

**ДЕРЖАВНИЙ ТОРГОВЕЛЬНО-ЕКОНОМІЧНИЙ УНІВЕРСИТЕТ
ВІННИЦЬКИЙ ТОРГОВЕЛЬНО-ЕКОНОМІЧНИЙ ІНСТИТУТ**

СИСТЕМА УПРАВЛІННЯ ЯКІСТЮ

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

Кафедра економічної кібернетики та інформаційних систем

ЗАТВЕРДЖЕНО

Постанова вченої ради

29.05.2023

протокол № 07, п. 8

ВВЕДЕНО В ДІЮ

Наказ від 29.05.2023 № 70

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

РОБОЧА ПРОГРАМА

Ступінь вищої освіти	«магістр» /	«master»
Галузь знань	12 «Інформаційні технології» /	«Information technologies»
Спеціальність	126 «Інформаційні системи та технології» /	«Information systems and technologies»
Освітня програма	«Інформаційні технології у бізнесі» /	«Information technologies in business»

Розробник: Романюк Вадим, доктор технічних наук, професор

Гарант освітньої програми «Інформаційні технології у бізнесі» — Романюк Вадим, доктор технічних наук, професор

Обговорено та схвалено на засіданні кафедри економічної кібернетики та інформаційних систем від 04.05.2023, протокол № 08; на засіданні вченої ради факультету економіки, менеджменту та права від 16.05.2023, протокол № 05.

Рецензенти:

Кузьміна Олена, кандидат технічних наук, доцент

Вапняр Олена, директор ТОВ «Універсальний сервіс», м. Вінниця

Редактор: Фатєєва Т.

Комп'ютерна верстка: Шуляк Н.

Підп. до друку 01.06.2023 р. Формат 60x84/16. Папір офсетний

Друк ксероксний. Ум. друк. арк. 1,33.

Обл.-вид. арк. 0,95. Тираж 2. Зам. № 216.

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

I. ЗАГАЛЬНІ ПОЛОЖЕННЯ

Мета вивчення дисципліни.

Дисципліна «Інженерія даних та знань для інформаційних систем» належить до циклу професійної підготовки освітньої програми «Інформаційні технології у бізнесі». Дисципліна «Інженерія даних та знань для інформаційних систем» належить до обов'язкових компонентів освітньої програми «Інформаційні технології у бізнесі».

Метою вивчення дисципліни «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 дисципліни 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.

Результат вивчення навчальної дисципліни та її місце в освітньому процесі.

Вивчення дисципліни включає лекційні, лабораторні заняття та самостійну роботу, що сприяє закріпленню необхідних теоретичних знань та допомагає набуттю практичних навичок. Дисципліна «Інженерія даних та знань для інформаційних систем» викладається на першому курсі у другому семестрі загальним обсягом 270 годин / 9 кредитів.

Результатом вивчення дисципліни «Інженерія даних та знань для інформаційних систем» буде набуття студентами 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.

— **фахові компетентності (СК):**

СК01. Здатність розробляти та застосувати ІСТ, 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.

Програмні результати навчання здобувачів з навчальної дисципліни «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.

Міждисциплінарні зв'язки: 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 «М'які обчислення», «Математичні методи і моделі складних економічних систем».

Знання, отримані здобувачами вищої освіти під час вивчення дисципліни «Інженерія даних та знань для інформаційних систем», є базою для опанування дисциплін циклу професійної підготовки, а також можуть бути застосовані під

час проходження виробничої (переддипломної) практики, підготовки випускної кваліфікаційної роботи.

У результаті вивчення цієї дисципліни здобувач вищої освіти зможе застосовувати набуті навички у роботі за фахом, а саме:

1) брати участь у інноваційних та дослідницьких проєктах, що стосуються аналізу даних;

2) проєктувати та створювати програмні системи зберігання та обробки великих масивів даних, розробляти високонавантажені системи обробки даних, використовувати методи штучного інтелекту та машинного навчання для оптимізації інформаційних систем;

3) застосовувати методи та алгоритми обчислювального інтелекту та інтелектуального аналізу даних в задачах класифікації, прогнозування, кластерного аналізу, пошуку асоціативних правил з використанням програмних інструментів підтримки багатовимірного аналізу даних;

4) організовувати командну роботу, управляти проєктами, підбирати команду проєкту, ефективно працювати в групі, визначати та розподіляти завдання з метою вирішення різноманітних дослідницьких та практичних бізнес-завдань;

5) знати і застосовувати на практиці архітектури та стандарти розподілених обчислень, концепції та технології паралельної обробки інформації при розробці та експлуатації паралельного та розподіленого програмного забезпечення інформаційних систем.

