

**ДЕРЖАВНИЙ ВИЩИЙ НАВЧАЛЬНИЙ ЗАКЛАД
«УЖГОРОДСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ»
ФАКУЛЬТЕТ СУСПІЛЬНИХ НАУК
КАФЕДРА ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ СИСТЕМ**



РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

**«ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ СТРУКТУР ДАНИХ»
«INTELLIGENT ANALYSIS OF DATA STRUCTURES»**

Рівень вищої освіти	Третій (освітньо-науковий)
Галузь знань	12 Інформаційні технології
Спеціальність	121 Інженерія програмного забезпечення
Освітня програма	Інженерія програмного забезпечення
Статус дисципліни	Обов'язкова
Мова навчання	Українська

Ужгород 2024

Робоча програма навчальної дисципліни «Інтелектуальний аналіз структур даних» («Intelligent analysis of data structures») розроблена для здобувачів третього рівня вищої освіти галузі знань 12 Інформаційні технології, спеціальності 121 Інженерія програмного забезпечення, освітньо-наукової програми «Інженерія програмного забезпечення».

Розробник: Повхан І.Ф., д.т.н., проф., професор кафедри програмного забезпечення систем ДВНЗ «УжНУ»

Робочу програму розглянуто та затверджено на засіданні кафедри програмного забезпечення систем

протокол № 6 від «31» травня 2024 р.

Завідувач кафедри Юрій БІЛАК

Схвалено науково-методичною комісією факультету інформаційних технологій

протокол № 8 від «20» серпня 2024 р.

Т.в.о. голови науково-методичної комісії Ігор ПОВХАН

ОПИС НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Найменування показників	Розподіл годин за навчальним планом	
	Денна форма навчання	Заочна форма навчання
Кількість кредитів ЄКТС – 4	Рік підготовки:	
Загальна кількість годин – 120	1	1
Кількість модулів – 2	Семестр:	
Тижневих годин: для денної форми навчання: аудиторних – 3.2 самостійної роботи здобувача – 3	2	2
	Лекції:	
	24	8
	Практичні (семінарські):	
	24	6
Вид підсумкового контролю: екзамен	Лабораторні:	
	-	-
Форма підсумкового контролю: письмова або тестова	Самостійна робота:	
	72	106

2. МЕТА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Основною метою вивчення дисципліни «Інтелектуальний аналіз структур даних» є формування у здобувачів вищої освіти третього освітньо-наукового рівня системи фундаментальних знань, дослідницьких компетентностей та практичних навичок, необхідних для розв'язання комплексних наукових та прикладних проблем у сфері аналізу великих обсягів інформації в задачах теорії штучного інтелекту широкого спектру. Даний курс орієнтований на глибоке розуміння внутрішньої будови, математичних основ та алгоритмічної ефективності методів інтелектуального аналізу. Дисципліна передбачає перехід від рівня користувача інструментів аналізу даних до рівня дослідник-архітектор, здатного критично оцінювати існуючі підходи, пропонувати власні модифікації алгоритмів та розробляти нові підходи/моделі для роботи зі структурованими та неструктурованими даними.

У результаті вивчення дисципліни «Інтелектуальний аналіз структур даних» аспіранти повинні знати та вміти:

- Сучасну класифікацію методів інтелектуального аналізу (ІА) – від класичних статистичних критеріїв до глибоких нейромережевих архітектур (CNN, RNN, Transformer, GNN) та ансамблевих підходів (Bagging, Boosting, Stacking).
- Математичне підґрунтя методів: теорію ймовірностей, лінійну алгебру, методи оптимізації (градієнтний спуск та його модифікації), функції втрат та регуляризацію.
- Принципи оцінки якості моделей: метрики класифікації (Accuracy, Precision, Recall, F1, AUC-ROC), регресії (MAE, MSE, RMSE), кластеризації (Silhouette, Dunn index) та методи валідації (Cross-Validation, Bootstrapping).
- Методи попередньої обробки даних: техніки очищення (виявлення викидів), нормалізації/стандартизації, кодування категоріальних змінних, балансування вибірок (SMOTE, ADASYN) та методи боротьби з пропусками (імпутація).
- Стратегії зменшення розмірності даних: метод головних компонент (PCA), метод t-SNE, UMAP та автоенкодера, а також критерії вибору дискретних інформативних ознак (Feature Selection).
- Особливості роботи з різними типами структур даних: текстовими (NLP), зображеннями (CV), часовими рядами, графовими структурами та просторовими даними.
- Сучасний інструментарій (бібліотеки Python: Pandas, NumPy, Scikit-learn, PyTorch/TensorFlow) та платформи для великих даних (Spark MLlib).
- Критично обирати метод ІА, обґрунтовуючи свій вибір виходячи з типу даних, розміру вибірки, обчислювальних ресурсів та специфіки наукової задачі.

