

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ДЕРЖАВНИЙ ТОРГОВЕЛЬНО-ЕКОНОМІЧНИЙ УНІВЕРСИТЕТ
ВІННИЦЬКИЙ ТОРГОВЕЛЬНО-ЕКОНОМІЧНИЙ ІНСТИТУТ**

СИСТЕМА УПРАВЛІННЯ ЯКІСТЮ

Сертифікована на відповідність ДСТУ ISO 9001:2015 (ISO 9001:2015, IDT)

Кафедра інформаційних систем та технологій

ЗАТВЕРДЖЕНО
Рішення вченої ради
02.06.2025
протокол № 08, п. 8

ВВЕДЕНО В ДІЮ
Наказ від 02.06.2025 № 104

**ІНЖЕНЕРІЯ ДАНИХ ТА ЗНАНЬ ДЛЯ ІНФОРМАЦІЙНИХ СИСТЕМ /
DATA AND KNOWLEDGE ENGINEERING
FOR INFORMATION SYSTEMS**

РОБОЧА ПРОГРАМА

Ступінь вищої освіти	«магістр» /	«master»
Галузь знань	F «Інформаційні технології» /	«Information Technologies»
Спеціальність	F6 «Інформаційні системи і технології» /	«Information Systems and Technologies»
Освітня програма	«Інформаційні технології у підприємстві» /	«Information Technologies in Entrepreneurship»

Розробник: Романюк Вадим, доктор технічних наук, професор

Гарант освітньої програми «Інформаційні технології у підприємстві» –
Романюк Вадим, доктор технічних наук, професор _____

Обговорено та схвалено:

на засіданні кафедри інформаційних систем та технологій від 10.03.2025,
протокол № 09;

на засіданні вченої ради факультету економіки, менеджменту та права від
06.04.2025, протокол № 07.

Рецензенти:

Гуральник Артем, кандидат технічних наук.

Мерінов Володимир, директор ТОВ «Універсальний сервіс»,
м. Вінниця.

Редактор: Фатєєва Т.

Комп'ютерна верстка: Шуляк Н.

Підп. до друку 12.06.2025 р. Формат 60х84/16. Папір офсетний

Друк ксероксний. Ум. друк. арк. 1,33.

Обл.-вид. арк. 0,95. Тираж 2. Зам. № 249.

Редакційно-видавничий відділ ВТЕІ ДТЕУ
21000, м. Вінниця, вул. Хмельницьке шосе, 25

I. ЗАГАЛЬНІ ПОЛОЖЕННЯ

Мета вивчення освітнього компонента.

Освітній компонент «Інженерія даних та знань для інформаційних систем» належить до обов'язкових компонентів освітньої програми «Інформаційні технології у підприємстві».

Метою вивчення освітнього компонента «Інженерія даних та знань для інформаційних систем» є формування у студентів системи знань з ефективного застосування методів аналізу даних, машинного навчання та оволодіння практичними навичками вирішення прикладних завдань інтелектуального аналізу даних для побудови, супроводу та оптимізації інформаційних систем.

Завданням освітнього компонента оволодіння основними поняттями аналізу, інженерії даних та знань, навчання машин; набуття фахових навичок формалізації задачі прийняття рішень для інформаційних систем та вибору відповідного методу рішення залежно від початкових даних; набуття практичних навичок побудови основних моделей аналізу даних та машинного навчання; ознайомлення з сучасними програмними засобами розробки інтелектуальних систем; набуття практичних навичок по візуалізації даних та презентації отриманих результатів моделювання інтелектуальних систем.

Предметом є концептуальні, методологічні та методичні засади теорії та практики інтелектуального аналізу даних та використання їх на практиці в галузі інформаційних систем та технологій.

Результати вивчення освітнього компонента, його місце в освітньому процесі.

Результатом вивчення освітнього компонента «Інженерія даних та знань для інформаційних систем» є формування комплексу компетентностей:

— інтегральна компетентність:

Здатність розв'язувати задачі дослідницького та інноваційного характеру у сфері інформаційних систем та технологій.

— загальні компетентності (ЗК):

ЗК01. Здатність до абстрактного мислення, аналізу та синтезу.

ЗК03. Здатність спілкуватися з представниками інших професійних груп різного рівня (з експертами з інших галузей знань/видів економічної діяльності).

ЗК04. Здатність розробляти проекти та управляти ними.

ЗК05. Здатність оцінювати та забезпечувати якість виконуваних робіт.

ЗК06. Здатність прогнозувати горизонт прийнятності бізнес-рішень.

— фахові (спеціальні) компетентності (СК):

СК01. Здатність розробляти та застосувати ICT, необхідні для розв'язання стратегічних і поточних задач.

СК02. Здатність формулювати вимоги до етапів життєвого циклу сервіс-орієнтованих інформаційних систем.

СК03. Здатність проектувати інформаційні системи з урахуванням особливостей їх призначення.

СК04. Здатність розробляти математичні, інформаційні та комп'ютерні моделі об'єктів і процесів інформатизації.

СК05. Здатність використовувати сучасні технології аналізу даних для оптимізації процесів в інформаційних системах.

СК08. Балансувати складність та стабільність бізнес-моделей.

Програмні результати навчання здобувачів з освітнього компонента «Інженерія даних та знань для інформаційних систем»:

РН01. Відшуковувати необхідну інформацію в науковій і технічній літературі, базах даних, інших джерелах, аналізувати та оцінювати цю інформацію.

РН04. Управляти процесами розробки, впровадження та експлуатації у сфері ІСТ, які є складними, непередбачуваними і потребують нових стратегічних та командних підходів.

