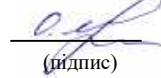


Міністерство освіти і науки України
Національний аерокосмічний університет
«Харківський авіаційний інститут»

Кафедра математичного моделювання та штучного інтелекту (№ 304)

ЗАТВЕРДЖУЮ

Гарант освітньої програми


(підпис)

Олена МЕДВЕДЬ

(ім'я та прізвище)

« 29 » серпня 2025 р.

СИЛАБУС ОBOB'ЯЗKОВОЇ НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

АНАЛІЗ ТА ВІЗУАЛІЗАЦІЯ ДАНИХ

(назва навчальної дисципліни)

Галузь знань: В «Культура, мистецтво та гуманітарні науки»

(шифр і найменування галузі знань)

Спеціальність: В11 Філологія

(код і найменування спеціальності)

Освітня програма: Прикладна лінгвістика

(найменування освітньої програми)

Рівень вищої освіти: другий (магістерський)

Силабус введено в дію з 01.09.2025

Харків 2025 рік

Розробник: Пічугіна О.С., професор каф. 304, д.ф.-м.н., професор

(прізвище та ініціали, посада, науковий ступінь і вчене звання)



(підпис)

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

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 ПРИЗВИЩЕ)

Силабус навчальної дисципліни погоджено на випусковій кафедрі прикладної лінгвістики (№703)

Протокол № 1 від «29» серпня 2025 р.

Завідувач кафедри прикладної лінгвістики

канд. філол. наук, доцент

(назва кафедри, науковий ступінь та вчене звання завідувача)



(підпис)

Вікторія РИЖКОВА

(ім'я та ПРИЗВИЩЕ)

Погоджено з представником здобувачів освіти:

Здобувач другого (магістерського)
рівня вищої освіти за ОП
«Прикладна лінгвістика»

(посада)



(підпис)

Ірина ВОРОБЕЙ

(ім'я та ПРИЗВИЩЕ)

1. Загальна інформація про викладача



ПІБ: Пічугіна Оксана Сергіївна

Посада: професор кафедри математичного моделювання та штучного інтелекту (304)

Науковий ступінь: д.ф.-м.н.

Вчене звання: професор

Перелік дисциплін, які викладає:

1. Системи та методи прийняття рішень
 2. Інтелектуальні системи підтримки прийняття рішень
 3. Електронний документообіг та електронне урядування
 4. Аналіз та візуалізація даних
 5. Цифровізація та вебометрія документно-інформаційних потоків
 6. Сучасна теорія геометричного проєктування
 7. Мінор «ШІ мовою Python». Дисципліна 2
- Напрями наукових досліджень: дослідження операцій, комбінаторна оптимізація, математичне моделювання, чисельні методи, машинне навчання, аналіз даних.

Контактна інформація: o.pichugina@khai.edu

2. Опис навчальної дисципліни

Форма здобуття освіти	Денна
Семестр	1
Мова викладання	Українська
Тип дисципліни	Обов'язкова
Обсяг дисципліни: кредити ЄКТС / кількість годин	6 кредитів ЄКТС / 180 годин (64 аудиторних; СРЗ – 116)
Види навчальної діяльності	Практичні заняття, самостійна робота
Види контролю	Поточний контроль, модульний контроль, семестровий контроль – іспит
<u>Пререквізити</u>	<u>Англійська мова. Практика мови</u> <u>Математичні основи лінгвістики</u> <u>Теорія алгоритмів та обчислювальних процесів</u> <u>Основи інформатики і обчислювальної техніки</u> <u>Обчислювальна техніка та програмування</u>

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

Мета – формування у здобувачів вищої освіти цілісного уявлення про сучасні підходи до аналізу даних та набуття практичних навичок роботи з інструментами статистики, машинного навчання та обробки природної мови. Дисципліна спрямована на опанування базових понять машинного навчання: лінійних регресій і методів оцінювання їх параметрів, описової статистики, ймовірнісних розподілів та розвідувального аналізу даних, а також основ NLP. Фокус зроблено на практичній роботі в Excel, R та Python.

