

Міністерство освіти і науки України

Харківський національний університет імені В. Н. Каразіна

Кафедра економічної кібернетики та прикладної економіки



РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Вступ до Big data

(назва навчальної дисципліни)

рівень вищої освіти Другий (магістерський)

галузь знань С «Соціальні науки, журналістика, інформація та міжнародні відносини»

спеціальність С1 Економіка та міжнародні економічні відносини

освітні програми Прикладна економіка

спеціалізація С1. 01 Економіка
(шифр і назва)

вид дисципліни обов'язкова
(обов'язкова / за вибором)

факультет Економічний

2025 / 2026 навчальний рік

Програму рекомендовано до затвердження вченою радою економічного факультету

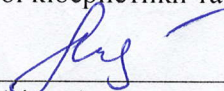
« 26 » серпня 2025 року, протокол № 18

Розробник програми: Кирило МАНАХОВ

Програму схвалено на засіданні кафедри економічної кібернетики та прикладної економіки

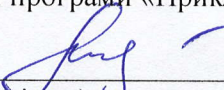
Протокол від « 26 » серпня 2025 року № 1

Завідувач кафедри економічної кібернетики та прикладної економіки


_____ Тамара МЕРКУЛОВА
(підпис) (прізвище та ініціали)

Програму погоджено з гарантом освітньо-професійної програми «Прикладна економіка»

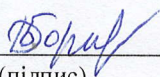
Гарант освітньо-професійної програми «Прикладна економіка»


_____ Тамара МЕРКУЛОВА
(підпис) (прізвище та ініціали)

Програму погоджено науково-методичною комісією економічного факультету

Протокол від « 26 » серпня 2025 року № 1

Голова науково-методичної комісії економічного факультету


_____ Дар'я ЗАГОРСЬКА
(підпис) (прізвище та ініціали)

ВСТУП

Програма навчальної дисципліни «Вступ до Big data» складена відповідно до освітньо-професійної програми другого (магістерського) рівня вищої освіти спеціальності C1 Економіка та міжнародні економічні відносини, освітня програма «Прикладна економіка».

1. Опис навчальної дисципліни

1.1. Формування системи теоретичних знань та практичних навичок з роботи з великими даними (Big Data).

1.2. Оволодіння принципами, методами та інструментами роботи з великими обсягами даних, а також застосування цих знань для вирішення типових завдань аналізу даних у бізнесі за допомогою сучасних технологій.

1.3. Кількість кредитів – 5 кредити.

1.4. Загальна кількість годин – 150 годин.

1.5. Характеристика навчальної дисципліни	
Нормативна / за вибором За вибором	
Денна форма навчання	Заочна (дистанційна) форма навчання
Рік підготовки	
1-й	1-й
Семестр	
1-й	1-й
Лекції	
32 год.	6 год.
Лабораторні заняття	
10 год.	6 год.
Самостійна робота, у тому числі	
108 год.	138 год.
Контрольні роботи	
2 год.	2 год.

1.6. Перелік компетентностей, що формує дана дисципліна

Інтегральна компетентність: Здатність визначати та розв'язувати складні економічні задачі та проблеми, приймати відповідні аналітичні та управлінські рішення у сфері економіки або у процесі навчання, що передбачає проведення досліджень та/або здійснення інновацій за невизначених умов та вимог.

ЗК1. Здатність генерувати нові ідеї (креативність).

ФК3. Здатність збирати, аналізувати та обробляти статистичні дані, науково-аналітичні матеріали, які необхідні для розв'язання комплексних економічних проблем, робити на їх основі обґрунтовані висновки.

ФК4. Здатність використовувати сучасні інформаційні технології, методи та прийоми дослідження економічних та соціальних процесів, адекватні встановленим потребам дослідження.

ФК8. Здатність оцінювати можливі ризики, соціально-економічні наслідки управлінських рішень.

ФК13*. Здатність до пошуку, агрегуванню та аналізу великих масивів інформації релевантної для вирішення професійних задач.

1.7. Перелік результатів навчання, що формує дана дисципліна

1. Формулювати, аналізувати та синтезувати рішення науково-практичних проблем.

2. Розробляти, обґрунтовувати і приймати ефективні рішення з питань розвитку соціально-економічних систем та управління суб'єктами економічної діяльності.

8. Збирати, обробляти та аналізувати статистичні дані, науково-аналітичні матеріали, необхідні для вирішення комплексних економічних завдань.