- Самостійно реалізовувати (з «нуля» або модифікувати) базові алгоритми машинного навчання без використання високорівневих бібліотек для глибшого розуміння їхньої роботи.
- Виконувати повний цикл EDA (Exploratory Data Analysis) – візуалізацію, виявлення аномалій, перевірку статистичних гіпотез щодо розподілу даних.
- Будувати пайплайни автоматичної обробки даних (Data Pipelines) з урахуванням принципів MLOps для подальшого масштабування дослідження.
- Розв'язувати задачі оптимізації гіперпараметрів (GridSearch, RandomSearch, Bayesian Optimization) для покращення точності моделей.
- Інтерпретувати результати роботи складних моделей, використовуючи методи SHAP, LIME або аналіз вагів для пояснення прогнозів.
- Протистояти переповненню (Overfitting) та недостатньому навчанню (Underfitting), застосовуючи відповідні техніки регуляризації, аугментації та дропау.
- Формулювати висновки та представляти результати аналізу у вигляді наукових звітів, публікацій та візуальних дашбордів, здатних переконати як фахову, так і нефармову аудиторію.

Для досягнення поставленої мети у процесі вивчення дисципліни «Інтелектуальний аналіз структур даних» вирішуються такі концептуальні завдання:

- Поглиблення теоретичних знань: систематизація знань про сучасні парадигми інтелектуального аналізу, включаючи глибоке навчання, ансамблеві методи, графові нейронні мережі та методи обробки просторово-часових рядів. Особливу увагу приділено аналізу теоретичних меж застосовності моделей (bias-variance tradeoff, VC-розмірність) та питанням інтерпретабельності чорних скриньок.
- Розвиток дослідницьких навичок: формування здатності самостійно формулювати наукові гіпотези, обирати оптимальні стратегії попередньої обробки (preprocessing) та інженерії ознак (feature engineering) для специфічних предметних областей, а також коректно планувати експерименти для верифікації отриманих результатів.
- Алгоритмічна компетентність: опанування технік оптимізації обчислювальних процесів, розуміння принципів розподілених обчислень (MapReduce, Spark) та ефективної роботи з великими даними, що виходять за межі оперативної пам'яті (Out-of-core computing).
- Критичне оцінювання: вироблення навичок порівняльного аналізу класичних статистичних методів та сучасних методів штучного інтелекту для уникнення типових логічних помилок (прокляття розмірності, перехресна кореляція, data leakage) та забезпечення відтворюваності наукових результатів.

У результаті опанування курсу аспірант має бути здатним не лише застосовувати пакети аналізу даних, але й проектувати архітектуру систем підтримки прийняття рішень. Це передбачає вміння будувати гібридні моделі, які поєднують переваги символічних та субсимвольних методів, адаптуючи їх під специфіку наукової галузі дослідника.

Важливим аспектом є також формування академічної доброчесності в контексті аналізу даних, що включає коректне управління даними (data management), забезпечення конфіденційності та етичні аспекти використання прогностичних алгоритмів.

Кінцевим результатом вивчення дисципліни «Інтелектуальний аналіз структур даних» є підготовка фахівця, здатного інтегрувати методи інтелектуального аналізу у власне дисертаційне дослідження як на етапі обробки емпіричних даних, так і для створення нових теоретичних узагальнень. Отримані знання стануть фундаментом для написання високоякісних наукових публікацій у виданнях, що індексуються у міжнародних наукометричних базах, та для подальшої професійної діяльності у сфері Data Science.

Відповідно до освітньо-наукової програми, вивчення дисципліни сприяє формуванню таких компетентностей:

Інтегральна компетентність (ІК):

Здатність продукувати нові ідеї, розв'язувати комплексні проблеми професійної та/або дослідницько-інноваційної діяльності у сфері інженерії програмного забезпечення та з дотичних до неї міждисциплінарних напрямів, застосовувати методологію наукової та педагогічної діяльності, проводити власне наукове дослідження, результати якого мають наукову новизну, теоретичне та практичне значення.

Спеціальні (фахові, предметні) компетентності (СК):

СК08 – Здатність здійснювати та організовувати науково педагогічну діяльність у закладах вищої освіти.

СК10 – Здобуття ґрунтовних знань зі спеціальності, за якою аспірант проводить дослідження, зокрема засвоєння основних концепцій, розуміння теоретичних і практичних проблем, історії розвитку та сучасного стану наукових знань за відповідною спеціальністю, володіння термінологією з досліджуваного напрямку в обсязі кредитів ЄКТС відповідно до стандарту вищої освіти.

СК11 – Здатність досліджувати, розробляти, проектувати, моделювати та реалізовувати інтелектуальні системи підтримки прийняття рішень.

3. ПЕРЕДУМОВИ ДЛЯ ВИВЧЕННЯ НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Обов'язковою передумовою вивчення навчальної дисципліни «Інтелектуальний аналіз структур даних» є здобутий другий рівень вищої освіти (магістр). Необхідними є базові знання з інженерії програмного забезпечення, комп'ютерних мереж, баз даних, алгоритмів, математичного моделювання та методів штучного інтелекту.

