

ХМЕЛЬНИЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ

ЗАТВЕРДЖУЮ

Декан факультету _____ ФІТ _____

Тетяна ГОВОРУЩЕНКО

Ім'я, ПРІЗВИЩЕ

2025 р.

РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Аналітика великих даних

Назва дисципліни

Призначення Робочої програми

Для освітніх програм різних спеціальностей

Рівень вищої освіти

Другий (магістерський)

Мова навчання

Українська

Обсяг дисципліни, кредитів ЄКТС

8

Статус дисципліни

Вибіркова

Факультет

Інформаційних технологій

Кафедра

Комп'ютерної інженерії та інформаційних систем

Форма здобуття освіти	Обсяг дисципліни		Кількість годин						Форма семестрового контролю
			Аудиторні заняття					Самостійна робота (в т.ч. ІРС)	Залік
	Кредити ЄКТС	Години	Разом	Лекції	Лабораторні роботи	Практичні заняття	Семінарські заняття		
Д	8	240	82	32	34	16		158	+

Робоча програма складена на основі освітніх програм підготовки магістра та стандарту вищої освіти спеціальності.

Робоча програма складена _____ канд. фіз.-мат. наук, доцент, Тетяна КИСІЛЬ

Підпис

Науковий ступінь, вчене звання, Ім'я, ПРІЗВИЩЕ

Схвалена на засіданні кафедри _____ комп'ютерної інженерії та інформаційних систем

Назва

Протокол від 18.08 2025 № 1.

Зав. кафедри комп'ютерної інженерії та інформаційних систем

Назва

Підпис

Ольга ПАВЛОВА

Ім'я, ПРІЗВИЩЕ

Хмельницький 2025

3. Пояснювальна записка

Дисципліна «Аналітика великих даних» належить до циклу вибірових дисциплін і забезпечує вивчення методів збору, обробки, аналізу та візуалізації великих обсягів даних із використанням сучасних технологій, інструментів та підходів. У процесі навчання студенти здобувають практичні навички роботи з базами даних, аналітичними платформами, бібліотеками для аналізу даних, алгоритмами машинного навчання та прогнозової аналітики. Дисципліна формує готовність застосовувати сучасні методи для вирішення прикладних завдань у сфері Big Data, розвиває мовну підготовку через роботу з англійськими технічними джерелами, документацією та стандартами. Окрім технічних умінь, студенти розвивають soft skills (комунікацію, командну взаємодію, ініціативність, здатність брати відповідальність і приймати рішення на основі аналітичних даних).

Мета дисципліни. Формування у здобувачів вищої освіти теоретичних знань та практичних навичок щодо методів, засобів та технологій аналізу великих даних, здатності застосовувати сучасні інструменти, алгоритми машинного навчання та засоби візуалізації для вирішення аналітичних завдань

Предмет дисципліни. Методи, технології та програмні засоби збору, обробки, аналізу, моделювання та візуалізації великих даних із використанням сучасних платформ та бібліотек аналітики.

Завдання дисципліни. формування практичних навичок роботи зі структурованими та неструктурованими даними; засвоєння принципів архітектур Big Data та технологій розподіленої обробки даних; ознайомлення з інструментами машинного навчання, хмарних сервісів і сучасних аналітичних платформ; розвиток умінь створювати аналітичні моделі, будувати візуалізації та готувати звіти; формування аналітичного мислення, уміння працювати в команді, приймати обґрунтовані рішення та ефективно використовувати інструменти аналітики

Результати навчання. Після вивчення дисципліни здобувач повинен:

знати: основні поняття, концепції та архітектури Big Data; методи збору, обробки, аналізу та візуалізації великих обсягів даних; інструменти машинного навчання та прогнозової аналітики;

володіти: базовими навичками роботи з сучасними аналітичними платформами та бібліотеками; інструментами побудови моделей, аналізу даних і створення інтерактивних звітів;

вміти: застосовувати методи аналізу та прогнозування даних для вирішення практичних завдань; обґрунтовувати вибір інструментів і технологій для реалізації аналітичних проєктів; ефективно працювати з документацією, проводити дослідження та аналізувати отримані результати;

демонструвати: ініціативність, здатність до самостійного навчання та критичного мислення; навички комунікації та роботи в команді; готовність приймати відповідальні рішення під час розв'язання складних аналітичних завдань.

4. Структура залікових кредитів дисципліни

Назва розділу (теми)	Кількість годин, відведених на:			
	лекції	лабораторні	практичні	СРС
Тема 1. Сутність Big Data та інструментарій для обробки та аналізу	6	4	2	27
Тема 2. Основні методи інтелектуального аналізу даних (Data Mining)	20	24	12	90
Тема 3. Основні технології обробки	6	6	2	41
Разом за семестр:	32	34	16	158

