

ХМЕЛЬНИЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ



ЗАТВЕРДЖУЮ

Дека́н факультету інформаційних технологій

проф. Тетяна ГОВОРУЩЕНКО

«05» Вересня 2024 р.

РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Методи та засоби аналізу великих та малих даних

Назва дисципліни

Галузь знань – 12 Інформаційні технології

Спеціальність – 122 Комп'ютерні науки

Рівень вищої освіти – Другий магістерський

Освітньо-професійна програма – Комп'ютерні науки

Обсяг дисципліни – 5 кредитів ЄКТС, **Шифр дисципліни** – ОПП.04

Мова навчання – українська

Статус дисципліни – обов'язкова (цикл професійної підготовки)



Факультет - Інформаційних технологій

Кафедра – Комп'ютерних наук

Курс	Семестр	Обсяг дисципліни		Кількість годин						Форма семестрового контролю		
				Аудиторні заняття				Індивідуальна робота студента	Самостійна робота, у т.ч. ІРС	Курсовий проект	Курсова робота	Залік
		Разом	Лекції	Лабораторні роботи	Практичні заняття							
1	1	5	150	51	34	17			99			+
Разом		5	150	51	34	17			99			1

Робоча програма складена на основі освітньо-професійної програми «Комп'ютерні науки» та стандарту вищої освіти зі спеціальності 122 – Комп'ютерні науки другого (магістерського) рівня вищої освіти.

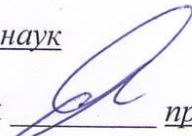
Програма складена

к.т.н., доц. Руслан БАГРІЙ

асистент каф. КН Олександр ОВЧАРУК

Схвалено на засіданні кафедри комп'ютерних наук

Протокол від 30 серпня 2024 р., №1. Зав. Кафедри  проф. Олександр БАРМАК

Робоча програма розглянута та схвалена вченою радою факультету інформаційних технологій

Голова вченої ради факультету  проф. Тетяна ГОВОРУЩЕНКО

Хмельницький 2024

Пояснювальна записка

Дисципліна «Методи та засоби аналізу великих та малих даних» викладається у напрямку підготовки спеціальності 122 – «Комп'ютерні науки» входить до складу обов'язкової частини навчального плану підготовки магістерського рівня вищої освіти. Викладається відповідно до робочого навчального плану підготовки магістрів.

Забезпечення візуалізації первинних даних з метою їх подальшої обробки, включно, до звичайних статистичних методів та аналогічних методик, а також новітніми засобами машинного навчання, штучного інтелекту тощо на сьогодні є актуальним науково-прикладним завданням яке є невід'ємною складовою функціонування сучасного підприємства.

Слід відзначити, що сучасні підприємства та організації повинні аналізувати як надвеликі обсяги даних так і малі об'єми даних для забезпечення своєї конкурентоспроможності. Це стосується як комерційних підприємств, які отримують великі неструктуровані дані щодо переваг та запитів своїх клієнтів, так й комунальних підприємств, громадських організацій, що мають обробляти значні дані з засобів Інтернет-речей, наприклад, лічильників тепла, звітність про рух транспортних засобів, наукових установ тощо. Однак, отримання таких даних, їх первинна обробка – це не єдині завдання, що надають конкурентоспроможності у сучасних умовах. На сьогодні існує необхідність у швидкому реагуванні на отримані великі масиви даних чи їх зміну для формування керівних рішень. Тому, дисципліна «Методи та засоби аналізу великих та малих даних» є перспективною та актуальною, оскільки розглядає рішення відбиття, консолідації та візуалізації даних, перетворюючи їх в ефективний засіб ведення бізнесу.

Передбачені програмою лабораторні заняття служать для закріплення теоретичних знань і набуття практичних навичок застосування методів та інструментів збору, обробки, аналізу, моніторингу та інтерпретації великих масивів даних, що виникають під час здійснення бізнес-процесів підприємства.

Пререквізити – вихідна.

Кореквізити – розподілені високопродуктивні інформаційно-обчислювальні системи та технології, кваліфікаційна робота.

Відповідно до Стандарту вищої освіти із зазначеної спеціальності та освітньої програми дисципліна має забезпечити:

Компетентності:

ЗК07. Здатність генерувати нові ідеї (креативність).

ФК01. Усвідомлення теоретичних засад комп'ютерних наук.

ФК02. Здатність формалізувати предметну область певного проєкту у вигляді відповідної інформаційної моделі.

