

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ЗАХІДНОУКРАЇНСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ФАКУЛЬТЕТ КОМП'ЮТЕРНИХ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ

ЗАТВЕРДЖУЮ:

В.о. декана факультету
комп'ютерних інформаційних
технологій

Ігор ЯКИМЕНКО

“ ” 2023 р.

ЗАТВЕРДЖУЮ:

В.о. проректора з
науково-педагогічної роботи

Віктор ОСТРОВЕРХОВ

“ ” 2023 р.

ЗАТВЕРДЖУЮ:

Директор навчально-наукового
інституту новітніх освітніх
технологій

Святослав ПИТЕЛЬ

“ ” 2023 р.

РОБОЧА ПРОГРАМА

з дисципліни «Інтелектуальний аналіз даних»

ступінь вищої освіти – бакалавр

галузь знань – 12 “Інформаційні технології”

спеціальність – 122 „Комп’ютерні науки”

освітньо-професійна програма – „Комп’ютерні науки”

Кафедра інформаційно-обчислювальних систем і управління

Форма навчання	Курс	Семестр	Лекції (год.)	Лабораторні заняття (год.)	ІРС (год.)	Тренінг, КПІЗ (год.)	Самост. робота студ. (год.)	Разом (год.)	Екз. (сем.)
Денна	3	5	28	42	4	10	66	150	5
Заочна	3	5, 6	8	4	–	–	138	150	6

Тернопіль – ЗУНУ
2023


31.08.2023

Робоча програма складена на основі освітньо-професійної програми «Комп'ютерні науки» підготовки бакалавра галузі знань 12 «Інформаційні технології» спеціальності 122 «Комп'ютерні науки», затвердженої Вченою радою ЗУНУ (протокол № 9 від 26 травня 2021 р.).

Робочу програму склала доцент кафедри ІОСУ, к.т.н. Христина ЛІП'ЯНІНА-ГОНЧАРЕНКО

Робоча програма затверджена на засіданні кафедри інформаційно-обчислювальних систем і управління, протокол № 1 від 29 серпня 2023 р.

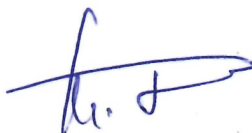
Завідувач кафедри



Мирослав КОМАР

Розглянуто та схвалено групою забезпечення спеціальності „Комп'ютерні науки”, протокол № 1 від 29 серпня 2023 р.

Голова групи
забезпечення спеціальності,
д-р техн. наук, доцент



Мирослав КОМАР

Гарант освітньо-професійної
програми "Комп'ютерні науки",
к.т.н, доцент



Христина
ЛІП'ЯНІНА-ГОНЧАРЕНКО

СТРУКТУРА РОБОЧОЇ ПРОГРАМИ НАВЧАЛЬНОЇ ДИСЦИПЛІНИ " ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ДАНИХ "

1. Опис дисципліни " Інтелектуальний аналіз даних "

Дисципліна «Інтелектуальний аналіз даних»	Галузь знань, спеціальність, СВО	Характеристика навчальної дисципліни
Кількість кредитів – 5	Галузь знань – 12 “Інформаційні технології”	Статус дисципліни: обов’язкова дисципліна циклу професійної підготовки Мова навчання: Українська
Кількість залікових модулів – 4	Спеціальність – 122 «Комп’ютерні науки»	Рік підготовки: 3 Семестр: <i>Денна</i> – 5 <i>Заочна</i> – 5, 6
Кількість змістових Модулів – 2	Освітньо- професійна програма «Комп’ютерні науки»	Лекції: <i>Денна</i> –28 год. <i>Заочна</i> – 8 год. Лабораторні заняття: <i>Денна</i> – 42 год. <i>Заочна</i> – 4 год.
Загальна кількість годин – 150	Ступінь вищої освіти – бакалавр	Самостійна робота: <i>Денна</i> – 66 год., <i>Заочна</i> – 138 год. Тренінг, КПЗ: <i>Денна</i> – 10 год. Індивідуальна робота: <i>Денна</i> – 4 год.
Тижневих годин – 10, з них аудиторних – 5 год.		Вид підсумкового контролю – екзамен

2. Мета і завдання дисципліни " Інтелектуальний аналіз даних "

2.1. Мета вивчення дисципліни

Метою дисципліни „ Інтелектуальний аналіз даних ” є вивчення методів сучасної обробки даних – інтелектуального аналізу даних (Data Mining), а також виявлення в необроблених масивах даних раніше невідомих, практично корисних знань і закономірностей, необхідних для прийняття рішень.