4. ОЧІКУВАНІ РЕЗУЛЬТАТИ НАВЧАННЯ

Відповідно до освітньо-наукової програми «Інженерія програмного забезпечення», вивчення навчальної дисципліни повинно забезпечити досягнення здобувачами таких програмних результатів навчання (РН):

Програмні результати навчання	Шифр РН
Мати передові концептуальні та методологічні знання з інженерії програмного забезпечення та дотичних до неї міждисциплінарних напрямів, а також дослідницькі навички, достатні для проведення наукових і прикладних досліджень на рівні останніх світових досягнень з відповідного напрямку, отримання нових знань та/або здійснення інновацій.	РН01
Планувати і виконувати експериментальні та/або теоретичні дослідження з інженерії програмного забезпечення та дотичних міждисциплінарних напрямів з використанням сучасних інструментів та дотриманням норм академічної і професійної етики, критично аналізувати результати власних досліджень і результати інших дослідників у контексті усього комплексу сучасних знань щодо досліджуваної проблеми.	РН02
Пропонувати нові ефективні методи і моделі розроблення, впровадження, супроводу та забезпечення якості програмного забезпечення та управління відповідними процесами на всіх етапах життєвого циклу.	РН03
Застосовувати сучасні інструменти і технології пошуку, оброблення та аналізу інформації, зокрема, статистичні методи аналізу даних великого обсягу та/або складної структури, спеціалізовані бази даних та інформаційні системи для покращення ефективності програмних систем.	РН05
Розробляти та досліджувати концептуальні, матема-	РН07

тичні і комп'ютерні моделі процесів і систем для отримання нових знань та/або створення інноваційних продуктів у інженерії програмного забезпечення та дотичних міждисциплінарних напрямках.	
Глибоко розуміти загальні принципи та методи інженерії програмного забезпечення, а також методологію наукових досліджень, застосовувати їх у власних дослідженнях та у викладацькій практиці.	RH08
Формулювати та вирішувати задачі оптимізації, адаптації, прогнозування, керування та прийняття рішень щодо процесів, засобів та ресурсів розробки, впровадження, супроводу та експлуатації програмного забезпечення.	RH09
Здійснювати дослідження та проектування всіх етапів управління процесами життєвого циклу складних систем на основі сучасних методів інтелектуального аналізу даних.	RH15
Вміти здійснювати дослідження, моделювати та проектувати засоби програмного забезпечення для прийняття рішень в умовах ризику та невизначеності для розв'язування прикладних задач цифрової трансформації.	RH16

Очікувані результати навчання, які повинні бути досягнуті здобувачами після опанування навчальної дисципліни «Інтелектуальний аналіз структур даних»:

Програмні результати навчання	Шифр РН
Впроваджувати передові концепції та методології безпечної розробки програмних моделей; критично аналізувати сучасні технології ШІ, моделі аналізу даних; обґрунтовувати вибір методів та інструментів для наукових і прикладних досліджень.	RH01
Адресно планувати та виконувати комплексні програмні моделі методів ШІ та дотичних міждисциплінарних напрямів з використанням сучасних інструментів – дотримуючись норм академічної і професійної етики.	RH02
Розробляти та удосконалювати моделі, методи, ін-	RH03

струменти аналізу даних, програмні прототипи складних структур даних; виконувати фазифікацію, агрегування, дефазифікацію, тестування та верифікацію отриманих результатів.	
Застосовувати та модифікувати сучасні інструменти і технології аналізу складних структур даних заданих в природній формі, для покращення ефективності інформаційних систем.	PH05
Досліджувати і розробляти моделі та методи обробки великих масивів дискретних даних для створення інноваційних продуктів у інженерії програмного забезпечення.	PH07
Вміти проводити глибоку структурну декомпозицію моделей аналізу структур даних на основі методології наукових досліджень, застосовувати отримані результати у власних дослідженнях та у викладацькій практиці.	PH08
Ефективно розв'язувати комплексні задачі оптимізації, адаптації, прогнозування, керування та прийняття рішень щодо процесів глибокого аналізу даних, виділення дискретних ознак і атрибутів.	PH09
Реалізовувати комплексні дослідження/аналіз великий масивів дискретних даних на основі сучасних методів інтелектуального аналізу даних (моделей/інструментів ІІІ).	PH15
Реалізовувати всі етапи комплексного дослідження масивів/структур даних, моделювати/аналізувати та проектувати програмні моделі/реалізації для прийняття рішень в умовах ризику та невизначеності (для розв'язування прикладних задач цифрової трансформації).	PH16

5. ЗАСОБИ ДІАГНОСТИКИ ТА КРИТЕРІЇ ОЦІНЮВАННЯ РЕЗУЛЬТАТІВ НАВЧАННЯ

Загальними засобами оцінювання та методами демонстрування результатів навчання є: практичні роботи; теоретичні модульні контролі; аналіз масивів да-

них/вихідного коду; звіти зі структурною схемою, інструкцією користувача, лістингом і скріншотами; презентація та захист результатів; екзамен.

Форми контролю та критерії оцінювання результатів

Форми поточного контролю: виконання, тестування, демонстрація та захист практичних програмних робіт; письмове або тестове виконання теоретичних модульних контрольних робіт.

Форма підсумкового семестрового контролю: екзамен.

Екзамен проводиться у письмовій або тестовій формі та охоплює теоретичний і практичний зміст двох модулів навчальної дисципліни. До складання підсумкового контролю допускаються здобувачі, які виконали, оформили та захистили обидві практичні роботи.

Особливості використання засобів діагностики та контролю за умов дистанційного навчання

В умовах використання формату онлайн-навчання (дистанційного навчання) із застосуванням корпоративної мережі Google Meet названі засоби, методи і форми визначаються за домовленістю зі студентським колективом і, в залежності від зручного виду взаємодії, застосовуються з допомогою існуючих функцій групових чатів та відео-конференцій.

Для ефективного засвоєння тематики є можливість демонстрації необхідних матеріалів на робочому столі комп'ютерного технічного засобу під час занять.