5. Програма навчальної дисципліни¹

5.1 Зміст лекційного курсу

Номер лекції	Перелік тем лекцій, їх анотації	Кількість годин
	Тема 1. Сутність Big Data та інструментарій для обробки та аналізу	6
1	Лекція 1. Big Data і Data Science: сутність та інструментарій. Data Science та Big Data: основні поняття та виклики. Види великих даних. Процес Data Science. Екосистема великих даних та Data Science. Літ.: [1-5]	2
2	Лекція 2. Класифікація методів аналізу великих даних Основні напрямки роботи з великими даними. Джерела Великих даних. Основні класи методів обробки великих даних. Літ.: [1-5]	2
3	Лекція 3. Метадані великих даних Метадані різних типів даних. Метадані у соціальних мережах. Метадані сховищ даних. Робота з метаданими. Літ.: [1-5]	2
	Тема 2. Основні методи інтелектуального аналізу даних (Data Mining)	20
4	Лекція 4. Методи очищення даних Збір даних. Підготовка даних. Очищення даних Літ.: [1-5]	2
5	Лекція 5. OLAP-аналіз Літ.: [1-5]	2
6	Лекція 6. Методи візуалізації великих даних Дослідження даних. Дашборд: приклади і застосування. Інструменти візуалізації даних Літ.: [1-5]	2
7	Лекція 7. Методи групування даних Моделювання/аналіз даних. Метрики. Класифікація і кластеризація даних Літ.: [1-5]	2
8	Лекція 8. Методи групування даних: класифікація даних Класичні технології класифікації. Класифікація методом опорних векторів. Байєсівська класифікація. Класифікація за допомогою методу найближчого сусіда. Класифікація CBR-методом. Алгоритми обмеженого перебору. Класифікація за допомогою штучних нейронних мереж. Класифікація за допомогою дерев рішень. Методи нечіткої логіки. Програмне забезпечення задач класифікації Літ.: [1-5]	2
9	Лекція 9. Методи групування даних: кластеризація даних Класичні технології кластеризації. Ієрархічні методи кластерного аналізу. Неієрархічні методи кластерного аналізу. Нові алгоритми кластерного аналізу. Програмне забезпечення задач класифікації Літ.: [1-5]	2
10	Лекція 10. Методи аналізу та моделювання залежностей. Секвенційний аналіз Загальні положення. Статистичні методи. Методи асоціативних правил. Алгоритм AprioriALL. Алгоритм GSP Літ.: [1-5]	2
11	Лекція 11 Регресійний аналіз Лінійна регресія. Нелінійна регресія. Множинна регресія Літ.: [1-5]	2
12	Лекція 12. Методи аналізу динаміки Основні поняття та визначення. Принципи та етапи аналізу і моделювання динаміки.	2

	Програмне забезпечення задач дослідження динаміки. Аналіз та моделювання тенденцій динаміки. Дослідження сезонності Літ.: [1-5]	
13	Лекція 13. Ансамблі моделей Означення. Види ансамблей. Алгоритм Бегтінг. Алгоритм бустінга. Алгоритм стекінг. Літ.: [1-5]	2
	Тема 3. Основні технології обробки	6
14	Лекція 14. Нейронні мережі та їх застосування при аналізі великих даних Поняття та можливості нейрокомп'ютерних технологій. Нейронні мережі прямого розповсюдження. Метод зворотного розповсюдження помилки. Програмні засоби реалізації нейрокомп'ютерних технологій. Сучасна практика та перспективні напрямки застосування нейротехнологій Літ.: [1-5]	2
15	Лекція 15. Технології Text Mining Інтелектуальний аналіз тексту. Методи інтелектуального аналізу даних (Data Mining). Методи автоматичного вилучення ключових термінів. Застосування технології Text майнінгу Літ.: [1-5]	2
16	Лекція 16 Технології Map Reduce Призначення та області застосування Map Reduce. Основні елементи Map Reduce. Історія розвитку головної технології Big Data. Архітектура Map Reduce. Принцип роботи Map Reduce. Основні переваги моделі Map Reduce. Недоліки і альтернативи Map Reduce. Літ.: [1-5]	2
	Разом за семестр:	32

5.2 Зміст практичних занять

№ п/п	Тема практичного заняття	Кількість годин
1	Застосування методів очищення даних.	2
2	Python: імпорт і попередня обробка даних	2
3	Python: OLAP-аналіз	2
4	Класифікація (Дерево рішень)	2
5	Кластерний аналіз	2
6	Асоціація	2
7	Застосування методів аналізу динаміки.	2
8	Штучний інтелект та нейромережі	2
	Разом за семестр	16

5.3 Зміст лабораторних занять

№ п/п	Тема лабораторного заняття	Кількість годин
1	Застосування методів очищення даних.	4
2	Python: імпорт і попередня обробка даних	4
3	Python: OLAP-аналіз	4
4	Класифікація (Дерево рішень)	4
5	Кластерний аналіз	4
6	Асоціація	4
7	Застосування методів аналізу динаміки.	4
8	Побудова найпростішої нейромережі	4
9	Підсумкове заняття	2
	Разом:	34

5.4 Зміст самостійної (у т.ч. індивідуальної) роботи здобувача вищої освіти

Самостійна робота студентів усіх форм здобуття освіти полягає у систематичному опрацюванні програмного матеріалу з відповідних джерел інформації, підготовці до практичних занять, підготовці до виконання і захисту лабораторних робіт. До послуг студентів сторінка кафедри у Модульному середовищі для навчання, де розміщені Робоча програма дисципліни та необхідні матеріали з її навчально-методичного забезпечення та контролю результатів навчання.