2.2. Завдання вивчення дисципліни

Завдання дисципліни “ Інтелектуальний аналіз даних ” полягає у опануванні базових принципів побудови моделей даних, ознайомлення з концепцією Knowledge Discovery in Data (виявлення знань в даних) і Data Mining («видобування» знань), ефективному використанні методів здобуття знань з великих масивів даних, ознайомленні з основними типами задач, що можуть бути вирішені за допомогою методів інтелектуального аналізу даних та отримані практичні навички з використання інструментальних засобів інтелектуального аналізу даних при вирішенні прикладних задач і навчитися інтерпретувати отримані результати.

2.3. Найменування та опис компетентностей, формування яких забезпечує вивчення дисципліни:

СК2. Здатність до виявлення статистичних закономірностей недетермінованих явищ, застосування методів обчислювального інтелекту, зокрема статистичної, нейромережевої та нечіткої обробки даних, методів машинного навчання та генетичного програмування тощо.

СК11. Здатність до інтелектуального аналізу даних на основі методів обчислювального інтелекту включно з великими та погано структурованими даними, їхньої оперативної обробки та візуалізації результатів аналізу в процесі розв’язування прикладних задач.

2.4. Передумови для вивчення дисципліни

Курс базується на знаннях, одержаних при вивченні дисциплін «Вища математика», «Дискретна математика», «Чисельні методи та програмування», «Теорія ймовірності і математична статистика», «Основи комп’ютерних наук», «Алгоритми і структури даних», «Бази і сховища даних» та «Сучасні парадигми програмування».

2.5. Результати навчання:

ПР1. Застосовувати знання основних форм і законів абстрактно-логічного мислення, основ методології наукового пізнання, форм і методів вилучення, аналізу, обробки та синтезу інформації в предметній області комп’ютерних наук.

ПР3. Використовувати знання закономірностей випадкових явищ, їх властивостей та операцій над ними, моделей випадкових процесів та сучасних програмних середовищ для розв’язування задач статистичної обробки даних і побудови прогнозних моделей.

ПР4. Використовувати методи обчислювального інтелекту, машинного навчання, нейромережевої та нечіткої обробки даних, генетичного та

еволюційного програмування для розв'язання задач розпізнавання, прогнозування, класифікації, ідентифікації об'єктів керування тощо.

ПР12. 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 на основі технологій DataMining, TextMining, WebMining.

3. Програма навчальної дисципліни «Інтелектуальний аналіз даних»

Змістовий модуль 1 – Основи інтелектуального аналізу даних

Тема 1. Концепції інтелектуального аналізу даних та DataMining

Визначення необхідності аналізу даних. Цілі та основні завдання процесу інтелектуального аналізу даних. Технології інтелектуального аналізу даних. Ітераційний характер процесу інтелектуального аналізу даних та його основні кроки. Вплив якості даних на процес інтелектуального аналізу даних. Зв'язок між зберіганням даних та інтелектуальним аналізом даних. Концепції великих даних та науки про дані..

Тема 2. Попередня обробка та аналіз даних

Важливість належного представлення необроблених даних. Методи збору та введення даних до системи. Характеристики необроблених даних: типи, обсяг, структура. Процеси перетворення необроблених даних. Методи нормалізації та згладжування даних. Встановлення відмінностей та зв'язків між даними. Методи оцінки адекватності та повноти даних, зокрема методи поведінки та аугментації даних при обмежених вибірках.

Тема 3. Стратегії обробки даних з відсутністю інформації та залежністю від часу

Проблема відсутності даних і стратегії їх обробки. Методи заповнення пропусків у даних. Аналіз даних з неповним набором інформації. Виявлення та опрацювання викидів у даних.