Зокрема, у разі потреби, під час онлайн-заняття можна надати доступ до свого екрану, щоб показати презентації або іншу тематичну інформацію на робочому столі.

Планування лекційних і лабораторних занять, модульних контрольних робіт, а також підсумкова перевірка знань у формі екзамену здійснюється заздалегідь за допомогою прив'язки до гугл-календаря. Синхронізація запланованих заходів виконується автоматично на всіх зручних для їх проведення пристроях.

Розподіл балів, які отримують здобувачі вищої освіти

Змістовий модуль	Практична робота	Теоретичний модульний контроль	Максимум за модуль
Модуль 1 (Т1–Т6)	Практична робота №1 – 30 балів	Модульний контроль №1 – 70 балів	100 балів
Модуль 2 (Т7–Т12)	Практична робота №2 – 30 балів	Модульний контроль №2 – 70 балів	100 балів

Оцінювання окремих видів навчальної роботи з дисципліни

Вид діяльності здобувача вищої освіти		
	Кількість	Максимальна кількість балів (сумарна)

Практичні заняття	3 заняття по 10	30
Модульна контрольна робота №1	1	70
Разом		100
Практичні заняття	3 заняття по 10	30
Модульна контрольна робота №2	1	70
Разом		100

Критерії оцінювання підсумкового семестрового контролю.

Оцінювання здійснюється окремо за кожним змістовим модулем за 100-бальною шкалою. Результат модуля 1 формується як сума балів за практичну роботу №1 (30 балів) і теоретичний модульний контроль №1 (70 балів). Результат модуля 2 формується як сума балів за практичну роботу №2 (30 балів) і теоретичний модульний контроль №2 (70 балів). Підсумкова семестрова оцінка та оцінка за екзамен визначаються як середнє арифметичне результатів двох модулів.

- Практична робота (30 балів): коректність програмної реалізації та обчислень – 12 балів; повнота виконання завдання – 6 балів; тестування і верифікація результатів – 4 бали; якість звіту та програмної документації – 4 бали; презентація і захист – 4 бали.
- Теоретичний модульний контроль (70 балів): у письмовій формі – 5 описових питань по 14 балів; у тестовій формі – 20 тестових завдань по 3,5 бала. Максимальна сума за кожний теоретичний модульний контроль становить 70 балів.

Підсумкова оцінка = (Модуль 1 + Модуль 2) / 2
--

Підсумкова екзаменаційна оцінка визначається за накопичувальною системою як середнє арифметичне результатів першого і другого змістових модулів.

Переведення даних 100-бальної шкали оцінювання у оцінки за національною шкалою та шкалою ЄКТС

Сума балів	Оцінка ЄКТС	Оцінка за національною шкалою	
		екзамен, диф. залік	залік
90 -100	A	відмінно	зараховано
82 - 89	B	добре	
74 - 81	C		
64 - 73	D	задовільно	
60 - 63	E		
35 - 59	FX	незадовільно	не зараховано
0 - 34	F		

Відповідно до Положення про порядок та методику проведення семестрових (курсівих) екзаменів і заліків в Ужгородському національному університеті,

затвердженого наказом ректора ДВНЗ «УжНУ» від 08.05.2015 р. № 698/01-17, знання здобувачів оцінюється як з теоретичної, так і з практичної підготовки за такими критеріями:

оцінку «відмінно» (90-100 балів, А) заслуговує здобувач, який:

всебічно і глибоко володіє навчально-програмовим матеріалом;

вміє самостійно виконувати завдання, передбачені програмою, використовує набуті знання і вміння у нестандартних ситуаціях;

засвоїв основну і ознайомлений з додатковою літературою, яка рекомендована програмою;

при виконанні практичного завдання застосовує системні знання навчального матеріалу, передбачені навчальною програмою.

оцінку «добре» (82-89 балів, В) – заслуговує здобувач, який:

повністю опанував і вільно (самостійно) володіє навчально-програмовим матеріалом, в тому числі застосовує його на практиці, має системні знання в достатньому обсязі відповідно до навчально-програмового матеріалу, аргументовано використовує їх у різних ситуаціях;

має здатність до самостійного пошуку інформації, а також до аналізу, постановки і розв'язування проблем професійного спрямування;

під час відповіді допустив деякі неточності, які самостійно виправив, добирає переконливі аргументи на підтвердження вивченого матеріалу;

оцінку «добре» (74-81 бал, С) – заслуговує здобувач, який:

в цілому навчальну програму засвоїв, але відповідає з певною кількістю помилок;

вміє порівнювати, узагальнювати, систематизувати інформацію під керівництвом викладача, в цілому самостійно застосовувати на практиці, контролювати власну діяльність;

опанував навчально-програмовий матеріал, успішно виконав завдання, передбачені програмою, засвоїв основну літературу, яка рекомендована програмою;

оцінку «задовільно» (64-73 бали, D) – заслуговує здобувач, який:

знає основний навчально-програмовий матеріал в обсязі, необхідному для подальшого навчання і використання його у майбутній професії;

виконує завдання непогано, але зі значною кількістю помилок;

ознайомлений з основною літературою, яка рекомендована програмою;

допускає помилки при виконанні завдань, але під керівництвом викладача знаходить шляхи їх усунення.