Номер тижня	Вид самостійної роботи	Кількість годин
1-2	Опрацювання лекційного матеріалу. Підготовка до ЛР1 та ПЗ1. Підготовка до захисту ЛР1.	18
3-4	Опрацювання лекційного матеріалу. Підготовка до ЛР2 та ПЗ2. Підготовка до захисту ЛР2.	18
5-6	Опрацювання лекційного матеріалу. Підготовка до ЛР3 та ПЗ3. Підготовка до захисту ЛР3.	18
7-8	Опрацювання лекційного матеріалу. Підготовка до ЛР4 та ПЗ4. Підготовка до захисту ЛР4.	18
9-10	Опрацювання лекційного матеріалу. Підготовка до ЛР5 та ПЗ5. Підготовка до захисту ЛР5.	18
11-12	Опрацювання лекційного матеріалу. Підготовка до ЛР6 та ПЗ6. Підготовка до захисту ЛР6.	18
13-14	Опрацювання лекційного матеріалу. Підготовка до ЛР7 та ПЗ7. Підготовка до захисту ЛР7.	18
15-16	Опрацювання лекційного матеріалу. Підготовка до ЛР8 та ПЗ8. Підготовка до захисту ЛР8.	18
17	Підготовка до ліквідації заборгованості	14
Разом семестр:		158

Примітки: ЛР – лабораторна робота, ПЗ– практичне заняття.

6. Технології та методи навчання

Процес навчання з дисципліни ґрунтується на використанні традиційних та сучасних технологій та методів навчання, зокрема: лекції (з використанням методів візуалізації, проблемного й інтерактивного навчання, мотиваційних прийомів, інформаційно-комунікаційних технологій); практичні заняття (з використанням інструктування, демонстрування, розв'язування типових і прикладних задач, аналізу кейсів, ситуативних завдань, елементів дискусії тощо); лабораторні заняття (з використанням методів комп'ютерного моделювання, методів проєктної діяльності, пояснення, дискусія тощо); самостійна робота (опрацювання теоретичного матеріалу, підготовка до виконання практичних робіт, поточного та підсумкового контролю), з використанням інформаційно-комп'ютерних технологій та технологій дистанційного навчання; лабораторні заняття: з використанням методів комп'ютерного моделювання, методів проєктної діяльності, пояснення, дискусія тощо.

7. Методи контролю

Поточний контроль здійснюється під час аудиторних лабораторних занять, а також у дні проведення контрольних заходів, встановлених робочою програмою і графіком освітнього процесу, в т.ч. з використанням Модульного середовища для навчання. При цьому використовуються такі методи поточного контролю:

- усне опитування перед допуском до лабораторного заняття;
- оцінювання результатів захисту лабораторних робіт;
- підсумкова контрольна робота

Підсумкова семестрова оцінка виставляється за результатами поточного контролю. Здобувач вищої освіти, який набрав з будь-якого виду навчальної роботи, суму балів нижчу за 60 відсотків від максимального балу, вважається таким, який **має** академічну заборгованість. Ліквідація академічної заборгованості із семестрового контролю здійснюється у період екзаменаційної сесії або за графіком, встановленим деканатом відповідно до «Положення про контроль і оцінювання результатів навчання здобувачів вищої освіти у ХНУ».

8. Політика дисципліни

Політика навчальної дисципліни загалом визначається системою вимог до здобувача вищої освіти, що передбачені чинними положеннями Університету про організацію і навчально-методичне забезпечення освітнього процесу. Зокрема, проходження інструктажу з техніки безпеки; відвідування занять з дисципліни є обов'язковим. За об'єктивних причин (підтверджених документально) теоретичне навчання за погодженням із лектором може відбуватись в он-лайн режимі. Успішне опанування дисципліни і формування фахових компетентностей і програмних результатів навчання передбачає необхідність підготовки до лабораторних занять (вивчення теоретичного матеріалу з теми, підготовку до усного опитування для допуску до заняття (наведені у Методичних рекомендаціях до лабораторних занять)), активно працювати на занятті, брати участь у дискусіях щодо прийнятих рішень при виконанні здобувачами лабораторних робіт тощо.

Здобувачі вищої освіти мають дотримуватися встановлених термінів виконання всіх видів навчальної роботи відповідно до робочої програми навчальної дисципліни. Термін захисту лабораторної роботи вважається своєчасним, якщо студент захистив її на поточному або наступному після виконання роботи занятті. Пропущене лабораторне заняття студент зобов'язаний відпрацювати у встановлений викладачем термін, але не пізніше, ніж за два тижні до кінця теоретичних занять у семестрі.

Здобувач вищої освіти, виконуючи самостійну роботу або індивідуальну роботу з дисципліни, має дотримуватися політики доброчесності (заборонені списування, плагіат (в т.ч. із використанням мобільних девайсів)). У разі виявлення порушення політики академічної доброчесності в будь-яких видах навчальної роботи здобувач вищої освіти отримує незадовільну оцінку і має повторно виконати завдання з відповідної теми (виду роботи), що передбачені робочою програмою. Будь-які форми порушення академічної доброчесності **не допускаються**.