Критерії оцінювання результатів навчання

Критерієм успішного проходження здобувачем освіти підсумкового оцінювання може бути досягнення ним мінімальних порогових рівнів оцінок за кожним запланованим результатом навчання навчальної дисципліни. Мінімальний пороговий рівень оцінки варто визначати за допомогою якісних критеріїв і трансформувати його в мінімальну позитивну оцінку використовуваної числової (рейтингової) шкали.

Рівні компетентності	За шкалою ДТЕУ	Критерії оцінювання
1	2	3
Високий (дослідницький)	90-100	Має обґрунтовані та всебічні знання з дисципліни, 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 викладача виконує елементарні завдання.
	1-34	Ознайомлений з навчальним матеріалом на рівні розпізнавання та відтворення окремих фактів.

Підсумковий контроль — екзамен. До екзамену допускаються всі здобувачі, які набрали за результатами поточної роботи протягом семестру 60 балів.

Результат підсумкового контролю (екзамен) з дисципліни «Управлінські бізнес-рішення на основі soft computing» для здобувачів очної форми навчання визначається як середньоарифметична сума балів поточної роботи та екзамену.

Кращим здобувачам, які повністю виконали програму з освітнього компоненту, 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), про що робить запис в екзаменаційному листі здобувача.

Здобувач вищої освіти, який не погоджується з оцінкою, отриманою під час підсумкового (семестрового) контролю, має право звернутися із проханням переглянути оцінку, одержану на екзамені (згідно Положення про апеляцію результатів підсумкового контролю знань студентів № 32 від 07.02.2022).

Обсяг дисципліни в кредитах та його розподіл (тематичний план)

Назва теми	Кількість годин				Форми контролю
	Усього годин / кредитів	з них			
		лекції	лабор. заняття	самот. робота студ.	
Тема 1. Введення до інженерії даних та знань	15	2	2	11	РПЗ, Т
Тема 2. Первинний аналіз даних. Основні поняття і визначення	15	2	4	9	РПЗ, Т
Тема 3. Візуальний аналіз даних в інформаційних системах	15	2	4	9	РПЗ, Т
Тема 4. Визначення важливості ознак	15	2	4	9	РПЗ, Т
Тема 5. Інженерія даних та знань на основі класифікації	15	2	4	9	РПЗ, Т
Тема 6. Узагальнений метричний класифікатор	15	2	4	9	РПЗ, Т
Тема 7. Лінійна класифікація та метод найменших квадратів	15	2	2	11	РПЗ, Т
Тема 8. Логістична регресія для оптимізації процесів в інформаційних системах	15	2	4	9	РПЗ, Т
Тема 9. Композиції алгоритмів	15	2	4	9	РПЗ, Т
Тема 10. Оптимізація випадкового лісу	15	2	4	9	РПЗ, Т
Тема 11. Побудова і витягування ознак	15	2	2	11	РПЗ, Т
Тема 12. Перетворення, вибір та відбір ознак об'єктів та засобів інформаційних систем	15	2	4	9	РПЗ, Т
Тема 13. Метод головних компонент	15	2	2	11	РПЗ, Т
Тема 14. Кластеризація об'єктів інформаційних систем	15	2	2	11	РПЗ, Т
Тема 15. Експоненціальне згладжування в аналізі даних часових рядів	15	2	2	11	РПЗ, Т
Тема 16. Експоненціальне згладжування за моделлю Хольта — Вінтерса	15	2	2	11	РПЗ, Т
Тема 17. Побудова моделей SARIMA для моделювання об'єктів і процесів інформатизації	15	2	2	11	РПЗ, Т
Тема 18. Прогнозування часових рядів з використанням алгоритму XGBoost	15	2	2	11	РПЗ, Т
Разом	270 / 9	36	54	180	
Підсумковий контроль — екзамен					

Умовні позначення:

РПЗ – розв'язання практичних завдань; Т – тестування.

II. ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Зміст дисципліни (теми програми):

Тема 1. Введення до інженерії даних та знань

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

Тема 2. Первинний аналіз даних. Основні поняття і визначення

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.

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

Тема 3. Візуальний аналіз даних в інформаційних системах

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.

Тема 4. Визначення важливості ознак

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

Тема 5. Інженерія даних та знань на основі класифікації

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

Тема 6. Узагальнений метричний класифікатор

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 7. 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 8. 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.

Тема 9. Композиції алгоритмів

Композиції алгоритмів: ансамблі, бутстреп (bootstrap), беггінг (bagging), випадковий ліс (Random Forest).

