балів × 1 теоретичне запитання + 10 балів × 2 практичних запитання) = 52 бали.

Рейтингова шкала з дисципліни дорівнює:

$R_c = R_{\text{ком.практ}} + R_{\text{МКР}} = 48 \text{ балів} + 52 \text{ балів} = 100 \text{ балів}$.

Календарний контроль: провадиться двічі на семестр як моніторинг поточного стану виконання вимог силабусу.

На першій атестації (8-й тиждень) студент отримує «зараховано», якщо його поточний рейтинг не менше 50 % від максимальної кількості балів (20 балів), яку може отримати студент до першої атестації.

На другій атестації (14-й тиждень) студент отримує «зараховано», якщо його поточний рейтинг не менше 50 % від максимальної кількості балів (35 балів), яку може отримати студент до другої атестації.

Семестровий контроль: **залік**

Умови допуску до семестрового контролю:

За семестрового рейтингу (R_c) не менше 60 балів та зарахуванні усіх робіт комп'ютерного практикуму аспірант отримує залік «автоматом» відповідно до таблиці (Таблиця відповідності рейтингових балів оцінкам за університетською шкалою). В іншому разі він має виконати залікову контрольну роботу.

Необхідною умовою допуску до виконання залікової контрольної роботи є виконання і захист комп'ютерного практикуму.

Аспірант може спробувати підвищити свою оцінку шляхом написання залікової контрольної роботи, при цьому його бали, отримані за семестр, анулюються («жорстка» система оцінювання).

Склад та критерії оцінювання залікової контрольної роботи:

Завдання на **залікову контрольну роботу** складається з 4 запитань – 2 теоретичних та 2 практичних. Відповідь на кожне теоретичне та практичне запитання оцінюється 25 балами.

Критерії оцінювання кожного теоретичного запитання контрольної роботи:

24-25 балів – відповідь правильна, повна, добре аргументована;

21-23 бали – відповідь правильна, розгорнута, але не дуже добре аргументована;
17-20 балів – в цілому відповідь правильна, але має недоліки;
12-16 балів – у відповіді є незначні помилки;
1-11 бали – у відповіді є суттєві помилки;
0 балів – немає відповіді або відповідь неправильна.

Критерії оцінювання практичного запитання контрольної роботи:

24-25 балів – відповідь правильна, розрахунки виконані у повному обсязі;
21-23 бали – відповідь правильна, але не дуже добре підкріплена розрахунками;
17-20 балів – в цілому відповідь правильна, але має недоліки;
12-16 балів – у відповіді є незначні помилки;
1-11 бали – у відповіді є суттєві помилки;
0 балів – немає відповіді або відповідь неправильна.

Максимальна кількість балів за модульну контрольну роботу:

25 балів × 2 теоретичних запитання + 25 балів × 2 практичних запитання = 100 балів.

Таблиця відповідності рейтингових балів оцінкам за університетською шкалою:

Кількість балів	Оцінка
100-95	Відмінно
94-85	Дуже добре
84-75	Добре
74-65	Задовільно
64-60	Достатньо
Менше 60	Незадовільно
Не виконані умови допуску	Не допущено

9. Додаткова інформація з дисципліни (освітнього компонента)

Перелік запитань, які виносяться на семестровий контроль буде оприлюднено на останньому занятті.

Робочу програму навчальної дисципліни (силабус):

Складено к.т.н., доц., Онай М.В.

Ухвалено кафедрою ПЗКС (протокол №8 від 25.01.2023)

Погоджено Методичною комісією факультету прикладної математики (протокол №6 від 27.01.2023)