У межах вивчення навчальної дисципліни здобувачам вищої освіти передбачено визнання і зарахування результатів навчання, набутих шляхом неформальної освіти, що розміщені на доступних платформах (**за наявності такого переліку, доцільно вказати рекомендовані курси**), які сприяють формуванню компетентностей і поглибленню результатів навчання, визначених робочою програмою дисципліни, або забезпечують вивчення відповідної теми та/або виду робіт з програми навчальної дисципліни (детальніше у Положенні про порядок визнання та зарахування результатів навчання здобувачів вищої освіти у ХНУ).

9. Оцінювання результатів навчання студентів у семестрі

Оцінювання академічних досягнень здобувача вищої освіти здійснюється відповідно до «Положення про контроль і оцінювання результатів навчання здобувачів вищої освіти у ХНУ». При поточному оцінюванні виконаної здобувачем роботи з кожної структурної одиниці і отриманих ним результатів викладач виставляє йому певну кількість балів із встановлених Робочою програмою для цього виду роботи. При цьому кожна структурна одиниця навчальної роботи може бути зарахована, якщо здобувач набрав не менше 60 відсотків (мінімальний рівень для позитивної оцінки) від максимально можливої суми балів, призначеної структурній одиниці.

При оцінюванні результатів навчання здобувачів вищої освіти з будь-якого виду навчальної роботи (структурної одиниці) рекомендується використовувати наведені нижче узагальнені критерії:

Таблиця – Критерії оцінювання навчальних досягнень здобувача вищої освіти

Оцінка та рівень досягнення здобувачем запланованих ПРН та сформованих компетентностей	Узагальнений зміст критерія оцінювання
Відмінно (високий)	Здобувач вищої освіти глибоко і у повному обсязі опанував зміст навчального матеріалу, легко в ньому орієнтується і вміло використовує понятійний апарат; уміє пов'язувати теорію з практикою, вирішувати практичні завдання, впевнено висловлювати і обґрунтовувати свої судження. Відмінна оцінка передбачає логічний виклад відповіді мовою викладання (в усній або у письмовій формі), демонструє якісне оформлення роботи і володіння спеціальними приладами та інструментами, прикладними програмами. Здобувач не вагається при видозміні запитання, вміє робити детальні та узагальнюючі висновки, демонструє практичні навички з вирішення фахових завдань. При відповіді допустив дві–три несуттєві <i>похибки</i> .
Добре (середній)	Здобувач вищої освіти виявив повне засвоєння навчального матеріалу, володіє понятійним апаратом, орієнтується у вивченому матеріалі; свідомо використовує теоретичні знання для вирішення практичних задач; виклад відповіді грамотний, але у змісті і формі відповіді можуть мати місце окремі неточності, нечіткі формулювання правил, закономірностей тощо. Відповідь здобувача вищої освіти будується на основі самостійного мислення. Здобувач вищої освіти у відповіді допустив дві–три <i>несуттєві помилки</i> .
Задовільно (достатній)	Здобувач вищої освіти виявив знання основного програмного матеріалу в обсязі, необхідному для подальшого навчання та практичної діяльності за професією, справляється з виконанням практичних завдань, передбачених програмою. Як правило, відповідь здобувача вищої освіти будується на рівні репродуктивного мислення, здобувач вищої освіти має слабкі знання структури навчальної дисципліни, допускає неточності і <i>суттєві помилки</i> у відповіді, вагається при відповіді на видозмінене запитання. Разом з тим, набув навичок, необхідних для виконання нескладних практичних завдань, які відповідають мінімальним критеріям оцінювання і володіє знаннями, що дозволяють йому під керівництвом викладача усунути неточності у відповіді.
Незадовільно (недостатній)	Здобувач вищої освіти виявив розрізнені, безсистемні знання, не вміє виділяти головне і другорядне, допускається помилок у визначенні понять, перекручує їх зміст, хаотично і невпевнено викладає матеріал, не може використовувати знання при вирішенні практичних завдань. Як правило, оцінка «незадовільно» виставляється здобувачеві вищої освіти, який не може продовжити навчання без додаткової роботи з вивчення навчальної дисципліни.

Структурування дисципліни за видами навчальної роботи і оцінювання результатів навчання студентів *денної* форми здобуття освіти у семестрі

Аудиторна робота								Контрольні заходи	Семестровий контроль
Лабораторні заняття №:								Контрольна робота	Залік
1*	2	3	4	5	6	7	8	КР	
Кількість балів за кожний вид навчальної роботи (мінімум-максимум)									
6-10	6-10	6-10	6-10	6-10	6-10	6-10	6-10	12-20	За рейтингом
48-80								12-20	60-100**

Примітки: *За набрану з будь-якого виду навчальної роботи з дисципліни кількість балів, нижче встановленого мінімуму, здобувач отримує незадовільну оцінку і має її перездати у встановлений викладачем (деканом) термін. Інституційна оцінка встановлюється відповідно до таблиці «Співвідношення інституційної шкали оцінювання і шкали оцінювання ЄКТС».