Тема 4. Аналіз великих наборів даних та методи зменшення розмірності

Значущість великих обсягів даних. Проблеми збереження та використання великих датасетів. Стратегії для скорочення обсягу даних. Методи вибору та вилучення функцій. Підходи до зниження значення для збереження інформативності даних. Застосування міри ентропії для ранжування функцій. Визначення важливості функцій на основі інформаційної цінності. Аналіз головних компонентів. Методи пошуку головних компонентів.

Тема 5. Методи візуалізації

Різноманітні методи та підходи до візуалізації даних. Інструменти для графічного відображення інформації, які сприяють візуальному аналізу залежностей, шаблонів та трендів у даних. Графіки, діаграми, розподіли та інші засоби для ефективного представлення результатів аналізу даних.

Змістовий модуль 2 – Статистичні методи та моделювання

Тема 6. Статистичний аналіз

Основні концепції та методи статистичного аналізу даних. Метод описової статистики та інференційної статистики. Статистичних тестів для зроблення висновків на основі вибіркового даних. Ця тема охоплює методи порівняння та оцінки статистичних різниць між різними наборами даних. Досліджуються підходи для виявлення значущих різниць між групами даних та їхній вплив на аналіз результатів. Методи оцінки результатів.

Тема 7. Методи оцінки результатів

Вступ до методів оцінки результатів. Метрики точності. F-міра та інші гармонічні метрики. Матриця помилок (Confusion Matrix). ROC-крива та AUC (Area Under the Curve). Перехресна перевірка (Cross-Validation).

Тема 8. Лінійний дискримінантний аналіз

Лінійний дискримінантний аналіз як метод класифікації та розділення категорій на основі лінійних комбінацій змінних. Підходи до побудови та використання лінійного дискримінантного аналізу. Баєсівський підхід до класифікації даних. Метрики точності класифікації: F-Measure, Harmonic Mean, матриця помилок, ROC-крива та AUC.

Тема 9. Регресійні моделі

Різні підходи до регресійного аналізу, які використовуються для передбачення залежних змінних на основі незалежних. Проста лінійна регресія, множинна лінійна регресія, нелінійна регресія, регресія з вибілковими точками, регресія з поліномами, регресія зі згладжуванням, регуляризована регресія (Lasso, Ridge, Elastic Net). Методи оцінки результатів у регресійних моделях.

Тема 10. Прогнозування часових рядів

Основні терміни та поняття, пов'язані з аналізом та прогнозуванням часових рядів. Методи визначення стаціонарності. Методи аналізу та моделювання часових рядів. Прогнозування сезонних та трендових компонентів. Авторегресія (AR). Ковзне середнє (MA). Авторегресія з ковзним середнім (ARMA). Авторегресія з інтегрованим ковзним середнім (ARIMA). Сезонна авторегресія з інтегрованим ковзним середнім (SARIMA). Авторегресія з екзогенними факторами (ARX). Екзогенна модель для ідентифікації (X12-ARIMA). Авторегресія з виправленням аномалій (ARIMA-X). Експоненційне згладжування (Exponential Smoothing). Модель Гарча (GARCH). Методи оцінки результатів.

Тема 11. TextMining і WebMining

Специфіка WebMining. Класифікація основних підзадач WebMining. Можливості WebMining за допомогою HITS, LOGSOM та алгоритмів проходження шляхів. Незалежний від запитів ранжування веб-сторінок та основні характеристики алгоритму PageRank. Структура TextMining, що визначає етапи переробки та дистиляції. Методологія семантичної індексації.