оцінку «задовільно» (60-63 бали, E) – заслуговує здобувач, який:

володіє основним навчально-програмовим матеріалом в обсязі, необхідному для подальшого навчання і використання його у майбутній професії, а виконання завдань задовольняє мінімальні критерії. Знання мають репродуктивний характер.

оцінка «незадовільно» (35-59 балів, FX) – виставляється здобувачу, який:

виявив суттєві прогалини в знаннях основного програмового матеріалу, допустив принципові помилки у виконанні передбачених програмою завдань.

оцінка «незадовільно» (35 балів, F) – виставляється здобувачу, який:

володіє навчальним матеріалом тільки на рівні елементарного розпізнавання і відтворення окремих фактів або не володіє зовсім;

допускає грубі помилки при виконанні завдань, передбачених програмою;

не може продовжувати навчання і не готовий до професійної діяльності після закінчення університету без повторного вивчення даної дисципліни.

При виставленні оцінки враховуються результати навчальної роботи студента протягом семестру.

Академічна доброчесність: Політика щодо академічної доброчесності учасників освітнього процесу формується на основі дотримання принципів академічної доброчесності з урахуванням норм Положення про академічну доброчесність в ДВНЗ «Ужгородський національний університет, затвердженого вченою радою університету від 23.02.2017 р., протокол № 3, <https://www.uzhnu.edu.ua/uk/infocentre/get/12223>.

Правила перезарахування кредитів у випадку мобільності: Політика щодо перезарахування кредитів ЄКТС у випадку мобільності формується з урахуванням норм Положення про порядок визнання (перезарахування) кредитів ЄКТС для учасників програм академічної мобільності у Державному вищому навчальному закладі «Ужгородський національний університет», затвердженого вченою радою університету від 30.05.2019 р., протокол № 6, <https://www.uzhnu.edu.ua/uk/infocentre/get/21266>.

Правила визнання результатів навчання, отриманих у неформальній та/або інформальній освіті: Політика щодо визнання результатів навчання, отриманих у неформальній та/або інформальній освіті, формується з урахуванням норм Положення про порядок визнання в Державному вищому навчальному закладі «Ужгородський національний університет» результатів навчання, здобутих протягом неформальної та / або інформальної освіти, затвердженого вченою радою університету від 18.12.2023 р., протокол № 11, <https://www.uzhnu.edu.ua/uk/infocentre/get/69166>.

6. ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

6.1. Зміст навчальної дисципліни

Змістовий модуль 1. дисципліни «Інтелектуальний аналіз структур даних»:

Тема 1. Теоретико-методологічні засади інтелектуального аналізу даних. Аналітичне мислення та прийняття рішень на основі даних. Формалізація задач, кла-

сифікація типів даних (структуровані, напівструктуровані, неструктуровані). Процес KDD (Knowledge Discovery in Databases).

Тема 2. Структури даних для великих обсягів інформації (Big Data). Моделі зберігання та організації даних у розподілених системах. Робота з таблицями, базами даних та сховищами даних. Огляд сучасних архітектур для обробки великих даних.

Тема 3. Візуалізація та глибокий аналіз даних (EDA). Сучасні методи візуалізації багатовимірних даних. Інтерактивні дашборди та бізнес-аналітика. Виявлення аномалій, пропущених значень та викидів.

Тема 4. Задачі прогнозування та регресійного аналізу. Методи структурної та параметричної ідентифікації невідомих залежностей. Лінійні та нелінійні регресійні моделі. Оцінювання якості прогнозування.

Тема 5. Методи класифікації даних. Алгоритми наївного Басса, метод k -найближчих сусідів, дерева рішень, випадковий ліс. Градієнтний бустинг (XGBoost, LightGBM, CatBoost). Оцінка ефективності класифікаторів.

Тема 6. Алгоритми кластеризації та сегментації. Метрики подібності та відстані. K-means, ієрархічна кластеризація, DBSCAN, Gaussian Mixture Models. Відновлення даних та зменшення розмірності (PCA, t-SNE, UMAP).

Змістовий модуль 2. дисципліни «Інтелектуальний аналіз структур даних»:

Тема 7. Пошук патернів та асоціативних правил. Аналіз транзакційних даних, алгоритм Apriori та його оптимізації. Виявлення прихованих закономірностей та кореляцій у великих масивах даних

Тема 8. Машинне навчання та глибинне навчання (Deep Learning). Архітектури нейронних мереж: повнозв'язні, згорткові (CNN), рекурентні (RNN, LSTM). Автокодувальники, генеративно-змагальні мережі (GAN). Трансформери та увага.

Тема 9. Інтелектуальна обробка природної мови (NLP). Методи токенізації, векторизації (Word2Vec, GloVe, FastText). Сучасні мовні моделі (BERT, GPT). Тематичне моделювання (LDA), аналіз тональності, розпізнавання іменованих сутностей.

Тема 10. Гібридні системи обчислювального інтелекту. Нейро-нечіткі системи, еволюційні алгоритми (генетичні алгоритми, диференціальна еволюція). Гібридизація методів для підвищення якості аналізу.

Тема 11. Поточкова обробка даних (Data Stream Mining). Алгоритми для роботи з даними, що надходять у реальному часі. Еволюційні системи, адаптивні моделі. Оцінювання концептуального дрейфу (concept drift).