Тема 10. 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 11. Побудова і витягування ознак

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

Тема 12. Перетворення, вибір та відбір ознак об'єктів та засобів інформаційних систем

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 13. Метод головних компонент

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

Тема 14. 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 15. Експоненціальне згладжування в аналізі даних часових рядів

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

Тема 16. Експоненціальне згладжування за моделлю Хольта — Вінтерса

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

Тема 17. Побудова моделей SARIMA для моделювання об'єктів і процесів інформатизації

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

Тема 18. Прогнозування часових рядів з використанням алгоритму XGBoost

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

Структура навчальної дисципліни

Результати навчання	Навчальна діяльність	Робочий час студента, год.
1	2	3
<p>Знати: предмет, завдання і структуру курсу; зв'язок курсу з іншими дисциплінами; як використовується інженерія даних та знань для підтримки та розвитку сучасних інформаційних систем.</p> <p>Вміти: інсталювати Jupyter Notebook; запускати і налаштовувати параметри Jupyter Notebook; визначити основні статистичні характеристики даних, в тому числі за допомогою бібліотеки Pandas у Jupyter Notebook.</p>	<p>Тема 1. Введення до інженерії даних та знань</p> <p>Лекція № 1</p> <p>План лекції</p> <ol style="list-style-type: none"> 1. Предмет, завдання і структура курсу. 2. Зв'язок курсу з іншими дисциплінами. 3. Важливість інженерії даних та знань для підтримки та розвитку сучасних інформаційних систем. <p>Рекомендовані джерела: Основна: 1, 2, 5, 6. Додаткова: 7, 8, 10. Інтернет-ресурси: 11, 16.</p>	2
	Самостійна робота студентів. Вивчення та доповнення матеріалу лекції. Підготовка до виконання завдань лабораторної роботи та оформлення звіту.	11
	<p>Лабораторне заняття № 1</p> <p>Завдання до заняття:</p> <ol style="list-style-type: none"> 1. Інсталювати Jupyter Notebook. 2. Завантажити таблицю даних (датафрейм). 3. Визначити основні статистичні характеристики даних. 	2
<p>Знати: поняття даних, об'єкта, множини, атрибутів та їх основних властивостей; предмет і завдання аналізу даних, в тому числі для інформаційних систем; поняття класифікації, класифікатора.</p> <p>Вміти: виконувати перетворення датафреймів; виконувати вибірки та групування даних.</p>	<p>Тема 2. Первинний аналіз даних.</p> <p>Основні поняття і визначення</p> <p>Лекція № 2</p> <p>План лекції</p> <ol style="list-style-type: none"> 1. Первинний аналіз даних з Pandas. 2. Використання Jupyter Notebook. 3. Основні поняття і визначення. Приклади задач. 4. Ознаки, вектори ознак. Об'єкти, класи. 5. Сортування, індексація та вилучення даних. 6. Класифікація. Класифікатор. Навчання, види навчання “з вчителем” і “без вчителя”. <p>Рекомендовані джерела: Основна: 1, 2, 5, 6. Додаткова: 7, 8, 10. Інтернет-ресурси: 11, 16.</p>	2
	Самостійна робота студентів. Вивчення та доповнення матеріалу лекції. Підготовка до виконання завдань лабораторної роботи та оформлення звіту.	9

1	2	3
	<p>Лабораторне заняття № 2 Завдання до заняття:</p> <ol style="list-style-type: none"> 1. Виконати вибірки по об'єктам датафрейму із лабораторного заняття № 1. 2. Виконати вибірки по ознакам (атрибутам). 3. Виконати перетворення датафрейму. 	2
	<p>Лабораторне заняття № 3 Завдання до заняття:</p> <ol style="list-style-type: none"> 1. Виконати групування даних датафрейму із лабораторного заняття № 2. 2. Скоріювати датафрейм і видалити вказані об'єкти та ознаки у скопійованому датафреймі. 	2
<p>Знати: призначення візуального аналізу даних; поняття та призначення гістограм, кореляційних матриць, групованих даних. Вміти: використовувати бібліотеки matplotlib, seaborn та plotly для візуального аналізу даних.</p>	<p align="center">Тема 3. Візуальний аналіз даних в інформаційних системах</p> <p>Лекція № 3 План лекції</p> <ol style="list-style-type: none"> 1. Візуальний аналіз даних з Python. 2. Бібліотеки matplotlib, seaborn та plotly. 3. Візуалізація гістограм, кореляційних матриць, групованих даних інформаційних систем. 4. Попарне порівняння ознак об'єктів інформаційних систем. <p>Рекомендовані джерела: Основна: 1, 3, 4, 6. Додаткова: 8, 9, 10. Інтернет-ресурси: 11, 12, 16.</p>	2
	Самостійна робота студентів. Вивчення та доповнення матеріалу лекції. Підготовка до виконання завдань лабораторної роботи та оформлення звіту.	9
	<p>Лабораторне заняття № 4 Завдання до заняття:</p> <ol style="list-style-type: none"> 1. Завантажити таблицю даних (датафрейм). 2. Побудувати графіки та гістограми розподілів кількісних ознак об'єктів. 	2
	<p>Лабораторне заняття № 5 Завдання до заняття:</p> <ol style="list-style-type: none"> 1. Побудувати графіки попарного порівняння ознак об'єктів із лабораторного заняття № 4. 	2
<p>Знати: статистики розподілу кількісних ознак; задачу визначення важливості ознак. Вміти: визначати статистики розподілу кількісних ознак; будувати та аналізувати кореляційні</p>	<p align="center">Тема 4. Визначення важливості ознак</p> <p>Лекція № 4 План лекції</p> <ol style="list-style-type: none"> 1. Статистики розподілу кількісних ознак. 2. Розподіл кількісної ознаки за двома категоріальними ознаками. 3. Задача визначення важливості ознак. 4. Мозаїчні графіки розподілів ознак. <p>Рекомендовані джерела: Основна: 1, 3, 4, 6.</p>	2