**4. Структура залікового кредиту з дисципліни
«Інтелектуальний аналіз даних»**

Денна форма навчання

Тема	<i>Кількість годин</i>					
	Лекції	Лабораторні заняття	Індивідуальна робота	Тренінг, КПЗ	Самостійна робота	Контрольні заходи
<i>Змістовий модуль 1 – Основи інтелектуального аналізу даних</i>						
Тема 1. Концепції інтелектуального аналізу даних та DataMining	2	-	-	5	8	Опитування під час заняття
Тема 2. Попередня обробка та аналіз даних	2	4	-		6	Опитування під час заняття
Тема 3. Стратегії обробки даних з відсутністю інформації та залежністю від часу	2	4	1		6	Опитування під час заняття
Тема 4. Аналіз великих наборів даних та методи зменшення розмірності	2	4	-		6	Опитування під час заняття
Тема 5. Методи візуалізації	2	4	1		6	Опитування під час заняття
<i>Змістовий модуль 2 – Статистичні методи та моделювання</i>						
Тема 6. Статистичний аналіз	2	4	1	5	6	Опитування під час заняття
Тема 7. Методи оцінки результатів	2	-	-		4	Опитування під час заняття
Тема 8. Лінійний дискримінантний аналіз	4	4	-		6	Опитування під час заняття
Тема 9. Регресійні моделі	6	6	1		6	Опитування під час заняття
Тема 10. Прогнозування часових рядів	2	6	-		6	Опитування під час заняття
Тема 11. TextMining і WebMining	2	6	-		6	Опитування під час заняття
Разом	28	42	4	10	66	

Заочна форма навчання

Тема	Кількість годин			
	Лекції	Лабораторні заняття	Тренінг, КПЗ	Самостійна робота
Змістовий модуль 1 – Основи інтелектуального аналізу даних				
Тема 1. Концепції інтелектуального аналізу даних та DataMining	4	2	-	14
Тема 2. Попередня обробка та аналіз даних				10
Тема 3. Стратегії обробки даних з відсутністю інформації та залежністю від часу				10
Тема 4. Аналіз великих наборів даних та методи зменшення розмірності				10
Тема 5. Методи візуалізації				14
Змістовий модуль 2 – Статистичні методи та моделювання				
Тема 6. Статистичний аналіз	4	2	-	10
Тема 7. Методи оцінки результатів				10
Тема 8. Лінійний дискримінантний аналіз				10
Тема 9. Регресійні моделі				20
Тема 10. Прогнозування часових рядів				20
Тема 11. TextMining і WebMining				10
Разом	8	4	-	138

5. Тематика лабораторних занять

Лабораторна робота 1: Введення до обробки та аналізу даних з використанням Python

Мета: Ознайомити студентів з основами використання мови програмування Python для обробки та аналізу даних, навчити роботі з Jupyter Notebook, завантаженню та читанню даних з різних форматів, а також надати загальний огляд бібліотеки Pandas для ефективних маніпуляцій з даними.

Питання для обговорення:

- Вступ до Python для аналізу даних.
- Робота з Jupyter Notebook та основними командами.
- Завантаження та читання даних з різних джерел (CSV, Excel, тощо).
- Огляд бібліотеки Pandas для маніпуляцій з даними.

Лабораторна робота 2: Дослідницький аналіз даних у Python.

Мета: Метою даної лабораторної роботи є вивчення основних кроків дослідницького аналізу даних з використанням мови програмування Python та відповідних бібліотек. У ході виконання лабораторної роботи студенти будуть ознайомлені з ключовими етапами обробки та аналізу даних, такими як збереження та завантаження даних, видалення зайвої інформації, очищення від пропущених значень, виявлення викидів та візуалізація даних.

Питання для обговорення:

- Збереження і та визначення контрольної точки
- Вивчення набору даних
- Видалення неактуальних стовпців
- Перейменування стовпців
- Видалення повторюваних рядків
- Видалення пропущених або нульових значень.
- Виявлення викидів

Лабораторна робота 3: Візуалізація даних

Мета: Вивчити методи візуалізації даних, ознайомити студентів з бібліотеками Matplotlib та Seaborn для створення графіків, діаграм та більш складних графічних представлень, а також вивчити способи візуалізації залежностей між різними ознаками.

Питання для обговорення:

- Використання бібліотеки Matplotlib для створення графіків та діаграм.
- Використання бібліотеки Seaborn для більш складних графічних представлень.
- Візуалізація залежностей між даними.

Лабораторна робота 4: Основи статистичного аналізу даних

Мета: Ознайомити студентів з основними поняттями описової статистики, розподілів даних, методами візуалізації, а також навчити їх проводити перевірку гіпотез та визначати р-значення для статистичних висновків.