10. Застосовувати сучасні інформаційні технології та спеціалізоване програмне забезпечення у соціально-економічних дослідженнях та в управлінні соціально-економічними системами.

17*. Застосовувати математичне та комп'ютерне моделювання для аналізу теоретичних та прикладних проблем в економіці та бізнесу.

1.8. Пререквізити: засвоєння дисциплін «Інформатика», «Економетрика», «Інтелектуальні системи аналізу даних»

2. Тематичний план навчальної дисципліни

Розділ 1. Основи аналізу даних з використанням Python

Тема 1. Використання pandas для аналізу даних

Тема 2. Вибір даних та переформатування

Розділ 2. Моделювання та регресійний аналіз

Тема 3. Аналіз даних з використанням нелінійної регресії

Розділ 3. Обробка даних та підготовка до аналізу

Тема 4. Обробка пропущених значень та категоріальних змінних

Тема 5. Крос-валідація та аналіз витоків даних

Розділ 4. Основи та технології Big Data

Тема 6. Основні поняття та технології Big Data

Тема 7. Архітектура систем Big Data

Тема 8. Hadoop екосистема

Тема 9. NoSQL бази даних

Розділ 5. Обробка та візуалізація даних

Тема 10. Масштабовані обчислення з використанням MapReduce

Тема 11. Поточкова обробка даних (Stream Processing)

Тема 12. Візуалізація великих обсягів даних та бізнес-інтелект

3. Структура навчальної дисципліни

Назви розділів	Кількість годин											
	денна форма						заочна форма					
	усього	у тому числі					усього	у тому числі				
		л	п	лаб.	інд.	с.р.		л	п	лаб.	інд.	с.р.
1	2	3	4	5	6	7	8	9	10	11	12	13
Розділ 1. Основи аналізу даних з використанням Python												
Разом за розділом 1	29	6		2		21	28	1		1		26
Розділ 2. Моделювання та регресійний аналіз												
Разом за розділом 2	29	6		2		21	28	1		1		26
Розділ 3. Обробка даних та підготовка до аналізу												
Разом за розділом 3	29	6		2		21	29	1		2		26
Розділ 4. Основи та технології Big Data												
Разом за розділом 4	29	6		2		21	28	1		1		26
Розділ 5. Обробка та візуалізація даних												
Разом за розділом 5	34	8		2		24	37	2		1		34
Усього годин	150	32		10		108	150	6		6		138

4. Теми лабораторних занять

№ з/п	Назва теми	Кількість годин	
		денна	заочна
1	Використання pandas для аналізу даних	1	
2	Вибір даних та переформатування	1	1
3	Аналіз даних з використанням нелінійної регресії	1	
4	Обробка пропущених значень та категоріальних змінних	1	1
5	Крос-валідація та аналіз витоків даних	1	1
6	Hadoop екосистема	1	1
7	NoSQL бази даних	1	
8	Масштабовані обчислення з використанням MapReduce	1	1

9	Візуалізація великих обсягів даних та бізнес-інтелект	2	1
	Разом	10	6

5. Завдання для самостійної роботи

№ з/п	Види, зміст самостійної роботи	Кількість годин	
		денна	заочна
	Розділ 1		
	Збір та підготовка даних за обраною тематикою	21	26
	Розділ 2		
	Аналіз даних з використанням нелінійної регресії	10	13
	Застосування моделей регресії на прикладах даних про ціни на житло.	11	13
	Розділ 3		
	Методи обробки категоріальних змінних: видалення, порядкове кодування та One-Hot кодування.	10	13
	Важливість крос-валідації для оцінки моделей машинного навчання.	11	13
	Розділ 4		
	Значення великих даних у сучасній економіці та бізнесі.	10	13
	Додаткові інструменти Hadoop: Hive, Pig, HBase та інші компоненти для ефективного обробки даних.	11	13
	Розділ 5		
	Огляд технологій та інструментів для потокової обробки даних: Apache Kafka, Apache Flink, Apache Storm.	12	17
	Вивчення сучасних інструментів та методів візуалізації даних: Tableau, Power BI, D3.js.	12	17
	Разом	108	138

6. Індивідуальні завдання

Індивідуальні завдання не передбачені навчальним планом.

7. Методи навчання

Форми та методи навчання: читання лекцій, проведення практичних занять та виконання самостійної роботи.

8. Методи контролю

Методи контролю передбачають поточний контроль, виконання самостійних та дві контрольні роботи. Підсумковий контроль проводиться у формі екзамену, що проводиться у письмовій формі і включає практичне завдання (дві задачі, 40 балів).