Оцінювання результатів захисту лабораторної роботи

Виконана й оформлена відповідно до встановлених Методичними рекомендаціями вимог лабораторна робота комплексно оцінюється викладачем при її захисті з урахуванням таких критеріїв: самостійність та правильність

виконання; повнота відповіді та знання методів аналізу великих даних, вміння працювати з технічною документацією та відкритими наборами даних, вміння користуватись бібліотеками для аналізу й обробки **даних** (Pandas, NumPy, Scikit-learn, TensorFlow).

Результат виконання і захисту здобувачем вищої освіти кожної лабораторної роботи оцінюється відповідно до таблиці Критеріїв оцінювання навчальних досягнень здобувача вищої освіти.

У випадку виявлення здобувачем рівня знань, нижчого ніж 60 відсотків від максимального балу, встановленого Робочою програмою для кожної структурної одиниці, лабораторна робота йому **не зараховується** і для її захисту він має детальніше опрацювати матеріал з теми роботи, методику її виконання, виправити грубі помилки та повторно вийти на її захист у призначений для цього викладачем час.

Оцінювання контрольної роботи

Контрольна робота, передбачена Робочою програмою, складається із 4 питань, кожне з яких є рівнозначним. Питання для контрольної роботи обираються із запропонованих в пункті 10 Робочої програми. Відповідь на кожне питання оцінюється від 3 до 5 балів згідно Критеріїв оцінювання, наведених вище. За всю контрольну роботу здобувач може отримати від 12 до 20 балів.

Здобувач виконує контрольну роботу письмово на передостанньому занятті. При отриманні негативної оцінки (від 0 до 11 балів) контрольну слід перездати на **останньому аудиторному занятті** з дисципліни.

Розподіл балів при оцінюванні завдань контрольної роботи

Номер питання	1	2	3	4
Кількість балів (мінімум-максимум)	3-5	3-5	3-5	3-5
Кількість отриманих балів (всього)	12-20			

Підсумкова семестрова оцінка за інституційною шкалою і шкалою ЄКТС визначається в автоматизованому режимі після внесення викладачем результатів оцінювання у балах з усіх видів навчальної роботи до електронного журналу. Співвідношення інституційної шкали оцінювання і шкали оцінювання ЄКТС наведені нижче у таблиці «Співвідношення».

Семестровий залік виставляється на останньому занятті за умови якщо загальна сума балів, яку накопичив здобувач з дисципліни (іншого освітнього компонента) за результатами **поточного** контролю, знаходиться у межах від 60 до 100 балів. При цьому за інституційною шкалою ставиться оцінка **«зараховано»**, а за шкалою ЄКТС – буквене позначення оцінки, що відповідає набраній студентом сумі балів відповідно до таблиці Співвідношення. Присутність здобувача у цьому випадку не є обов'язковою.

Таблиця – Співвідношення інституційної шкали оцінювання і шкали оцінювання ЄКТС

Оцінка ЄКТС	Рейтингова шкала балів	Інституційна оцінка (рівень досягнення здобувачем вищої освіти запланованих результатів навчання з навчальної дисципліни)	
		Залік	Іспит/диференційований залік
A	90-100	Зараховано	Відмінно/Excellent – високий рівень досягнення запланованих результатів навчання з навчальної дисципліни, що свідчить про безумовну готовність здобувача до подальшого навчання та/або професійної діяльності за фахом
B	83-89		Добре/Good – середній (максимально достатній) рівень досягнення запланованих результатів навчання з навчальної дисципліни та готовності до подальшого навчання та/або професійної діяльності за фахом
C	73-82		
D	66-72		
E	60-65		Задовільно/Satisfactory – Наявні мінімально достатні для подальшого навчання та/або професійної діяльності за фахом результати навчання з навчальної дисципліни
FX	40-59	Незараховано	Незадовільно/Fail – Низка запланованих результатів навчання з навчальної дисципліни відсутня. Рівень набутих результатів навчання є недостатнім для подальшого навчання та/або професійної діяльності за фахом
F	0-39		Незадовільно/Fail – Результати навчання відсутні

10. Питання для самоконтролю результатів навчання

1. Що таке Pandas? Призначення та застосування.
2. Назвіть основні класи методів обробки великих даних. Дайте стислу характеристику.
3. Суть Text Mining та його завдання.
4. Суть Map Reduce та його завдання.
5. Дайте визначення інтелектуального аналізу даних.
6. Що таке розвідувальний аналіз.
7. Що таке якість даних?
8. Які цілі підготовки даних до аналізу?
9. Який атрибут називається цільовим?
10. Що таке значимий і незначимий атрибут?