1	2	3
матриці; будувати та аналізувати мозаїчні графіки розподілів ознак.	Додаткова: 8, 9, 10. Інтернет-ресурси: 11, 12, 16.	
	Самостійна робота студентів. Вивчення та доповнення матеріалу лекції. Підготовка до виконання завдань лабораторної роботи та оформлення звіту.	9
	Лабораторне заняття № 6 Завдання до заняття: 1. Побудувати і проаналізувати розподіл кількісної ознаки за двома категоріальними ознаками за даними лабораторного заняття № 5.	2
	Лабораторне заняття № 7 Завдання до заняття: 1. Побудувати і проаналізувати кореляційні матриці за даними лабораторного заняття № 6. 2. Побудувати і проаналізувати мозаїчні графіки розподілів ознак.	2
Знати: класифікацію на основі дерев рішень; класифікацію на основі методу найближчих сусідів. Вміти: виконувати класифікацію з використанням методу найближчих сусідів; оцінювати якість та ефективність класифікатора.	Тема 5. Інженерія даних та знань на основі класифікації Лекція № 5 План лекції 1. Класифікація, дерева рішень і метод найближчих сусідів. 2. Метричні методи класифікації. 3. Метод найближчих сусідів і його узагальнення. 4. Підбір числа k за критерієм змінного контролю. Рекомендовані джерела: Основна: 2, 3, 5, 6. Додаткова: 7, 8, 10. Інтернет-ресурси: 11, 13, 16.	2
	Самостійна робота студентів. Вивчення та доповнення матеріалу лекції. Підготовка до виконання завдань лабораторної роботи та оформлення звіту.	9
	Лабораторне заняття № 8 Завдання до заняття: 1. Для заданого датафрейму виконати побудову класифікатора з використанням методу найближчих сусідів.	2
	Лабораторне заняття № 9 Завдання до заняття: 1. Оцінити якість та ефективність класифікатора із лабораторного заняття № 8.	2

1	2	3
<p>Знати: поняття узагальненого метричного класифікатора.</p> <p>Вміти: виконувати класифікацію з використанням дерев рішень; оптимізувати дерево рішень з метою підвищення якості та ефективності класифікації.</p>	<p>Тема 6. Узагальнений метричний класифікатор</p> <p>Лекція № 6</p> <p>План лекції</p> <ol style="list-style-type: none"> 1. Узагальнений метричний класифікатор, поняття відступу. 2. Метод потенційних функцій, градієнтний алгоритм. 3. Відбір еталонів й оптимізація метрики. 4. Складні випадки для дерев рішень. Складні випадки для методу найближчих сусідів. <p>Рекомендовані джерела: Основна: 2, 3, 5, 6. Додаткова: 7, 8, 10. Інтернет-ресурси: 11, 13, 16.</p> <p>Самостійна робота студентів. Вивчення та доповнення матеріалу лекції. Підготовка до виконання завдань лабораторної роботи та оформлення звіту.</p> <p>Лабораторне заняття № 10 Завдання до заняття:</p> <ol style="list-style-type: none"> 1. Для заданого датафрейму виконати побудову класифікатора на основі дерева рішень. 2. Оцінити якість та ефективність дерева рішень. <p>Лабораторне заняття № 11 Завдання до заняття:</p> <ol style="list-style-type: none"> 1. Використати метод найближчих сусідів та дерево рішень для задачі прогнозування. 	<p>2</p> <p>9</p> <p>2</p> <p>2</p>
<p>Знати: поняття лінійної моделі класифікації і регресії; поняття лінійної моделі регресії; принцип метода найменших квадратів; принцип метода максимальної правдоподібності.</p> <p>Вміти: використовувати лінійну класифікацію для дослідження якості вказаного об'єкта досліджень; визначати найбільш впливові ознаки об'єкта; ранжувати ознаки за їх впливовістю.</p>	<p>Тема 7. Лінійна класифікація та метод найменших квадратів</p> <p>Лекція № 7</p> <p>План лекції</p> <ol style="list-style-type: none"> 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. <p>Рекомендовані джерела: Основна: 1, 2, 4, 5. Додаткова: 8, 9, 10. Інтернет-ресурси: 11, 13, 16.</p> <p>Самостійна робота студентів. Вивчення та доповнення матеріалу лекції. Підготовка до виконання завдань лабораторної роботи та оформлення звіту.</p> <p>Лабораторне заняття № 12 Завдання до заняття:</p> <ol style="list-style-type: none"> 1. Дано результати спостережень щодо якості вказаного об'єкта досліджень. 	<p>2</p> <p>11</p> <p>2</p>