РН08. Розробляти моделі інформаційних процесів та систем різного класу, використовувати методи моделювання, формалізації, алгоритмізації та реалізації моделей з використанням сучасних комп'ютерних засобів.

РН09. Розробляти і використовувати сховища даних, здійснювати аналіз даних для підтримки прийняття рішень.

РН11. Розв'язувати задачі цифрової трансформації у нових або невідомих середовищах на основі спеціалізованих концептуальних знань, що включають сучасні наукові здобутки у сфері інформаційних технологій, досліджень та інтеграції знань з різних галузей.

РН12. Використовувати системи штучного інтелекту для побудови первинних модельних наближень еволюції бізнес-процесів з подальшою модельною гібридизацією і побудовою найбільш стійкої моделі.

Міждисциплінарні зв'язки: робоча програма упорядкована відповідно до освітньо-професійної програми підготовки магістрів, базується на знаннях технічного фахівця в галузі інформаційних технологій та вивченні таких розділів, як «Алгоритми та обчислювальна складність», «Математика в ІТ», а також освітніх компонентів «Управлінські бізнес-рішення на основі soft computing», «Математичні методи і моделі складних економічних систем».

Критерії оцінювання результатів навчання

Критерієм успішного проходження здобувачем вищої освіти підсумкового оцінювання може бути досягнення ним мінімальних порогових рівнів оцінок за кожним запланованим результатом навчання освітнього компонента. Мінімальний пороговий рівень оцінки варто визначати за допомогою якісних критеріїв і трансформувати його в мінімальну позитивну оцінку використовуваної числової (рейтингової) шкали (табл. 1.1). Підсумковим контролем є екзамен.

Таблиця 1.1. Критерії оцінювання результатів навчання

Рівні компетентності	За шкалою ДТЕУ	Критерії оцінювання
1	2	3
Високий (дослідницький)	90-100	Має обґрунтовані та всебічні знання з освітнього компонента, вміє узагальнювати та систематизувати набуті знання; самостійно знаходить джерела інформації та працює з ними; проводить власні дослідження, може використовувати набуті знання та вміння при розв'язанні задач.
Достатній (частково-пошуковий)	82-89	Володіє навчальним матеріалом, вміє зіставляти та узагальнювати, виявляє творчий інтерес до предмету, виконує завдання з повним поясненням та обґрунтуванням, але допускає незначні помилки; може усвідомити нові для нього факти, ідеї.
	75-81	Володіє визначеним програмою навчальним матеріалом; розв'язує завдання, передбачені програмою, з частковим поясненням.
Елементарний (репродуктивний)	69-74	Володіє навчальним матеріалом на репродуктивному рівні; може самостійно розв'язати та пояснити розв'язання завдання.
	60-68	Ознайомлений з навчальним матеріалом, відтворює його на репродуктивному рівні; виконує елементарні завдання за зразком або відомим алгоритмом.
Низький (фрагментарний)	35-59	Ознайомлений та відтворює навчальний матеріал на рівні окремих фактів та фрагментів матеріалу; під керівництвом НПП виконує елементарні завдання.
	1-34	Ознайомлений з навчальним матеріалом на рівні розпізнавання та відтворення окремих фактів.

Для очної (денна, вечірня) форми навчання поточна робота оцінюється в 100 балів, підсумковий контроль (екзамен) оцінюється в 100 балів.

До екзамену допускаються всі здобувачі вищої освіти, які набрали за результатами поточної роботи протягом семестру 60 балів.

Результат підсумкового контролю (екзамен) з освітнього компонента для здобувачів очної форми навчання визначається як середньоарифметична сума балів поточної роботи та екзамену.

Кращим здобувачам, які повністю виконали програму з освітнього компонента, виявили активність в науково-дослідній роботі за відповідною тематикою, стали призерами студентських олімпіад, виступали на

конференціях та за результатами поточної роботи набрали 90 і більше балів, науково-педагогічний працівник має право виставити результат екзамену без опитування (при усному екзамені) чи виконання екзаменаційного завдання (при письмовому екзамені).

Результат підсумкового контролю (екзамен) з освітнього компонента для здобувачів заочної форми навчання оцінюється в 100 балів, відповідно до Положення про організацію освітнього процесу від 17.06.2024 № 08 (зі змінами від 25.11.2024, протокол № 12).

Згідно з цим же Положенням здобувач вищої освіти, який не погоджується з оцінкою, отриманою під час підсумкового контролю, має право в день оголошення результатів звернутися із заявою на ім'я директора з проханням апеляційного перегляду оцінки.