ФК03. Здатність використовувати математичні методи для аналізу формалізованих моделей предметної області.

ФК04. Здатність збирати і аналізувати дані (включно з великими), для забезпечення якості прийняття проєктних рішень.

ФК06. Здатність застосовувати існуючі і розробляти нові алгоритми розв'язування задач у галузі комп'ютерних наук.

ФК07. Здатність розробляти програмне забезпечення відповідно до сформульованих вимог з урахуванням наявних ресурсів та обмежень.

ФК09. Здатність розробляти та адмініструвати бази даних та знань.

Програмні результати навчання:

ПРН02. Мати спеціалізовані уміння/навички розв'язання проблем комп'ютерних наук, необхідні для проведення досліджень та/або провадження інноваційної діяльності з метою розвитку нових знань та процедур.

ПРН06. Розробляти концептуальну модель інформаційної або комп'ютерної системи.

ПРН07. Розробляти та застосовувати математичні методи для аналізу інформаційних моделей.

ПРН08. Розробляти математичні моделі та методи аналізу даних (включно з великим).

ПРН09. Розробляти алгоритмічне та програмне забезпечення для аналізу даних

(включно з великими).

ПРН11. Створювати нові алгоритми розв'язування задач у сфері комп'ютерних наук, оцінювати їх ефективність та обмеження на їх застосування

ПРН12. Проектувати та супроводжувати бази даних та знань.

ПРН16. Виконувати дослідження у сфері комп'ютерних наук.

ПРН18. Збирати, формалізувати, систематизувати і аналізувати потреби та вимоги до інформаційної або комп'ютерної системи, що розробляється, експлуатується чи супроводжується

Мета дисципліни – здатність забезпечити організацію обчислювальних процесів витягнення та інтерпретації прихованих знань з використанням пакетів прикладних програм різних типів (tensor toolbox, Nadoor, Map reduce та ін.); здатність до розробки інтерфейсного програмного забезпечення для поєднання джерела великих даних (ВД) з відповідним програмним середовищем.

Предмет дисципліни. Методи, технології та програмні засоби для аналізу первинних даних і візуалізації їх результатів для підтримки рішень.

Завдання дисципліни – формування компетентностей та забезпечення програмних результатів навчання для аналізу, дослідженню та візуалізації великих масивів даних із використанням сучасних інформаційних технологій.

Результати навчання. Після вивчення дисципліни студент має: досконало знати професійну термінологією та сучасний стан розвитку Business Intelligence (BI), знати методи та технології застосування у рішенні завдань обробки великих даних; проводити аналіз, дослідження та візуалізацію великих масивів даних із використанням сучасних інформаційних технологій; аналізувати та обирати оптимальні рішення під час проектування розподілених баз даних, систем збору великих даних, застосовувати засоби Business Intelligence для обробки та візуалізації бізнес-даних та бізнес-аналітики; адаптовувати методи та технології Business Intelligence для роботи з малими обсягами даних; використовувати інструментальні засоби Business Intelligence для аналізу, моніторингу та візуалізації бізнес-даних та побудови бізнес-аналітики.

Структура залікових кредитів дисципліни

Назва теми	Кількість годин, відведених на:		
	лекції	лабор. роботи	СРС
Розділ 1. Основні положення концепції Big Data.	4	4	20
Розділ 2. Обґрунтування вибору та особливості застосування інструментальних засобів Бі	12	4	25
Розділ 3. Основи проектування ефективних рішень у галузі Бі.	18	9	54
Разом за семестр:	34	17	99