1	2	3
	2. Виконати побудову залежності якості об'єкта від його ознак за допомогою лінійної регресії.	
<p>Знати: принцип роботи лінійного класифікатора; метод логістичної регресії.</p> <p>Вміти: використовувати логістичну регресію для дослідження якості вказаного об'єкта досліджень; визначити найбільш впливові ознаки об'єкта; ранжувати ознаки за їх впливовістю.</p>	<p align="center">Тема 8. Логістична регресія для оптимізації процесів в інформаційних системах</p> <p>Лекція № 8 План лекції</p> <ol style="list-style-type: none"> 1. Лінійний класифікатор. Логістична регресія як лінійний класифікатор. 2. Принцип максимальної правдоподібності та логістична регресія в інформаційних системах. 3. L2-регуляризація логістичної функції втрат. 4. Випадок лінійно розділеної вибірки. Випадок лінійно нерозділеної вибірки. <p>Рекомендовані джерела: Основна: 1, 2, 4, 5. Додаткова: 8, 9, 10. Інтернет-ресурси: 11, 13, 16.</p> <p>Самостійна робота студентів. Вивчення та доповнення матеріалу лекції. Підготовка до виконання завдань лабораторної роботи та оформлення звіту.</p> <p>Лабораторне заняття № 13 Завдання до заняття:</p> <ol style="list-style-type: none"> 1. Визначити найбільш впливові ознаки об'єкта із лабораторного заняття № 12. <p>Лабораторне заняття № 14 Завдання до заняття:</p> <ol style="list-style-type: none"> 1. Проранжувати ознаки за їх впливовістю (значущістю) на основі даних лабораторного заняття № 13. 	<p align="center">2</p> <p align="center">9</p> <p align="center">2</p> <p align="center">2</p>
<p>Знати: поняття ансамблю, бутстрепу (bootstrap), беггінгу (bagging), випадкового лісу (Random Forest).</p> <p>Вміти: будувати класифікатор на основі ансамблю класифікаторів; будувати класифікатор на основі бутстрепу.</p>	<p align="center">Тема 9. Композиції алгоритмів</p> <p>Лекція № 9 План лекції</p> <ol style="list-style-type: none"> 1. Композиції алгоритмів. 2. Ансамблі. 3. Бутстреп (bootstrap). 4. Беггінг (bagging). 5. Випадковий ліс (Random Forest). <p>Рекомендовані джерела: Основна: 1, 2, 3, 5, 6. Додаткова: 8, 9. Інтернет-ресурси: 11, 16.</p> <p>Самостійна робота студентів. Вивчення та доповнення матеріалу лекції. Підготовка до виконання завдань лабораторної роботи та оформлення звіту.</p>	<p align="center">2</p> <p align="center">9</p>