11. Що таке відбір атрибутів?
12. У чому полягає завдання класифікації? Наведіть практичний приклад.
13. Що таке навчання з учителем і без учителя?
14. Завдання класифікації є описовим або прогнозуючим і чому?
15. Навіщо потрібна навчальна і тестова вибірки для вирішення завдання класифікації?
16. Які існують підходи для поділу вихідної вибірки на навчальну і тестову?
17. Метод класифікації 0R.
18. Метод класифікації за одним правилом 1R.
19. Метод класифікації PRISM.
20. Наївна Баєсова класифікація NaiveBayes.
21. Метод побудови дерев рішень CART
22. Метод побудови дерев рішень Id3 та C4.5.
23. Метод опорних векторів SMO.
24. Метод k найближчих сусідів kNN.
25. Як оцінити якість побудованої моделі класифікації?
26. Що таке матриця помилок? Як її інтерпретувати?
27. Що означають параметри чутливість, специфічність, точність? Як їх розрахувати?
28. Що таке параметр Каппа? Що він показує?
29. Що таке аналіз витрати-вигоди?
30. Як порівняти роботу двох класифікаторів?
31. Що таке ансамблі класифікаторів і адаптивний бустінг? Для чого вони застосовуються?
32. У чому полягає завдання регресії? Наведіть приклад.
33. Чим завдання регресії схоже і чим відрізняється від завдання класифікації?
34. Завдання регресії є описовим або прогнозуючим і чому?
35. Метод лінійної регресії.
36. Метод опорних векторів для вирішення завдання регресії.
37. Метод найближчих сусідів для вирішення завдання регресії.
38. Як оцінити якість побудованої моделі для завдання регресії?
39. Прогнозування часових рядів.
40. У чому полягає завдання кластеризації? Наведіть практичний приклад.
41. Що таке навчання з учителем і без учителя? До якого типу відноситься завдання кластеризації?
42. Завдання кластеризації є описовим або прогнозуючим і чому?
43. Чим визначається «схожість» об'єктів при вирішенні задачі кластеризації?
44. Що таке однорівнева і ієрархічна кластеризація?
45. Що таке чітка і нечітка кластеризація?
46. Які є підходи до розрахунку відстані між кластерами?
47. Алгоритми ієрархічна кластеризація?
48. Метод кластеризації k-середніх.
49. Метод нечіткої кластеризації fuzzy c-means.
50. Метод кластеризації k-медоїд.
51. Ієрархічні методи кластеризації.
52. Імовірнісний метод кластеризації
53. Метод кластеризації COBWEB.
54. Метод кластеризації заснований на щільності розташування об'єктів DBSCAN.
55. Як оцінити якість побудованої моделі для завдання кластеризації?
56. У чому полягає завдання пошуку асоціативних правил? Наведіть практичний приклад.
57. Що таке частий набір?
58. Що таке сильне асоціативне правило?
59. З яких двох кроків складається пошук асоціативних правил?
60. У чому полягає принцип Apriori?
61. Як формуються правила зі знайдених частих наборів?
62. Опишіть алгоритм Apriori
63. Що означають параметри support, confidence, lift, conviction, які застосовуються в алгоритмі Apriori?
64. Опишіть алгоритм ECLAT
65. Опишіть алгоритм FPGrowth
66. На які дві групи підрозділяються методи Data Mining за принципом роботи з вихідними навчальними даними?
67. Чи правильне таке формулювання: "Асоціація є часткою випадкової послідовності з часовим лагом, що дорівнює нулю"?
68. Що перебуває в основі так званої інформаційної піраміди?
69. На які категорії підрозділяються задачі Data Mining, залежно від моделей, що використовуються?
70. Для вирішення яких задач застосовуються ієрархічні алгоритми?
71. У чому подібність задач класифікації й прогнозування?
72. Який з напрямів включає виявлення закономірностей у діях користувача web-вузла або їхньої групи?
73. Які фактори впливають на результат класифікації в наївнобайєсовському підході?
74. Які переваги характерні для використання байєсовських мереж?
75. Які задачі вирішуються за допомогою методу опорних векторів?
76. Що являє собою група синапсів нейрону?

77. Визначите головну функцію штучного нейрона.
78. У чому полягає навчання мереж, що самоорганізуються?
79. Які існують традиційні методи візуалізації?
80. Назвіть характеристики однієї з основних тенденцій в області візуалізації.
81. Що означає інтегрованість сховища даних?
82. Назвіть основні концепції сховища даних.
83. Якщо набір даних упорядкований та у ньому присутня сезонна або циклічна компонента, то яку мінімальну кількість даних необхідно мати для можливості аналізу?
84. Що повинна робити якісна програма очищення даних?
85. Які характеристики повинна мати якісна програма очищення даних?
86. На якому етапі перетинається робота фахівця предметної області й фахівця з видобутку даних?
87. Назвіть характеристики, властиві SAS Enterprise Miner.
88. Рішення яких задач передбачають алгоритми аналізу даних в PolyAnalyst?
89. Чи існує необхідність тимчасового або постійного копіювання даних для аналізу в системі KXEN?
90. Який компонент KXEN дозволяє виявити природні групи (кластери) у наборі даних?
91. Назвіть слабкі сторони використання готового програмного забезпечення Data Mining.