Програма навчальної дисципліни Зміст лекційного курсу*

Номер лекції	Перелік тем лекцій, їх анотації	Кількість годин
	<i>Перший семестр</i>	
1	Вступ до великих даних. Огляд курсу. Поняття обробки, здобуття даних. Хеш-функція. Індеси Знайомство з технологіями Hadoop та Spark. Літ.: [1-3]	2
2	Застосування алгоритмів Map Reduce. Розподілені файлові системи. Map Reduce (Згортка-Відображення). Алгоритми, побудовані на MapReduce. Модель комунікативної вартості. Вступ до PySpark. Основи машинного обчислення засобами PySpark. Літ.: [4]	2
3	Основні терміни та визначення. Особливості технологій глибинного аналізу даних (Data Mining), сховищ даних (Data Warehousing). Літ.: [7]	2
4	Основні терміни та визначення. Бізнес-аналітики (Business Analytics) та місце Бі у сучасних бізнес-процесах підприємств та організацій. Класифікація технологій Бі. Літ.: [7]	2
5	Визначення джерел та засобів отримання первинних даних. Основи технологій моніторингу, реєстрації та обробки великих даних (Big Data). Літ.: [5]	2
6	Визначення джерел та засобів отримання первинних даних. Особливості технологій Інтернет-речей (IoT) та засоби машинного навчання (Machine Learning). Особливості застосування технологій Бі у рішенні завдань рівня розумного будинку, міста чи підприємства. Поняття «Індустрія 4.0». Літ.: [6]	2
7	Інструментальні засоби Бі. Технології Microsoft Power BI. Приклад застосування Microsoft Power BI для візуалізації бізнес-даних та створення звітів. Літ.: [6]	2
8	Інструментальні засоби Бі. Особливості технологій хмарних обчислень у рішенні завдань Бі. Літ.: [8]	2
9	Технології СУБД та архітектура проектних рішень на основі застосування інструментальних засобів Бі. Літ.: [3] [5]	2
10	Розподілені обчислювальні системи. Моніторинг стану розподілених	2

	обчислювальних систем на базі відкритих програмних засобів Nagios, Icinga та Zabbix. Технології Grafana. Агрегація даних засобами Бі. Літ.: [3] [5]	
11	Застосування технологій хмарних обчислень. Особливості автоматизації отримання замовником ресурсів хмарних систем. Автоматизація процесів розгортання та підтримки Бі. Поняття та особливості технологій застосування чат-ботів. Літ.: [3] [5]	2
12	Особливості супроводження та підтримки Бі-систем. Розгортання та супроводження розподілених систем. Сучасні засоби веб-технологій у рішенні завдань Бі. Перспективи розвитку систем та технологій Бі. Літ.: [5]	2
13	Платформа Elasticsearch для організації розподіленого оброблення великих обсягів даних у межах індексування та пошуку будь-яких типів документів. Мультитенантні веб- застосунки. Літ.: [12]	2
14	Засоби Kibana для побудови користувацьких панелей відображення і візуалізації необхідних аспектів даних. Літ.: [12]	2
15	Особливості стеку технологій Elasticsearch. Застосування засобів Kibana у рішенні завдань Бі. Літ.: [12]	2
16	Системний підхід для побудови складних інформаційних систем. Роль бізнес-аналітика у сучасній компанії, що виконує розробку програмного забезпечення на замовлення (аутсорсинг). Літ.: [8]	2
17	Перспективи розвитку систем та технологій Бі. Місце та значення технологій Бі під час оброблення великих масивів даних. Літ.: [8]	2
	Разом за семестр:	17

Зміст лабораторних занять
Перелік лабораторних занять для студентів денної форми навчання

№ з/п	Тема лабораторного заняття	Кількість годин
Перший семестр		
1	Вступ до мови Scala та інтерфейс командного рядка Літ.: [1-3]	2
2	Вступ до PySpark. Основи машинного обчислення засобами PySpark Літ.: [4]	2
3	Порівняння характеристик сучасних рішень інструментальних засобів Бі. Літ.: [5]	2
4	Робота із середовищем Microsoft Power BI. Літ.: [5]	2
5	Моніторинг стану розподілених інформаційних систем на прикладі систем Nagios, Icinga та Zabbix. Літ.: [6]	2
6	Розробка онтології сервісів та видів фактів. Розгортання сервісів у хмарі. Літ.: [5]	2
7	Проектування архітектури та засобів Бі на основі відкритих рішень. Робота з технологією Elasticsearch та Kibana. Літ.: [5, 12]	2

8	Технології Grafana. Агрегація даних засобами БІ. Літ.: [6]	3
Разом за семестр:		17

Зарахування результатів неформальної освіти

Окремі результати вивчення курсу можуть бути зараховані у випадку отримання студентом результатів навчання у неформальній освіті, що підтверджені відповідним документом (сертифікат, свідоцтво, освітня програма тощо):

– Як результат виконання ЛР №3-4 зараховується онлайн курс Power BI Fundamentals (<https://app.datacamp.com/learn/skill-tracks/power-bi-fundamentals>)

– Як результат виконання ЛР №7 зараховується онлайн курс Streaming Data with AWS Kinesis and Lambda (<https://app.datacamp.com/learn/courses/streaming-data-with-aws-kinesis-and-lambda>)

Зміст самостійної (у т.ч. індивідуальної) роботи

Самостійна робота студентів денної форми навчання полягає у систематичному опрацюванні програмного матеріалу, підготовці до виконання і захисту лабораторних робіт, тестування з теоретичного матеріалу тощо. Керівництво самостійною роботою та контроль за виконанням індивідуального завдання здійснює викладач згідно з розкладом консультацій у позаурочний час.