Завдання – поетапне ознайомлення здобувачів із ключовими концепціями машинного навчання (регресія, кореляція, прогноз) та їх реалізацією на реальних вибірках; формування вміння будувати парні й множинні лінійні регресійні моделі, оцінювати їхні параметри різними, аналізувати адекватність моделей за допомогою статистичних критеріїв і візуалізації. Дисципліна спрямована на розвиток компетентностей в описовій статистиці (генеральна сукупність, вибірка, вибіркові середнє та дисперсія, ймовірнісні розподіли), здатності виконувати EDA за допомогою таблиць, гістограм, box- та scatter-графіків, кореляційних і щільнісних графіків, а також опанування основних етапів і інструментів NLP: токенизація, лематизація, стемінг, побудова векторних і мовних моделей, текстова класифікація.

Інтегральна компетентність:

здатність розв'язувати складні задачі аналізу та інтерпретації числових і текстових даних, що передбачає застосування статистичних методів, базових алгоритмів машинного навчання й обробки природної мови, засобів візуалізації та програмних інструментів Excel, R, Python; характеризується умінням працювати з реальними статистичними даними, оцінювати якість моделей і робити обґрунтовані прогнози висновки для підтримки прийняття рішень.

Загальні компетентності (ЗК)

Після закінчення цієї програми здобувач освіти буде здатен:

ЗК 2. Здатність бути критичним і самокритичним.

ЗК 3. Здатність до пошуку, опрацювання та аналізу інформації з різних джерел.

ЗК 6. Здатність спілкуватися іноземною мовою.

ЗК 7. Здатність до абстрактного мислення, аналізу та синтезу.

ЗК 8. Навички використання інформаційних і комунікаційних технологій.

ЗК 11. Здатність проведення досліджень на належному рівні.

Спеціальні компетентності (ФК)

Після закінчення цієї програми здобувач освіти буде здатен:

ФК4. Здатність здійснювати науковий аналіз і структурування мовного / мовленнєвого/літературного матеріалу з урахуванням класичних і новітніх методологічних принципів.

ФК6. Здатність застосовувати поглиблені знання з обраної філологічної спеціалізації для вирішення професійних завдань, а саме: здатність забезпечувати відбір, аналіз, оцінку, систематизацію, моніторинг, організацію, зберігання, розповсюдження та надання в користування інформації та знань у будь-яких форматах; здатність ефективної організації навчально-виховного процесу; здатність до напрацювання умінь, навичок та здібностей, що забезпечують успішність навчально-виховного процесу; здатність використовувати методи систематизації, пошуку, збереження, класифікації інформації для різних типів контенту; здатність використовувати сучасні прикладні комп'ютерні технології та програмне забезпечення для вирішення завдань спеціальності; здатність проектувати та створювати інформаційні системи, ресурси, продукти, системи та послуги; здатність застосовувати принципи проектування баз даних та інформаційних систем; здатність до проведення дескриптивного лінгвістичного аналізу, статистичного аналізу, методів візуалізації даних та удосконалення навичок роботи в Excel та програмуванні на Python та R.

Програмні результати навчання (ПРН):

Після закінчення цієї програми здобувач освіти буде здатен:

ПРН 2. Упевнено володіти державною та іноземною мовами для реалізації письмової та усної комунікації, зокрема в ситуаціях професійного й наукового спілкування; презентувати результати досліджень державною та іноземною мовами.

ПРН 4. Оцінювати й критично аналізувати соціально, особистісно та професійно значущі проблеми і пропонувати шляхи їх вирішення у складних і непередбачуваних умовах, що потребує застосування нових підходів та прогнозування.

ПРН 11. Здійснювати науковий аналіз мовного, мовленнєвого й літературного матеріалу, інтерпретувати та структурувати його з урахуванням доцільних методологічних принципів, формулювати узагальнення на основі самостійно опрацьованих даних.