Тема 12. Методи пояснювального ШІ (Explainable AI). Інтерпретація моделей «чорної скриньки» (LIME, SHAP). Побудова довірених та прозорих систем прийняття рішень. Етика та відповідальність у застосуванні ШІ.

6.2. Структура навчальної дисципліни

Назви тем	Денна форма			
	усього	у тому числі		
		л	прак	с.р.
1	2	3	4	5
Тема 1. Теоретико-методологічні засади інтелектуального аналізу даних. Аналітичне мислення та прийняття рішень на основі даних.	9	2	2	5
Тема 2. Структури даних для великих обсягів інформації (Big Data). Моделі зберігання та організації даних у розподілених системах.	9	2	2	5
Тема 3. Візуалізація та глибокий аналіз даних (EDA). Сучасні методи візуалізації багатовимірних даних.	9	2	2	5
Тема 4. Задачі прогнозування та регресійного аналізу. Методи структурної та параметричної ідентифікації невідомих залежностей.	9	2	2	5
Тема 5. Методи класифікації даних. Алгоритми наївного Баеса, метод k-найближчих сусідів, дерева рішень, випадковий ліс.	11	2	2	7
Тема 6. Алгоритми кластеризації та сегментації. Метрики подібності та відстані. K-means, ієрархічна кластеризація, DBSCAN, Gaussian Mixture Models.	9	2	2	5
Тема 7. Пошук патернів та асоціативних правил. Аналіз транзакційних даних, алгоритм Apriori та його оптимізації.	9	2	2	5
Тема 8. Машинне навчання та глибинне навчання (Deep Learning). Архітектури нейронних мереж: повнозв'язні, згорткові (CNN), рекурентні (RNN, LSTM).	14	2	2	10
Тема 9. Інтелектуальна обробка природної мови (NLP). Методи токенізації, векторизації (Word2Vec, GloVe, FastText).	9	2	2	5
Тема 10. Гібридні системи обчислювального інтелекту. Нейро-нечіткі системи, еволюційні алгоритми (генетичні алгоритми, диференціальна ево-	14	2	2	10

Назви тем	Денна форма			
	усього	у тому числі		
		л	прак	с.р.
1	2	3	4	5
люція).				
Тема 11. Потокова обробка даних (Data Stream Mining). Алгоритми для роботи з даними, що надходять у реальному часі.	9	2	2	5
Тема 12. Методи пояснювального ШІ (Explainable AI). Інтерпретація моделей «чорної скриньки» (LIME, SHAP).	9	2	2	5
Всього	120	24	24	72

Назви тем	Денна форма				Заочна форма			
	усього	у тому числі			усього	у тому числі		
		л	прак	с.р.		л	прак	с.р.
1	2	3	4	5	6	7	8	9
Тема 1. Теоретико-методологічні засади інтелектуального аналізу даних. Аналітичне мислення та прийняття рішень на основі даних.	9	2	2	5				
Тема 2. Структури даних для великих обсягів інформації (Big Data). Моделі зберігання та організації даних у розподілених системах.	9	2	2	5				
Тема 3. Візуалізація та глибокий аналіз даних (EDA). Сучасні методи візуалізації багатовимірних даних.	9	2	2	5				
Тема 4. Задачі прогнозування та регресійного аналізу. Методи структурної та параметричної ідентифікації невідомих залежностей.	9	2	2	5				
Тема 5. Методи класифікації даних. Алгоритми наївного Баєса, метод k-найближчих сусідів, дерева рішень, випадковий ліс.	11	2	2	7				

Назви тем	Денна форма				Заочна форма			
	усього	у тому числі			усього	у тому числі		
		л	прак	с.р.		л	прак	с.р.
1	2	3	4	5	6	7	8	9
Тема 6. Алгоритми кластеризації та сегментації. Метрики подібності та відстані. K-means, ієрархічна кластеризація, DBSCAN, Gaussian Mixture Models.	9	2	2	5				
Тема 7. Пошук патернів та асоціативних правил. Аналіз транзакційних даних, алгоритм Apriori та його оптимізації.	9	2	2	5				
Тема 8. Машинне навчання та глибинне навчання (Deep Learning). Архітектури нейронних мереж: повнозв'язні, згорткові (CNN), рекурентні (RNN, LSTM).	14	2	2	10				
Тема 9. Інтелектуальна обробка природної мови (NLP). Методи токенизації, векторизації (Word2Vec, GloVe, FastText).	9	2	2	5				
Тема 10. Гібридні системи обчислювального інтелекту. Нейро-нечіткі системи, еволюційні алгоритми (генетичні алгоритми, диференціальна еволюція).	14	2	2	10				
Тема 11. Поточкова обробка даних (Data Stream Mining). Алгоритми для роботи з даними, що надходять у реальному часі.	9	2	2	5				
Тема 12. Методи пояснювального ШІ (Explainable AI). Інтерпретація моделей «чорної скриньки» (LIME, SHAP).	9	2	2	5				
Всього	120	24	24	72	120	8	6	106