**Обсяг освітнього компонента в кредитах та його розподіл
(тематичний план)**

Назва теми	Кількість годин				Форми контролю	Бальна оцінка
	Усього годин / кредитів	з них				
		лекції	лабораторні заняття	самостійна робота здобувачів		
Тема 1. Введення до інженерії даних та знань. Первинний аналіз даних	20	2	4	14	РПЗ, Т	6
Тема 2. Візуальний аналіз даних в інформаційних системах. Визначення важливості ознак	20	2	4	14	РПЗ, Т	6
Тема 3. Інженерія даних та знань на основі класифікації. Узагальнений метричний класифікатор	20	2	4	14	РПЗ, Т	6
Тема 4. Лінійна класифікація та метод найменших квадратів	20	2	4	14	РПЗ, Т	6
Тема 5. Логістична регресія для оптимізації процесів в інформаційних системах	20	2	4	14	РПЗ, Т	6
Тема 6. Композиції алгоритмів	20	2	4	14	РПЗ, Т	6
Тема 7. Оптимізація випадкового лісу	20	2	4	14	РПЗ, Т	6
Тема 8. Перетворення, вибір та відбір ознак об’єктів та засобів інформаційних систем	20	2	4	14	РПЗ, Т	6
Тема 9. Метод головних компонент	20	2	4	14	РПЗ, Т	6
Тема 10. Кластеризація об’єктів інформаційних систем	20	2	4	14	РПЗ, Т	6
Тема 11. Експоненціальне згладжування в аналізі даних часових рядів. Побудова моделей SARIMA для моделювання об’єктів і процесів інформатизації	20	2	4	14	РПЗ, Т	6
Тема 12. Прогнозування часових рядів з викори-станням методів і засобів штучного інтелекту	20	2	4	14	РПЗ, Т	6
Індивідуальне завдання	30			30	ІЗ	28
Разом	270 / 9	24	48	198		100
Підсумковий контроль	екзамен					

Перелік умовних позначень форм контролю та оцінка їх у балах:

РПЗ – розв’язання практичних завдань – 4 бали.

Т – тестування – 2 бали.

ІЗ – індивідуальні завдання – 28 балів (курси на платформі Prometheus або на інших сервісах – 10 балів; участь у наукових заходах – 18 балів).

Загальна сума за поточну навчальну роботу (аудиторну та самостійну) за семестр – 100 балів.

II. ПРОГРАМА ОСВІТНЬОГО КОМПОНЕНТА

Зміст освітнього компонента (теми програми):

Тема 1. Введення до інженерії даних та знань.

Первинний аналіз даних

Предмет, завдання і структура курсу. Зв'язок курсу з іншими дисциплінами. Важливість інженерії даних та знань для підтримки та розвитку сучасних інформаційних систем.

Первинний аналіз даних з Pandas. Використання Jupyter Notebook. Основні поняття і визначення. Приклади задач. Ознаки, вектори ознак. Об'єкти, класи. Сортування, індексація та вилучення даних.

Розробка і використання сховищ даних для аналізу даних з метою підтримки прийняття рішень. Використання методів і засобів штучного інтелекту для побудови сховищ даних. Зв'язок з датафреймами.

Класифікація. Класифікатор. Навчання, види навчання “з вчителем” і “без вчителя”.

Тема 2. Візуальний аналіз даних в інформаційних системах.

Визначення важливості ознак

Візуальний аналіз даних з Python. Бібліотеки matplotlib, seaborn та plotly. Візуалізація гістограм, кореляційних матриць, групованих даних інформаційних систем. Попарне порівняння ознак об'єктів інформаційних систем.

Статистики розподілу кількісних ознак. Розподіл кількісної ознаки за двома категоріальними ознаками. Задача визначення важливості ознак. Мозаїчні графіки розподілів ознак.

Тема 3. Інженерія даних та знань на основі класифікації.

Узагальнений метричний класифікатор

Класифікація, дерева рішень і метод найближчих сусідів. Метричні методи класифікації. Метод найближчих сусідів і його узагальнення. Підбір числа k за критерієм змінного контролю.

Узагальнений метричний класифікатор, поняття відступу. Метод потенційних функцій, градієнтний алгоритм. Відбір еталонів й оптимізація метрики.

Складні випадки для дерев рішень. Складні випадки для методу найближчих сусідів.

Тема 4. Лінійна класифікація та метод найменших квадратів

Лінійні моделі класифікації і регресії. Метод найменших квадратів. Метод максимальної правдоподібності. Розкладання помилки на зміщення і розкид. Регуляризація лінійної регресії.

Тема 5. Логістична регресія для оптимізації процесів в інформаційних системах

Лінійний класифікатор. Логістична регресія як лінійний класифікатор. Принцип максимальної правдоподібності та логістична регресія в інформаційних системах.

L2-регуляризація логістичної функції втрат. Випадок лінійно розділеної вибірки. Випадок лінійно нерозділеної вибірки.

Аналіз даних щодо потенційного банкрутства суб'єкта підприємницької діяльності. Аналіз даних щодо потенційних інвестицій або запозичень з боку суб'єкта підприємницької діяльності. Використання сховищ даних для побудови логістичної регресії.

Тема 6. Композиції алгоритмів

Композиції алгоритмів: ансамблі, бутстреп (bootstrap), беггінг (bagging), випадковий ліс (Random Forest). Визначення оптимальної кількості слабких класифікаторів або простих алгоритмів у композиції алгоритмів.

Тема 7. Оптимізація випадкового лісу

Структура дерев рішень. Алгоритм побудови дерева рішень. Навчання дерева рішень. Алгоритм Random Forest. Програмна реалізація дерева рішень. Обробка пропусків. Переваги та недоліки дерев рішень.

Тема 8. Перетворення, вибір та відбір ознак об'єктів та засобів інформаційних систем

Постановка задачі витягування ознак. Витягування ознак для текстів, зображень, геоданих, дати та часу, часових рядів, веб-об'єктів.

Перетворення ознак. Нормалізація і зміна розподілу. Взаємодії. Заповнення пропусків у даних інформаційних систем.

Вибір ознак об'єктів та засобів інформаційних систем. Статистичні підходи. Відбір з використанням моделей. Перебір.

Тема 9. Метод головних компонент

Інтуїція, теорія і особливості застосування методу головних компонент. Приклади використання методу головних компонент. Відсіювання малозначущих ознак. Зниження розмірності.