11. Навчально-методичне забезпечення

Освітній процес з дисципліни «Аналітика великих даних» повністю і в достатній кількості забезпечений необхідною навчально-методичною літературою. В модульному середовищі для навчання подано матеріали лекцій, практичних та лабораторних робіт

12. Матеріально-технічне та програмне забезпечення дисципліни (за потреби)

Інформаційна та комп'ютерна підтримка: ПК, смартфон або інший мобільний пристрій, проєктор. Програмне забезпечення: програми Microsoft Office або аналогічні, доступ до мережі Інтернет, робота з презентаціями,

Спеціалізоване ПЗ: Anaconda Jupyter Notebook / JupyterLab, PySpark – є безкоштовними для особистого та навчального використання або open-source

13. Рекомендована література:

1. Hans Weber. Big Data and Artificial Intelligence: Complete Guide to Data Science, AI, Big Data and Machine Learning. Режим доступу: <https://www.amazon.com/dp/B08FP9YZXN?tag=uuid10-20>.
2. Jay M. Patel. Getting Structured Data from the Internet: Running Web Crawlers/Scrapers on a Big Data Production Scale 1st ed. Edition. Режим доступу: <https://www.amazon.com/dp/1484265750?tag=uuid10-20>.
3. Dursun Delen. Predictive Analytics: Data Mining, Machine Learning and Data Science for Practitioners, 2nd Edition (FT Press Analytics) 2nd Edition. Режим доступу: <https://www.amazon.com/dp/0136738516?tag=uuid10-20>
4. Dinesh Sachdev. Thinking Big: Developers Guide for Big Data Engineering & Analytics. Режим доступу: <https://www.amazon.com/dp/B08TQD9GWQ?tag=uuid10-20>
5. Талах М.В., Т-16 Технології обробки Big Data. Навчальний посібник/ Талах М.В., – Чернівці: Чернівецький нац.ун-т, 2024. – 454 с.

Додаткова література

1. Технології оброблення великих даних: конспект лекцій з дисципліни «Технології оброблення великих даних» [Електронний ресурс] : навч. посіб. для студ. спеціальності 121 «Інженерія програмного забезпечення» (освітня програма «Інженерія програмного забезпечення мультимедійних та інформаційно-пошукових систем»)/ Л.М. Олещенко; КПІ ім. Ігоря Сікорського. – Електронні текстові дані (1 файл: 5,55 Мбайт). – Київ: КПІ ім. Ігоря Сікорського, 2021. – 227 с
2. Д.В. Ланде, І.Ю. Субач, А.Я. Гладун. Оброблення надвеликих масивів даних (Big Data) : навчальний посібник. Київ 2021. – 168 с.
3. Черняк О. І. Інтелектуальний аналіз даних : підручник / О. І. Черняк, П. В. Захарченко. – К.: Знання, 2014. – 599 с.

Офіційна документація та довідники

Apache Hadoop – <https://hadoop.apache.org/>
 Apache Spark – <https://spark.apache.org/>
 TensorFlow для аналізу даних – <https://www.tensorflow.org/>
 Pandas & NumPy – <https://pandas.pydata.org/> | <https://numpy.org/>
 Plotly / Dash / Power BI для візуалізації – <https://plotly.com/python/>

YouTube-канали українською та англійською

ITVDN – відеоуроки українською та російською: <https://itvdn.com/ua>
Prometheus – безкоштовні лекції: <https://www.youtube.com/@PrometheusUkraine>
Corey Schafer (англ.) – курси з Python, Pandas, NumPy: <https://www.youtube.com/c/Coreyms>
freeCodeCamp (англ.) – повний курс [Data Science with Python](#)

Інтерактивні ресурси для тренувань

<https://bigdatauniversity.com/> (IBM Big Data University)
<https://analytics.google.com/> (Google Analytics Academy)
<https://www.kaggle.com/datasets> (реальні великі набори даних)
<https://cloudskillsboost.google/> (BigQuery, Spark, Dataflow практикум)

14. Інформаційні ресурси

1. Модульне середовище для навчання. URL : <https://msn.khmn.edu.ua/course/view.php?id=8241>
2. Електронна бібліотека ХНУ. URL: <http://library.khmn.edu.ua/>
3. Інституційний репозитарій ХНУ. URL : <https://elar.khmn.edu.ua/home>

АНАЛІТИКА ВЕЛИКИХ ДАНИХ

Тип дисципліни	Вибіркова
Рівень вищої освіти	Другий (магістерський)
Мова викладання	Українська
Семестр	—
Кількість призначених кредитів ЄКТС	8,0
Форми здобуття освіти, для яких викладається дисципліна	Очна (денна)

Результати навчання. Після вивчення дисципліни студент повинен: знати: основні поняття, концепції та архітектури Big Data; методи збору, обробки, аналізу та візуалізації великих обсягів даних; інструменти машинного навчання та прогнозувальної аналітики; володіти: базовими навичками роботи з сучасними аналітичними платформами та бібліотеками; інструментами побудови моделей, аналізу даних і створення інтерактивних звітів; вміти: застосовувати методи аналізу та прогнозування даних для вирішення практичних завдань; обґрунтовувати вибір інструментів і технологій для реалізації аналітичних проєктів; ефективно працювати з документацією, проводити дослідження та аналізувати отримані результати; демонструвати: ініціативність, здатність до самостійного навчання та критичного мислення; навички комунікації та роботи в команді; готовність приймати відповідальні рішення під час розв'язання складних аналітичних завдань.

