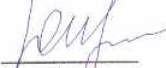


Міністерство освіти і науки України
Національний аерокосмічний університет
«Харківський авіаційний інститут»

Кафедра математичного моделювання та штучного інтелекту (№ 304)

ЗАТВЕРДЖУЮ

Гарант освітньої програми


(підпис)

Юнна ЩЕРБАКОВА
(ім'я та ПРІЗВИЩЕ)

« 29 » 08 2025 р.

**СИЛАБУС ОBOB'ЯЗKОВОЇ
НАВЧАЛЬНОЇ ДИСЦИПЛІНИ**

Аналіз даних

(назва навчальної дисципліни)

Галузь знань: 12 «Інформаційні технології»
(шифр і найменування галузі знань)

Спеціальність: 124 «Системний аналіз»
(код і найменування спеціальності)

Освітня програма: «Системний аналіз і управління»
(найменування освітньої програми)

Рівень вищої освіти: *перший (бакалаврський)*

Силабус уведено в дію з 01.09.2025

Харків – 2025 р.

Розробник: Базілевич К.О., доц. каф. 304, к. т. н., доц.
(прізвище та ініціали, посада, науковий ступінь і вчене звання)


(підпис)

Силабус навчальної дисципліни розглянуто на засіданні кафедри № 304
математичного моделювання та штучного інтелекту
(назва кафедри)

Протокол № 1 від « 21 » серпня 2025 р.

В. о. завідувача кафедри к.ф.-м.н., доцент
(науковий ступінь і вчене звання)


(підпис)

Олексій КАРТАШОВ
(ім'я та ПРІЗВИЩЕ)

Погоджено з представником здобувачів освіти:

Здобувач другого (магістерського)

рівня вищої освіти за ОП

«Системний аналіз і управління»


(підпис)

Володимир ТИТАРЕНКО
(ім'я та ПРІЗВИЩЕ)

1. Загальна інформація про викладача



ПІБ: *Базілевич Ксенія Олексіївна*

Посада: *доцент каф. 304*

Науковий ступінь: *канд. техн. наук*

Вчене звання: *доцент*

Перелік дисциплін, які викладає:

Інтелектуальний аналіз даних; Проєктування програмного забезпечення; Сучасні методи проєктування програмного забезпечення

Напрями наукових досліджень:

Інтелектуальний аналіз даних

Контактна інформація:

k.bazilevych@khai.edu

2. Опис навчальної дисципліни

Форма здобуття освіти	Денна
Семестр	5
Мова викладання	Українська
Тип дисципліни	Обов'язкова
Обсяг дисципліни: кредити ЄКТС/ кількість годин	4,0 кредити ЄКТС / 120 годин (64 аудиторних, з яких: лекції – 32, лабораторні – 32; СРЗ – 56);
Види навчальної діяльності	Лекції та лабораторні заняття, самостійна робота
Види контролю	Поточний контроль, модульний контроль, семестровий контроль – залік
Пререквізити	«Програмування та алгоритмічні мови», «Теорія ймовірностей»

3. Мета та завдання навчальної дисципліни, переліки компетентностей та очікуваних результатів навчання

Мета – надання теоретичних та практичних аспектів аналізу даних, спрямованих на пошук у необроблених даних раніше невідомих, практично корисних знань та закономірностей, необхідних для прийняття рішень, та розробки інформаційних технологій аналізу даних.

Завдання – ознайомлення з теоретичними аспектами технології аналізу даних, формування у студентів базових навичок застосування методів аналізу даних з використанням інструментальних засобів.

Компетентності, які набуваються:

Інтегральна компетентність:

Здатність розв'язувати складні спеціалізовані задачі та практичні проблеми системного аналізу у професійній діяльності або в процесі навчання, що передбачають застосування теоретичних положень та методів системного аналізу та інформаційних технологій і характеризується комплексністю та невизначеністю умов.

Загальні компетентності (ЗК)

Після закінчення цієї програми здобувач освіти буде здатен:

- до абстрактного мислення, аналізу та синтезу (ЗК1);
- застосовувати знання у практичних ситуаціях (ЗК2);
- розуміти предметну область та бути здатним до професійної діяльності (ЗК3);
- спілкуватися державною мовою як усно, так і письмово (ЗК 4);
- вчитися й оволодівати сучасними знаннями (ЗК 6).