Тема 10. Кластеризація об'єктів інформаційних систем

Кластеризація. Метод k -means. Вибір числа кластерів для k -means. Метод k -means++. Ієрархічна кластеризація в інформаційних системах.

Метрики якості кластеризації. Відкриті проблеми кластеризації об'єктів інформаційних систем та напрямки їх можливого вирішення.

Тема 11. Експоненціальне згладжування в аналізі даних часових рядів.

Побудова моделей SARIMA для моделювання об'єктів і процесів інформатизації

Аналіз часових рядів з Python. Оцінка тренду. Експоненціальне згладжування в аналізі даних часових рядів. Подвійне експоненціальне згладжування.

Модель Хольта — Вінтерса. Враховування сезонних компонентів. Оцінювання тренду. Побудова довірчих інтервалів. Крос-валідація на часових рядах. Оптимальні параметри моделі Хольта — Вінтерса.

Стаціонарність часового ряду. Білий шум. Нестационарні часові ряди. Автокореляція. Частинна автокореляція. Позбавляння від нестационарності і побудова моделей SARIMA для моделювання об'єктів і процесів інформатизації.

Тема 12. Прогнозування часових рядів з використанням методів і засобів штучного інтелекту

Лінійні та нелінійні моделі на часових рядах. Вилучення ознак з використанням методів і засобів штучного інтелекту. Побудова Python-функції для прогнозування часового ряду з використанням алгоритму XGBoost та методів і засобів штучного інтелекту. Визначення аномалій та оптимізація параметрів Python-функції.

Структура освітнього компонента

Результати навчання	Навчальна діяльність	Робочий час здобувача, год.
1	2	3
Розуміти: предмет, завдання і структуру курсу; зв'язок курсу з іншими дисциплінами; як використовується інженерія даних та знань для підтримки та розвитку сучасних інформаційних систем; поняття даних, об'єкта, множини, атрибутів та їх основних властивостей; предмет і завдання аналізу даних, в тому числі для інформаційних систем; поняття і завдання сховищ даних; поняття класифікації, класифікатора. Могти: інсталювати Jupyter Notebook; запускати і налаштовувати параметри Jupyter Notebook; визначити основні статистичні характеристики даних, в тому числі за допомогою бібліотеки Pandas у Jupyter Notebook; виконувати перетворення датафреймів; виконувати вибірки та групування даних; використовувати методи і засоби штучного інтелекту для побудови сховищ даних.	Тема 1. Введення до інженерії даних та знань. Первинний аналіз даних. Лекція № 1. План лекції. 1. Предмет, завдання і структура курсу. Зв'язок курсу з іншими дисциплінами. Важливість інженерії даних та знань для підтримки та розвитку сучасних інформаційних систем. 2. Первинний аналіз даних з Pandas. Використання Jupyter Notebook. Основні поняття і визначення. Приклади задач. 3. Ознаки, вектори ознак. Об'єкти, класи. Сортування, індексація та вилучення даних. 4. Використання методів і засобів штучного інтелекту для побудови сховищ даних. Зв'язок з датафреймами. 5. Класифікація. Класифікатор. Навчання, види навчання “з вчителем” і “без вчителя”. Рекомендовані джерела: Основні: 1, 2, 5, 6. Додаткові: 7, 8, 10. Інтернет-ресурси: 11, 16.	2
	Самостійна робота здобувачів. Вивчення та доповнення матеріалу лекції. Підготовка до виконання завдань лабораторної роботи та оформлення звіту.	14
	Лабораторне заняття № 1. Завдання до заняття: 1. Інсталювати Jupyter Notebook. 2. Завантажити таблицю даних (датафрейм). 3. Визначити основні статистичні характеристики даних.	2
	Лабораторне заняття № 2. Завдання до заняття: 1. Виконати вибірки по об'єктам датафрейму із лабораторного заняття № 1. 2. Виконати вибірки по ознакам (атрибутам). 3. Виконати перетворення датафрейму. 4. Виконати групування даних датафрейму. 5. Скоріювати датафрейм і видалити вказані об'єкти та ознаки у скопійованому датафреймі.	2

1	2	3
Розуміти: призначення візуального аналізу даних; поняття та призначення гістограм, кореляційних матриць, групованих даних; статистики розподілу кількісних ознак; задачу визначення важливості ознак. Могти: використовувати бібліотеки matplotlib, seaborn та plotly для візуального аналізу даних; визначати статистики розподілу кількісних ознак; будувати та аналізувати кореляційні матриці; будувати та аналізувати мозаїчні графіки розподілів ознак.	Тема 2. Візуальний аналіз даних в інформаційних системах. Визначення важливості ознак. Лекція № 2. План лекції. 1. Візуальний аналіз даних з Python. Бібліотеки matplotlib, seaborn та plotly. 2. Візуалізація гістограм, кореляційних матриць, групованих даних інформаційних систем. Попарне порівняння ознак об'єктів інформаційних систем. 3. Статистики розподілу кількісних ознак. 4. Розподіл кількісної ознаки за двома категоріальними ознаками. 5. Задача визначення важливості ознак. Мозаїчні графіки розподілів ознак. Рекомендовані джерела: Основні: 1, 3, 4, 6. Додаткові: 8, 9, 10. Інтернет-ресурси: 11, 12, 16.	2
	Самостійна робота здобувачів. Вивчення та доповнення матеріалу лекції. Підготовка до виконання завдань лабораторної роботи та оформлення звіту.	14
	Лабораторне заняття № 3. Завдання до заняття: 1. Завантажити таблицю даних (датафрейм). 2. Побудувати графіки та гістограми розподілів кількісних ознак об'єктів. 3. Побудувати графіки попарного порівняння ознак об'єктів.	2
	Лабораторне заняття № 4. Завдання до заняття: 1. Побудувати і проаналізувати розподіл кількісної ознаки за двома категоріальними ознаками за даними лабораторного заняття № 3. 2. Побудувати і проаналізувати кореляційні матриці за даними лабораторного заняття № 3. 3. Побудувати і проаналізувати мозаїчні графіки розподілів ознак.	2