ПРН 15. Обирати оптимальні дослідницькі підходи й методи для аналізу конкретного лінгвістичного чи літературного матеріалу.

4. Зміст навчальної дисципліни

Модуль

Змістовний модуль 1. Вступ у машинне навчання. Регресії, описова статистика, розвідувальний аналіз даних

Тема 1. Штучний інтелект і машинне навчання: вступ

Стисла анотація. Розглянуто поняття штучного інтелекту (AI) та машинного навчання (ML), їхню еволюцію, основні методи та сфери застосування. Висвітлюється роль статистичних підходів у розвитку AI, а також перехід до сучасних алгоритмів машинного навчання — регресії, класифікації, кластеризації, дерев рішень. Розкривається суть моделі навчання з учителем, приклад побудови лінійної регресії, ідея прогнозування на основі експериментальних даних. У практичній частині студенти знайомляться з інструментами Excel та Python для реалізації базових аналітичних операцій - матричних обчислень, використання надбудов «Solver» і «Аналіз даних».

Тема лекції 1. Штучний інтелект і машинне навчання: сутність, історія, приклади

Тема лекції 2. Типологія методів машинного навчання

Тема практичного заняття 1. Реалізація базових матричних операцій, аналіз даних та оптимізація в Excel

Тема практичного заняття 2. Реалізація аналогічних процедур у Python

Самостійна робота здобувача освіти: опрацювання матеріалу лекції, виконання індивідуального практичного завдання, підготовка до захисту практичного завдання.

Тема 2. Парна та множинна лінійні регресії. Метод найменших квадратів і перевірка адекватності моделі

Стисла анотація: Поняття кореляційного поля та стохастичної залежності між фактором і показником. Означення регресії, мета побудови регресійної моделі й прогнозу. Парна лінійна регресія, вибір форми залежності, поняття експериментальних і теоретичних значень показника, похибок спостережень. Методи оцінювання параметрів парної регресії: метод «натягнутої нитки», метод сум, метод найменших квадратів (МНК). Множинна лінійна регресія, векторно-матричний запис моделі, матрична форма МНК, система нормальних рівнянь та умови її розв'язності. Регресійна та залишкова суми квадратів, коефіцієнт детермінації, дисперсія й середньоквадратичне відхилення адекватності. Перевірка адекватності моделі реальним даним за критерієм Фішера, формулювання нульової та альтернативної гіпотез, інтерпретація результатів. Реалізація побудови регресії та аналізу адекватності в Excel за допомогою лінії тренду та вбудованих функцій. Приклад прогнозування зарплати за досвідом роботи.

Тема лекції 3. Парна лінійна регресія та основні методи оцінювання параметрів

Тема лекції 4. Множинна лінійна регресія, МНК у матричній формі та перевірка адекватності

Тема практичного заняття 3: Побудова парної лінійної регресії в Excel різними методами

Тема практичного заняття 4: Використання функції ЛИНЕЙН(), оцінка якості моделі та перевірка адекватності

Самостійна робота здобувача освіти: опрацювання матеріалу лекції, виконання індивідуального практичного завдання, підготовка до захисту практичного завдання.

Тема 3. Статистичні основи аналітики практичних задач

Стисла анотація: Поняття невизначеності в практичних рішеннях та роль імовірності для її кількісної оцінки. Вступ до описової статистики: середнє, медіана, мода, мінімум і максимум, розмах, квантілі та квартилі, стандартне відхилення та дисперсія, їх інтерпретація на прикладах бізнес-показників (прибуток, виручка, продуктивність). Розрізнення «генеральної сукупності» та «вибірки», причини використання вибірок та огляд основних методів вибіркового відбору (випадкова, систематична, кластерна вибірка). Базові поняття теорії ймовірностей: випадкові події та їх ймовірність. Основні ймовірнісні розподіли, що використовуються в бізнес-аналітиці: нормальний, біноміальний, розподіл Пуассона, а також умовна ймовірність та її прикладні інтерпретації. Практичні приклади реалізації обчислень у R та Python (mean,

median, quantile, sd, var, summary(), describe(), pnorm(), pbinom(), ppois()), робота з data frame та побудова кореляційних полів.