Спеціальні компетентності (ФК)

Після закінчення цієї програми здобувач освіти буде здатен:

- до комп'ютерної реалізації математичних моделей реальних систем і процесів; проектувати, застосовувати і супроводжувати програмні засоби моделювання, прийняття рішень, оптимізації, обробки інформації, інтелектуального аналізу даних (ФК 6);
- представляти математичні аргументи і висновки з них з ясністю і точністю і в таких формах, які підходять для аудиторії як усно, так і в письмовій формі (ФК 9).

Програмні результати навчання (ПРН):

- Застосовувати методи і засоби роботи з даними і знаннями, методи математичного, логіко-семантичного, об'єктного та імітаційного моделювання, технології системного і статистичного аналізу (ПРН 12);
- Розуміти і застосовувати на практиці методи статистичного моделювання та прогнозування, оцінювати вихідні дані (ПРН 14).

4. Зміст навчальної дисципліни

МОДУЛЬ 1

Змістовий модуль 1

Тема 1. *Поняття вимірювання в статистичному дослідженні*

Стисла анотація: Предмет вивчення і задачі дисципліни «Аналіз даних». Первинні статистики. Статистичні гіпотези. Вимірювальні шкали: номінативна шкала, рангова шкала, шкала інтервалів, шкала відносин. Правила ранжирування, перевірка правильності ранжирування, випадок однакових рангів. Форми обліку результатів спостереження. Основи підготовки даних. Представлення результатів аналізу даних у вигляді інформативних звітів, що містять обґрунтовані висновки та рекомендації.

Теми лекцій і лабораторних робіт: «Введення до предмету», «Типи шкал вимірювання та аналіз первинних статистик».

Самостійна робота здобувача освіти: опрацювання матеріалу лекцій, виконання індивідуальних завдань відповідно варіанту, підготовка до захисту лабораторних робіт.

Тема 2. *Дисперсійний аналіз*

Стисла анотація: Постановка задачі однофакторного дисперсійного аналізу. Основні посилення дисперсійного аналізу. Статистичні оцінки факторної та залишкової дисперсій. Формулювання та перевірка гіпотези про рівність дисперсій. Постановка задачі двофакторного дисперсійного аналізу. Статистичні оцінки відповідних дисперсій. Формулювання та перевірка гіпотези про рівність дисперсій.

Теми лекцій і лабораторних робіт: «Однофакторний дисперсійний аналіз», «Двофакторний дисперсійний аналіз».

Самостійна робота здобувача освіти: опрацювання матеріалу лекцій, виконання індивідуальних завдань відповідно варіанту, підготовка до захисту лабораторних робіт.

Тема 3. *Кореляційний аналіз*

Стисла анотація: Коефіцієнт детермінації як універсальна характеристика ступеню тісноти статистичного зв'язку. Кореляційне відношення. Дослідження лінійної залежності за допомогою парного коефіцієнта кореляції. Множинні та часткові коефіцієнти кореляції.

Теми лекцій і лабораторних робіт: «Елементи кореляційного аналізу даних».

Самостійна робота здобувача освіти: опрацювання матеріалу лекцій,

виконання індивідуальних завдань відповідно варіанту, підготовка до захисту лабораторних робіт.

Модульний контроль 1

Змістовий модуль 2

Тема 4. Множинний регресійний аналіз

Стисла анотація: Регресійний аналіз. Основні ідеї регресійної моделі. Модель лінійної регресії. Оцінка якості моделі, інтерпретація та оцінки коефіцієнтів регресії, рівень значущості коефіцієнтів. Обмеження регресійної моделі. Нелінійні залежності. Причини виникнення, приклади різних форм нелінійних залежностей в регресійних моделях. Модель логістичної регресії. Інтерпретація коефіцієнтів логістичної регресії.

Теми лекцій і лабораторних робіт: «Множинна регресія».

Самостійна робота здобувача освіти: опрацювання матеріалу лекцій, виконання індивідуальних завдань відповідно варіанту, підготовка до захисту лабораторних робіт.

Тема 5. Елементи Data Mining

Стисла анотація: Знайомство з Data Mining. Класифікація, кластеризація, проста та складна класифікація, багатовимірна класифікація, штучна та природна класифікація. Постановка задачі автоматичної класифікації, відстані між окремими об'єктами та класами об'єктів. Функціонали якості розбиття на класи та екстремальна постановка задачі кластер-аналіза. Паралельні кластер-процедури. Послідовні кластер-процедури. Оцінка результатів дослідження. Повний цикл обробки даних.

Теми лекцій і лабораторних робіт: «Використання інструментів Data Mining для виконання задач класифікації, кластеризації та регресії», «Повний цикл обробки даних».

Самостійна робота здобувача освіти: опрацювання матеріалу лекцій, виконання індивідуальних завдань відповідно варіанту, підготовка до захисту лабораторних робіт.