1	2	3
Розуміти: класифікацію на основі дерев рішень; класифікацію на основі методу найближчих сусідів; поняття узагальненого метричного класифікатора. Могти: виконувати класифікацію з використанням методу найближчих сусідів; оцінювати якість та ефективність класифікатора; виконувати класифікацію з використанням дерев рішень; оптимізувати дерево рішень з метою підвищення якості та ефективності класифікації.	Тема 3. Інженерія даних та знань на основі класифікації. Узагальнений метричний класифікатор. Лекція № 3. План лекції. 1. Класифікація, дерева рішень і метод найближчих сусідів. Метричні методи класифікації. 2. Метод найближчих сусідів і його узагальнення. Підбір числа k за критерієм змінного контролю. 3. Узагальнений метричний класифікатор, поняття відступу. 4. Метод потенційних функцій, градієнтний алгоритм. Відбір еталонів й оптимізація метрики. 5. Складні випадки для дерев рішень. Складні випадки для методу найближчих сусідів. Рекомендовані джерела: Основні: 2, 3, 5, 6. Додаткові: 7, 8, 10. Інтернет-ресурси: 11, 13, 16.	2
	Самостійна робота здобувачів. Вивчення та доповнення матеріалу лекції. Підготовка до виконання завдань лабораторної роботи та оформлення звіту.	14
	Лабораторне заняття № 5. Завдання до заняття: 1. Для заданого датафрейму виконати побудову класифікатора з використанням методу найближчих сусідів. 2. Оцінити якість та ефективність класифікатора.	2
	Лабораторне заняття № 6. Завдання до заняття: 1. Для заданого датафрейму виконати побудову класифікатора на основі дерева рішень. 2. Оцінити якість та ефективність дерева рішень. 3. Використати метод найближчих сусідів та дерево рішень для задачі прогнозування.	2

1	2	3
Розуміти: поняття лінійної моделі класифікації і регресії; поняття лінійної моделі регресії; принцип метода найменших квадратів; принцип метода максимальної правдоподібності. Могти: використовувати лінійну класифікацію для дослідження якості вказаного об'єкта досліджень; визначати найбільш впливові ознаки об'єкта; ранжувати ознаки за їх впливовістю.	Тема 4. Лінійна класифікація та метод найменших квадратів. Лекція № 4. План лекції. 1. Лінійні моделі класифікації і регресії. 2. Метод найменших квадратів. 3. Метод максимальної правдоподібності. 4. Розкладання помилки на зміщення і розкид. 5. Регуляризація лінійної регресії. Рекомендовані джерела: Основні: 1, 2, 4, 5. Додаткові: 8, 9, 10. Інтернет-ресурси: 11, 13, 16.	2
	Самостійна робота здобувачів. Вивчення та доповнення матеріалу лекції. Підготовка до виконання завдань лабораторної роботи та оформлення звіту.	14
	Лабораторне заняття № 7. Завдання до заняття: 1. Дано результати спостережень щодо якості вказаного об'єкта досліджень. 2. Виконати побудову залежності якості об'єкта від його ознак за допомогою лінійної регресії. 3. Виконати регуляризацію лінійної регресії за наданим шаблоном і пояснити її значення.	2
	Лабораторне заняття № 8. Завдання до заняття: 1. Вилучити 5 % даних результатів спостережень щодо якості вказаного об'єкта досліджень із лабораторного заняття № 7. 2. Виконати побудову нової залежності якості об'єкта від його ознак за допомогою лінійної регресії. 3. Вилучити 10 % даних результатів спостережень щодо якості вказаного об'єкта досліджень із лабораторного заняття № 7. 4. Виконати побудову нової залежності якості об'єкта від його ознак за допомогою лінійної регресії. 5. Вилучити 20 % даних результатів спостережень щодо якості вказаного об'єкта досліджень із лабораторного заняття № 7. 6. Виконати побудову нової залежності якості об'єкта від його ознак за допомогою лінійної регресії.	2

1	2	3
Розуміти: принцип роботи лінійного класифікатора; метод логістичної регресії. Могти: використовувати логістичну регресію для дослідження якості вказаного об'єкта досліджень; визначати найбільш впливові ознаки об'єкта; ранжувати ознаки за їх впливовістю.	Тема 5. Логістична регресія для оптимізації процесів в інформаційних системах. Лекція № 5. План лекції. 1. Лінійний класифікатор. Логістична регресія як лінійний класифікатор. Принцип максимальної правдоподібності та логістична регресія в інформаційних системах. 2. L2-регуляризація логістичної функції втрат. Випадок лінійно розділеної вибірки. Випадок лінійно нерозділеної вибірки. 3. Аналіз даних щодо потенційного банкрутства суб'єкта підприємницької діяльності. 4. Аналіз даних щодо потенційних інвестицій або запозичень з боку суб'єкта підприємницької діяльності. 5. Використання сховищ даних для побудови логістичної регресії. Рекомендовані джерела: Основні: 1, 2, 4, 5. Додаткові: 8, 9, 10. Інтернет-ресурси: 11, 13, 16.	2
	Самостійна робота здобувачів. Вивчення та доповнення матеріалу лекції. Підготовка до виконання завдань лабораторної роботи та оформлення звіту.	14
	Лабораторне заняття № 9. Завдання до заняття: 1. Визначити найбільш впливові ознаки об'єкта із лабораторного заняття № 8. 2. Проранжувати ознаки за їх впливовістю (значущістю) на основі даних лабораторного заняття № 8.	2
	Лабораторне заняття № 10. Завдання до заняття: 1. Побудувати логістичну регресію бінарного показника потенційного банкрутства суб'єкта підприємницької діяльності. 2. Побудувати логістичну регресію бінарного показника потенційних інвестицій або запозичень з боку суб'єкта підприємницької діяльності.	2