Самостійна робота дисципліни «Методи та засоби аналізу великих та малих даних» також передбачає виконання Індивідуальної самостійної роботи (ІСР) до 2 індивідуальних завдань, що має на меті закріплення та розвиток вмій і навичок, набутих при виконанні циклу лабораторних робіт. Оцінка, яка виставляється за виконання ІСР, виводиться як середня з оцінювання таких елементів (опис критеріїв наведено вище):

- якість та рівень розв'язання задач роботи;
- якість наповнення й оформлення звіту.

Зміст самостійної роботи студентів денної форми навчання

Тема	Вид самостійної роботи <i>П'ятий семестр</i>	Кількість годин
1	Опрацювання теоретичного матеріалу, підготовка до виконання лабораторної роботи №1.	12
2	Опрацювання теоретичного матеріалу, підготовка до захисту лабораторної роботи №1 та виконання лабораторної роботи №2.	12
3	Опрацювання теоретичного матеріалу, підготовка до захисту лабораторної роботи №2 та виконання лабораторної роботи №3.	12
4	Опрацювання теоретичного матеріалу, підготовка до захисту лабораторної роботи №3 та виконання лабораторної роботи №4. Виконання завдання для самостійної роботи №1 - Робота із середовищем Tableau.	12
5	Опрацювання теоретичного матеріалу, підготовка до захисту лабораторної роботи №4 та виконання лабораторної роботи №5.	12
6	Опрацювання теоретичного матеріалу, підготовка до захисту лабораторної роботи №5 та виконання лабораторної роботи №6.	12
7	Опрацювання теоретичного матеріалу, підготовка до захисту лабораторної роботи №6 та виконання лабораторної роботи №7. Виконання завдання для самостійної роботи №2 - Обробка та аналіз потоків великих обсягів даних у реальному часі.	12
8	Опрацювання теоретичного матеріалу, підготовка до захисту лабораторної роботи №7 та виконання і захисту лабораторної роботи №8. Підготовка до підсумкового контролю.	15
Всього		99

Технології та методи навчання

Процес навчання з дисципліни ґрунтується на використанні традиційних та сучасних технологій, зокрема: лекції (з використанням методів візуалізації); лабораторні заняття (з використанням тренінгів, майстер-класів), самостійна робота (індивідуальні завдання).

Необхідні інструменти, обладнання, програмне забезпечення: комп'ютер (надається для використання в лабораторіях кафедри КН), Google Chrome (безкоштовне), Microsoft Power BI (безкоштовна версія), Tableau Public (безкоштовна версія) або інші безкоштовні рішення.

Методи контролю

Поточний контроль здійснюється під час лекційних та лабораторних занять, а також у дні проведення контрольних заходів, встановлених робочою програмою і графіком навчального процесу. При цьому використовуються такі методи поточного контролю:

- захист лабораторних робіт;
- презентація і захист індивідуальних завдань.

При виведенні підсумкової семестрової оцінки враховуються результати як поточного контролю, так і підсумкового контрольного заходу. Студент, який набрав позитивний середньозважений бал за поточну роботу і не здав підсумковий контрольний захід, вважається невстигаючим.

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». Кожний вид роботи з дисципліни оцінюється за інституційною чотирибальною шкалою і виставляється в електронному журналі обліку успішності.

Засвоєння студентом теоретичного матеріалу з дисципліни оцінюється під час захисту лабораторних робіт та підсумковому контрольному заході.

Поточний контроль передбачає захист лабораторних робіт. Поточний контроль здійснюється під час лабораторних занять, а також у дні проведення контрольних заходів, встановлених робочою програмою і графіком навчального процесу.

Політика проведення лабораторних робіт з навчальної дисципліни «Методи та засоби аналізу великих та малих даних»:

1. Студент має бути присутнім на кожному занятті.
2. Студент може бути відсутнім на занятті, якщо здав відповідну лабораторну роботу або з поважної причини: хвороба, тощо.
3. В разі потреби при виконанні лабораторної роботи студент може скористатися засобами зв'язку для пошуку інформації в інтернеті, тощо.
4. На одному занятті захищається тільки одна лабораторна робота. В разі бажання студента здати ще одну роботу, студент має стати в кінець черги. За одне заняття можна здати не більше 2 робіт.
5. Перескладання лабораторної роботи можливе на наступному занятті.
6. Можливе перезарахування деяких лабораторних робіт елементами неформальної освіти згідно Робочої програми дисципліни.