Модульний контроль 2

5. Індивідуальні завдання

Робота розрахункова (РР) «Проведення власного дослідження із застосуванням методів класифікації»

1. Обрати статистичні дані з будь-якої предметної області для аналізу. Приклади додаються в файлі з завданням.

2. Обрати самостійно мету дослідження: сформулювати постановку задачі та хід дослідження, з урахуванням того, які класи будуть отримані в результаті розрахунків.

3. Провести розрахунки за допомогою двох методів для вирішення задачі класифікації, порівняти результати розрахунків.

4. Розробити алгоритмічну модель та блок-схему кожного методу.

5. Розробити програмний додаток, що вирішує описані задачі та підтримує завантаження статичних даних для розрахунків.

6. Оформити звіт по виконаних завданнях, зробити висновки.

6. Методи навчання

1. Пояснювально-ілюстративний (інформаційно-рецептивний) метод: та метод проблемного виконання (лекція).

2. Репродуктивний (лабораторні роботи).

3. Частково-пошуковий (евристичний) та дослідницький: (самостійна робота).

4. Дисципліна «Аналіз даних» передбачає лекційні (в т.ч. з використанням мультимедійного обладнання) і практичні заняття для виконання лабораторних робіт під керівництвом викладача та самостійну роботу студента за підручниками і матеріалами, що забезпечує закріплення теоретичних знань, сприяє набуттю практичних навичок і розвитку самостійного наукового мислення. Передбачено регулярні індивідуальні консультації.

7. Методи контролю

Визначення рівня засвоєння студентом навчального матеріалу дисципліни здійснюється шляхом проведення поточних (захист лабораторних робіт) і підсумкових контролів (захист змістовного модуля, залік).

Поточний контроль має на меті систематично відстежувати, наскільки студент розуміє та засвоює матеріал, вміє виконувати лабораторні роботи, самостійно вирішувати завдання та писати відповідний код, готувати звіти та представляти результати дослідження в усній або письмовій формі.

Підсумковий контроль покликаний оцінити глибину знань студента з усього курсу, його розуміння зв'язків між різними темами, здатність застосовувати отримані знання на практиці та самостійно вирішувати задачі з аналізу даних, які виникають в рамках дисципліни.

8. Критерії оцінювання та розподіл балів, які отримують здобувачі освіти

Таблиця 8.1 – Розподіл балів, які отримують здобувачі освіти

Складові навчальної роботи	Бали за одне заняття (завдання)	Кількість занять (завдань)	Сумарна кількість балів
Змістовний модуль 1			
Виконання і захист лабораторних робіт	0...5	5	0...25
Модульний контроль	0...10	3	0...30
Змістовний модуль 2			
Виконання і захист лабораторних	0...5	3	0...15
Модульний контроль	0...15	2	0...30
Усього за семестр			0...100

Семестровий контроль (*залік*) проводиться у разі відмови здобувача освіти від балів підсумкового контролю й за наявності допуску до *заліку*. Під час складання семестрового *заліку* здобувач освіти має можливість отримати максимум 100 балів. Білет для *заліку* складається з одного теоретичного і двох практичних питань. У першому запитанні студент має продемонструвати теоретичні знання. У другому запитанні – показати навички складання і виконання комп'ютерного додатку для вирішення задачі аналізу даних за заданою темою. У третьому пункті – продемонструвати вміння проводити перевірку отриманих результатів за допомогою стандартних функцій та бібліотек. Максимальна кількість балів за кожне запитання та їх суми показані в табл. 8.2.

Таблиця 8.2 – Шкали оцінювання відповідей на запитання білету

Складові білету	Складові оцінки	Бали за одне питання	Сумарна кількість балів
Пункт 1. Теоретичне запитання	Надано визначення поняття	0...10	0...20
	Наведено власний приклад використання	0...10	
Пункт 2. Практичне запитання	Складено алгоритм розрахунку (блок-схема) задачі	0...15	0...50
	Створено проект додатку для вирішення задачі	0...20	
	Отримано правильні результати розрахунків	0...15	
Пункт 3. Практичне запитання	Проведено перевірку отриманих результатів за допомогою стандартних функцій та бібліотек	0...30	0...30
Підсумкова оцінка			0...100

Таблиця 8.3 – Шкали оцінювання: бальна і традиційна

Сума балів	Оцінка за традиційною шкалою	
	Іспит, диференційний залік	Залік
90 – 100	Відмінно	Зараховано
75 – 89	Добре	
60 – 74	Задовільно	
0 – 59	Незадовільно	Не зараховано

Критерії оцінювання роботи здобувача освіти протягом семестру

Задовільно (60-74): мати володіння базовими знаннями та навичками, необхідними для досягнення програмних результатів навчання з аналізу даних; виконати та успішно здати лабораторні роботи, що охоплюють основні методи та алгоритми; написати модульні роботи, що демонструють розуміння ключових концепцій та технік аналізу даних; мати вміння застосовувати основні інструменти та бібліотеки для попередньої обробки даних, знати застосування алгоритмів аналізу та оцінки їхньої ефективності.