1	2	3
Розуміти: поняття ансамблю, бутстрепу (bootstrap), беггінгу (bagging), випадкового лісу (Random Forest). Могти: будувати класифікатор на основі ансамблю класифікаторів; будувати класифікатор на основі бутстрепу.	Тема 6. Композиції алгоритмів. Лекція № 9. План лекції. 1. Композиції алгоритмів. Ансамблі. 2. Бутстреп (bootstrap). 3. Беггінг (bagging). 4. Випадковий ліс (Random Forest). 5. Визначення оптимальної кількості слабких класифікаторів або простих алгоритмів у композиції алгоритмів. Рекомендовані джерела: Основні: 1, 2, 3, 5, 6. Додаткові: 8, 9. Інтернет-ресурси: 11, 16.	2
	Самостійна робота здобувачів. Вивчення та доповнення матеріалу лекції. Підготовка до виконання завдань лабораторної роботи та оформлення звіту.	14
	Лабораторне заняття № 11. Завдання до заняття: 1. Дано результати спостережень щодо кредитного скорингу. 2. Побудувати класифікатор щодо надійного кредитування на основі ансамблю класифікаторів. 3. Виконати підлаштування значимих параметрів побудованого класифікатора. 4. Вказати способи оптимізувати кількість слабких класифікаторів у даному ансамблі.	2
	Лабораторне заняття № 12. Завдання до заняття: 1. Побудувати класифікатор щодо надійного кредитування із лабораторного заняття № 11 на основі бутстрепу. 2. Виконати підлаштування значимих параметрів побудованого класифікатора. 3. Побудувати класифікатор щодо надійного кредитування із лабораторного заняття № 11 на основі бутстрепу з іншим параметром інтервальної оцінки. 4. Виконати підлаштування значимих параметрів побудованого класифікатора.	2

1	2	3
Розуміти: структуру дерев рішень; алгоритм побудови дерева рішень; алгоритм побудови випадкового лісу; переваги та недоліки дерев рішень. Могти: будувати класифікатор на основі беггінгу; будувати класифікатор з використанням алгоритму випадкового лісу; виконувати підлаштування параметрів класифікатора; визначати оптимальний класифікатор.	Тема 7. Оптимізація випадкового лісу. Лекція № 7. План лекції. 1. Структура дерев рішень. Алгоритм побудови дерева рішень. 2. Навчання дерева рішень. 3. Алгоритм Random Forest. 4. Програмна реалізація дерева рішень. Обробка пропусків. 5. Переваги та недоліки дерев рішень. Рекомендовані джерела: Основні: 1, 2, 3, 5, 6. Додаткові: 8, 9. Інтернет-ресурси: 11, 16.	2
	Самостійна робота здобувачів. Вивчення та доповнення матеріалу лекції. Підготовка до виконання завдань лабораторної роботи та оформлення звіту.	14
	Лабораторне заняття № 13. Завдання до заняття: 1. Побудувати класифікатор щодо надійного кредитування із лабораторного заняття № 12 на основі беггінгу. 2. Виконати підлаштування значимих параметрів побудованого класифікатора.	2
	Лабораторне заняття № 14. Завдання до заняття: 1. Побудувати класифікатор щодо надійного кредитування із лабораторного заняття № 12 з використанням алгоритму випадкового лісу. 2. Виконати підлаштування значимих параметрів побудованого класифікатора. 3. Виконати оптимізацію спроектованого алгоритму випадкового лісу. 4. Визначити оптимальний класифікатор серед побудованих на лабораторних заняттях № 11 — 14.	2

1	2	3
Розуміти: поняття і призначення витягування ознак (feature extraction, feature engineering); поняття і призначення перетворення ознак (feature transformation); поняття і призначення відбору ознак (feature selection). Могти: виконувати витягування ознак для текстів; виконувати витягування ознак для зображень; виконувати витягування ознак для геоданих; виконувати витягування ознак для дати та часу; виконувати витягування ознак для часових рядів; виконувати витягування ознак для веб-об'єктів; виконувати перетворення ознак; використовувати відбір ознак для підвищення якості класифікатора.	Тема 8. Перетворення, вибір та відбір ознак об'єктів та засобів інформаційних систем. Лекція № 8. План лекції. 1. Постановка задачі витягування ознак. 2. Витягування ознак для текстів, зображень, геоданих, дати та часу, часових рядів, веб-об'єктів інформаційних систем. 3. Перетворення ознак. 4. Нормалізація і зміна розподілу. Взаємодії. Заповнення пропусків у даних інформаційних систем. 5. Вибір ознак об'єктів та засобів інформаційних систем. Статистичні підходи. Відбір з використанням моделей. Перебір. Рекомендовані джерела: Основні: 1, 3, 5, 6. Додаткові: 7, 9. Інтернет-ресурси: 11, 14, 16.	2
	Самостійна робота здобувачів. Вивчення та доповнення матеріалу лекції. Підготовка до виконання завдань лабораторної роботи та оформлення звіту.	14
	Лабораторне заняття № 15. Завдання до заняття: 1. Виконати витягування ознак для даного фрагмента тексту. 2. Виконати витягування ознак для даного зображення.	2
	Лабораторне заняття № 16. Завдання до заняття: 1. Виконати витягування ознак для вказаних геоданих. 2. Виконати витягування ознак для вказаних дати та часу. 3. Виконати витягування ознак для даного часового ряду. 4. Виконати витягування ознак для даного веб-об'єкта. 5. Підвищити якість класифікатора за допомогою відбору ознак.	2