Зміст навчальної дисципліни. Data Science та Big Data: основні поняття та виклики. Види великих даних. Процес Data Science. Екосистема великих даних та Data Science. Основні напрямки роботи з великими даними. Джерела Великих даних. Основні класифікації методів обробки великих даних. Метадані різних типів даних. Метадані у соціальних мережах. Метадані сховищ даних. Робота з метаданими. Збір даних. Підготовка даних. Очищення даних. OLAP-аналіз. Дослідження даних. Дашборд: приклади і застосування. Інструменти візуалізації даних. Моделювання/аналіз даних. Метрики. Класифікація і кластеризація даних. Класичні технології класифікації. Класифікація методом опорних векторів. Байєсівська класифікація. Класифікація за допомогою методу найближчого сусіда. Класифікація CBR-методом. Алгоритми обмеженого перебору. Класифікація за допомогою штучних нейронних мереж. Класифікація за допомогою дерев рішень. Методи нечіткої логіки. Програмне забезпечення задач класифікації. Класичні технології кластеризації. Ієрархічні методи кластерного аналізу. Неієрархічні методи кластерного аналізу. Нові алгоритми кластерного аналізу. Програмне забезпечення задач класифікації. Методи асоціативних правил. Алгоритм AprioriALL. Алгоритм GSP. Лінійна регресія. Нелінійна регресія. Множинна регресія. Принципи та етапи аналізу і моделювання динаміки. Програмне забезпечення задач дослідження динаміки. Аналіз та моделювання тенденцій динаміки. Дослідження сезонності. Види ансамблей. Алгоритм Беггінг. Алгоритм бустинга. Алгоритм стекінг. Поняття та можливості нейрокомп'ютерних технологій. Нейронні мережі прямого розповсюдження. Метод зворотного розповсюдження помилки. Програмні засоби реалізації нейрокомп'ютерних технологій. Сучасна практика та перспективні напрямки застосування нейротехнологій. Інтелектуальний аналіз тексту. Методи інтелектуального аналізу даних (Data Mining). Методи автоматичного вилучення ключових термінів. Застосування технології Text майнінгу. Призначення та області застосування Map Reduce. Основні елементи Map Reduce. Історія розвитку головної технології Big Data. Архітектура Map Reduce. Принцип роботи Map Reduce. Основні переваги моделі Map Reduce. Недоліки і альтернативи Map Reduce

Запланована навчальна діяльність: Мінімальний обсяг навчальних занять в одному кредиті ЄКТС навчальної дисципліни для другого (магістерського) рівня вищої освіти за денною формою здобуття освіти становить 8 годин на 1 кредит ЄКТС.

Форми (методи) навчання: лекції (з використанням методів візуалізації, проблемного й інтерактивного навчання, мотиваційних прийомів, інформаційно-комунікаційних технологій); практичні заняття (з використанням інструктування, демонстрування, розв'язування типових і прикладних задач, аналізу кейсів, ситуаційних завдань, елементів дискусії тощо); лабораторні заняття (з використанням методів комп'ютерного моделювання, методів проєктної діяльності, пояснення, дискусія тощо); самостійна робота (опрацювання теоретичного матеріалу, підготовка до виконання практичних робіт, поточного контролю), з використанням інформаційно-комп'ютерних технологій та технологій дистанційного навчання; лабораторні заняття: з використанням методів комп'ютерного моделювання, методів проєктної діяльності, пояснення, дискусія тощо

Форми оцінювання результатів навчання: оцінювання лабораторних та контрольних робіт.

Вид семестрового контролю: залік.

Навчальні ресурси:

1. Hans Weber. Big Data and Artificial Intelligence: Complete Guide to Data Science, AI, Big Data and Machine Learning. Режим доступу: <https://www.amazon.com/dp/B08FP9YZXN?tag=uid10-20>.
2. Jay M. Patel. Getting Structured Data from the Internet: Running Web Crawlers/Scrapers on a Big Data Production Scale 1st ed. Edition. Режим доступу: <https://www.amazon.com/dp/1484265750?tag=uid10-20>.
3. Dursun Delen. Predictive Analytics: Data Mining, Machine Learning and Data Science for Practitioners, 2nd Edition (FT Press Analytics) 2nd Edition. Режим доступу: <https://www.amazon.com/dp/0136738516?tag=uid10-20>
4. Dinesh Sachdev. Thinking Big: Developers Guide for Big Data Engineering & Analytics. Режим доступу: <https://www.amazon.com/dp/B08TQD9GWQ?tag=uid10-20>
5. Талах М.В., Т-16 Технології обробки Big Data. Навчальний посібник/ Талах М.В., – Чернівці: Чернівецький нац.ун-т, 2024. – 454 с.
6. Модульне середовище для навчання. Доступ до ресурсу: <https://msn.khmn.edu.ua/course/view.php?id=8241>
7. Електронна бібліотека ХНУ. Доступ до ресурсу: <http://library.khmn.edu.ua/>

Викладачі: канд. фіз.-мат.наук, доцент, Кисіль Тетяна Миколаївна