Добре (75-89): мати володіння знаннями та навичками, достатніми для досягнення програмних результатів навчання з аналізу даних; виконати та успішно здати лабораторні роботи, включаючи реалізацію та порівняння різних алгоритмів; написати модульні роботи, що демонструють глибоке розуміння концепцій та технік аналізу даних.

Додатково до вимог для оцінки «Задовільно»: вміння обґрунтовано вибирати відповідні алгоритми для розв'язання конкретних задач аналізу даних; вміння проводити експерименти з різними параметрами алгоритмів та інтерпретувати результати; вміння критично оцінювати результати аналізу даних та пропонувати шляхи для покращення моделей; вміння візуалізувати результати аналізу даних, використовуючи відповідні графіки та діаграми для їх інтерпретації.

Відмінно (90-100): володіння глибокими знаннями, уміннями та навичками, що дозволяють самостійно та обґрунтовано розв'язувати складні задачі аналізу даних; досконале виконання та захист всіх лабораторних робіт; успішно написані модульні роботи, що демонструють здатність до самостійного дослідження та аналізу проблем у; вміння самостійно використовувати широкий спектр інструментів та бібліотек для повного циклу аналізу даних: від збору та попередньої обробки до побудови та оцінки складних моделей; вміння реалізовувати власні алгоритми або модифікації існуючих алгоритмів для вирішення специфічних задач; вміння представляти результати аналізу даних у вигляді чітких та інформативних звітів, що містять обґрунтовані висновки та рекомендації.

9. Політика навчального курсу

Відвідування занять. З огляду на інтерактивний формат курсу, відвідування практичних занять є обов'язковим. Студенти, які не можуть регулярно відвідувати заняття, повинні протягом тижня узгодити з викладачем індивідуальний графік відпрацювання пропущених занять. Пропущені заняття відпрацьовуються на найближчій консультації протягом тижня після пропуску у формі співбесіди за питаннями з плану заняття. В окремих випадках дозволяється відпрацювання у письмовій формі шляхом виконання індивідуального письмового завдання.

Дотримання вимог академічної доброчесності здобувачами освіти під час вивчення навчальної дисципліни. Під час вивчення навчальної дисципліни здобувачі освіти мають дотримуватися загальноприйнятих морально-етичних норм і правил поведінки, вимог академічної доброчесності, передбачених Положенням про академічну доброчесність Національного аерокосмічного університету «Харківський авіаційний інститут» (<https://khai.edu/assets/files/polozhennya/polozhennya-pro-akademichnu-dobrochesnist.pdf>). Очікується, що роботи здобувачів освіти будуть їх оригінальними дослідженнями або міркуваннями. Відсутність посилань на використані джерела, фабрикавання джерел, списування, втручання в роботу інших здобувачів освіти становлять, але не обмежують, приклади можливої академічної недоброчесності. Виявлення ознак академічної недоброчесності в письмовій роботі здобувача освіти є підставою для її незарахування викладачем незалежно від масштабів плагіату чи обману.

Вирішення конфліктів. Порядок і процедури врегулювання конфліктів, пов'язаних із корупційними діями, зіткненням інтересів, різними формами дискримінації, сексуальними домаганнями, міжособистісними стосунками та іншими ситуаціями, що можуть виникнути під час навчання, а також правила етичної поведінки регламентуються Кодексом етичної поведінки в Національному аерокосмічному університеті «Харківський авіаційний інститут» (<https://khai.edu/ua/university/normativna-baza/ustanovchi-dokumenti/kodeks-etichnoi-povedinki/>).

10. Методичне забезпечення

1. Аналіз даних [Електронний ресурс] : навч. метод. забезп. / Н. С. Бакуменко. – Х. : Нац. аерокосм. ун-т ім. М. Є. Жуковського «Харк. авіац. ін-т», 2019. – 28 с. Режим доступу: http://library.khai.edu/library/fulltexts/doc/_002_Anali_Dani.pdf

Увесь науково методичний комплект з дисципліни розміщено на освітньому порталі Mentor: <https://mentor.khai.edu/course/view.php?id=3156>.