Тема лекції 5. Генеральна сукупність, вибірка та описова статистика в бізнес-аналітиці

Тема лекції 6. Дисперсія, стандартне відхилення, ймовірність і розподіли в бізнес-задачах

Тема практичного заняття 5: Описова статистика, вибірки та базові обчислення в R і Python

Тема практичного заняття 6: Ймовірність, розподіли, умовна ймовірність і візуалізація

Самостійна робота здобувача освіти: опрацювання матеріалу лекції, виконання індивідуального практичного завдання, підготовка до захисту практичного завдання.

Тема 3. Розвідувальний аналіз даних (Exploratory Data Analysis, EDA)

Стисла анотація теми. EDA - це початковий етап аналітики даних, який включає дослідження, опис й візуалізацію даних з метою виявлення їх структури, аномалій та закономірностей. Розглядаються основні методи описової статистики та візуалізації: таблиці, гістограми, box-плоти, scatter-плоти, кореляційні та щільнісні графіки. Вивчаються засоби одномірного та багатомірного, зокрема двомірного, аналізу та візуалізації даних. Досліджуються підходи до обробки, інтерпретації та вибору оптимального засобу візуалізації числових та категоріальних даних

Тема лекції та практичного заняття 7: Описова статистика та табличний аналіз у Excel, R та Python

Тема лекції та практичного заняття 8: Графічний розвідувальний аналіз даних у Excel, R та Python

Самостійна робота здобувача освіти: опрацювання матеріалу лекції, виконання індивідуального практичного завдання, підготовка до захисту практичного завдання.

Модульний контроль 1

Змістовний модуль 1. Обробки природної мови

Тема 4. Вступ до обробки природної мови (Natural Language Processing)

Стисла анотація теми. Тема присвячена основам обробки природної мови (Natural Language Processing, NLP). NLP поєднує лінгвістику, штучний інтелект і машинне навчання. Розглядаються етапи аналізу мови, методи попередньої обробки тексту, формування векторних представлень, класичні статистичні й нейронні підходи до NLP, а також приклади реалізації в Excel, Python та R. Вивчається, як теоретично і практично виконувати токенізацію, лематизацію, стемінг, вилучення ознак, побудову моделей класифікації текстів і застосування Word2Vec. Опановуються основні NLP бібліотеки у Python і R: NLTK, Spacy, Scikit-learn, Gensim та Text2Vec. Здійснюється ознайомлення з методами візуалізації результатів NLP-аналізу з використанням Python засобів (Matplotlib, Seaborn, WordCloud, Plotly) та R (ggplot2, tidytext, LDAvis).

Тема практичного заняття 9: Ознайомлення з основними сферами застосування NLP

Тема лекції та практичного заняття 10:Лінгвістичні основи та структура мови

Тема лекції та практичного заняття 11:Вирішення проблем неоднозначності у NLP

Тема лекції та практичного заняття 12:Робота з корпусами та регулярними виразами

Тема лекції та практичного заняття 13:Використання NLP-бібліотек Python та R

Тема лекції та практичного заняття 14:Побудова preprocessing-pipeline

Тема лекції та практичного заняття 15:Побудова векторних та мовних моделей

Тема лекції та практичного заняття 16:Текстова класифікація та Word2Vec
Самостійна робота здобувача освіти: опрацювання матеріалу лекції, виконання індивідуального практичного завдання, підготовка до захисту практичного завдання.

Модульний контроль 2

5. Індивідуальні завдання

Індивідуальне завдання не передбачено навчальним планом.

6. Методи навчання

Словесні, наочні, практичні.

7. Методи контролю

Поточний контроль (теоретичне опитування й розв'язання практичних завдань), модульний контроль (тестування за розділами курсу) та підсумковий (семестровий) контроль (іспит).