1	2	3
	<p>Лабораторне заняття № 15 Завдання до заняття:</p> <ol style="list-style-type: none"> Дано результати спостережень щодо кредитного скорингу. Побудувати класифікатор щодо надійного кредитування на основі ансамблю класифікаторів. Виконати підлаштування параметрів класифікатора. 	2
	<p>Лабораторне заняття № 16 Завдання до заняття:</p> <ol style="list-style-type: none"> Побудувати класифікатор щодо надійного кредитування із лабораторного заняття № 15 на основі бутстрепа. Виконати підлаштування параметрів класифікатора. 	2
<p>Знати: структуру дерев рішень; алгоритм побудови дерева рішень; алгоритм побудови випадкового лісу; переваги та недоліки дерев рішень. Вміти: будувати класифікатор на основі беггінгу; будувати класифікатор з використанням алгоритму випадкового лісу; виконувати підлаштування параметрів класифікатора; визначати оптимальний класифікатор.</p>	<p>Тема 10. Оптимізація випадкового лісу Лекція № 10 План лекції</p> <ol style="list-style-type: none"> Структура дерев рішень. Алгоритм побудови дерева рішень. Навчання дерева рішень. Алгоритм Random Forest. Програмна реалізація дерева рішень. Обробка пропусків. Переваги та недоліки дерев рішень. <p>Рекомендовані джерела: Основна: 1, 2, 3, 5, 6. Додаткова: 8, 9. Інтернет-ресурси: 11, 16.</p>	2
	<p>Самостійна робота студентів. Вивчення та доповнення матеріалу лекції. Підготовка до виконання завдань лабораторної роботи та оформлення звіту.</p>	9
	<p>Лабораторне заняття № 17 Завдання до заняття:</p> <ol style="list-style-type: none"> Побудувати класифікатор щодо надійного кредитування із лабораторного заняття № 15 на основі беггінгу. Виконати підлаштування параметрів класифікатора. 	2
	<p>Лабораторне заняття № 18 Завдання до заняття:</p> <ol style="list-style-type: none"> Побудувати класифікатор щодо надійного кредитування із лабораторного заняття № 15 з використанням алгоритму випадкового лісу. Виконати підлаштування параметрів класифікатора. Визначити оптимальний класифікатор серед побудованих на лабораторних заняттях № 15 — 18. 	2

1	2	3
<p>Знати: поняття і призначення витягування ознак (feature extraction, feature engineering).</p> <p>Вміти: 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.</p>	<p align="center">Тема 11. Побудова і витягування ознак</p> <p>Лекція № 11</p> <p>План лекції</p> <ol style="list-style-type: none"> 1. Постановка задачі витягування ознак. 2. Витягування ознак для текстів, зображень, геоданих, дати та часу, часових рядів, веб-об'єктів інформаційних систем. <p>Рекомендовані джерела: Основна: 1, 3, 5, 6. Додаткова: 7, 9. Інтернет-ресурси: 11, 14, 16.</p> <p>Самостійна робота студентів. Вивчення та доповнення матеріалу лекції. Підготовка до виконання завдань лабораторної роботи та оформлення звіту.</p> <p>Лабораторне заняття № 19</p> <p>Завдання до заняття:</p> <ol style="list-style-type: none"> 1. Виконати витягування ознак для даного фрагмента тексту. 2. Виконати витягування ознак для даного зображення. 	<p align="center">2</p> <p align="center">11</p> <p align="center">2</p>
<p>Знати: поняття і призначення перетворення ознак (feature transformation); поняття і призначення відбору ознак (feature selection).</p> <p>Вміти: виконувати перетворення ознак; використовувати відбір ознак для підвищення якості класифікатора.</p>	<p align="center">Тема 12. Перетворення, вибір та відбір ознак об'єктів та засобів інформаційних систем</p> <p>Лекція № 12</p> <p>План лекції</p> <ol style="list-style-type: none"> 1. Перетворення ознак. 2. Нормалізація і зміна розподілу. Взаємодії. Заповнення пропусків у даних інформаційних систем. 3. Вибір ознак об'єктів та засобів інформаційних систем. Статистичні підходи. Відбір з використанням моделей. Перебір. <p>Рекомендовані джерела: Основна: 1, 3, 5, 6. Додаткова: 7, 9. Інтернет-ресурси: 11, 14, 16.</p> <p>Самостійна робота студентів. Вивчення та доповнення матеріалу лекції. Підготовка до виконання завдань лабораторної роботи та оформлення звіту.</p> <p>Лабораторне заняття № 20</p> <p>Завдання до заняття:</p> <ol style="list-style-type: none"> 1. Виконати витягування ознак для вказаних геоданих. 2. Виконати витягування ознак для вказаних дати та часу. 	<p align="center">2</p> <p align="center">9</p> <p align="center">2</p>