Питання для обговорення:

- Описові статистики: середнє, медіана, мода, дисперсія, тощо.
- Розподіли даних та їх візуалізація.
- Перевірка гіпотез та р-значення.

Лабораторна робота 5: Регресійний аналіз та прогнозування

Мета: Вивчити базові поняття регресійного аналізу, ознайомити студентів з методами простої та багатофакторної регресії, навчити оцінювати якість моделей та використовувати бібліотеку Scikit-learn для побудови прогнозуючих моделей.

Питання для обговорення:

- Вступ до регресійного аналізу.
- Проста та багатофакторна регресія.
- Оцінка якості моделі.

- Використання бібліотеки Scikit-learn для побудови моделей.

Лабораторна робота 6: Баєсівський аналіз

Мета: Представити студентам введення до баєсівської статистики, навчити їх використовувати баєсівський підхід для моделювання невизначеності, а також познайомити з баєсівською регресією та класифікацією.

Питання для обговорення:

- Введення до баєсівської статистики.
- Використання баєсівського підходу для моделювання невизначеності.
- Баєсівська регресія та класифікація.

Лабораторна робота 7: Лінійний дискримінантний аналіз

Мета: Ознайомити студентів з методом лінійного дискримінантного аналізу (LDA), навчити їх застосовувати LDA для зменшення розмірності та класифікації даних, а також порівнювати його з методом головних компонент (PCA).

Питання для обговорення:

- Вступ до методу лінійного дискримінантного аналізу (LDA).
- Застосування LDA для зменшення розмірності та класифікації.
- Порівняння з методом головних компонент (PCA).

Лабораторна робота 8: Прогнозування часових рядів

Мета: Вивчити методи аналізу та прогнозування часових рядів, навчити студентів застосовувати моделі ARIMA та експоненційного згладжування для прогнозування майбутніх значень часових рядів.

Питання для обговорення:

- Огляд методів аналізу та прогнозування часових рядів.
- Застосування моделей ARIMA та експоненційного згладжування.
- Оцінка точності прогнозів.

Лабораторна робота 9: Text Mining і Web Mining

Мета: Ознайомити студентів з методами витягнення та аналізу даних з веб-сторінок, навчити обробляти та аналізувати текстові дані, а також використовувати методи аналізу настроїв тексту та виявлення ключових слів.

Питання для обговорення:

- Витягнення та аналіз даних з веб-сторінок.
- Обробка та аналіз текстових даних.
- Аналіз настроїв тексту та виявлення ключових слів.

Лабораторне заняття 10: Обробка та аналіз неструктурованими даних

Мета: Розглянути способи роботи з неструктурованими даними, такими як зображення, аудіо та текст, навчити студентів використовувати відповідні бібліотеки та техніки для обробки цих даних.

Питання для обговорення:

- Робота з неструктурованими даними: роль та значення неструктурованих даних у сучасному аналізі інформації.

- Використання бібліотек для обробки неструктурованих даних: переваги та особливості бібліотек для роботи з різними типами даних (зображення, аудіо, текст).

6. Комплексне практичне індивідуальне завдання

Індивідуальні завдання з дисципліни «Інтелектуальний аналіз даних» виконується самостійно кожним студентом. КПЗ є науковим дослідженням за варіантами, прикладну область обирає студент самостійно. Метою виконання КПЗ є оволодіння навичками застосування методів інтелектуального аналізу даних при розв'язуванні прикладних проблем. КПЗ оформлюється згідно з встановленими вимогами.