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

Таблиця 8.1 – Розподіл балів, які отримують здобувачі

Складові навчальної роботи	Бали за одне заняття (завдання)	Кількість занять (завдань)	Сумарна кількість балів
Змістовний модуль 1			
Участь у лекції	0...8	2	0...16
Виконання і захист практичних робіт	0...7	4	0... 28
Модульний контроль 1	0...6	1	0...6
Змістовний модуль 2			
Участь у лекції	0...8	2	0...16
Виконання і захист практичних робіт	0...7	4	0... 28
Модульний контроль 2	0...6	1	0...6
Усього за семестр			0...100

Семестровий контроль (іспит) проводиться у разі відмови здобувача від балів поточного тестування й за наявності допуску до іспиту. Під час складання семестрового іспиту здобувач має можливість отримати максимум 100 балів. Білет для іспиту складається з двох теоретичних і двох практичних завдань, максимальна кількість балів за кожне виконане теоретичне завдання – 20 балів, за кожне виконане практичне завдання – 30 балів (що складає в сумі 100 балів).

Таблиця 8.2 – 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 протягом семестру

Задовільно (60 - 74). Мати мінімум знань та умінь. Мати стійкі навички виконання елементарних дій з опрацювання даних на комп'ютері, здати практичні роботи, здати 2 модульні контрольні роботи.

Добре (75 - 89). Вільно володіти навчальним матеріалом, вміти застосовувати знання на практиці. Мати стійкі навички виконання елементарних дій з опрацювання даних на комп'ютері, здати усі практичні роботи, здати 2 модульні контрольні роботи.

Відмінно (90 - 100). Здати всі контрольні точки з оцінкою «відмінно». Досконально знати всі теми та вміти застосовувати одержані знання.

9. Політика навчального курсу

Відвідування занять та робота в семестрі. Відпрацювання пропущених занять відбувається відповідно до розкладу консультацій, за попереднім погодженням з викладачем.

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

Сторінка дисципліни знаходиться за посиланням:
<https://mentor.khai.edu/course/view.php?id=8547>

11. Рекомендована література

Базова

1. Бахрушин В.Є. Методи аналізу даних : навчальний посібник для студентів. – Запоріжжя : КПУ, 2011. – 268 с. ISBN 978-966-414-103-8.
2. Hodeghatta U.R., Nayak U. Practical Business Analytics Using R and Python: Solve Business Problems Using a Data-driven Approach. – Berkeley, CA: Apress Publishers, 2023.– 706 с. <https://link.springer.com/10.1007/978-1-4842-8754-5>
3. Evergreen S. Effective Data Visualization: The Right Chart for the Right Data. – Los Angeles: Sage Publications, 2019.– 352 с. <https://www.amazon.com/Effective-Data-Visualization-Right-Chart/dp/1544350880>

Допоміжна

1. Healy K. Data Visualization: A Practical Introduction. – Princeton, NJ: Princeton University Press, 2018.– 296 с. <https://dokumen.pub/data-visualization-a-practical-introduction-0691181616-9780691181615.html>
2. Hammond M. Python for Linguists. – Cambridge: Cambridge University Press, 2020.– 310 с. <https://doi.org/10.1017/9781108642408>. ISBN 978-1-108-49344-4.
3. Грицюк П.М., Остапчук О.П. Аналіз даних: навчальний посібник.– Рівне: НУВГП, 2008. – 218 с.
4. Тиркусова Н.В., Боровик В.О., Глущенко Л.О. Аналіз даних : навч. посіб.– Суми : СумДУ, 2008.– 204 с.
5. Литвин В.В., Нікольський Ю.В., Пасічник В.В. Аналіз даних та знань. Навчальний посібник.– Львів : Магнолія, 2006.– 276 с.

12. Інформаційні ресурси

1. Google Colab <https://colab.google/>
2. Pyomo <https://www.pyomo.org/>
3. Using R language with Anaconda <https://www.anaconda.com/docs/getting-started/working-with-conda/packages/using-r-language>
4. Microsoft Excel <https://www.microsoft.com/uk-ua/microsoft-365/excel>