1	2	3
Розуміти: метод головних компонент та його призначення. Могти: використовувати метод головних компонент для зменшення розмірності даних.	Тема 9. Метод головних компонент. Лекція № 9. План лекції. 1. Інтуїція, теорія і особливості застосування методу головних компонент на практиці. Приклади використання методу головних компонент. 2. Відсіювання малозначущих ознак. 3. Зниження розмірності. Рекомендовані джерела: Основні: 1, 3, 4, 6. Додаткові: 7, 10. Інтернет-ресурси: 11, 14, 16.	2
	Самостійна робота здобувачів. Вивчення та доповнення матеріалу лекції. Підготовка до виконання завдань лабораторної роботи та оформлення звіту.	14
	Лабораторне заняття № 17. Завдання до заняття: 1. Знизити розмірність набору даних № 1.	2
	Лабораторне заняття № 18. Завдання до заняття: 1. Знизити розмірність набору даних № 2.	2
Розуміти: призначення кластеризації; метод <i>k-means</i> та його призначення; метод <i>k-means++</i> та його призначення; метрики якості кластеризації. Могти: метод <i>k-means</i> для кластеризації; метод <i>k-means++</i> для кластеризації; виконувати ієрархічну кластеризацію вказаного набору даних для вказаного числа кластерів.	Тема 10. Кластеризація об'єктів інформаційних систем. Лекція № 10. План лекції. 1. Кластеризація. Метод <i>k-means</i> . 2. Вибір числа кластерів для <i>k-means</i> . Ієрархічна кластеризація в інформаційних системах. 3. Метрики якості кластеризації. 4. Відкриті проблеми кластеризації об'єктів інформаційних систем та напрямки їх можливого вирішення. Рекомендовані джерела: Основні: 1, 3, 4, 6. Додаткові: 7, 10. Інтернет-ресурси: 11, 14, 16.	2
	Самостійна робота здобувачів. Вивчення та доповнення матеріалу лекції. Підготовка до виконання завдань лабораторної роботи та оформлення звіту.	14
	Лабораторне заняття № 19. Завдання до заняття: 1. Виконати кластеризацію вказаного набору даних для вказаного числа кластерів за допомогою методів <i>k-means</i> та <i>k-means++</i> .	2
	Лабораторне заняття № 20. Завдання до заняття: 1. Виконати ієрархічну кластеризацію вказаного набору даних для вказаного числа кластерів.	2

1	2	3
Розуміти: поняття та призначення аналізу часового ряду; поняття та призначення експоненціального згладжування; поняття оцінки тренду; модель Хольта — Вінтерса; як вираховуються сезонні компоненти; якість прогнозування; поняття стаціонарності часового ряду; поняття нестаціонарності часового ряду; поняття автокореляції; поняття частинної автокореляції; методи позбавлення від нестаціонарності і побудови моделей SARIMA. Могти: виконувати експоненціальне згладжування часового ряду; виконувати експоненціальне згладжування часового ряду за моделлю Хольта — Вінтерса; виконувати прогнозування згладженого часового ряду; визначати стаціонарність часового ряду; обчислювати автокореляційну функцію; позбавлятися від нестаціонарності часового ряду і будувати SARIMA-модель.	Тема 11. Експоненціальне згладжування в аналізі даних часових рядів. Побудова моделей SARIMA для моделювання об'єктів і процесів інформатизації. Лекція № 11. План лекції. 1. Аналіз часових рядів з Python. Оцінка тренду. 2. Експоненціальне згладжування в аналізі даних часових рядів. Подвійне експоненціальне згладжування. Модель Хольта — Вінтерса. 3. Вираховування сезонних компонентів. Крос-валідація на часових рядах. Оптимальні параметри моделі Хольта — Вінтерса. 4. Стаціонарність часового ряду. Нестаціонарні часові ряди. Автокореляція. Частинна автокореляція. 5. Позбавлення від нестаціонарності і побудова моделей SARIMA для моделювання об'єктів і процесів інформатизації. Рекомендовані джерела: Основні: 1, 2, 3, 5, 6. Додаткові: 8, 9, 10. Інтернет-ресурси: 11, 15, 16.	2
	Самостійна робота здобувачів. Вивчення та доповнення матеріалу лекції. Підготовка до виконання завдань лабораторної роботи та оформлення звіту.	14
	Лабораторне заняття № 21. Завдання до заняття: 1. Дано впорядковані дані щодо невинуваткових моментів часу. Зобразити даний часовий ряд. 2. Виконати експоненціальне згладжування цього часового ряду. 3. Виконати подвійне експоненціальне згладжування цього часового ряду.	2
	Лабораторне заняття № 22. Завдання до заняття: 1. Виконати потрібне експоненціальне згладжування часового ряду із лабораторного заняття № 21 за моделлю Хольта — Вінтерса. 2. Побудувати SARIMA-модель для цього часового ряду.	2