Варіанти КПЗ з дисципліни «Інтелектуальний аналіз даних»

№	Тема дослідження
1.	Провести обробку та очищення текстових даних з соціальних мереж для подальшого аналізу.
2.	Визначити ключові патерни в аудіозаписах інтерв'ю пацієнтів для виявлення психічних станів.
3.	Використати нейромережі для розпізнавання об'єктів на зображеннях з відеокамер в реальному часі.
4.	Проаналізувати та класифікувати тексти новин на позитивні та негативні відгуки з використанням методів обробки текстових даних.
5.	Розробити систему виявлення шахрайства на основі аналізу транзакцій та моделей машинного навчання.
6.	Вивчити особливості аналізу геолокаційних даних для прогнозування руху транспорту в місті.
7.	Аналізувати медичні дані для виявлення зв'язків між генетичними факторами та захворюваннями.
8.	Дослідити вплив погодних умов на споживання електроенергії та розробити прогнозування споживання.
9.	Використовувати аналіз даних для виявлення патернів у фінансових ринках та здійснення успішних інвестицій.
10.	Аналізувати дані зі спортивних змагань для прогнозування результатів команд і гравців.
11.	Використати методи важливості функцій для визначення ключових характеристик в аналізі медичних даних.
12.	Проаналізувати фактори, що впливають на вартість нерухомості, застосувавши методи ранжування функцій.
13.	Застосувати методи важливості функцій для визначення ключових змінних у моделі прогнозування споживчих відгуків.
14.	Порівняти різні методи оцінки важливості функцій у контексті класифікації зображень для медичної діагностики.
15.	Визначити найважливіші атрибути в економічних даних для прогнозування зміни валютного курсу.
16.	Використати методи важливості функцій для визначення ключових характеристик в аналізі медичних даних.
17.	Проаналізувати фактори, що впливають на вартість нерухомості, застосувавши методи ранжування функцій.
18.	Побудувати інтерактивну візуалізацію геоданих для аналізу руху пасажирського транспорту у місті.
19.	Створити візуалізацію економічних показників для аналізу впливу подій на ринок.
20.	Створити динамічну візуалізацію для аналізу змін температур на глобальному рівні протягом останніх десяти років.

№	Тема дослідження
21.	Провести аналіз даних про продажі товарів з різних регіонів та з'ясувати, чи є статистично значуща різниця в середніх значеннях.
22.	Використати аналіз дисперсії для визначення впливу факторів на результати спортивних змагань.
23.	Вивчити залежності між різними ознаками з даних про клієнтів та визначити фактори, що впливають на їхні відгуки.
24.	Провести кореляційний аналіз між даними з різних показників економічного розвитку та знайти сильні залежності.
25.	Застосувати аналіз часових рядів для вивчення тенденцій у популярності певних музичних жанрів протягом останнього десятиліття.
26.	Застосувати баєсівський підхід до визначення ймовірності здоров'я пацієнта за його медичними даними.
27.	Використовувати баєсівський метод для побудови прогнозу ризику захворювання на пандемію в майбутньому.
28.	Застосувати баєсівський підхід до оцінки ймовірності виникнення природної катастрофи.
29.	Вивчити баєсівський підхід для визначення ймовірності подій на фондовому ринку та вибору оптимальної стратегії.
30.	Розробити систему оцінки ймовірності відновлення втрачених даних у випадку технічних збоїв за допомогою баєсівського методу.
31.	Застосувати t-тест для перевірки гіпотези про рівність середніх значень доходів чоловіків та жінок.
32.	Використати непараметричний критерій Манна-Уїтні для порівняння рейтингів двох різних рекламних кампаній.
33.	Провести ANOVA для порівняння середніх витрат на рекламу у трьох різних містах.
34.	Визначити, чи є статистично значуща різниця у результатах змагань між командами з різних регіонів.
35.	Використовувати bootstrap метод для оцінки середнього прибутку підприємства з великою кількістю варіацій.
36.	Побудувати модель лінійної регресії для прогнозування вартості нерухомості залежно від різних факторів.
37.	Розробити модель поліноміальної регресії для аналізу зв'язку між рівнем забрудненості повітря та показниками здоров'я населення.
38.	Використати логістичну регресію для аналізу впливу різних факторів на вірогідність придбання товару.
39.	Провести аналіз неоднорідності даних та використати регресію з вагами для моделювання залежності.
40.	Вивчити методи регресії з різною структурою даних та порівняти їх результати у прогнозуванні майбутніх показників.
41.	Використати LDA для зменшення розмірності даних та класифікації зображень різних видів птахів.
42.	Застосувати LDA для аналізу даних про різні види вина та визначення ключових характеристик.
43.	Вивчити вплив зміни параметрів LDA на ефективність класифікації даних з медичних обстежень.
44.	Порівняти результати лінійного дискримінантного аналізу з методом головних компонент у зменшенні розмірності даних.
45.	Застосувати LDA для класифікації різних видів спортивних рекордів залежно від географічних факторів.
46.	Побудувати модель ARIMA для прогнозування цін на акції певної компанії на фондовому ринку.
47.	Використати експоненційне згладжування для прогнозування продажів товарів на наступні місяці.