6.3. Теми практичних занять

№ з/п	Назва теми	Денна	Заочна
1	Практична робота №1. Дослідницький аналіз та візуалізація багатовимірних даних. Опанування інструментів EDA (Pandas, Plotly, Seaborn). Побудова інтерактивних дашбордів, виявлення аномалій, кореляцій та прихованих закономірностей у реальних наборах даних.		2
2	Практична робота №2. Побудова та оцінювання регресійних моделей. Реалізація лінійної та поліноміальної регресії, гребеневої регресії (Ridge), LASSO. Оцінювання якості за допомогою метрик MAE, RMSE, R^2 . Відбір значущих ознак.		2
3	Практична робота №3. Реалізація класифікаторів на основі дерев рішень та ансамблів. Побудова дерев рішень, випадкового лісу, градієнтного бустингу (XGBoost, LightGBM). Налаштування гіперпараметрів. Порівняння якості моделей на реальних задачах (наприклад, кредитний скоринг або діагностика).		2
4	Практична робота №4. Застосування алгоритмів кластеризації для сегментації даних. Реалізація K-means, ієрархічної кластеризації, DBSCAN. Вибір оптимальної кількості кластерів. Інтерпретація отриманих сегментів та візуалізація методом PCA/t-SNE.		2
5	Практична робота №5. Виявлення асоціативних правил у транзакційних даних. Реалізація алгоритму Apriori та FP-Growth. Аналіз ринкового кошика, побудова асоціативних правил, оцінка їх достовірності та підтримки.		2
6	Практична робота №6. Інтелектуальна обробка текстових даних (NLP). Токенізація, лематизація, векторизація (Word2Vec, TF-IDF). Побудова моделі класифікації текстів (наприклад, визначення тональності відгуків або тематичне моделювання LDA).		2
7	Практична робота №7. Розробка нейромережевої моделі на основі трансформерів. Побудова та навчання моделі з механізмом уваги (Transformer) для задачі класифікації або генерації тексту. Використання попередньо навчених моделей (BERT, GPT) для тонкого налаштування (fine-tuning).		2
8	Практична робота №8. Побудова гібридної системи обчислю-		2

	вального інтелекту. Реалізація нейро-нечіткої системи (ANFIS) або гібриду генетичного алгоритму з нейронною мережею для оптимізації параметрів складної функції.		
9	Практична робота №9. Поточковий аналіз даних у реальному часі. Розробка конвеєра обробки поточкових даних (Apache Kafka & Python). Реалізація адаптивної моделі класифікації з виявленням концептуального дрейфу (concept drift).		2
10	Практична робота №10. Інтерпретація моделей чорної скриньки з використанням SHAP/LIME. Застосування методів пояснювального ШІ для аналізу рішень складних моделей (бустинг, нейромережі). Візуалізація важливості ознак та побудова локальних пояснень.		2
11	Практична робота №11. Побудова графа знань та семантичний пошук. Створення онтології, наповнення графа знань (Neo4j або RDF). Побудова SPARQL-запитів для інтелектуального пошуку та логічного виводу.		2
10	Практична робота №12. Розробка та захист дослідницького прототипу ШІ-системи. Комплексна робота: вибір предметної області, підготовка даних, побудова, навчання та оцінка моделі. Оформлення результатів у вигляді звіту або наукової публікації. Презентація та захист прототипу перед аудиторією.		2
	Загальна кількість		24

6.4. Самостійна робота

№ з/п	Назва теми	Денна
1	Тема 1. Формалізація задач, класифікація типів даних (структуровані, напівструктуровані, неструктуровані). Процес KDD (Knowledge Discovery in Databases).	5
2	Тема 2. Робота з таблицями, базами даних та сховищами даних. Огляд сучасних архітектур для обробки великих даних.	5
3	Тема 3. Інтерактивні дашборди та бізнес-аналітика. Виявлення аномалій, пропущених значень та викидів.	5
4	Тема 4. Лінійні та нелінійні регресійні моделі. Оцінювання якості прогнозування.	5
5	Тема 5. Градієнтний бустинг (XGBoost, LightGBM, CatBoost). Оцінка	7

	ефективності класи-фікаторів.	
6	Тема 6. Відновлення даних та зменшення розмірності (PCA, t-SNE, UMAP).	5
7	Тема 7. Виявлення прихованих закономірностей та кореляцій у великих масивах даних.	5
8	Тема 8. Автокодувальники, генеративно-змагальні мережі (GAN). Трансформери та увага.	10
9	Тема 9. Сучасні мовні моделі (BERT, GPT). Тематичне моделювання (LDA), аналіз тональності, розпізнавання іменованих сутностей.	5
10	Тема 10. Гібридизація методів для підвищення якості аналізу.	10
11	Тема 11. Еволюційні системи, адаптивні моделі. Оцінювання концептуального дрейфу (concept drift).	5
12	Тема 12. Побудова довірених та прозорих систем прийняття рішень. Етика та відповідальність у застосуванні ШІ.	5
	Загальна кількість	72

7. РЕКОМЕНДОВАНІ ДЖЕРЕЛА ІНФОРМАЦІЇ

Основна література

1. Géron, A. (2022). Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow (3rd ed.). O'Reilly Media.
2. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press.
3. Gupta, D., Bhattacharyya, S., Khanna, A., & Sagar, K. (Eds.). (2020). Intelligent data analysis: From data gathering to data comprehension. Wiley. <https://doi.org/10.1002/9781119544485>
4. Gutiérrez, G. (2025). Data structures & algorithms for AI: Building smart systems. Independently Published.
5. Han, J., Kamber, M., & Pei, J. (2011). Data mining: Concepts and techniques (3rd ed.). Morgan Kaufmann.
6. Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning: Data mining, inference, and prediction (2nd ed.). Springer.
7. Kelleher, J. D., & Tierney, B. (2018). Data science. MIT Press.
8. Mishra, A. (2025). Data structure and algorithms for AI & ML (AI course). Independently Published.