При виконанні лабораторних робіт студент має дотримуватись положень академічної доброчесності; використання напрацювань інших студентів (компоненти програмного забезпечення та програмні коди, графічні матеріали лабораторних робіт та фрагменти звітів з лабораторних робіт) є підставою для зниження оцінки чи перескладання лабораторної роботи.

Студент до виконання лабораторної роботи має бути підготовлений, зокрема: ознайомлений із темою, порядком виконання та відповідними теоретичними відомостями. Для встановлення рівня готовності студента до виконання роботи викладач здійснює опитування.

Пропущене лабораторне заняття студент зобов'язаний відпрацювати не пізніше, ніж за два тижні до кінця теоретичних занять у семестрі.

Семестрова підсумкова оцінка визначається як середньозважена з усіх видів навчальної

роботи, виконаних і зданих студентом позитивно, з урахуванням коефіцієнта вагомості і розраховується в автоматизованому режимі за відповідною програмою. Вагові коефіцієнти змінюються залежно від структури дисципліни і важливості окремих видів її робіт. При виведенні підсумкової семестрової оцінки враховуються результати як поточного контролю, так і підсумкового контрольного заходу, який проводиться письмово з усього матеріалу дисципліни. Студент, який набрав позитивний середньозважений бал за поточну роботу і не здав підсумковий контрольний захід, вважається невстигаючим.

Оцінювання знань здобувача вищої освіти здійснюється за такими критеріями:

Оцінка за інституційною шкалою	Узагальнений критерій оцінювання
Відмінно	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 державною мовою (в усній або у письмовій формі), демонструє якісне оформлення роботи і володіння спеціальними приладами та інструментами. Здобувач вищої освіти не вагається при видозміні запитання, вміє робити детальні та узагальнюючі висновки. При відповіді допустив дві–три несуттєві похибки.
Добре	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 несуттєві помилки.
Задовільно	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, здобувача вищої освіти має слабкі знання структури дисципліни, допускає неточності і суттєві помилки 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	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 здобувачу вищої освіти, який не може продовжити навчання без додаткової роботи з вивчення дисципліни.

Структурування дисципліни за видами робіт і оцінювання результатів навчання студентів денної форми навчання у семестрі за ваговими коефіцієнтами

Аудиторна робота								Самостійна, індивідуальна робота	Семестровий контроль, іспит
<i>1 семестр</i>									
Лабораторні роботи №:								Тематичний контроль:	Підсумковий контрольний захід
1	2	3	4	5	6	7	8		
ВК: 0,6								-	0,4

Умовні позначення: ВК – ваговий коефіцієнт.

Підсумкова семестрова оцінка за інституційною шкалою і шкалою ЄКТС встановлюється в автоматизованому режимі після внесення викладачем усіх оцінок до електронного журналу. Співвідношення інституційної шкали оцінювання і шкали оцінювання ЄКТС наведені у таблиці.

Співвідношення інституційної шкали оцінювання і шкали оцінювання ЄКТС

Оцінка ЄКТС	Інституційна інтервальна шкала балів	Інституційна оцінка, критерії оцінювання		
A	4,75–5,00	5	Зараховано	Відмінно – глибоке і повне опанування навчального матеріалу і виявлення відповідних умінь та навичок
B	4,25–4,74	4		Добре – повне знання навчального матеріалу з кількома незначними помилками
C	3,75–4,24	4		Добре – в загальному правильна відповідь з двома-трьома суттєвими помилками
D	3,25–3,74	3		Задовільно – неповне опанування програмного матеріалу, але достатнє для практичної діяльності за професією
E	3,00–3,24	3		Задовільно – неповне опанування програмного матеріалу, що задовольняє мінімальні критерії оцінювання
FX	2,00–2,99	2	Незараховано	Незадовільно – безсистемність одержаних знань і неможливість продовжити навчання без додаткових знань з дисципліни
F	0,00–1,99	2		Незадовільно – необхідна серйозна подальша робота і повторне вивчення дисципліни