1	2	3
Розуміти: лінійні та нелінійні моделі на часових рядах; принцип прогнозування часового ряду з використанням алгоритму XGBoost; як визначати аномалії та оптимізувати параметри Python-функції для прогнозування часового ряду. Могти: будувати Python-функції для прогнозування часового ряду з використанням алгоритму XGBoost та методів і засобів штучного інтелекту; визначати аномалії та оптимізувати параметри Python-функції для прогнозування часового ряду.	Тема 12. Прогнозування часових рядів з використанням методів і засобів штучного інтелекту. Лекція № 12. План лекції. 1. Лінійні та нелінійні моделі на часових рядах. Вилучення ознак з використанням методів і засобів штучного інтелекту. 2. Побудова Python-функції для прогнозування часового ряду з використанням алгоритму XGBoost та методів і засобів штучного інтелекту. 3. Визначення аномалій та оптимізація параметрів Python-функції. Рекомендовані джерела: Основні: 1, 2, 3, 5, 6. Додаткові: 8, 9, 10. Інтернет-ресурси: 11, 15, 16.	2
	Самостійна робота здобувачів. Вивчення та доповнення матеріалу лекції. Підготовка до виконання завдань лабораторної роботи та оформлення звіту.	14
	Лабораторне заняття № 23. Завдання до заняття: 1. Виконати прогнозування згладженого часового ряду із лабораторного заняття № 22.	2
	Лабораторне заняття № 24. Завдання до заняття: 1. Для заданого нестационарного часового ряду позбавитися від його нестационарності і побудувати SARIMA-модель з використанням методів і засобів штучного інтелекту.	2
	Індивідуальне завдання	30
ВСЬОГО:		270 / 9

III. РЕКОМЕНДОВАНІ ДЖЕРЕЛА

Основні:

1. De S., Dey S., Bhatia S., Bhattacharyya S. Chapter 1 — An introduction to data mining in social networks, in: Hybrid Computational Intelligence for Pattern Analysis, Advanced Data Mining Tools and Methods for Social Computing / Editors: S. De, S. Dey, S. Bhattacharyya, S. Bhatia. — Academic Press, 2022, pp. 1 — 25.
2. Akyildirim E. Big data analytics, order imbalance and the predictability of stock returns / E. Akyildirim, A. Sensoy, G. Gulay, S. Corbet, H. N. Salari // Journal of Multinational Financial Management. — 2021. — Vol. 62. — Article ID 100717.
3. Pal A. Deep-learning-based visual data analytics for smart construction management / A. Pal, S.-H. Hsieh // Automation in Construction. — 2021. — Vol. 131. — Article ID 103892.
4. Kamm S. A survey on machine learning based analysis of heterogeneous data in industrial automation / S. Kamm, S. S. Veekati, T. Müller, N. Jazdi, M. Weyrich // Computers in Industry. — 2023. — Vol. 149. — Article ID 103930.
5. Demirbaga, Ü., Aujla, G. S., Jindal, A., Kalyon, O. Big Data Analytics. Theory, Techniques, Platforms, and Applications. Springer, Cham, 2024.
6. Paluszek, M., Thomas, S. Data for Machine Learning in MATLAB. In: MATLAB Machine Learning Recipes, pp. 21–48. Apress, Berkeley, CA, 2024.

Додаткові:

7. Romanuke V. V. Deep clustering of the traveling salesman problem to parallelize its solution. Computers & Operations Research, vol. 165, Art. no. 106548, 2024.
8. Mussabayev R., Mussabayev R. Superior Parallel Big Data Clustering Through Competitive Stochastic Sample Size Optimization in Big-Means. In: Nguyen, N. T., Chbeir, R., Manolopoulos, Y., Fujita, H., Hong, T.-P., Nguyen, L. M., Wojtkiewicz, K. (eds.). Intelligent Information and Database Systems. ACIIDS 2024. Lecture Notes in Computer Science, vol. 14796, pp. 224–236. Springer, Singapore, 2024.
9. Badshah A., Daud A., Alharbey R., Banjar A., Bukhari A., Alshemaimri B. Big data applications: overview, challenges and future. Artificial Intelligence Review, vol. 57, Art. no. 290, 2024.
10. Walha A., Ghazzi F., Gargouri F. Data integration from traditional to big data: main features and comparisons of ETL approaches. The Journal of Supercomputing, vol. 80, pp. 26687–26725, 2024.

Internet-ресурси:

11. Аналіз даних [Електронний ресурс]. — Режим доступу: https://uk.wikipedia.org/wiki/Аналіз_даних. — Назва з екрану.
12. Кореляційний аналіз [Електронний ресурс]. — Режим доступу: https://uk.wikipedia.org/wiki/Кореляційний_аналіз. — Назва з екрану.
13. Регресійний аналіз [Електронний ресурс]. — Режим доступу: https://uk.wikipedia.org/wiki/Регресійний_аналіз. — Назва з екрану.

14. Кластерний аналіз [Електронний ресурс]. — Режим доступу: https://uk.wikipedia.org/wiki/Кластерний_аналіз. — Назва з екрану.
15. Аналіз часових рядів [Електронний ресурс]. — Режим доступу: https://uk.wikipedia.org/wiki/Аналіз_часових_рядів. — Назва з екрану.
16. Dimensionality reduction [Електронний ресурс]. — Режим доступу: https://en.wikipedia.org/wiki/Dimensionality_reduction. — Назва з екрану.