11. Рекомендована література

Базова

1. Бахрушин В. Є. Аналіз даних. – Запоріжжя: ГУ «ЗІДМУ», 2006. – 170 с.
2. Гирко В. Л. Багатомірний статистичний аналіз. – К.: Вища школа, 1988. – 320 с.
3. Корольок В. С., Боровских Ю. В. Асимптотичний аналіз розподілу статистик. – К.: Наукова думка, 1984. – 301 с.
4. Паніотто В. І., Максименко В. С., Харченко Н. М. Статистичний аналіз соціологічних даних. – К.: Вид. дім «КМ Академія», 2004. – 270с.
5. Leskovec J., Rajaraman A., Ullman J. Mining of Massive Datasets. 3rd Edition. Cambridge: Cambridge University Press, 2020. 603 p. URL: <http://www.mmids.org/>
6. Марченко О. О., Россада Т. В. Актуальні проблеми Data Mining: Навчальний посібник для студентів факультету комп'ютерних наук та кібернетики. Київ. 2017. 150 с.
7. Zhang S. Challenges in KNN Classification // IEEE Transactions on Knowledge & Data Engineering. – 2022. – Vol. 34, No. 10. – P. 4663–4675. – DOI: 10.1109/TKDE.2021.3049250.
8. Zaichenko Yu. P. Fundamentals of Designing Intelligent Systems: A Textbook, Kyiv: Slovo, 2004, 352.
9. Закон України «Про вищу освіту» від 01.07.2014 р. № 1556-VII // *Відомості Верховної Ради України*. – 2014. – № 37–38. – Ст. 2004. Режим доступу: <https://zakon.rada.gov.ua/laws/show/1556-18>

Допоміжна

1. Жлуктенко В. І., Наконечний С.І., Савіна С.С. Теорія ймовірностей і математична статистика. – К.: КНЕУ, 2001. – 336 с.
2. Іващенко П. О., Семеняк І.В., Іванов В.В. Багатовимірний статистичний аналіз. – Харків: Основа, 1992. – 144 с.

12. Інформаційні ресурси

1. James G., Witten D., Hastie T., Tibshirani R. An Introduction to Statistical Learning : with Applications in R. – 2nd ed. – Springer, 2021. [Електронний ресурс]. URL: <https://www.casact.org/sites/default/files/2022-12/James-G.-et-al.-2nd-edition-Springer-2021.pdf>
2. Analysis of Variance (ANOVA) [Електронний ресурс]. URL: <https://numiqo.com/tutorial/anova>
3. Correlation analysis [Електронний ресурс]. URL: <https://numiqo.com/tutorial/correlation>
4. Understanding Correlation: Measuring Relationships in Data [Електронний ресурс]. URL: <https://www.datacamp.com/tutorial/correlation>

5. Mean squared error [Электронный ресурс]. URL: <https://www.britannica.com/science/mean-squared-error>
6. Root Mean Square Error (RMSE) [Электронный ресурс]. URL: <https://itwiki.dev/data-science/ml-reference/ml-glossary/root-mean-square-error>
7. Mean Absolute Error [Электронный ресурс]. URL: <https://www.sciencedirect.com/topics/engineering/mean-absolute-error>
8. Mean Absolute Percentage Error (MAPE): What You Need To Know [Электронный ресурс]. URL: <https://arize.com/blog-course/mean-absolute-percentage-error-mape-what-you-need-to-know/>
9. MdAPE - Median Absolute Percentage Error [Электронный ресурс]. URL: <https://support.numxl.com/hc/en-us/articles/115001223503-MdAPE-Median-Absolute-Percentage-Error>
10. Mean Squared Error [Электронный ресурс]. URL: <https://www.geeksforgeeks.org/mean-squared-error/>
11. HODSON, Timothy O. Root mean square error (RMSE) or mean absolute error (MAE): When to use them or not. Geoscientific Model Development Discussions, v. 2022, p. 1–10, 2022.
12. What is the Mean Absolute Error? [Электронный ресурс]. URL: <https://www.deepchecks.com/glossary/mean-absolute-error/>
13. What Is MAPE? Your Guide to Mean Absolute Percentage Error [Электронный ресурс]. URL: <https://www.indeed.com/career-advice/career-development/what-is-mape>
14. How to make your data cumulative [Электронный ресурс]. URL: <https://helpcenter.flourish.studio/hc/en-us/articles/8761522991503-How-to-make-your-data-cumulative>