1	2	3
	<p>Лабораторне заняття № 21 Завдання до заняття:</p> <ol style="list-style-type: none"> 1. Виконати витягування ознак для даного часового ряду. 2. Виконати витягування ознак для даного веб-об'єкта. 3. Підвищити якість класифікатора за допомогою відбору ознак. 	2
<p>Знати: метод головних компонент та його призначення. Вміти: використовувати метод головних компонент для зменшення розмірності даних.</p>	<p>Тема 13. Метод головних компонент Лекція № 13 План лекції</p> <ol style="list-style-type: none"> 1. Інтуїція, теорія і особливості застосування методу головних компонент. 2. Приклади використання методу головних компонент. 3. Відсіювання малозначущих ознак. <p>Рекомендовані джерела: Основна: 1, 3, 4, 6. Додаткова: 7, 10. Інтернет-ресурси: 11, 14, 16.</p>	2
	<p>Самостійна робота студентів. Вивчення та доповнення матеріалу лекції. Підготовка до виконання завдань лабораторної роботи та оформлення звіту.</p>	11
	<p>Лабораторне заняття № 22 Завдання до заняття:</p> <ol style="list-style-type: none"> 1. Застосувати метод головних компонент до вказаного набору даних. 	2
<p>Знати: призначення кластеризації; метод k-means та його призначення; метод k-means++ та його призначення; метрики якості кластеризації. Вміти: метод k-means для кластеризації; метод k-means++ для кластеризації; виконувати ієрархічну кластеризацію вказаного набору даних для вказаного числа кластерів.</p>	<p>Тема 14. Кластеризація об'єктів інформаційних систем Лекція № 14 План лекції</p> <ol style="list-style-type: none"> 1. Кластеризація. Метод k-means. 2. Вибір числа кластерів для k-means. Ієрархічна кластеризація в інформаційних системах. 3. Метрики якості кластеризації. 4. Відкриті проблеми кластеризації об'єктів інформаційних систем та напрямки їх можливого вирішення. <p>Рекомендовані джерела: Основна: 1, 3, 4, 6. Додаткова: 7, 10. Інтернет-ресурси: 11, 14, 16.</p>	2
	<p>Самостійна робота студентів. Вивчення та доповнення матеріалу лекції. Підготовка до виконання завдань лабораторної роботи та оформлення звіту.</p>	11

1	2	3
	<p>Лабораторне заняття № 23 Завдання до заняття:</p> <ol style="list-style-type: none"> 1. Виконати кластеризацію вказаного набору даних у лабораторному занятті № 23 для вказаного числа кластерів за допомогою методів k-means та k-means++. 2. Виконати ієрархічну кластеризацію вказаного набору даних для вказаного числа кластерів. 	2
<p>Знати: поняття та призначення аналізу часового ряду; поняття та призначення експоненціального згладжування; поняття оцінки тренду. Вміти: виконувати експоненціальне згладжування часового ряду.</p>	<p>Тема 15. Експоненціальне згладжування в аналізі даних часових рядів Лекція № 15 План лекції</p> <ol style="list-style-type: none"> 1. Аналіз часових рядів з Python. 2. Оцінка тренду. 3. Експоненціальне згладжування в аналізі даних часових рядів. 4. Подвійне експоненціальне згладжування. <p>Рекомендовані джерела: Основна: 1, 2, 3, 5, 6. Додаткова: 8, 9, 10. Інтернет-ресурси: 11, 15, 16.</p> <p>Самостійна робота студентів. Вивчення та доповнення матеріалу лекції. Підготовка до виконання завдань лабораторної роботи та оформлення звіту.</p>	2
<p>Знати: модель Хольта — Вінтерса; як вираховуються сезонні компоненти; якість прогнозування. Вміти: виконувати експоненціальне згладжування часового ряду за моделлю Хольта — Вінтерса; виконувати прогнозування згладженого часового ряду.</p>	<p>Тема 16. Експоненціальне згладжування за моделлю Хольта — Вінтерса Лекція № 16 План лекції</p> <ol style="list-style-type: none"> 1. Модель Хольта — Вінтерса. 2. Вираховування сезонних компонент. 3. Оцінювання тренду. 4. Побудова довірчих інтервалів. 5. Крос-валідація на часових рядах. 6. Оптимальні параметри моделі Хольта — Вінтерса. <p>Рекомендовані джерела: Основна: 1, 2, 3, 5, 6. Додаткова: 8, 9, 10. Інтернет-ресурси: 11, 15, 16.</p>	2
		11
	<p>Лабораторне заняття № 24 Завдання до заняття:</p> <ol style="list-style-type: none"> 1. Дано впорядковані дані щодо невідповідних моментів часу. Зобразити даний часовий ряд. 2. Виконати експоненціальне згладжування цього ряду. 	2