9. Molnar, C. (2022). Interpretable machine learning: A guide for making black box models explainable. Leanpub. <https://christophm.github.io/interpretable-ml-book/>
10. Murphy, K. P. (2012). Machine learning: A probabilistic perspective. MIT Press.
11. Provost, F., & Fawcett, T. (2013). Data science for business: What you need to know about data mining and data-analytic thinking. O'Reilly Media.
12. Russell, S., & Norvig, P. (2021). Artificial intelligence: A modern approach (4th ed.). Pearson.
13. Shalev-Shwartz, S., & Ben-David, S. (2014). Understanding machine learning: From theory to algorithms. Cambridge University Press.
14. Silberschatz, A., Korth, H. F., & Sudarshan, S. (2019). Database system concepts (7th ed.). McGraw-Hill.
15. Smith, J. (2026). Mastering data structures and algorithms with Python and AI: Learn, implement & solve real problems using Python and artificial intelligence. Independently Published.
16. VanderPlas, J. (2016). Python data science handbook: Essential tools for working with data. O'Reilly Media.
17. Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). Data mining: Practical machine learning tools and techniques (4th ed.). Morgan Kaufmann.
18. Zgurovsky, M., & Zaichenko, Y. (2025). Системи та методи штучного інтелекту [Systems and methods of artificial intelligence]. Akademperiodyka.
19. Zhu, X., & Wu, X. (2020). Artificial intelligence for data science: Theory, algorithms, and applications. Springer.
20. Навчальний посібник ВНТУ. (n.d.). Machine Learning та інтелектуальний аналіз даних [Machine learning and intelligent data analysis]. Вінницький національний технічний університет [Vinnytsia National Technical University].

Допоміжна література

1. Abnar, S., & Zuidema, W. (2020). Quantifying attention flow in transformers. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 1040–1050. <https://doi.org/10.18653/v1/2020.acl-main.97>
2. Alom, M. Z., Taha, T. M., Yakopcic, C., Westberg, S., Sidike, P., Nasrin, M. S., Van Essen, B. C., Awwal, A. A. S., & Asari, V. K. (2019). A state-of-the-art survey on deep learning theory and architectures. Electronics, *8*(3), 292. <https://doi.org/10.3390/electronics8030292>
3. Biecek, P., & Burzykowski, T. (2021). Explanatory model analysis: Explore, explain, and examine predictive models. Chapman & Hall/CRC.

4. Bischl, B., Binder, M., Lang, M., Pielok, T., Richter, J., Coors, S., Thomas, J., Ullmann, T., Becker, M., Boulesteix, A. L., & Deng, D. (2021). Hyperparameter optimization: Foundations, algorithms, best practices, and open challenges. *Data Mining and Knowledge Discovery*, *35*(3), 1–35. <https://doi.org/10.1007/s10618-021-00747-9>
5. Borovicka, T., Jirina, M., Kordik, P., & Jirina, M. (2012). Selecting representative data sets. In M. J. Zaki, M. Goebel, & S. K. Das (Eds.), *Advances in data mining* (pp. 1–14). Springer.
6. Hamilton, W. L. (2020). *Graph representation learning*. Morgan & Claypool Publishers.
<https://doi.org/10.2200/S01045ED1V01Y202009AIM046>
7. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778. <https://doi.org/10.1109/CVPR.2016.90>
8. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, *60*(6), 84–90. <https://doi.org/10.1145/3065386>
9. Leskovec, J., Rajaraman, A., & Ullman, J. D. (2020). *Mining of massive datasets* (3rd ed.). Cambridge University Press. <https://doi.org/10.1017/9781108627305>
10. Li, Y., & Qu, H. (2018). Interactive data visualization: A review of recent advances and challenges. *Journal of Visualization*, *21*(2), 1–25. <https://doi.org/10.1007/s12650-018-0469-3>
11. Liu, S., Wang, X., Liu, M., & Zhu, J. (2021). Towards better analysis of machine learning models: A visual analytics perspective. *Visual Informatics*, *5*(4), 27–38. <https://doi.org/10.1016/j.visinf.2021.12.001>
12. Mehta, P., Bukov, M., Wang, C. H., Day, A. G. R., Richardson, C., Fisher, C. K., & Schwab, D. J. (2019). A high-bias, low-variance introduction to machine learning for physicists. *Physics Reports*, *810*, 1–124. <https://doi.org/10.1016/j.physrep.2019.03.001>
13. Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using Siamese BERT-networks. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*, 3982–3992. <https://doi.org/10.18653/v1/D19-1410>
14. Schmidt, M., & Lipson, H. (2009). Distilling free-form natural laws from experimental data. *Science*, *324*(5923), 81–85. <https://doi.org/10.1126/science.1165893>

15. Zhang, C., Bengio, S., Hardt, M., Recht, B., & Vinyals, O. (2021). Understanding deep learning (still) requires rethinking generalization. *Communications of the ACM*, *64*(3), 107–115. <https://doi.org/10.1145/3446776>