Питання для самоконтролю результатів навчання

1. Поняття машинного навчання та використання його у Data Science.
2. Інструменти Python для машинного навчання.
3. Процес моделювання. Створення нових показників і вибір моделі.
4. Тренування моделі. Перевірка адекватності моделі. Прогнозування нових спостережень.
5. Типи машинного навчання.
6. Контрольоване та неконтрольоване навчання.
7. Частково контрольоване навчання.
8. Загальні методи обробки великих об'ємів даних. Правильний вибір алгоритма, структури даних, вибір інструментів.
9. Приклади використання технологій обробки великих масивів даних та їх вплив на розвиток компаній.
10. Основні концепції Map Reduce і Spark.
11. Інструменти Hive, Sqoop.
12. Вивчення інструменту обробки великих даних – Apache Spar.
13. Алгоритми розв'язку задач машинного навчання, застосування на практиці алгоритми Spark MLlib.
14. Аналіз та генерації зображень, відео, тексту, звуку та інших видів даних за допомогою таких інструментів, як Python, Keras і TensorFlow.
15. Розгортання кластерів HDInsight. Типи кластерів HDInsight.
16. Управління кластерами HDInsight. Управління кластерами HDInsight за допомогою Power Shell.
17. Трансформація великих даних, управління великими даними.
18. Паралельний аналіз операцій. Використання контексту обчислення
19. RxLocalParallel та rxExcel.

20. Управління наборами даних. Категоризація даних. Імпорт даних в Azure Machine Learning. Дослідження та перетворення даних у Azure Machine Learning.
21. Підготовка даних до використання в Azure Machine Learning. Попередня обробка даних. Обробка неповних наборів даних.
22. Використання функції інженерії та вибору.
23. Робочі процес Azure Machine Learning. Оцінка моделей.
24. Графові бази даних. Сфери застосування графових баз даних. Neo4j графова база даних. Мова запитів до графів Cypher.
25. Моделі асоціацій між сутностями та характеристиками для різних категорій NoSQL баз даних.
26. Використання простору даних для моделювання Великих даних.
27. Використання моделей Azure
28. Machine Learning. Розгортання та публікація моделей.

Методичне забезпечення

Навчальний процес з дисципліни «Методи та засоби аналізу великих та малих даних» повністю і в достатній кількості забезпечений необхідною навчально-методичною літературою.

Рекомендована література

Основна

1. Jugnesh Kumar, Anubhav Kumar, Rinku Kumar. Big Data and Analytics: The key concepts and practical applications of big data analytics (English Edition). BPB Publications, 2024. – 232 p
2. Feeney K. et al. Engineering Agile Big-Data Systems //Kevin Feeney, Jim Davies, James Welch, Sebastian Hellmann, Christian Dirschl, Andreas Koller, Pieter Francois, Arkadiusz Marciniak. — River Publishers, 2018. – 436 p.
3. Raheem N. Big Data: A Tutorial-Based Approach. Taylor & Francis Group LLC, CRC Press, 2019. — 203 p
4. Russo, Marco, и Alberto Ferrari. The Definitive Guide to DAX: Business Intelligence for Microsoft Power BI, SQL Server Analysis Services, and Excel Second Edition. 2nd edition, Microsoft Press, 2019.
5. Kelly, Nicholas. Delivering Data Analytics: A Step-By-Step Guide to Driving Adoption of Business Intelligence from Planning to Launch. 1st edition, Kogan Page, 2021.
6. Leskovec, Jure, и др. Mining of Massive Datasets. 3rd edition, Cambridge University Press, 2020.
7. Paul D. Business Analysis / D. Paul, J. Cadle, D. Yeates. – [4rd edition]. – BCS Learning & Development Publication, 2020. – 272 p.

Допоміжна

8. Wang, Shouhong, и Hai Wang. Business Database Technology: Theories and Design Process of Relational Databases, SQL, Introduction to OLAP, Overview of NoSQL Databases. 2nd edition, Universal Publishers, 2022.
9. Karambelkar, Hrishikesh Vijay. Apache Hadoop 3 Quick Start Guide: Learn about Big Data Processing and Analytics. 1st edition, Packt Publishing, 2018.

Інформаційні ресурси

1. Модульне середовище для навчання. Доступ до ресурсу: <https://msn.khnu.km.ua>.
2. Електронна бібліотека університету. Доступ до ресурсу: <http://library.khmnu.edu.ua/>
3. Репозитарій ХНУ. Доступ до ресурсу: <http://elar.khmnu.edu.ua/jspui/?locale=uk>