1	2	3
	Самостійна робота студентів. Вивчення та доповнення матеріалу лекції. Підготовка до виконання завдань лабораторної роботи та оформлення звіту.	11
	Лабораторне заняття № 25 Завдання до заняття: 1. Виконати подвійне експоненціальне згладжування часового ряду із лабораторного заняття № 24.	2
Знати: поняття стаціонарності часового ряду; поняття нестаціонарності часового ряду; поняття автокореляції; поняття частинної автокореляції; методи позбавлення від нестаціонарності і побудови моделей SARIMA. Вміти: визначати стаціонарність часового ряду; обчислювати автокореляційну функцію; позбавлятися від нестаціонарності часового ряду і будувати SARIMA-модель.	Тема 17. Побудова моделей SARIMA для моделювання об'єктів і процесів інформатизації Лекція № 17 План лекції 1. Стаціонарність часового ряду. 2. Нестаціонарні часові ряди. 3. Автокореляція. Частинна автокореляція. 4. Позбавлення від нестаціонарності і побудова моделей SARIMA для моделювання об'єктів і процесів інформатизації. Рекомендовані джерела: Основна: 1, 2, 3, 5, 6. Додаткова: 8, 9, 10. Інтернет-ресурси: 11, 15, 16.	2
	Самостійна робота студентів. Вивчення та доповнення матеріалу лекції. Підготовка до виконання завдань лабораторної роботи та оформлення звіту.	11
Знати: 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. Вміти: будувати Python-функції для прогнозування	Тема 18. Прогнозування часових рядів з використанням алгоритму XGBoost Лекція № 18 План лекції 1. Лінійні та нелінійні моделі на часових рядах. Вилучення ознак. 2. Побудова Python-функції для прогнозування часового ряду з використанням алгоритму XGBoost. 3. Визначення аномалій та оптимізація параметрів Python-функції. Рекомендовані джерела: Основна: 1, 2, 3, 5, 6. Додаткова: 8, 9, 10. Інтернет-ресурси: 11, 15, 16.	2
	Лабораторне заняття № 26 Завдання до заняття: 1. Виконати потрібне експоненціальне згладжування часового ряду із лабораторного заняття № 25 за моделлю Хольта — Вінтерса.	2

1	2	3
<p>часового ряду з використанням алгоритму XGBoost; визначати аномалії та оптимізувати параметри Python-функції для прогнозування часового ряду.</p>	<p>Самостійна робота студентів. Вивчення та доповнення матеріалу лекції. Підготовка до виконання завдань лабораторної роботи та оформлення звіту.</p>	<p>11</p>
	<p>Лабораторне заняття № 27 Завдання до заняття:</p> <ol style="list-style-type: none"> 1. Виконати прогнозування згладженого часового ряду із лабораторного заняття № 26. 2. Для заданого нестационарного часового ряду позбавитися від його нестационарності і побудувати SARIMA-модель. 	<p>2</p>
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. Tagliaferri L. Python machine learning projects / L. Tagliaferri, M. Morales, E. Birbeck, A. Wan. — New York City: DigitalOcean, 2019. — 135 p.
6. Deisenroth M. P. Mathematics for machine learning / M. P. Deisenroth, A. A. Faisal, C. S. Ong. — Cambridge: Cambridge University Press, 2019. — 411 p.

Додаткові

7. Romanuke V. V. Two-layer perceptron for classifying flat scaled-turned-shifted objects by additional feature distortions in training / V. V. Romanuke // Journal of Uncertain Systems. — 2015. — Vol. 9, No. 4. — P. 286 — 305. <https://doi.org/10.20535/1810-0546.2017.6.110724>
8. Romanuke V. V. Optimal training parameters and hidden layer neurons number of two-layer perceptron for generalized scaled objects classification problem / V. V. Romanuke // Information Technology and Management Science. — 2015. — Vol. 18. — P. 42 — 48. <https://doi.org/10.1515/itms-2015-0007>
9. Romanuke V. V. Training data expansion and boosting of convolutional neural networks for reducing the MNIST dataset error rate / V. V. Romanuke // Research Bulletin of NTUU “Kyiv Polytechnic Institute”. — 2016. — No. 6. — P. 29 — 34. <https://doi.org/10.20535/1810-0546.2016.6.84115>
10. Romanuke V. V. An infinitely scalable dataset of single-polygon grayscale images as a fast test platform for semantic image segmentation / V. V. Romanuke // KPI Science News. — 2019. — No. 1. — P. 24 — 34. <https://doi.org/10.20535/kpi-sn.2019.1.157259>

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. Romanuke V. V. Interval uncertainty reduction via division-by-2 dichotomization based on expert estimations for short-termed observations / V. V. Romanuke // Journal of Uncertain Systems. — 2018. — Vol. 12, No. 1. — P. 3 — 21. [Електронний ресурс]. — Режим доступу: <http://www.worldacademicunion.com/journal/jus/jusVol12No1paper01.pdf>. — Назва з екрану.