9. Схема нарахування балів

Поточний контроль, самостійна робота, індивідуальні завдання	Екзаменаційна робота (письмова)	Сума
Самостійні роботи		
60	40	100

Критерії оцінювання

- 1) Дві контрольні роботи (по 15 балів кожна) складаються з 2 практичних завдань, які виконуються та здаються в електронному вигляді.
- 2) Поточний контроль та самостійна робота складається з 10 практичних завдань по 6 балів кожне

Оцінка	Критерії оцінювання
10-8	Завдання виконано в повному обсязі та чітко з дотриманням усіх вимог. Своєчасно представлено на перевірку.
7-6	Завдання виконано повністю, допущено незначні помилки. АБО Завдання виконано в повному обсязі та чітко з дотриманням усіх вимог. Представлено на перевірку із запізненням.
6-4	Завдання виконано не повністю, допущено значні помилки. Своєчасно представлено на перевірку. АБО Завдання виконано повністю, допущено незначні помилки. Представлено на перевірку із запізненням.
3-0	Завдання виконано не повністю, допущено значні помилки. Представлено на перевірку зі значним запізненням.

3) Екзаменаційна робота (40 балів)

Робота складається з 2-х практичних завдань.

Звіт по кожному завданню включає: Скрипт з завантаженими та обробленими без помилок екзаменаційними даними та документ звіту - файл у форматі MS Word «Прізвище.docx»

Кількість балів	Критерії оцінювання
18-20	Здобувач вищої освіти правильно обирає метод розв'язання задачі, володіє різнобічними вміннями, навичками та прийомами рішення завдань. Завдання виконане без помилок. Звіт містить розгорнуті авторські коментарі. Код скрипту чітко структурований
15-17	Здобувач вищої освіти правильно застосовує теоретичні знання та положення при рішенні практичної задачі, володіє необхідними вміннями та навичками роботи з програмами. Виконав завдання з окремими незначними помилками. Звіт містить стислі авторські коментарі. Код скрипту чітко структурований
12-14	Здобувач вищої освіти правильно застосовує теоретичні знання та положення при рішенні практичної задачі, володіє необхідними вміннями та навичками роботи з програмами. Виконав завдання з окремими незначними помилками. Звіт не містить авторських коментарів. Код скрипту структурований не чітко
9-11	Здобувач вищої освіти при розв'язанні практичної задачі допустив значну помилку. Не досить вільно володіє вміннями та навичками роботи з програмами. Звіт не містить авторських коментарів. Код скрипту структурований не чітко
6-8	Здобувач вищої освіти при розв'язанні практичної задачі допустив значну помилку. Звіт не сформовано. Код скрипту містить незначні помилки
3-5	Здобувач вищої освіти при розв'язанні практичної задачі допустив значну помилку. Звіт не сформовано. Код скрипту містить значні помилки
0-2	Здобувач вищої освіти не уміє застосовувати знання на практиці. Не вирішив завдання взагалі або допустив грубих помилок. Звіт не сформовано. Код скрипту містить значні помилки

Шкала оцінювання

Сума балів за всі види навчальної діяльності протягом семестру	Оцінка
	для чотирирівневої шкали оцінювання
90 – 100	відмінно
70-89	добре
50-69	задовільно
1-49	незадовільно

10. Рекомендована література

Основна література

1. Віктор Майєр-Шонбергер, Кеннет Кук'єр – Big Data: A revolution that will transform how we live, work, and think.
2. Том Вайт – Hadoop: The definitive guide.
3. Білл Чемберс, Матеї Захарія – *Spark: the definitive guide*
4. Прамод Дж. Садаларе, Мартін Фаулер – NoSQL distilled: A brief guide to the rmerging world of polyglot persistence.
5. Фостер Провост, Том Фосетт – *Data science for business*.

Допоміжна література

1. Мартін Клеппманн – Designing data-intensive applications.
2. Метью А. Рассел – Mining the social web.
3. Бенджамін Вільямс – Big Data fundamentals: concepts, drivers & techniques.
4. Кеті О'Ніл, Джо Тібер – Weapons of math destruction: How big data increases inequality and threatens democracy.
5. Стівен Лінч – Big Data and analytics: Transforming business using machine learning.

Посилання на інформаційні ресурси в Інтернеті, відео-лекції, інше методичне забезпечення

1. <https://moodle.karazin.ua/>