№	Тема дослідження
48.	Розробити систему прогнозування зміни погоди на основі аналізу попередніх метеоданих.
49.	Провести аналіз та прогнозування руху трафіку на дорогах міста з використанням часових рядів.
50.	Вивчити можливість прогнозування вартості криптовалюти на наступний рік за допомогою методів аналізу часових рядів.
51.	Витягнути та проаналізувати дані зі статей на новинних порталах для визначення ключових тем обговорення.
52.	Застосувати методи Text Mining для аналізу соціальних мереж та визначення популярних трендів.
53.	Розробити систему виявлення фейків та маніпуляцій в новинах за допомогою аналізу текстових даних.
54.	Провести аналіз текстів відгуків користувачів на веб-сайтах для визначення задоволеності клієнтів.
55.	Використовувати Text Mining для аналізу відгуків пацієнтів про якість медичних послуг та виявлення проблем.

7. Самостійна робота

№ з/п	Тематика	Кількість годин	
		ДФН	ЗФН
1	Концепції інтелектуального аналізу даних та DataMining	8	14
2	Попередня обробка та аналіз даних	6	10
3	Стратегії обробки даних з відсутністю інформації та залежністю від часу	6	10
4	Аналіз великих наборів даних та методи зменшення розмірності	6	10
5	Методи візуалізації	6	14
6	Статистичний аналіз	6	10
7	Методи оцінки результатів	4	10
8	Лінійний дискримінантний аналіз	6	10
9	Регресійні моделі	6	20
10	Прогнозування часових рядів	6	20
11	TextMining і WebMining	6	10
Разом:		66	138

8. Тренінг з дисципліни

Тематика: Розробка ПЗ з використанням методів інтелектуального аналізу даних.

№ з/п	Вид роботи	Порядок проведення тренінгу
1	Вступна частина	ознайомлення студентів з темою тренінгового заняття і видача завдання
2	Практична частина	виконання завдань студентами згідно з індивідуальним завданням; оформлення короткого звіту
3	Підведення підсумків	Презентація та оцінювання результатів виконаних завдань

9. Засоби оцінювання та методи демонстрування результатів навчання

У процесі вивчення дисципліни «Інтелектуальний аналіз даних» використовуються наступні засоби оцінювання та методи демонстрування результатів навчання:

- поточне тестування та опитування;
- оцінювання результатів лабораторних робіт;
- ректорська контрольна робота;
- оцінювання виконання завдань тренінгу;
- оцінювання результатів КППЗ;
- екзамен.

10. Критерії, форми поточного та підсумкового контролю

Підсумковий бал (за 100-бальною шкалою) з дисципліни «Інтелектуальний аналіз даних» визначається як середньозважена величина, залежно від питомої ваги кожної складової залікового кредиту:

Заліковий модуль 1	Заліковий модуль 2	Заліковий модуль 3	Екзамен
20 %	20 %	20 %	40 %
1. Виконання та захист лабораторних робіт (5 робіт по 5 балів) – 25 балів 2. Модульна контрольна робота – 75 балів	1. Виконання та захист лабораторних робіт (5 робіт по 5 балів) – 25 балів 2. Ректорська контрольна робота – 75 балів	1. Виконання завдань під час тренінгу – 20 балів 2. Написання та захист КППЗ – 80 балів	1. Тестові завдання (25 тестів по 2 бали) – 50 балів 2. Завдання 1 – 25 балів 3. Завдання 2 – 25 балів

Шкала оцінювання:

За шкалою ЗУНУ	За національною шкалою	За шкалою ECTS
90-100	Відмінно	A (відмінно)
85-89	Добре	B (дуже добре)
75-84		C (добре)
65-74	Задовільно	D (задовільно)
60-64		E (достатньо)
35-59	Незадовільно	FX (незадовільно, з можливістю повторного складання)
1-34		F (незадовільно, з обов'язковим повторним курсом)

11. Інструменти, обладнання та програмне забезпечення, використання яких передбачає навчальна дисципліна

№	Найменування	Номер теми
1.	Середовище Python	1-11
2.	Мультимедійне обладнання	1-11

РЕКОМЕНДОВАНІ ДЖЕРЕЛА ІНФОРМАЦІЇ

1. Гороховатський В. О. Методи інтелектуального аналізу та оброблення даних : навч. посіб. / В. О. Гороховатський, І. С. Творошенко ; М-во освіти і науки України, Харків. нац. ун-т радіоелектроніки. – Харків : ХНУРЕ, 2021. – 92 с.
2. Бахрушин В. Є. Методи аналізу даних: Навч. посібник / В. Є. Бахрушин. – Запоріжжя: КПУ, 2019. – 268 с.
3. Інтелектуальний аналіз даних: Комп'ютерний практикум [Електронний ресурс] : навч. посіб. для студ. спеціальності 122 «Комп'ютерні науки та інформаційні технології», спеціалізацій «Інформаційні системи та технології проектування», «Системне проектування сервісів» / О. О. Сергеев-Горчинський, Г. В. Іщенко ; КПІ ім. Ігоря Сікорського.– Київ : КПІ ім. Ігоря Сікорського, 2020. – 73 с.
4. Акіменко В.В. Прикладні задачі інтелектуального аналізу даних (DATA MINING). К.: КНУ ім. Тараса Шевченка, 2020. 152 с.
5. Jiajun, Z., Zong, C., & Xia, R. (2022). Text Data Mining. Springer.
6. Li, B., Yue, L., Jiang, J., Chen, W., Li, X., Long, G., Fang, F., & Yu, H. (Ред.). (2022). Advanced Data Mining and Applications. Springer International Publishing. <https://doi.org/10.1007/978-3-030-95408-6>
7. Park, L. A. F., Gomes, H. M., Doborjeh, M., Boo, Y. L., Koh, Y. S., Zhao, Y., Williams, G., & Simoff, S. (Ред.). (2022). Data Mining. Springer Nature Singapore. <https://doi.org/10.1007/978-981-19-8746-5>
8. Tan, Y., & Shi, Y. (Ред.). (2022). Data Mining and Big Data. Springer Nature Singapore. <https://doi.org/10.1007/978-981-19-8991-9>
9. Ufuk Baytar, C. (Ред.). (2022). The Future of Data Mining. Nova Science Publishers. <https://doi.org/10.52305/kcin5931>
10. Cao, J. (2023). E-Commerce Big Data Mining and Analytics. Springer Nature Singapore. <https://doi.org/10.1007/978-981-99-3588-8>
11. Konys, A., & Nowak-Brzezińska, A. (Ред.). (2023). Knowledge Engineering and Data Mining. MDPI. <https://doi.org/10.3390/books978-3-0365-6789-1>
12. Mai, T. T., Crane, M., & Bezbradica, M. (Ред.). (2023). Educational Data Mining und Learning Analytics. Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-658-39607-7>
13. Olson, D. L., & Lauhoff, G. (2023). Deskriptives Data-Mining. Springer Nature Switzerland. <https://doi.org/10.1007/978-3-031-21274-1>
14. Shah, K., Shah, N., Sawant, V., & Parolia, N. (2023). Practical Data Mining Techniques and Applications. Auerbach Publications. <https://doi.org/10.1201/9781003390220>
15. Zhang, H. (2023a). Handbook of Mobility Data Mining, Volume 1: Data Preprocessing and Visualization. Elsevier.
16. Zhang, H. (2023b). Handbook of Mobility Data Mining, Volume 2: Mobility Analytics and Prediction. Elsevier.
17. Zhang, H. (2023c). Handbook of Mobility Data Mining, Volume 3: Mobility Data-Driven Applications. Elsevier.